

AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Uğur ERCAN

VERİ MADENCİLİĞİ İLE HANEHALKI YAĞ TÜKETİMİNİN MODELLENMESİ

Ekonometri Ana Bilim Dalı
Doktora Tezi

Antalya, 2016

AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Uğur ERCAN

VERİ MADENCİLİĞİ İLE HANEHALKI YAĞ TÜKETİMİNİN MODELLENMESİ

Danışman

Doç. Dr. Sezgin IRMAK

Ekonometri Ana Bilim Dalı

Doktora Tezi

Antalya, 2016

T.C.
Akdeniz Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğüne,

Uğur ERCAN'ın bu çalışması, jürimiz tarafından Ekonometri Ana Bilim Dalı Doktora Programı tezi olarak kabul edilmiştir.

Başkan : Doç. Dr. Murat Alper BAŞARAN (İmza)

Üye (Danışmanı) : Doç. Dr. Sezgin IRMAK (İmza)

Üye : Doç. Dr. Mehmet MERT (İmza)

Üye : Yrd. Doç. Dr. Mustafa Koray ÇETİN (İmza)

Üye : Yrd. Doç. Dr. Mehmet Özer DEMİR (İmza)

Tez Başlığı: Veri Madenciliği ile Hanehalkı Yağ Tüketiminin Modellenmesi

Onay: Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Tez Savunma Tarihi : 18/07/2016

Mezuniyet Tarihi : 11/08/2016

(İmza)

Prof. Dr. Zekeriya KARADAVUT

Müdür

AKADEMİK BEYAN

Doktora Tezi olarak sunduğum “Veri Madenciliği ile Tahmin Edici Modelleme” adlı bu çalışmanın, akademik kural ve etik değerlere uygun bir biçimde tarafımdan yazıldığını, yararlandığım bütün eserlerin kaynakçada gösterildiğini ve çalışma içerisinde bu eserlere atıf yapıldığını belirtir; bunu şerefimle doğrularım.

Uğur ERCAN



İÇİNDEKİLER

ŞEKİLLER LİSTESİ	v
TABLOLAR LİSTESİ	vii
KISALTMALAR LİSTESİ	ix
ÖZET	xi
SUMMARY	xii
ÖNSÖZ	xiii
GİRİŞ.....	1

BİRİNCİ BÖLÜM

VERİ MADENCİLİĞİ

1.1	Veri Madenciliği Kavramı.....	4
1.2	Veri Madenciliği ve İstatistik	6
1.3	Veri Madenciliğinin Temelleri ve Gelişimi.....	7
1.4	Veri Madenciliği ve İlişkili Olduğu Alanlar.....	10
1.5	Veri Madenciliği Süreci.....	12
1.5.1	Veri Tabanlarında Bilgi Keşfi Süreci ile İlgili Metodolojiler	13
1.5.1.1	Fayyad vd. Göre Veri Tabanlarında Bilgi Keşfi Süreci Aşamaları.....	13
1.5.1.2	SEMMA Metodolojisi.....	14
1.5.1.3	Two Crows'a Göre Veri Madenciliği Süreci	15
1.5.1.4	CRISP-DM Metodolojisi	16
1.5.1.5	Cabena vd.'e Göre Veri Madenciliği Süreci	19

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ YÖNTEMLERİ

2.1	Veri Madenciliği Yöntemlerinin Sınıflandırılması.....	20
2.2	Karar Ağaçları	21
2.2.1	Karar Ağacı Algoritmaları.....	24
2.2.2	Karar Ağaçlarının Oluşturulması.....	25
2.2.3	Bölünmeleri Bulma	25
2.2.3.1	Bölünme Kriterleri	27
2.2.3.1.1	Safsızlık Tabanlı Kriterler.....	28
2.2.3.1.2	Normalize Edilmiş Safsızlık Tabanlı Kriterler	31
2.2.3.1.3	Uzaklık Ölçüsü	31
2.2.3.1.4	İkili Kriterler	32
2.2.4	Durdurma Kriterleri	34
2.2.5	Karar Ağaçlarının Budanması	35
2.2.5.1	Azaltılmış Hata Budaması.....	36

2.2.5.2	Maliyet Karmaşıklık Budama	37
2.2.5.3	Minimum Hata Budaması	37
2.2.5.4	Kötümser (Hata) Budama.....	38
2.2.5.5	Hata Tabanlı Budama.....	39
2.2.5.6	Kritik Değer Budama	40
2.2.5.7	Optimal Budama	40
2.2.5.8	Minimum Tanım Uzunluğu Budama	41
2.2.6	Karar Ağacı Endükleyicileri (Algoritmaları)	41
2.2.6.1	ID3, C4.5 ve C5.0 Algoritmaları.....	42
2.2.6.2	CART (C&RT) Algoritması.....	43
2.2.6.3	CHAID Algoritması	45
2.2.6.4	QUEST Algoritması.....	46
2.2.7	Karar Ağaçlarının Etkinliğinin Değerlendirilmesi	47
2.2.8	Karar Ağaçlarının Avantajları ve Dezavantajları	47
2.3	Birliktelik Kuralları	48
2.3.1	Temel Kavramlar ve Tanımlar	50
2.3.2	Birliktelik Kurallarında İlginçlik Ölçütleri.....	51
2.3.2.1	Kapsam Değeri	51
2.3.2.2	Destek Değeri	51
2.3.2.3	Güven Değeri	52
2.3.2.4	Kaldırma Oranı.....	52
2.3.2.5	Kaldıraç Oranı	53
2.3.3	Sık Öğe Kümelerinin Bulunması.....	53
2.3.3.1	Apriori Algoritması	54
2.3.3.2	FP – Growth Algoritması	55
2.4	Kümeleme Analizi.....	57
2.4.1	Benzerlik ve Uzaklık Ölçüleri	59
2.4.1.1	Benzerlik Ölçüleri	59
2.4.1.2	Uzaklık Ölçüleri	59
2.4.2	Kümeleme Yöntemlerinin Sınıflandırılması	61
2.4.2.1	Bölümlendirme Metotları.....	61
2.4.2.2	Hiyerarşik Modeller	65
2.4.2.3	Yoğunluk Tabanlı Metotlar.....	68
2.4.2.4	Izgara (Grid) Tabanlı Metotlar.....	68
2.4.2.5	Model Tabanlı Metotlar	68
2.5	Yapay Sinir Ağları.....	69
2.5.1	YSA'nın Özellikleri.....	72
2.5.2	YSA'nın Dezavantajları	73

2.5.3	Geleneksel Algoritmalar ile YSA'ların Karşılaştırılması	74
2.5.4	YSA'nın Kullanım Alanları.....	74
2.5.5	Yapay Nöron Modeli	75
2.5.6	YSA Mimarileri.....	78
2.5.6.1	İleri Beslemeli YSA	78
2.5.6.2	Yinelemeli YSA	80
2.5.7	YSA'da Öğrenme	83
2.5.7.1	YSA'da Öğrenme Stratejileri	84
2.5.7.2	Temel Öğrenme Kuralları	85
2.5.8	YSA'da Ağ Seçimi	86

ÜÇÜNCÜ BÖLÜM

HANEHALKI YAĞ TÜKETİMİ

3.1	Yağlar	88
3.1.1	Zeytinyağı.....	88
3.1.2	Margarin ve Diğer Bitkisel Katı Yağlar	91
3.1.3	Bitkisel Sıvı Yağlar	94
3.1.4	Tereyağı.....	96
3.2	Tüketim.....	100
3.2.1	Hanehalkı Tüketim Harcaması	101
3.2.1.1	Hanehalkı Tüketimine Etki Eden Faktörler	104
3.2.1.2	Hanehalkı Gıda Tüketimine Etki Eden Faktörler.....	105
3.2.1.2.1	Gelir	105
3.2.1.2.2	Hanehalkı Büyüklüğü ve Kompozisyonu, Kır-Kent Durumu	107
3.2.1.2.3	Eğitim ve Meslek	108
3.2.1.2.4	Yaş	109
3.2.1.2.5	Cinsiyet	110
3.3	Hanehalkı Gıda ve Yağ Tüketimi Konusunda Yapılmış Çalışmalar.....	111

DÖRDÜNCÜ BÖLÜM

HANEHALKI YAĞ TÜKETİMİNİN MODELLENMESİ

4.1	Uygulamanın Amacı.....	149
4.2	Uygulamada Kullanılan Veriler.....	149
4.2.1	Veri Ön Hazırlama Süreci	149
4.2.1.1	Fert Karakteristiği Önışleme ve Temizleme Aşaması.....	150
4.2.1.2	Tüketim Karakteristiği Önışleme ve Temizleme Aşaması	152
4.2.1.2.1	Yeni Değişkenlerin Oluşturulması.....	153
4.2.1.2.2	Enflasyondan Arındırma İşlemi	153

4.2.1.3	Hane Karakteristiklerine Ait Veri Önişleme ve Temizleme Aşaması	154
4.3	Uygulamada Kullanılan Yöntem ve Veri Madenciliği Yazılımı.....	158
4.4	Analiz Sonucunda Elde Edilen Bulgular	161
4.4.1	Hedef Değişkenlere Ait Temel İstatistikler	161
4.4.2	Zeytinyağı Tüketimini Etkileyen Faktörler	164
4.4.3	Yenilebilir Sıvı Yağ Tüketimini Etkileyen Faktörler	177
4.4.4	Margarin ve Diğer Bitkisel Yağ Tüketimini Etkileyen Faktörler.....	188
4.4.5	Tereyağı Tüketimini Etkileyen Faktörler	200
SONUÇ		211
KAYNAKÇA.....		217
EK 1-Enflasyondan Arındırma İşleminde Kullanılan Programın		
Arayüzü ve Kodları		234
EK 2-Zeytinyağı Karar Ağacı		237
EK 3-Yenilebilir Sıvı Yağ Karar Ağacı		243
EK 4-Margarin ve Diğer Bitkisel Yağlar Karar Ağacı		248
EK 5-Tereyağı Karar Ağacı.....		254
ÖZGEÇMİŞ		259

ŞEKİLLER LİSTESİ

Şekil 1.1 Veri Tabanı Endüstrisinin Gelişimi.....	8
Şekil 1.2 Veri Madenciliğinin İlişkili Olduğu Alanlar	11
Şekil 1.3 Veri Madenciliği İçin Kullanılan Metodolojiler	12
Şekil 1.4 Veri Tabanlarında Bilgi Keşfi Süreci	13
Şekil 1.5 Veri Tabanlarında Bilgi Keşfi Süreci	14
Şekil 1.6 Two Crows’a Göre Veri Madenciliği Süreci	15
Şekil 1.7 CRISP-DM’e Göre Veri Madenciliği Süreci	16
Şekil 1.8 Cabena vd. Göre Veri Madenciliği Süreci	19
Şekil 2.1 Veri Madenciliği Taksonomisi	20
Şekil 2.2 Karar Ağacı Örneği	23
Şekil 2.3 İyi ve Kötü Bölünmeler	26
Şekil 2.4 Kümeleme Analizi.....	57
Şekil 2.5 Kümeleme Algoritmalarının Sınıflandırılması.....	61
Şekil 2.6 Kümelerin K-Means Algoritması ile Oluşturulması	63
Şekil 2.7 Dendogram Örneği	66
Şekil 2.8 Agglomerative ve Divisive Dendogram Yapısı	68
Şekil 2.9 Sinir Hücresi (Nöron) Yapısı	71
Şekil 2.10 Yapay Nöron Modeli.....	75
Şekil 2.11 İleri Beslemeli Çok Katmanlı Sinir Ağı	79
Şekil 2.12 Yinelemeli Yapay Sinir Ağı	80
Şekil 2.13 Jordan Ağı	81
Şekil 2.14 Elman Ağı	82
Şekil 2.15 Hopfield Ağı.....	82
Şekil 4.1 IBM SPSS Modeller Ana Ekranı.....	159
Şekil 4.2 IBM SPSS Modeller Type Modülü	159
Şekil 4.3 Kurulmuş Bir Model Örneği	160
Şekil 4.4 IBM SPSS Modeller CHAID Karar Ağacı Çıktısı	160
Şekil 4.5 Zeytinyağı Karar Ağacı-1.....	166
Şekil 4.6 Zeytinyağı Karar Ağacı-2.....	167
Şekil 4.7 Zeytinyağı Karar Ağacı-3.....	168
Şekil 4.8 Yenilebilir Sıvı Yağ Karar Ağacı-1.....	179
Şekil 4.9 Yenilebilir Sıvı Yağ Karar Ağacı-2.....	180

Şekil 4.10 Yenilebilir Sıvı Yağ Karar Ağacı-3.....	181
Şekil 4.11 Yenilebilir Sıvı Yağ Karar Ağacı-4.....	181
Şekil 4.12 Margarin ve Diğer Bitkisel Yağlar Karar Ağacı-1	190
Şekil 4.13 Margarin ve Diğer Bitkisel Yağlar Karar Ağacı-2.....	191
Şekil 4.14 Margarin ve Diğer Bitkisel Yağlar Karar Ağacı-3.....	192
Şekil 4.15 Margarin ve Diğer Bitkisel Yağlar Karar Ağacı-4.....	193
Şekil 4.16 Tereyağı Karar Ağacı-1.....	201
Şekil 4.17 Tereyağı Karar Ağacı-2.....	202
Şekil 4.18 Tereyağı Karar Ağacı-3.....	203



TABLOLAR LİSTESİ

Tablo 1.1 İstatistiksel Analiz ve Veri Madenciliği Karşılaştırması.....	7
Tablo 1.2 Veri Madenciliği Değişiminin Adımları	10
Tablo 2.1 Karar Ağacı Endükleyicilerinin Genel Değerlendirilmesi	42
Tablo 2.2 Benzerlik Ölçütleri	59
Tablo 2.3 Uzaklık Ölçüleri	60
Tablo 2.4 K-Means Algoritmasında Kullanılan Küme Tahmini ve Popüler Başlangıç Yöntemleri.....	64
Tablo 2.5 Biyolojik Sinir Hücresi ile Yapay Sinir Hücresinin Karşılaştırılması	71
Tablo 2.6 Geleneksel Algoritmalar ile YSA'nın Karşılaştırılması	74
Tablo 2.7 Toplama Fonksiyonları	76
Tablo 2.8 Aktivasyon Fonksiyonları	77
Tablo 2.9 Temel Olarak Sinir Ağlarının Uygulama Alanları.....	86
Tablo 3.1 Dünya Zeytinyağı Üretim ve Tüketimi	90
Tablo 3.2 Türkiye Margarin Üretimi.....	93
Tablo 3.3 Türkiye Bitkisel Ham Yağ Üretimi.....	95
Tablo 3.4 Türkiye Bitkisel Likit Yağ İhracatı	95
Tablo 3.5 Dünya Bitkisel Ham Yağ Üretimi.....	95
Tablo 3.6 Türkiye Bitkisel Sıvı Yağ Tüketimi	96
Tablo 3.7 Bazı Ülkelerde Kişi Başına Tereyağı Tüketimi	97
Tablo 3.8 Türkiye'de Tereyağı Üretim, Tüketim Verileri.....	98
Tablo 3.9 Dünya Tereyağı Arz Kullanım ve Ticareti: 2009-2013	98
Tablo 4.1 Hanehalkı Bireylerinin Yaş Aralıkları	150
Tablo 4.2 Hanehalkı Reisinin Sağlık Sigorta Durumu	151
Tablo 4.3 Hanehalkı Reisinin Eğitim Durumu.....	152
Tablo 4.4 Hanede Kullanılan Birinci Yakıt Türü	154
Tablo 4.5 Çalışmanın Analizinde Kullanılan Karakteristikler	155
Tablo 4.6 Zeytinyağı Tüketimine Ait Tanımlayıcı İstatistikler.....	161
Tablo 4.7 Diğer Yenilebilir Sıvı Yağ Tüketimine Ait Temel İstatistikler.....	162
Tablo 4.8 Margarin ve Diğer Bitsiksel Yağ Tüketimine Ait Tanımlayıcı İstatistikler.....	163
Tablo 4.9 Tereyağı Tüketimine Ait Tanımlayıcı İstatistikler.....	163
Tablo 4.10 Gelir Grupları	164
Tablo 4.11 Zeytinyağı Tüketimini Etkileyen Karakteristikler	165

Tablo 4.12 Hanehalkı Zeytinyağı Tüketim Durumu	165
Tablo 4.13 Yenilebilir Sıvı Yağ Tüketimini Etkileyen Karakteristikler	178
Tablo 4.14 Hanehalkı Yenilebilir Sıvı Yağ Tüketim Durumu	178
Tablo 4.15 Margarin ve Diğer Bitkisel Yağ Tüketimini Etkileyen Karakteristikler.....	189
Tablo 4.16 Hanehalkı Margarin ve Diğer Bitkisel Yağ Tüketim Durumu.....	189
Tablo 4.17 Tereyağı Tüketimini Etkileyen Karakteristikler	200
Tablo 4.18 Hanehalkı Tereyağı Tüketim Durumu	200



KISALTMALAR LİSTESİ

AID	Automatic Interaction Detector
ART	Adaptive Resonance Theory
BDT	Binary Decision Tree
BYSD	Bitkisel Yağ Sanayicileri Derneği
CART (C&RT)	Classification and Regression Tree
CCP	Cost Complexity Pruning
CHAID	Chi-Square Automatic Interaction Detector
CLARA	Clustering Large Applications
CRISP-DM	Cross Industry Standard Process for Data Mining
CVP	Critical Value Pruning
DİE	Devlet İstatistik Enstitüsü
DIP	Depth Impurity Pruning
DPT	Devlet Planlama Teşkilatı
EBP	Error Based Pruning
FSP	Food Stamp Program
HHB	Hanehalkı Büyüklüğü
HHRCINS	Hanehalkı Reisinin Cinsiyeti
HHREGIT	Hanehalkı Reisinin Eğitim Durumu
HHRMEDDUR	Hanehalkı Reisinin Medeni Durumu
HHRSIG	Hanehalkı Reisinin Sigorta Durumu
HHRYAS	Hanehalkı Reisinin Yaşı
HHTIPI	Hanehalkı Tipi
KKD	Knowledge Discovery in Databases
LVQ	Learning Vector Quantization Network
MDL	Minimum Tanım Uzunluğu Budama
MEB	Milli Eğitim Bakanlığı
MEP	Minimum Error Pruning
MML	Minimum Message Length
MusKonut	Müstakil Konut Sahipliği
MÜMSAD	Mutfak Ürünleri ve Margarin Sanayicileri Derneği
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing

OP	Optimal Pruning
PAM	Partitioning Around Methods
PEP	Pessimistic Pruning
QUEST	Quick, Unbiased, Efficient Statistical Tree
REP	Reduced Error Pruning
SGK	Sosyal Güvenlik Kurumu
SSE	Sum of Squared Errors
SSK	Sosyal Sigortalar Kurumu
SOM	Self Organizing Map
TEAE	Tarımsal Ekonomi Araştırma Enstitüsü
TEPGE	Tarımsal Ekonomi ve Politika Geliştirme Enstitüsü
TÜİK	Türkiye İstatistik Kurumu
USK	Ulusal Süt Konseyi
VBP	Validation Based Pruning
VTKB	Veri Tabanlarında Bilgi Keşfi
VM	Veri Madenciliği
YSA	Yapay Sinir Ağı

ÖZET

Hanehalkına ait hane, fert ve tüketim karakteristiklerini içeren “Hanehalkı Bütçe Anketi” veri seti ile yapılmış birçok çalışma mevcuttur. Bu çalışmanın amacı TÜİK tarafından derlenen Hanehalkı Bütçe Anketi verileri kullanılarak, hanehalkının tüketmiş olduğu zeytinyağı, tereyağı, margarin ve diğer bitkisel yağlar ile diğer yenilebilir sıvı yağ tüketimini etkileyen hanehalkı karakteristiklerine ilişkin faktörlerin ve hanehalklarının tüketim durumlarının Karar Ağacı yöntemi ile belirlenmesidir.

Çalışmanın birinci bölümünde veri madenciliği kavramı ve bilgi keşfi sürecinde kullanılan metodolojiler özetlenmiştir, ikinci bölümünde veri madenciliği modelleme çeşitleri ile ilgili bilgiler verilmiştir. Üçüncü bölümde genel olarak katı ve sıvı yağlar ile ilgili bilgilere yer verilmiştir. Ayrıca bu bölümde hanehalkı genel tüketimi, gıda tüketimi ve yağ tüketimi harcamalarını etkileyen faktörler incelenmiş ve bu alanda daha önce yapılmış çalışmalar detaylı bir şekilde incelenmiştir. Çalışmanın uygulama kısmında ilk önce 2009-2012 yılları arasında TÜİK tarafından derlenen hanehalkı bütçe anketi verilerinin veri birleştirme, ön işleme aşamaları ile gelir ve harcama verilerinin enflasyondan arındırma işlemlerine yer verilmiştir. Buradan elde edilen veri setleri ile veri madenciliği yöntemlerinden kümeleme analizi ve karar ağacı yöntemleri kullanılarak Türkiye hanehalkının belirlenen yağ tüketimi kategorisinde tüketim profili çıkarılmıştır.

Yapılan çalışma sonucunda hanehalkının tüketmiş olduğu dört farklı yağ türünün tüketimini etkileyen faktörler karar ağacı yöntemi ile belirlenmiştir. Yağ türlerinin tüketimini etkileyen faktörler genel olarak birbirine benzetmekle beraber, hangi yağ türünü hangi karakteristiğin hangi şekilde etkilediği farklılıklar göstermektedir. Meydana gelen karar ağacında bu detaylar görülmektedir. Zeytinyağı ve tereyağı gibi sağlıklı, besin değeri yüksek ve pahalı olan yağ türlerini etkileyen en önemli karakteristiğin gelir olduğu görülürken, margarin ve diğer bitkisel yağlar ile diğer yenilebilir sıvı yağ gibi daha ucuz yağ türlerinin tüketimini etkileyen en önemli karakteristiğin hanehalkı büyüklüğü olduğu görülmüştür. Her bir yağ türünde meydana gelen karar ağacının alt düğümlerine ve detaylarına inildiğinde tüketimi etkileyen faktörlerin değiştiği görülmüş ve bunların detayları çalışmada incelenmiştir.

Anahtar Kelimeler: Hanehalkı Bütçe Anketi, Hanehalkı Yağ Tüketimi, Veri Madenciliği, Karar Ağaçları

SUMMARY

MODELING HOUSEHOLD OIL CONSUMPTION BY USING DATA MINING

There are a lot of researchs made with HSB data set which contains characteristics of household and personal consumption of households. The aim of this research is to identify the factors on household characteristics which affect the consumption of olive oil, butter, margarine, other vegetable oils and consumable oils consumed by households and consumption situations of households with decision tree method by using HSB datas compiled by TÜİK.

In the first part of the research, data mining concept and methodologies used in knowledge discovery are summarized. In the second part, informations about data mining modelling type; in the third part, generally the informations about fats and oils are given. And also, in this part, the factors affect the general household consumption, food consumption and the oil consumption expenditures are examined, and the studies in this field are scrutinized.

In the execution part of the research, databinding and preprocessing phases of HBS datas compiled by TÜİK between the years 2009-2012 and net of income and expenditure datas of inflation were included first. With the data sets obtained, by using the clustering analysis and decision tree method which are data mining methods, the consumption profile in determined oil consumption cathegory of Turkey household is profiled.

In consequence of the research, factors affect the consumption of four different kinds of oil consumed by the household are determined via decision-tree method. Despite the factors affect the consumption of fat kinds usually look alike, it shows differences on which fat kind is affected by which characteristic in which way. In the decision tree, those details are seen. While the most important characteristic on consumption of healthy, nutritious and expensive oils like olive oil and butter is the income, it's the size of the household on cheaper oils like margarines, other vegetable and consumable oils. When looking down the nodes and the details of a decision tree took place on each kinds of oil, it is seen that the factors which affect the consumption and the characteristics are changed, the details of these factors are examined in the research.

Keywords: Household Budget Survey, Household Oil Consumption, Data Mining, Decision Trees

ÖNSÖZ

Akademik eğitim hayatının en önemli süreci olan doktora aşamasında desteğini her zaman yanımda hissettiğim, insanlığı ve akademisyenliği ile saygı duyduğum değerli hocam, tez danışmanım Doç. Dr. Sezgin IRMAK’a teşekkürlerimi sunuyorum.

Tez süreci boyunca, tez izleme komitelerinde olumlu tavırları ve yapıcı eleştirileri ile tez çalışmamın ilerlemesine ve gelişim göstermesine katkı sağlayan değerli hocalarım Doç. Dr. Mehmet MERT ve Yrd. Doç. Dr. Koray ÇETİN’e teşekkürlerimi sunuyorum. Doktora tez süreci boyunca verdiği desteklerden dolayı TÜİK Uzmanı Turaç YAVUZ’a teşekkürlerimi sunuyorum.

Doktora tez çalışmamı finansal olarak destekleyen Akdeniz Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimine teşekkürlerimi sunuyorum. Tez çalışmasında kullanılan “Hanehalkı Bütçe Anketi” araştırması verilerini aldığım Türkiye İstatistik Kurumuna (TÜİK) teşekkürlerimi sunuyorum.

Akademisyenlik hayatına adım atmamda bana önder olan, desteğini her zaman yanımda hissettiğim ve beni her zaman cesaretlendiren değerli aileme teşekkürlerimi sunuyorum.

Uğur ERCAN

Antalya, 2016

GİRİŞ

20. yüzyılda diyot, transistor, tristör gibi elektronik devre elemanlarının keşfi ile başlayan süreç, elektronik uygulamalarla büyük gelişim göstermiştir. Bilimde yaşanan ilerlemelerin sonucunda bilgisayar ve birçok elektronik alet keşfedilerek insanlığın kullanımına sunulurken, bilgisayar teknolojisinde meydana gelen değişimler güçlü ve uygun fiyatlı bilgisayarların, veri toplama ekipmanlarının ve depolama ortamlarının geliştirilmesine yol açmıştır. Gerçekleştirilen her işlemin ileride kullanılmak amacıyla bir veri olarak saklanması zaman içerisinde zorunlu hale gelirken, veri toplama yöntemlerinin ve veri depolama ortamlarının gelişmesi, yapılan işlemlerin birer veri ya da bilgi kaydı olarak veri tabanlarına daha kolay ve hızlı olarak depolanmasına imkân sağlamıştır. Depolanan bu veriler zaman içerisinde büyük boyutlara ulaşmıştır.

Veri tabanları içerisinde biriken ve saklanan veri yığınları içerisinde geleneksel yöntemlerle görülemeyen, önemli, anlamlı ilişkiler ve örüntüler içermektedir. Kayıtlı bu verilerden daha kesin ve doğru bir şekilde tahmin, sınıflandırma ve/veya betimleme yapabilmek için yeni bir analiz alanı ihtiyacı doğmuştur. Bu ihtiyaçları gidermek için bilgisayar, veri tabanı, istatistik, makine öğrenmesi, matematik, yapay zekâ, veri görselleştirme gibi birçok bilim alanının ve tekniğin birleşimiyle bir metodoloji geliştirilmiştir, bu “Veri Madenciliği”dir.

Veri tabanlarında biriken büyük miktarlardaki verilerin ne tür ilişkiler içerdiği, ne anlama geldiği ile ilgili düşünceler ve “Veri Tabanlarında Bilgi Keşfi” kavramı 1989 yılında gerçekleştirilen Veri Tabanlarında Bilgi Keşfi Çalıştayında ortaya çıkarılmıştır (Fayyad vd., 1996: 39). Çoğu araştırmacı tarafından VTBK süreci ile veri madenciliği terimlerinin eş anlamlı olarak da kullanılmasına rağmen Veri Madenciliği (Data Mining), Veri Tabanlarında Bilgi Keşfi sürecinin en önemli adımını oluşturmaktadır (Akpınar, 2000: 2). Birçok insan tarafından ortak olarak kullanılan Veri Madenciliği ya da Bilgi Madenciliği tanımı basitçe, büyük miktardaki verilerden bilginin ayıklanması anlamına gelir (Han ve Kamber, 2006: 5). Literatürde, verilerden bilgi madenciliği, bilgi çıkarma, veri örüntü analizi, veri arkeolojisi, veri tarama ve bilgi hasatı gibi kavramlarla ifade edilebilmektedir (Han ve Kamber, 2006: 5, Fayyad vd., 1996: 39).

Veri tabanları, veri yığınları veya herhangi bir ortamda biriktirilen veriler ham hali sadece birer veri kayıdır, dolayısıyla veriler ham hali ile büyük bir anlam ifade etmezler. Veriler belirli amaçlar doğrultusunda işlenip kullanıldığında bir anlam ifade etmeye başlar. İşlenen verilerden elde edilen bilgiler işletme kârının yükseltilmesinde, ekonomik kaynakların

yönetilmesinde ve daha karlı bir şekilde değerlendirilmesinde, sağlık alanında insanlığa hizmet etmek gibi birçok alanda kullanılmaktadır. Bu nedenle veri madenciliği ve veri madenciliği teknikleri sağlık, mühendislik, ekonomi, eğitim, işletme gibi birçok alanda uygulanmıştır ve uygulanmaya devam etmektedir. Veri Madenciliği alanının bu kadar popüler ve tercih edilir olmasının en önemli nedenlerinden bazıları; istatistikî tekniklerden elde edilen sonuçlara göre daha iyi tahmin yüzdesine ve doğruluk derecesine sahip olması, kullanılan yöntemlerin çeşitliliği ve bu yöntemlerin belirli uygulamalardan bağımsız olarak birçok alandaki verilere kullanılabilmesi örnek gösterilebilir. Veri madenciliği yöntemleri eldeki mevcut veriler ile geleceğe yönelik tahmin yapmada, verilerin betimlenmesinde ve doğrulanmasında, veriler arasındaki ilişkilerin ve büyüklüklerinin belirlenmesinde kullanılabilir.

TÜİK tarafından ilk olarak 1954 yılında derlen hanehalkı bütçe anketi verileri, 2002 yılından itibaren düzenli olarak derlenmektedir. Bu veriler ile Tüketici fiyat indekslerinde kullanılacak maddelerin seçimi ve ağırlıklarının elde edilmesi, hanelerin tüketim kalıplarında zaman içinde meydana gelen değişikliklerin izlenmesi, milli gelir hesaplamalarında özel nihai tüketim harcamaları tahminlerine yardımcı olacak verilerin derlenmesi, Yoksulluk sınırının belirlenmesi vb. diğer sosyoekonomik göstergelerin elde edilmesi, asgari ücret tespit çalışmaları için gerekli verilerin derlenmesi amaçlanmaktadır (TÜİK, 2011a).

Hanehalkının sahip olduğu sosyoekonomik, demografik, fert karakteristikleri kullanarak herhangi bir tüketim karakteristiğinin tüketimini etkileyen faktörlerin belirlenmesi için yapılan analizlerin çoğu ekonometrik modeller ve istatistikî yöntemler kullanarak oluşturulmuş ve çözülmüştür.

Sınıflandırma amacıyla veri madenciliği sürecinin kullanıldığı çalışma dört ana bölümden oluşmaktadır.

Birinci bölümde veri madenciliği kavramı, gelişim süreci, ilişkili olduğu alanlar, veri madenciliği süreci ve bilgi keşfi sürecinde kullanılan metodolojiler anlatılmıştır.

İkinci bölümde en önemli veri madenciliği yöntemlerinden olan karar ağaçları, birliktelik kuralları, kümeleme analizi, yapay sinir ağları alt başlıklar halinde detaylı olarak incelenmiştir. Tezin uygulama kısmının yöntem bölümünü içeren “Karar Ağaçları” kısmı detaylı bir şekilde incelenmiştir.

Üçüncü bölümde yağ tüketim istatistikleri, hanehalkı toplam tüketimini, gıda ve yağ tüketimini etkileyen faktörler incelenmiştir. Ayrıca hanehalkı gıda ve yağ tüketimine ait geçmiş çalışmalar ayrıntılı olarak incelenmiş ve açıklanmıştır.

Tezin uygulamasının yer aldığı dördüncü bölümde ise uygulamanın amacı, uygulamada kullanılan veriler ve bu verilerin ön işleme aşamaları, analizde kullanılan yöntem ve veri madenciliği yazılımı ve analiz sonucunda elde edilen bulgulara yer verilmiştir. Uygulamada dört amaç belirlenmiştir bunlar; zeytinyağı, margarin ve diğer bitkisel yağlar, tereyağı ve yenilebilir sıvı yağ tüketimini etkileyen faktörlerin ve hanehalklarının tüketim durumlarının karar ağacı yöntemi ile belirlenmesidir. Her bir yağ grubu için tüketimi etkileyen faktörler belirlenmiş ve hangi koşullara göre değişim gösterdikleri belirtilmiştir. Sonuç değerlendirmesi ve öneriler kısmı ile tezin dördüncü bölümü tamamlanmıştır.



BİRİNCİ BÖLÜM

VERİ MADENCİLİĞİ

1.1 Veri Madenciliği Kavramı

Birçok insan tarafından ortak olarak kullanılan veri madenciliği ya da bilgi madenciliği tanımı basitçe, büyük miktardaki verilerden bilginin ayıklanması anlamına gelir (Han ve Kamber, 2006: 5). Literatürde, verilerden bilgi madenciliği, bilgi çıkarma, veri örüntü analizi, veri arkeolojisi, veri tarama ve bilgi hasatı gibi kavramlarla ifade edilebilmektedir (Han ve Kamber, 2006: 5, Fayyad vd., 1996: 39).

Veri madenciliğinin tanımı birçok araştırmacı, akademisyen ve kuruluş tarafından farklı şekillerde yapılmıştır. Bu tanımlarına bakacak olursak;

- Veri madenciliği büyük veritabanlarında daha önceden bilinmeyen desenlerin keşfidir (IBM, 1996: 1).
- Veri madenciliği, büyük veri tabanlarından bilgiler çıkarmak için makine öğrenmesi, örüntü tanıma, istatistik, veri tabanları ve görselleştirme tekniklerini bir araya getiren disiplinler arası bir alandır (Cabena vd., 1998: 12).
- Veri madenciliği, geçerli tahminler yapmak için kullanılabilir veri desenleri ve veri ilişkilerini keşfetmek için çeşitli araçlar kullanan bir veri analiz sürecidir (Two Crows Corporation, 1999: 1).
- Veri madenciliği, umulmadık ilişkileri bulmak ve veri sahibine hem yararlı hem de anlaşılabilir yeni bir yolla özetlemek için görsel veri setlerinin analizidir (Hand vd., 2001: 1).
- Veri desenlerini çıkarmak için akıllı yöntemlerin uygulandığı bir süreçtir (Zhang ve Zhang, 2002: 5).
- Veri madenciliği, veri tabanlarında saklı olan bilgiyi ortaya çıkarmak ve yapıyı açıklamak için çok sayıda veri analizi aracını kullanan bir süreçtir (Oğuzlar, 2003: 67).
- Veri madenciliği önemli iş kararları almak için geçerli, faydalı ve anlaşılabilir desen veya modelleri belirleme süreci olarak tanımlanmaktadır (Kuonen, 2014).
- Veri madenciliği belirli bir veri koleksiyonundan çeşitli modeller, özetler ve türetilmiş değerleri keşfetme sürecidir (Kantardzic, 2011: 6).
- Veri madenciliği değişkenler arasındaki ilginç, genellikle beklenmedik, önceden bilinmeyen ilişkileri ve yapıları ortaya çıkarmak için tasarlanmıştır (Zhao ve Luan, 2006: 9).

- Verilerde önceden bilinmeyen anlamlı ve değerli bilgiler, örüntüler elde etme işlemidir (Yıldırım vd., 2008: 430).
- Veri madenciliği, büyük hacimdeki verilerin içerisinde yer alan, gizli ve önemli örüntülerin, desenlerin, bilgilerin ve ilişkilerin açığa çıkarılıp stratejik karar destek amaçlı kullanımudur (Koyuncugil ve Özgülbaş, 2009: 21).
- Veri madenciliği, veritabanlarında veya veri ambarlarında bulunan veriler arasında var olan, bilinmeyen, istatistik gibi klasik yöntemlerle görülemeyen ve sıradan olmayan ilişkileri, örüntüleri, yapıları veya eğilimleri ortaya çıkarmak amacıyla istatistik, matematik, makine öğrenimi ve bilgisayar uygulamaları alanlarının birleşimi tekniklerin kullanılarak analiz edilmesi, sonuçların anlamlı bir şekilde özetlenmesi ve görselleştirilmesi sürecidir (Irmak, 2009: 5).
- Veri tabanlarında saklanan büyük miktardaki verilerden anlamlı veri korelasyonlarını, desenlerini ve eğilimlerini örüntü tanıma teknolojileri, istatistiksel ve matematiksel teknikler kullanarak ortaya çıkaran keşfetme sürecidir (Gartner Group, 2014).

Veri madenciliği aşağıdaki durumlarda kullanılabilir (Brusilovsky, 2015);

- İş problemi yapılandırılmadığında,
- Doğru ve kesin tahmin, açıklamadan daha önemli olduğunda,
- Aralık, nominal, ordinal, sayı ve metin değişkenlerinin karışımı ve sayısal olmayan değişkenlerin sayısını içeren veri gerekli olduğunda,
- Değişkenler arasında ilgisiz, alakasız ve gereksiz öznitelikler var olduğunda,
- Değişkenler arasındaki ilişkiler, karakterize edilemeyen doğrusalsızlıklarıyla doğrusal olmadığında,
- Veri heterojen olduğunda, (aykırı değerlerin, kaldıraç noktalarının (leverage points) ve kayıp değerlerin büyük bir yüzdesi ile)
- Örneklem boyutu oldukça büyük olduğunda

Veri madenciliğinin ne olduğu ya da ne olmadığı sorusuna aşağıdaki cevaplar verilebilir (Two Crows Corporation, 1999: 1).

- Veri madenciliği bir araçtır, sihirli bir değnek değildir bu nedenle üzerinde çalışılan konuyu bilme zorunluluğunu ortadan kaldırmaz. Yani, bu verileri yorumlamak ve anlamak yine kendi işiniz ile ilgilidir.
- Veri madenciliği, verinin içindeki desenlerin ve ilişkilerinin bulunmasında iş analistlerine yardımcı olmaktadır. Organizasyon için desenin değerini anlatmak veri madenciliğinin görevi değildir.

- Veri madenciliği ile ortaya çıkarılan desenlerin ve ilişkilerin mutlaka gerçek dünyada doğrulanması gerekmektedir.
- Veri madenciliğinde tercih edilen yöntemler ve optimizasyon teknikleri modelin doğruluğunu ve hızını etkileyecektir.
- Veri madenciliğinin diğer bir amacı, vasıflı iş analistlerini veya yöneticilerini değiştirmek değil onlara yaptıkları işi geliştirmek için yeni ve güçlü bir araç vermektir.

1.2 Veri Madenciliği ve İstatistik

Fayyad vd. İstatistik ile Veri Tabanlarında Bilgi Keşfi arasında birçok ortak nokta bulunduğunu bildirmişlerdir (Fayyad vd., 1996: 40). Aslında istatistik, veri madenciliğinin yapı taşlarından biridir ve veri madenciliğinin çekirdeğini oluşturmaktadır. Veri madenciliği ise istatistiğin bir uzantısı ve genişlemesi olarak görülebilir (Zhao ve Luan, 2006: 10; Kdnuggets, 2015). Bu nedenle veri madenciliği birçok yönden istatistikle yakın ilişkilidir. Hem istatistik hem de veri madenciliği, veri setinin deşifre edilmesi ve analizi ile ilgili alanlardır. Her iki yöntem de belirsizliklerle mücadele etmek ve gelecekteki olaylarla ilgili bilgi vermek, bir olayı etkileyen önemli faktörleri belirlemek ve türetilen modeli uygulayarak gelecekteki olayları daha iyi tahmin etmek için geliştirilmiş araçlardır (Zhao ve Luan, 2006: 10).

Veri Madenciliği, istatistik biliminin teknolojiyle bütünleşmesi sonucu meydana gelen bir araçtır. İstatistiksel açıdan bakıldığında, veri madenciliği büyük ve karmaşık veri setlerinin bilgisayar ile otomatik olarak keşfi ve analizi olarak görülebilir. Her ne kadar kullanıcı için kolaylık sağlayıcı ara yüzler hazırlanmış olsa bile programı sağlıklı kullanmak ve isabetli kararlar alabilmek açısından en azından temel istatistik bilgisi ihtiyacı önemlidir (Ganesh, 2002: 1; Özmen, 2001: 2-3).

Veri madenciliği ile istatistik arasındaki ortak noktalar ve farklılıklar bazı kişiler ve kurumlar tarafından belirtilmiştir, bu belirden ifadeler şu şekilde sıralanabilir;

- İstatistiksel analiz sırasında kullanılan veri, genellikle ilgili çalışmayı gerçekleştirmek için toplanmıştır. Verinin nasıl ve ne kadar toplanacağı istatistik çalışmasının bir parçasıdır. İstatistikçiler, nihai amaca ulaşmak için yeterli olacak en az sayıda verinin toplanmasına verimlilik açısından dikkat ederler ve özen gösterirler, veri madenciliğinde ise veri önceden veri madenciliğinden başka bir amaç için toplanmıştır (Esenkar, 2014).
- Veri madenciliğinde önemli bir nokta verinin temizlenmesi, birleştirilmesi gibi ön işlemlere tabi tutulmasıdır (Tüzüntürk, 2010: 73).

- Veri madenciliği geleneksel istatistikten farklıdır, istatistik çıkarsama odaklıdır yani hipotez oluşturulur ve verilere karşı doğrulanır. Aksine Veri madenciliği, keşif odaklıdır, desen ve hipotezler otomatik olarak veriden elde edilir (Zhang ve Zhang, 2002: 2).
- Zhao ve Luan'a göre veri madenciliği ile istatistiğin farklı yönleri; teoremin rolü, genellenebilirlik, hipotez testi ve güven düzeyi olarak belirtmişlerdir (Zhao ve Luan, 2006: 10-13).

Yukarıda belirtilen ortak noktalar ve farklılıklar ayrı ayrı belirtilmiş olmasına karşın belirtilen ifadeler aslında iki yöntemin karşılaştırılmasıdır. Yukarıda belirtilen karşılaştırmaların yanı sıra Moss ve Atre'nin yapmış olduğu karşılaştırma Tablo 1.1'de gösterilmiştir (Moss ve Atre, 2003: 381).

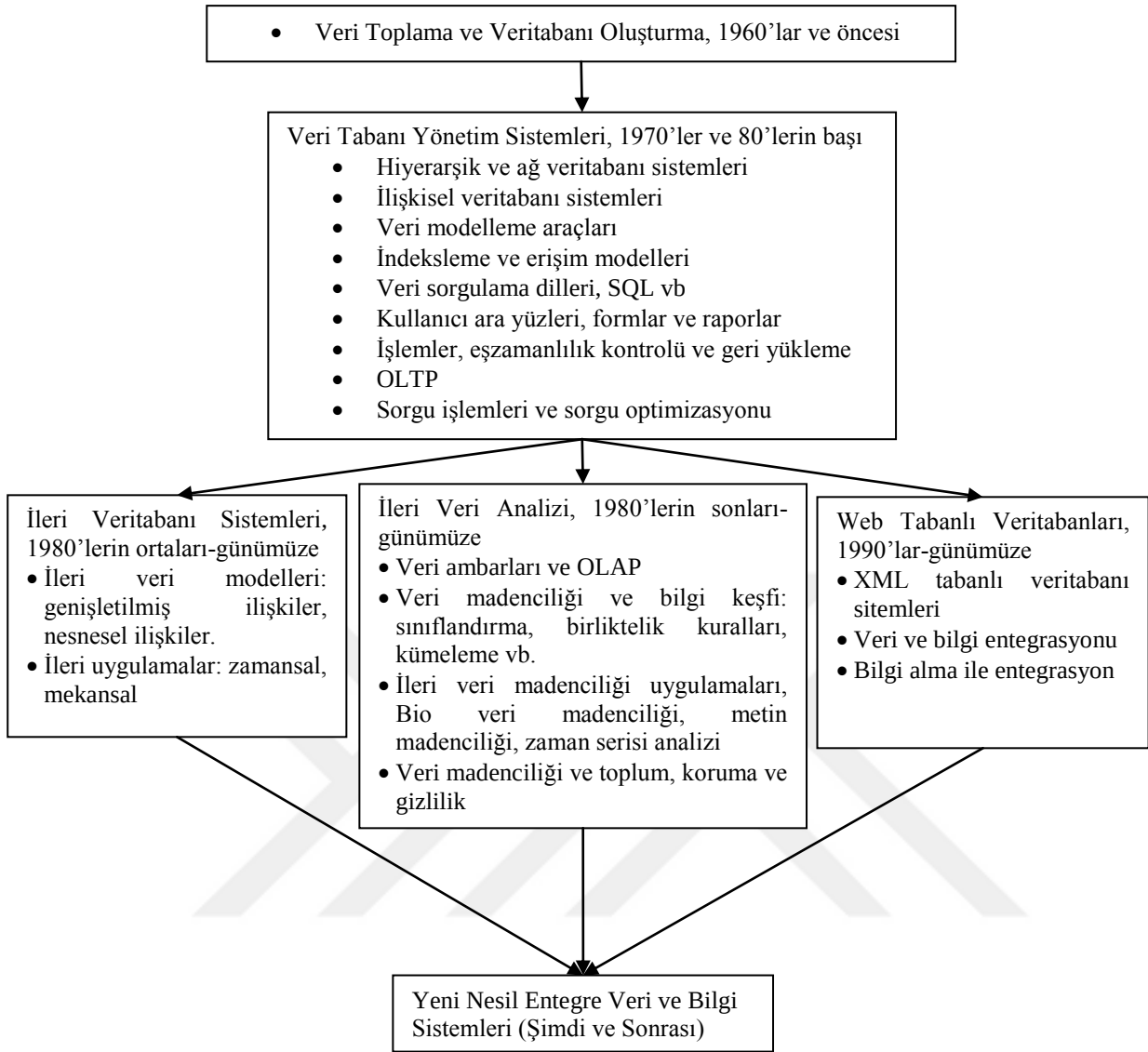
Tablo 1.1 İstatistiksel Analiz ve Veri Madenciliği Karşılaştırması

	İstatistiksel Analiz	Veri Madenciliği
1	İstatistikçiler genellikle bir hipotezle başlarlar.	Veri madenciliği bir hipoteze gerek duymaz.
2	İstatistikçiler kendi hipotezlerini eşleştirmek için kendi denklemlerini geliştirmek zorundadırlar.	Veri madenciliği algoritmaları denklemleri otomatik olarak geliştirir.
3	İstatistiksel analiz sadece sayısal veri kullanır.	Veri madenciliği sayısal verilerin haricinde başka veri tipleri de kullanır. (Örn; ses, metin,...)
4	İstatistikçiler, analizleri sırasında kirli veriyi bulur ve filtrelerler.	Veri madenciliği temiz ve iyi belgelenmiş verilere dayanır.
5	İstatistikçiler kendi sonuçlarını yorumlar ve bu sonuçları müdürlere ve şirket yöneticilerine iletirler.	Veri madenciliği sonuçlarını yorumlamak kolay değildir. Veri madenciliği sonuçlarını analiz etmek ve sonuçlarını aktarmak için mutlaka bir istatistikçi süreçte yer almalıdır.

Kaynak: Moss ve Atre, 2003: 381

1.3 Veri Madenciliğinin Temelleri ve Gelişimi

Veri madenciliği, bilgi teknolojisinin doğal evriminin sonucu olarak görülebilir. Veri tabanı endüstrisinin gelişimini Şekil 1.1'deki gibi özetleyebiliriz.



Şekil 1.1 Veri Tabanı Endüstrisinin Gelişimi

Kaynak: Han ve Kamber, 2006: 2

Şekil 1.1'e göre, 1960'lar ve öncesinde yapılan veri toplama ve ilkel veri tabanı oluşturma işlemleri veri madenciliğinin temellerini oluşturmuştur. 1970'ler ilkel hiyerarşik ve ağ veri tabanından ilişkisel veri tabanı (veri ilişkisel tablo yapılarında saklanır) sistemlerinin geliştirilmesi ilerlemiştir. Bu gelişmelere ilave olarak sorgu dilleri, kullanıcı ara yüzleri, optimize sorgu işleme, hareket (işlem-transaction) yönetimi araçları geliştirilmeye başlanmıştır. Bu şekilde kullanıcılar veriye esnek ve rahat bir şekilde erişim sağlamışlardır. OLTP (çevrimiçi işlem gerçekleştirme/online transaction processing) verimli depolama, geri alma ve büyük miktarda veri yönetimi için önemli bir araç olarak ilişkisel teknolojinin geniş kabulü ve değişimi için önemli ölçüde katkıda bulunmuştur. 1980'lerin ortalarından itibaren ilişkisel teknolojinin benimsenmesi, yeni ve güçlü veritabanı sistemleri geliştirme ve araştırma faaliyetlerinin artış göstermesi ile birlikte veritabanı teknolojisi karakterize

edilmiştir. Bunlar genişletilmiş-ilişkisel, nesne yönelimli, nesne-ilişkisel ve tündengelim modelleri gibi gelişmiş veri modellerinin gelişimini teşvik etmiştir. Mekânsal, zamansal, multimedya, aktif, akış ve sensör, bilimsel ve mühendislik veri tabanları, bilgi tabanları ve ofis bilgi üsleri de dâhil olmak üzere uygulamaya yönelik veri tabanı sistemleri geliştirilmiştir. Geçmiş otuz yılda bilgisayar teknolojisinde meydana gelen büyük değişmeler güçlü ve uygun fiyatlı bilgisayarların geliştirilmesine, veri toplama ekipmanlarına ve depolama ortamlarının geliştirilmesine yol açmıştır. Bu teknoloji, veritabanı ve bilişim sektörü, veritabanları ve işlem yönetimi, bilgi alma ve veri analizi için mevcut bilgi depoları ve büyük veri tabanları için büyük destek sağlamıştır. Karar verme ve yönetimi kolaylaştırmak için, tek bir sitede birleşik bir şema altında heterojen verilerden oluşan veri havuzu mimarisi, veri ambarları kavramı ortaya çıkmıştır. Veri ambarı teknolojisi, veri temizleme, veri entegrasyonu, ve OLAP (çevrimiçi analitik işleme/online analytical processing) içermektedir (Han ve Kamber, 2006: 1-3). OLAP, veriler üzerinde analizlerin yapılmasını sağlayan bir sistemdir. Bu yaklaşıma göre veri ambarları üstünden sorgular çekerek istenilen analizlere ulaşılması sağlanılır. İlişkisel veritabanlarına oranla daha hızlı cevapların alınması mümkündür, çünkü veriler ilişkilerinden arındırılmış bir şekilde tutulmaktadır (Cebesoy, 2014). OLAP araçları çok boyutlu analiz ve karar vermeyi desteklemesine rağmen derinlemesine analiz için veri sınıflandırma, kümeleme ve veri karakterizasyonu gibi ek veri analiz araçları gereklidir (Han ve Kamber, 2006: 3).

Yapay zekâ ve makine öğrenmesi alanlarında popüler olan Veri Tabanlarında Bilgi Keşfi (VTBK) kavramı (KDD/Knowledge Discovery in Databases) ilk kez 1989 (Piatetsky-Shapiro 1991) Veri Tabanlarında Bilgi Keşfi çalıştayında vurgulanmıştır. Fayyad vd.'lerine göre VTBK, verilerden yararlı bilgiyi keşfetme sürecinin bütünü ifade ederken, veri madenciliği ise bu süreçte yer alan belli bir aşama olarak ifade edilmiştir (Fayyad vd., 1996: 39). 90'lı yıllarda popüler olmaya başlayan veri madenciliği zaman içerisinde birçok gelişme katetmiştir. Bu gelişimi incelediğimizde; Agrawal vd. (1996) birliktelik kurallarında kullanılan ve hızlı bir algoritma olan Apriori'yi önerirken, Fayyad vd. (1996) veri tabanlarında bilgi keşfinin süreçlerine ve veri madenciliğinin bu süreçteki yerine yönelik bir akış sunmuşlar ve veri madenciliğinin temel özelliklerini irdemişlerdir. Cabena vd. (1997) veri madenciliğinin diğer iş zekâsı çözümleriyle mukayesesini yapıp, aralarındaki hiyerarşiyi ortaya koymuş, Berson vd. (1999) VM'nin en yaygın kullanıma sahip alanı olan müşteri ilişkileri yönetimi kapsamında veri madenciliği yöntemleri ve uygulamalarına yer vermişlerdir, yeni nesil yöntemlerden olan Karar ağacı yöntemlerinin başlıcalarından CART ve CHAID'i incelemişlerdir (Koyuncugil, 2006: 7).

Veri tabanı endüstrisinin değişim adımlarının başka bir ifadesi Tablo 1.2’de gösterilmiştir.

Tablo 1.2 Veri Madenciliği Değişiminin Adımları

Değişimsel Adım	İşletme/Ticari Soru	Kullanılan Teknolojiler	Ürün Sağlayıcılar	Özellikleri
Veri Toplama (1960’lar)	Son Beş yıldaki gelirim ne?	Bilgisayarlar, diskler, teypler (bantlar)	IBM, CDC	Geçmişe yönelik, statik veri dağıtımı
Veri Girişi (1980’ler)	Geçen Mart’da İngiltere’de birim satış neydi?	İlişkisel Veri Tabanları (RDBMS) Yapısal Sorgulama Dilleri (SQL) ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Kayıt düzeyinde, geçmişe yönelik dinamik veri dağıtımı
Veri Ambarları ve Karar Destekleme (1990’lar)	Geçen Mart’da İngiltere’de birim satış neydi? Boston’a kadar inceleyin	Çevrimiçi Analitik Süreç (OLAP) Çok Boyutlu Veri Tabanları Veri Ambarları	Pilot, Comshare, Arbor, Cognos, Microstrategy	Çoklu düzeyde, geçmişe yönelik veri iletimi
Veri Madenciliği (Günümüzde gelişmekte olan)	Gelecek ay Boston’da birim satış miktarı ne olabilir? Neden?	İleri Algoritmalar Çok İşlemcili Bilgisayarlar Büyük Veri Tabanları	Pilot, Lockheed, IBM, SGI,	Beklenen, etkin veri dağıtımı

Kaynak: Thearling, 2014

Tablo 1.3’e göre veri madenciliği teknikleri uzun bir araştırma ve ürün geliştirme sürecinin sonucudur. Bu değişim ilk olarak işletme verilerinin bilgisayarlara depolanmasıyla başlamıştır, veri girişlerinin gelişimiyle devam etmiş ve veri madenciliği günümüzde de kullanıcılara verilerini yönetme olanağı sağlayan bir süreç haline gelmiştir. VM üç teknoloji tarafından desteklenmektedir, bunlar: Büyük Miktarda Veri Toplama, Çok İşlemcili Güçlü Bilgisayarlar, Veri Madenciliği Algoritmalarıdır (Thearling, 2014).

1.4 Veri Madenciliği ve İlişkili Olduğu Alanlar

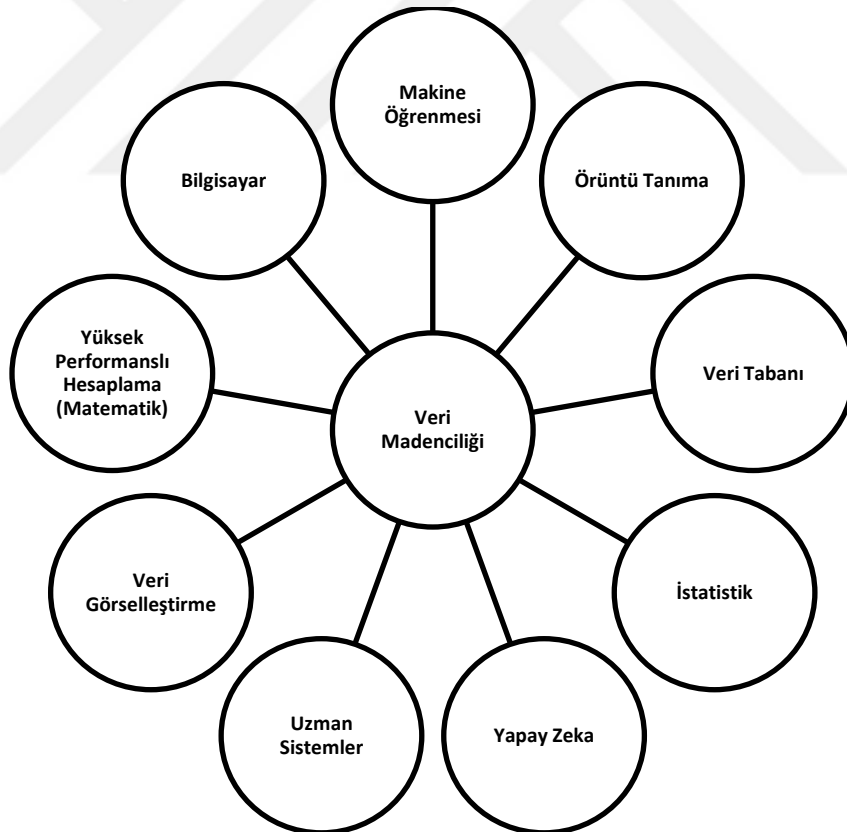
Veri madenciliği ve ilişkili olduğu alanlar ile ilgili birçok görüş vardır, bunları ayrıntılı olarak belirtecek olursak;

- Veri madenciliği problemi ve ilgili çözümlerin kökleri klasik veri analizindedir. Veri madenciliğinin birçok disiplinde kökenleri vardır, bunlardan en önemli ikisi istatistik ve makine öğrenimidir. İstatistiğin kökleri matematiğe dayanmaktadır, bu nedenle veri madenciliğinde matematiksel titizliğe vurgu vardır, yani bir şeyin pratikten önce teorik gerekçelerle mantıklı bir şekilde kurulması gerekmektedir. Buna karşılık, makine

öğrenimi topluluğunun köklerinin çoğu bilgisayar uygulamalarındandır. Bu da uygulamacı bir yöne, yani etkinliğinin biçimsel bir kanıtı için beklemeksizin ne kadar iyi performans gösterdiğini görmek amacıyla bir şeyi test etme isteğine neden olmaktadır (Kantardzic, 2011: 4).

- Veri Tabanlarında Bilgi Keşfi (KDD), makine öğrenmesi, örüntü tanıma, veri tabanı, istatistik, yapay zekâ, uzman sistemler, veri görselleştirme ve yüksek performanslı hesaplama gibi birçok farklı araştırma alanının kesiştiği, gelişmeye devam eden çok disiplinli bir alandır (Fayyad vd., 1996: 39).
- Veri madenciliği disiplinler arası bir alandır, bunun doğası gereği veri tabanı, makine öğrenmesi, istatistik, paralel ve dağıtık hesaplama, veri görselleştirme, bilgi geri kazanımı (information retrieval) gibi disiplinlerin katkısını almıştır. Hiç şüphesi birincil olan ilk üç alan; veri tabanı, makine öğrenmesi ve istatistiktir (Zhou, 2003: 1).

İncelenen kaynaklar neticesinde veri madenciliğinin ilgili olduğu tüm alanlar Şekil 1.2’de gösterilmiştir.



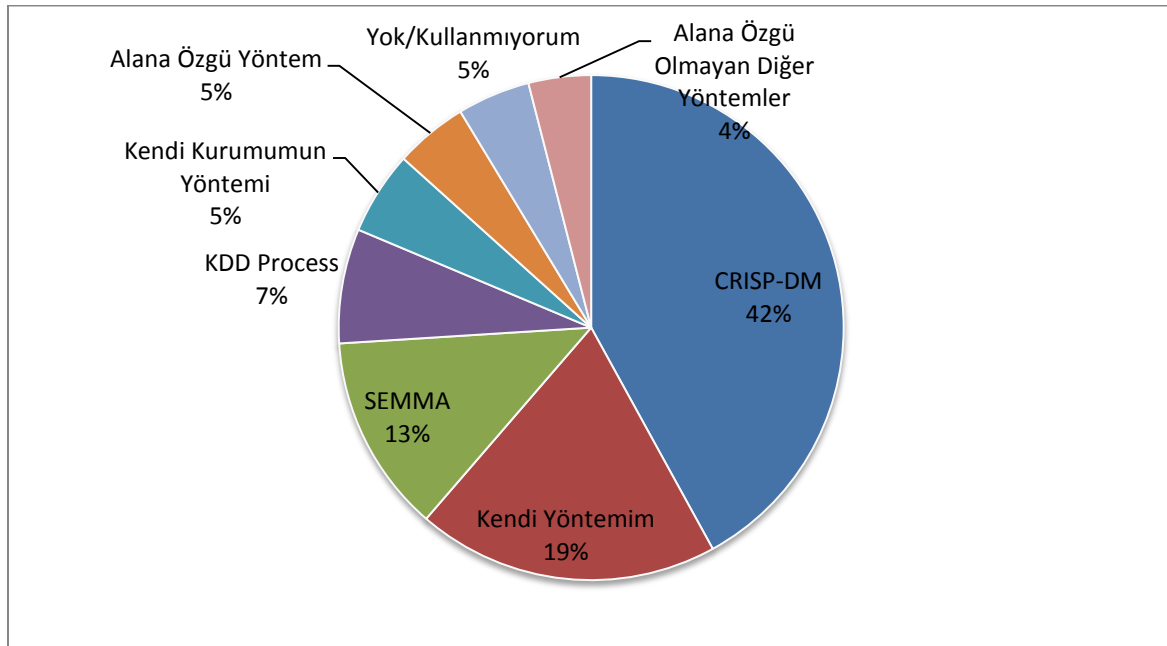
Şekil 1.2 Veri Madenciliğinin İlişkili Olduğu Alanlar

Kaynak: Fayyad vd., 1996: 39; Kantardzic, 2011: 4; Zhou, 2003: 1

1.5 Veri Madenciliği Süreci

Veri madenciliği, Veri Tabanlarında Bilgi Keşfi, Veri Madenciliği ve Bilgi Keşfi, Bilgi Keşfi kavramları, büyük hacimli veri kaynaklarından yararlı bilginin çıkarılması için kullanılan kullanılan terimlerdir (IBM, 1996: 1; Mariscal vd., 2010: 137). Daha önceden de belirtildiği gibi, bilgi çıkarma sürecinin tamamı Veri Tabanlarında Bilgi Keşfi Süreci olarak adlandırılırken, veri madenciliği tüm VTBK sürecinin sadece bir adımıdır (Fayyad vd., 1996: 42; Mariscal vd., 2010: 137). Genel olarak endüstri ve basın dünyasında ise veri madenciliği, tüm VTBK (KDD) sürecini belirtmek için kullanılır. Bu yüzden, bu alanda söz ederken her iki terim de belirsiz olarak kullanılabilir. Son yıllarda ise VTBK (KDD) sürecini temsil etmek için “Veri Madenciliği ve Bilgi Keşfi” en uygun isim olarak öne sürülmüştür (Mariscal vd., 2010: 137).

VTBK keşfi kavramının ortaya çıkmasının ardından Clementine, IBM Intelligent Miner, Weka ve DBMiner gibi uygulamalar, veri madenciliği algoritmalarının uygulanmasını kolaylaştırmak için geliştirilmiştir. Veri madenciliği süreci ve yöntemleri açısından bakıldığında 2000 yılında en önemli kilometre taşı olarak, Veri Madenciliği için Çapraz Endüstri Standardı Süreci (CRISP-DM) ileri sürülmüştür (Mariscal vd., 2010: 137-138). Veri madenciliğinde birçok metodoloji kullanılmaktadır. Kdnuggets tarafından 2007 Ağustos yılında sorulan “Veri Madenciliği için hangi metodolojiyi kullanıyorsunuz?” sorusuna verilen cevaplar Şekil 1.3’de gösterilmiştir.



Şekil 1.3 Veri Madenciliği İçin Kullanılan Metodolojiler

Kaynak: Kdnuggets, 2014

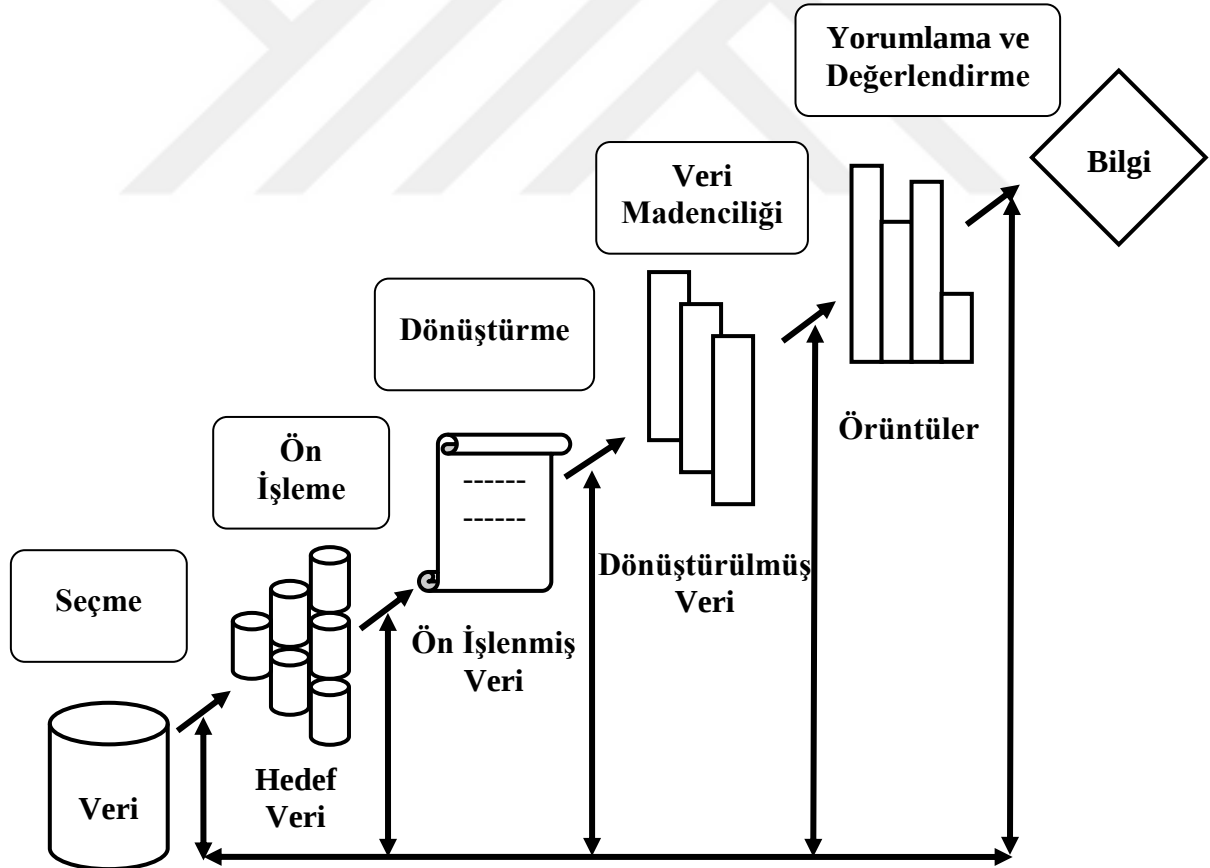
Kdnuggets.com'a göre veri madenciliği geliştirmek için kullanılan en yaygın metodoloji CRISP-DM olarak bildirilmiştir. Ayrıca SAS tarafından SEMMA metodolojisi de yaygın olarak kullanılmaktadır. (Kdnuggets, 2014)

Veri Madenciliği Süreç modelleri ve metodolojilerinin çeşitliliği açısından bakıldığında tam olarak kesin bir metodoloji yoktur. Genel olarak kabul görmüş bazı standartlar vardır. Bunlar endüstri standardı süreci CRISP-DM ve SEMMA'dır. İki yaklaşımda da her aşama her analiz için gerekli olmasa da bu işlem veri araştırma ile başlayarak, uygulama gibi gerekli adımları kapsar (Olson ve Delen, 2008: 9). CRISP-DM yaklaşımı ile SEMMA yaklaşımı ile tamamen uyumludur. İki yaklaşımda bilgi keşfi işlemine yardımcı olmaktadır (Olson ve Delen, 2008: 22).

1.5.1 Veri Tabanlarında Bilgi Keşfi Süreci ile İlgili Metodolojiler

Veri tabanlarında bilgi keşfi (VTBK) süreci ile geliştirilmiş ve uygulanmakta olan birçok metodoloji vardır. Bu kısımda VTBK sürecinde kullanılan metodolojilere yer verilmiştir.

1.5.1.1 Fayyad vd. Göre Veri Tabanlarında Bilgi Keşfi Süreci Aşamaları



Şekil 1.4 Veri Tabanlarında Bilgi Keşfi Süreci

Kaynak: Devedzic, 2002: 621

Uygulamada, bilgi keşfi süreci (KDD) asla sorunsuz çalışmaz. Aksine doğası gereği zaman alıcı, artan ve döngüsel bir süreçtir, dolayısıyla birçok tekrarlama ve geri besleme

olabilir. Bireysel aşamalar tek başına tekrar edilebilir ve tüm işlem genellikle farklı veri setleri için tekrarlanır. Şekil 1.4. geri besleme, tekrarlama döngüleri görülebilmektedir, bu şekil incelendiğinde süreç daha iyi anlaşılacaktır. Veri tabanlarında bilgi keşfi sürecinde yer alan aşamalar (Devedzic, 2002: 621);

- 1) **Seçme:** Veritabanından belirlenen hedef verilerin üretilmesi için yapılan işlemdir. Burada ki amaç hedef veri setini oluşturacak verileri, veri tabanından tipik olarak ayıklama işlemidir.
- 2) **Önişleme:** Bu aşama hedef veri setinden gürültüden arındırılarak ön işlenmiş veri kümesi ve ön işlenmiş veri dizileri oluşturmaktır.
- 3) **Dönüştürme:** Bir sonraki aşama, istenen veri madenciliği görevini yerine getirmek için ön-işlenmiş verilerin uygun bir forma dönüşümdür.
- 4) **Veri Madenciliği:** İstenen veri madenciliği işlemlerini çalıştırmak ve desenler kümesini oluşturmak için çalıştırılan bir prosedür aşamasıdır. Fakat tüm desenler ve örüntüler kullanışlı değildir.
- 5) **Yorumlama ve Değerlendirme:** Keşfedilen tüm desenleri yorumlama ve değerlendirme işlemi, gerekli görülen yararlı desenlerin ayrıştırılması ve geri kalanının göz ardı edilmesi işlemi kullanıcıya aittir. Ayrıştırılan yararlı desenler keşfedilen bilgiyi temsil etmektedir.

1.5.1.2 SEMMA Metodolojisi

Veri madenciliği sürecini planlayan metodolojilerden bir diğeri SAS Enstitüsü tarafından geliştirilen SEMMA'dır. SEMMA ile VTBK süreci arasında iki temel fark vardır bunlar; birinci olarak SEMMA, VTBK sürecinin ilk adımı uygulama alanının öğrenme sürecini atlayarak doğrudan örneklem adımı ile başlar, ikinci olarak VTBK süreci öğrenilen, keşfedilen bilginin kullanılması aşamasını içerirken SEMMA bu adımı içermez (Mariscal vd., 2010: 144). Şekil 1.5. SEMMA Metodolojisinin adımlarını göstermektedir.



Şekil 1.5 Veri Tabanlarında Bilgi Keşfi Süreci

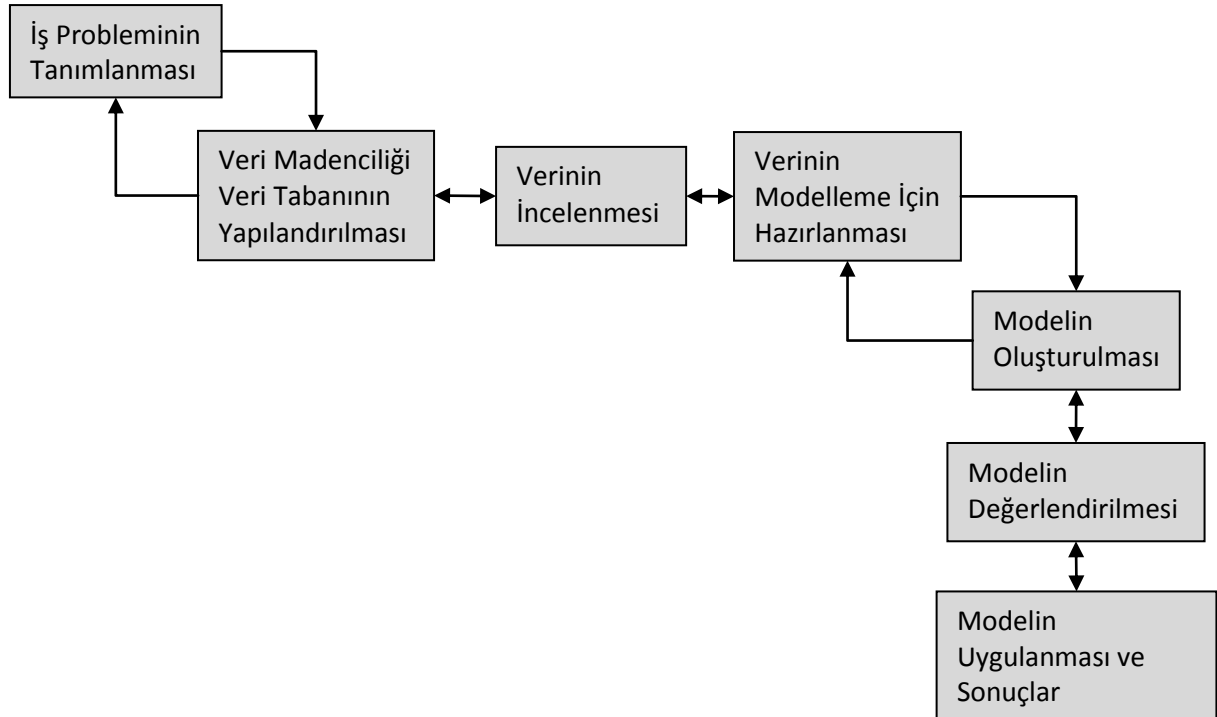
Kaynak: Mariscal vd., 2010: 144

SEMMA metodolojisinde yer alan adımlar ve açıklamaları (Olson ve Delen, 2008: 20-21);

- 1) **Örnekleme:** Önemli bilgileri içerecek kadar büyük aynı zamanda kolay yönetebilmek ve kolay işlenebilecek kadar küçük verinin çıkarılması işlemidir.

- 2) **İnceleme:** Kullanıcı, veri setinden daha iyi bir anlayış kazanmak için beklenmeyen eğilimleri, anomalileri bulmak için araştırma yapar. İnceleme aşaması, keşfetme sürecinde arındırma ve yönlendirmeye yardımcı olur. Görsel keşif net eğilimleri ortaya çıkarmıyorsa faktör analizi, uyum analizi ya da kümeleme analizi gibi istatistiksel araçlar ile veriler keşfedilebilir.
- 3) **Düzeltilme:** Model oluşturma sürecine odaklı olarak değişkenlerin kullanıcı tarafından seçimi, oluşturulması ve dönüşümü aşamalarını içerir. İnceleme aşamasına bağlı olarak verileri işlemek ve yeni değişkenleri tanıtmak gerekebilir. Aynı zamanda aykırı değişkenleri aramak ve değişkenlerin sayısını azaltmak için gerekli olabilir.
- 4) **Modelleme:** İstenen sonucu güvenli bir şekilde tahmin etmek için kullanıcı değişken kombinasyonları ile verinin modellenmesi aşamasıdır. Modelleme teknikleri; yapay sinir ağları, karar ağaçları, kaba küme analizleri, vektör makineleri, zaman serileri, bellek tabanlı akıl yürütme olarak sayılabilir. Her bir model tipi belirli bir güce sahiptir ve veriye bağlı olarak belirli veri madenciliği durumlarına uygundur.
- 5) **Değerlendirme:** Kullanıcının, veri madenciliği sürecinin bulgularının güvenilirliğini, kullanılabilirliğini ve yararlılığını değerlendirdiği yerdir. Veri madenciliği sürecinin son adımında, kullanıcı hangi modelin ne kadar iyi bir performansla tahmin ettiğini değerlendirir.

1.5.1.3 Two Crows'a Göre Veri Madenciliği Süreci



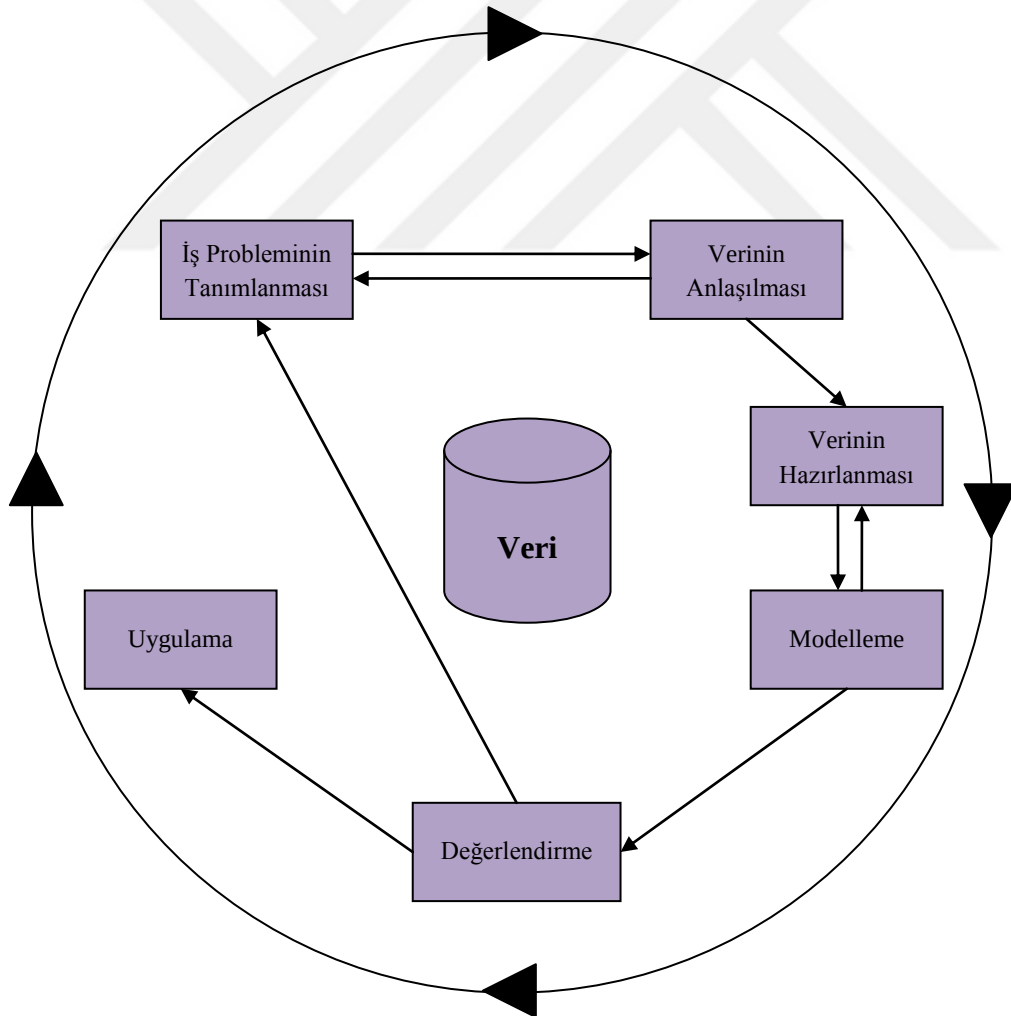
Şekil 1.6 Two Crows'a Göre Veri Madenciliği Süreci

Kaynak: Mariscal vd., 2010: 146

Şekil 1.6 Two Crows'a göre veri madenciliği sürecini göstermektedir. Bu süreç, iş probleminin tanımlanması adımı ile başlar, veri madenciliği veri tabanının yapılandırılması, verinin incelenmesi, verilenin modelleme için hazırlanması, modelin oluşturulması, modelin değerlendirilmesi adımları ile devam eder, modelin uygulanması ve sonuçlarını dağıtma adımı ile sona erer. Two Crows metodolojisi benzer için farklı isimler kullanmasına rağmen orijinal VTBK metodolojisine çok yakındır. Yukarıda sayılan metodolojilere ek olarak İnsan merkezli, Anand ve Buchner metodolojileri sayılabilir (Mariscal vd., 2010: 145-146).

1.5.1.4 CRISP-DM Metodolojisi

90'lı yılların ortalarında, veri madenciliği projelerindeki ihtiyaçları ve ortak sorunlara karşılık olarak Daimler Chrysler, SPSS, SIG, TeraData, OHRA ve NCR firmaları, Çapraz Endüstri Standardı Süreci (Cross Industry Standard Process for Data Mining) CRISP-DM'i geliştirmişlerdir. CRISP-DM metodolojisinin başarısının en önemli faktörü endüstriden, araçtan ve uygulamadan bağımsız olmasıdır (Mariscal vd., 2010: 146).



Şekil 1.7 CRISP-DM'e Göre Veri Madenciliği Süreci

Kaynak: Olson ve Delen, 2008: 10

Bu süreç modelinin geliştirilmesi güçlü sanayi desteğine sahiptir, aynı zamanda Avrupa Komisyonu tarafından finanse edilen ESPRIT programı tarafından desteklenmektedir (Cios vd., 2007: 12). CRISP-DM modeli endüstri kuruluşları tarafından yaygın olarak kullanılmaktadır. Bu model döngüsel olarak tasarlanmış 6 aşamadan oluşmaktadır. Bu aşamaları Şekil 1.7’de görebiliriz (Olson ve Delen, 2008: 9).

1) İşletme Probleminin Tanımlanması

Bu aşama işletme (iş) hedeflerini belirleme, mevcut durumun değerlendirilmesi, veri madenciliği hedeflerini oluşturmak ve bir proje planı geliştirme aşamalarını içerir (Olson ve Delen, 2008: 11). Veri madenciliği ve bilgi keşfi için birinci ve en önemli ön şart verilerinizi ve iş probleminizi anlamaktır. Problemi doğru tanımlamadan sorunları çözmeye çalışmak, madencilik için veriyi hazırlamak ve sonuçları doğru bir şekilde yorumlamak mümkün olmayacaktır. Veri madenciliğini iyi bir şekilde yapmak ve kullanmak için hedefleri açık ve net bir şekilde tanımlamak gereklidir (Two Crows, 1999: 22). Bu da yeni bir bilgiye duyulan yönetimsel ihtiyaçla ve yapılacak çalışmanın iş hedeflerinin tanımlanmasıyla başlar. Sonrasında böyle bir bilgi bulmak için veri toplama, verilerin analiz ve raporlama faaliyetlerinden sorumlu olan bir plan geliştirilmelidir. Bu aşamada çalışmayı desteklemek için hazırlık bütçesi oluşturulmalıdır (Olson ve Delen, 2008: 11).

2) Verinin Anlaşılması

İş hedefleri ve proje planı oluşturulduktan sonra verinin anlaşılması ve veri gereksinimleri göz önünde bulundurulur (Olson ve Delen, 2008: 9). Bu adım başlangıç verisinin toplanması ve veriyi tanıma ile başlar. Başlangıç verilerin toplanması, verinin tanımlanması, verinin keşfi ve veri kalitesinin doğrulanması aşamalarını içerir (Cios vd., 2007: 13).

3) Verinin Hazırlanması

Bu adım nihai veri kümesini oluşturmak için gerekli tüm faaliyetleri kapsar (Cios vd., 2007: 13). Mevcut veri kaynakları belirlendikten sonra, verilerin seçilmesi, temizlenmesi, istenen formda inşa edilmesi ve biçimlendirilmesi gerekir (Olson ve Delen, 2008: 9). Sıralı bir şekilde yazacak olursak; veri seçimi, verilerin temizlenmesi, verilerin inşası, verinin entegrasyonu ve veri alt basamakları biçimlendirme aşamalarını içerir (Cios vd., 2007: 13-14). Verinin hazırlanması aşaması bilgi keşfi sürecinin %50’si ile %90 zamanını kapsar (Two Crows, 1999: 23). Veri ön işlemenin amacı, daha kaliteli veri için seçilen verilerin temizlemesidir. Bazı seçilmiş veriler, farklı kaynaklardan seçildiği için farklı formatta olabilirler. Seçilen veriler web metin, metin dosyaları, ses gibi dosyalar ise tutarlı bir elektronik biçime dönüştürülmesi gerekmektedir. Filtreleme, birleştirme (toplama), uygun

olmayan veya hatalı girilmiş verileri ayıklamak ve eksik verileri doldurma genel olarak veri temizleme anlamına gelir. Filtreleme ile aykırı ve fazlalık veri için seçilen veri incelenir (Olson ve Delen, 2008: 12-13).

4) Modelleme

Modelleme aşaması, veri madenciliği yazılımları ile değişik durumlar için sonuçlar üretme aşamasıdır. İlk olarak kümeleme analizi ve verilerin görsel olarak keşfi yapılır. Verinin türüne bağlı olarak değişik modeller daha sonra uygulanabilir. Eğer amaç verinin gruplandırılması ise ve gruplar belirliyse diskriminant analizi uygun olabilir. Eğer amaç tahmin ise ve veri sürekli ise regresyon analizi, eğer veri sürekli değilse lojistik regresyon uygun olabilir. Her iki amaç için de yapay sinir ağları kullanılabilir. Verinin sınıflandırılması için karar ağaçları da başka bir teknik olarak kullanılabilir (Olson ve Delen, 2008: 15). Modelleme aşamasının alt adımları şu şekilde sıralanabilir; modelleme tekniği/tekniklerinin seçimi, test tasarımı üretimi, modellerin oluşturulması ve oluşturulan modellerin değerlendirilmesi aşamalarını içerir (Cios vd., 2007: 14).

5) Değerlendirme

Veri değerlendirme (yorumlama) aşaması çok önemlidir, çünkü verilerden bilginin özümsemesidir. İki konu önemlidir, birincisi veri madenciliği aşamasında keşfedilen bilgi desenlerinden iş değerinin nasıl anlaşılacağı, ikincisi veri madenciliği sonuçlarını göstermek için hangi görselleştirme araçlarının kullanılacağıdır. Keşfedilen bilgi desenlerinden iş değerini belirleme bulmaca oynamaya benzer. Düzgün bilgi desenleri yorumlamak amacıyla uygun bir görselleştirme aracı seçmek önemlidir. Pasta grafikleri, histogramlar, kutu diyagramları, dağılım diyagramları ve dağılımlar gibi birçok görselleştirme aracı mevcuttur (Olson ve Delen, 2008: 18). Değerlendirme aşamasını adım adım yazacak olursak; sonuçlarının değerlendirilmesi, süreç yorumlama/inceleme ve bir sonraki adımın belirlenmesi aşamalarını içerir (Cios vd., 2007: 14).

6) Uygulama

Modelin oluşturulması çoğunlukla projenin sonu anlamına gelmemektedir. Genellikle elde edilen bilginin müşterinin kullanabileceği şekilde organize edilmesi ve sunulması gerekir. Gereksinimlere bağlı olarak uygulama aşaması bir rapor oluşturma gibi basit ya da tekrarlanabilen veri madenciliği sürecinin uygulaması gibi karmaşık olabilir. Birçok durumda uygulama adımlarını gerçekleştirecek olan bir veri analisti değil, kullanıcı olacaktır (Olson ve Delen, 2008: 34). Uygulama aşaması; plan dağıtımı, plan izleme ve bakımı, nihai raporun üretimi ve süreç alt basamaklarının gözden geçirilmesi aşamalarını içerir (Cios vd., 2007: 14).

1.5.1.5 Cabena vd.'e Göre Veri Madenciliği Süreci

Şekil 1.8 Cabena vd. göre VM sürecini göstermektedir. İş/İşletme amaçlarının belirlenmesi, veri hazırlama (veri seçimi, veri ön işleme ve veri dönüşümü aşamalarını içerir), veri madenciliği, sonuçların analizi ve bilginin özüm senmesi aşamalarını içerir (Mariscal vd., 2010: 145).



Şekil 1.8 Cabena vd. Göre Veri Madenciliği Süreci

Kaynak: Tarapanoff vd., 2001

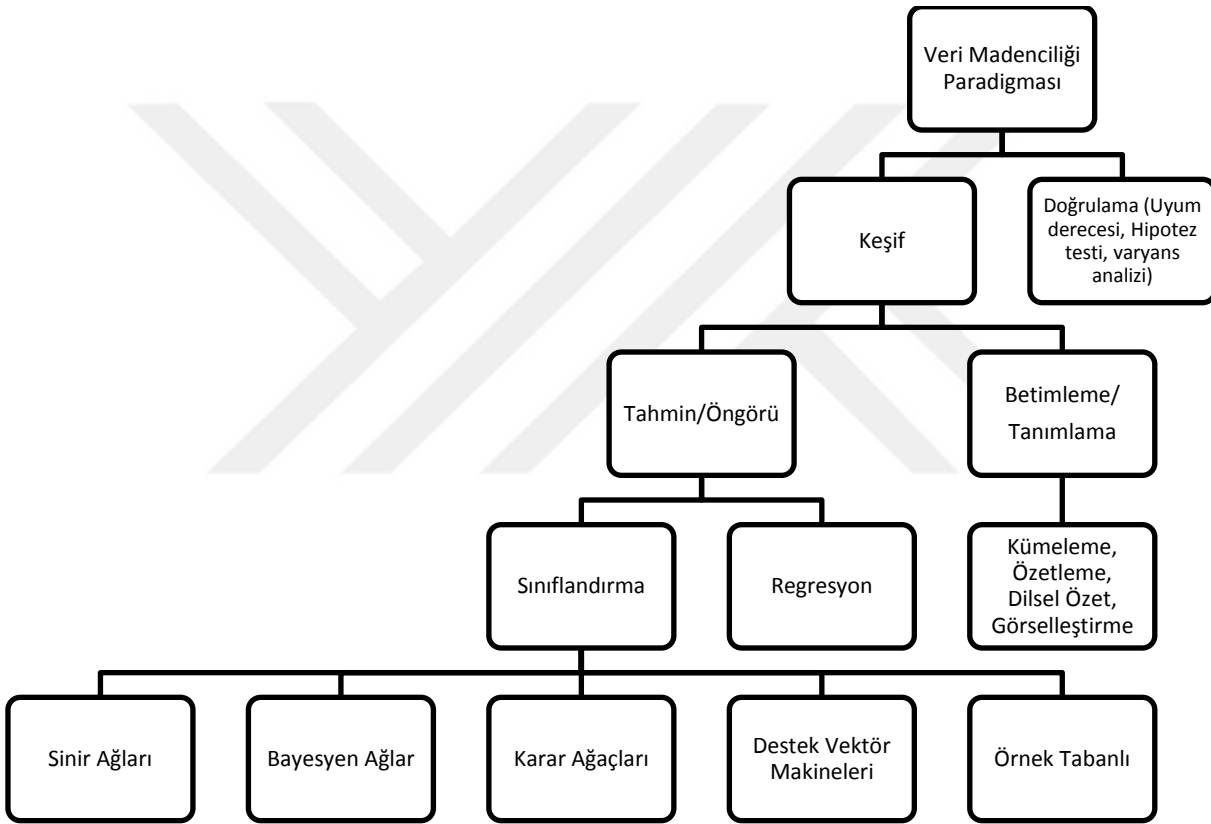
Veri Tabanlarında Bilgi Keşfi sürecinde kullanılan ve yukarıda detayları anlatılan bu beş metodolojinin haricinde Anand & Buchner, Human-Centered, 5A, 6 Sigma, Marbán vd, Cios vd, DMIE, RAMSYS ve KDD Roadmap gibi birçok metodoloji mevcuttur (Mariscal vd., 2010: 159).

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ YÖNTEMLERİ

2.1 Veri Madenciliği Yöntemlerinin Sınıflandırılması

Farklı amaçlar ve hedefler için kullanılan Veri Madenciliğinde birçok yöntem bulunmaktadır, bunun için bir taksonomi oluşturulmuştur. Maimon ve Rokach'a göre veri madenciliği taksonomisi ve detayları Şekil 2.1'de gösterilmiştir.



Şekil 2.1 Veri Madenciliği Taksonomisi

Kaynak: Maimon ve Rokach, 2005a: 6

Veri madenciliğinin iki ana türü olan keşif amaçlı (sistem özerk olarak yeni kurallar ve desenler bulur) ve doğrulama amaçlı (sistem kullanıcının hipotezini doğrular) türleri arasında seçim yapmak önemlidir. Keşif metotları verinin içinden desenleri, örüntüleri otomatik olarak tespit etme, belirleme amaçlıdır. Keşif metotları, Tahmin (Öngörü) ve Betimleme (Tanımlama) yöntemleri olarak iki dala ayrılır. Betimleyici (Açıklayıcı) metotlar verinin yorumlanmasına yöneliktir. Tahmin edici metotlarda amaç, örnekle ilgili bir ya da daha fazla

değişken değerlerini tahmin edebilen otomatik olarak bir davranış modeli oluşturmaya yöneliktir. Doğrulama metotları bir dış kaynak tarafından önerilen hipotezin değerlendirilmesini ele alır. Bu metot, uyum iyiliği testi, hipotez testi (t testi gibi), varyans analizi (ANOVA) gibi geleneksel istatistikî metotları içerir (Maimon ve Rokach, 2005a: 5-6).

Berry ve Linoff'a göre veri madenciliğinin amaçları; Sınıflandırma (Classification), Tahmin (Estimation), Öntahmin (Prediction), Benzer Gruplama (Affinity grouping), Kümeleme (Clustering) ile Tanımlama ve Ayrılama/Ayrıntılı İnceleme (Description & Profiling)'dir (Berry ve Linoff, 2004: 8).

Kimball ve Merz, veri madenciliğinin amaçlarını dört farklı gruba indirgemiş ve gruplandırmıştır, bunlar Sınıflandırma, Kümeleme, Tahmin-Öntahmin ve Benzer Gruplama'dır. Westphal ve Blaxton da veri madenciliğinin amaçlarını dört farklı grupta incelemiştir, bunlar Sınıflama, Tahmin, Kümeleme, Tanımlama'dır (Akın, 2008: 54).

Akpınar'a göre veri madenciliğinde kullanılan yöntemler, tahmin edici (Predictive) ve tanımlayıcı (Descriptive) yöntemler olmak üzere iki ana başlık altında incelenmektedir. Tahmin edici modeller amaç, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi, kurulan modelden yararlanarak sonuçları bilinmeyen veriler için sonuçların tahmin edilmesidir. Tanımlayıcı modellerde ise amaç, karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin, desenlerin tanımlanması sağlamaktır. Veri madenciliği modellerini; Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler, Kümeleme, Sınıflama ve Regresyon olmak üzere üç ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı Örüntü modelleri tanımlayıcı modellerdir (Akpınar, 2000: 5).

2.2 Karar Ağaçları

Karar ağaçları, sınıflandırma ve regresyon problemlerine uygulanabilen, etkili, parametrik olmayan, yoğun hesaplama gerektiren bir yöntemdir (Rao vd., 2005: 303). Bu yöntem, geniş bir kayıt koleksiyonunu basit ardışık bir dizi kurallar uygulayarak daha küçük kayıtlar kümesine bölen bir yapıdır. Birbirini izleyen her bölme ile ortaya çıkan setlerinin üyeleri birbirine daha çok benzemektedir. Bir karar ağacı modeli geniş heterojen popülasyonu küçük parçalara bölmek için bir kurallar kümesi içerir (Berry ve Linoff, 2004: 166). Bir karar ağacı, boş olmayan düğümler (nodes) ve kenarlar (edges) kümesi içeren, sonlu bir graf gibi tasvir edilebilir.

$$Graph = (Nodes, Edges)$$

Bu graf aşağıdaki özellikleri taşımalıdır (Barros vd., 2015: 8);

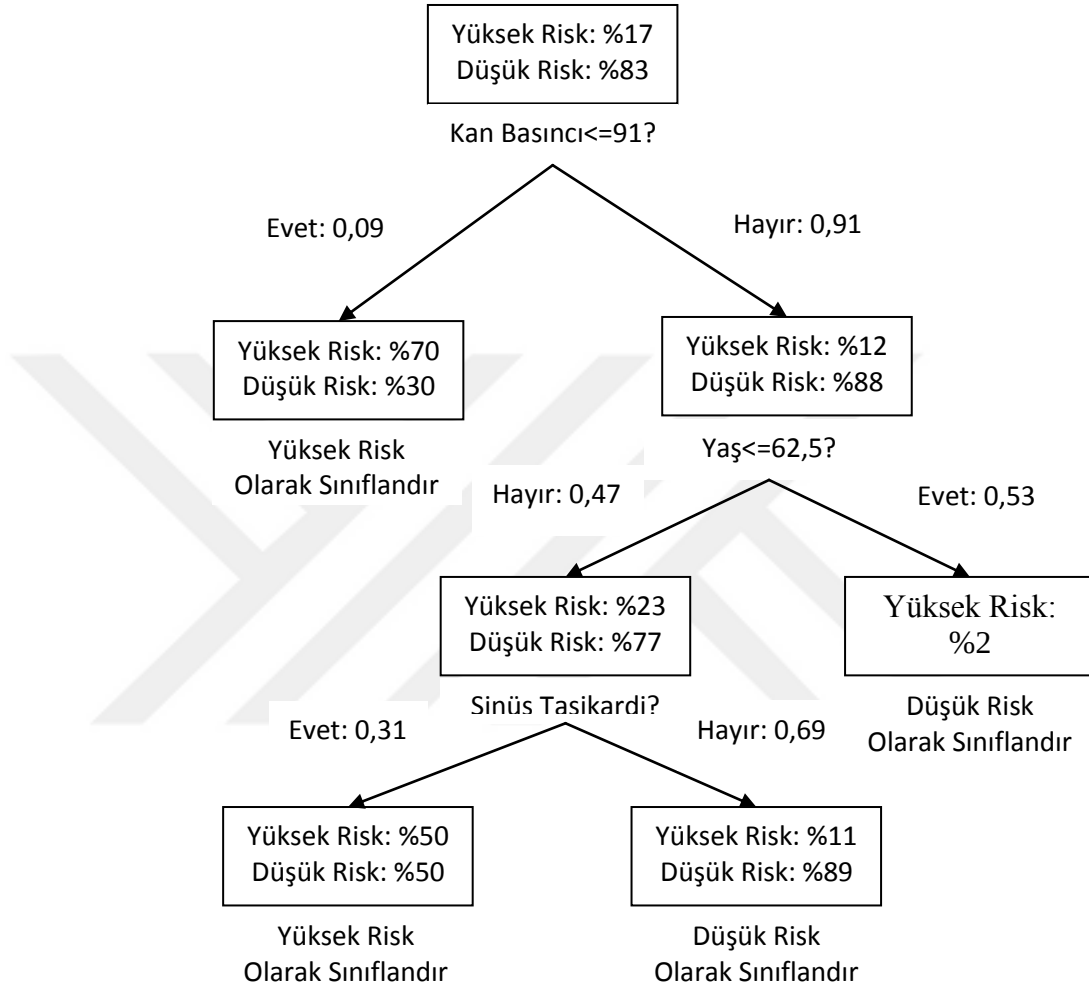
1. Graf yönlendirilmelidir.
2. Grafın içinde herhangi bir döngü olmamalıdır, başka bir ifade ile graf döngüsel olmamalıdır.
3. Yalnızca kök düğüme herhangi bir kenar gelmemektedir.
4. Diğer tüm düğümlerin, tam olarak gelen bir kenarı vardır.
5. Kök düğümünden diğer düğümlere sadece benzersiz (eşsiz) bir tek yol vardır.
6. V düğümünden W düğüme giden bir yol düşünelim ($V \neq W$), V düğümü, W düğümünün atası/ebeveynidir (üst düğümü). W düğümü, V düğümünün soyundan gelmektedir. W düğüme çocuk düğüm ya da alt düğüm denir. Soyundan gelen düğümü olmayan başka bir ifade ile alt düğümü olmayan düğümlere yaprak ya da terminal düğüm denir. Kök hariç diğer tüm düğümlere iç düğüm denir.

Derinlik ve genişlik kavramları karar ağaçları ile ilgili önemli kavramlardır. Ağacın ortalama derinliği, kök düğümünden terminal düğüme kadar seviyelerin ortalama sayısıdır. Ağacın her seviyesindeki iç düğümlerin ortalama sayısı ise ağacının ortalama genişliği olarak ifade edilir. Karar ağacının hem derinlik ve hemde genişliği ağacın karmaşıklık göstergeleridir. Bunların yüksek değerlere sahip olması karar ağacının daha karmaşık olduğunu gösterir (Barros vd., 2015: 8).

Bir sınıflandırma veya regresyon probleminde eğitim kayıtlarının veri seti verilir bu eğitim veri tabanı olarak da isimlendirilebilmektedir. Her kayıt birçok öznitelikten oluşmaktadır. Öznitelikler, sayısal öznitelikler ve sayısal olmayan öznitelikler (kategorik öznitelikler) olarak ayrılmaktadır. Bağımlı öznitelik olarak adlandırılan bir tane seçkin öznitelik vardır. Geriye kalan diğer öznitelikler belirleyici öznitelikler olarak isimlendirilir; bunlar sayısal ya da kategorik olabilirler. Bağımlı öznitelik kategorik ise problem bir sınıflandırma problemi olarak adlandırılır ve bağımlı öznitelik sınıf etiketi olarak çağrılır. Bağımlı öznitelik sayısal ise problem regresyon problemi olarak isimlendirilir. Sınıflandırmanın amacı, belirleyici öznitelikler açısından bağımlı özneliğin dağılımının kısa ve özlü bir modelini inşa etmektir. Çıkan model, belirleyici özniteliklerin değerleri bilinen, ama bağımlı özneliğinin değeri bilinmeyen değerleri veritabanına atamak için kullanılır. Sınıflandırma bilimsel deneyler dâhil tıbbi tanı, dolandırıcılık tespiti, kredi onay ve hedef pazarlama gibi geniş bir uygulama alanına sahiptir. Sinir ağları (Bishop, 1995; Ripley, 1996), genetik algoritmalar (Goldberg, 1989), Bayes yöntemleri (Cheeseman'ın & Stutz, 1996), log-doğrusal modeller ve diğer istatistiksel yöntemler (Agresti, 1990; Chirstensen 1997, James, 1985), karar tabloları (Kohavi, 1995), ve ağaç yapılı modelleri, sözde sınıflandırma ağaçları (Breiman, Friedman,

Olshen, & Stone, 1984; Quinlan, 1986) gibi birçok sınıflandırma modeli literatürde önerilmiştir (Ye, 2003: 3-4).

Karar Ağacına örnek olarak Şekil 2.2’de bir hastanede kalp krizi hastaları için oluşturulan bir karar ağacını göz önüne alalım.



Şekil 2.2 Karar Ağacı Örneği

Kaynak: Salford System, 2001: 11

Tıbbi personel mümkün olduğunca basit ve hızlı bir şekilde, ikinci öldürücü kalp krizi riskinin yüksek olduğu hastayı belirlemek istemektedir. Düşük riskli bir hasta standart bir sağlık biriminde kalacak, yüksek riskli hasta ise sürekli takip edilmesi için yoğun bakım ünitesinde kalacaktır. Analiz için kullanılacak veri, 215 hastanın tıbbi kaydından oluşmaktadır. Her kayıt, hasta kabulünü izleyen ilk 24 saat içerisinde yaş, taşikardi varlığı, kan basıncı gibi temel değişkenleri içeren, ayrıca kan analizine dayalı enzim konsantrasyonu gibi karmaşık 19 ölçüm içermektedir. Kabulü izleyen 30 gün içinde, yüksek risk grubu 37 hasta ölmüş, düşük risk grubu 178 hasta yaşamına devam etmektedir. Bu verilerin sistematik

analizine dayanarak, CART (Classification and Regression Tree) ters bir ağaç şeklinde hazırlanan yukarıdaki şemayı (Şekil 2.2) üretmiştir (Salford System, 2001: 10).

Karar Ağacını okumaya ilk önce en yüksek yerden (Kök-Root) başlanır. Kalp krizi hastasının sistolik kan basıncı 91 veya altına düşerse hastayı yüksek riskli olarak görüyoruz. Bu karar sadece tek değişken temelinde yapılmıştır. İlk olarak yaş'a bakılır, yaş 62,5'un altında ise hasta düşük riskli olarak sınıflandırılır. İkinci olarak 62,5 yaş üstü hastalar sinüs taşikardi temelinde yüksek ve düşük risk olacak şekilde bölünür (ayrıştırılır). Sinüs taşikardi var ise hasta yüksek riskli olarak sınıflandırılır. Mevcut diğer veriler hiçbir sınıflamada kullanılmamıştır. Sınıflandırma ve regresyon ağaçları (CART), bir kök düğüm ile başlar ve evet/hayır soruları süreci boyunca soyundan gelen (torun/alt) düğümleri oluşturur. Bazı düğümler terminaldir, bu sınıflandırma işleminin sona erdiğini göstermektedir, diğer düğümler terminal düğüme ulaşana kadar bölünmeye devam edecektir. Her bölünme bir ebeveyn (üst) düğümü, tek bir değişkene ilişkin bir soruya dayalı olarak iki çocuk (alt/torun) düğüme ayırır (Salford System, 2001: 11-12).

CART metodolojisi teknik açısından ikili özyinelemeli bölümlenme olarak bilinir. Bu süreç ikilidir çünkü ebeveyn düğümler her zaman iki alt düğüme ayrılır ve özyinelemelidir.

2.2.1 Karar Ağacı Algoritmaları

Karar Ağaçları gelişim sürecini incelediğimizde; (Rao vd., 2005: 304)

- 1960 yılında Morgan ve Sonquist (1963) tarafından AID'in (Automatic Interaction Dedection) oluşturulmasıyla gelişmeye başlamıştır.
- 1970'lerde, Morgan ve Messenger (1973) tarafından sınıflandırma ağaçları üretmek/oluşturmak için Thaid (Teta AID) oluşturulmuştur.
- THAID kökenli bir metot olan CHAID (CHi-squared Automatic Interaction Detector) Kass (1980) tarafından geliştirilmiştir.
- 1980'li yıllarda istatistikçiler Breiman, Friedman, Olshen, Stone (1984) tarafından, gelişmiş bir program olan veriyi ağaçlara uydurma CART (Classification And Regression Trees) geliştirilmiştir.
- Makine öğrenmesi topluluğu, sınıflandırma için karar ağaçları oluşturan birçok program üretmişlerdir. Bunlar arasında Quinlan'ın son derece popüler olan ID3 ve ID3 (Quinlan, 1979) kökeninden gelen C4.5 ve See5/C5.
- Ayrıca istatistikçiler başka ağaç tabanlı yöntemler geliştirmişlerdir, örneğin QUEST (Quick Unbiased Efficient Statistical Tree) metodu Loh ve Shih (1997) tarafından geliştirilmiştir.

Karar ağaçları oluşturulurken (yapılandırılırken) kullanılan algoritmaya göre ağacın yapısı, görüntüsü değişmektedir. Değişik ağaç yapıları farklı sınıflandırma sonuçları verebilmektedir. Karar ağaçlarına bağlı olarak geliştirilen çok sayıda algoritma bulunmaktadır, bu algoritmalar birçok açıdan farklılık göstermektedirler. Bunlar kök düğüm, düğüm ve bölünme kriteri seçimlerinde izledikleri yol olarak sayılabilir (Tapkan vd., 2011: 249).

2.2.2 Karar Ağaçlarının Oluşturulması

Karar ağacı algoritmaları, verilen bir veri setinden otomatik olarak bir karar ağacı oluşturan algoritmalarlardır. Genellikle hedef, genelleme hatasını minimize ederek optimal (en uygun karar) ağacını elde etmektir. Düğüm sayısını en aza indirmek veya ortalama derinliği en aza indirmek gibi diğer hedef fonksiyonları da tanımlanabilir (Maimon ve Rokach, 2005a: 151). Karar ağacı algoritmaları, ağacın oluşturulması ve budama olarak adlandırılan iki aşamadan oluşmaktadır. Birçok karar ağacı oluşturma algoritmalarında ağaç yukarıdan aşağıya doğru büyür. Veritabanında kök düğümden başlayarak her düğümde bölünme koşulunu seçmek için “bölünme seçim metodu” uygulanmıştır ve bu süreç özyinelemeli olarak devam eder. (Sumathi ve Sivanandam, 2006: 157). Karar ağaçlarının oluşturulmasına ait algoritma aşağıda verilmiştir (Ye, 2003: 7).

Input: Node n, partition D, Split selection method SS

Output: Decsion tree for D rooted at node n

Top – Down Decision Tree Induction Schema

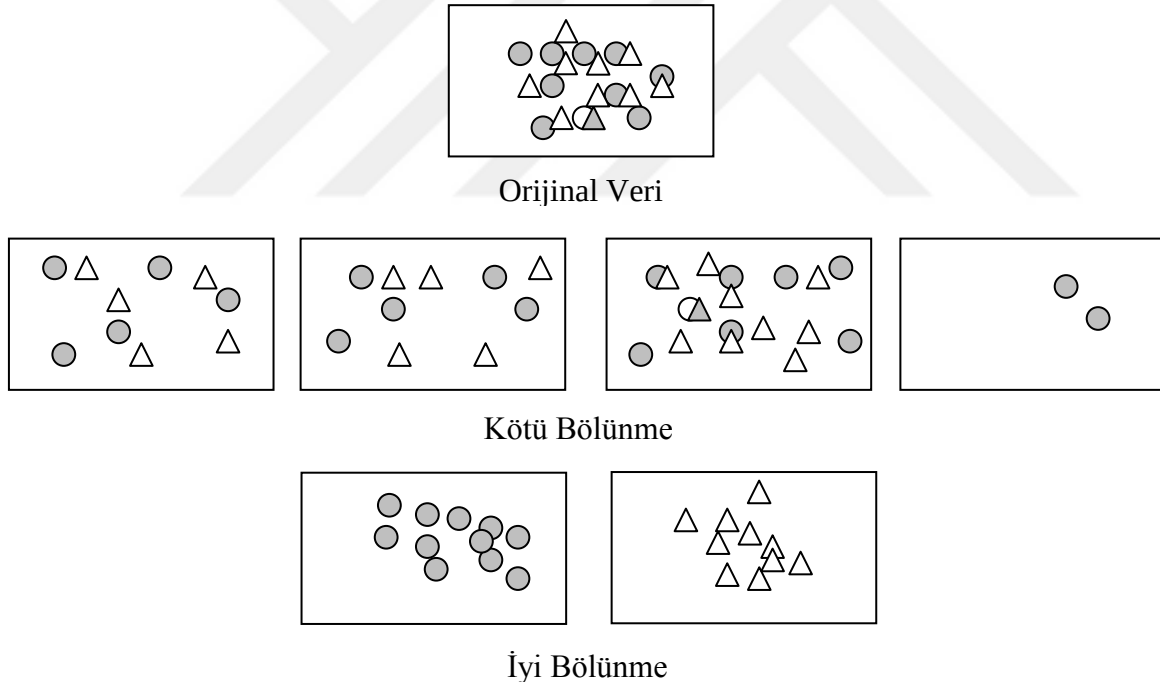
BuildTree(Node n, dataset D, split selection method SS)

1. Apply SS to D to find the splitting criterion
2. If n splits
3. Use best split to partition D into D_1 and D_2
4. BuildTree(n_1 , D_1 , SS)
5. BuildTree(n_2 , D_2 , SS)
6. endif

2.2.3 Bölünmeleri Bulma

İlk işlem hangi giriş alanının en iyi bölünmeyi sağlayacağına karar vermektir. Bütün girdi öznitelikleri denendikten sonra iki ya da daha fazla alt düğüm oluşturmak için ilk bölünme en iyi bölünme olarak belirlenir. Hiçbir bölünme olmadığında (kayıtların sayısı az olduğunda) ya da herhangi bir bölünme iyileşmeye yol açmadığında algoritma durur ve bu

düğüm bir yaprak düğüm haline dönüşür. Aksi takdirde algoritma bölünmeleri gerçekleştirir ve meydana gelen alt düğümler için bu süreç devam eder. Yani algoritma kendi kendini tekrarlamaktadır bu yüzden rekürsif olarakda isimlendirilebilmektedir. Uygun bir bölünme kriteri seçimi, girdi özniteliğinin türüne değil hedef özniteliğin türüne bağlıdır. Hedef öznitelik kategorik ise Gini, Bilgi Kazancı veya Ki Kare, sürekli ise Varyans Azaltma veya F Testi girdi özniteliğinin sürekli ya da kategorik olup olmamasına bakılmaksızın bölünme için uygundur (Berry ve Linoff, 2004: 173). Karar ağaçlarının endüksiyonunda başlıca meselenin düğümü alt kümelerle ayırmak için hangi özniteliğin (özniteliklerin) seçilmesi gerektiğini daha önce ifade etmiştik. Paralel eksen karar ağaçlarında (tek-değişkenli) ise problem giriş verisini daha iyi ayıran özniteliği seçmektir. Böyle bir öznitelik temelinde bir karar kuralı oluşturulur ve giriş verisi bu kuralın sonuçlarına göre filtre edilir. Oblik karar ağaçları için (çoklu-değişkenli) ise hedef iyi ayırma gücünde bir öznitelikler kombinasyonu bulmaktır. İki durumda da, iki strateji de sıralı öznitelikler ile nicelik bakımından ilgilenir (Barros vd., 2015: 11).



Şekil 2.3 İyi ve Kötü Bölünmeler

Kaynak: Berry ve Linoff, 2004: 172

Düğümün her yeni nesli hedef değişkene göre atalarından daha fazla saflığa sahip olacak şekilde, daha küçük gruplar halinde tekrar tekrar bölünmektedir. Şekil 2.3 iyi ve kötü bölünmeleri göstermektedir.

2.2.3.1 Bölünme Kriterleri

En iyi bölünme, kayıt setlerinin saflığını en yüksek derecede artırır. İyi bir bölünme işlemi benzer büyüklüklerde düğümler oluşturur ya da en azından az kayıtlar içeren düğümler oluşturmaz. Bölünme işlemi gerçekleştirilirken birçok bölünme ölçütü çeşitli kategorilere ayrılmıştır. Berry ve Linoff'a göre sınıflandırma problemlerinde, hedef değişkenin kategorik ya da nümerik olmasına göre değişen birçok saflık ölçütü vardır (Berry ve Linoff, 2004: 172-178). Bunlar;

- Kategorik değişkenler için; Gini kriteri (Popülasyon Çeşitliliği), Entropi (Bilgi Kazancı), Bilgi Kazanç Oranı, Ki Kare Testi,
- Nümerik hedef değişkenler için; Varyans Azaltma ve F Testi kullanılan ölçütlerdir.

Rokach ve Maimon bölünme kriterlerini, değişkenin tek ve çok olmasına göre ayırmıştır. Karar ağacı endükleyicilerinin çoğunda ayrık bölünme fonksiyonları tek değişkenlidir. Tek değişkenli denmesinin anlamı, bir iç düğümün tek bir öznitelik değerine göre bölünmüş olmasıdır. Algoritma (endükleyici) bölünme işlemini gerçekleştirmek için en iyi özniteligi arar. Birçok tek değişkenli bölünme kriteri vardır ve bunlar farklı şekilde karakterize edilebilirler. Örneğin:

- Ölçmenin kökenine göre olan tek değişkenli bölünme kriterleri: bilgi kuramı, bağımlılık ve uzaklık,
- Ölçmenin yapısına göre olan tek değişkenli bölünme kriterleri: safsızlık (impurity) tabanlı kriterler, normalleştirilmiş safsızlık tabanlı kriterler ve ikili kriterler

gibi kriterler yaygın olarak kullanılmaktadır (Rokach ve Maimon, 2014: 61).

Barros vd. ise tek değişkenli bölünme ölçütlerini aşağıdaki gibi kategorilere ayırmışlardır;

- Bilgi temelli kriterler;
- Uzaklık temelli kriterler;
- Diğer sınıflandırmalar kriterleri;
- Regresyon kriterleri.

Bu kategorilerin hatları bazen belirsizdir ve ne şekilde olursa olsun bir taksonomi oluşturmazlar. Verilen bir kategoride sunulan çoğu kriterler başka kategorilerdeki kriterlerin tahminleri olarak gösterilebilir (Barros vd., 2015: 11).

Çok değişkenli bölünme kriterlerinde, birçok öznitelik tek düğüm ayırma testine katılabilir. En iyi çok değişkenli kriterleri bulma, en iyi tek değişkenli kriterleri bulmaya göre daha çok karmaşıktır. Dahası bu tip kriterler ağacın performansını bariz bir şekilde

artırılmasına karşın tek değişkenli kriterlere göre daha az popülerdir. Çok değişkenli bölünme ölçütlerinin büyük bir kısmı, girdi özniteliklerin doğrusal (lineer) bileşimlerine dayanmaktadır. En iyi doğrusal kombinasyon bulmak için aç gözlü arama (greedy search), lineer programlama, lineer diskriminant analizi gibi yöntemler kullanarak yapılabilir (Maimon ve Rokach, 2005a: 157).

2.2.3.1.1 Safsızlık Tabanlı Kriterler

“k” ayrık (kesikli) değerlere sahip rasgele verilen “x” değişkeni, $P = (p_1, p_2, \dots, p_k)$ şeklinde dağıtılmış olsun, safsızlık ölçütü fonksiyonu $\phi: [0, 1]^k \rightarrow \mathbb{R}$, aşağıdaki koşulları sağlar;

- $\phi(P) \geq 0$
- En az bir i için $p_i=1$ ise $\phi(P)$ minimumdur. (Eğer $\exists i, p_i=1$ ise $\phi(P)$ minimumdur.)
- Eğer $\forall i$ için, $1 \leq i \leq k, p_i = 1/k$ ise $\phi(P)$ maksimumdur.
- $\phi(P)$, P 'nin bileşenlerine göre simetriktir.
- $\phi(P)$ kendi aralığında düzgündür ve her yerde diferansiyellenebilir (türevlenebilir).

Belirtmek gerekir ki, eğer olasılık vektörü 1'in (x değişkeni yalnızca bir değer alır) bir bileşenine sahipse o zaman değişken saf olarak adlandırılır. Diğer bir yandan, eğer tüm bileşenleri eşitse, safsızlık düzeyi (seviyesi) maksimuma ulaşır. “S” verilmiş bir eğitim kümesi olsun, hedef özniteliğin olasılık vektörü “y” şu şekilde tanımlanır;

$$P_y(S) = \left(\frac{|\sigma_{y=c_1} S|}{|S|}, \dots, \frac{|\sigma_{y=c_{|dom(y)|}} S|}{|S|} \right)$$

Kesikli öznitelik a_i kaynaklı bölünmenin iyiliği, “ $v_{i,j} \in \text{dom}(a_i)$ ” değerlerine göre S'in bölümlere ayrılmasından sonra hedef özniteliğin safsızlığının azalması olarak tanımlanır.

$$\Delta\Phi(a_i, S) = \phi(P_y(S)) - \sum_{j=1}^{|\text{dom}(a_i)|} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot \phi(P_y(\sigma_{a_i=v_{i,j}} S))$$

(Barros vd., 2015: 13; Fayyad ve Irani; 1992: 105; Rokach ve Maimon 2014: 62)

Bilgi Kazancı – Entropi:

Quinlan'ın ilk karar ağacı algoritması olan ID3, veri setinin bölünmesi için Bilgi Kazancını ya da eş değer olarak Entropiyi kullanır (Aggarwal, 2015: 93). Maksimum

Olabilirlik Yöntemi ile yakından ilişkili olan bilgi kazancı, safsızlık ölçütü gibi entropi ölçüsü kullanan safsızlık tabanlı bir kriterdir (Rokach ve Maimon, 2014: 62).

$$BilgiKazancı(a_i, S) = Entropi(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot Entropy(y, \sigma_{a_i=v_{i,j}} S)$$

$$Entropi(y, S) = \sum_{c_j \in dom(y)} -\frac{|\sigma_{y=c_j} S|}{|S|} \cdot \log_2 \frac{|\sigma_{y=c_j} S|}{|S|}$$

Entropi kullanmak her zaman daha küçük bir karar ağacı oluşturmayı garanti etmez, fakat deneyimler entropinin diğer kriterlere göre daha az dallara sahip ağaç ürettiğini göstermektedir. Deneyimler küçük ağaçların, büyük olanlara göre daha doğru (hassas) tahmin yapma eğiliminde olduğunu göstermektedir (Bramer, 2007: 57).

Gini İndeksi:

İtalyan ekonomist ve istatistikçi Corrado Gini'den adını alan Gini indeksi, popüler bir bölünme ölçütüdür. Biyolog ve çevre bilimciler tarafından kullanılan bu kriter, aynı popülasyondan rastgele seçilen iki türün aynı sınıfta olma olasılığını vermektedir (Berry ve Linoff, 2004: 178). Gini indeksi, hedef özniteliklerin değerlerinin olasılık dağılımları arasındaki farklılıkları ölçen safsızlık tabanlı bir kriterdir. Gini indeksi Breiman (1984) ve Gelfand (1981) tarafından kullanılmıştır (Rokach ve Maimon, 2014: 62-63).

$$Gini(y, S) = 1 - \sum_{c_j \in dom(y)} \left(\frac{|\sigma_{y=c_j} S|}{|S|} \right)^2$$

a_i özniteliğini seçmek için değerlendirme kriteri aşağıdaki gibi tanımlanır;

$$GiniGain(a_i, S) = Gini(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot Gini(y, \sigma_{a_i=v_{i,j}} S)$$

Gürültülü (aykırı) veriler için iyi çalışan Gini İndeksi, CART tarafından bölünme kriteri olarak kullanılmaktadır (Aggarwal, 2015: 94; Timofeev, 2004: 10). CART yazarlarının, bilgi kazancı yerine gini indeksini tercih etmelerinin nedeni gini indeksinin

simetrikleştirilmiş maliyetleri içerecek şekilde genişletilebilir olması ve bilgi kazancından daha hızlı hesaplanmasıdır (Kantardzic, 2011: 189).

Olabilirlik Oranı (Likelihood Ratio) - Ki Kare:

Ki kare, birden çok ayrık örnek arasında bazı oluşumların beklenen ve gözlenen frekanslar arasındaki standartlaştırılmış farkların karelerinin toplamı olarak tanımlanır (Berry ve Linoff, 2004: 180).

Olabilirlik oranı;

$$G^2(a_i, S) = 2 \cdot \ln 2 \cdot |S| \cdot \text{InformationGain}(a_i, S)$$

şeklinde tanımlanır.

Bu oran, bilgi kazancı kriterinin istatistiksel öneminin ölçülmesi için yararlıdır. Sıfır hipotezinde (H_0) girdi ve hedef öznitelikleri koşullu (şartlı) olarak bağımsızdır. Eğer H_0 hipotezi kabul edilirse, test istatistiği $(\text{dom}(a_i) - 1) \cdot (\text{dom}(y) - 1)$ serbestlik derecesi ile χ^2 olacak şekilde dağıtılmaktadır (Rokach ve Maimon, 2014: 63).

DKM Kriteri:

Sınıf özniteliği ikili olduğu durumlar için dizayn edilmiş olan DKM bölünme kriteri adını, Dietterich, Kearns, ve Mansour'dan alır (Aggarwal, 2015: 95). DKM kriteri safsızlık tabanlı (impurity based) bir bölünme kriteridir. Safsızlık tabanlı fonksiyon şu şekilde tanımlanır;

$$DKM(y, S) = 2 \cdot \sqrt{\left(\frac{|\sigma_{y=c_1} S|}{|S|}\right) \cdot \left(\frac{|\sigma_{y=c_2} S|}{|S|}\right)}$$

DKM kriteri diğer safsızlık tabanlı kriterlere (Bilgi Kazancı ve Gini İndeksi) göre, belirli bir hata elde etmek için daha küçük ağaç gerektirdiği teorik olarak ispatlanmıştır (Rokach ve Maimon, 2014: 63). Böylelikle DKM ölçütünün, Entropi ve Gini indeksine göre daha çok tercih edilebilir olduğu gösterilmiştir (Donskoy, 2013: 17). Yazarlar, tahmin doğruluğunun belirli bir seviyesi için DKM tabanlı ağacın beklenen boyutunun, C4.5 ya Gini kullanılarak oluşturulan ağaçlardan daha küçük olduğunu kanıtlamıştır (Aggarwal, 2015: 95). Çeşitli yazarlar karar ağacı bölünme kriteri olarak uygulanan DKM'nin dengeli olmayan (unbalanced) veri setleri üzerinde gelişmiş performans gösterdiğini bildirmişlerdir (Cieslak ve Chawla, 2008: 241-242).

2.2.3.1.2 Normalize Edilmiş Safsızlık Tabanlı Kriterler

Safsızlık tabanlı kriterler, geniş alan değerlere sahip özniteliklere karşı yanlıdır. Açıkçası bu kriter çok değere sahip girdi özniteliklerini, az değere sahip girdi özniteliklerine tercih etmektedir. Örneğin, ulusal güvenlik numarasını gösteren bir girdi özniteliği büyük olasılıkla en yüksek bilgi kazancını elde edecektir. Ancak, bu özniteliği karar ağacına eklemek zayıf (poor) genelleştirilmiş doğruluğa neden olacaktır. Bu nedenle safsızlık tabanlı ölçütleri normalleştirmek yararlıdır (Rokach ve Maimon, 2014: 63-64).

Kazanç Oranı:

Quinlan yeni bir ölçü oluşturmak için bilgi kazancını normalize etmiştir, bu yeni ölçüt Kazanç Oranı (gain ratio) olarak bilinmektedir. Kazanç oranı, Quinlan'ın C4.5 karar ağacında kullanılmaktadır (Aggarwal, 2015: 94). Kazanç oranı, bilgi kazancını aşağıdaki gibi normalleştirir. (Quinlan, 1993)

$$GainRatio|(a_i, S) = \frac{InformationGain(a_i, S)}{Entropy(a_i, S)}$$

Payda sıfır olduğunda tanımsız olan bu oran, payda çok küçük olduğunda öznitelikler lehine eğilim gösterebilir. Dolayısıyla, bu oranın iki aşama da gerçekleştirilebilir olduğu önerilir. İlk olarak, bütün öznitelikler için bilgi kazancı hesaplanır. Yalnızca en az ortalama kazanç kadar performans gösteren öznitelikler dikkate alınır ve en iyi kazanç oranını elde eden öznitelik seçilir. Quinlan, hem doğruluk açısından hem de sınıflandırıcı karmaşıklığı açısından kazanç oranının, basit bilgi kazancı ölçütüne göre daha iyi performans eğilimi gösterdiğini göstermiştir (Rokach ve Maimon, 2014: 64).

2.2.3.1.3 Uzaklık Ölçüsü

López de Mántaras normalize bilgi mesafesinin bir bölünme kriteri olarak kullanılmasını önermiştir. Bu fonksiyon bir uzaklık metriğidir. Negatif olmama, simetrik olma ve üçgen eşitsizliği gibi özellikleri sağlamaktadır, ayrıca bu fonksiyon [0,1] aralığı arasında normalize edilir (Aggarwal, 2015: 95).

$$\frac{\Delta\Phi(a_i, S)}{-\sum_{v_{i,j} \in dom(a_i)} \sum_{c_{k,j} \in dom(y)} \frac{\left| \sigma_{a_i=v_{i,j} \vee y=c_{k,j}} S \right|}{|S|} \cdot \log_2 \frac{\left| \sigma_{a_i=v_{i,j} \vee y=c_{k,j}} S \right|}{|S|}}$$

Uzaklık ölçütü, kazanç oranında olduğu gibi safsızlık ölçütünü normalleştirir, fakat bunu farklı bir şekilde gerçekleştirir (Rokach ve Maimon, 2014: 64).

AUC Bölünme Kriteri:

AUC metrik değerinin bir bölünme kriteri olarak kullanılabileceği Ferri C. vd. tarafından 2002 yılında belirtilmiştir. ROC (Receiver Operator Characteristic) eğrisinin dışbükey gövdesinin altında maksimum alanı elde eden öznitelik bölünme kriteri olarak seçilir. AUC tabanlı bölme kriterinin, ROC eğrisi altında kalan alan ve sınıflandırma doğruluğu açısından diğer bölme kriterlerine göre daha iyi bir performans sergilediği gösterilmiştir. Safsızlık kriterlerinin aksine bu kriter, bölünmeden sonra çocuk düğümlerin ağırlıklı safsızlığı ile ebeveyn düğümün safsızlığı arasında bir karşılaştırma yapmaz (Ferri vd., 2002: 142; Rokach ve Maimon, 2014: 66).

2.2.3.1.4 İkili Kriterler

İkili karar ağaçları sıklıkta çeşitli nedenler için kullanılır. Birçok öznitelik doğal olarak ikilidir ve ikili ağaçları yorumlamak kolaydır (Aggarwal, 2015: 95). İkili kriterler, ikili ağaçları oluştururken kullanılır. Bu ölçütler, girdi öznitelik alanının iki alt alana bölünmesine dayalıdır. “ $\beta(a_i, \text{dom}_1(a_i), \text{dom}_2(a_i), S)$ ”, S örneği üzerinde a_i öz niteliği için ikili ölçüt olarak tanımlansın, burada “ $\text{dom}_1(a_i)$ ” ve “ $\text{dom}_2(a_i)$ ” a_i ’nin iki alt alanıdır. Öznitelik alanını, birbirini dışlayan ve ayrıntılı iki alt alana optimal bölünmesi için elde edilen değer özniteliklerin karşılaştırılması için kullanılır (Rokach ve Maimon, 2014: 64-65).

$$\beta^*(a_i, S) = \max_{\forall \text{dom}_1(a_i); \text{dom}_2(a_i)} \beta(a_i, \text{dom}_1(a_i), \text{dom}_2(a_i), S)$$

Twoing Kriteri:

Twoing bölünme kriteri, Gini İndeksine bir alternatif olarak Sınıflandırma ve Regresyon Ağaçlarında (CART) tanımlanmıştır ve çok sınıflı problemlerde kullanır. Her düğümden sınıflar, ayrık ve karşılıklı ayrıntılı sınıflar içeren iki süper sınıfa ayrılır (Kantardzic, 2011: 190). Hedef özniteliğin alanı nispeten geniş olduğunda Gini indeksi sorunlar ile karşılaşabilir. Bu gibi durumlarda, Twoing kriteri olarak adlandırılan ikili kriterler kullanmak da mümkündür. Bu kriter aşağıdaki gibi tanımlanmıştır;

$$Twoing\beta(a_i, dom_1(a_i), dom_2(a_i), S)$$

$$= 0,25 \cdot \frac{|\sigma_{a_i \in dom_1(a_i)} S|}{|S|} \cdot \frac{|\sigma_{a_i \in dom_2(a_i)} S|}{|S|} \cdot \left(\sum_{c_i \in dom(y)} \left| \frac{|\sigma_{a_i \in dom_1(a_i) \vee y=c_i} S|}{|\sigma_{a_i \in dom_1(a_i)} S|} - \frac{|\sigma_{a_i \in dom_2(a_i) \vee y=c_i} S|}{|\sigma_{a_i \in dom_2(a_i)} S|} \right| \right)^2$$

Hedef özniteliği ikili olduğunda, Gini İndeksi ve Twoing Kriteri eşdeğerdir. Çok sınıflı problemler için Twoing kriteri, eşit, dengeli şekilde bölünmüş bölünmelere sahip öznitelikleri tercih etmektedir (Rokach ve Maimon, 2014: 65). Twoing bölünme kriteri bize daha dengeli ağaçları oluşturmak için izin verse de, bu algoritma Gini İndeksine göre daha yavaş çalışır (Kantardzic, 2011: 190).

Ortogonal Kriteri:

Fayyad ve Irani tarafından sunulan bu ikili bölünme kriterinde, bölünme kriteri olarak ikili bölünmelerden gelen sınıf olasılığı vektörleri arasındaki açının kosinüsünü kullanılmıştır (Aggarwal, 2015: 95).

Bu ikili bölünme kriteri şu şekilde tanımlanmaktadır;

$$ORT(a_i, dom_1(a_i), dom_2(a_i), S) = 1 - \cos\theta(P_{y,1}, P_{y,2})$$

Burada $\theta(P_{y,1}, P_{y,2})$, $P_{y,1}$ ve $P_{y,2}$ vektörleri arasındaki açıdır. Bu vektörler sırasıyla, $\sigma_{a_i \in dom_1(a_i)} S$ ve $\sigma_{a_i \in dom_2(a_i)} S$ bölümlerindeki hedef özniteliğin olasılık dağılımlarını temsil etmektedir. Ortogonal kriterin, belirli problem kümeleri için Gini indeksi ve Bilgi Kazancı kriterlerine göre daha iyi performans sergilediği gösterilmiştir (Fayyad ve Irani, 1992: 108; Rokach ve Maimon, 2014: 65). Bu ölçüt $[0,1]$ aralığında değer alırken maksimum değerini ise vektörler ortogonal olduğunda alacaktır, başka bir ifade ile bu vektörlerin dik olması demektir. Vektörler ortogonal olduğunda aralarındaki açı 90 derece olacaktır, $\cos(90^\circ)=0$ olduğundan maksimum değer olan “1”e ulaşacaktır. Minimum değer ise “0”dır, bu da vektörler arasındaki açının “0” olmasıdır ($\cos(0^\circ)=1$). Bu durumda ise vektörler birbirine paraleldir (Fayyad ve Irani, 1992: 108).

Kolmogorov-Smirnov Kriteri:

Friedman ve Rounds tarafından ikili sınıf problemlerini incelemek için Kolmogorov-Smirnov uzaklığına dayalı ikili bir ölçüttür sunmuşlardır (Barros vd., 2015: 16). İkili hedef özniteliğini, $dom(y) = \{c_1, c_2\}$ olarak varsayarsak, bu kriteri şu şekilde tanımlayabiliriz;

$$KS(a_i, \text{dom}_1(a_i), \text{dom}_2(a_i), S) = \left| \frac{|\sigma_{a_i \in \text{dom}_1(a_i) \text{ ve } y=c_1} S|}{|\sigma_{y=c_1} S|} - \frac{|\sigma_{a_i \in \text{dom}_2(a_i) \text{ ve } y=c_2} S|}{|\sigma_{y=c_2} S|} \right|$$

Kolmogrov-Smirnov kriteri, çoklu sınıf ve eksik veri değerlerine sahip hedef öznitelikler ile başa çıkmak için Utgoff ve Clouse (1996) tarafından genişletilmiştir (Rokach ve Maimon, 2014: 66).

2.2.4 Durdurma Kriterleri

Yukarıdan aşağıya ağaç endükleyicilerinde her bölünme yeni düğümler üretir ve bu yeni düğümler özyineli olarak yeni bölünme noktalarını gösterir. Düğüm sınıf saflığına ulaştığında veya düğüm sadece tek bir öge içerdiğinde bölünme algoritması duracaktır (Aggarwal, 2015: 100). Ayrıca bir ağacın büyüme aşaması, durdurma kriteri tetiklenene kadar sürer (Barros vd., 2015: 29).

Aşağıda belirtilen maddeler literatürde kullanılan durdurma kriterleridir.

1. Eğitim kümesindeki tüm örnekler tek bir “y” değerine aittir. Yani sınıf homojenliğine ulaşmak da denebilir. Tüm örnekler aynı sınıfa ait olduğunda bu homojenliğe ulaşılmış olur, böylece bu düğümün yeniden bölünmesine gerek yoktur.
2. Öznitelik homojenliğine ulaşma; belirli bir düğüme ulaşan tüm örnekler aynı öznitelik değerlerine sahip olduğunda (aynı sınıf değeri gerekli olmasada) bölünme durur.
3. Bir parametre veri parçalanma problemini önlemek için “terminal olmayan bir düğüm için örneklerin minimum sayısını” belirtebilir.
4. Maksimum ağaç derinliğine ulaştığında bölünme durur.
5. Terminal düğümdeki durumların (olguların) sayısı ebeveyn düğümlerin için minimum durumların sayısından daha azdır.
6. Eğer düğüm bölünürse, bir veya daha fazla alt düğümdeki durumların sayısı çocuk düğümler için minimum durumların sayısından daha az olacaktır.
7. En iyi bölme kriteri, belli bir eşik değerden çok büyük değildir.
8. Bir parametre zayıf bölünmeleri önlemek için “bölünme kriter eşiği” olarak belirlenebilir.
9. Veri seti boyutu minimuma ulaştığında bölünme durur.
10. Düğüm sayısı maksimuma ulaştığında bölünme durur.

(Aggarwal, 2015: 100; Barros vd., 2015: 29; Rokach ve Maimon, 2014: 69)

2.2.5 Karar Ağaçlarının Budanması

Bir veya daha fazla alt ağacın atılması (çıkarılması) ve alt ağacın yapraklarla değiştirilmesi karar ağacını sadeleştirir ve bu karar ağacı budanmasında temel görevdir (Kantardzic, 2011: 184). Bir karar ağacı oluşturulduktan sonra eğitim verisi içerisindeki gürültüden veya aykırı değerlerden dolayı bazı dallar anomaliler yansıtır. Ağaç budama yöntemleri (pruning methods) aşırı uyum (overfitting) problemini çözmek için kullanılan yöntemlerdir. Bu yöntemler en az güvenilir dalları çıkarmak için istatistiksel yöntemler kullanır (Han ve Kamber, 2006: 304). Genellikle budama yöntemlerinde herhangi bir düğüm için, elde edilen alt karar ağacının yerine bir son düğüm yerleştirilir. Bu işlem, kestirim hata yüzdesi belirli bir tolerans altında ve alt karar ağacı için son düğümden yüksek olduğu durumda yapılır (Irmak, 2009: 40). Yaprak ile alt ağacı değiştirme işleminde kullanılan algoritmanın, tahmin hata oranını düşürmek ve sınıflandırma model kalitesini artırması beklenir (Kantardzic, 2011: 184).

Sıkı bir durdurma kriterileri uygulamak, küçük ve aşırı uyumlu olmayan, yetersiz uyumlu (underfitted) karar ağaçları oluşturma eğilimi gösterir. Diğer yandan, gevşek durdurma kriterleri uygulanması, eğitim kümesine aşırı uyumlu büyük, geniş karar ağaçları oluşturma eğilimindedir. Bu sorunu çözmek için Breiman vd. (1984) bir budama yöntemi geliştirmişlerdir. Gevşek bir durdurma kriteri ile karar ağacının eğitim kümesine aşırı uyumluluğu sağlanır, sonra aşırı uyumlu olan bu ağacın genelleme doğruluğuna katkısı olmayan alt dalları kaldırılarak küçük bir ağaç olacak şekilde kesilir. Budama metodlarının uygulanmasının bir karar ağacının genelleme performansını arttırdığı çeşitli çalışmalarda (özellikle gürültülü alanlarda) gösterilmiştir (Rokach ve Maimon, 2014: 69). Budanmış ağaçlar küçük ve daha az kamaşık olma eğilimi gösterirler bu sebeple anlaşılması daha kolaydır (Han ve Kamber, 2006: 304).

Karar ağaçlarında budama işleminin nasıl çalıştığına dair iki farklı yaklaşım vardır bunlar, prepruning (forward pruning) ve postpruning (backward pruning)'dir (Bramer, 2007: 125). Karar ağaçlarının budanması için çeşitli teknikler vardır. Bu teknikler yukarıdan aşağıya veya aşağıdan yukarıya doğru düğümlerin ziyareti ile gerçekleşir (Rokach ve Maimon, 2014: 70).

Literatürde birçok budama metodu bulunmaktadır, bunlar;

- Azaltılmış Hata Budaması (REP, Reduced Error Pruning)
- Maliyet-Karmaşıklık Budama (CCP, Cost Complexity Pruning)
- Minimum Hata Budaması (MEP, Minimum Error Pruning)
- Kötümser Budama (PEP, Pessimistic Pruning)

- Hata Tabanlı Budama (EBP, Error Based Pruning)
- Optimal Budama (OP, Optimal Pruning)
- Minimum Tanım Uzunluğu Budama (MDL, Minimum Description Length Pruning)
- Minimum Mesaj Uzunluğu (MML, Minimum Message Length)
- Kritik Değer Budaması (CVP, Critical Value Pruning)
- Derinlik Safsızlığı Budama (DIP, Depth Impurity Pruning)
- Doğrulama Tabanlı Budama (VBP, Validation Based Pruning)

olarak sayılabilir (Barros vd., 2015: 30-35; Grabczewski, 2014: 62-63; Aggarwal, 2015: 97-99; Rokach ve Maimon, 2014: 70-74).

Çeşitli çalışmalarda farklı budama tekniklerinin performansları kıyaslanmıştır. Sonuçlar, maliyet karmaşıklığı budaması, azaltılmış hata budaması gibi bazı metotların aşırı budama eğiliminde olduğunu göstermiştir, başka bir ifade ile daha küçük ancak daha az doğru (daha az hassas, accurate) karar ağacı oluştururlar. Hata tabanlı budama, kötümser hata budaması ve minimum hata budaması gibi diğer metotlar ise daha az budamaya yanlıdır. Karşılaştırmaların çoğunda, hiç bir budama metodunun diğer kalan budama metotlarına göre daha iyi, üstün bir performans göstermediği anlaşılmıştır (Rokach ve Maimon, 2014: 73).

2.2.5.1 Azaltılmış Hata Budaması

Azaltılmış Hata Budaması (REP) metodu karar ağaçlarının budanması için basit bir strateji olarak Quinlan tarafından öne sürülmüştür (Barros vd., 2015: 30). Azaltılmış budama yöntemi, ağacın büyümesi ve budanması için aynı iyilik kriterlerini (goodness criteria) kullanırken iki safha için farklı veri örneklerini kullanılır. Başlangıç ağacı endüksiyon öncesi eğitim veri kümesi, büyüme veri kümesi ve budama veri kümesi olmak üzere ayrılmıştır. Başlangıç ağacının endüksiyonu büyüme veri kümesini kullanılarak yapılır (Aggarwal, 2015: 98). Azaltılmış Hata Budamanın, T 'den verilen bir altağacın iyiliğini (goodness) değerlendirmek için budama setini kullanmasındaki amaç, budama setindeki sınıflandırma hatası dikkate alınarak her terminal olmayan " t " düğümünü değerlendirmektir. Altağac $T^{(t)}$ ile yaprak düğümü değiştirdiğimizde eğer hata azalırsa o zaman $T^{(t)}$ mutlaka budanmalıdır. Quinlan bir koşul öne sürer: Bir düğüm " t " eğer budama setinde daha alt sınıflandırma hatasına uyum sağlayan ağaç içeriyorsa budama yapılmaz. Bu koşulun uygulanabilir sonucu şudur ki, REP aşağıdan yukarı tasarımda çalıştırılmalıdır. REP budanmış T ağacı ilginç ve elverişli bir özellik sunar. Orijinal T ağacının budamasından sonuçlanan ağaç, en küçük ve doğru ağaçtır. Bu elverişli özelliğin yanı sıra REP'in diğer bir avantajı doğrusal karmaşıklığıdır. Böylece her bir düğüm, T de sadece bir kere ziyaret edilir. Özellikle küçük

veri setleri için kısmen ciddi olan bir dezavantaj budama seti kullanma ihtiyacıdır (Barros vd., 2015: 30-31). Bu yöntemin sayılabilecek diğer dezavantajından birisi büyük miktarda eğitim verisine ihtiyaç duymasıdır. Bir diğeri ise büyüme ve budama için aynı kriterleri kullanarak yetersiz budama eğilimi göstermesidir (Aggarwal, 2015: 98).

2.2.5.2 Maliyet Karmaşıklık Budama

Maliyet Karmaşıklık Budaması çok adımlı bir budama stratejisi olup, tek düğüm ile alt ağaçların değiştirilmesini hedeflemektedir (Aggarwal, 2015: 97). Aynı zamanda en zayıf bağlantı budama veya hata karmaşıklığı budaması olarak bilinir (Rokach ve Maimon, 2014: 70). Maliyet Karmaşıklık Budama, CART sisteminin postpruning stratejisidir ve iki adımdan oluşur (Barros vd., 2015: 33). İlk aşamada $T_0, T_1, T_2, \dots, T_k$ ağaç dizisi, eğitim verisi üzerinde oluşturulur, burada T_0 budama öncesindeki orijinal ağaç, T_k ise kök ağaçtır. İkinci aşamada, ağaçlardan bir tanesi genelleme hatasının tahminine dayalı olarak budanmış ağaç olarak seçilir. T_{i+1} ağacı, uygun yapraklı T_i öncel ağacındaki bir ya da daha fazla alt ağaçların değiştirilmesi ile elde edilir. Budanmış alt ağaçlar, her bir budanmış yaprak düğüm için en düşük hata artış oranını elde edenlerdir:

$$\alpha = \frac{\varepsilon(\text{pruned}(T, t), S) - \varepsilon(T, S)}{|\text{leaves}(T)| - |\text{leaves}(\text{pruned}(T, t))|}$$

S örneği üzerindeki T ağacının hata oranı: $\varepsilon(T, S)$

T 'deki yaprakların sayısı: $|\text{leaves}(T)|$

T 'deki uygun bir yaprak ile t düğümünün yer değiştirilmesi ile elde edilen ağaç: $\text{pruned}(T, t)$

İkinci aşamada budanmış ağaçların genelleme hataları tahmin edilir. Daha sonra en iyi budanmış ağaç seçilir. Eğer veri seti yeterince büyükse, veri setini eğitim seti ve budama seti olacak şekilde bölünmesi önerilir. Ağaçlar, eğitim seti kullanılarak inşa edilir ve budama setinde değerlendirilir. Veri kümesi yeterince büyük değilse, hesaplama karmaşıklığı etkilerine rağmen, çapraz doğrulama yönteminin kullanılması önerilir (Rokach ve Maimon 2014: 70). Maliyet Karmaşıklık Budama hem ağacın karmaşıklığını hemde hataların sayısını hesaba katar. Ağacın büyüklüğü ağacın karmaşıklığını temsil etmek için kullanılır (Cai, 2006: 14-15).

2.2.5.3 Minimum Hata Budaması

Minimum Hata Budaması, Niblett ve Bratko tarafından önerilmiştir. Bağımsız bir veri kümesi üzerinde beklenen hata oranını en aza indiren bir tek ağaç arayan aşağıdan yukarıya

doğru bir yaklaşımdır (Cai, 2006: 16). Safsızlık kriteri kullanan Minimum Hata Budama (Minimum Error Pruning-MEP), ağacın büyüme aşaması gibi çalışmaktadır fakat bölünme yerine birleştirme kullanır. Bir yaprak düğümün ebeveyn ögesinden başlayarak, onun çocuklarının hata oranlarının ağırlıklı toplamı boyutu ile beklenen hata oranını karşılaştırır. Ebeveyn düğümde beklenen hata oranı düşük ise alt ağacı yaprak haline dönüştürülür. Bu işlem tüm tüm yaprakların ebeveyn düğümlerine ulaşınca kadar tekrarlanır, sonunda ağaç optimize edilmiş olur (Aggarwal, 2015: 98). J. Mingers bu yöntemin çeşitli dezavantajları olduğuna işaret etmiştir. Birincisi tüm sınıflar olasılıkla eşittir, bu pratikte nadiren doğrudur. İkincisi, bu yöntem sadece tek ağaç üretir. Üçüncü olarak, sınıf sayısı kuvvet bir şekilde budama derecesini etkiler, bu da kararsız sonuçlara yol açar (Cai, 2006: 17).

2.2.5.4 Kötümser (Hata) Budama

Kötümser (Hata) Budama (Pessimistic Error Pruning-PEP) Quinlan tarafından önerilmiştir, eğitim kümesi hem ağacın oluşturulmasında hem de budanmasında kullanılır (Barros vd., 2015: 31). Bu nedenle PEP yöntemi ikinci bir veri seti gerektirmez (Aggarwal, 2015: 98). PEP, budama seti veya çapraz doğrulama gereksinimi yerine kötümser istatistik korelasyon testi kullanır (Rokach ve Maimon, 2014: 71). Belirgin hata oranı yani eğitim setindeki hata oranı, iyimser yanlı ve en iyi budanmış ağacı seçmek için kullanılamaz. Bu nedenle Quinlan daha gerçekçi olan binom dağılımı için süreklilik düzeltmesini tanıtmıştır (Esposito vd., 1997: 478).

$$\varepsilon'(T, S) = \varepsilon(T, S) + \frac{|leaves(T)|}{2 \cdot |S|}$$

Ancak bu düzeltme hala iyimser bir hata oranı üretmektedir. Sonuç olarak Quinlan bir iç düğüm t'nin hata oranı referans ağaca göre bir standart hata içerisinde ise bu iç düğüm t'nin budanmasını önermektedir,

$$\varepsilon'(pruned(T, t), S) \leq \varepsilon'(T, S) + \sqrt{\frac{\varepsilon'(T, S) \cdot (1 - \varepsilon'(T, S))}{|S|}}$$

Son durum, oranlar için istatistiksel güven aralığına dayanmaktadır. Genellikle son durum kullanılır. T alt ağacı temsil eder, t ise T'nin bir iç düğüm olan köküdür. S ise, t düğümüne atıfta bulunan eğitim setinin bölümünü gösterir. Eğer bir iç düğüm budanmış ise soyundan gelen düğümler budama işleminden çıkarılır böylece nispeten hızlı bir budama ile

sonuçlanır (Rokach ve Maimon, 2014: 71-72). Kötümser Budama süreci yukarıdan aşağıya doğru çalışan bir yöntemdir ve yüksek hızlı çalışmaktadır (Esposito vd., 1997: 479). Bu yöntem (PEP), Azaltılmış Hata Budama (REP) yöntemine göre çok daha hızlı çalışmaktadır ve aynı zamanda daha yüksek doğruluk sağlamaktadır (Cai, 2006: 14).

2.2.5.5 Hata Tabanlı Budama

Quinlan tarafından geliştirilen (Barros vd., 2015: 34) Hata Tabanlı Budama (Error Based Pruning-EBP) beklenen hata oranının çok daha kötümser tahminine dayalı olduğu için kötümser budamanın (PEP) iyileştirilmiş halidir ve C4.5 algoritmasına uygulanmıştır (Esposito vd., 1999: 292). Kötümser budamada olduğu gibi hata oranı, oranlar için istatistik güven aralığının üst sınırı kullanılarak tahmin edilir.

$$\varepsilon_{UB}(T, S) = \varepsilon(T, S) + Z_{\alpha} \cdot \sqrt{\frac{\varepsilon(T, S) \cdot (1 - \varepsilon(T, S))}{|S|}}$$

Burada;

S eğitme kümesi üzerinde, T ağacının yanlış sınıflandırma oranı: $\varepsilon(T, S)$

Standart normal birikimli dağılımın tersi: Z, istenilen anlamlılık düzeyi: α

“t” düğümü tarafından kökleşmiş alt ağaç: $subtree(T, t)$

“t”nin en sık görülen çocuk düğümü: $maxchild(T, t)$, (yani S’deki örneklerin çoğu bu belli çocuk düğüme ulaşır.)

“t” düğümüne ulaşan S’deki tüm örnekler: S_t dir.

Prosedür, tüm düğümler üzerinde aşağıdan yukarıya doğru sırayla ziyareti gerçekleştirir ve aşağıdaki değerleri karşılaştırır;

1. $\varepsilon_{UB}(subtree(T, t), S_t)$
2. $\varepsilon_{UB}(pruned(subtree(T, t), t), S_t)$
3. $\varepsilon_{UB}(subtree(T, maxchild(T, t)), S_{maxchild(T, t)})$

En küçük değere göre bu yöntem ya ağacı olduğu gibi bırakır ya “t” düğümünü budar, ya da $maxchild(T, t)$ tarafından köklü alt ağaç ile “t” düğümünü yer değiştirir (Rokach ve Maimon, 2014: 72). Hata Tabanlı Budama, Kötümser Hata Budamanın aksine aşağıdan yukarıya doğru çalışır ve Kötümser Hata Budamasına göre daha kötümserdir ve küçük ağaçlar üretmektedir (Barros vd., 2015: 34; Grabczewski, 2014: 59-60). EBP ve PEP, ağaçların yapılandırılması ve basitleştirilmesi için eğitim setinde bilgi kullanır (Esposito vd., 1999: 292). Hata Tabanlı Budamanın önemli bir özelliği ve avantajı, ilginç alt dalları göz ardı

etmeden gereksiz, kullanışsız dalların budanmasına izin veren aşılama operasyonudur (grafting operation) (Barros vd., 2015: 35).

2.2.5.6 Kritik Değer Budama

Kritik Değer Budama (Critical Value Pruning-CVP) stratejisi, Maliyet Karmaşıklık Budama stratejisi gibi aynı aşamalı yaklaşımı izler. Fakat budama için yeni bir iyilik ölçüsü tanımlamak yerine, ağacın büyümesi için kullanılan aynı kriter yeniden kullanır. Büyüme aşamasında her düğüm için bölünme kriterleri değerlerini kaydeder (Aggarwal, 2015: 98). Bu yöntemde kritik değer olarak adlandırılan eşik değeri bir düğümün önemini veya gücünü hesaplamak için ayarlanır, düğüm kritik değere ulaşmadığı sürece budama yapılır. Bir düğüm budama koşullarını sağlasa bile, bu düğümün çocukları budama koşulunu sağlamadıkları zaman bu düğüm tutulur yani budanmaz. Eğer daha büyük bir kritik değer seçilirse daha şiddetli budama olacağından dolayı daha küçük bir ağaç elde edilecektir (Cai, 2006: 17-18). Mingers kritik değer budamayı iki ana adımda, aşağıdaki gibi açıklamaktadır (Barros vd., 2015; s.33, Esposito vd., 1997: 480).

1. Kritik değerleri artırmak için alt ağacı budamak (Budanmış ağaçların kümesini oluşturma).
2. Budanmış ağaç dizisi arasından, bir bütün olarak ağacın önemini (G istatistiğine dayalı olarak) ve tahmin yeteneğini veya doğruluğunu (budama setine dayalı olarak) ölçerek en iyi ağacı seçmek.

Bu yöntemin dezavantajı yetersiz budamaya (underprune) güçlü eğilimi olması ve düşük öngörü doğruluğuna sahip ağaçları seçmesidir (Cai, 2006: 18). Diğer bir dezavantaj ise Kritik Değer Budamanın, Azaltılmış Hata Budamasında olduğu gibi bir budama kümesine (budama seti) ihtiyacının olmasıdır (Barros vd., 2015: 33).

2.2.5.7 Optimal Budama

Breiman vd.. Optimal Budamanın matematiksel özelliklerini doğrulamış, ayrıca “k” aday arasından, belirli bir optimal budanmış alt ağacı seçmek için bir algoritma tanıtmışlardır (Cai, 2006: 18). Bratko ile Bohanec çalışmalarında, daha önceki yöntemlerin optimal altı çözümlere yol açtığını göstermiş ve optimum budamayı bulma konusunu başka bir şekilde ele alarak Optimal Budamayı garanti eden bir algoritma (OPT) tanıtmışlardır. Dinamik programlamaya dayalı olan OPT algoritması, $O = (s^2)$ sürede (Burada s: Başlangıç karar ağacının yapraklarının sayısını göstermektedir) budanmış ağaç dizisini üretmektedir. Almuallim Optimal Budamayı (OPT) düzelterek yeni bir versiyonunu oluşturmuştur, bu yeni algoritma OPT-2 olarak isimlendirilmektedir (Almuallim, 1996: 348).

2.2.5.8 Minimum Tanım Uzunluğu Budama

MDL (Minimum Description Length), düğümün genelleştirilmiş doğruluğunu değerlendirmek için kullanılabilir. Bu yöntem ağacı kodlamak için gereken bit sayısı vasıtasıyla karar ağacının boyutunu ölçmektedir (Rokach ve Maimon, 2014: 73). MDL presibi “en iyi” sınıflandırma ağacı olduğunu belirtmektedir, bu ağaç en az sayıda bit ile kodlanmıştır. Bu nedenle ilk önce, herhangi bir ikili karar ağacını kodlamayı sağlayan kodlama şemasını tanımlamak gerekmektedir. Bir kodlama şeması göz önüne alındığında, en az kod uzunluğuna sahip alt ağacı seçerek belirli bir sınıflandırma ağacını budanabilir (Ye, 2003: 15). Bir “t” yaprağındaki bölünmenin maliyeti;

$$Cost(t) = \sum_{c_i \in dom(y)} |\sigma_{y=c_i} S_t| \cdot \ln \frac{|S_t|}{|\sigma_{y=c_i} S_t|} + \frac{|dom(y)| - 1}{2} \ln \frac{|S_t|}{2} + \ln \frac{\pi^{\frac{|dom(y)|}{2}}}{\Gamma(\frac{|dom(y)|}{2})}$$

şeklinde tahmin edilebilir. “t” düğümüne ulaşmış örnekler S_t , bir iç düğümün bölünme maliyeti kendi alt (çocuk) düğümlerinin maliyeti toplamına dayalı olarak hesaplanır (Rokach ve Maimon, 2014: 73).

Ayrıca, Wallaces ve Patrick (1993) Minimum Mesaj Uzunluk Budama Yöntemini (MML) önermiştir, Kearns ve Mansour (1998) budama algoritmasını teorik olarak kanıtlamıştır (Rokach ve Maimon, 2014: 73). Bir diğer budama yöntemi Fournier ve Crémilleux tarafından sunulan Derinlik Safsızlık Budamadır. (DIP-Depth Impurity Pruning). Karar ağaçları literatüründe ayrıca Doğrulama Tabanlı Budama adında bir budama tekniği de bulunmaktadır (Grabczewski, 2014: 62-63).

2.2.6 Karar Ağacı Endükleyicileri (Algoritmaları)

Karar ağaçlarını oluşturmak için kullanılan birçok algoritma vardır. Bunlar, Quinlan (1986) tarafından geliştirilen ID3, yine Quinlan (1993) tarafından geliştirilen ve ID3’ün geliştirilmiş bir versiyonu olan C4.5, Breiman (1984) tarafından geliştirilen CART, Kass (1980) tarafından geliştirilen CHAID, Loh ve Shih (1997) tarafından geliştirilen QUEST olarak sayılabilir. (Rokach ve Maimon, 2014: 77-80)

Tablo 2.1, Karar ağacı endükleyicilerinin genel bir değerlendirmesini tablo olarak göstermektedir.

Tablo 2.1 Karar Ağacı Endükleycilerinin Genel Değerlendirilmesi

Algoritma	Girdi Değişkeni	Çıktı Değişkeni	Tahmin Türü	Meydana Gelen Dal Sayısı	Bölünme Kriteri
CHAID	Kategorik, Nümerik	Kategorik, Nümerik	Sınıflandırma, Regresyon	≥ 2	Ki Kare, F, Likelihood Ratio
QUEST	Kategorik, Nümerik	Kategorik	Sınıflandırma	$= 2$	Ki Kare, F, Levene
CART	Kategorik, Nümerik	Kategorik, Nümerik	Sınıflandırma, Regresyon	$= 2$	Gini, Twoing, Least Squared Deviation
C4.5/C5.0	Kategorik, Nümerik	Kategorik	Sınıflandırma	≥ 2	Gain Ratio

Kaynak: Palmer vd., 2011: 379

2.2.6.1 ID3, C4.5 ve C5.0 Algoritmaları

J. Ross Quinlan tarafından Sydney Üniversitesinde geliştirilmiş olan ID3 algoritması, ilk olarak 1975 yılında yine Quinlan tarafından Makine Öğrenmesi isimli kitabında sunulmuştur. ID3 algoritması, Kavram Öğrenme Sistemi (CLS-Concept Learning System) algoritmasına dayanmaktadır (Prabhu ve Venatesan, 2007: 25). ID3 algoritması karar ağaçlarının oluşturulmasında kullanılan çok basit bir algoritmadır ve en önemli özelliği basitliğidir. Bölünme ölçütü olarak Bilgi Kazancını (Information Gain) kullanır, tüm örnekler tek hedef özelliğin bir değerine ait olduğunda ya da en iyi bilgi kazancı sıfırdan büyük olmadığında ağacın büyümesi durmaktadır. ID3 algoritmasında budama prosedürü bulunmamakla beraber eksik değer ve nümerik veri ile işlem yapmamaktadır. ID3 algoritmasının bazı dezavantajları vardır, bunlar;

- ID3 optimal çözümü garanti etmez, aç gözlü strateji (Greedy Strategy) kullandığı için yerel optimum değerlere sıkışabilirsiniz. Yerel optimumlardan kaçınmak için arama sırasında geri izleme (backtracking) kullanılabilir.
- ID3 eğitim verilerine aşırı uyumlu olabilir. Aşırı uyumluluktan kaçınmak için, büyük ağaçlar yerine küçük ağaçlar tercih edilmelidir. Algoritma genellikle küçük ağaçlar üretir, fakat her zaman mümkün olan en küçük ağacı üretmeyebilir.
- ID3 nominal veriler için tasarlanmıştır. Sürekli veriler, nominal hale getirildikten sonra kullanılabilir.

C4.5, ID3 gibi Quinlan tarafından geliştirilmiştir ve ID3'ün geliştirilmiş bir türevidir. Bölünme ölçütü olarak Kazanç Oranını kullanmaktadır (Gain Ratio). Bölünmesi gereken örnek sayısı, belirli bir eşik değerinden aşağıda olduğunda bölünme işlemi sona ermektedir. C4.5 algoritmasında ağacın büyüme aşamasından sonra Hata Tabanlı Budama (Error Based Pruning) gerçekleştirilir. C4.5 sayısal verileri işleyebilir. Düzeltilmiş kazanç oranı kriterini kullanarak, eksik değerler içeren bir eğitim setinden endüklenebilir. C4.5 algoritmasında ID3'e göre çeşitli iyileştirmeler yapılmıştır, bunlar; (Grabczewski, 2014: 18; Rokach ve Maimon, 2014: 77-78)

- C4.5 bir budama prosedürü kullanır
- C4.5 eksik veriler ile çalışabilir.
- Sürekli öznitelikler ile çalışabilme. C4.5 sürekli öznitelikleri, özneliğin değer aralığını iki alt kümeye bölerek (ikili bölme) ele alır. Özellikle, kazanç oranı kriterlerini maksimuma yükselten en iyi eşiği arar. Eşiğin üzerindeki her değer ilk alt kümeyi oluşturur ve tüm diğer değerler ikinci alt kümeyi oluşturur.
- Düğüm safsızlık ölçümü modifiyesi

ID3 ve C4.5 sezgisel yukarıdan aşağıya doğru aç gözlü (geri dönülmez) arama yaklaşımını benimser, karar ağaçları yukarıdan aşağıya doğru özyinelemeli böl ve yönet şeklinde oluşturulur (Han ve Kamber, 2006: 292; Wong ve Leung, 2002: 10-12). ID3, gürültü nedeniyle aşırı uyumu önlemek için ki-kare testi kullanırken C4.5, aşırı uyumu önlemek için ki-kare testini kullanmaz, aksine C4.5 ağacın büyümesine olanak sağlar ve daha sonra gereksiz dalları budar (Wong ve Leung, 2002: 10-12).

C4.5 algoritması 1997 yılında değiştirilmiştir, meydana gelen bu yeni algoritma C5.0'dır. C5.0 "boosting" işlemi adı verilen tahmin doğruluğu performansını artıran bir prosedürü desteklemektedir. Ayrıca C5.0'a tarih gibi yeni veri tipleri, "uygulanamaz" değerler ile çalışma, değişken yanlış sınıflandırma maliyetleri kavramı, öznitelikleri ön filtreleme mekanizması gibi özellikler eklenmiştir (Kantardzic, 2011: 188). C5.0 algoritması bellek ve hesaplama süresi açısından C4.5'den çok daha verimlidir. Ayrıca Weka veri madenciliği aracında J48, C4.5 algoritmasının açık kaynak kodlu bir Java uygulamasıdır (Rokach ve Maimon, 2014: 78).

2.2.6.2 CART (C&RT) Algoritması

CART (Classification and Regression Tree), Breiman ve arkadaşları (Leo Breiman, Jerome Friedman, Richard Olshen ve Charles Stone (BFOS)) tarafından ikili ağaçların karakterize edilmesi için 1984 yılında geliştirilmiştir. Geliştirdikleri bu algoritma makine

öğrenmesi, parametrik olmayan istatistik, veri madenciliği ve yapay zekâ gelişiminde önemli bir kilometre taşı temsil etmektedir. CART, hedef ve belirleyici değişkenlerin nominal ve sürekli olduğu verileri işleyebilen, ikili-öz yinelemeli bir algoritmadır. Veri ham hali ile işlenir ve herhangi bir ağırlıklandırma gerekli değildir, önerilmez. CART’da veri, kök düğümünden başlayarak iki alt düğüme ayrılır, her alt düğüm kendi içinde iki alt düğüme ayrılır. Ağaç herhangi bir durdurma kuralı kullanmadan maksimum boyuta ulaşır, veri eksikliğinden dolayı bölünme gerçekleşmez ve ağacın büyüme süreci böylelikle durmuş olur. Maksimum boyutlu ağaç, Maliyet Karmaşıklığı Budama yöntemi ile köke doğru geri biçimde budanır (Wu ve Kumar, 2009: 179-181). CART, çoklu bölünmelere izin veren ilk sistem olup, yerel optimal noktalardan kaçmak için hiçbir yerleşik mekanizması olmayan deterministik bir algoritmadır (Barros vd., 2015: 25). Bölünme ölçütü olarak Gini İndeksi kullanılır. ID3 ve C4.5 gibi açgözlü arama yaklaşımını benimser ve karar ağaçları yukarıdan aşağıya doğru öz yinelemeli böl-yönet şeklinde inşa edilmiştir (Han ve Kamber, 2006: 292: 302). Çok sınıflı problemlerde ise Twoing Bölünme Ölçütünü kullanmaktadır (Kantardzic, 2011: 190). CART’ın önemli bir özelliği Regresyon Ağaçları oluşturmak için kullanılmasıdır. Regresyon Ağaçlarında yapraklar, bir sınıfa değil gerçek bir sayıyı öngörür. Regresyon durumunda CART, tahmin hata karesini minimize eden bölünmeleri arar (Rokach ve Maimon, 2014: 79). Regresyon ağaçları, sınıflandırma ağaçlarına göre biraz daha basittir çünkü CART’da kullanılan büyüme ve budama kriterleri aynıdır (Kantardzic, 2011: 191).

Sadece bir ağaç üretmek için değil, her biri en uygun ağaç olmaya aday iç içe budanmış ağaç dizisi üretmek için tasarlanmış olan CART mekanizması otomatik sınıf dengeleme, otomatik eksik değer işleme yapabilirken, maliyet duyarlı öğrenme, dinamik özellik inşası ve olasılık ağaç tahminine izin vermektedir (Wu ve Kumar, 2009: 181).

CART’ın istatistikçiler için bir değerlendirmesini yapacak olursak (Nisbet vd., 2009: 144-145);

- Nonparametrik, veri dağılım özelliklerini gerektirmez.
- Nihai modelleme değişkenleri önceden seçilmez, fakat algoritma tarafından otomatik olarak seçilir.
- Tutarlı olması için verilerin belirli bir matematiksel fonksiyon ile dönüştürülmesine gerek yoktur.
- Çok karmaşık etkileşim kalıpları analiz edilebilir.
- Girdi uzayındaki aykırı (gürültülü) değerlerden etkilenmez.
- Sadece lokal olarak, çıkış değişkenindeki aykırı değerlerden etkilenir.
- Kategorik ve sürekli değişkenlerin herhangi bir bileşimini kabul edebilir.

- Kategorik bağımlı bir değişken üzerinde, tabakalı örnekleri için ayarlanabilir.
- Eksik değerleri işleyebilir, olgular silinmez.

2.2.6.3 CHAID Algoritması

CHAID (Chi-Square Automatic Interaction Detector), Ki Kare Otomatik Etkileşim Dedektörü için kullanılan bir kısaltmadır. İkili Karar Ağaçları'nın (BDT) bölünmesi için istatistik metodları kullanan uygulamalar son yüzyılın 70'lerinde geliştirilmiştir. AID (Otomatik Etkileşim Dedektörü) algoritmasının evrimleşmiş halidir (Donskoy, 2013: 26). CHAID, bölümlendirme amaçlı kullanılan etkili bir istatistiksel tekniktir. Bir istatistiksel testin anlamlılığını kriter olarak kullanarak, bir potansiyel ön kestirici değişkenin tüm değerlerini değerlendirir. (Oğuzlar, 2004: 81) Başlangıçta sadece nominal verileri işlemek için tasarlanmış olan CHAID algoritması 1980 yılında Kass tarafından geliştirilmiştir (Rokach ve Maimon, 2014: 79). Bir değişkeni birden fazla böldüğü için CART'dan farklıdır. Sınıflandırma problemleri için, her adımda en iyi bölünmeyi belirlemek için Ki-kare testine kullanmaktadır. Regresyon problemleri (hedef değişken sürekli ise) için ise F-Testini kullanmaktadır (Nisbet vd., 2009: 146: 247). Nominal değişkenler için Pearson Ki Kare kriterleri, Sıralı-Rank değişkenleri için Olabilirlik-Oran testi kullanılır. Değerlerin istatistiksel önemi ayırtedilebilir çiftleri, değerlerin homojen grubunda birleştirilir ve ayırtedilebilir çiftler bulunduğu sürece, süreç iteratif bir şekilde tekrar eder. Homojen değerler gruplarını bölen değişken, bölünme için seçilir. İkili Karar Ağacı sentezinin durdurulması aşağıdaki durumlardan herhangi biri oluştuğunda gerçekleşir (Donskoy, 2013: 26):

- 1) Maksimum ağaç derinliğine ulaşıldığında,
- 2) Her terminal düğüm verilen eşik değeri sayısından daha küçük nokta sayıları içerdiğinde. Kayıp değişken değerleri (eğer bunlar varsa) ayrı gruplarda birleştirilirler.

CHAID optimaliteyi garanti etmeyen, sezgisel ve ileri aşamalı (forward stepwise) bir yöntemdir. Değişkenlerin hepsi aynı anda değil, sırayla göz önüne alınır, bu nedenle tek bir optimal çözümü garanti etmez (van Diepen ve Franses, 2006: 816-817). CHAID, her bir girdi özneliği için hedef özneliğine göre en az anlamlı derecede farklı olan değerler çifti bulur. Anlamlı farklılık bir istatistik testinden elde edilen " p " değeri yolu ile ölçülür. Seçilen her bir çift için, elde edilen " p " değerinin belli bir birleştirme eşik değerinden büyük olup olmadığını kontrol eder. Eğer cevap pozitif ise değerleri birleştirir ve ilave potansiyel çiftleri birleştirmek için arama yapar. Bu süreç anlamlı çiftler bulunmayana kadar tekrarlanır. Böylece mevcut düğümü bölmek için kullanılan en iyi girdi özneliği seçilir, öyle ki her bir çocuk (alt) düğüm bir seçilen özneliğin değerlerinin homojen bir grubundan oluşur. Eğer en iyi girdi özneliğin

düzeltilmiş “ p ” değeri belli bir bölme eşik değerinden küçük değil ise bölme gerçekleşmez. CHAID eksik değerlerin hepsini tek bir geçerli kategori gibi muamele ederek işlem yapar ve budama işlemi gerçekleştirmez (Kuzey, 2012: 77-78).

CHAID’in avantajlarını sıralayacak olursak (Nisbet vd., 2009: 147);

- Hızlıdır.
- Geniş karar ağaçları oluşturur, çünkü CART gibi ikili bölünme oluşturma ile sınırlı değildir. Pazar araştırmalarında popülerdir.
- Tek dala bağlı çok sayıda terminal düğüm noktaları üretebilir. Bu terminal düğüm noktaları basit bir iki yönlü olasılık tablosu içinde/ile her bir değişken için çoklu kategori kullanarak uygun bir şekilde özetlenebilir.

CHAID’in dezavantajları (Nisbet vd., 2009: 147);

- Değişkeni birden fazla bölme ile alt parçalara ayırdığı için, algoritmada güvenilir sonuçlar elde etmek için büyük miktarda veri gereklidir.
- CHAID ağacı gerçek olamayacak kadar kısa ve ilginç olabilir, çünkü çoklu bölünmeyi gerçek iş koşulları ile ilişkilendirmek zordur.
- Gerçek değişkenler analiz öncesi kategorik hale gelmeye zorlanır ancak değerler arasındaki sıralamanın korunması gereken durumlarda kategorize etmeye çalışmak kullanışlı olmayabilir.

2.2.6.4 QUEST Algoritması

QUEST (Quick, Unbiased, Efficient Statistical Tree) Hızlı, Tarafsız, Verimli İstatistiksel Ağaç algoritması Loh ve Shih tarafından 1997 yılında geliştirilmiştir. Tek değişkenli ve doğrusal kombinasyonlu bölünmeleri desteklemektedir (Rokach ve Maimon, 2014: 80). Bölünme için giriş değişkeni ve hedef değişkeni arasındaki bağlantı sıralı ya da sürekli değişkenlerin homojenlik dağılımı F-kriterleri ANOVA (Varyans Analizi) temelinde veya Levene testi temelinde, nominal değişkenler için Pearson Ki Kare- X^2 kriterleri temelinde hesaplanır. Çoklu sınıf hedef değişkenleri için iki süpersınıfın birleştirilmesinde kümeleme analizi kullanılmaktadır. Bölünme için hedef özellikli istatistiksel bağlantının en büyük tahmincisine sahip olan değişken seçilirken, budanma için bir çapraz geçerlilik kullanılmaktadır (Donskoy, 2013: 27). Giriş özniteliği için optimal bölünme noktasını bulmak için Kuadratik Diskriminant Analizi uygulanır. QUEST önemsiz yanlılığa sahiptir ve ikili karar ağacı üretmektedir.

ID3, C4.5, C5.0, CHAID, CART, QUEST algoritmalarının dışında;

- Muller ve Wysotzki tarafından 1994 yılında gerçekleştirilen, CAL5

- Loh ve Vanichsetakul tarafından 1988 yılında gerçekleştirilen, FACT
- Brodley ve Utgoff 1995 yılında gerçekleştirilen, LMDT
- Holte tarafından 1993 yılında gerçekleştirilen, T1
- Rastogi ve Shim tarafından 2000 yılında gerçekleştirilen, PUBLIC
- Friedman tarafından 1991 yılında gerçekleştirilen, MARS

Algoritmaları literatürde yer almaktadır (Rokach ve Maimon, 2014: 80-81).

2.2.7 Karar Ağaçlarının Etkinliğinin Değerlendirilmesi

Bütün olarak düşünüldüğünde bir karar ağacının etkinliği, karar ağacına bir test kümesinin (karar ağacını oluştururken kullanılmamış olan kayıtlar koleksiyonu) uygulanmasıyla ve doğru sınıflandırma yüzdelерinin gözlenmesiyle belirlenir. Bu işlem karar ağacının bütünü için bir sınıflandırma hata yüzdesi sağlar ve bu ayrıca karar ağacının bağımsız dallarının kalitesi hakkında da bilgi sağlaması açısından önemlidir. Karar ağacında yer alan her yol bir kuralı temsil eder ve bazı kurallar diğerlerinden daha anlamlıdır. Son düğüm veya dallanma düğümlerinin her birinde aşağıdaki unsurlar ölçülebilir (Irmak, 2009: 41).

1. Düğüme giren kayıt sayısı
2. Her sınıfta yer alan kayıtların yüzdesi
3. Mevcut düğümün son düğüm olması durumunda buradaki kayıtların nasıl sınıflandırılacağı
4. Mevcut düğümde sınıflandırılan kayıtların sınıflama doğruluğu yüzdesi
5. Eğitim seti (training set) ve kontrol seti (test set) arasındaki varyans dağılımı

2.2.8 Karar Ağaçlarının Avantajları ve Dezavantajları

Literatürde, bir sınıflandırma aracı olarak karar ağaçlarının çeşitli avantajları işaret edilmiştir, bunlar; (Rokach ve Maimon, 2014: 81)

- Karar ağaçları kendini açıklayıcıdır ve yoğun olduğu zaman da takip etmesi kolaydır. Başka bir ifadeyle, eğer karar ağacının makul sayıda yaprakları var ise profesyonel olmayan kullanıcılar tarafından anlaşılabilir. Ayrıca, karar ağaçları kurallar kümesine dönüştürülebilir. Bu nedenle bu gösterim anlaşılabilir olarak kabul edilir.
- Karar ağaçları hem nominal hem de sayısal girdi özniteliklerini işleyebilir.
- Karar ağacının gösterimi herhangi bir kesikli (ayrık) değerli sınıflandırıcıyı temsil etmek için yeterince zengindir.
- Karar ağaçları hatalar içeren veri setlerini işleme yeteneğine sahiptir.

- Karar ağaçları eksik değerler içeren veri setlerini işleme yeteneğine sahiptir.
- Karar ağaçları parametrik olmayan bir yöntem olarak kabul edilir. Bu, “Karar ağaçları uzay dağılımı ve sınıflandırıcı yapısı hakkında hiçbir varsayıma sahip değildir.” demektir.
- Sınıflandırma maliyeti yüksek olduğu zaman, karar ağaçları sadece kökten yaprağa tek bir yol boyunca özelliklerin değerlerini istediğinden çekici (cazip) olabilir.

Diğer yandan karar ağaçları bazı dezavantajları vardır bunlar (Rokach ve Maimon, 2014: 82-83);

- ID3 ve C4.5 gibi algoritmaların çoğu hedef özniteliğin sadece kesikli değerler olmasını gerektirmektedir.
- Karar ağacı “böl ve yönet” metodunu kullanırlar, eğer birkaç yüksek alakalı öznitelik varsa iyi bir performans gösterme eğilimindedir, fakat daha çok karmaşık etkileşimler varsa daha az performans gösterme eğilimindedirler. Bunun nedenlerinden biri, diğer sınıflandırıcıların bir karar ağacı kullanarak sunması oldukça zorlayıcı olabilen bir sınıflandırıcıyı daha kompakt bir şekilde tanımlayabilmesidir. Bu fenomenin basit bir örneği karar ağaçlarının çoğaltma (replication) problemidir. Bir kavramı temsil etmek için karar ağaçlarının çoğu, örnek uzayı birbirini dışlayan bölgelere böldüklerinden bazı durumlarda tanımlayıcıyı temsil etmek için karar ağacı aynı alt ağacın tekrarlarını içermelidir.
- Karar ağaçlarının açgözlü özelliği dikkat edilmesi gereken bir başka dezavantaja neden olur. Eğitim kümesine, alakasız (irrelevant) özniteliklere ve gürültüye karşı olan aşırı duyarlılık karar ağaçlarını dengesizleştirir. Köke yakın bir bölünmedeki küçük bir değişiklik bütün alt ağacı değiştirecektir.
- Parçalanma problemi, verinin daha küçük parçalara bölünmesine neden olur. Bu durum genelde birçok özellik yol boyunca test edildiğinde yaşanır. Çoğaltma her zaman parçalanmayı işaret etsede, parçalanma herhangi bir çoğaltma olmadan gerçekleşebilir.
- Bir diğer problem, eksik değerler ile başa çıkmak için gerekli çabayı ifade eder. Eksik değerler ile başa çıkma yeteneği avantaj olarak kabul edilmekle birlikte, bu yeteneği elde etmek için aşırı çaba gösterme bir dezavantaj olarak kabul edilir.

2.3 Birliktelik Kuralları

Birliktelik Kuralları veri setlerindeki elemanlar arasındaki beklenmedik ilişkilerin ortaya çıkarılması, karakterize edilmesi, aralarındaki bağlantının büyüklüğünün tespit

edilmesi için kullanılan bir yöntemdir. Ayrıca Birliktelik Kuralları bu elemanların kombinasyonları arasındaki yakınlıkları açıklayan kuralları tanımlar (Maimon ve Rokach, 2005a: 307; Ye, 2003: 37).

Birliktelik Kuralları büyük miktarda kurallar dizisini bulmak üzere tüm veri setini tarar. Birçok birliktelik kuralları keşif sistemleri, Apriori algoritmasında olduğu gibi sık öge kümesi stratejisini kullanır. (Agrawal & Srikant, 1994). Bu yaklaşım belirli bir minimum destek gerektirir. İlk olarak destek eşiğine eşit ya da destek eşiğini aşan tüm veri setleri keşfedilir. Birliktelik kuralları bundan sonra oluşturulur. Bu strateji düşük (seyrek) yoğunluktaki veri setleri için çok etkilidir (Ye, 2003: 37).

Örnek olarak;

- Kahve ve kola satın alan müşteriler, %80 ihtimalle çikolata satın almaktadırlar.
- Kahve ve kola alan müşteriler, %30 oranında diyet ürünlerini almayı tercih etmemektedirler.

gösterilebilir.

Veri madenciliği literatüründe genellikle “Pazar Sepet Analizi” olarak bilinmektedir. Pazar sepet verilerinde ilk kez Agrawal ve Srikant tarafından uygulanmıştır (1995). Fakat Birliktelik Kuralları sadece bu alanda kullanılmamaktadır. Birliktelik kurallarının uygulandığı alanlar aşağıda listelenmiştir (Maimon ve Rokach, 2005a: 307).

- Nüfus sayımında (Brin vd., 1997)
- Dilbilimsel verilerde (Aumann ve Lindell, 2003)
- Sigortacılık verilerinde (Castelo ve Giudici, 2003)
- Medikal görüntüleme (Gamberger vd., 1999)
- Başarısızlık plan nedenleri (Zaki, 2001)
- Web kişiselleştirme (Mobasher vd., 2002)
- Metin verileri (Brin vd., 1997, Delgado vd., 2002)
- Yayın veritabanları (Lee vd., 2001)

Birliktelik kurallarında ilk adım sık öge setlerini tespit etmektir. Bu öge setleri sıklıkla bir araya gelen ve araştırılması gereken öge setleridir. Arama uzayının üstel bir ölçek olmasından dolayı bu aşama hesaplama gücü açısından çok zordur, etkili bir algoritma ve veri yapılarının kullanımına ihtiyaç vardır. Gerçek zamanlı bir veri ile uğraşırken bu faktörler gerçekten çok önemli bir hale gelmektedir (Tanıar, 2008: 61).

2.3.1 Temel Kavramlar ve Tanımlar

$I=\{i_1, i_2, i_3, \dots, i_m\}$ öge kümesi (item set), $D=\{T_1, T_2, T_3, \dots, T_m\}$ işlemleri içeren veri tabanı (T, transaction) ve her işlem öğelerin bir kümesi yani $T \subseteq I$ olacak şekilde tanımlansın. Her işlemin (T, transaction) TID adı verilen bir tanımlayıcısı ve A öğelerin bir kümesi olsun. Bir işlem (T), A öge setini ancak ve ancak $A \subseteq T$ olduğunda içerir. Bir birliktelik kuralı, öncül (antecedent) ve ardıl (consequent) adı verilen iki öge kümesinden oluşur. Ardıl öge genellikle tek bir öge içerecek şekilde sınırlandırılmıştır. Kurallar genellikle;

$$A (\text{öncül/antecedent}) \Rightarrow B (\text{ardıl/consequent})$$

şeklinde gösterilir, burada $A \subset I$, $B \subset I$ ve $A \cap B = \emptyset$ 'dir (Han ve Kamber, 2006: 230; Maimon ve Rokach, 2005: 300, Ye, 2003: 27-28).

Birliktelik kurallarının öncül ve ardıl öge setleri arasındaki yakınlığı gösterdiğini daha önce ifade etmiştik. Birliktelik kuralında ilişkiyi tanımlamak için frekans tabanlı bir istatistik eşlik etmektedir. Bu ilişkileri tanımlamak için destek (support) ve güven (confidence) adı verilen iki istatistikî sayısal değer kullanılmaktadır (Ye, 2003: 27-28). Analistler, ya yüksek güven veya yüksek destek ya da her ikisini birden tercih edebilirler. Güçlü kurallar belirli asgari destek ve güven kriterlerini karşılayan ya da aşacak şekildedir (Larose, 2005: 184).

Birliktelik kurallarının amacı D kümesinden kurallar oluşturmaktır, kuralların;

- Destek değeri (support), kullanıcı tarafından belirlenen en küçük destek (minsupport) değerinden büyük ya da eşit olmalıdır. ($\text{support} \geq \text{minsupport threshold}$)

$$\text{Support}(A \cup B) \geq \text{Minsupp}$$

- Güven değeri (confidence), belirlenen en küçük güven (minconfidence) değerinden büyük ya da eşit olmalıdır. ($\text{confidence} \geq \text{minconf threshold}$)

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)} \geq \text{Minconf}$$

Birliktelik kuralı oluşturmada iki farklı yaklaşım mevcuttur (Zhang ve Zhang, 2002: 26);

- 1) Destek değeri, kullanıcı tarafından belirlenen en küçük destek (minsupport) değerinden büyük ya da eşit kümeleri bulmadır. Yani tüm yaygın öğeleri belirlemektir.

- 2) Kural oluşturmaktır. Destek değeri (confidence), minconf (minconfidence) değerinden büyük ya da eşit olan kuralları oluşturmaktır.

Birliktelik kurallarını keşfetmek için bir algoritma ile her olası öncül ve ardıl öge setleri arasındaki birliktelikleri bulmak, bunların destek ve güven değerlerini hesaplamak ve belirtilen kısıtlamaları sağlamayan tüm birliktelikleri atmak istersek burada ki muhtemel kombinasyonların sayısı çok büyük olacaktır. Eğer burada 1000 öge seti var ise 2^{1000} kadar değişik kombinasyon mevcuttur. Bu sadece öncüllerin sayısıdır (Ye, 2003: 28). Bu sebepten dolayı birliktelik analizinin uygulanmasında ele alınması gereken iki konu vardır. Birincisi, büyük veri setinden örüntülerin ve desenlerin çıkartılmasının hesaplama açısından oldukça maliyetli olmasıdır. İkincisi ise bulunan bazı örüntülerin ve desenlerin sahte olma ihtimalidir, çünkü bu örüntüler tesadüfen ortaya çıkmış olabilir (Tan vd., 2006: 328).

2.3.2 Birliktelik Kurallarında İlginçlik Ölçütleri

Bir birliktelik kuralı, Öncül (LHS) $\Rightarrow$ Ardıl (RHS) olarak tanımlansın.

2.3.2.1 Kapsam Değeri

Birliktelik kuralının Kapsam Değeri (coverage), öncül öge setinin toplam öge setine oranı olarak tanımlanır. Kapsam değeri [0,1] arasında bir değer alır ve 1'e yakın olan bir değer ilginç, kayda değer bir birliktelik kuralı olarak kabul edilebilir. Kapsam değeri aşağıdaki gibi formülize edilebilir; (Maimon ve Rokach, 2005a: 621)

$$Kapsam|Coverage = \frac{n(LHS)}{N} = P(LHS) = \frac{\text{Öncül öge seti eleman sayısı}}{\text{Toplam Eleman Sayısı}}$$

2.3.2.2 Destek Değeri

Birliktelik kuralının Destek Değeri (support), öncül ve ardıl öge setlerini birlikte içeren işlemlerin tüm veri tabanındaki işlem sayısına oranı olarak tanımlanır (Ye, 2003: 28). Destek değeri aşağıdaki gibi formülize edilebilir;

$$Destek|Support = \frac{n(LHS \cap RHS)}{N} = P(LHS \cap RHS)$$

$$= \frac{\text{Öncül ve Ardıl öge setlerini birlikte içeren işlemlerin sayısı}}{\text{Toplam işlem sayısı}}$$

Destek değeri, bir kuralın veri setinde ne kadar sıklıkla gözlemlendiğinin, ayrıca bunun bir sonucu olarak bu kuralın ne kadar önemli olduğunun göstergesi olarak kabul edilebilir (Maimon ve Rokach, 2005a: 621).

2.3.2.3 Güven Değeri

Bir birliktelik kuralının Güven Değeri (confidence), öncül ve ardıl öge setini birlikte içermeye olasılığının öncül öge setini içermeye olasılığına bölümüdür (Ye, 2003: 28). Güven değeri aşağıdaki şekilde ifade edilir;

$$\text{Güven|Confidence} = \frac{n(LHS \cap RHS)}{n(LHS)} = \frac{P(LHS \cap RHS)}{P(LHS)}$$

$$= \frac{\text{Öncül ve Ardıl öge setlerini birlikte içeren işlemlerin sayısı}}{\text{A öge setini içeren işlem sayısı}}$$

Güven değeri, [0,1] arasında bir değer alır. 1'e yakın bir güven değeri, birliktelik kuralının önemli olduğunu gösterir. Güven ve Destek değerleri, birliktelik kuralı çıkarma işlemlerinde yaygın olarak kullanılır ve ayrıca Agrawal ve Srikant'ın sık öge seti ölçütü olarak bilinir (Maimon ve Rokach, 2005a: 621).

2.3.2.4 Kaldırma Oranı

İlginçliğin en popüler tarafsız ölçüsü kaldırma oranıdır (lift ratio). Kaldırma oranı; kuralın güven değerinin, ardıl öge setini içeren işlem sayısının veritabanındaki toplam işlem sayısına oranının bölümüdür (Ye, 2003: 31). Birlikteliğin önemli bir ölçüsüdür ve kapsama değerinden bağımsızdır, bu oran aşağıda belirtildiği gibi ifade edilir (Maimon ve Rokach, 2005a: 622);

$$\text{Lift} = \frac{P(LHS \cap RHS)}{P(LHS)P(RHS)} = \frac{\text{Confidence}}{P(RHS)}$$

Lift değeri R^+ 'da (R^+ : Reel Pozitif Sayılar Uzayı) herhangi bir değer alabilir ve bu ilginçlik ölçütü için aşağıdaki çıkarımlar elde edilir (Maimon ve Rokach, 2005a: 622);

1. Lift değeri 1'e yaklaşıyorsa, öncül ve ardıl ögeler birbirinden bağımsızdır.
2. Lift değeri $+\infty$ 'a yaklaşıyorsa, 2 farklı durum vardır, bunlar;
 - $RHS \subseteq LHS$ ya da $LHS \subseteq RHS$ bu iki durumdan herhangi biri sağlandığı zaman kuralın ilginç (kayda değer) olmadığı sonucuna varılır.

- $P(RHS)$ değeri 0'a yakın ya da $P(RHS|LHS)$ değeri 1'e yakın ise, ilk durumda kuralın önemli olmadığını, ikinci durumda ise kuralın ilginç olduğunun iyi bir göstergesidir.
3. Lift değerinin 0'a eşit olması, $P(RHS|LHS) = 0 \Leftrightarrow P(RHS \cap LHS) = 0$ demektir. Bu durumda kural önemli değildir.

2.3.2.5 Kaldıraç Oranı

Kaldıraç oranı (leverage ratio), birliktelik kuralının güven ve kapsam değerlerini içeren, birlikteliğin önemli bir ölçüsüdür. Kaldıraç (Leverage) değeri $[-1,1]$ arasında bir değer alır. Bu değer 0 ya da 0'ın altında ise öncül ve ardıl öğelerin birbirlerinden bağımsız yani ilişkisiz olduğunu gösterirken, değer 1'e yakın ise güçlü bir birliktelik kuralının göstergesidir. Kaldıraç oranı aşağıdaki gibi ifade edilir (Maimon ve Rokach, 2005a: 622);

$$Leverage = P\left(\frac{RHS}{LHS}\right) - (P(LHS) \cdot P(RHS))$$

Kaldırma katsayıları (Leverage) yaygın olarak kullanılmasına rağmen, bir kuralın ne kadar ilginç olacağı konusunda her zaman iyi bir ölçüt değildir. Düşük frekanslı ve yüksek kaldırma oranına sahip bir birliktelik, alternatif kurallı yüksek frekans ve düşük kaldırma oranına sahip birliktelikten daha az önemli olabilir. Çünkü sonra gelen (latter applies) daha fazla bireyi etkiler ve bundan dolayı öncül ve ardıl arasındaki etkileşimden kaynaklanan bireylerin sayısındaki toplam artış çok daha büyüktür. Tek bir değerde hacmin ve etkinin gücünü yakalayan ölçü; kaldırma katsayılarıdır (Ye, 2003: 32). Ayrıca Birliktelik Kurallarında kullanılan kaldırma oranı, finasta kullanılan kaldırma oranından farklıdır.

2.3.3 Sık Öğe Kümelerinin Bulunması

Sık öğe kümelerinin belirlenmesi bilgi keşfi ve veri madenciliği ile ilgilenen çalışmacıların karşılaştığı en önemli konulardan biridir. Büyük veritabanlarından sık öğe kümelerinin belirlenmesi için geliştirilmiş birçok algoritma mevcuttur. Apriori sık öğe kümelerinin bulunmasında en çok kullanılan algoritmadır. Verimlilik için bu yaklaşımın, Hash tabanlı algoritma ve OPUS tabanlı algoritma gibi çeşitli varyasyonları inşa edilmiştir (Zhang ve Zhang, 2002: 33).

2.3.3.1 Apriori Algoritması

Birliktelik kurallarının çıkarılmasında kullanılan bir algoritmadır. Agrawal ve Srikant (1994) tarafından önerilmiştir. Algoritmanın ismi, adından anlaşılacağı üzere bu algoritma sık öge kümesi hakkında ön bilgi kullanmaktadır. “K+1” adet öge kümesi bulmak için “K” adet öge seti kullanılır. Kendinden önceki çıkarımlara bağlı olduğu için, latince, önce anlamına gelen “prior” kelimesinden gelmektedir (Taniar, 2008: 63).

Apriori algoritması sık öge kümelerini bulabilmek için veri tabanını birçok kez tarar ve algoritma aşağıda belirtildiği çalışır (Diri, 2016: 16).

1. İlk taramada bir elemanlı minsup sağlayan sık öge kümeleri bulunur.
2. Bir önceki taramada bulunan sık öge kümeleri, aday kümeler adı verilen yeni potansiyel sık öge kümelerini üretmek için kullanılır.
3. Aday kümelerin destek değerleri tarama sırasında hesaplanır ve aday kümelerinden minsup değerini sağlayan kümeler sık öge kümeleri olarak ayrılırlar.
4. Sık öge kümeleri bir sonraki geçiş sırasında aday küme olarak adlandırılır. Bu süreç yeni sık öge kümesi bulunmayana kadar devam eder.

Apriori Algoritması (Ding ve Sundarraaj, 2006: 3)

APRIORI(D, S_{min})

1. $C_1 \leftarrow \{\{i\} | i \in I\}$
2. $k \leftarrow 1$
3. *while* $C_k \neq \emptyset$ *do*
4. *for each* transaction $T \in D$ *do*
5. *for each* candidate itemset $c_k \in C_k$ *do*
6. *if* $c_k \subseteq T$ *then*
7. $s(c_k) \leftarrow s(c_k) + 1$
8. $L_k \leftarrow \{c_k | c_k \in C_k, s(c_k) \geq c_{min}\}$
9. $C_{k+1} \leftarrow \emptyset$
10. *for each* $I_1, I_2 \in L_k, |I_1 \cap I_2| = k - 1$ *do*
11. $I \leftarrow I_1 \cup I_2$
12. *if* $\forall J \subset I, |J| = k$ *and* $J \in L_k$ *then*
13. $C_{k+1} \leftarrow C_{k+1} \cup I$
14. $k \leftarrow k + 1$
15. *return* L_k

2.3.3.2 FP – Growth Algoritması

Han, Pei ve Yin tarafından 2000 yılında geliştirilmiştir (Han vd., 2000: 1). Sık örüntüleri bulmak için kullanılan FP-Growth birliktelik (association) algoritması çoğu algorithmadan daha etkili bir şekilde çalıştığı ve maliyeti azalttığı görülmüştür. Bütün veritabanını daha küçük ve daha yoğun bir veri yapısı, sık örüntü ağacı içinde tutması bunun en büyük nedenidir. Apriori tabanlı algorithmalardan farklı olarak tüm veritabanı iki kez taranır. Birinci taramada tüm öğelerin destek değerinin hesaplanır, ikinci tarama da ise ağaç yapısı oluşturulur (Özdoğan, 2010: 27).

FP-Growth Algoritmasının Genel Yapısı (Birant vd., 2010: 5)

Algoritma FPGrowth(VT, mindestek)

```

Boş liste tanımla: F[];
foreach Hareket  $H_i$  in VT do
    foreach Nesne  $n_j$  in  $H_i$  do
         $F[n_j]++$ 
    end
end
foreach Nesne  $n$  in F do
    if  $F[n] < \text{mindestek}$  then
         $n$  nesnesini F listesinden sil
    end
end
Sırala F[];
FPtree ağaç yapısının kök düğümünün tanımla: kök;
foreach Hareket  $H_i$  in VT do
     $H_i$  kaydını F listesine göre sırala;
    AğacaEkle( $H_i$ , kök);
end
foreach nesne  $n_i$  in N do
    Growth(kök,  $n_i$ , mindestek);
end
End

```

Growth algoritmasının genel yapısı (Birant vd., 2010: 5)

Algoritma Growth(düğüm, n, mindestek)

```

    if düğüm tek bir yol, Y, içeriyorsa then
        foreach düğümleri_kombinasyonu dk in Y do
            örüntü ö = dk  $\cup$  n
            destek = min(dk düğümlerinin destek değerleri)
            if ö.destek > mindestek then
                Çıktı(ö);
            end
        end
    end
else
    foreach  $\alpha_i$  in düğüm do
        örüntü ö =  $\alpha_i \cup n$ 
        destek = min( $\alpha_i$  düğümlerinin destek değerleri)
        if ö.destek > mindestek then
            Çıktı(ö);
        end
    end
    nesne – koşullu örüntüleri oluştur ;
    nesne – koşullu FPtree ağacı oluştur;
    if FPtree  $\neq \emptyset$  then
        Growth(FPtree, ö, mindestek);
    end
end
end
end

```

End

FP-Growth algoritması ana görevin kendi içinde daha küçük görevlere ayrılmasına olanak vermektedir ve oluşturulan FPtree veri yapısı asıl veri kümesinden daha büyük olmamaktadır (Birant vd., 2010: 5).

Apriori ve FP-Growth haricinde birçok birliktelik kuralı algoritması vardır bunlar; (Data Mining Articles, 2014)

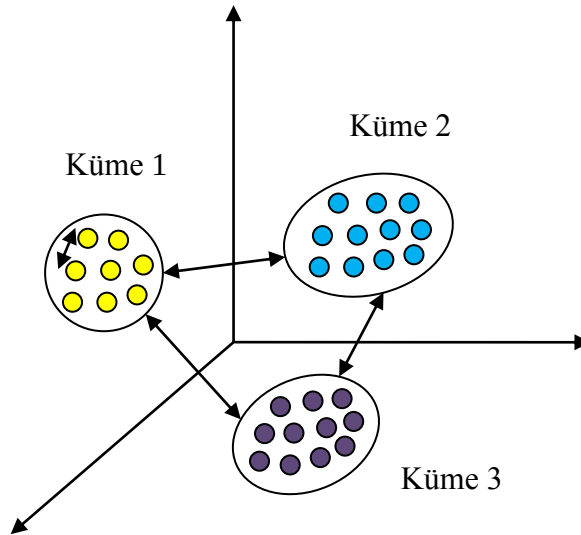
- **AIS**, Rakesh Agrawal, Tomasz Imielinski, Arun Swami, 1993
- **SETM**, Maurice Houtsma, Arun Swami, 1995
- **DHP – Direct Hashing and Pruning**, J. S. Park, M. Chen, P.S. Yu., 1995
- **Partition**, Ashok Savasere, Edward Omiecinski, Shamkant Navathe, 1995

- **ECLAT – Echivalence Class Clustering and Bottom-up Lattice Traversal**, Mohammed Javeed Zaki, Srinivasan Parthasarathy, Mitsunori Ogihara, Wei Li, 1997
- **PASCAL**, Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, L. Lakhal, 2000
- **H-MINE**, Jian Pei, Jiawei Han Lu, Shojiro Nishio, Shiwei Tang, Dongqing Yang, 2001
- **RELIM (Recursive Elimination)**, Christian Borgelt, 2005

algoritmalarıdır.

2.4 Kümeleme Analizi

Kümeleme analizi verileri benzerlik veya yakınlık özelliklerine göre farklı segmentler ya da kümeler halinde gruplandırma süreci olarak tanımlanabilir. İyi bir kümeleme analizi grup içerisinde yüksek benzerlik gösterirken, gruplar arasında düşük benzerlik gösterir. Yani kümeler, kendi içerisinde homojen iken kendi aralarında ise heterojen yapıdadır. İstatistiksel bir teknik olan kümeleme analizi, sosyal alanlar dâhil 40 yılı aşkın bir süredir birçok alanda çalışılmıştır. Çünkü kümeleme analizi, temel bir veri analiz adımı ve genel bir kavramdır. Önceden tanımlanmış bir sınıf etiketi olmamasından dolayı Kümeleme Analizi denetimsiz bir öğrenmedir. Makine öğrenmesi ve istatistiksel örüntü tanıma alanlarında çalışılmıştır (Ye, 2003: 248). Kümelemenin amacı tanımlayıcı/açıklayıcıdır.



Şekil 2.4 Kümeleme Analizi

Kaynak: Pandre, 2016

Şekil 2.4’de görüldüğü gibi küme içi uzaklıklar minimize, kümeler arası uzaklıklar maksimize edilir.

İyi bir kümeleme analizi aşağıdaki özelliklere sahip olmalıdır. (Han ve Kamber, 2006: 385-386; Kogan vd., 2006: 29; Xu vd., 2005: 741)

- Algoritma özneliklerin tüm tipleri için kullanılabilir olmalıdır.
- Yüksek boyutlu verilerle çalışma becerisine sahip olmalıdır.
- Büyük veri setleri için ölçeklenebilir olmalıdır.
- Düzgün şekilli olmayan kümeleri bulma yeteneği olmalıdır.
- Gürültülü veriler ile başa çıkma yeteneğine sahip olmalıdır.
- Anlaşılabilir, yorumlanabilir ve kullanışlı sonuçlar üretmelidir.
- Girdi parametrelerini saptamak için kullanılan alan bilgisinde minimum gereksinime sahip olmalıdır. Kümeleme analizinde, bazı algoritmalar kullanıcı tarafından belirli girdi parametrelerine ihtiyaç duyarlar. Ancak bu parametreleri belirlemek zordur.
- Kısıt tabanlı kümeleme: Çeşitli kısıtlamalar altında kümeleme gerçekleştirebilmeli
- Veri giriş sırasından etkilenme: Bazı kümeleme algoritmaları veri giriş sıralamasına karşı hassastır. Yani verilen bir bilgi objesi önemli ölçüde girdi nesnelerinin sunum sıralamasına bağlı olan farklı kümelemelere dönüşebilir.
- Veri setinin güncellenmesinden etkilenme: Bazı kümeleme algoritmaları sonradan eklenen verileri var olan kümeleme yapılarına dâhil edemez, bu yüzden en baştan bir kümeleme yapılmalıdır.

Tipik bir kümeleme faaliyeti aşağıdaki 5 adımı içerir (Ye, 2003: 249).

1. Uygun desen gösterimi
2. Uygun bir mesafe ya da benzerlik ölçüsü ile nesneler arası yakınlığın tanımı
3. Kümeleme
4. Veri soyutlama (data abstraction)
5. Çıktının değerlendirilmesi

Formal olarak olarak kümeleme yapısı, alt kümelerin bir kümesi şeklinde temsil edilmektedir (Maimon ve Rokach, 2005a: 269).

$$C = C_1, \dots, C_k, \quad S = \bigcup_{i=1}^k C_i$$

$$i \neq j, \quad C_i \cap C_j = \emptyset$$

2.4.1 Benzerlik ve Uzaklık Ölçüleri

Kümeleme analizi ile benzer nesneleri ve örnekleri gruplarken, iki nesnenin birbirine benzeyip benzemediğinin belirlemek için bir ölçü çeşitinin kullanılması gerekmektedir. Bu ilişkiyi tahmin etmek için kullanılan iki ana ölçü birimi vardır, bunlar benzerlik ölçüleri ve uzaklık ölçüleridir. Birçok kümeleme metodu herhangi bir nesne çifti arasındaki benzerlik ya da farklılığı belirlemek amacıyla uzaklık ölçülerini kullanmaktadır. Değişkenlerin nümerik, ikili (binary), nominal, ordinal veya karışık tipli (mixed types) olmasına göre çeşitli uzaklık ölçüleri vardır (Maimon ve Rokach, 2005a: 270-272).

2.4.1.1 Benzerlik Ölçüleri

Uzaklık fonksiyonunun alternatif bir kavramı Benzerlik Fonksiyonudur. Bu fonksiyon simetrik olmalıdır. Benzerlik fonksiyonun (dikotom benzerlik fonksiyonu) hedef aralığı [0,1] arasındadır. Değişkenler ikili (binary) ve nominal olduğu zaman uzaklık ölçülerinin yerine benzerlik ölçüleri kullanılır (Maimon ve Rokach, 2005b: 65). Birimlerin birbirlerine olan benzerliklerini göstermekte kullanılan bu ölçümler maksimum "1" değerini alabilirler. Benzerlik ölçülerinin değerleri arttıkça birimler arasındaki benzerlikler artarken, değer azaldıkça da birimler arasındaki benzerlikler azalmaktadır (Vatansever, 2008: 87). Yaygın olarak kullanılan bazı benzerlik ölçüleri Tablo 2.2’de gösterilmiştir.

Tablo 2.2 Benzerlik Ölçütleri

Kosinüs Ölçüsü	$s(x_i, x_j) = \frac{x_i^T \cdot x_j}{\ x_i\ \cdot \ x_j\ }$
Pearson Korelasyon Ölçüsü	$s(x_i, x_j) = \frac{(x_i - \bar{x}_i)^T \cdot (x_j - \bar{x}_j)}{\ x_i - \bar{x}_i\ \cdot \ x_j - \bar{x}_j\ }$
Genişletilmiş Jaccard Ölçüsü	$s(x_i, x_j) = \frac{x_i^T \cdot x_j}{\ x_i\ ^2 + \ x_j\ ^2 - x_i^T \cdot x_j}$
Dice Katsayı Ölçüsü	$s(x_i, x_j) = \frac{2x_i^T \cdot x_j}{\ x_i\ ^2 + \ x_j\ ^2}$

Kaynak: Maimon ve Rokach, 2005b: 65-66

2.4.1.2 Uzaklık Ölçüleri

Uzaklık ölçülerinin değerleri arttıkça birimler arasındaki benzerlik azalırken, değerler azaldıkça da birimler arasındaki benzerlik artmaktadır. Uzaklık ölçüsünde “0” maksimum

benzerliği ifade etmektedir. (Vatansever, 2008: 87). Geçerli bir uzaklık ölçüsü simetrik olmalı ve denk vektörler durumunda minimum değeri almalıdır (Genellikle sıfır) (Maimon ve Rokach, 2005a: 270).

Uzaklık Fonksiyonun Özellikleri:

i, j gibi iki nesne arasındaki uzaklığı belirtmek için; $d(i, j)$ notasyonu kullanılacaktır. Uzaklık ölçüsü genellikle metrik uzaklık ölçüsü olarak bilinir ve aşağıdaki özellikleri sağlar (Han ve Kamber, 2006: 389):

- 1) $d(i, j) \geq 0$; Uzaklık negatif değildir.
- 2) $d(i, i) = 0$; Her birimin kendisine olan uzaklığı sıfırdır.
- 3) $d(i, j) = d(j, i)$; Uzaklık fonksiyonu simetriktir.
- 4) $d(i, j) \leq d(i, k) + d(k, j)$; Üçgen eşitsizliği.

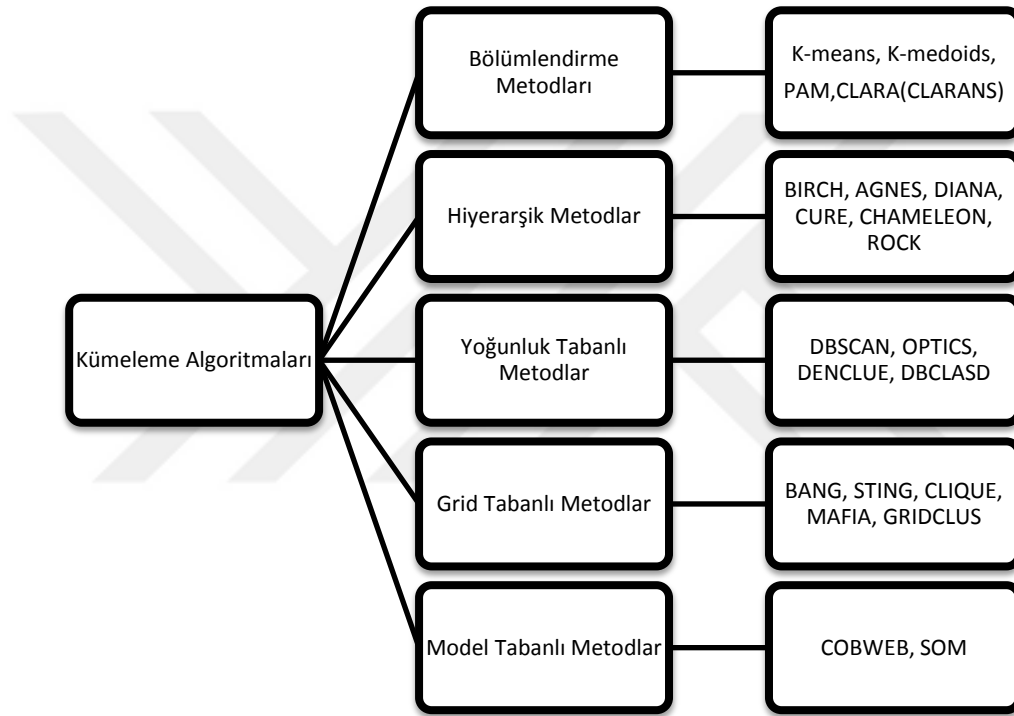
Yaygın olarak kullanılan bazı uzaklık ölçüleri Tablo 2.3’de gösterilmiştir.

Tablo 2.3 Uzaklık Ölçüleri

Öklid Uzaklık Ölçüsü	$d(i, j) = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$
Manhattan/City Block Ölçüsü	$d(i, j) = \sum_{k=1}^m x_{ik} - x_{jk} $
Minkowski Uzaklık Ölçüsü	$d(i, j) = \sqrt[p]{\sum_{k=1}^m x_{ik} - x_{jk} ^p}, p > 0$
Pearson Uzaklık Ölçüsü	$d(i, j) = \sqrt{\frac{\sum_{k=1}^m (x_{ik} - x_{jk})^2}{s_k^2}}$
Mahalanobis Uzaklık Ölçüsü	$d(i, j) = (x_i - x_j)^T \sum (x_i - x_j)^{-1}$
Canberra Uzaklık Ölçüsü	$x_i, x_j \geq 0, d(i, j) = \sum_{i=1}^n \frac{ x_i - x_j }{(x_i + x_j)}$
Çebişev Ölçüsü (Tchebyshev)	$d(i, j) = \max_{i=1..n} x_i - x_j $

2.4.2 Kümeleme Yöntemlerinin Sınıflandırılması

Farley ve Raftery (1998) kümeleme algoritmalarını iki ana gruba ayırmıştır, bunlar hiyerarşik ve bölümlendirme metotlarıdır, Han ve Kamber (2001) ise bu iki gruba 3 kategori daha ekleyerek kümeleme metodlarını beş ana gruba ayırmıştır (Maimon ve Rokach, 2005a: 278). Literatürde birçok kümeleme algoritması mevcuttur. Kümeleme yöntemlerini net bir şekilde sınıflandırmak kolay değildir, çünkü birçok kategoriden özelliğe sahip yöntemlerden dolayı kategoriler çakışabilir (Han ve Kamber, 2006: 398). Genel olarak büyük (majör) kümeleme algoritmaları aşağıdaki gibi sınıflandırılabilir.



Şekil 2.5 Kümeleme Algoritmalarının Sınıflandırılması

Kaynak: Aggarwal ve Reddy, 2014: 128, 133; Han ve Kamber, 2006: 398-433; Kogan vd., 2006: 27-29

2.4.2.1 Bölümlendirme Metotları

Bölümlendirme metotları, “n” elemanlı bir veri tabanını, $k \leq n$ olacak şekilde “k” adet bölüme ayırır. Ayrılan her bölüm bir kümeyi gösterir ve kümeler aşağıdaki özellikleri sağlamak zorundadır.

- 1) Her küme en az bir eleman içermelidir.
- 2) Her eleman sadece bir kümeye ait olmalıdır.

İki numaralı koşul “Bulanık Bölümlendirme Metotunda” (Fuzzy Partitioning Method) esnetilebilir. Verilen bir “k” (k: oluşturulacak olan kümelerin sayısı) sayısı ile kümeleme işlemi başlar, daha sonra nesneleri bir gruptan diğer gruba taşıyarak bölümlendirmeyi

iyileştiren yinelemeli yer değiştirme tekniğini kullanır (Han ve Kamber, 2006: 398-399). Bölümleme işlemi sonucu küme içi elemanlar homojen bir yapı gösterir. Homojenlik, küme içindeki noktalar ile küme merkezinin arasındaki mesafeyi en aza indirecek şekilde uygun bir skor fonksiyonu tarafından yapılır (Hand vd., 2001: 296-297).

K-Means Kümeleme Algoritması:

1957 yılında Cox tarafından tanıtılan bu algoritma, Ball-Hall (1967) ve Mac-Queen (1967) tarafından “K-Means” olarak isimlendirilmiştir (San vd., 2004: 242). K-Means kümeleme, hata karalarının toplamı (SSE) amaç fonksiyonunu minimize etme amacına sahip bir optimizasyon problemidir (Aggarwal ve Reddy, 2014: 90). Algoritmanın çalışmaya başlaması için, kümeleme işlemi bitene kadar değişmeyecek bir sayıya ihtiyaç duyulmaktadır. Bu sayı “k” ile gösterilir ve “k” sayısı meydana gelecek küme sayısını belirtir. Bu yüzden bu algoritmaya k-means adı verilmiştir. K-Means algoritması “n” adet veri içeren bir veri tabanını verilerin özelliklerine göre $k \leq n$ olacak şekilde “k” adet kümeye, kümeler içi benzerliği maksimum, kümeler arası benzerliği minimum yapacak şekilde böler. Küme benzerliği, küme merkezi olarak kabul edilen bir nesne ile kümedeki diğer nesneler arasındaki uzaklıkların ortalama değeri ile ölçülmektedir (Han ve Kamber, 2006; s.402; Demiralay ve Çamurcu, 2005: 4). K-Means algoritmasında Öklid uzaklığı, Manhattan uzaklığı ve Kosinüs benzerliği kullanılmaktadır. En çok tercih edilen popüler ölçü Öklid uzaklığıdır (Aggarwal ve Reddy, 2014: 89). K-Means Kümeleme Algoritması aşağıda verilmiştir (Maimon ve Rokach, 2005: 281).

Input: S (instance set), K (number of cluster)

Output: clusters

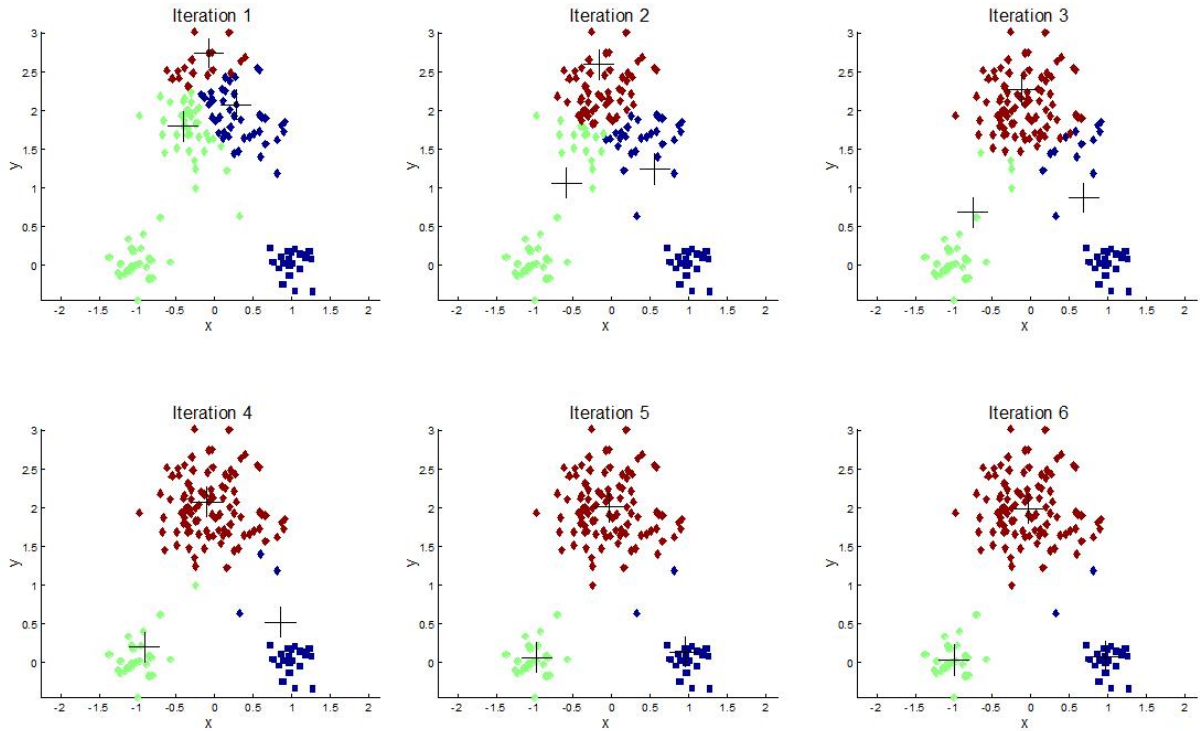
1. Initialize K cluster centers.
2. while termination condition is not satisfied do
3. Assign instances to the closest cluster center.
4. Update cluster centers based on the assignment.
5. end while

K-Means algoritmasının işlem basamakları aşağıda verilmiştir (Han ve Kamber, 2006: 402; Aggarwal ve Reddy, 2014: 89).

1. Nesneler arasından “k” adet rastgele nokta seçilir. Bu noktaların her biri başlangıçta kümenin ortalamasını veya merkezini temsil etmektedir.
2. Her bir nesne, daha sonra seçilen belirli bir yakınlık ölçümüne dayalı en yakın merkeze atanır.

3. Her küme için yeni ortalama hesaplanır ve her küme için küme merkezleri güncellenir.
4. Algoritma, küme merkezleri değişmeyinceye veya alternatif yakınsama kriteri sağlanana kadar 2. ve 3. adımları iteratif bir şekilde tekrarlar.

Şekil 2.6 K-Means tekniğiyle yapılan bir kümeleme işleminin aşamalarını gösterirken, artı işareti (+) ile belirtilen yerler küme merkezlerini göstermektedir.



Şekil 2.6 Kümelerin K-Means Algoritması ile Oluşturulması

Kaynak: Pandre, 2016

K-Means Algoritmasının Özelliklerini belirtecek olursak (San vd., 2004: 243; Kogan vd., 2006: 29).

- 1) Büyük veri setlerini işlemede etkilidir.
- 2) Genellikle yerel optimumları bulur.
- 3) Sadece sayısal veriler üzerinde çalışır.
- 4) Oluşan kümeler dışbükeydir.
- 5) Sonuçlar büyük ölçüde başlangıç küme merkezine bağlıdır.
- 6) Hesaplanan yerel optimum, küresel optimumdan farklı olabilir.
- 7) İyi bir “k” sayısını nasıl seçeceğimiz açık değildir.
- 8) Gözlem dışı/istisnai/gürültü (outlier) sonuçlara duyarlıdır.
- 9) Temel algoritma ölçeklenebilir değildir.
- 10) Oluşan kümeler dengeli olmayabilir (Dengesiz).

K-Means algoritmasının performansını etkileyebilecek önemli faktörler, başlangıç merkezlerinin (centroid) seçimi ve K kümelerinin sayısının tahminidir. K-Means algoritmasında kullanılan küme tahmini ve popüler başlangıç yöntemlerş Tablo 2.4’de gösterilmektedir.

Tablo 2.4 K-Means Algoritmasında Kullanılan Küme Tahmini ve Popüler Başlangıç Yöntemleri

	Küme Sayısının Tahmini	Popüler Başlangıç Yöntemleri
1	Calinski–Harabasz İndeksi	Hartigan ve Wong
2	GAP İstatistiği	Milligan
3	Akaike Bilgi Kriteri (AIC)	Bradley ve Fayyad
4	Bayes Bilgi Kriteri (BIC)	K-means++
5	Duda ve Hart	
6	Silhouette Katsayısı	
7	Newman ve Girman	
8	ISODATA	

Kaynak: Aggarwal ve Reddy, 2014: 91-93

K-Medoids Algoritması:

Kaufman ve Rousseeuw tarafından 1987 yılında geliştirilmiştir (Kaufman ve Rousseeuw, 1990: 68). K-Medoids algoritmasında küme merkezleri, kümedeki nesnelerin ortalama değerleri ile değil, küme merkezine en yakın nesne (medoids) bulunarak belirlenmektedir (Akın, 2008: 82). K-Medoids algoritmasının amacı K-Means algoritmasının amacı gibi, önceden tanımlanmış bir amaç fonksiyonu en aza indiren bir küme çözümü bulmaktır. SSE yerine mutlak hata kriterini en aza indirmeyi hedeflemektedir (Aggarwal ve Reddy, 2014: 93). Aykırı ve gürültülü değerler varlığında K-Means tekniğine göre daha güçlüdür, çünkü medoid aykırı ve gürültülü değerlerden daha az etkilenmektedir (Han ve Kamber, 2006: 407).

K-Medoids metodunun ilk versiyonları PAM (Partitioning Around Medoids) ve CLARA (Clustering LARge Applications) algoritmalarıdır (Kogan vd., 2006: 39). Küme sayısı kullanıcı tarafından belirlenen ve küçük veritabanlarında iyi sonuç veren PAM algoritması, öncelikle rastgele seçtiği “k” adet sayıyı küme merkezi olarak alır. Kümeye her yeni eleman eklendiğinde kümenin elemanlarını deneyerek kümenin gelişmesine en fazla katkıda bulunabilecek noktayı tespit edince bulduğu noktayı yeni merkez, eski merkezi ise sıradan küme elemanı olacak şekilde yer değiştirme işlemi yapmaktadır (Akın, 2008: 82-83; Dinçer, 2006: 32-33). K-Medoids, K-Means algoritmasına göre daha sağlamdır, fakat K-

Medoids algoritmasının hesaplama karmaşıklığı daha yüksektir. Bu durum da büyük veri setleri için uygun olmadığını gösterir. Bu nedenle PAM, büyük veri tabanlarında kullanılma ihtiyacını gidermek üzere geliştirilmiştir. Geliştirilen bu yeni yöntem CLARA'dır (Aggarwal ve Reddy, 2014: 94). K-Medoids Kümeleme Algoritması aşağıda verilmiştir (Han ve Kamber, 2006: 406).

Algorithm: *k*-medoids. PAM:

Input: *k*: the number of clusters,

D: a data set containing *n* objects.

Output: A set of *k* clusters.

Method:

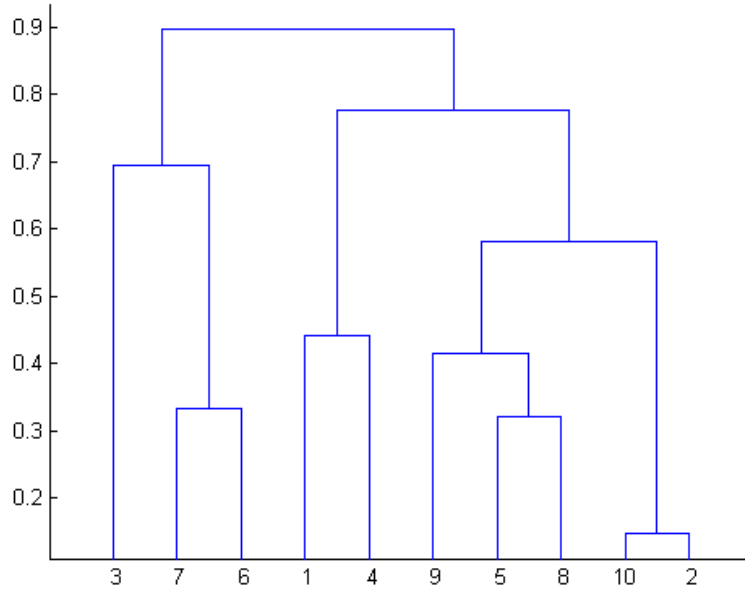
1. arbitrarily choose *k* objects in *D* as the initial representative objects or seeds;
2. repeat
3. assign each remaining object to the cluster with the nearest representative object;
4. randomly select a nonrepresentative object, o_{random} ;
5. compute the total cost, *S*, of swapping representative object, o_j , with o_{random} ;
6. if $S < 0$ then swap o_j with o_{random} to form the new set of *k* representative objects;
7. until no change;

K-Medoids algoritması aşağıdaki gibi özetlenebilir (Akın, 2008: 83).

1. “*k*” küme sayısının belirlenmesi.
2. Başlangıç medoidleri olarak *k* nesnelerinin seçimi.
3. En yakın medoid *x*'e sahip kümeye, kalan nesneleri atamak
4. Amaç fonksiyonunu hesaplamak. (Hata kareler kriteri: en yakın medoidler için bütün nesnelerin uzaklıklarının toplamı)
5. Tesadüfi olarak *y* noktasının seçimi.
6. Eğer *x* ile *y*'nin yer değiştirmesi amaç fonksiyonunu minimize ederse bu iki noktanın (*x* ile *y*) yerini değiştirmek.
7. Değişiklik olmayana kadar 3. Adım ile 6. Adım arası işlemler tekrarlanır.

2.4.2.2 Hiyerarşik Modeller

Hiyerarşik kümeleme verileri “Dendogram” adı verilen hiyerarşik bir ayrıştırma oluşturarak kümeleme esasına dayanır (Ye, 2003: 255). Bu yöntem kümeleri özyinelemeli olarak aşağıdan-yukarıya yâda yukarıdan-aşağıya doğru oluşturmaktadır (Maimon ve Rokach, 2005a: 278).



Şekil 2.7 Dendrogram Örneği

Kaynak: Math Works, 2016

Hiyerarşik modeller iki alt grupta incelenir:

1. Aşağıdan-Yukarıya Doğru Yaklaşım (Yığılmacı veya Toplayıcı, Agglomerative)

Agglomerative yaklaşımda her nesne başlangıçta kendi başına bir küme temsil etmektedir, istenilen küme yapısı elde edilinceye kadar kümeler arda birleştirilir (Maimon ve Rokach, 2005a: 278). İki küme arasındaki mesafenin ölçümüne bağlı olarak Tek Bağlantı Kümeleme (Single Link Clustering/ Nearest Neighbor Method), Tam Bağlantı Kümeleme (Complete Link Clustering/ Maximum Method or the Furthest Neighbor Method) ve Ortalama Bağlantı Kümeleme (Average Link Clustering/ Minimum Variance Method) olarak ayrılmaktadır (Ye, 2003: 255).

Tek Bağlantı Tekniği:

Tek bağlantı tekniğinde gruplar, iki küme arasında birbirine en benzer (en yakın) üyelerin birleştirilmesiyle oluşturulur (Johnson ve Wichern, 1992: 682). Bu teknikte önce birbirine en yakın iki birim (gözlem) bir kümeye yerleştirilir. Daha sonra diğer en yakın uzaklık tespit edilerek ilk oluşturulan kümeye bu gözlem eklenir veya iki gözlemden oluşan yeni bir küme oluşturulur. İşlem tüm gözlemlerin bir kümeye yerleştirilmesine kadar devam eder. Bu teknikte eğer i ve j 'inci birimler birleştirilmiş ise birleştirilen kümenin k 'ıncı küme ile ilişkisi uzaklık ölçütü olarak,

$$d_{k(i,j)} = \min(d_{ki}, d_{kj})$$

şeklinde ifade edilir. Burada;

$d_{k(i,j)}$; k.kümenin daha önce oluşan i. ve j. kümelerle olan uzaklığını,

d_{ki} ; k'ncü kümenin i'inci kümeye olan uzaklığını,

d_{kj} ; k'ncü kümenin j'ncü kümeye olan uzaklığını gösterir (Çelik, 2013: 181).

Tam Bağlantı Tekniği:

Bu yöntem bir farklılık hariç Tek Bağlantı Tekniğine çok benzemektedir. Bu fark, her aşamada kümelerdeki iki eleman arasındaki uzaklığın en yüksek olanının kullanılmasıdır (Johnson ve Wichern, 1992: 686).

Bu teknikteki uzaklıklar,

$$d_{k(i,j)} = \max(d_{ki}, d_{kj})$$

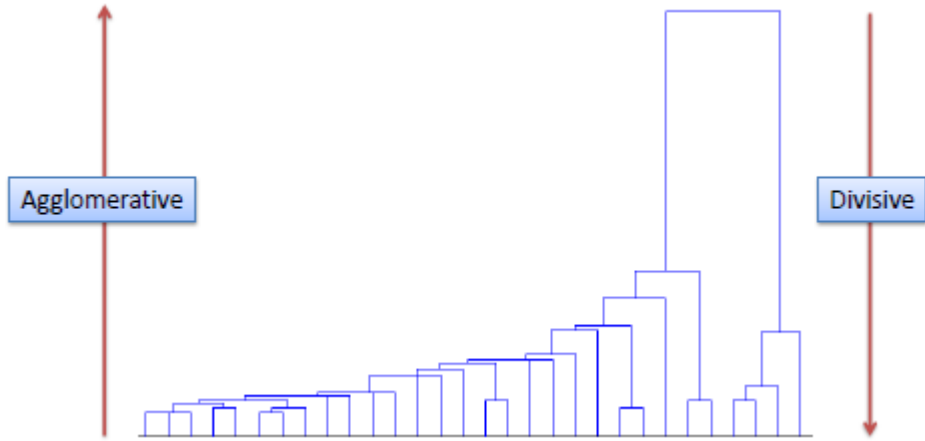
şeklinde gösterilir (Çelik, 2013: 182).

Ortalama Bağlantı Tekniği:

Bu teknikte işleme tek bağlantı ve tam bağlantı tekniklerinde olduğu gibi başlanır fakat kümeleme kriteri olarak bir küme içindeki birim ile diğer küme içindeki birimler arasındaki ortalama uzaklıklar kullanılır. Ortalama bağlantı tekniğinde kümeler küçük varyanslar ile birbirlerine bağlıdır (Çelik, 2013: 182).

2. Yukarıdan-Aşağıya Doğru Yaklaşım (Bölümleyici veya Ayırmacı, Divisive)

Bu yaklaşımda tüm nesneler başlangıçta tek bir kümeye aittir, daha sonra kümeler kendi alt kümelerini oluşturacak şekilde ayrılır. Bu süreç istenen küme yapıları edilinceye kadar devam eder. Hiyerarşik algoritmalarının bazı dezavantajları vardır bunlar: iyi ölçeklenememesi yani zaman karmaşıklığının olması, büyük girdi çıktı maliyetlerine neden olması ve yapıpı bitirilen bir işlemin geri alınamamasıdır (Maimon ve Rokach, 2005a: 278-279).



Şekil 2.8 Agglomerative ve Divisive Dendogram Yapısı

Kaynak: Sayad, 2015

2.4.2.3 Yoğunluk Tabanlı Metotlar

Bu metotlar, nesnelerin doğal dağılımını bir yoğunluk fonksiyonu aracılığı ile tespit ederek bir eşik yoğunluğunu aşan bölgeleri küme olarak adlandırırlar (Dinçer, 2006: 40). Düzgün şekilli olmayan kümeleri bulma, büyük veri setlerine ölçekleyebilme, gürültülü ve istisnai verileri bulma ve bunlardan etkilenmeme gibi gereksinimleri gidermek için öne sürülmüş bir kümeleme paradigmasıdır. Yoğunluk tabanlı kümeleme, kümelerin sayısı ve dağılımı hakkında hiçbir varsayım yapmadığı için parametrik olmayan bir yöntem olarak kabul edilebilir. DBSCAN, OPTICS, DENCLUE Yoğunluk tabanlı kümeleme teknikleri olarak sayılabilir (Aggarwal ve Reddy, 2014: 112-115, 124).

2.4.2.4 Izgara (Grid) Tabanlı Metotlar

Izgara tabanlı algoritmalar veri uzayını ızgara yapıları oluşturmak için sonlu sayıda hücrelere ayırırlar, daha sonra grid yapısındaki hücrelerden kümeleri oluştururlar. Çevrelerinden daha yoğun veri noktaları içeren bölgeler kümelere karşılık gelmektedir. Izgara tabanlı kümeleme algoritmaları büyük boyutlu veri setlerinin madenciliğinde etkilidir. Bu yöntemin avantajı özellikle çok büyük veri kümeleri için zaman karmaşıklığında önemli bir azalma göstermesidir. GRIDCLUS, BANG, STING, CLIQUE ve MAFIA gibi birçok ızgara tabanlı algoritma yöntemi mevcuttur (Aggarwal ve Reddy, 2014: 128, 133).

2.4.2.5 Model Tabanlı Metotlar

Model tabanlı metotlar veri ile bazı matematiksel modeller arasındaki uyumu optimize ederler. Geleneksel kümeleme yöntemlerinin aksine nesnelerin gruplarını tanımlarlar, bir kavram ya da sınıfı temsil eden her bir grup için karakteristik tanımlamaları bulur. Karar

ağaçları ve yapay sinir ağları en çok kullanılan endüksiyon yöntemleridir (Maimon ve Rokach, 2005a: 283).

2.5 Yapay Sinir Ağları

Beyin son derece karmaşık, doğrusal olmayan, paralel bir bilgisayardır (işlemcidir). Algı, örüntü tanıma ve motor kontrol gibi bir takım işlemleri günümüz bilgisayarlarından kat kat daha hızlı yapmak için nöron adı verilen yapısal bileşenlere sahiptir (Haykin, 2009: 1). Yapay Sinir Ağı (YSA) insan beyninin hesaplama modelinin özetidir. İnsan beyninde tahmini olarak 10^{11} adet nöron vardır, bu nöronlar tahminen 10^{15} adet bağlantı ile birbirlerine bağlanırlar. Beyine benzeyen bir YSA, nöronlardan ve bağlantı birimlerinden oluşmaktadır. Bir YSA'yı görüntülediğimizde aynı graf'ta olduğu gibi nöronlar düğümleri, bağlantılar ise kenarları göstermektedir. YSA, adından da anlaşılacağı üzere yönlü bağlantılar aracılığıyla bağlanan düğümlerden oluşan bir ağ yapısıdır. Her düğüm bir işlem birimini, düğümler arası bağlantılar ise bağlı düğümler arası nedensel ilişkiyi temsil eder. Birçok tanım ve yaklaşım olmasına rağmen; YSA, basit işlem birimlerinden oluşan deneyimsel bilgi depolayan ve bu bilgiyi kullanılabilir yapmak için doğal bir eğilime sahip olan büyük paralel dağıtılmış bir işlemcidir (Kantardzic, 2011: 200).

YSA tahmin ve öngörü problemlerinde sıklıkla kullanılır ve aşağıda belirtildiği gibi iki bakımdan beyni andırmaktadır (Haykin, 2009: 2).

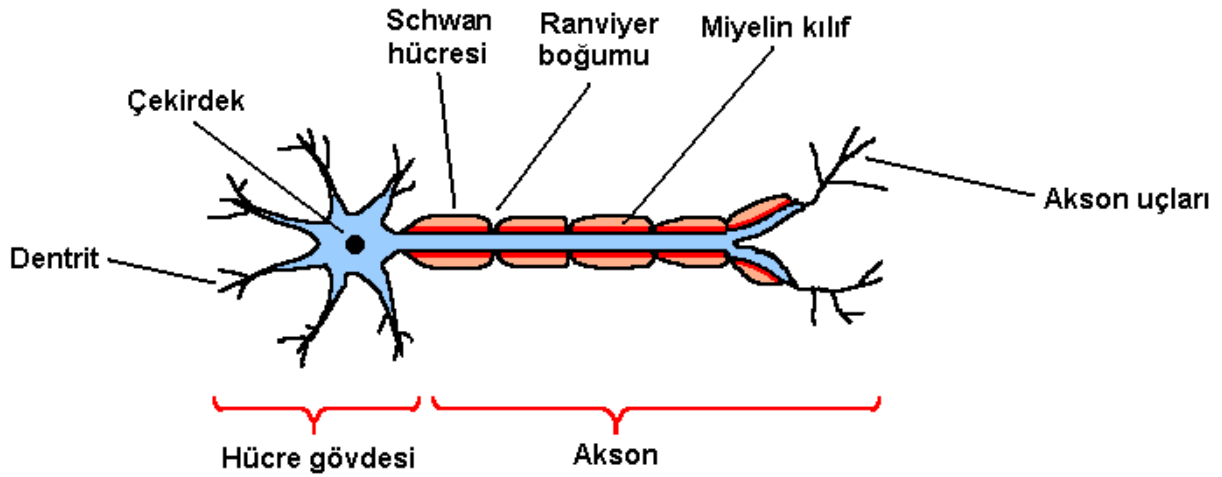
1. Bilgi, bir öğrenme süreci boyunca ağ'dan elde edilir.
2. Sinaptik ağırlıklar olarak bilinen sinirler arası bağlantılar, güçlüdür ve edinilen bilgiyi depolamak için kullanılır.

Öğrenme sürecini gerçekleştirmek üzere kullanılan prosedüre öğrenme algoritması denir, bu fonksiyon istediğimiz amaca ulaşmak için ağın sinaptik ağırlıklarını düzenler. YSA, standart bir veri madenciliği aracı olarak kabul edilmiştir ve desen sınıflandırma, zaman serisi analizi, tahmin ve kümeleme gibi birçok veri madenciliği görevleri için kullanılmaktadır (Maimon ve Rokach, 2005a: 419).

YSA ve biyolojik sinir ağları iki ortak özelliğe sahiptir, birincisi bilginin paralel işlenmesi, ikincisi öğrenme ve deneyimlerden genelleme yapmadır. YSA'nın birçok önemli özelliği, onu veri madenciliği için uygun ve önemli kılmaktadır. YSA, temel veri üretme ve spesifik model yapıları hakkında gerçek dışı ön varsayımlar gerektirmez. Tarihsel olarak sinir ağları alanından birçok araştırmacı biyoloji, matematik, nörobilim, psikoloji, fizik, bilgisayar bilimleri ve bilişsel bilim gibi alanlarında yararlanmışlardır. McCulloch ve Pitts (1943) çalışması sinir ağının modern temeli ve sinir ağı alanının kökeni olarak kabul edilir. Onların

araştırmaları, bir nöronun nasıl çalıştığını açıklamak için matematiksel bir model kullanmaları ilk adımdır. Geliştirdikleri nöron modelinin ana özelliği nöron çıktısını belirlemek için, giriş sinyallerinin ağırlıklı toplamını bir eşik değeri ile karşılaştırmasıdır. Bu bize basit sinir ağının herhangi bir aritmetik ya da mantıksal işlevi hesaplayabileceğini göstermiştir. 1949 yılında Hebb “Davranışın Organizasyonu” kitabında davranışın nöron etkisiyle açıklanabilir olduğunu göstermiştir. Hebb ilk öğrenme kuralını öne sürmüştür, yani biyolojik nöronların öğrenmesi mekanizmasını öne sürmüştür. 1950’lerde Rosenblatt ve diğer araştırmacılar perceptron adı verilen biyolojik nöron modellerini, yani sinir ağlarının bir sınıfını geliştirmişlerdir. Perceptron ve ilişkili öğrenme kuralı (Rosenblatt, 1958) sinir ağları araştırmaları için büyük ilgi uyandırmıştır. Aynı zamanda, Widrow ve Hoff (1960) yeni bir öğrenme algoritması geliştirmiş, ADALINE (Adaptive Linear Neuron) adı verilen ve perceptron’lara benzeyen sinir ağlarında uygulamışlardır. Perceptron’larda kullanılan katı sınırlayıcı fonksiyon yerine doğrusal transfer fonksiyonunu kullanmışlardır. Widrow-Hoff öğrenme kuralı, günümüz popüler sinir ağı öğrenme metotlarının temelidir. Minsky ve Papert (1969) “Perceptronlar” adlı kitabında, Perceptron’ların sınırlandırılmasında ve diğer ilgili ağlarda lineer olmayan ayrılabilir problemlerin geniş bir sınıfının çözümüne dikkat çekmiştir. 1980’li yıllarda Hopfield (1982) istatistiksel mekanik kullanarak özyineli ağı tekrarlayan bir sınıfının işlemlerini açıklamak ve ilişkisel bellek olarak YSA’nın eğitilebileceğini göstermiştir. Kısıtlı bir optimizasyon problemi olan “Gezgin Satıcı Problemi”nde Hopfield Ağları başarıyla kullanılmaktadır. 1982 yılında Kohonen kendiliğinden organize olabilen bir YSA geliştirmiştir. YSA’da en önemli gelişme etkili bir eğitim algoritmasının geliştirilmesi olmuştur, bu geri yayılım algoritması olarak bilinir. Geri yayılım algoritmasının temeli Widrow ve Hoff’un öğrenme kuralından gelmektedir, Werbos (1974) ile resmîyet kazanmıştır, Parker (1985), Rumelhart, Hinton ve Williams (1986) tarafından geliştirilmiştir (Maimon ve Rokach, 2005a: 419-422). Her bir sinir hücresi, dış kaynaklardan (komşulardan) aldığı girdileri, dış ünitelere yaymak üzere bir çıktı değeri üretir.

Şekil 2.9 Bir sinir hücresinin (nöron) yapısını göstermektedir. Şekil 2.9’da görülen “Dentritler” sinyalleri alır, hücre çekirdeği girdi sinyallerini işler, aksonlar işlenmiş girdileri çıktıya dönüştürür, sinaplar ise diğer sinir hücreleri ile elektrokimyasal bağlantıları oluşturur.



Şekil 2.9 Sinir Hücresi (Nöron) Yapısı

Kaynak: Biyodoc, 2014

Tablo 2.5, biyolojik sinir hücresi ile yapay sinir hücresinin temel yapılarını karşılaştırmaktadır.

Tablo 2.5 Biyolojik Sinir Hücresi ile Yapay Sinir Hücresinin Karşılaştırılması

Biyolojik Sinir Hücresi	Yapay Sinir Hücresi
Nöron	İşlem Elemanı
Akson	Çıktı
Dentrit	Toplama Fonksiyonu
Çekirdek	Aktivasyon Fonksiyonu
Sinaps	Ağırlıklar

Kaynak: Yapay Sinir Ağları, 2014

Sınıflandırma veya tahmin için kullanılan YSA'ların izlediği adımları belirtecek olursak (Berry ve Linoff, 2004: 219).

1. Giriş ve çıkış özelliklerini belirleme.
2. Küçük bir aralık ((-1,+1) aralığında) içinde olacak şekilde girileri ve çıkışları dönüştürme.
3. Uygun bir topoloji ile ağ kurma.
4. Ağı eğitme.
5. Ağırlıklar kümesini seçmek için doğrulama setini kullanma. (Bu hatayı en aza indirir)
6. Test kümesini kullanarak ağı değerlendirme. (Ne kadar iyi performans gösterdiğini görmek için)

7. Bilinmeyen girişler ile sonuçların tahmini için ağ tarafından üretilen modeli uygulama.

2.5.1 YSA'nın Özellikleri

Karmaşık problemlerin çözümünde kullanılabilen Yapay Sinir Ağları birçok özelliğe sahiptir (Haykin, 2009: 2-4; Kantardzic, 2011: 200-201; Kulluk, 2009: 63-64; Öztemel, 2012: 31-33);

1. Doğrusal Olmama (Nonlinearity): Temel bir birim olarak yapay nöron doğrusal ya da doğrusal olmayabilir, fakat YSA tamamı ile doğrusal değildir. Aktivasyon fonksiyonu olarak genelde doğrusal olmayan bir fonksiyon seçilir, bu sebepten dolayı YSA'ların doğrusal olmadığı sonucuna ulaşılabilir.
2. Örnekten Öğrenme (Learning From Examples): Bir YSA eğitim kümesini işleyerek ya da örneklerden öğrenerek kendi ara bağlantı ağırlıklarını düzenler ve geliştirir.
3. Uyarlanabilirlik (Adaptivity): YSA çevrede meydana gelen değişikliklere karşı ara bağlantı ağırlıklarını uyumlu hale getiren bir yeteneğe sahiptir. Belirli ortamda çalıştırmak için eğitilmiş bir YSA, çevre koşullarında meydana gelen değişiklikler ile başa çıkmak için yeniden kolayca eğitilebilir. Dahası durağan olmayan bir ortamda çalıştığında, gerçek zamanlı olarak parametrelerini adapte eden bir YSA dizayn edilebilir.
4. Kanıtsal/Delilsel Tepki (Evidential Response) Desen sınıflandırması bağlamında sinirsel bir ağ, yalnızca hangi desenin seçileceği hakkında bilgi vermez. Bunun yanında yapılacak tercihe olan güvenin belirlenmesi için de tasarlanabilir. Bu son bilgi muğlak desenlerin reddedilmesi için kullanılabilir (eğer ortaya çıkarsalar) ve bu şekilde ağın sınıflandırma performansı da artar.
5. Hata Toleransı (Fault Tolerance): Donanım şeklinde uygulanan bir YSA doğası gereği hata toleransına ve sağlam hesaplama yeteneğine sahiptir. Bu da YSA'nın performansının olumsuz durumlarda kademeli olarak azalacağı anlamına gelir.
6. VLSI Uyarlanabilirlik (VLSI Implementability): Bir sinir ağının güçlü paralel doğası bazı görevlerin hesaplanması için onu potansiyel hızlı hale getirir. Bu özellik VLSI (Very Large Scale Integrated) teknolojisi kullanarak uygulanması için YSA'yı son derece uygun kılar.
7. Dağıtık belleğe sahiptirler. YSA'larda bilgi ağa dağılmış bir şekilde bulunur. Hücreler arası bağlantı değerleri ağın bilgisini gösterir. Tek bir bağlantının kendi başına anlamı yoktur. Ağın tamamı öğrendiği için olay bütünü karakterize etmektedir, bu sebeple

bilgiler ağı dağıtılmış durumdadır. Bu durum bize dağıtık bir belleğin oluştuğunu göstermektedir.

8. Eksik Bilgi ile Çalışma: Geleneksel sistemlerin aksine YSA'lar eğitildikten sonra eksik bilgiler ile çalışabilir, gelen yeni örneklerde eksik bilgi olmasına rağmen sonuç üretebilirler.
9. Yerel Bilgi İşleme: YSA'da her bir işlem birimi, ele alınan problemin sadece gerekli kısmı ile ilgilenmektedir ve problemin bir parçasını işlemektedir. Hücreler basit işlem yapmalarına karşın sağlanan görev dağılımı ile karmaşık problemler çözülebilmektedir.
10. Analiz ve Tasarım Tekdüzeliliği (Uniformity of Analysis and Design): Nöronların bir formu veya başka formları tüm YSA'lar için ortak bir elemanı temsil etmektedir, yani hepsi için ortaktır. Bu ortak nokta YSA'nın farklı uygulamalarında öğrenme algoritmalarını ve teorilerini paylaşmayı mümkün kılar. Modüler ağlar, modülün sorunsuz entegrasyonu ile inşa edilebilir.
11. Örüntü tamamlama yapabilirler. Ağı eksik bilgiler içeren örüntü veya şekil verildiğinde, ağ eksik bilgileri bulması konusunda son derece etkili çözümler üretebilmektedir.
12. Dereceli bozulma gösterme. YSA'lar hatalara karşı toleranslı olmaları bozulmalarında dereceli olmasına sebep olur, yani yavaş yavaş bozulurlar. Bu bozulma bilginin bozukluğundan yada hücrenin bozulmasından kaynaklanır.
13. Bağlamsal Bilgiler (Contextual Information): Ağıdaki her nöron, ağıdaki diğer tüm nöronların küresel aktivitesinden potansiyel olarak etkilenir. Sonuç olarak, bağlamsal bilgiler bir sinir ağı tarafından doğal olarak ele alınmıştır.

2.5.2 YSA'nın Dezavantajları

Birçok önemli avantaja sahip olan Yapay Sinir Ağları bazı dezavantajları vardır (Kulluk, 2009: 65; Öztemel, 2012: 34-35);

1. Ağın davranışları açıklanamamaktadır.
2. Donanım bağımlı çalışırlar.
3. Probleme uygun ağ yapısının belirlenmesinde belirli bir kural yoktur, genellikle deneme yanılma yoluyla yapılır.
4. Bazı ağlarda ağın parametre değerlerinin belirlenmesinde bir kural yoktur.
5. Ağın eğitiminin ne zaman bitirileceğine karar vermek için geliştirilmiş bir yöntem bulunmamaktadır. Ağın örnekler üzerindeki hatasının belirli bir değerin altına

indirilmesi eğitimin tamamlanması anlamına gelmektedir, ayrıca optimum neticeler veren bir mekanizma henüz bulunmamaktadır.

6. Ağın öğreneceği problemin ağa gösterimi önemlidir. YSA'lar sadece nümerik bilgiler ile çalışabilmektedirler, bu nedenle problemler önce nümerik değerlere çevrilmek zorundadırlar ve bu kullanıcının yeteneğine bağlıdır. Uygun bir gösterim mekanizması problemin çözümünü engelleyebileceği gibi verimini de düşürecektir dolayısıyla ağın performansını doğrudan etkileyecektir.
7. Bazı ağlar hariç, kararlılık analizi yapılamamaktadır.
8. Farklı sistemlere uyarlanması zor olabilir.

2.5.3 Geleneksel Algoritmalar ile YSA'ların Karşılaştırılması

Tablo 2.6 Geleneksel algoritmalar ile YSA'nın karşılaştırılmasını göstermektedir.

Tablo 2.6 Geleneksel Algoritmalar ile YSA'nın Karşılaştırılması

Geleneksel Algoritmalar	Yapay Sinir Ağları
Çıkışlar, koyulan kurallara girişlerin uygulanması ile elde edilir.	Öğrenme esnasında giriş çıkış bilgileri verilerek, kurallar koyulur.
Hesaplama; merkezi, eş zamanlı ve ardışıldır.	Hesaplama; toplu, eş zamansız ve öğrenmeden sonra paraleldir.
Bellek paketlenmiş ve hazır bilgi depolanmıştır.	Bellek ayrılmış ve ağa yayılmıştır.
Hata toleransı yoktur.	Hata toleransı vardır.
Nispeten hızlıdır.	Yavaş ve donanımına bağımlıdır.
Bilgiler ve algoritmalar kesindir.	Deneyimden yararlanır.

Kaynak: Çayıroğlu, 2015

2.5.4 YSA'nın Kullanım Alanları

Yapay sinir ağları başlıca sınıflandırma ve örüntü tanıma olmak üzere birçok alanda kullanılmaktadır. YSA'ların çok boyutlu, gürültülü, karmaşık, kesin olmayan, eksik, kusurlu, hata olasılığı yüksek sensör verilerinin olması ve problemi çözmek için matematiksel modelin ve algoritmaların bulunmadığı durumlarda kullanıldıkları görülmektedir. (Öztemel, 2012: 35-36)

Muhtemel fonksiyon kestirimleri, Sınıflandırma, İlişkilendirme, Örüntü tanıma, Zaman serileri analizleri, Regresyon analizi, Öngörü, Sinyal filtreleme, Veri sıkıştırma, Örüntü tanıma, Doğrusal olmayan sinyal işleme, Doğrusal olmayan sistem modelleme,

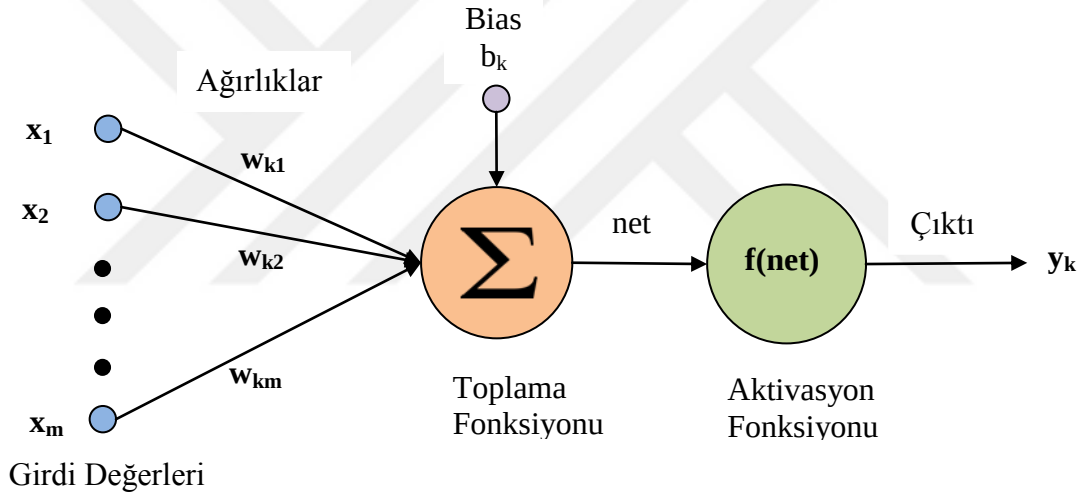
Optimizasyon, Kontrol ve robotik, Nörofizyoloji, Öğrenme ve algılama, Yapay Zekâ, Dinamik Sistemler gibi alanlarda YSA'lar kullanılmaktadır (Öztemel, 2012: 36; Sunay, 2010: 31; Çayıroğlu, 2015; Maimon ve Rokach, 2005a: 437).

2.5.5 Yapay Nöron Modeli

Nöron, sinir ağının çalışması için temel olan bir bilgi işlem birimidir. Bir sinir modelinin temel üç elemanı vardır, bunlar (Kantardzic, 2011: 201):

- 1) Sinaps veya bağlantılar kümesi, her biri kendi ağırlığı ya da gücü ile karakterize edilir. “ x_i ” girdilerin kümesi, her birinin ağırlığı “ w_{ki} ” olarak belirlenir.
- 2) Toplama (Birleştirme) fonksiyonu.
- 3) Aktivasyon (Transfer) fonksiyonu.

Şekil 2.10 yapay sinir hücresinin (nöron) yapısını göstermektedir.



Şekil 2.10 Yapay Nöron Modeli

Kaynak: Kantardzic, 2011: 202

Toplama Fonksiyonu: Tüm girişleri tek bir değer olacak şekilde birleştirir. Her girişin kendine ait bir ağırlığı vardır. En yaygın birleştirme işlemi ağırlıklı toplamdır, her giriş kendi ağırlığı ile çarpılır, oluşan sonuçlar birbirine eklenerek toplanır ve böylece ağırlıklı toplam elde edilmiş olur. Ayrıca minimum, maksimum, mantıksal OR, AND gibi birleştirme fonksiyonları da bulunmaktadır. Birleştirme fonksiyonlarının seçiminde birçok esneklik olmasına rağmen “standart ağırlıklı toplam” birçok durumda iyi çalışmaktadır. Bu esneklik sinir ağlarının ortak bir özelliği olarak karşımıza çıkmaktadır (Berry ve Linoff, 2004: 222).

$x_i, i=1, \dots, m$ girdiler olmak üzere her bir girdi karşılık gelen ağırlığı w_{ki} ile çarpılır, burada “k” belirli bir nöronun indisidir. Burada ağırlıklı toplam YSA’da net olarak ifade edilir (Kantardzic, 2011: 202):

$$net_k = x_1 w_{k1} + \dots + x_m w_{km} + b_k$$

Burada $w_{k0} = b_k$ ve $x_0 = 1$ olarak alırsak net toplam şu şekilde ifade edilebilir.

$$net_k = x_0 w_{k0} + x_1 w_{k1} + \dots + x_m w_{km} + b_k = \sum_{i=0}^m x_i w_{ki}$$

Tablo 2.7 Bazı Toplama Fonksiyonu örneklerini göstermektedir.

Tablo 2.7 Toplama Fonksiyonları

$Net = \sum_{i=1}^N X_i \cdot W_i$	Ağırlık değerleri girdiler ile çarpılır ve bulunan değerler birbirleriyle toplanarak Net girdi hesaplanır.
$Net = \prod_{i=1}^N X_i \cdot W_i$	Ağırlık değerleri girdiler ile çarpılır ve daha sonra bulunan değerler birbirleriyle çarpılarak Net Girdi Hesaplanır.
$Net = \max(X_i \cdot W_i)$	“n” adet girdi içinden, ağırlıklar girdilerle çarpıldıktan sonra içlerinden en büyüğü Net girdi olarak kabul edilir.
$Net = \min(X_i \cdot W_i)$	“n” adet girdi içinden, ağırlıklar girdilerle çarpıldıktan sonra içlerinden en küçüğü Net girdi olarak kabul edilir.
$Net = \sum_{i=1}^N \text{Sgn}(X_i \cdot W_i)$	“n” adet girdi içinden, girdilerle ağırlıklar çarpıldıktan sonra pozitif ile negatif olanların sayısı bulunur. Büyük olan sayı hücrenin net girdisi olarak kabul edilir.
$Net = Net_{Eski} + \sum_{i=1}^N X_i \cdot W_i$	Hücreye gelen bilgiler ağırlıklı olarak toplanır. Daha önce hücreye gelen bilgilere yeni hesaplanan girdi değerleri eklenerek hücrenin net girdisi hesaplanır.

Kaynak: Öztemel, 2012: 50

Transfer Fonksiyonu: İsmi birleştirme fonksiyonunun değerini alıp transfer etmesinden dolayı almaktadır, ayrıca transfer fonksiyonu birleştirme fonksiyonundan gelen sonuçla çıktı değerini hesaplar (Berry ve Linoff, 2004: 223). Bir nöronun çıktı aralığı $[0,1]$ yada alternatif olarak $[-1,1]$ ’dir. Yapay bir nöronda görülen b-bias, pozitif ya da negatif olmasına bağlı

olarak net aktivasyon fonksiyonunu düşürme ya da yükseltme etkisine sahiptir (Haykin, 2009: 11).

Bir nöron, “f” aktivasyon fonksiyonu yardımıyla “net_k” değerini alarak bir çıktı üretir ve “y_k” olarak isimlendirilir. Literatürde aktivasyon fonksiyonu yerine transfer fonksiyonu kavramı da kullanılabilmektedir (Kantardzic, 2011: 202).

$$y_k = f(net_k)$$

Literatürde birçok aktivasyon fonksiyonu bulunmaktadır, bunlardan bazıları Tablo 2.8’de gösterilmiştir.

Tablo 2.8 Aktivasyon Fonksiyonları

Katı Limit	$\varphi = \begin{cases} 1, & \text{Eğer } net \geq 0 \\ 0, & \text{Eğer } net < 0 \end{cases}$
Simetrik Katı Limit	$\varphi = \begin{cases} 1, & \text{Eğer } net \geq 0 \\ -1, & \text{Eğer } net < 0 \end{cases}$
Lineer (Doğrusal)	$\varphi = net$
Doygun Doğrusal	$\varphi = \begin{cases} 1, & \text{Eğer } net > 1 \\ net, & \text{Eğer } 0 \leq net \leq 1 \\ 0, & \text{Eğer } net < 0 \end{cases}$
Simetrik Doygun Doğrusal	$\varphi = \begin{cases} 1, & \text{Eğer } net > 1 \\ net, & \text{Eğer } 0 \leq net \leq 1 \\ -1, & \text{Eğer } net < -1 \end{cases}$
Log-Sigmoid	$\varphi = \frac{1}{1 + e^{-net}}$
Hiperbolik Tanjant Sigmoid	$\varphi = \frac{e^{net} - e^{-net}}{e^{net} + e^{-net}}$

Kaynak: Kantardzic, 2011: 203

YSA’lar yönlendirilmiş graflar ile gösterilir. Modelin işlevsel detayları hakkında herhangi bir ödün vermeden sinyal akış çizgilerini (signal flows graph) kullanarak modelin görünüşünü kolaylaştırılabilir. İşaret akış çizgileri, Mason (1953, 1956) tarafından iyi tanımlanmış kurallar seti ile lineer ağaçlar için geliştirilmiştir. Sinyal akış grafiği, yönlendirilmiş bağlantılardan oluşan bir ağdır, bu bağlantılar birbirlerine düğümler aracılığıyla bağlanırlar. Sinyallerin akışı için 3 temel kural vardır bunlar; (Haykin, 2009: 14).

- 1) Bir sinyal bağlantı yerine sadece ok tarafından belirtilen yöndeki bağlantı boyunca akar. İki çeşit bağlantı mevcuttur bunlar sinaptik bağlantı ve aktivasyon bağlantısı.
- 2) Bir düğümdeki sinyaller, düğüme giren sinyallerin toplamına eşittir.

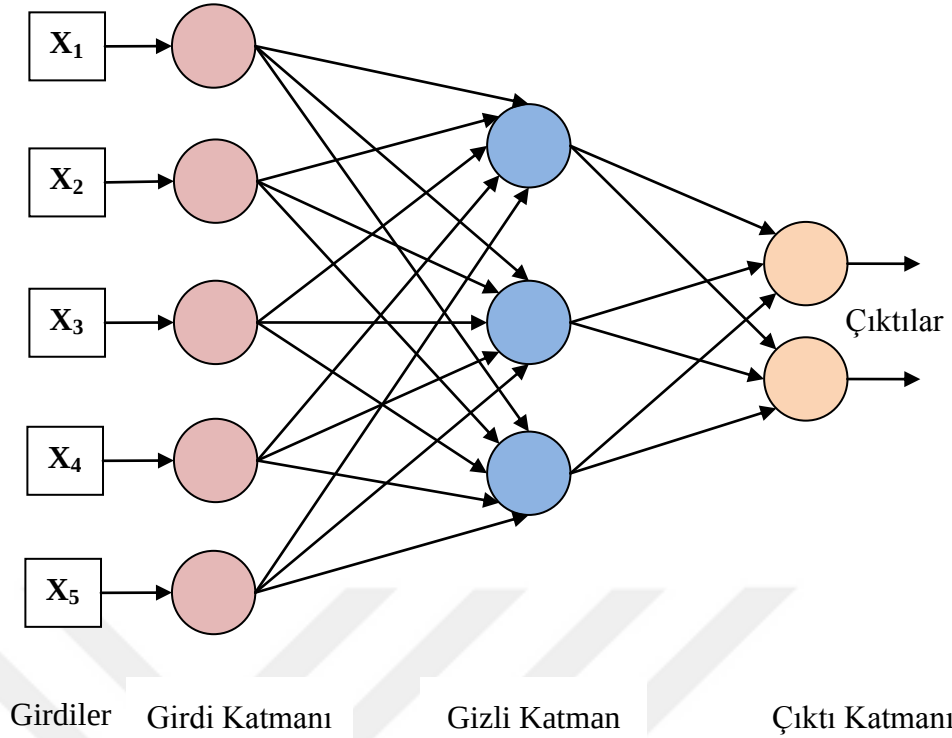
3) Bir düğümden, birbirinden bağımsız transfer fonksiyonlarına sahip sinyaller gidebilir.

2.5.6 YSA Mimarileri

Yapay sinir ağları, yapay sinir hücrelerinin birbirine bağlanmasıyla oluşan yapılardır. Bir yapay sinir ağında, birbirleriyle bağlantılı sinir hücrelerinin yer aldığı girdi katmanı (input layer), çıktı katmanı (output layer) ve gizli (ara) katman (hidden layer) olmak üzere temelde üç katman bulunmaktadır. Girdi Katmanı (input layer) yapay sinir ağına dış dünyadan girdilerin geldiği ilk katmandır. Bu katmanda girdiler genelde herhangi bir işleme uğramadan alt katmanlara iletilmektedir. Gizli (Ara) Katman(lar) (hidden layer(s)) ara katmanda bulunan nöronların dış ortamla bağlantıları yoktur. Yalnızca girdi katmanından gelen sinyalleri alarak, çıktı katmanına sinyal gönderirler. Bazı YSA’larda ara katman bulunmadığı gibi birden fazla gizli katman bulunabilmektedir, diğer bir ifade ile gizli katman sayısı ağdan ağa değişebilmektedir. Gizli katmanlardaki nöron sayıları giriş ve çıkış sayısından bağımsızdır, ayrıca birden fazla gizli katman olan ağlarda gizli katmanların kendi aralarındaki hücre sayıları da farklı olabilmektedir. Gizli katmanların ve bu katmanlardaki nöronların sayısının artması hesaplama karmaşıklığını ve süresini arttırmasına rağmen yapay sinir ağının daha karmaşık problemlerin çözümünde de kullanılabilmesini sağlar. Çıktı Katmanı (output layer) ara katmanlardan gelen bilgileri işleyerek ağın çıktılarını üreten katmandır. Bu katmanda üretilen çıktılar dış dünyaya gönderilir. Geri beslemeli ağlarda bu katmanda üretilen çıktı kullanılarak ağın yeni ağırlık değerleri hesaplanır (Çayıroğlu, 2015; Öztemel, 2012: 52-53; Yavuz ve Deveci, 2012: 172). Sinir ağları, bağlantıların türüne göre 3 farklı kategoriye ayrılırlar, bunlar: İleri Beslemeli Tek Katmanlı, İleri Beslemeli Çok Katmanlı ve Yinelemeli’dir (Haykin, 2009: 21).

2.5.6.1 İleri Beslemeli YSA

İleri beslemeli ağlarda dış ortamdan gelen veriler ilk önce girdi katmanına girer, daha sonra sırayla varsa gizli (ara) katmana ve en son olarak çıktı katmanına gelerek buradan dış ortama iletilirler. İleri Beslemeli ağlarda herhangi bir döngü ya da geri besleme olmadığı için, yayılma işlemi girişten çıkışa doğru olmaktadır. İleri beslemeli bir sinir ağında aynı katmandaki düğümler arasında hiçbir bağlantı bulunmamaktadır, spesifik bir katmanın çıkışları diğer katmana girdi olarak bağlanmaktadır. Bu gösterim modülerlik nedeniyle tercih edilir çünkü aynı katmandaki düğümler aynı özelliğe sahiptir (Kantardzic, 2011: 205). Şekil 2.11, İleri beslemeli çok katmanlı sinir ağını göstermektedir.



Şekil 2.11 İleri Beslemeli Çok Katmanlı Sinir Ağı

Kaynak: Han ve Kamber, 2006: 328

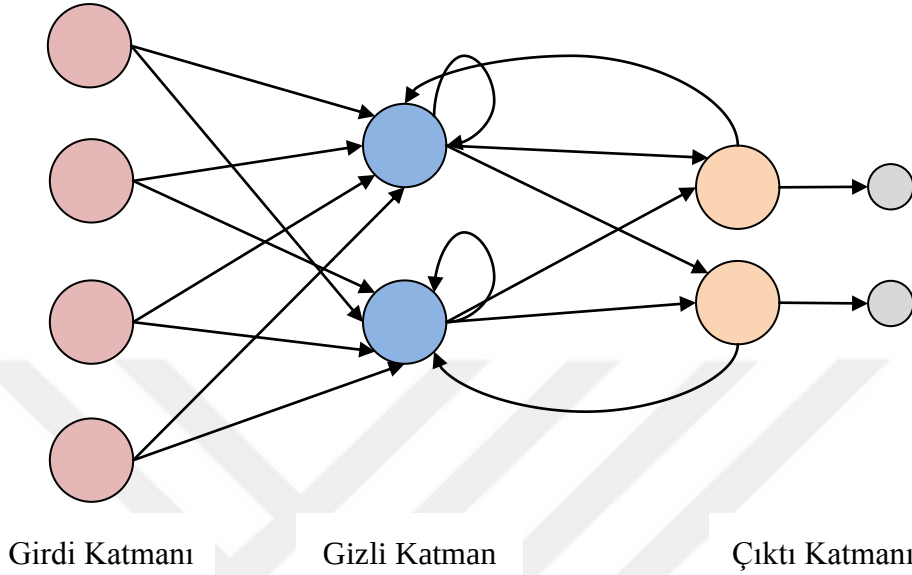
İleri Beslemeli YSA'lar olmak üzere 2'ye ayrılmaktadır.

- **Tek Katmanlı Basit Sinir Ağları:** Bu tip sinir ağlarında sadece çıkış katmanı bulunmaktadır.
- **Çok Katmanlı Sinir Ağları (Gizli Katmanlı/Hidden Layer) :** Bu tip sinir ağlarında bir ya da daha fazla sayıda gizli (ara) katman bulunmaktadır. Girdi katmanlarından gelen ve genellikle ham sinyallerin özelliklerini belirlemek, ağırlıklandırmak, sonuçları çıktı katmanına yönlendirmek için gizli katmanlara ihtiyaç duyulmaktadır. "Gizli Katman" denilmesinin sebebi istenilen çıktı değerini saklamasıdır. Gizli katmandaki sinir hücreleri, ağın giriş ve çıkış davranışlarının incelenmesi yoluyla gözlemlenememektedir (Sunay, 2010: 30).

İleri beslemeli YSA'lar sinir ağları modellerinde en çok tercih edilen (Özellikle Çok Katmanlı YSA) YSA mimarileridir. Wong, Bodnovich, ve Selvi (1997) İleri Beslemeli YSA mimarisinin %95 oranında tercih edildiğini belirlemişlerdir. İleri Beslemeli YSA'lar, girdi değişkenleri ile bir ya da daha fazla çıkış değişkenleri arasındaki ilişkileri modellemek için idealdir ve uygundur (Maimon ve Rokach, 2005a: 422).

2.5.6.2 Yinelemeli YSA

Veri akışının sadece ileriye doğru değil aynı zamanda geriye doğru da olabildiği YSA mimarisidir. En az bir geri besleme döngüsünün olması nedeniyle İleri beslemeli YSA'dan ayrılmaktadır. Şekil 2.12 Yinelemeli bir YSA'yı göstermektedir.



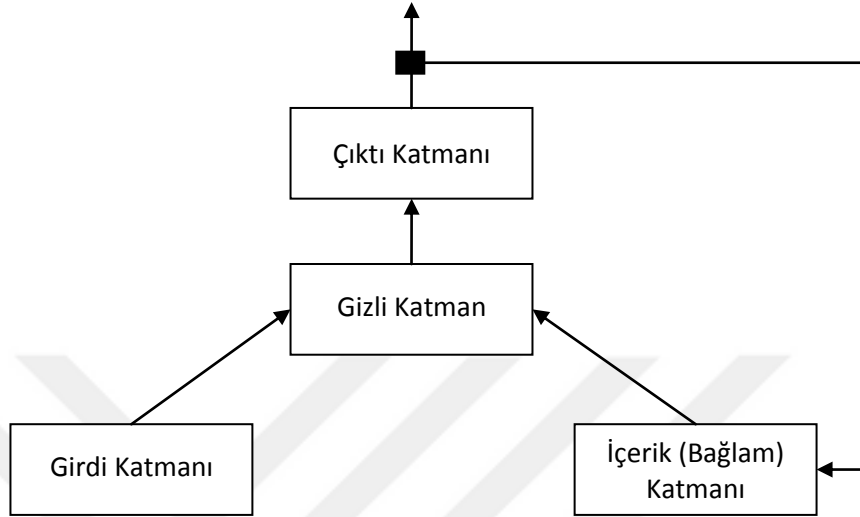
Şekil 2.12 Yinelemeli Yapay Sinir Ağı

Kaynak: Kurenkov, 2016

Bu tip YSA'lar da geri besleme döngülerinde nöronun çıktısı kendi girişine beslenir. Geri beslemeli YSA'da, bir ya da daha fazla hücrenin çıkışı kendisine veya diğer hücrelere giriş olarak verilmektedir. Geri besleme, genelde bir geciktirme elemanı üzerinden yapılır ve bir katmandaki hücreler arasında olabileceği gibi katmanlar arasındaki hücreler arasında da olabilir. Bu yapısı ile lineer olmayan dinamik bir davranış gösterir. Geri beslemenin yapılış şekline göre farklı yapıda ve davranışta geri beslemeli YSA yapıları elde edilebilir. Ayrıca geri beslemeli ağlarda gecikmeler söz konusudur ve ağ yapısı çıkışlar girişlere bağlanarak ileri beslemeli ağlardan elde edilir. Bu ağlara örnek olarak Hopfield, Elman ve Jordan ağları verilebilir (Kulluk S., 2009: 75-76). En basit yapıya sahip olan ve kullanılması en kolay olan ağ Elman ağıdır. Yinelemeli YSA'lar iki şekilde olabilir (Öztemel, 2012: 165).

1. Tam geri dönüşümlü ağlar: Bu ağlar rastgele ileri ve geri bağlantıları olan ağlardır ve bu bağlantıların hepsi eğitilebilirler.
2. Kısmi geri dönüşümlü ağlar: Bu ağlarda ağın işlem elemanlarına ek olarak içerik elemanları da bulunmaktadır. Temelde ileri beslemeli ağlardır ve ileri bağlantılar eğitilebilirler. Geri dönüşüm sadece içerik elemanları üzerinden yapılır ve bu bağlantılar eğitilemezler.

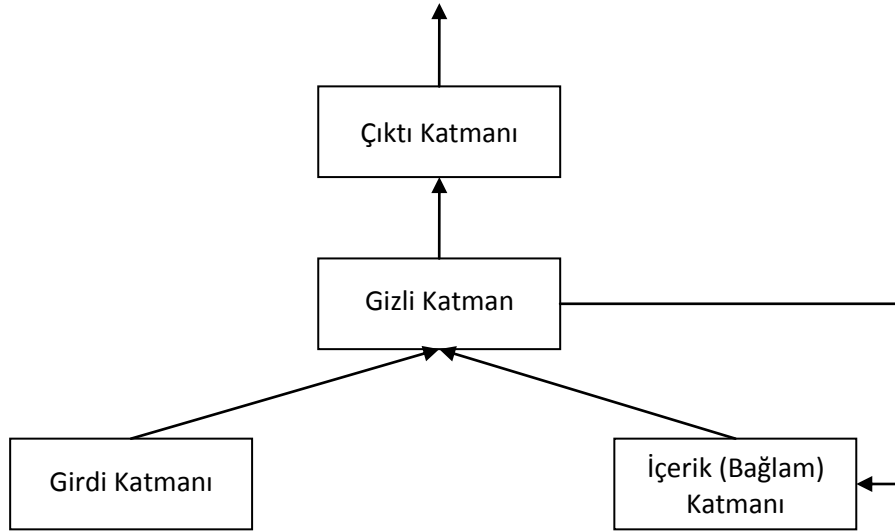
En eski yinelemeli ağ Jordan ağıdır. Girdi katmanı, gizli katman, çıktı katmanı ve durum birimlerinden (State Units) oluşmaktadır (Kröse ve Van Der Smagt, 1996: 48). Durum birimleri, çıktı katmanından aldıkları aktivasyon değerlerini bir sonraki iterasyona aktarmakla görevlidir (Heaton Research, 2015). Şekil 2.13 Jordan ağını göstermektedir.



Şekil 2.13 Jordan Ağı

Kaynak: Paola: 2015

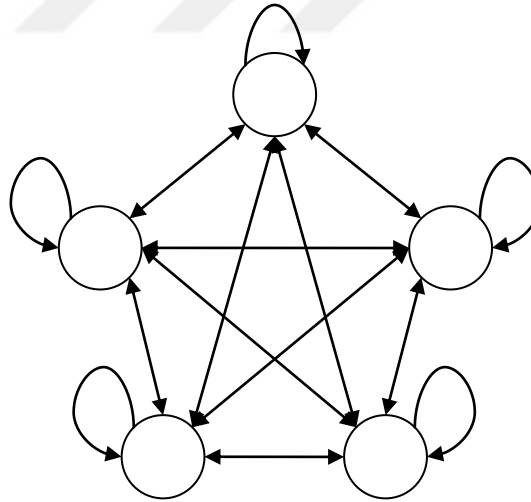
Elman ağı, Elman tarafından 1990 yılında geliştirilmiştir, girdi katmanı, gizli katman, çıktı katmanı ve içerik (bağlam) katmanından oluşmaktadır. Ekstra girdi birimi olan içerik katmanının aktivasyon değerleri, geri besleme yoluyla gizli katmandan gelmektedir. Elman ağı, Jordan ağına benzemekle beraber iki farklılığı bulunmaktadır. Bunlardan ilki geri besleme yaptıkları aktivasyon değerlerini çıktı katmanından almak yerine ara katmandan almaları, ikincisi içerik elemanlarının kendilerine bağlantılarının bulunmamasıdır (Kröse ve Van Der Smagt, 1996: 48-49). Elman ağı temelde ileri beslemeli ağlar gibi standart geri yayılım algoritması ile eğitilmektedir. Algoritmanın uygulanmasında ağın sadece ileri besleme bağlantıları eğitilmektedir. Geri besleme bağlantıları ise kullanıcının önceden deneysel olarak belirlediği değerlerde sabit kabul edilmektedir, ayrıca geri besleme bağlantılarının ağırlık değerleri eğitme başarısı üzerinde oldukça etkili olup, uygun değerlerin seçilmesi son derece önemlidir (Kalinlı, 2002: 27). Şekil 2.14 Elman ağının yapısını göstermektedir.



Şekil 2.14 Elman Ağı

Kaynak: Kröse ve Van Der Smagt, 1996: 49

Hopfield ağı, işlem elemanlarının tamamının hem girdi hem de çıktı elemanları olduğu tek katmanlı ve geri dönüşümlü bir ağıdır (Öztemel, 2012: 170). Hopfield ağı hiçbir dış girişe sahip değildir. Ağın davranışları özerk ve sadece kendi dinamiklerine bağlıdır (Dreyfus, 2005: 274). Şekil 2.15 Hopfield ağının yapısını göstermektedir.



Şekil 2.15 Hopfield Ağı

Kaynak: Mehrotra, 1996: 227

Genellikle ikili (0 veya 1) ve bipolar (-1 veya +1) girişleri kabul eden Hopfield ağında bağlantılar iki yönlüdür yani bilgi her iki yönde de akmaktadır ve simetriktr. Bu ağda, ağın bağlantı değerleri bir enerji fonksiyonu olarak saklanır. Her iki yönde akan bilgiye uygulanan ve her bağlantı için hesaplanan bir ağırlık değeri vardır (Baş, 2006: 75). Hopfield ağında, i

nöronundan j nöronuna olan bağlantının ağırlığı, j nöronundan i nöronuna olan bağlantının ağırlığına eşittir (Dreyfus, 2005: 274).

2.5.7 YSA'da Öğrenme

Bir YSA'nın öğrenmesi, ağıdaki tüm birimleri birbirine bağlayan kenarların ağırlıklarının belirlenmesi işidir. Amaç eğitim setini kullanarak, YSA'da bulunan çıktı ile istenilen çıktının birbirine en yakın olmasını sağlayacak şekilde ağırlıkları hesaplamaktır. Geri yayılım metodu adı verilen bir yöntem kullanılarak gerekli ağırlık düzenlemeleri yapılır. Geri yayılım 3 aşamada çalışır: (Berry ve Linoff, 2004: 228-229; Kantardzic, 2011: 208)

1. YSA bir eğitim seti alır ve mevcut ağırlıkları kullanarak çıkış ya da çıkışları hesaplar,
2. Geri yayılım, hesaplanan sonuç ile beklenen sonuç arasındaki farkı alarak hata hesaplanır.

$$\begin{array}{ccc}
 \text{örnek}_k & \text{Girdiler} & \text{Çıktı} \\
 & x_{k1}, \dots, x_{kn} & d_k \\
 \\
 & y_k = f\left(\sum_{i=1}^m x_i w_{ki}\right), & \text{ağda hesaplanan değer} \\
 \\
 & e_k(n) = d_k(n) - y_k(n), & \text{hata değeri}
 \end{array}$$

YSA'da öğrenmenin amacı, hata fonksiyonunu minimum yapmaktır.

$$E(n) = \frac{1}{2} e_k^2(n), \quad \text{hata fonksiyonu}$$

3. Hata ağ üzerinden geri beslenerek ağırlıklar yeniden düzenlenir.

$$\Delta w_{kj}(n) = \eta \cdot e_k(n) \cdot x_j(n)$$

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

(Kantardzic, 2011: 208)

YSA'lar klasik programlama da olduğu gibi belirli bir algoritma çerçevesinde programlanmazlar, eğitilirler (Yakut, 2012: 61).

2.5.7.1 YSA'da Öğrenme Stratejileri

Belirli bir amaç için yapılandırılmış ağ eğitime (öğrenmeye) hazırdır. İşlem ilk olarak rastgele seçilen ağırlıklar ile başlar. Daha sonra eğitim (öğrenme) başlar (Andersen ve McNeill, 1992: 10). Öğrenme stratejilerinin denetimli ve denetimsiz olmak üzere ikiye ayrıldığı görülmüştür (Zurada, 1992: 56; Andersen ve McNeill, 1992: 10). Haykin öğrenme stratejilerini Öğretmenli ve Öğretmensiz olmak üzere ikiye ayırmıştır. Öğretmenli öğrenme denetimli öğrenme olarak bilinmektedir, öğretmensiz öğrenme denetimsiz öğrenme ve takviyeli öğrenme olarak ikiye ayrılmaktadır (Haykin, 2009: 34).

Öğretmenli Öğrenme: Öğretmenli öğrenme stratejisinde, öğrenen sistemin öğrenebilmesine bir öğretmen yardımcı olmaktadır (Öztemel, 2012: 25). Bu öğrenme stratejisinde girdi ve çıktılar birlikte kullanılmaktadır. Ağ girdileri işler ve elde ettiği çıktıları beklenen çıktılar ile karşılaştırır. Daha sonra hatalar geriye doğru yayılır ve sistem ağı kontrol etmek için ağırlıkları değiştirir. Bu öğrenme modelinde eğitim sağlayan veri seti “eğitim seti” olarak adlandırılır. Ağın eğitimi sırasında aynı veri seti birçok kez işlenir ve ağırlıklar düzenlenir (Andersen ve McNeill, 1992: 10). LVQ ağı ve Çok Katmanlı Algılayıcı Ağı örnek olarak verilebilir (Öztemel, 2012: 25; Taylor, 2006: 70). Ayrıca Öğretmenli öğrenme algoritmalarına örnek olarak; Widrow ve Hoff tarafından geliştirilen delta kuralı, Rumelhart ve McClelland tarafından geliştirilen genelleştirilmiş delta kuralı algoritması örnek olarak verilebilir (Vural, 2007: 29).

Öğretmensiz Öğrenme: Bu öğrenme stratejisinde sistemin öğrenmesine yardımcı olan herhangi bir öğretmen yoktur. Sisteme sadece girdi değerleri gösterilir, örneklerdeki parametreler arasındaki ilişkileri sistemin kendi kendisine öğrenmesi beklenir. Daha çok sınıflandırma problemleri için kullanılan bir öğrenme yöntemidir. Sadece, sistemin öğrenmesi bittikten sonra çıktıların ne anlama geldiğini gösteren etiketlendirmenin kullanıcı tarafından yapılması gerekmektedir (Öztemel, 2012: 25). Bu ağlar kendi ağırlıklarını ayarlamak için dışarıdan düzenlemeye gerek duymazlar. Bunu yerine kendi iç performanslarını izlemektedir. Bu ağlar girdi sinyallerinde düzen ve eğilim aramaktadır. Doğru veya yanlış olup olmadığı belirtilmeden, ağ kendisini nasıl organize edeceği ile ilgili bilgilere sahip olmaktadır. Bu bilgi, ağ topolojisinin ve öğrenme kurallarının içerisine yerleştirilmiştir (Andersen ve McNeill, 1992: 28). ART ve SOM ağları örnek olarak gösterilebilir (Taylor, 2006: 80-81).

Takviyeli Öğrenme: Bu öğrenme stratejisinde öğrenen sisteme bir öğretmen yardımcı olur. Fakat öğretmen her girdi seti için olması gereken çıktı setini sisteme göstermek yerine sistemin kendisine gösterilen girdilere karşılık çıktısını üretmesini bekler. Üretilen çıktının doğru veya yanlış olduğunu gösteren bir sinyal üretir. Sistem, öğretmenden gelen bu sinyali

dikkate alarak öğrenme sürecini devam ettirir. Boltzmann kuralı bu stratejiye örnek olarak verilebilir (Kulluk, 2009: 77-78).

Karma Stratejiler: Üst kısımda anlatılan üç stratejiden birkaçını birlikte kullanarak öğrenme gerçekleştiren ağlarda bulunmaktadır. Bu ağlar, kısmen danışmanlı kısmen de danışmansız olarak öğrenme gerçekleştiren ağlardır. Radyal Tabanlı YSA (RBN) ve Olasılık Tabanlı YSA (PBNN) bunlara örnek olarak verilebilir (Öztemel, 2012: 25).

2.5.7.2 Temel Öğrenme Kuralları

Birçok öğrenme kuralları ortak olarak kullanılmakta olup, en çok bilinen ve en eski olanı Hebb kuralıdır. Diğer kuralların birçoğu Hebb kuralının çeşitli varyasyonlarıdır. Bazı araştırmacılar biyolojik öğrenme modelini ana hedef olarak belirlemiştir, diğer araştırmacılar ise doğanın öğrenmeyi nasıl işleyeceğini kendi algılarının uyarlamaları ile denemektedir. Her iki şekilde de nöral işlemenin nasıl çalıştığını insanın anlaması oldukça sınırlıdır. Öğrenme şu anda geliştirilen öğrenme yasalarından daha karmaşıktır (Andersen ve McNeill, 1992: 29). Literatürde mevcut öğrenme algoritmalarının çoğu Hebb, Delta, Kohonen ve Hopfield olmak üzere dört farklı öğrenme kuralından esinlenilerek geliştirilmiştir.

Hebb Kuralı: İlk ve şüphesiz en iyi bilinen ve diğer öğrenme kurallarının temelini oluşturan öğrenme kuralı Donald Hebb tarafından tanıtılmıştır. Hebb'in temel kuralı, eğer nöron diğer bir nörondan bilgi alırsa ve her iki hücre de aktif ise (matematik olarak aynı işareti taşıyorsa) nöronlar arasındaki ağırlık güçlendirilmelidir (Andersen ve McNeill, 1992: 29).

Hopfield Kuralı: Bu kural, Hebb kuralına benzemektedir. Yapay sinir ağı elemanlarının ne kadar kuvvetlendirilmesi veya zayıflatılması gerektiği belirlenir. Eğer beklenen çıktı ve girdiler ikisi de aktif veya pasif ise öğrenme katsayısına göre ağırlık değerlerini arttır ya da ağırlık değerlerini azalt şeklindedir. Öğrenme katsayısı genel olarak 0-1 arasında kullanıcı tarafından atanan sabit ve pozitif bir değerdir (Öztemel, 2012: 26; Andersen ve McNeill, 1992: 29).

Delta Kuralı: Bu kural, Hebb kuralının bir başka varyasyonudur. Bu kurala göre beklenen çıkış değeri ile gerçekleşen çıkış değeri arasındaki farkı azaltmak için giriş bağlantılarının ağırlık değerlerinin sürekli değiştirilmesi ilkesine dayanarak geliştirilmiştir. Bu kural, ağırlık ortalaması karesel hatasını minimize edecek şekilde sinaptik ağırlıkları değiştirmektedir (Andersen ve McNeill, 1992: 29). Ağırlık ürettiği çıktı ile beklenen çıktı arasındaki hatanın karelerinin ortalamasını en küçüklemek hedeflenmektedir (Öztemel, 2012: 27).

Kohonen Kuralı: Biyolojik sistemlerdeki öğrenmeden esinlenerek Kohonen tarafından geliştirilmiştir. Bu kuralda nöronlar öğrenmek ya da ağırlıklarını güçlendirmek için rekabet

ederler. En büyük çıkışa sahip işlem elemanı kazanan olarak deklare edilir. Bu işlem elemanı (nöron) diğer komşu işlem elemanlarını inhibe etme yeteneğine sahip olmaktadır. Sadece kazanan nöron için bir çıkışa izin verilir ve bu nöronun komşu nöronların ağırlıklarını ayarlamasına izin verilir (Andersen ve McNeill, 1992: 29) .

2.5.8 YSA'da Ağ Seçimi

Tüm YSA'lar nöronlar, bağlantılar ve transfer fonksiyonu kavramlarına bağlı olduğundan sinir ağlarının farklı yapıları ve mimarileri arasında benzerlikler bulunmaktadır. Varyasyonların çoğu çeşitli öğrenme kurallarından ve bu kuralların bir ağ topolojisini nasıl değiştirdiğinden kaynaklanmaktadır. Aşağıdaki tablo da en sık görülen YSA'lar kabaca özetlenmiştir. Bu kategoriler özel olmamakla birlikte, bazı özel uygulamalar için ağ mimarileri karışıklıklarını ayırmak içindir. Temel olarak sinir ağlarının uygulamaları 5 ana kategoriye ayrılmaktadır. Bunlar (Andersen ve McNeill, 1992: 31):

- Tahmin
- Sınıflandırma
- Veri ilişkilendirme
- Veri kavramsallaştırma
- Veri filtrelemedir.

Tablo 2.9 Temel olarak sinir ağlarının uygulama alanlarını göstermektedir.

Tablo 2.9 Temel Olarak Sinir Ağlarının Uygulama Alanları

Kullanım Amacı	Ağ	Kullanma Nedeni
Tahmin	• Geri Yayılma • Delta Bar Delta • Genişletilmiş Delta Bar Delta • Yönlendirilmiş rastgele arama • Yüksek Mertebeden Sinir Ağları • Geri Yayılma İçerisine Kendi Kendine Organize Olabilen Haritalar (Self Organizing Map into Backpropagation)	Girdi değerlerini kullanarak bazı çıktı değerlerinin tahmini Örneğin; Piyasadaki en iyi hisse senedini almak, hava tahmini, insanların kanser risklerini belirlemek vb
Sınıflandırma	• Learning Vector Quantization (LVQ) • Olasılıksal Sinir Ağı (PBNN) • Karşı Yayılım (Counter Propagation)	Girdi değerlerini kullanarak sınıflandırma belirlemek
Veri İlişkilendirme	• Hopfield	Sınıflandırma gibi fakat

	• Boltzmann Makinesi • Hamming Ağı • Çift Yönlü İlişkisel Bellek • Mekan/Uzaysal-Zamansal Örüntü Tanıma	verilerin içindeki hataların bulunması
Veri Kavramsallaştırma	• Adaptif Rezonans Ağı (ART) • Kendiliğinden Organize Olabilen Haritalar (SOM)	Girdilerin analizi ile gruplama ilişkiler çıkarılması
Veri Filtreleme	• Devirdaim (Recirculation)	Giriş sinyalini düzgünleştirmek

Kaynak: Andersen ve McNeill, 1992: 31

ÜÇÜNCÜ BÖLÜM

HANEHALKI YAĞ TÜKETİMİ

3.1 Yağlar

Yağlar insan vücudunun beslenmesi, yaşam döngüsünün devam ettirilebilmesi ve günlük enerji ihtiyacının karşılanması açısından son derece önemli kaynaklardır. Yetişkin bir insanın günlük aktiviteleri için 2000-3000 kalori gerekli olduğu ifade edilirken, dengeli ve sağlıklı beslenmenin gereği olarak da bu miktarın 650-900 kalorisinin yağlardan karşılanması gerekmektedir. Bunun sağlanması için bir insanın günde yaklaşık 95 gram yağ tüketmesi gerekir. Bu miktarda yağ, sıvı olarak yemeklerle, katı yağ olarak kahvaltılarda ve peynir, süt vb. besinlerden karşılanabilmektedir. Buna göre dolaylı alımların dışında günde 63 gram doğrudan yağ alınması uygun görülmektedir. Bu ise kişi başına yılda yaklaşık 23 kg yağ tüketilmesi demektir (Kolsarıcı vd., 2006: 66). Amerikan Kalp Birliği Beslenme Komitesine göre günlük alınan enerjinin %25-%35'i yağlardan alınmalı, bunlar için balık, fındık ve bitkisel yağlar gibi ürünler kullanılabileceğini belirtmiştir. Günlük alınan kaloringin %7'den azı doymuş yağlardan elde edilebilirken (140 kalori ya da 16 gr yağ), günlük alınan kaloringin %1'den azı trans yağlardan elde edilebilir (20 kalori ya da 2 gr yağ) (Heart Org, 2015).

Ülkemizde yağ tüketimi ağırlıklı olarak bitkisel yağlardan oluşmaktadır ve hayvansal yağlara olan talep düşüktür, bunun nedeni ise hayvansal yağ ürünlerinin pahalı olmasıdır. Gıda sektöründe talebin yaklaşık yüzde 13'ünü zeytinyağı hariç diğer bitkisel yağ sanayi karşılamaktadır (Kaderli, 2006: 16). Türkiye'de ise yağ tüketimi gelişmiş ülkelerdeki tüketim değerlerinin gerisinde olsa da, yüksek değerlere sahiptir (Polat, 2010: 1). 2000 yılında, toplam dünya tüketiminin yağlar (83,3 milyon ton) %76,2'sini oluşturmuştur. Bu gıda olarak kişi başına tüketim yaklaşık 12,9 kg (sıvı yağ: 10,4 kg, katı yağ: yağlar 2,5 kg) olduğu anlamına gelir (Sayılı vd., 2005: 716). Türkiye'de 2003 yılı verilerine göre zeytinyağı hariç bitkisel yağ tüketimi, ham yağ eşdeğeri olarak 17,2 kg iken AB-15'de bu değer 30,3 kg, ABD'de 29,3 kg, Kanada'da 27 kg civarında olmuştur. Ülkemizde sıvı bitkisel yağ 750.000 ton, margarin ise 450.000 ton olarak tüketildiği tahmin edilmektedir. Bitkisel yağ üretimi ise Türkiye'de gıda sanayi içinde % 8,5 civarında bir paya sahiptir (Polat, 2010: 1-2).

3.1.1 Zeytinyağı

Zeytinyağı vücut için gerekli, ancak sentez edilemeyen temel yağ asitleri ile sadece yağda eriyebilen E vitamininin kaynağını oluşturmaktadır. Yüksek kalori değerinin yanı sıra, meyve suyu gibi natürel tüketilebilen tek yağ olma özelliği ile de insan beslenmesinde çok

önemli bir konuma sahiptir. Zeytinyağı A, D, E ve K vitaminleri ile kalsiyum, fosfor, potasyum, kükürt, magnezyum, az miktarda demir, bakır ve manganez gibi mineraller içermektedir (Tariş, 2015). Yalnızca fiziksel işlemlerle elde edilen zeytinyağı, yüksek miktarda tekli doymamış yağ, antioksidan, fenolik bileşikler içermektedir (Gimeno vd., 2002: 114). Zeytinyağı herhangi bir kimyasal işlemde geçirilmeden üretilebildiği için tüketilirken de vitamin varlığını, temel yağ asitlerini ve diğer besleyici önemdeki doğal maddeleri muhafaza edebilme özelliğine sahiptir (Tunalıoğlu vd., 2003: 49).

Zeytinyağı tüketimi günümüzde insanlar tarafından yağ gereksinimlerinin karşılanmasının sağlıklı bir yolu olarak görülmektedir. Zeytinyağının kökleri çok eskiye dayanmaktadır ve içerdiği antioksidanlar, karotenoidler, fenolik bileşikler ve vitaminler bakımından beslenme ve sağlık üzerindeki önemi büyüktür. Zeytinyağı, içerdiği bileşikler nedeni ile hücre zarı akışkanlığı, lipid peroksidasyonu ve yaşlanmaya neden olan serbest radikalleri bağlayıcı özelliği ile de oksidasyonu engelleyerek kötü kolesterolü azaltmakta ve iyi kolesterolü artırmaktadır. Bu bilgiler doğrultusunda yağlardan alınan enerjinin, günlük kalori ihtiyacının %30'unu aşmaması ve bunun da %60 ile %70'inin zeytinyağından sağlanacak şekilde beslenilmesi önerilmektedir (TZOB, 2010: 280).

Zeytinyağı, yağların içinde en kolay hazmedileni olup, başta kalp-damar hastalıkları olmak üzere sindirim sistemi, kemik yapısı beyin ve sinir dokuları üzerinde çok önemli fonksiyonları bulunmaktadır. Kandaki kötü huylu kolestrolü (LDL) temizleyen zeytinyağı, iyi huylu kolesterolü de (HDL) yükseltmektedir. Çocukların raşitizmden korunmasına yardımcı olmaktadır (Tariş, 2015).

Dünya'da ve Türkiye'de Zeytinyağı Üretimi ve Tüketimi:

Zeytinyağı üretimi çoktan aza doğru sırasıyla İspanya, İtalya, Yunanistan, Tunus, Türkiye, Suriye, Fas ve Portekiz gibi Akdeniz'i çevreleyen ülkelerde toplanmıştır. Zeytin ağacı, dünya üzerindeki beş kıtada, Kuzey ve Güney yarı küresinde özellikle Akdeniz iklimine sahip 30-45 derece enlemler arasında yayılmış, kuzey yarı kürede 30 ve güney yarı kürede de 8 ülkedeki sınırlı bir alanda yetiştiriciliği yapılan bir Akdeniz bitkisidir. Dünyadaki zeytin ağacı varlığının %98'i Akdeniz havzası olarak da adlandırılan bu bölgede hâkim durumdadır. Akdeniz havzasında bulunan yukarıda belirtilen bu sekiz ülke, dünya zeytinyağı üretiminin %90'dan fazlasını tek başına sağlamaktadır (Öztürk vd., 2009: 35-36). Tablo 3.1'de Türkiye ve Dünyada 2013 yılına ait zeytinyağı üretim, kişi başı tüketim ve toplam tüketim miktarları görülmektedir.

Tablo 3.1 Dünya Zeytinyağı Üretim ve Tüketimi

Ülke Adı	Üretim (1.000 Ton)	Tüketim (1.000 Ton)	Zeytinyağı Tüketimi (Kişi Başı/Kg)
İspanya	618,2	486,9	10,4
İtalya	415,5	550	9,2
Yunanistan	357,9	180	16,3
Tunus	220	40	3,7
Türkiye	195	150	2
Suriye	175	160,5	7
Fas	100	129	3,9
Cezayir	66	60,5	1,5
Portekiz	59,2	74	7,1
Ürdün	21,5	20	3,1
İsrail	18	19,5	2,4
Arjantin	17	6	0,1
Mısır	16,5	12	0,1
Filistin	15,5	13,5	3,2
Libya	15	15	2,4
Şili	15	6	0,3
Lübnan	14	20	4,5
Arnavutluk	12	13	4,7
Avustralya	9,5	37	1,6
Fransa	5,1	113,1	1,7
ABD	4	287	0,9
Karadağ	0,5	0,5	0,8
Lüksemburg	0	3	5,5
İsviçre	0	13,5	1,7
Kanada	0	37	1,1
İngiltere	0	62	1
Almanya	0	60,5	0,8

Kaynak: Olive Oil Market, 2015, International Olive Council, 2015

Tablo 3.1’den görüldüğü üzere zeytinyağı üretiminin en çok yapıldığı ülkelerde genellikle zeytinyağı tüketiminin de yüksek olduğu görülmektedir. Zeytinyağı üretiminin en çok yapıldığı ülkeler sırasıyla İspanya, İtalya, Yunanistan, Tunus, Türkiye ve Suriye, en çok

tüketimin yapıldığı ülkelere sırasıyla İtalya, İspanya, ABD, Yunanistan, Suriye ve Türkiye olarak görülmektedir. İtalya ve Fas'ta üretim tüketimi karşılamazken Tunus ve Yunanistan'da üretim tüketimin üzerindedir. Kişi başı zeytinyağı tüketiminin en yüksek olduğu ülkelere sırasıyla Yunanistan, İspanya, İtalya, Portekiz, Suriye ve Lüksemburg olarak görülmektedir. Kişi başı zeytinyağı tüketimi Yunanistan'da 16,3 litre, İspanya'da 10,3 litre, İtalya'da 9,2 litre, Portekiz'de 7,1 litre, Suriye'de 7 litre, Lüksemburg'da 5,5 litre olarak görülmektedir. Bu ülkelerde kişi başı zeytinyağı tüketimi 5 litrenin üzerindedir. Türkiye Akdeniz havzasında yer alan ve zeytinyağı üreticisi olan bir ülke olmasına rağmen yeterli tüketim düzeyine ulaşamamıştır, ülkemizde kişi başı zeytinyağı tüketimi 2 litre olarak görülmektedir. Kişi başı zeytinyağı tüketiminin yüksek olduğu ülkelerin zeytinyağı üreten ülkeler olması son derece anlamlıdır, tablodan da görüldüğü üzere kişi başı en çok tüketim genellikle en çok üretimin yapıldığı ülkelerde gerçekleşmektedir. Lüksemburg bu konuda bir istisna oluşturmaktadır, çünkü üretim miktarı sıfır iken kişi başı tüketim miktarı 5,5 litredir. Bu durum Lüksemburg'da yaşayan bireylerin kişi başı gelirinin yüksek ve eğitim düzeyinin yüksek olmasından kaynaklanabilir.

3.1.2 Margarin ve Diğer Bitkisel Katı Yağlar

İnsan tüketimine uygun bitkisel ve/veya hayvansal yağlardan elde edilen, süt yağı içeriğine göre tanımlanan, temel olarak yağ içinde su emülsiyonu tipinde, süt ve/veya süt ürünleri içerebilen, şekillendirilebilen bir ürün olarak tanımlanmaktadır. Margarin, 2 Ekim 1869 yılında Fransız kimyager Mège Mouriès tarafından elde edilmiştir. Tamamen hayvansal orijinli don yağı, mide özsuğu ve süt karışımının soğutulması ile ürünü elde etmiş ve patentini almıştır. Margarine ilk olarak bitkisel yağ karıştırılması 1877'de %5-10 oranında gerçekleşmiştir (Başaran, 2015). Margarinin bu yollarla üretimi 40 yıl kadar sürmüş, Alman kimyager W. Norman'ın hidrojenasyon prosesini keşfi ile önemli bir yol katetmiştir. 1909 yılında likit yağların hidrojen gazı ile sertleştirilmiş, 1940 yılında kristalizasyon prosesinde yaşanan gelişmeler ile margarin tam olarak tereyağı benzeri bir ürün haline gelmiştir. Margarin üretiminde yaşanan bu gelişmeler 20. yüzyılın ikinci yarısında da devam etmiştir, 1960 yılında fraksiyon, 1980 yılında inferesterifikasyon tekniklerinin gelişmesiyle devam etmiştir (Mümsad, 2015).

Türk standartlarında bitkisel margarin, çeşitli bitkisel yağların kısmi olarak hidrojene edilmeleri ile elde edilen sertleştirilmiş rafine yağlardan veya bu yağlara çeşitli rafine bitkisel yağların karıştırılmasından elde edilen ve/veya içerisinde emülsiyon halinde su, pastörize fermente yağsız süt, pastörize taze süt, süt tozu ve peynir tozu ile katkı maddeleri bulunabilen

ürün olarak tanımlanmaktadır (BYSD, 2016b). Margarin hem kahvaltılık hem de yemeklik olarak kullanılan bir yağdır. Kahvaltı malzemesi olarak tereyağı ile margarin ikame ürün olurken, yemek, pasta, börek vb. yiyeceklerin hazırlanmasında margarine ikame olan ürünler tereyağı ve bitkisel sıvı yağlardır (Türkiye Kalkınma Bankası A.Ş., 1999: 3).

Margarin üretiminde kullanılan bitkisel yağlar; haşhaş, pamuk, mısır, soya, kanola, ayçiçeği, palm, zeytin, hindistan cevizi, badem ve fındıktır (Başaran, 2015; BYSD, 2016b). Bu nedenle margarin üretimi bitkisel sıvı yağ ve yağlı tohumlar sektöründen etkilenmektedir. Margarin üretimi genel olarak margarin hammaddesi olan pamuk ve soya yağının üretimine bağlıdır. Son yıllarda sağlıklı olarak kabul edilen bitkisel sıvı yağ talebi ve tüketimi hızla artarken, margarin pazarı azalan bir hızla büyüme göstermektedir. Margarin tüketiminde kırsal kesimin payı artarken, kent kesiminin payı azalmaktadır. Tereyağının daha pahalı ve az dayanıklı olması tüketici tercihinde margarini ön plana çıkarmaktadır, bu durum ise sgenellikle dar gelirli tüketici grubunda etkilidir. (Türkiye Kalkınma Bankası A.Ş., 1999: 4-5).

Tablo 3.2. Türkiye margarin üretim verilerini göstermektedir (Mümsad, Türkiye Margarin Üretim İstatistikleri, 2016). Tablo 3.2 incelendiğinde paket ve kase margarin üretimi yıllar içerisinde dalgalanma göstermektedir, fakat 2006 yılından itibaren üretimde azalma görülmektedir. Pastacılık yağ ve margarinleri ile endüstriyel yağların üretimi ise zamanla giderek artmaktadır.

Tablo 3.2 Türkiye Margarin Üretimi

	2005 (Ton)	2006 (Ton)	2007 (Ton)	2008 (Ton)	2009 (Ton)	2010 (Ton)	2011 (Ton)	2012 (Ton)	2013 (Ton)	2014 (Ton)	2015 (Ton)
Paket	125.748	126.743	124.726	121.166	113.374	105.426	92.376	88.733	86.126	84.164	88.501
Kase	39.015	32.367	33.449	32.151	34.410	31.254	28.751	28.148	28.566	30.015	32.276
Toplam Sofra Margarin	164.763	159.110	158.175	153.317	147.784	136.680	121.127	116.881	114.692	114.179	120.777
Yemeklik Margarin	64.849	54.533	45.410	51.282	36.126	36.632	28.147	36.594	40.430	31.762	22.460
Pastacılık Yağ ve Margarinleri	58.199	83.977	84.810	90.479	98.007	105.509	106.006	108.941	125.231	160.958	156.896
Endüstriyel Yağlar	199.997	236.654	265.926	240.733	235.991	253.294	249.838	243.397	290.282	316.492	334.262
Endüstriyel Margarin	258.196	320.631	350.736	331.212	333.998	358.803	355.844	352.338	415.513	477.450	491.158
Toplam İç Pazar	487.808	534.274	554.321	535.811	517.908	532.115	505.118	505.813	570.635	623.391	634.395
Toplam İhracat	144.185	106.363	114.257	96.034	91.000	102.452	121.287	109.000	153.447	156.620	152.058
Genel Toplam	631.993	640.637	668.578	631.845	608.908	634.567	626.405	614.979	724.082	780.011	786.453

Kaynak: Mümsad, Türkiye Margarin Üretim İstatistikleri, 2016

3.1.3 Bitkisel Sıvı Yağlar

Bitkisel yağlar, yağlı tohumla sahip bitkilerden ve onun çekirdeklerinden elde edilir. Ülkemizin bulunduğu coğrafi konum, iklim, toprak yapısı ve sulama imkânlarıyla birçok bitkinin yetişebileceği bir konumdur. Bu nedenle ülkemizde bitkisel yağ üretiminin ve çeşitliliğin yüksek olması beklenir.

Dünya genelinde tarımı yapılan yağlı tohumların soya fasulyesi, kanola, pamuk tohumu, yer fıstığı, ayçiçeği ve palm çekirdeği olduğu görülürken ülkemizde ise ayçiçeği, çığit, soya, yerfıstığı, haşhaş, susam, kolza ve aspir bitkilerini sayabiliriz. Pamuk tohumu (çığit) aslında yağ bitkisi değildir fakat ülkemizin bitkisel yağ sanayine önemli katkı sağlaması bakımından bu gruptandır (Kooperatifçilik Bölge Müdürlüğü, 2015: 3-4).

Dünyada ve Türkiye’de yağlı tohumlar, bitkisel yağ ve mamulleri sektörünün, son yirmi-yirmibeş yıl içerisinde büyük bir ivme kazandığı görülmektedir. Küresel yağlı tohumlar ve bitkisel yağ pazarının geçen yıllar içerisinde yüksek bir büyüme eğilimi içerisinde olduğu görülmektedir. Bitkisel yağlar gıda, yem (hammadde veya katkı maddesi), enerji ve kimyasal sektörlerde yoğun olarak kullanılan stratejik bir ürün haline gelmiştir. Bununla birlikte son yıllarda biyoyakıt talebinin artması da yağlı tohum üretimini etkilemiştir ve tüketimi giderek artan önemli bir ürün haline gelmiştir. (Kooperatifçilik Bölge Müdürlüğü, 2015: 3-4; Kooperatifçilik Bölge Müdürlüğü, 2014: 3)

Ülkemizde son yıllarda bitkisel yağ tüketimi artış göstermesine rağmen yıllar itibarıyla yağ ihtiyacını karşılayacak düzeyde üretim gerçekleştirilememiştir. İklim ve toprak özellikleri dikkate alındığında, yağlı tohumlu bitkilerin üretimi bakımından büyük bir potansiyele sahip olmasına rağmen, yağlı tohum üretim maliyetlerinin yüksekliği, dolayısıyla yetiştirildikleri bölgelerdeki alternatif ürünlerle rekabet edememesi, ürün paritelerinin yağlı tohumlar aleyhinde olması, dünya ham yağ fiyatlarının daha düşük olması ve uygulanan politikaların etkin olmaması bunun başlıca nedenleri arasında gösterilebilir. Bu nedenle bitkisel yağ sanayi hammadde yönünden %70 düzeyinde dışa bağımlıdır (Top ve Uçum, 2012: 6-7).

1997-2004 yılları arasındaki 8 yıllık bitkisel yağ tüketim verileri incelendiğinde bu yüzdelerin %37-ayçiçek yağı, %17-palm yağı (genelde margarin sanayinde), %16-pamuk yağı, %13-soya yağı, %7-mısır yağı, %5,5-zeytinyağı olduğu görülmektedir. Türkiye’de kişi başına yıllık toplam bitkisel yağ tüketimi 17,5 kg iken bu oran dünyada 14,8 kg’dır ve dünya ortalamasının üzerindedir. AB ülkelerinde kişi başı bitkisel yağ tüketimi 19,2 kg ve ABD’de ise bu oran 27,8 kg olarak gerçekleşmiştir. Ülkemiz, AB ve ABD gibi gelişmiş ülkelerin tüketim değerlerini altındadır (TEAE, 2005: 13).

Tablo 3.3 Türkiye bitkisel ham yağ üretim verilerini göstermektedir (BYSD, İstatistikler, 2016a) Tablo 3.3 incelendiğinde 2004 yılından 2008 yılına kadar bitkisel hamyağ üretimi dalgalı bir seyir izlerken, 2009 yılından itibaren üretim giderek artmıştır.

Tablo 3.3 Türkiye Bitkisel Ham Yağ Üretimi

2004 (Bin Ton)	2005 (Bin Ton)	2006 (Bin Ton)	2007 (Bin Ton)	2008 (Bin Ton)	2009 (Bin Ton)	2010 (Bin Ton)	2011 (Bin Ton)	2012 (Bin Ton)	2013 (Bin Ton)	2014 (Bin Ton)
438	470	568	513	570	506	619	655	680	815	755

* 2013 yılı itibariyle Aspir, Keten ve Ketencik yağı verileri dahil edilmiştir.

Kaynak: BYSD, İstatistikler, 2016a

Tablo 3.4 Türkiye bitkisel likit yağ ihracat verilerini göstermektedir. Tablo 3.4 incelendiğinde 2004 yılından 2009 yılına kadar bitkisel likit yağ üretimi dalgalı bir seyir izlerken, 2010 yılından itibaren üretim giderek artmıştır.

Tablo 3.4 Türkiye Bitkisel Likit Yağ İhracatı

2004 (Bin Ton)	2005 (Bin Ton)	2006 (Bin Ton)	2007 (Bin Ton)	2008 (Bin Ton)	2009 (Bin Ton)	2010 (Bin Ton)	2011 (Bin Ton)	2012 (Bin Ton)	2013 (Bin Ton)	2014 (Bin Ton)
67	98	272	118	265	267	182	448	600	618	724

Kaynak: BYSD, İstatistikler, 2016a

Tablo 3.5, 2010-2014 yılları arasındaki Dünya bitkisel ham yağ üretim verilerini göstermektedir.

Tablo 3.5 Dünya Bitkisel Ham Yağ Üretimi

Yağ Çeşitleri	2010 (Milyon Ton)	2011 (Milyon Ton)	2012 (Milyon Ton)	2013 (Milyon Ton)	2014 (Milyon Ton)
Soya Yağı	41	43	43	47	47
Pamuk Yağı	5	5	5	5	5
Ayçiçek Yağı	12	15	14	16	15
Kolza Yağı	24	24	25	26	27
Palm Yağı	49	52	56	59	63
Diğer Yağlar	18	18	18	18	19
Toplam	149	157	161	171	176

Kaynak: BYSD, İstatistikler, 2016a

Tablo 3.5 dikkatli incelendiğinde en çok üretilen yağ çeşitinin palm yağı ve soya yağı olduğu görülmektedir. Yıllara göre pamuk yağı üretiminin sabit olduğu, ayçiçeği yağının dalgalı bir üretim miktarının olduğu görülmüştür. Pamuk, kolza ve soya yağı üretimi yıllar içerisinde giderek artma eğiliminde olduğu görülmüştür.

Tablo 3.6 incelendiğinde Türkiye’de en çok tüketilen bitkisel yağın sırasıyla ayçiçeği yağı, kanola yağı ve mısır yağı olduğu görülmektedir.

Tablo 3.6 Türkiye Bitkisel Sıvı Yağ Tüketimi

	Ayçiçeği Yağı (Bin Ton)	Soya Yağı (Bin Ton)	Palm Yağı (Bin Ton)	Mısır Yağı (Bin Ton)	Kanola Yağı (Bin Ton)	Fındık Yağı (Bin Ton)	Pamuk Yağı (Bin Ton)
2010/11	765	30	5	50	85	35	20
2011/12	780	5	5	45	70	25	50
2012/13	850	10	5	45	60	20	20
2013/14	900	10	5	40	50	10	20

Kaynak: Tekçe, 2014: 4

3.1.4 Tereyağı

Süt yağının dayanıklı bir forma dönüştürülerek saklanması ve tüketime sunulma çabaları oldukça eski zamanlara dayanmaktadır. Tereyağına ilişkin Batılı kaynaklarda M.Ö. 3000 olarak belirtilen ilk bilgilere, M.Ö. 8000’de Anadolunun doğusunda yerleşen Urartu kaynaklarında da rastlanmıştır. 19 yy.’ın ortalarına doğru ise süte derin kaplarda kaymak bağlatma usulünün uygulanmaya başlanmasıyla tereyağı işletmeleri kurulmaya başlanmıştır (MEB, 2013: 1). Türk Gıda Kodeksine göre tereyağı, ağırlıkça en az %80, en fazla %90 oranında süt yağı, en fazla %2 oranında yağsız süt kuru maddesi ve en fazla %16 oranında su içeriğine sahip ürün olarak tanımlanmıştır. Sadeyağ ise, süt ve/veya süt ürünlerinden elde edilen, su ve yağsız kuru madde unsurlarının tamamına yakın bölümü uzaklaştırılmış, ağırlıkça en az %99 oranında süt yağı içeriğine sahip ürün olarak tanımlanmaktadır (Türk Gıda Kodeksi, 2015).

Tereyağı, süt yağının yoğun hale (konsantre) getirilmesidir. Tereyağının bileşiminde süt kuru maddesini oluşturan öğelerin tümü bulunmaktadır, fakat bu oranlar arasında farklılıklar mevcuttur. Süt yağı içermiş olduğu esansiyel yağ asitlerinden dolayı beslenme açısından önemli özelliklere sahiptir. Özellikle içerisinde yaşamsal öneme sahip yağ asitlerini bulundurması, sindirilme yeteneğinin yüksek olması, yağda çözünen vitaminleri içermesi ve

vücut sıcaklığında çözünmüş halde bulunması nedeniyle de önemli bir süt bileşenidir. Süt yağının bu özellikleri de tereyağının besin değerini artırmaktadır (MEB, 2013: 3-4).

Dünyada kişi başına düşen tereyağı tüketiminin giderek arttığı görülmektedir, 1990-94 yılları arasında 1,34 kg, 1995-99 yılları arasında 1,16 kg, 2000-04 yılları arasında 1,26 kg, 2005-08 yılları arasında 1,3 kg, 2009 yılında ise 1,3 kg olduğu görülmüştür. Türkiye’de tereyağı tüketiminin kişi başına tüketim miktarı, 1990-94 yıllarında 2,32 kg, 1995-99 yıllarında 2,08 kg, 2000-04 yıllarında 1,82 kg, 2005-08 yıllarında 2,05 kg, 2009 yılında ise 2,2 kg olduğu görülmüştür. (Örmeci Kart ve Demircan, 2014: 90) 2013 yılında kişi başı yıllık tereyağı tüketiminin 1,42 kg olduğu görülmüştür (USK, 2014: 52).

2008 yılına ait bazı ülkelerdeki tereyağı tüketimleri İsviçre (5,7 kg), AB 25/27 (3,9 kg), Avustralya (4,1 kg), Rusya (2,8 kg), ABD (2,5 kg) ve Arjantin (0,74 kg) dır (Ambalajlı Süt ve Süt Ürünleri Sanayicileri Derneği, 2010: 32).

Tablo 3.7 incelendiğinde, 2013 yılında kişi başına tereyağı tüketiminin en fazla olduğu ülkeler arasında ilk sıralarda Yeni Zelanda, AB, Avusturya, Kanada, ABD ve Rusya yer aldığı görülmektedir.

Tablo 3.7 Bazı Ülkelerde Kişi Başına Tereyağı Tüketimi

	2009 (kg/yıl)	2010 (kg/yıl)	2011 (kg/yıl)	2012 (kg/yıl)	2013 (kg/yıl)
Avusturya	3,16	3,26	3,43	3,56	3,64
Kanada	2,64	2,64	2,78	2,78	2,7
AB	3,94	3,82	3,91	3,99	3,99
Japonya	0,59	0,67	0,65	0,61	0,57
Meksika	1,98	1,96	1,86	1,87	1,95
Yeni Zelanda	5,11	4,58	4,53	4,71	4,88
ABD	2,21	2,19	2,14	2,5	2,44
Arjantin	1,03	0,89	0,83	0,9	0,94
Brezilya	0,41	0,38	0,4	0,4	0,41
Rusya	2,94	2,23	2,3	2,34	2,44

Kaynak: TEPGE, 2014: 5

2009-2014 yılları arasında Türkiye’de tereyağı üretim, ithalat, toplam yurtiçi kullanım ve ihracat verileri içeren Tablo 3.8, incelendiğinde üretim ve toplam yurtiçi kullanım yıllar arasında giderek artmıştır.

Tablo 3.8 Türkiye’de Tereyağı Üretim, Tüketim Verileri

	2009 (Ton)	2010 (Ton)	2011 (Ton)	2012 (Ton)	2013 (Ton)	2014/a (Ton)
Üretim	31.137	32.978	34.928	38.128	41.515	45.203
İthalat	13.477	11.751	10.288	15.157	19.618	25.392
Toplam Yurtiçi Kullanım	42.811	44.234	44.087	52.788	60.184	69.562
İhracat	180	357	725	443	637	916

Kaynak: TEPGE, 2014: 20

Yalnızca 2010-2011 toplam yurtiçi tereyağı kullanım verisi yatay bir seyir izlemiştir. Tereyağı ithalatı 2009-2011 yılları arasında azalma göstermesine karşın, 2012-2014 yılları arasında artış göstermektedir. İhracat verilerini incelediğimizde ise 2009-2011 yılları arasında bir artış görülürken, 2012 yılında bir azalma, daha sonraki yıllarda ise artış gösterdiği görülmüştür.

Tablo 3.9 Avustralya, AB-27, Japonya Yeni Zelanda, Rusya ve Dünya genelindeki üretim, ithalat, ihracat ve tüketim verilerini içermektedir.

Tablo 3.9 Dünya Tereyağı Arz Kullanım ve Ticareti: 2009-2013

	Başlangıç Stokları (Bin Ton)	Üretim (Bin Ton)	İthalat (Bin Ton)	İhracat (Bin Ton)	Yurtiçi Tüketim (Bin Ton)	Sene Sonu Stokları (Bin Ton)
2009						
Avustralya	13	118	18	87	60	2
AB-27	53	2,03	63	149	1,894	103
Japonya	23	81	0	0	74	30
Y. Zelanda	65	482	1	492	20	36
Rusya	12	246	107	4	349	12
ABD	54	713	17	30	694	60
Dünya	235	8.039	345	813	7.549	257
2010						
Avustralya	2	132	19	58	73	22
AB-27	103	1.98	41	154	1.934	36
Japonya	30	74	3	0	86	21
Y. Zelanda	36	441	1	429	20	29
Rusya	12	207	113	2	319	11
ABD	60	709	10	59	683	37
Dünya	257	8.179	302	735	7.827	176
2011						
Avustralya	22	121	19	42	78	42
AB-27	36	2.055	44	124	1.982	29
Japonya	21	63	15	0	83	16
Y. Zelanda	29	487	1	449	20	48
Rusya	11	217	116	2	330	12
ABD	37	821	12	65	757	48
Dünya	175	8.584	274	723	8,097	213

	Başlangıç Stokları (Bin Ton)	Üretim (Bin Ton)	İthalat (Bin Ton)	İhracat (Bin Ton)	Yurtiçi Tüketim (Bin Ton)	Sene Sonu Stokları (Bin Ton)
2012						
Avustralya	42	119	21	54	82	46
AB-27	29	2.1	55	120	2.031	33
Japonya	16	69	10	0	77	18
Y. Zelanda	48	527	0	509	21	45
Rusya	12	216	119	2	335	10
ABD	48	843	18	47	793	69
Dünya	213	8.914	302	764	8.419	246
2013*						
Avustralya	46	117	23	55	85	46
AB-27	33	2.09	55	120	2.033	25
Japonya	18	71	1	0	72	18
Y. Zelanda	45	506	1	475	22	55
Rusya	10	205	145	2	348	10
ABD	69	852	13	91	780	63
Dünya	246	9.096	332	782	8.651	241

Kaynak: TEPGE, 2014: 28

Karakaya ve Akbay'ın yapmış oldukları “İstanbul İlinde Tüketicilerin Süt ve Süt Ürünleri Tüketim Alışkanlıkları” isimli çalışmalarında İstanbul ilinde tüketicilerin süt ve süt ürünleri tüketimleri ve tercihlerini analiz etmişlerdir. Bu amaçla analizde değişik sosyoekonomik ve demografik gruplarda yer alan 400 aileden elde edilen veriler kullanılmıştır. Tüketicilerin tüketim miktarı ve tercihlerini etkileyen sosyoekonomik ve demografik faktörlerin etkilerini belirlemek amacıyla tanımlayıcı istatistiklerden, varyans analizlerinden ve Ki kare testinden yararlanılmıştır. Analizlerin sonucuna göre ailelerin aylık tereyağı tüketim değeri 0,59 kg/ay olarak görülmüştür. Açık süt, pastörize süt, paket süt, ayran, yoğurt, dondurma, kaymak, lor ve peynir tüketim değerleri de ayrıca detaylı olarak belirtilmiştir. Gıda harcamalarının, toplam harcamalar içerisindeki oranı yaklaşık olarak %22,5 olarak tespit edilmiştir. F- testi sonucuna göre, ailelerin ortalama gıda harcaması miktarlarında önemli bir farklılık saptanmıştır. Bireylerin gelirleri arttıkça ortalama gıda harcamaları da artmaktadır. Araştırma alanında gelir arttıkça süt ve süt harcamaları mutlak olarak artarken, oransal olarak azaldığı saptanmıştır. Süt ve süt ürünleri harcamalarının toplam gıda harcaması içerisindeki oranı %18,6 olarak saptanmıştır (Karakaya ve Akbay, 2013).

Örmeci Kart ve Demircan'nın yapmış oldukları “Dünyada ve Türkiye’de Süt ve Süt Ürünleri Üretimi, Tüketimi ve Ticaretindeki Gelişmeler” isimli çalışmalarında dünyada ve Türkiye’de süt ve süt ürünlerine ait üretim, tüketim ve dış ticaret istatistikleri 1990 yılından itibaren incelenmiştir. Süt ürünleri kapsamında en yaygın olarak tüketilen peynir, tereyağı ve yoğurda yer verilmiştir. Çalışmanın temel materyalini oluşturan söz konusu ürünlere ait

veriler FAO'dan alınmıştır. Elde edilen veriler basit ortalama, yüzde ve indeks hesapları yapıldıktan sonra yorumlanmıştır. Çalışmanın bulgularına göre, dünyada en fazla tereyağı tüketen ülkenin kişi başına 9,20 kg ile Yeni Zelanda olduğu görülmektedir. Yeni Zelanda'yı 7,6 kg ile Fransa, 6 kg ile İsviçre, 5,80 kg ile Almanya takip etmektedir. Türkiye'nin ise tereyağı tüketimi dünya ortalaması olan 1,30 kg'ın üstünde olmasına rağmen 2,20 kg ile birçok Avrupa ülkesinin gerisinde olduğu görülmüştür. Tereyağı ve diğer sütü ürünlerine ait ithalat, ihracat, üretim ve tüketim değerleri çalışmada detaylı olarak verilmiştir (Örmeci Kart ve Demircan, 2014).

3.2 Tüketim

Toplumda yaşayan bireylerin yaşamlarını devam ettirebilmeleri için yeme, içme, barınma, yeni ve bilinmeyenleri keşfetme, ihtiyaç üretme, alma, verme vb ihtiyaçlarını karşılamaları gerekmektedir. Bu eylemler insanın tüketim macerasının başlayıp gelişmesinde belirleyici olmuştur. Bireylerin bu gereksinimleri karşılayabilmesi için üretim ve tüketim yapması gerekmektedir. Bu anlamda insanoğlunun yaşam serüveni, tüketim macerasıyla birliktelik arz etmektedir. Yani insanın doğduğu anda başlayan ve ölene kadar geçirdiği sürede gerçekleşen bir eylemdir. (Karakaş, 2001: 12)

Tüketim kavramının ne anlama geldiğiyle ilgili ekonomi, iktisat, sosyoloji gibi alanlarda yapılmış birçok tanım bulunmaktadır. Tüketim kavramı ile ilgili yapılmış tanımlara bakacak olursak;

- Tüketim yiyecek, giysi, çocuk bakımı vb mal ve hizmetlerin hanehalkı tarafından kullanılmasıdır (Magrabi vd., 1991: 9) .
- Faydalı mal ve hizmetlerin insanların ihtiyaçlarının karşılanması amacıyla kullanılmasıdır (Üstünel, 1997: 30).
- Belirli bir ihtiyacı tatmin etmek için bir ürünü yâda hizmeti edinme, sahiplenme yâda yok etmedir (Odabaşı, 1999: 4)
- İnsanların mal ve hizmetleri ihtiyaç veya isteklerini karşılamak amacıyla satın alıp kullanmalarına tüketim denir (Ertek, 2009: 25).
- İnsan ihtiyaçlarını ve arzularını karşılamak üzere mal ve hizmetlerin kullanılmasıdır (TÜİK, 2011b: XXI).

Halk dilinde tüketim bir şeyin tüketilip yok edilmesi, helak olması anlamına gelir. Bir çiftçinin hayvanlarına yedirdiği yem ya da bir fabrikatörün enerji üretmek için tesislerinde kullandığı kömür tüketim olarak kabul edilir. Fakat ekonomi biliminde bu kavramın farklı bir tanımı vardır. Bu tanıma göre tüketim, yalnız ve yalnız hür insanların ihtiyaçlarının

karşılanması amacıyla mal ve hizmetlerin kullanılmasıdır. Söz konusu mal ve hizmetlerin sağlayacakları faydadan yararlanmanın dolaysız olması gerekir. O zaman hür olmayan insanların ve hayvanların yeme içme ihtiyaçlarının karşılanması için bazı malların kullanılmasını tüketim olarak sayılamaz. Bu sadece sahiplerine yarar sağlamak amacıyla üretimin bir aşaması olarak düşünülmesi gereken bir faaliyettir. Ayrıca tüketim amacıyla kullanılan malların yok olması, ortadan kalması da şart değildir. Örneğin, bir heykelin bazı insanlar tarafından seyredilmesi insana keyif verir, bu eylem insanların bazı ihtiyaçlarını karşılamaktadır. Burada dolaysız bir faydalanmada vardır ve ekonomik anlamda bir tüketim söz konusudur. Oysa heykelin kendisi ne yok olmuş ne de değişime uğramıştır. Ekonomi ilminde benimsenen anlamı ile tüketim, kıt malların faydalarının hür insanların ihtiyaçlarını dorudan doğruya karşılamak amacıyla kullanılması olarak tanımlanabilir (Üstünel, 1997: 30-31).

Tüketim harcamaları kavramı, bir mal veya hizmet satın almak için yapılan parasal ödeme veya fedakarlıkların tümüdür (TÜİK, 2011a). Hanehalkının satın aldığı mal için para ödemesi bir harcama, malın bir miktarının kullanılması ise bir tüketimdir.

Tüketimle ilgili uygulamalı çalışmaların başlangıcı 18. yüzyılın sonlarına kadar uzanmaktadır. 1795’de Davis’in, 1797’de Eden’in, 1850’de Frederick Le Play’in araştırmalarını 1855’de Edouard Ducpetiaux izlemiştir. Le Play ve Ducpetiaux’ının çalışmalardan yararlanan Alman istatistikçi Ernst Engel tüketim harcamalarına ilişkin araştırmaların modern iktisadi düşüncedeki önemini ortaya çıkarmıştır. Son zamanlarda ise tüketim çalışmaları Barteni Clements, Deaton, Goldberger, Houthakker, Leser, Muellbauer, Philips, Powel, Stone, Theil ve Working gibi bilim adamları tarafından incelenmiştir (Özer, 1999: 1).

Bireylerin ihtiyaçlarını dolaylı ve doğrudan karşılayan her şeye mal denir. Mallar, ekonomik mallar, serbest mallar, dayanıklı mallar, dayanıksız mallar, tüketim malları gibi çeşitli şekillerde sınıflandırılabilirler. Tüketim malları, ihtiyaçlarımızı doğrudan doğruya karşılamaya yarayan mallardır. Salça, zeytinyağı, televizyon, araba, bilgisayar tüketim mallarına örnek gösterilebilir (Üstünel, 1997: 29-30).

3.2.1 Hanehalkı Tüketim Harcaması

Hanehalklarının anket ayı içinde gıda ve alkolsüz içecekler, alkollü içecekler, sigara ve tütün, giyim ve ayakkabı, konut, su, elektrik, gaz ve diğer yakıtlar, mobilya, ev aletleri ve ev bakım hizmetleri, sağlık, ulaştırma, haberleşme, eğlence ve kültür, eğitim hizmetleri, lokanta ve oteller ile çeşitli mal ve hizmet ihtiyaçlarını karşılamak için yapmış oldukları peşin ya da

vadeli mal ve hizmet satın alışı ile hanehalklarının tarımsal faaliyetleri sonucu elde ettikleri mal üretimlerinden tükettiği maddeleri, hanede iktisaden faal olan fertlerin çalıştığı işyerinde ürettiği ya da satışa sunduğu mallardan hanede tüketilenlerin değeri, özel kişi veya kuruluşlara vermek üzere hediye veya yardım şeklinde satın aldığı mal veya hizmetler ile gayrimenkul veya mülk aynı geliri şeklinde gelen malların değeri kapsamıştır (TÜİK, 2011b: XXI).

Hanehalklarının gelirlerinin tüketimleri ve birikimleri arasında paylaşımı nasıl olduğu son derece önemlidir. Bu karar hanehalkının zamanla ekonomik refahını da etkileyecektir. Hanehalkının bugün büyük miktarda harcama yapması ileriki yaşamında daha az harcama yapmasına neden olabileceği açıktır. Hanehalklarının tüketim ve birikim kararlarının toplam etkisi, ekonominin büyüme hızının belirlenmesinde, üretim ve istihdam düzenin belirlenmesinde, ticaret dengesinin belirlenmesinde önemlidir (Sachs ve Larrain, 1993: 78). Ayrıca, hane bazında refah ve yaşam standartlarının ortaya konmasında dolayısıyla toplumsal statünün belirlenmesinde önemlidir. Toplam tasarruf tüketim kararları ücret, vergi, faiz gibi büyüklükleri de doğrudan ya da dolaylı olarak etkilediği için diğer makro değişken fonksiyonlarının da içinde bulunan önemli bir değişken olarak görülür. Hanehalkı tüketim harcamalarının incelenmesinin önemli bir nedeni de tüketim kalıplarından hanehalkı refah düzeyine ulaşılabilmesidir. Tüketim harcamaları ile bir bütün olarak hanehalkı tüketim kalıbının çıkarılması, bir ilin, bölgenin ya da ülkenin refah düzeyinin yaşam standardına ait bilgilerin ortaya konulması anlamına gelmektedir (Aktaş, 2008: 9).

Hanehalkı bütçe anketleri, hanelerin sosyoekonomik yapıları, yaşam düzeyleri, tüketim kalıpları hakkında bilgi veren ve toplumun ihtiyaçlarının belirlenmesi ve uygulanan sosyoekonomik politikaların geçerliliğinin test edilmesi amacıyla kullanılan en önemli veri kaynaklarından biridir. Türkiye İstatistik Kurumu 2002 yılından itibaren her yıl düzenli olarak hanehalkı bütçe anketi uygulamaktadır (TÜİK, 2011b: III).

Hanehalkı gelirin nasıl kullanıldığı, gelirin tüketim, mal ve hizmetler arasındaki paylaşımı, tasarrufu ekonomistlerin, iktisatçıların ve politikacıların ilgilendikleri, üzerinde araştırma yapıları önemli konulardandır. Hanehalkı tüketimine yönelik yapılan çalışmalar tüketici tercihlerinin belirlenmesinde ve gelecekteki talebin ne olacağının öngörülmesi bakımından son derece önemlidir. Ayrıca, hanehalkı tüketiminin sosyodemografik özelliklere göre gösterdiği değişimi ortaya koymak, gelir-talep esnekliklerini belirlemek, sosyal politikaların etkinliği için önemli araçlardır. (Fisunoğlu ve Şengül, 2011: 251)

Ekonomik ve sosyal değişimlerin yakından izlenmesi, ekonomik kalkınmanın belirli bir denge içinde hızlandırılması, halkın hayat seviyesinin öngörülen bir düzeye erişmesinin sağlanması amacıyla politikalar üretilmesi, uygulanan politikaların ve sonuçlarının

değerlendirilmesi bu tür anketlerden elde edilecek verilerle mümkün olmaktadır (TÜİK, 2011b: III).

Hanehalkı bütçe anketi ile aşağıda belirtilen hedefler gerçekleştirilmektedir (TÜİK, 2011a).

- Tüketici fiyat indekslerinde kullanılacak maddelerin seçimi ve temel yıl ağırlıklarının elde edilmesi ve güncellenmesi,
- Hanelerin tüketim yapısı ve tüketim kalıplarında zaman içinde meydana gelen değişikliklerin izlenmesi,
- Milli gelir hesaplamalarında özel nihai tüketim harcamaları tahminlerine yardımcı olacak verilerin derlenmesi,
- Asgari ücret tespit çalışmaları için ihtiyaç duyulan verilerin elde edilmesi,
- Harcamaya dayalı yoksulluk sınırının belirlenmesi, hanehalklarının yaşam seviyeleri, beslenme sorunları vb. diğer sosyo-ekonomik analizler için gerekli verilerin elde edilmesi amaçlanmaktadır.

Türkiye'de ilk çalışma 1933 yılında Ticaret Bakanlığı Konjonktür Yayın Müdürlüğü tarafından yapılmış olan ve Ankara'da memur, İstanbul'da işçi ailelerini kapsayan “Aile Bütçeleri Anketleri” ile başlamıştır. Türkiye İstatistik Kurumu, 1964 yılında ilk aşamada Adana, Ankara, İzmir ve İstanbul olmak üzere 4 büyük ilimizde aile bütçeleri anketi uygulanmasına başlamış, ikinci aşamada bulundukları bölgeyi temsil etme durumları esas tutularak Samsun, Antalya, Diyarbakır, Bursa, Ordu, Erzurum, Eskişehir il merkezleri kapsama alınmıştır. Bu anketin sonuçlarına dayalı olarak 11 il merkezi için 1968 bazlı tüketici fiyatları indeksi hazırlanmıştır. 1973-1974 yılları arasında ilk kez bir coğrafi kesimin tümünde, Türkiye İstatistik Kurumu tarafından Milli Eğitim Bakanlığı ve Devlet Planlama Teşkilatı işbirliğiyle anket uygulanmıştır. 1978-1979 yıllarında nüfusu 10 001'den fazla olan 40 yerleşim yerinde her ay değişen 822 hanehalkına uygulanan anketle bu alandaki çalışmalara devam edilmiştir. 1987 yılında Hanehalkı Gelir ve Tüketim Harcamaları Anketi, Türkiye genelini kapsayacak şekilde ilk kez uygulanmış olup ankette; bölge, nüfus tabakaları, kır ve kent ayrımında gelir ve tüketim farklılaşmasını gösteren bilgiler yer almıştır. Bu anketin sonuçları 1987 bazlı Tüketici Fiyatları Endeksinin baz yılı fiyatları ile ağırlıkların belirlenmesinde ve 1987 yılı gelir dağılımı analizlerinde kullanılmıştır. 1994 Hanehalkı Gelir ve Tüketim Harcamaları Anketi ise, daha önce uygulanan anketlerden farklı bir yöntemle tüketim harcamaları ve gelir dağılımı amaçlarına yönelik olarak ayrı ayrı düzenlenmiştir. Bu anketin sonuçları 1994 bazlı Tüketici Fiyatları Endeksinin baz yılı fiyatları ve ağırlıkların tespitinde kullanılmıştır. Kurumumuz, 2002 yılından itibaren daha küçük örnek hacmi ile her

yıl düzenli olarak Hanehalkı Gelir ve Tüketim Harcamaları Anketi uygulamayı planlamıştır. Avrupa Birliği uyum çalışmaları çerçevesinde ve 2003 yılından itibaren uygulaması başlatılacak harmonize tüketici fiyat indeksine baz oluşturması için, 2003 yılında yeni ismiyle “Hanehalkı Bütçe Anketi”nin tahmin boyutu yeniden ele alınmış ve Türkiye, Kent, Kır, İstatistiki Bölge Birimleri Sınıflaması Düzey-I ve her Düzey-I bölgesi için kent/kır ayrımı ile Düzey-II bazında tahmin verme amacına yönelik olarak örnek hacmi sadece bu yıl için genişletilerek uygulanmıştır. 2004 yılından itibaren, çalışmaların yıllık örnek hacmi yeniden küçültülerek yıllık anket sistemine devam edilmiştir (TÜİK, 2012).

Yapmış olduğumuz tez çalışmasında 2009-2012 yılları arasında Türkiye İstatistik Kurumu tarafından derlenen 4 yıllık Hanehalkı Bütçe Anketi verileri kullanılmıştır.

- 2009 Hanehalkı Bütçe Anketi 1 Ocak–31 Aralık 2009 tarihleri arasında yine bir yıl süre ile her ay değişen aylık ortalama 1 050, yıllık toplam 12 600 örnek hanehalkıyla yürütülmüştür.
- 2010 Hanehalkı Bütçe Anketi 1 Ocak – 31 Aralık 2010 tarihleri arasında bir yıl süre ile her ay değişen 1.104, yıllık toplam 13.248 örnek hanehalkına anket uygulanmış, anketi geçerli olan hanehalkı sayısı 10.082 olmuştur.
- 2011 Hanehalkı Bütçe Anketi 1 Ocak–31 Aralık 2011 tarihleri arasında yine bir yıl süre ile her ay değişen 1.104, yıllık toplam 13.248 örnek hanehalkına anket uygulanmış, anketi geçerli olan hanehalkı sayısı 9.918 olmuştur.
- 2012 Hanehalkı Bütçe Anketi ise 1 Ocak–31 Aralık 2012 tarihleri arasında yine bir yıl süre ile her ay değişen 1.104, yıllık toplam 13.248 örnek hanehalkına anket uygulanmış, anketi geçerli olan hanehalkı sayısı 9.987 olmuştur.

3.2.1.1 Hanehalkı Tüketimine Etki Eden Faktörler

Hanehalkı tüketim harcamalarına etki eden birçok faktör bulunmaktadır. Bu faktörlerin etkisi mal ve ürün gruplarına göre değişiklik gösterebilmektedir.

Tüketiciler satın alma kararlarını verirken gelir, servet düzeyi, faiz oranı, fiyatlar, beklentiler, gelir dağılımı, enflasyon, bireylerin eğitim durumları, meslekleri ve yaşları, tasarruflar, geçmiş tüketim, teknoloji ve sosyokültürel faktörler gibi çok sayıda faktörün etkisi altında kalmaktadırlar (Tarı ve Pehlivanoğlu, 2007: 194).

Odabaşına Göre, genel tüketici davranışını etkileyen faktörler; psikolojik faktörler, sosyo-kültürel faktörler ve demografik faktörler olmak üzere üçe ayrılır. Psikolojik faktörler; öğrenme, güdülenme, algılama, tutum ve kişilik olmak üzere beşe ayrılır. Sosyo-kültürel faktörler; danışma grupları, sosyal sınıf, aile, kişisel etkiler ve kültür olmak üzere beşe ayrılır.

Demografik faktörler; yaş, cinsiyet, eğitim, coğrafik yerleşim, meslek ve gelir olmak üzere altıya ayrılmaktadır (Odabaşı, 1996: 21)

Karafakıoğlu'na göre, tüketici davranışını etkileyen faktörler; kültürel faktörler, sosyal faktörler, kişisel faktörler ve psikolojik faktörler olmak üzere dörde ayrılmaktadır. Kültürel faktörler; kültür, alt kültürler ve sosyal sınıftır. Sosyal faktörler; statü ve roller, aile ve referans gruplarıdır. Kişisel faktörler; yaş, meslek, ekonomik durum, yaşam biçimi ve kişiliktir. Psikolojik faktörler ise; ihtiyaçlar, algılama, güdüler, inançlar, tutumlar ve öğrenme sürecidir (Karafakıoğlu, 2006: 94)

Aktaş'a göre, hanehalkı tüketim harcamalarını etkileyen faktörleri temelde ekonomik, demografik, psikolojik ve sosyal faktörler olarak dört gruba ayırmıştır. Ekonomik faktörler; gelir, gelir dağılımı, servet, geçmiş tüketim, fiyatlar ve para aldanımı, beklentiler, faiz haddi, tüketici kredileri, vergilerdeki değişiklikler, enflasyon olarak sayılabilir. Demografik faktörler; eğitim, meslek, yaş, hanehalkının bileşimi, hanehalkı büyüklüğü ve hanehalkının yaşadığı yerleşim biriminin büyüklüğü olarak sayılabilir. Psikolojik faktörler; motivasyon, algıda seçicilik, öğrenme, tutum ve inançlar, gösteriş etkisi, tasarruf motifleri, kişisel ve ulusal özellikler olarak sayılabilir. Sosyal faktörler; kültür, sosyal sınıf, referans grupları, aile ve sosyal statü-roller olarak sayılabilir (Aktaş, 2008: 19-40).

3.2.1.2 Hanehalkı Gıda Tüketimine Etki Eden Faktörler

Hanehalkı gıda tüketimine etki eden birçok faktör bulunmaktadır. İncelemiş olduğum makaleler, kitaplar, notlar ve elektronik kaynaklardan derlediğim bu faktörleri iki başlık altında topladım. Bunlar ekonomik faktörler ve demografik (sosyodemografik) faktörlerdir.

- 1) Ekonomik Faktörler; Gelir
- 2) Demografik Faktörler; Eğitim, Yaş, Meslek, Sosyal Statü, Cinsiyet, İstihdam Durumu, Medeni Durum, Irk, Hanehalkı Büyüklüğü, Kentleşme (Kır-Kent Durumu), Bölge, Gıda Damga Programı, Çocuk Beslenme Programı

3.2.1.2.1 Gelir

Hanehalkı harcamalarına etki eden en önemli faktörlerin başında “Gelir” gelmektedir. Hanehalkının yapmış olduğu bütün tüketim kalemleri TÜİK tarafından 12 Ana tüketim grubu altında toplanmaktadır. Gıda ve Alkolsüz İçecekler grubu da bu 12 tüketim grubundan biridir. Hanehalkının gıda tüketim harcamalarının en önemli belirleyicisi yine “Gelir”dir. Gelirin, birçok ekonomist ve iktisatçı tarafından yapılmış tanımı vardır ve ana tüketim gruplarında ya da alt tüketim kalemlerinde gelirin etkilerini detaylı olarak anlatan çalışmalar mevcuttur.

Gelir hanehalkı harcamalarının miktarını ve seviyesini belirleyen en önemli ekonomik kavramdır. Tıpkı zaman gibi bir akış kavramıdır ve birim zaman başına miktarı ölçülmelidir. Ekonomistlere göre iki farklı gelir vardır bunlar; para geliri ve gerçek gelirdir. Para geliri, zaman birimi başına hane tarafından alınan para ya da satın alma gücüdür. Bir hanenin para geliri hanehalkı üyeleri tarafından tüm kaynaklardan gelen paraları kapsar. Tasarruf ve yatırım getirileri, sosyal güvenlik yardımları, kar/kazanç, emeklilik/emeklilik maaşı (earnings, pensions), refah yardımları ve diğer parasal yardımlar, hediyeler, sigortaları hasar tazminatları ve mirasta para gelirinin kapsamı içindedir. Gerçek gelir, belirli bir zaman süresi içinde hane tarafından elde edilen mal ve hizmetlerin akışıdır. Parasal olarak ifade edilen bu terim, tüketimin para değeridir. Tüketimin para değerini doğru ölçmek zordur ve gerekli veriler nadiren mevcuttur (Magrabi vd., 1991: 76-77).

Gelir, toplum açısından milli ekonomide belir bir dönemde (genellikle yılda bir) yeniden yaratılan değerlerin toplamıdır, bu milli gelir olarak tanımlanır. Firma açısından yapılan gelir tanımı ise, belirli bir süre içerisinde firmanın halka veya diğer firmalara sattığı mal ve hizmetlerden elde ettiği bedellerin toplamından, firmanın üretim giderleri çıkarılarak hesaplanır. Ailenin geliri, emek karşılığı alınan ücret, maaşlar, toprak kirası, sermaye geliri (veya faiz) veya teşebbüs geliri (veya kâr) şeklinde olabileceği gibi her birinden bir miktar bulunabilir. Bir başka tanımıyla gelir, bir kimsenin belirli bir dönemin başında ve sonunda eş zenginlikte kalabilmesi için o dönemde, isterse, tüketebileceği mal ve hizmetlerin toplamına denir (Üstünel, 1997: 42).

Alman iktisatçı Engel, 1857 yılında Saksonya Krallığı üretim ve tüketim koşulları üzerine bir çalışma yayınlamıştır. Bu çalışmasında gelir ve gıda harcamaları arasındaki ilişkiyi belirten ampirik bir yasa formülize etmiştir. 1857'den günümüze kadar Engel Yasası olarak bilinen yasa; “Gelir arttıkça, gıdaya harcanan para artarken ve gıdaya harcanan paranın payı azalmaktadır.” şeklindedir (Houthakker, 1957: 532). Engel kanunlarının ifade ettiği ilişkileri grafiksel olarak ortaya koyan eğriler Engel Eğrileri olarak ifade edilir. Gelir-tüketim eğrilerinden hareketle elde edilen engel eğrileri, fiyatlar sabit tutularak hanehalklarının çeşitli gelir düzeylerinde bazı mal ve servislerden satın alacakları miktarları gösteren fonksiyon yapısıdır. Ayrıca Engel Eğrileri farklı gelir gruplarında hanehalkları arasında tüketim kalıplarının nasıl değiştiğini gösterir, başka bir ifade ile özellikleri bakımından tüketim kalıplarını ortaya koymanın sistematik bir yoludur (Beyaz, 2007: 71).

3.2.1.2.2 Hanehalkı Büyüklüğü ve Kompozisyonu, Kır-Kent Durumu

Hanehalkı büyüklüğü, hanehalkı tüketimine etki eden önemli faktörlerden birisidir ve hanede yaşayan bireylerin toplam sayısı olarak ifade edilir. Hanehalkı büyüklüğünün artması ile hanede gıda harcamalarına ayrılan para miktarı artar (Magrabi vd., 1991: 105). Yapılan birçok çalışmada hanehalkı büyüklüğü ile gıda ve diğer tüketim kalemlerinin arasındaki ilişki incelenmiştir. Tezin literatür özeti kısmında bununla ilgili detaylar verilmiştir.

Hanehalkı bütçe anketlerindeki tüketim harcaması verileri hane düzeyinde toplandığından, hanehalkını meydana getiren bireylerin hane kaynaklarından aldığı paylarda bir tahsis kuralı temel alınarak bazı düzeltmeler yapılmalıdır. Hanelerde bulunan fertlerin, yaş ve cinsiyetine bağlı olarak tüketimleri birbirinden farklıdır. Ayrıca, hanedeki bazı harcamalar (gıda, konut, elektrik, su, ısınma vb.) hanehalkı büyüklüğünden bağımsız olarak ortak harcamalardır. Bu nedenle, yoksulluk göstergeleri “eşdeğer fert” üzerinden hesaplanmaktadır. Eşdeğer fert ölçeği, ailedeki yetişkin ve çocuk sayılarına bağlı olarak oluşturulmaktadır. Böylece, farklı büyüklük ve bileşimlerdeki (yetişkin ve çocuk sayısı itibarıyla) hanehalkları arasında eşit düzeyde karşılaştırmalar yapılması mümkün olmaktadır. (TÜİK, 2008: 41) Bu nedenle de hanehalkı düzeyinde toplanan gelirlerin bireysel gelirlere dönüştürülmesi gerekmektedir. Bunu, toplam hane gelirini, haneyi oluşturan fert sayısına bölerek yapmak doğru olmayacaktır. Bu hesaplamada, hanelerin yetişkin-çocuk bileşimlerindeki farklılıkları dikkate almak gerekmektedir. (TÜİK, 2015)

Eşdeğer fert ölçeğini oluşturmada FAO’nın ve OECD’nin kullandığı farklı yöntemler vardır. EUROSTAT’ın yoksulluk ve eşitsizlik göstergelerinin hesaplanmasında kullanılmasını önerdiği eşdeğerlik ölçeği ise OECD’nin modifiye edilmiş ölçeğidir. Buna göre;

- Hanehalkı reisi için = 1
- Hanede yaşayan 14 yaş ve üzeri her birey için= 0,5
- Hanede 14 yaş altındaki her birey için= 0,3

katsayıları dikkate alınır. (TÜİK, 2008: 43-44)

Eşdeğer fert ölçeğinin formüle edilmiş hali;

Eşdeğer Fert Başına Tüketim Harcaması= HH Tüketim Harcaması / [1+(0,5*(Hanede yaşayan 14 yaş ve üzeri birey sayısı-1) + 0,3*(Hanede 14 yaş altındaki birey sayısı))] şeklindedir. (Kızılgöl, 2012: 185)

Hanehalkı kompozisyonu, hanehalkı tüketim harcamalarını etkileyen diğer önemli bir faktördür. Dünya genelinde gittikçe artan oranda çekirdek aile yapısına doğru bir gidiş gözlenmesine rağmen aile içinde karar alma, dayanışma ve iletişim biçimlerindeki farklılığın devam ettiği de gözlenmektedir. Toplumumuzda da çekirdek aile yapısı yaygınlaşmakta

olmasına rağmen aile büyükleri ve diğer aile bireyleriyle süren iletişim, yardımlaşma ve kararlara katılma hanehalkının tüketim eylemlerini ve harcamalarını etkilemektedir. Burada ki diğer bir önemli nokta ise ailede ki çalışan kişi sayısıdır. Ailede çalışan kişi sayısının artması tüketim eylemlerinde ve harcamalarında bireysel kararların ağırlığının artması yanında, örneğin ev eşyası, taşınmaz mal edinimi, seyahat ve dinlenme gibi bazı tüketim eylemlerinde ortak kararlar alınmasını beraberinde getirebilir. (Altunışık vd., 2001: 35-36)

Hanede bulunan bebek ve çocuk olması durumunda hanehalkının harcamalarının sadece yetişkin veya yaşlılardan oluşan hanelerden farklı olacaktır. Evli çiftler ile evli ve çocuklu çiftleri karşılaştırdığımızda, çocuklu hanelerin çocuksuz hanelere göre ev gıda, eğlence, kıyafet ve eğitim tüketimlerine ayırdıkları bütçe daha büyük iken, sağlık hizmeti ve kişisel sigortaya ayırdıkları bütçe daha küçüktür. Çocuğun birçok harcamalarına ayrılan pay çocuğun yaşı ile birlikte artış eğilimindedir. Bu harcamalar, gıda (özellikle ev dışı gıda tüketimi), kıyafet, ulaşım, sağlık hizmeti, kişisel bakım, eğitim, tütün ve nakit para katkısını içermektedir (Magrabi vd., 1991: 181-182).

Evli ve çocuklu çiftler ile tek yetişkin birey karşılaştırıldığında, hem evli ve çocuklu çiftler hem de tek yaşayan bireyin olduğu hanelerde harcamaların büyük kısmının hane giderlerine ve özel ulaşım giderlerine harcandığı görülmüştür. Tek yaşayan bireyler, evli ve çocuklu bireylere göre hane giderleri, ev dışı gıda tüketimi, alkol ve tütün harcamalarında daha fazla bütçe ayırdığı görülmüştür (Magrabi vd., 1991: 187).

Hanehalkı tüketimine etki eden faktörlerden birisi de kentleşmedir. Ülkemizde kent nüfusu giderek artarken, köy nüfusu giderek azalmaktadır. Kentte yaşayan nüfusun artması, köyden şehre göç edenlerin yaşam tarzlarını da etkilemektedir. Bu da tüketim kalıplarını ve talebi etkilemektedir (Altunışık vd., 2001: 38). Kır ve kentte yaşayan bireylerin tüketim kalemlerine ayırdıkları bütçe ve harcama büyüklükleri birbirinden farklıdır. Gelir ve diğer özellikler sabit tutulduğunda kentte yaşayan haneler ile kırdaki yaşayan haneler kıyaslandığında, kırdaki yaşayan haneler özel araçla ulaşım, sağlık giderleri ve tütün ürünlerinde kentte yaşayan hanelere göre daha fazla bütçe ayırdıkları, hane giderlerine (housing), toplu taşıma, eğlence ve eğitime daha az bütçe ayırdıkları görülmüştür. Ev gıda tüketimine kırsal hanelerin daha fazla bütçe ayırdıkları görülmüştür, gelir ve diğer karakteristikler dikkate alındığında bunun tam tersi doğrudur (Magrabi vd., 1991: 192-193).

3.2.1.2.3 Eğitim ve Meslek

Tüketicilerin satın alma ve harcama kararların etkileyen bir diğer sosyo-demografik karakter eğitimidir. Tüketicilerin eğitim durumları ve düzeylerindeki gelişim tüketim

kalıplarını ve hayat tarzını etkileyen önemli bir kriterdir. Eğitim düzeyinin artması, beraberinde gelir artışını da getiriyorsa bireyin toplum içindeki sosyal statüsü de değişmekte ve dolayısıyla tüketim kalıpları ve harcama biçimi de farklılaşabilmektedir. (Altunışık vd., 2001: 40)

Gelir, meslek seçimi, yerleşim yeri, coğrafi hareketlilik, tüketim harcamaları, boş zaman etkinlikleri, hobiler, arkadaş ve birliktelik nitelikleri, yaşam tarzı, kişisel ve sosyal konulara yönelik tutumlar gibi olguların tümü eğitimle güçlü ilişkileri olduğu görünmektedir (Michael, 1975: 235).

Bireyin mesleği genelde gördüğü eğitim ve düzeyi ile ilişkilidir. Meslek, ferdin ve ailesinin sosyal statüsünü ve gelir seviyesini etkilemektedir. Geçerli bir mesleği olan bireyler yüksek gelir elde ederek daha iyi ekonomik koşullara erişebilecektir. Yüksek gelirin etkisi ile kaliteli ürünleri tercih edecektir. Meslek, fertlerin geçimini sağlayan, sosyal statülerini belirleyen ve kendine has kanunî ve ahlâkî kuralları olan görece sürekli bir faaliyet tarzı olarak tanımlanabilir. Çağımızın toplumlarında çoğu meslekler belirli bir eğitim seviyesinden sonra kazanılmaktadır (Eke, 1987: 377-401).

İşin sosyal ve ekonomik değerlendirmesi, çalışanın ve ailesinin toplumsal açıdan varlığını belirlemektedir. Bireyin geliri, fiziki varlığının sosyo-ekonomik temelini oluşturmakta, sosyal statü ve prestij kazandırmaktadır (Erkan, 1991: 109).

Geçici bir mesleğe sahip olma, mevsimlik bir işte çalışma, emekli olma bireyin kazancını etkileyecektir dolayısıyla bireyin tüketim harcamalarında etkili olacaktır (Çalışkan, 2003: 37). Aslında bireyin eğitimi, mesleği, geliri ve tüketim harcamaları arasında ilişki olduğu bir gerçektir.

3.2.1.2.4 Yaş

Yaş, hanehalkı tüketim harcamalarına etki eden bir diğer değişkendir. Brumberg-Ando ve Modigliani'nin Yaşam Döngüsü Hipotezi (Yaşam Boyu Gelir Hipotezi) kazanç profilinin şekli ve yaş dağılımının, toplam tüketim ve tasarruf eğilimini etkileyeceğini göstermiştir (Berg, 1996: 2). Yaşam Boyu Gelir Hipotezine göre birey yaşamının ilk yıllarında gelir düzeyi yetersiz olduğu için "net borçlu" durumundadır. Orta yaşlarında artan geliri ile tüketiminden geriye kalan kısmı ile tasarruf eder, bu tasarrufun bir kısmı ile önceki dönemdeki borçlarını öder ve kalan kısmını ise emeklilik sonrası yaşamı için saklar. Emeklilik sonrası dönemde ise, bireyin menfi tasarrufu söz konusudur; orta yaşlarında biriktirdiği ile tüketimini karşılar. Bu hipoteze göre nüfusun yaş yapısı tüketim harcamalarının belirlenmesinde önemli bir faktördür. Yaş dağılımının değişmesi gelirin, tüketimle tasarruf

arasındaki paylaşımını deęiřtirir. Toplam tüketim harcamaları, yař daęılımındaki deęiřmelerden en çok etkilenen makro ekonomik deęiřkendir (Çalıřkan, 2003: 38; Dönek, 1996: 93).

Tüketim harcamalarının aile bütçesindeki yeri, genç yařlarda zirveye ulařırken orta yařlardan ileri yařlara doęru gidildikçe azalmaya bařlar. Tasarruf edilen gelir ise, genç yařlardan orta yařlara doęru gidildikçe artarken, orta yařlarda zirveye ulařtıktan sonra ileri yařlarda düşer. Bu durum, tüketilen gelirin genç ve yařlı gruplarda yüksek, orta yařlarda düşük bir oran seyrettięini gösterir (Çalıřkan, 2003: 38).

Her yař grubundaki bireylerin istekleri, zevkleri ve tercihleri farklıdır, dolayısıyla bireylerin yařa baęlı olarak tüketim kalıpları da deęiřmektedir. Belirli yař gruplarında belirli ürünlere olan ilgi fazladır. Ev gıda harcamalarına yapılan harcamalar, dięer faktörler sabit iken genç ve yařlılar da nispeten düşük iken, 45-54 yař arası bireylerde ise en yüksek seviyede olduęu görölmüřtür. Evde gıda harcamalarında toplam bütçenin payı genellikle aile reisinin yaşı ile birlikte artar. Hanehalkı reisinin yaşı 65 veya üzerinde olduęu hanelerde toplam harcamaların büyük bir kısmı saęlık harcamalarına ayrılmıřtır. Ayrıca konut giderleri, ev gıda giderleri, toplu ulařım harcamalarına yüksek bir bütçe ayrıldıęı görölmüřtür. 65 yař altındaki bireylere göre alkol, tütün, ev dıřı gıda tüketimi, kıyafet, eęlence, eęitim ve özel araçla ulařım giderleri gibi tüketim kalemlerine daha az bütçe payı ayırmıřlardır. 0-17 yař arası bebek ve gençlerin olduęu hanelerde tüketim harcamalarının bebeęin yâda çocuęun yařına baęlı olarak deęiřtięi görölmüřtür. (Magrabi vd., 1991: 105, 168-174)

3.2.1.2.5 Cinsiyet

Hanehalkı tüketim harcamalarına etki eden dięer bir deęiřken cinsiyettir. Çaędař günümüz toplumunda cinsiyet ve tüketim harcamaları arasında yakın iliřki olduęu bir gerçektir. Birçok kültürde kimin ne tüketeceęi, eril yâda diřil rolleri gerçekleřtirmesine baęlıdır (Firat, 1991: 378). Hanehalkı kompozisyonunda cinsiyetlerin oranı tüketim üzerinde etkili olduęu gibi eřlerin cinsiyeti de tüketim harcamaları üzerinde etkili olmaktadır (Çalıřkan, 2003: 40) Birçok ailede belirli ürünlere yapılan harcamalarda kocanın rolü daha büyüktür. Televizyon, araba gibi yüksek harcama gerektiren ürünlerin satın alınmasında erkeęin rolü daha fazla iken, temizlik ürünleri, mutfak eřyaları, kadının kendi kıyafetleri gibi ürünlerin satın alınmasında kadının rolü daha büyüktür. Birlikte eřit olarak kara verdikleri durumlar ise ev seçimi, tatil yeri gibi sayılabilir. Ülkemizde 1992 yılında DPT tarafından yapılan arařtırmaya göre ailenin gelirinin hangi ihtiyaçları için harcanacaęına karar veren

kişilerin oranı; aile büyükleri-%5,06, evin beyi-%45,87, evin hanımı-%10,64, eşler birlikte-%38,06 ve bilinmeyen-%0,36 olarak görülmüştür (Odabaşı, 1996: 111-113).

Case ve Deaton'un Hindistan'ın kırsal kesiminden ve Güney Afrika'dan derlediği verilerle yaptıkları bir çalışmada belirli yaş gruplarında erkeklerin belirli yaş gruplarında bayanların daha çok harcama yaptıkları görülmüştür. Kırsal Hindistan'da 0-5, 6-15, 36-55 yaş gruplarında erkeklerin, 16-35 ve 56+ yaş gruplarında bayanların daha çok harcama yaptıkları görülmüştür. En çok harcamanın yapıldığı yaş aralığı 36-55 arasındır. Güney Afrika'da ise 0-5, 6-15 yaş aralıklarında bayanların, 16 ve üzeri tüm yaşlarda ise erkeklerin daha çok harcama yaptıkları görülmüştür. En çok harcamanın yapıldığı yaş aralı bura da 36-55 arasındır. Güney Afrika'da gıda harcamalarında 0-5 ve 16-35 yaş grupları hariç erkeklerin bayanlara göre daha fazla harcadıkları görülmüştür. 0-5 yaş grubunda bayanlar daha fazla harcama yaparken, 16-35 yaş arasında harcamalar eşittir. Tütün, alkol, kişisel harcamalar, eğitim harcamaları, çocuk kreşleri/yuvaları (day care), sağlık harcamaları, ayakkabı ve kıyafet harcamalarının da belirli yaş gruplarında farklılaştığı görülmüştür (Case ve Deaton, 2003: 14-15, 50, 53).

3.3 Hanehalkı Gıda ve Yağ Tüketimi Konusunda Yapılmış Çalışmalar

Tez çalışmasının bu bölümünde hanehalkı gıda ve yağ tüketimi konusu ile ilgili çalışmalar detaylı olarak incelenmiştir. Bu çalışmalar 1980 ile 2015 yılları arasında yapılmıştır. İncelen çalışma sayısı toplam 59 adettir.

Ray (1980); yapmış olduğu çalışmasında aile büyüklüğü dâhil edilerek oluşturulan AIDS modeli Hindistan bütçe verilerine uygulanarak dokuz farklı mal grubu için harcama, fiyat ve esneklikler tahmin edilmeye çalışmıştır. Ayrıca Hindistan harcama desenlerinin diğer gelişmekte olan ekonomilerle karşılaştırılması yapılmıştır. Çalışmada Hindistan Ulusal Örneklem Araştırma Ofisi tarafından Nisan 1952 ile Haziran 1969 yılları arasında kırsal, kentsel ve şehir alanlarında derlenen veriler kullanılmıştır. Tahmin yöntemi olarak Tam Bilgi Maksimum Olabilirlik yöntemi kullanılmıştır. (Ray, 1980)

Davis vd. (1983); yapmış oldukları çalışmalarında farklı ırklardaki düşük gelirli hane halklarının seçilen sosyoekonomik özelliklerinin, toplam ve grup gıda harcama kalıplarına etkilerini incelemişlerdir. Veriler 1979-1980 yılları arasında eğitilmiş anketörler vasıtasıyla derlenmiştir ve 300 haneye uygulanmıştır. Çalışmada Çift Logaritmik Fonksiyonel Form, sosyoekonomik faktörlerin ev gıda harcamalarındaki yanıtlarını açıklamak için kullanılmıştır. Çalışmada her bir ev için gıda harcama davranışı fonksiyonel olarak ifade edilmiştir. Bu form; $Q_f = F(I, FSP, HS, A, R, L, S)$ şeklinde tanımlanmıştır. Burada; Q_f - hanenin aylık gıda harcamasını, I - hanenin aylık gelirini, FSP - gıda damga programını, HS - hanedeki kişi

sayısını, A - hane reisinin yaşını, R - hane reisinin ırkını, L - hanenin yerleşim yerini (kırkent), S - hane reisinin eğitim düzeyini göstermektedir. Gelir, hanehalkı büyüklüğü ve FSP'nin hanenin gıda harcamalarına büyük ve pozitif bir etkisinin olduğu görülmüştür. Hanehalkı reisinin yaşı ve eğitim düzeyinin gıda harcamalarındaki değişimleri açıklamada bir etkisine rastlanmamıştır. Hanehalkı reisinin beslenme bilgisi gıda satın alma faaliyetlerinin etkinliğini artırırken, beslenme eğitimi hanenin gıda harcama verimliliğini artırmış dolayısıyla gıda harcamalarını azalttığı görülmüştür. (Davis vd., 1983)

Gould vd. (1991); yapmış oldukları çalışmalarında 1962-1987 yılları arasında derlenen üçer aylık zaman serisi verileri kullanılarak katı ve sıvı yağ talebinin modelini tahmin etmişlerdir. Gıda tercihlerinin demografik, sosyal ve ekonomik faktörlerden etkilendiğini belirtmişlerdir. ABD'de katı ve sıvı yağların tüketimlerini belirleyen değişkenlerin: göreceli fiyatlar, beğeniler, tercihler ve sosyoekonomik değişkenler olduğunu belirtmişlerdir. Deaton ve Muellbauer tarafından formüle edilen AIDS modeli bu analiz için temel olarak kullanılmıştır. Katı ve sıvı yağların tüketiminde yaş, eğitim seviyesi ve beyaz olmayan nüfusun tahmini etkileri genellikle istatistiksel olarak anlamlı olarak görülmüştür. (Gould vd., 1991)

Nayga (1994); çalışmasında gıda enerjisi, protein, A vitamini, C vitamini, tiamin, riboflavin, niasin, kalsiyum, fosfor ve demir tüketiminde sosyoekonomik ve demografik faktörlerin etkisini araştırmıştır. Analizde kullanılan sosyoekonomik ve demografik faktörler; kentleşme, bölge, ırk, etnik köken, cinsiyet, istihdam durumu, gıda damga katılımı, hanehalkı büyüklüğü, ağırlık, boy, yaş ve gelir olarak belirlenmiştir. ABD'de Ulusal Gıda Tüketimi Anketinden Nisan 1987 ile Ağustos 1988 yılları arasında derlenen, 6.259 gözlemden oluşan kesitsel veri analizde kullanılmıştır. Çalışmada Çoklu Regresyon Modeli kullanılmıştır. Kentte yaşayan bireylerin banliyölerde yaşayan bireylere göre riboflavin, kalsiyum ve fosforu daha fazla tükettikleri görülmüştür. Kuzeydoğuda yaşayan bireyler güneyde yaşayan bireylere göre daha fazla A, C vitamini ve kalsiyum tüketmektedir. Bir diğer farklılık ise ırkda görülmektedir, siyahî bireyler beyaz bireylere göre daha fazla C vitamini tüketirken, gıda enerjisi, A vitamini, tiamin, riboflavin, niasin, kalsiyum, fosfor ve demiri daha az tüketmektedirler. Gıda enerjisi, protein, tiamin, riboflavin, kalsiyum ve fosfor tüketimi yaş ve boy için istatistiksel olarak anlamlı besinlerdir. Yıllık gelir A, C vitamini ve kalsiyum tüketimini etkileyen önemli bir faktördür. Erkeklerin tüm besin maddelerini kadınlardan daha fazla tükettikleri görülmüştür. İşsiz bireylerin, çalışan bireylere göre daha fazla A ve C vitamini, riboflavin, kalsiyum ve demir tükettikleri görülmüştür. Gıda damga programı katılımcıları bu programa katılmayan bireylerle kalsiyum dışında aynı miktarda besinleri

tükettikleri görülmüştür. Hanehalkı büyüklüğünün A, C vitamini, tiamin ve niasin tüketimi ile negatif ilişkili olduğu görülmüştür (Nayga, 1994).

Ippolito ve Mathios (1995); yapmış oldukları çalışmalarında ABD'de kanser ve kalp hastalığı riskine neden olan yağ ve doymuş yağların tüketimlerini incelemişlerdir. Çalışmada kullanılan veriler ABD Tarım Bakanlığına ait 1977, 1985, 1986, 1987/88 ve 1989/90 yıllarına aittir. Verilerin analizinde temel istatistikî yöntemler kullanılmıştır. Çalışmanın sonuçlarına göre 1977-1990 arasında günlük doğrudan alınan yağ miktarı zaman içerisinde düşerken, alınan toplam yağ miktarında (doğrudan ve dolaylı alınan) düştüğü görülmüştür. Dolaylı olarak alınan yağların kaynağının et, kümes hayvanları, balık, süt, yumurta, tatlı, hububat (tahıl), meyve, atıştırmalık, peynir, içecekler gibi vb gıda ürünleri olduğu belirtilmiştir. 1977-1990 yılları arasında lise ve üzeri eğitim alanlarda yağ ve doymuş yağ tüketimi ile ilgili olarak meydana gelen değişim, eğitim düzeyi lise altında olan bireylerde meydana gelen değişime göre daha yüksek olduğu görülmüştür. (Ippolito ve Mathios, 1995)

Lanjouw ve Ravallion (1995); yapmış oldukları çalışmalarında yoksulluk ve hanehalkı arasındaki ilişkiyi analiz etmişlerdir. Çalışmada kullanılan veriler, 1991 yılında Pakistan'ın 300 kentsel ve kırsal kesiminden 4.794 haneden derlenmiştir. Çalışmada kullanılan bağımsız değişkenler hanehalkı harcaması, hanehalkı büyüklüğü, demografik değişkenlerdir (yetişkinlerin yaşı, çocukların yaşı, bebeklerin yaşı, çocukların cinsiyeti, anne ve babanın okuryazarlık durumu, bebeklerin, çocukların, yetişkinlerin hanehalkı içindeki oranları). Çalışmada Engel Eğrileri kullanılmıştır. Çalışmanın sonuçlarına göre, gelişmekte olan ülkelerde büyük hanehalklarının yoksul olma eğiliminde olduğu görülmüştür. Gelişmekte olan ülkelerde kişi başına gelir (veya kişi başına harcama) ile hanehalkı büyüklüğü arasında negatif bir ilişki vardır. (Lanjouw ve Ravallion, 1995)

Hupkens vd. (1997); yapmış oldukları çalışmalarında sosyal sınıf farklılıklarının yağ ve lif alımları üzerindeki etkisini araştırmışlardır. Çalışma Maastricht, Liege ve Aachen şehirlerinde, okula giden çocuğu olan çekirdek ailelerin bayanları üzerinde yapılmıştır. Analizde, Kasım 1993 ile Şubat 1994 (23 Aralık 1993-9 Ocak 1994 arası hariç) arasında derlenen 849 gözlem kullanılmıştır. Verilerin analizinde Varyans Analiz ve Ki Kare yöntemleri kullanılmıştır. Sosyal sınıfı belirlemede kullanılan kriter eğitim düzeyidir. Her sosyal sınıf için ve her şehir detaylı şekilde et ve et ürünleri, süt ve süt ürünleri, peynir, yenilebilir sıvı ve katı yağların günlük tüketimleri belirtilmiştir. Sosyal sınıflar ve şehirler dikkate alınarak detaylı karşılaştırmalar yapılmıştır. Çalışmanın sonucuna göre sosyal sınıf farklılıklarının yağ alımını etkilediği görülmüştür. (Hupkens vd., 1997)

Deaton ve Paxson (1998); yapmış oldukları çalışmalarında gıda harcamaları ile hanehalkı genişliği arasındaki ilişkiyi açıklamışlardır. Çalışmada kullanılan verilerin derlendiği ülkeler ve yılları ABD-1990, Büyük Britanya-1992, Fransa-1989, Tayvan-1990, Tayland-1992 Pakistan-1991, Güney Afrika-1993 yılıdır. Çalışmada Parametrik Olmayan, Yarı Parametrik ve Parametrik Engel Eğrileri kullanılarak analiz yapılmıştır. Kişi başına toplam hanehalkı harcamaları sabit tutulduğunda, kişi başına yapılan gıda harcaması kişi sayısı ile birlikte düşer. Buna göre ABD, Britanya ve Fransa’da gıda tüketimi hanehalkı ölçek artışına göre küçük bir miktarda düştüğü görülmüştür. Tayland, Pakistan ve Güney Afrika gibi fakir ülkelerde kişi başına tüketim harcamaları sabit tutulduğunda hanehalkı büyüklüğündeki bir birim artış gıdanın bütçe payı içerisindeki oranını %5 azatırken, kişi başına yapılan gıda harcamalarında %10’dan fazla azalmaya yol açmıştır. Teoriye göre yoksul ülkelerde kişi başına düşen gıda tüketimi ile hane halkı büyüklüğü arasında güçlü bir ilişki olmalıdır. Ancak yoksul ülkelerde hane halkı büyüklüğü ile kişi başına düşen gıda tüketimi arasında negatif bir ilişki bulunmaktadır (Deaton ve Paxson, 1998).

Roos vd. (1998); yapmış oldukları çalışmalarında Finli erkek ve bayanların gıda davranışlarının sosyal belirleyicilerini araştırmışlardır. Çalışmanın verileri Finladiya’da 1992 yılında derlenen Finmonica Risk Faktörü Anketi’nden elde edilmiştir. Sağlık ve diyet davranışları, sosyal faktörlerin çeşitli yönlerini içeren verilerin bulunduğu 6.051 gözlem yer almaktadır. Çalışmada kullanılan sosyoekonomik statünün belirlenmesinde eğitim karakteristiğinden yararlanılmıştır. Eğitim, gelir, mesleki sosyal sınıf, medeni durum, ebeveyn durumu, çalışma durumu gibi sosyoekonomik ve demografik karakteristikler çalışmada kullanılmıştır. Çalışmanın analizinde Ki-Kare, T-Testi ve Lojistik Regresyon Analizi kullanılan yöntemlerdir. Her iki cins içinde eğitim seviyesi ile gıda davranışı ilişkisi benzerdir. Çalışma durumu sadece kadınların gıda davranışı ile ilgilidir. Çocuklu kadınların gıda davranışı, kadınların geri kalan kısmına göre diyet doğrultusuna daha yakın olarak görülmüştür. Yüksek eğitilmiş kadın ve erkeklerin gıda davranışları, temel eğitim almış kadın ve erkeklere göre diyet doğrultusuna daha yakındır. Kadınların gıda davranışlarının belirleyicileri eğitim düzeyi, çalışma durumu, medeni durum, ebeveyn durumu olarak görülmüştür. Ebeveyn durumu sadece kadınların gıda davranışı ile ilişkili olarak bulunmuştur. Medeni durum hem erkek hem de kadın gıda davranışı ile ilişkilidir. (Roos vd., 1998)

Özer (1999); yapmış olduğu doktora tezinde 1987-1994 döneminde Türkiye’de meydana gelen iktisadi ve sosyal değişmelere bağlı olarak tüketici tercihlerinin yapısının nasıl oluştuğunu analiz etmiştir. Çalışmanın amacını detaylandırıldığında Doğrusal Harcama Sistemi ile Türkiye’de hanehalkı tüketim kalıplarını belirlemek, mal grupları itibarıyla Engel

eğrilerini ve esneklikleri tahmin etmek ve bunlar vasıtasıyla malları sınıflandırmak, 1987-1994 döneminde gerek ana mal grupları gerekse tali mal (gıda) grupları itibariyle tüketim kalıplarındaki yapısal değişimi araştırmak, tüketim kalıplarındaki yapısal değişime bağlı olarak tüketici refahında ortaya çıkan değişimi tespit etmek ve hanehalkı tüketim kalıplarını kır/kent ve coğrafi bölgeler itibariyle karşılaştırmak suretiyle aradaki farklılıkları (eğer varsa) ortaya koymaktır. Çalışmada kullanılan veriler TÜİK (Devlet İstatistik Enstitüsü, DİE) tarafından uygulanan 1987 ve 1994 Hanehalkı Gelir ve Tüketim Harcamaları Anketlerinden elde edilmiştir. 1987 yılına ait 26.400 adet, 1994 yılına ait 26.256 adet haneye ait veriler analizde kullanılmıştır. (Özer, 1999)

Huang ve Lin (2000); yapmış oldukları çalışmalarında hanehalkı bütçe anketi verilerinden bir talep sistemi tahmini için yeni bir yaklaşım geliştirmişlerdir. Üç farklı gelir seviyesine ayrılan hanehalkları için gıda taleplerini ve talep esnekliklerini analiz etmişlerdir. Çalışmanın verileri, 1987-88 yılında ABD’de yapılan Ulusal Gıda Tüketim Anketi’nden elde edilmiştir. Analizde 4.245 hanehalkına ait veriler kullanılmıştır. Hanehalkının tükettiği gıdalar sığır eti, domuz eti, kümes hayvanları, diğer etler, balık, süt ve süt ürünleri (mandıra ürünleri), tahıl, ekmek, katı ve sıvı yağlar, yumurtalar, sebzeler, meyveler, meyve suyu şeklinde onüç farklı gruba ayrılmıştır. Tüketicilerin gıda tercihlerini etkileyen sosyoekonomik ve demografik faktörler analize dâhil edilmiştir. Bunlar; kişi başı hanehalkı geliri, kişi başı ev gıda harcaması, hanehalkı reisi eğitim seviyesi, ev dışı gıda harcamasının payı (away-from-home share of the food budget), ırk, kentleşme (kır, kent, banliyö), bölge, hanehalkı kompozisyonu ve hanehalkı büyüklüğüdür. Çalışmanın analizinde AIDS modeli kullanılmıştır. Çalışmanın sonuçlarına göre her gıda kategorisindeki bütçe payı, kişi başına düşen gıda harcaması değişikliğinden etkilenir. Tahminlere göre toplam gıda harcamaları arttıkça sığır eti, diğer et, yumurta, tahıl ve ekmeğin bütçe paylarının azaldığı görülmüştür. Ayrıca kümes hayvanları, meyve, sebze ve meyve suyunun bütçe paylarının, toplam gıda harcamaları ile birlikte arttığı görülmüştür. Geniş haneler gıda harcamalarının büyük kısmını sığır eti ve domuz etine ayırırken, küçük kısmını ise ekmek ve meyve suyuna ayırdıkları görülmüştür. Hanehalkı reisinin eğitim düzeyinin üniversite olduğu hanelerin sığır eti, domuz eti, diğer et ve yumurtaya gıda bütçesinin daha az kısmını ayırırken, kümes hayvanları, meyve ve sebzeye gıda bütçesinin daha fazla kısmını ayırdığı görülmüştür. Diğer bir farklılık ırk değişkeninde görülmüştür. Siyahi hanehalkı reisleri, beyaz hanehalkı reislerine göre domuz eti, kümes hayvanları, diğer et, balık, yumurta ve meyve suyuna gıda bütçelerinin daha fazla kısmını ayırırken, ekmek, süt ve süt ürünleri (mandıra ürünleri) ve meyveye daha az kısmını ayırdıkları görülmüştür. Şehir ve banliyölerde yaşayan hanehalkları, kırsal kesimde yaşayan

hanehaklarına göre meyve ve meyve suyuna gıda bütçelerinin daha fazla kısmını ayırırken, yağ ve domuz etine gıda bütçelerinin daha az kısmını ayırdıkları görülmüştür. Ayrıca bölgesel ve mevsimsel farklılıkların gıda bütçe tahsisinde önemli rol oynadığı görülmüştür. Örneğin Batı'da yaşayan hanehalklarının diğer hanehaklarına göre, süt ve süt ürünleri (mandıra ürünleri) ve meyveye gıda bütçelerinin önemli bir kısmını ayırırken, domuz eti ve diğer etlere daha az kısmını ayırdıkları görülmüştür. Hanehalkı kompozisyonunun gıda bütçe tahsisini etkilediği görülmüştür, 2-5 yaş arası çocukların fazla olduğu hanelerde süt ve süt ürünlerine gıda bütçesinin daha fazla kısmını harcadıkları görülmüştür. (Huang ve Lin, 2000)

Jae vd. (2000); yapmış oldukları çalışmalarında kentsel ailelerin hazır gıda harcamaları ve aile karakteristikleri arasındaki ilişkiyi incelemişlerdir. Çalışmanın verileri 1996 yılında Kore'de yapılan Kore Aile Gelir ve Harcamaları Anketinden elde edilen 39.025 kentsel aileye ait gözlemden oluşmaktadır. Aylık toplam harcama, hanehalkı reisinin yaşı, çocuk sayısı, altı yaşından küçük çocuk sahipliği, hanehalkı reisinin eğitimi, karısının iş gücüne katılımı, ikamet yeri ve ev sahipliği durumu gibi sosyoekonomik ve demografik karakteristikler çalışmanın analizinde kullanılmıştır. Verilerin analizinde kullanılan yöntem Lojistik Regresyon modelidir. Çalışmanın bulgularına göre bazı aile karakteristiklerinin hazır gıda harcamalarını etkilediği görülmüştür. Çalışmada hazır gıda harcamalarının batılı ülkelerde yapılan çalışmalara göre farklılıklar gösterdiğine işaret edilmiştir. Aile geliri ve karısının iş gücüne katılımının kentsel Kore'de hazır gıda harcamalarında negatif ve önemli bir etkisi varken, Batılı ülkelerde tam tersi bir etkisi olduğu görülmüştür. Diğer şartlar sabitken, hanehalkı reisinin yaşı ve eğitimi, çocuk sayısı, altı yaşın altındaki çocuk varlığı ve ikamet yeri gibi karakteristiklerin Batılı ülkelerle benzer etkilere sahip olduğu görülmüştür. Sonuçlar göstermektedir ki, hanehalkı reisinin yaşı ve eğitimi, çocuk sayısı, altı yaşından küçük çocuk varlığı, eşinin işgücüne katılımı ve ikamet yeri hazır gıda harcama düzeyini tahmin etmede anlamlı bulunmuştur. Ev sahibi olmanın hazır gıda harcamalarına anlamlı etkisinin bulunmadığı sonucuna varılmıştır. (Jae vd., 2000)

Liu ve Chern (2001); yapmış oldukları çalışmalarında demografik değişkenlerin gıda tüketimi üzerindeki etkisini araştırmışlardır. Veri seti, Çin'in Jiangsu bölgesinde, Ulusal İstatistik Bürosu tarafından 1994 yılında, 800 haneden derlenen detaylı gıda tüketim bilgileri ve demografik karakteristikleri içermektedir. Veri setinde altı gıda ürünü yer almıştır, bunlar; pirinç, un, domuz eti, süt, taze sebze, taze meyvedir. Veri setinde yer alan demografik karakteristikler, buzdolabı sahipliği, hanehalkı genişliği, 17 yaşın altındaki çocuk sayısı, cinsiyet, yaş grubu, eğitim düzeyi, bölge, kır-kent durumudur. Çalışmada kullanılan modeller Working-Leser, LA/AIDS, LES ve QES'dir. Seçilen modellerdeki bilinmeyen parametreleri

tahmin etmek için, tek eşitlik ve tam talep sistemi için sırasıyla En Küçük Kareler Yöntemi (OLS) ve Görünüşte İlişkisiz En Regresyon Yöntemleri kullanılmıştır. Çalışmanın sonucuna göre Working-Leser ve LA/AIDS modelleri benzer sonuçlar göstermektedir. LES ve QES sonuçları oldukça yakın çıkmıştır. Fakat LA/AIDS ve LES sonuçları birbirinden farklıdır. Demografik değişkenler seçilen modelin performansını artırmıştır. Her demografik değişkenin etkisi seçilen modellerde farklı olduğu görülmüştür. LES ve QES için demografik değişkenlerin etkisi belirsizdir. Pirinç ve un gıda maddeleri için bölge ve kentleşme önemlidir. Working-Leser ve LA/AIDS modellerinde eğitim seviyesi süt tüketimi için önemli olduğu görülmüştür. Hanehalkı büyüklüğü ve 17 yaş altındaki çocukların sayısını LES ve QES modelleri için anlamlı iken Working-Leser ve LA/AIDS modelleri için anlamlı olmadığı görülmüştür. (Liu ve Chern, 2001)

Mattison vd. (2001); yapmış oldukları çalışmalarında üç farklı enerji düzenleme yaklaşımının bireylerin yağ alım derecesini nasıl etkilediğini ve her bir enerji düzenleme yaklaşımıyla ilgili olarak yağ alımı ile sosyo-ekonomik durum, demografik durum ve yaşam tarzı özellikleri arasındaki ilişki tahlilini karşılaştırmışlardır. Çalışmanın verileri, Malmö Diyet ve Kanser Çalışması'nın alt örneklem verilerinden alınmıştır. 7.055 bayan, 3.240 erkek bireyden (toplam: 10.295) oluşan bu veriler, 1995 ve 1996 yıllarında derlenmiştir. Her katılımcının yağ alımı için Üç değişken tanımlanmıştır. Bu değişkenler; Fatg: gram bazında günlük mutlak yağ alımı, Fat%: göreceli yağ alımı, yağlardan alınan enerjinin toplam enerji içindeki yüzdesi (FAT%), Fatres: toplam enerjinin toplam yağ üzerine regresyonundan kalan hata değerleridir. Katılımcıların vücut kitle oranları (BMI), yaş, cinsiyet, sosyoekonomik statüleri, eğitim seviyeleri, etnik köken, boş zaman fiziksel aktiviteleri, kohort durumu (cohort status), enerji alımı-bazal metabolizma hızı oranı, sigara içme alışkanlıkları ve günlük alkol tüketimi değişkenleri çalışmanın analizinde kullanılmıştır. Çalışmanın analizinde Lojistik Regresyon ve Çok Değişkenli Analiz kullanılmıştır. Çalışmanın sonuçlarına göre, yalnız yaşayan kadınlarda yüksek yağ alımı görülmüştür. Yüksek sosyoekonomik statüye sahip fakat eğitim düzeyi düşük olan kadınlarda yüksek yağ alımı arasında bir ilişki görülürken, FAT% ve FATres ile ilgili hiçbir ilişki görülmemiştir. Yalnız yaşayan ve eğitim düzeyi düşük erkeklerde, yüksek FAT% ve FATres arasında bir ilişki görülürken, yağ alımı arasında bir ilişki görülmemiştir. Yüksek yağ alımı ile sosyoekonomik statü ve eğitim arasındaki ilişkinin karmaşık olduğu görülmüştür. Yalnız yaşayan bireyler ile diğer tüm hane yapıları karşılaştırıldığında cinsiyetten bağımsız olarak yüksek yağ alımı olduğu görülmüştür. Her iki cinsiyet içinde, alkol kullanımı ile yüksek yağ alımı arasında ilişki olduğu görülmüştür. (Mattison vd., 2001)

Abdulai (2002); yapmış olduğu çalışmasında İsviçre’de gıda harcamalarını tahmin etmiştir. Çalışmada kullanılan veriler, Federal İstatistik Ofisi tarafından 1998 yılında 9.295 haneden derlenen Hanehalkı Harcama Anketi’nden elde edilmiştir. Çalışmanın analizinde kullanılan gıda ürünleri ekmek ve hububat (tahıl), et ve balık, süt, peynir ve yumurta, katı ve sıvı yağlar, meyveler ve sebzeler ile gıda dışı (şeker ve içecekler) ürünlerdir. Ayrıca sosyoekonomik ve demografik karakteristiklerde analizde yer almaktadır. Bunlar; hanehalkı büyüklüğü, aylık hanehalkı geliri, kişi başı aylık harcama tutarı, verinin toplandığı ay, bölge, katılımcının yaşı, eğitim durumu, mesleğidir. Çalışmada parametrelerin tahmini için QUAIDS modeli kullanılmıştır. Çalışmanın sonucuna göre modele giren demografik değişkenlerin çoğu önemli çıkmıştır. Tüm gıda grupları için, hanehalkı büyüklüğü katsayısı pozitif ve anlamlı iken gıda dışı grubunda bu katsayı anlamlı ama negatiftir. Bölgesel farklılıklar bütçe payları karşısında nispeten küçük bir etkiye sahiptir. Tüketicinin gelir ile harcamalarının beraber arttığı görülürken, gelirin artmasıyla beraber et, balık, sebze ve meyveye olan talebin arttığı görülmüştür. Hanehalkı büyüklüğünün, hanenin tüketim kalıplarını etkilediği görülmüştür. Harcama ve fiyatların belirli bir düzeyi için, geniş haneler pahalı nispeten ucuz malları tercih ederek kendi tüketim kalıplarını ayarlamak zorunda bırakılır. (Abdulai, 2002)

Fang ve Beghin (2002); yapmış oldukları isimli çalışmalarında Çin’in üç bölgesinde 1992 yılı ve 1994-1998 yılları arasındaki 10.000 hanehalkına ait kentsel hanehalkı verilerini kullanmışlardır. Linquad Eksik Talep Sistemine dayalı olarak, yemeklik bitkisel yağlar ve hayvansal yağların nihai talep tahminleri sağlamaya çalışmışlardır. Her bölgenin önemli bitkisel yağ talep fiyatının esnek olduğu görülmüştür. Tüm bitkisel yağların 1’den az pozitif gelir elastikiyeti vardır. Çinli tüketicilerin geliri arttıkça yağ tüketimlerini çeşitlendirdiği görülmüştür. (Fang ve Beghin, 2002)

Gould (2002); yapmış olduğu çalışmasında Çin’deki üç kentsel ildeki hanelerin evde gıda tüketim harcamalarının önemli belirleyicilerini tespit etmek istemiştir. Analizde kullanılan veriler 1995-1997 yılları arasında Çin Halk Cumhuriyeti Devlet İstatistik Bürosu tarafından yapılan yıllık anketten elde edilmiştir. Satın alma bilgileri ve miktarları, hanehalkı üyelerinin yaş, cinsiyet ve eğitim düzeyleri, hane geliri, iş gücüne katılım özelliği karakteristikleri veri setinde yer almaktadır. Toplam 5.273 haneden derlenen veriler analizde kullanılmıştır. Tahmin yöntemi olarak Tam Bilgi Maksimum Olabilirlik yöntemi kullanılmıştır. Çalışmanın sonucuna göre Çin gıda talebinin fiyat değişimlerine duyarlı olduğu görülmüştür. Toplam hanehalkı ev gıda harcamaları ile hanehalkı geliri, hanedeki yetişkin birey sayısı ve buzdolabı sahipliği arasında pozitif ilişkili olduğu görülmüştür. Bölge ve alt bölgeler genelinde harcamalarda belirgin farklılıklar bulunmuştur. Hanedeki yetişkin

birey sayısında meydana gelen artışa göre hanehalkının harcamalarını ayarlamaya çalıştığı görülmüştür. Örneğin, kümes hayvanları ve deniz ürünleri talebi düşerken, domuz eti ve diğer gıda maddelerinin talebi yükselmiştir. (Gould, 2002)

Raper vd. (2002); yapmış oldukları çalışmalarında ABD hanehalkının gıda harcamaları, yoksulluk ve yoksul olmama durumuna göre geçim miktarlarının, demografik değişkenlerin doğrusal kombinasyonları ile Doğrusal Harcama Sistemi dâhilinde analizini yapmışlardır. Çalışmada ABD İstatistik bürosu tarafından 1992 yılına ait Tüketici Harcamaları Anketi ve Ayrıntılı Aylık Tüketici Fiyat Endeksleri verileri kullanılmıştır. Veri setinde 3.577 haneye ait 7.154 gözlemden elde edilen veriler bulunmaktadır. Gıda ürünleri harcamaları dokuz harcama kategorisinde toplanmıştır. Bunlar; alkolsüz içecekler, ev dışı tüketilen gıdalar, et (sığır, domuz, kümes hayvanları, diğer et, deniz ürünleri, yumurta), evde tüketilen diğer gıdalar (ev dışında hazırlanmış ve ev içinde yenen gıdalar), tahıl ve unlu mamuller, mandıra (süt) ürünleri, meyve ve sebzeler (taze ve işlenmiş); tatlılar ve şekerler, katı ve sıvı yağlardır. Analizin sonuçlarına göre; yoksulluk sınırında 2.378 gözlem (1.189 hane), yoksul olmayan 4.776 gözlem (2.388 hane) bulunmaktadır. Pek çok demografik karakteristik yoksul ve yoksul olmayan haneler için geçim miktarları ile ilişkili ve istatistiksel olarak anlamlıdır. Hanehalkı reisinin siyahi yâda başka ırktan olması geçim miktarı ile negatif ilişkili iken, çocuk sahipliği geçim miktarı ile pozitif ilişkilidir. Hanelerin yoksulluk durumu için hanehalkı büyüklüğü yedi gıda grubunda istatistiksel olarak anlamlıdır ve hanehalkı büyüklüğü geçim miktarı ile negatif ilişkilidir. Yoksul olmayan haneler için, sadece iki gıda grubu istatistiksel olarak anlamlıdır. Bunlar süt/mandıra ürünleri ile katı ve sıvı yağlardır. İlginçtir ki, yoksul olan hanehalkı için her gıda kategorisinde siyahi hanehalkı reisi anlamlı değilken, yoksul olmayan hanehalkı için siyahi hanehalkı reisi anlamlıdır. Yoksul olmayan hanelerde ırk ve çocuk kompozisyonu geçim miktarının belirlenmesinde aile büyüklüğünden daha önemlidir. Yüksek gelir seviyelerinde ırk ve çocukların varlığı geçim miktarların önemli belirleyicileri olmaya devam ettiği görülmüştür. Bazı gıda gruplarında yoksul olan hanelerin, yoksul olmayan hanelere göre daha büyük harcama esnekliklerinin olduğu görülmüştür. Bunlar dokuz gıda grubunun altısıdır, alkolsüz içecekler, evde tüketilen diğer gıdalar, tahıl ve unlu mamüller, süt/mandıra ürünleri, meyveler ve sebzeler, katı ve sıvı yağlardır. (Raper vd., 2002)

Yen vd. (2002); yapmış oldukları çalışmalarında 1987-1988 yıllarına ait ABD Ulusal Gıda Tüketimi Araştırması kesitsel verileri kullanılmıştır. İki aşamalı prosedür yaklaşımıyla katı ve sıvı yağ talebinin denklemleri tahmin edilmiştir. ABD'de hanehalkı tarafından tüketilen sıvı ve katı yağların (tereyağı, margarin, hamur işlerinde kullanılan katı yağ, pişirme

yağı ve salata sosu) tüketimi için Sansürlü Sistem Tahmincisi, doğrusal olmayan Translog Talep Sistemine uygulanmıştır. Hanehalkı karakteristiklerini içeren 3.943 gözlem analizde kullanılmıştır. Üç gruba ayrılmış hanehalkı kompozisyonu (yaş grupları: 20 yaş altı, 21-64 yaş arası, 64 yaş üzeri), gelir, hanehalkı reisinin eğitimi, kır-kent durumu, bölge, ev sahipliği, ırk, etnik köken, yemek planlayıcısının cinsiyeti, gıda damga katılım durumu (food stamp participation status) gibi sosyoekonomik, demografik ve hanehalkı karakteristikleri kullanılmıştır. Tüketilen ürünlerin ortalama miktarları (haftada), tereyağı 0,61 lb, margarin için 0,76 lb, hamur işlerinde kullanılan katı yağ (shortening) için 0,90 lb, yemeklik yağ için 0,68 lb ve salata sosu için 0,73 lb'dir. Tereyağı kullanan hanehalkı sayısı 964 olup bütün hanelerin %24'ü, margarin kullanan hanehalkı sayısı 3.187 olup bütün hanelerin %81'i, hamur işlerinde kullanılan katı yağ kullanan hanehalkı sayısı 785 olup bütün hanelerin %20'si, pişirme yağı kullanan hanehalkı sayısı 1.452 olup bütün hanelerin %37'si, salata sosu kullanan hanehalkı sayısı 2.865 olup bütün hanelerin %73'ü kullanmaktadır. Gelir karakteristiği tereyağı, yemeklik yağ ve salata sosu tüketimini belirlemede pozitif rol oynarken, margarin tüketimini belirlemede negatif bir rol oynamaktadır. Eğitim karakteristiği margarin tüketimini belirlemede negatif bir rol oynarken, tereyağı tüketimini belirlemede etkisiz bir rol oynamaktadır. Eğitim karakteristiği salata sosu tüketiminde pozitif etkili iken, hamur işlerinde kullanılan bitkisel yağların tüketiminde negatif etkilidir. Yemek planlayıcılarının kadın olduğu hanelerde margarin, hamur işlerinde kullanılan katı yağlar, yemeklik yağ ve salata sosu tüketme daha olasıdır. Beyaz tenli kişilerin yaşadığı hanelerde diğer ırkların yaşadığı hanelere göre margarin ve salata sosu tüketmesi daha olası iken, yemeklik yağ tüketimi daha az olasıdır. Sıvı ve katı yağ tüketiminde bölgesel varyasyon etkilidir. Ayrıca kentleşme, etnik köken ve ev sahipliği bir ya da daha fazla malın tüketiminde önemli rol oynamaktadır. 20-64 yaş arası kişilerin yaşadığı hanelerde demografik esneklikler margarin ve salata sosu için önemlidir. (Yen vd., 2002)

Friel vd. (2003); yapmış oldukları çalışmalarında besin alımı ile sosyal varyasyon arasındaki ilişkiyi incelemişlerdir. Çalışmanın verileri İrlanda Cumhuriyeti kırsal ve kentsel alanda 6.539 yetişkine uygulanmış anketten derlenmiştir. 1998 yılında uygulanan Yaşam, Tutum ve Beslenme Anketi'ne ait veriler çalışmanın analizde kullanılmıştır. Bireylerin boy ve kiloları, yaş, cinsiyet, sosyal statüleri, eğitim durumu, tıbbi kart uygunluğu/niteliği (eligibility), medeni durum, hanehalkı büyüklüğü ve kır-kent durumu çalışmanın analizinde kullanılan sosyoekonomik ve demografik değişkenlerdir. ANCOVA, Kolmogorov-Smirnov ve Çok Değişkenli Regresyon Modeli çalışmanın analizinde kullanılmıştır. Çalışmada verileri derlenen gıda ürünleri; yağ, karbonhidrat, kolesterol, protein, alkol, lif, vitamindir (A, C, D,

B6, B12). Çalışmanın sonucuna göre, erkeklerin kadınlara göre önemli ölçüde daha yüksek toplam enerji (günlük) alımı gerçekleştirmiştir. Enerjinin önemli ölçüde büyük bir yüzdesini yağ, doymuş yağ, tekli doymamış yağ ve alkolden sağlamışlardır. Erkeklerin A ve B12 vitamini, riboflavin, kalsiyum ve çinko açısından daha yüksek ortalamaya sahip olduğu görülmüştür. Kadınlarda yaşın artması ile birlikte karbonhidrattan alınan enerji, kolesterol ve B6 vitamini dışında tüm besinlerin alımında azalma olduğu görülmüştür. Erkeklerde karbonhidrat ve proteinden alınan enerjinin yaş ile birlikte arttığı görülmüştür. Kadınlarda ve erkeklerde yaşın artmasıyla birlikte yağdan alınan enerjinin azaldığı görülmüştür. Sosyal statü azalırken, yağ ve proteinden elde edilen enerjinin toplam enerjiye olan katısının arttığı görülmüştür. Protein, toplam yağ, tekli doymamış yağ ve doymuş yağdan elde edilen enerji alımı, vitamin A ve B12 alımı ile kadınların eğitim düzeylerinin (arasında) ters ilişkili olduğu görülmüştür. Eğitimsiz yâda düşük eğitimli kadınların günlük lif, karbonhidrat ve alkolden alınan enerji, C vitamini, fosfor, folik asit, kalsiyum ve demir alımının önemli derecede düşük olduğu görülmüştür. Eğitimsiz yâda düşük eğitimli erkeklerin günlük çoklu doymamış yağ asitleri, karbonhidrat ve alkolden alınan enerji, lif, B6 ve C vitamini ve folik asit alımının önemli derecede düşük olduğu görülmüştür. Medikal kart statüsünün her iki cinsiyette de gıda ve enerji alımında istatistiksel olarak anlamlı bir değişim göstermediği görülmüştür. Medeni durum statüsünün, erkeklerin enerji ve besin alımları üzerinde etkili olduğu görülmüştür. Tek yaşayan erkeklerin, aile ile birlikte yaşayan erkeklere göre enerji ve gıda ürünlerinin yarısında anlamlı derecede daha düşük alım yaptıkları gözlenmiştir. Kentte yaşayan erkeklerin kırsal alanda yaşayan erkeklere kıyasla, yağdan alınan enerjinin düşük olduğu görülürken kadınlarda ise proteinden alınan enerjinin daha düşük olduğu görülmüştür. Enerji ve gıda alımında yaş, cinsiyet, sosyal statü, kır-kent durumu, medeni durum, eğitim durumu ve hanehalkı büyüklüğü gibi sosyoekonomik ve demografik karakteristiklerin etkili olduğu görülmüştür. (Friel vd., 2003)

Hulshof vd. (2003); yapmış oldukları çalışmalarında farklı sosyoekonomik statüdeki yetişkinlerin diyet alımlarını ve zaman içerisindeki eğilimlerini incelemişlerdir. Çalışmada Hollanda Ulusal Gıda Tüketimi Araştırmasının (DNFCS-1:1987/88, DNFCS-2:1992; DNFCS-3 = 1997/98) 19 yaş ve üzeri 6.008 erkek, 6.957 bayan bireyden derlenen verileri analizde kullanılmıştır. Eğitim, meslek ve mesleki pozisyonuna göre yüksek, orta, düşük ve çok düşük şeklinde sosyoekonomik statülerde sınıflandırılmıştır. Analizde cinsiyete ve yaşa göre düzenlenmiş Lojistik Regresyon Modeli kullanılmıştır. Araştırmada erkek ve bayanlar için bütün sosyoekonomik statüler ve anket dönemleri göz önüne alınarak günlük ekmek, patates, sebze, meyve, süt ve süt ürünleri, peynir, et ve kümes hayvanları, katı ve sıvı yağlar

vb gıda ürünlerinin günlük alınan miktarları (gram/gün) belirtilmiştir. Düşük sosyoekonomik statüye sahip bireylerin patates ve alkolsüz içecekleri yüksek oranda tükettikleri görülmüştür. Daha yüksek sosyoekonomik statüye sahip bireylerin meyve, peynir, tuzlu atıştırmalıklar, balık, hazır yenilebilen et (sadece erkekler), meyve suyu, şarap ve çayı yüksek oranda tükettikleri görülmüştür. Düşük sosyoekonomik statüye sahip olan erkekler, daha yüksek sosyoekonomik statüye sahip olan erkeklere göre patates, ekmek, şeker ve tatlılar, sürülebilir yağlar ve pişirme yağları (spreads and cooking oils), et, bira ve kahveyi daha çok tükettikleri görülmüştür. Düşük sosyoekonomik statüye sahip olan kadınların yüksek sosyoekonomik statüye sahip kadınlara göre patates, katı ve sıvı yağlar, et, alkolsüz içecekler ve kahveyi daha çok tükettikleri görülmüştür. Yüksek sosyoekonomik statüye sahip kadın ve erkeklerin sebze, peynir ve şarabı, düşük sosyoekonomik statüye sahip erkek ve kadınlara göre daha çok tükettikleri görülürken, yüksek sosyoekonomik statüye sahip erkek ve kadınların protein alımlarının yüksek, yağ alımlarının düşük miktarda olduğu görülmüştür. 1987/88 ile 1997/98 yılları arasında yağlardan sağlanan enerjinin düştüğü görülmüştür. (Hulshof vd., 2003)

McGee (2003); yapmış olduğu çalışmasında, Amerikan zeytinyağı tüketicisinin profilini belirlemiştir. Çalışmaya göre bu kişiler; başlıca alışveriş yapan, 25-54 yaşlar arasında, yüksek maaş (40.000\$ ve üzeri), eğitilmiş (Lise ve üzeri), şehir dışında lüks semtlerde yaşayan ve sağlık ilgisi olan bireylerdir. (McGee, 2003)

Lazaridis (2004); yapmış olduğu çalışmasında Yunanistan'da Hanehalkı Bütçe Anketi verileri ile LA/AIDS modelini kullanarak katı ve sıvı yağların üç tipi için (zeytinyağı, bitkisel yağ ve margarin) talebi etkileyen etmenleri belirlemiştir. Çalışmada kullanılan veriler Yunanistan Ulusal İstatistik Servisi tarafından 1993/1994 yılında 6.756 haneden derlenen aile bütçe anketine ait verilerdir. Çalışmada kullanılan değişkenler; gıdaya harcanan para, aile büyüklüğü, zeytinyağının fiyatı, bitkisel yağın fiyatı, katı yağların fiyatı, hanehalkı reisinin yaşı, cinsiyeti ve eğitim seviyesi, hane dışında gıdaya ayrılan yüzde, belirli yaş grubuna giren birey sayısı, kır-kent durumu, hanenin bulunduğu bölge ve araştırmanın hangi zamanda (ay) yapıldığıdır. Çalışmanın sonuçlarına göre zeytinyağı tüketimi ile gıda harcama tutarı ve hanehalkı reisinin yaşı arasında pozitif ilişki olduğu, aile büyüklüğü, eğitim seviyesi, hane dışında yapılan toplam gıda harcamalarının yüzdesi arasında negatif ilişki olduğu tespit edilmiştir. Ayrıca zeytinyağı tüketimini belirleyen ana faktörün gelenek (alışkanlıklar) olduğu görülmüştür. (Lazaridis, 2004)

Januszewska ve Viaene (2005); yapmış oldukları çalışmalarında üç hedef üzerinde durulmuştur. Birinci hedef, gelirle ilişkili sosyal gruplar arasında satın alınan gıda miktarının farklı olup olmadığıdır. İkinci hedef, gelir ile bütçede gıda harcamasına ayrılan pay arasındaki

ilişkiyi analiz etmektir. Üçüncü hedef ise, gelir grupları arasındaki harcamalar ve gıda harcamalarında sosyo-ekonomik değişkenlerin etkisini araştırmaktır. Çalışmada kullanılan veriler 2001 yılında Belçika’da yapılan Gesellschaft für Konsum und Absatzforschung (GfK)’dan elde edilen panel verilerdir. Toplam 507 hanehalkından derlenen veriler analiz edilmiştir. Kişi başına düşen gelir, aile büyüklüğü, çocuk sayısı, yaşam bölgesi, eğitim, meslek ve yaş grupları gibi sosyo-ekonomik hanehalkı karakteristikleri analizde kullanılmıştır. Otuz Dokuz gıda ürünü yedi kategoride gruplandırılmıştır. Analiz prosedürü üç aşamadan oluşmuştur. Birincisinde, ANOVA test yöntemi kullanılmıştır ve yedi gıda kategorisi için Üç farklı gelir grupları dikkate alınarak kişi başına satın alınan gıda miktarı ve kişi başına gıda harcamaları hesaplanmıştır. İkincisinde, kişi başına satın alınan gıda miktarı ve kişi başına gıda harcamaları gelir gruplarına göre sosyo-ekonomik değişkenler kullanılarak Regresyon Analizi ile hesaplanmıştır. Burada açıklayıcı değişkenler olarak gelir, bölge, eğitim, meslek ve yaş gibi sosyo-ekonomik değişkenler kullanılmıştır. Üçüncü durumda ise, gelir grupları ile bölge, eğitim, meslek ve yaş gibi değişkenler arasında ilişkiler Ki-Kare test yöntemi ile analiz edilmiştir. Üç gelir grubu arasında farklı besin grubu satın almada anlamlı fark yoktur. Tek önemli fark yüksek gelir grubu tarafından satın alınan balık miktarı büyüktür. Düşük gelir grubunda balık, mandıra (süt) ürünleri ve sebzeler için daha az para harcamaktadır. Yaş karakteristiği, satın alınan gıda miktarı üzerinde pozitif önemli etkiye sahiptir tek değişkendir. Çapraz tablolama analizi sonucularına göre eğitim, meslek ve gelir grupları arasında anlamlı ilişkiler olduğunu göstermektedir (Januszewska ve Viaene, 2005).

Sayılı vd. (2005); yapmış oldukları çalışmalarında Tokat ilinin merkez ilçesinde hanehalkı yemeklik yağ ve yağ tüketimi ile gelir-gider esnekliklerinin miktarını belirlemişlerdir. Veriler Tokat ili merkez ilçesinde hanelerdeki tüketicilerle yapılan anket ile toplanmıştır. Analizde 206 gözlemden oluşan bir veri seti kullanmışlardır. Çalışmada kullanılan yöntemler; Çift Logaritmik Form, Yarı Logaritmik Form ve Ters Logaritmik Form kullanılmıştır. Aile büyüklüğü, cinsiyet, yaş, eğitim seviyesi, meslek ve ikamet yeri, tüketilen katı ve sıvı yağ çeşitleri ve miktarları gibi bilgiler derlenerek analize dâhil edilmiştir. Gelir 4 gruba ayrılmıştır. Ortalama aylık hane geliri 787,70\$, aile büyüklüğü 4,14 kişi, ortalama yaş 40,31 olarak bulunmuştur. Sayılı vd., gıda talebi çalışmalarında eskiden tüketici tercihlerini belirleyen en önemli faktörün gelir olduğunu, günümüzde ise yaş, cinsiyet, aile büyüklüğü, eğitim düzeyi ve meslek grubu olduğunu belirtmişlerdir. En çok tercih edilen yağın sırayla tereyağı, zeytinyağı, mısır yağı, ayçiçeği yağı, margarin, fındık yağı, karışık bitkisel yağ ve soya yağı olduğunu belirtmişlerdir. Hane başına ortalama tüketim 5,62 kg iken kişi başına düşen tüketim 1,36 kg olarak bulunmuştur. Yağlar (ayçiçek yağı, mısırozü yağı, fındık yağı,

karışık bitkisel yağ ve soya yağı), toplam yenilebilir sıvı ve katı yağ tüketiminin çoğunluğunu oluşturmuştur. Bu tüketim 3,41 kg/ay olarak görülmüştür. Hanehalkı başına aylık katı ve sıvı yağ tüketiminin 0,98 kg tereyağı, 0,40 kg margarin, 0,83 kg zeytinyağı, 1,57 kg ayçiçeği yağı, 1,30 kg mısır yağı, 0,46 kg fındık yağı, 0,04 kg karışık bitkisel yağ ve 0,02 kg soya fasulyesi yağı olarak tespit etmişlerdir. Hanede yemeklik yağ ve yağ tüketime harcanan paranın 15,38\$ (Amerikan Doları) ve gelirin %1,95'ine karşılık geldiğini bulmuşlardır. En çok tercih edilen yağlar sırasıyla, tereyağı, zeytinyağı, mısıryağı, ayçiçek yağı, margarin, fındık yağı, karışık bitkisel yağ ve soya yağıdır. Hanenin gelir düzeyi arttıkça beklendiği gibi, tereyağı ve zeytinyağı tüketiminin artmakta olduğunu fakat diğer yağların tüketiminin azalmakta olduğunu göstermişlerdir. Yüksek gelire sahip hanelerin daha doğal ve sağlıklı yağları tüketmeyi tercih eden haneler olduğu anlamına geldiği belirtilmiştir. (Sayılı vd., 2005)

Tiryaki ve Akbay (2005); yapmış oldukları çalışmada sosyolekonomik gruplar itibariyle Türkiye'deki ailelerin zeytinyağı tüketim yapısını analiz etmişlerdir. Çalışmanın verileri Devlet İstatistik Enstitüsü tarafından 2003 yılında 12 ay süreyle derlenen, 25.764 aile ile yapılmış olan Hanehalkı Bütçe Anketi verilerinden elde edilmiştir. Verilerin analizinde Çapraz Tablolardan yararlanılmış ve F Testi kullanılmıştır. Çalışmanın sonucuna göre yüksek gelirli ailelerin daha fazla miktarda zeytinyağı tükettikleri ve zeytinyağı tüketimi amacıyla daha fazla miktarda harcama yaptıklarını, hem gelir grupları hem yerleşim yerleri hem de bölgeler itibariyle aylık ortalama hane halkı zeytinyağı tüketim miktarları ve harcama değerleri istatistiksel olarak gruplar arasında farklılık olduğu gibi kentsel ve kırsal yerleşim yerleri itibariyle de bölgeler arasında önemli farklılıklar olduğunu, hane halkı genişliği dikkate alınarak hesaplanan kişi başına yıllık zeytinyağı tüketim miktarının bölgeler itibariyle farklılık gösterdiğini göstermişlerdir. (Tiryaki ve Akbay, 2005)

Külekçi vd. (2006); yapmış oldukları “Dondurulmuş Gıda Tüketimini Etkileyen Sosyo-Ekonomik Özelliklerin Belirlenmesi; Erzurum İli Örneği” isimli çalışmalarında Erzurum ilindeki dondurulmuş gıda tüketimi ile sosyo-ekonomik özellikler arasındaki ilişkiler tespit etmişlerdir. Erzurum ili merkez ilçede yüksek, orta ve düşük gelirli ailelerin ikamet ettiği üç alt belediye sınırlarında 336 aile ile anketler yapılmıştır. Verilerin analizinde aritmetik ortalama, sıralama, gruplama, oran ve çaprazlama analiz yöntemleri kullanılmıştır. Çalışmanın sonucuna göre dondurulmuş gıda tüketiminin düşük olması nedenlerinin başında gelir faktörünün önemli rol oynadığı belirlenmiştir. Gelir düzeyinin düşüşü ile birlikte dondurulmuş gıda tüketiminin azaldığı görülmüştür. Eğitim seviyesi yüksek kişilerin, düşük veya eğitimsiz insanlara göre daha fazla dondurulmuş gıda tükettikleri görülürken, ailedeki

birey sayısı ile dondurulmuş gıda tüketim oranı arasında ters ilişki olduğu görülmüştür (Külekçi vd., 2006).

Ricciuto vd. (2006); yapmış oldukları çalışmalarında Kanada hanehalkları arasında sosyodemografik faktörler ile gıda seçimi arasındaki ilişkileri incelemek istemişlerdir. Çalışmanın verileri, Kanada İstatistik Dairesi tarafından 1996 yılında yapılan Hanehalkı Gıda Harcamaları Anketi'nden elde edilmiştir. Çalışmanın analizinde toplam 9.969 hanehalkından derlenen veriler kullanılmıştır. Çalışmada kullanılan değişkenler; sosyodemografik karakteristikler, satın alınan gıda miktarı ve tipleri, mağazalardan satın alınan içecekler, ödenen fiyat ve gıdanın satın alındığı mağaza tipidir (süper market, market, özel mağaza vb). Satın alınan gıdalar; tahıl ürünleri, sebzeler ve meyveler, süt ürünleri, et ve alternatifler (balık, kümes hayvanları, yumurta, fındık, fıstık ezmesi, bakliyat ürünleri vb), diğer gıdalardır (yağlar (tereyağı, margarin, pişirme ve salata yağları), şekerler, tatlılar, aperatifler, alkolsüz içecekler vb). Çalışmada tanımlayıcı istatistikler uygulanmıştır, ayrıca hanehalkı sosyodemografik özellikler ve gıda satın alımları arasındaki ilişkileri tahmin etmek için Çoklu Regresyon Analizi kullanılmıştır. Sosyodemografik karakterler hanehalkı büyüklüğü, hanehalkı kompozisyonu, kişi başına düşen gelir ve hanehalkı reisinin eğitim durumudur. Çalışmanın sonucuna göre hanehalkı büyüklüğü, hanehalkı kompozisyonu, gelir ve eğitim karakteristikleri birlikte, gıda satın alma varyasyonunun büyük bir kısmını (%21-29) açıklamıştır. İstatistiksel olarak en anlamlı değişkenlerin hanehalkı büyüklüğü ve gelir (kişi başına) olduğu görülmüştür. Hanehalkı büyüklüğü ile satın alınan tüm gıda gruplarının ilişkili olduğu görülmüştür. Aynı sayıda bireyin yaşadığı haneler dikkate alındığında, 65 yaş ve üzeri yaşlı yetişkinlerin, diğer gıda ürünleri grubu hariç diğer tüm gıda ürünleri grubunda (tahıl ürünleri, sebzeler ve meyveler, süt ürünleri, et ve alternatifler (balık, kümes hayvanları, yumurta, fındık, fıstık ezmesi, bakliyat ürünleri vb)) genç yetişkinlere göre önemli derecede daha fazla satın alım yaptıkları görülmüştür. Diğer gıda ürünlerinde (yağlar (tereyağı, margarin, pişirme ve salata yağları), şekerler, tatlılar, aperatifler, alkolsüz içecekler vb) ise genç yetişkinler, yaşlı yetişkinlere göre daha fazla satın alım yaptıkları görülmüştür. 65 yaş ve üzeri yaşlı yetişkinlerde sebze ve meyveye harcanan gelir payının diğer gıda gruplarına göre daha büyük olduğu görülmüştür. Bu tip hanelerde yüksek yağlı et satın alımının az olduğu görülmüştür. Evde 15 yaşından küçük bireylerin olduğu hanelerde süt ürünleri grubu hariç tüm gıda ürünlerinde satın alınan miktarların az olduğu görülmüştür. Yüksek gelirin, tüm gıda gruplarının satın alımı ile ilişkili olduğu görülmüştür. Gelir seviyesinin belli bir değerin altında olması durumunda sıvı süt alımının düştüğü görülmüştür, bunun bir eşik seviyesi olduğu söylenebilir. Besin grupları içerisinde, süt ve süt ürünleri ile et ve alternatif ürünler

grubu hariç diğer tüm gıda gruplarında satın alma kalıplarının benzer olduğu görülmüştür. Yüksek gelir seviyelerinde diğer gıda ürünlerini satın almada hafif bir düşüş eğilimi görülmüştür. Gelir belirli bir seviyenin üzerinde ise yağlı süt tercih edilirken, düşük gelirden az yağlı sütün tercih edildiği görülmüştür. Et ve alternatiflerin olduğu gıda grubunda az yağlı et ve kümes hayvanları, balık satın alımlarında yüksek gelirle birlikte az oranda yükseliş olduğu görülmüştür. Hanehalkı geliri, hanehalkı kompozisyonu ve gelir ne olursa olsun, eğitim değişkeninin sebze, meyve ve süt ürünlerinin daha fazla miktarda satın alımı ile ilişkili olduğu, et ve alternatifleri ile diğer gıdaların daha az miktarda satın alımı ile ilişkili olduğu görülmüştür. Eğitim düzeyinin artmasıyla et ve alternatif ürünlerinin satın alımlarının azaldığı görülmüştür. Analizin sonucunda, hanehalkı sosyodemografik özelliklerinin gıda satın alma kalıpları üzerinde güçlü bir etkiye sahip olduğu görülmüştür. Yaş, gelir ve eğitim seviyesinin, hanehalkı büyüklüğü ve kompozisyonunun gıda seçiminin ve satın alımlarının önemli belirleyicileri olarak görülmüştür. Yaşlı bireylerin, genç bireylere göre sağlık endişesiyle yağ, tuz ve şeker tüketimini sınırladıkları ve tüketmekten kaçındıkları görülmüştür. (Ricciuto vd., 2006)

Akbağ (2007); yapmış olduğu çalışmada, 2003 yılında (12 ay'a ait), kentsel alanlarda 18.278 haneden TÜİK tarafından derlenen hanehalkı bütçe anketi verilerini kullanarak Türkiye'deki hanelerin yemeklik yağ ve yağ tüketim modellerini incelemiştir. AIDS modeli kullanılarak gerçekleştirilen çalışmada beş ana yağ türü için talep parametreleri ve esneklikleri analiz edilmiştir. 2003 yılında katı ve sıvı yağlara ayrılan bütçenin toplam gıda harcamasının %6,1'ini oluşturduğunu ve aylık hanehalkı zeytinyağı tüketiminin 424 gr, bitkisel yağın 4.037 gr, katı yağın ise 1.372 gr olduğu belirtmiştir. Açıklayıcı değişkenler olarak; hanehalkı geliri, hanehalkı reisinin eğitim, yaş ve medeni durumu, eşinin çalışıp çalışmama durumu ve bölge karakteristikleri kullanılmıştır. Çalışmada "LA/AIDS" modeli kullanılmıştır. Model "Zellner Görünürde İlişkisiz Regresyon Yöntemi" kullanılarak tahmin edilmiştir. Hanehalkı geliri düşük, orta ve yüksek olarak üç gruba ayrılmıştır. Elde edilen sonuçlara göre ortalama olarak gıda harcamalarının toplam harcamalar içerisindeki payı %29,8, düşük gelir grubunda %38,7, orta gelir grubunda %31,7, yüksek gelir grubunda ise %21 olarak görülmüştür. Hanelerin %86,3'ünün sıvı ve katı yağ tükettiğini, %13,7'sinin ise hiç yağ tüketmediğini belirtmiştir. Gelir arttıkça ayçiçeği ve margarin tüketimi azalırken, zeytinyağı ve tereyağı tüketiminin arttığı görülmüştür. Düşük gelirli grupta ayçiçek yağı %61,71, mısır yağı %5,79, zeytinyağı %3,86, fındık yağı %1,65, diğer bitkisel yağlar %1,95, margarin %21,76, tereyağı %3,03 olarak, yüksek gelirli grupta ayçiçek yağı %46,45, mısır yağı %1,07, zeytinyağı %15,04, fındık yağı %2,48, diğer bitkisel yağlar %0,99, margarin

%17,19, tereyağı %5,29 olarak görülmüştür. Yüksek gelirli grupta zeytinyağının daha çok tüketilmesinin nedeninin bu grupta yer alan tüketicilerin sağlık (yağ ve kolesterol için) endişelerinin olduğunu gösterebilir. Hanehalkı büyüklüğü istatistiksel olarak anlamlı görülürken, zeytinyağı ve mısır yağı ile negatif, margarin ve ayçiçek yağı ile pozitif ilişkilidir. İstatistiksel olarak anlamlı çıkan bir diğer karakteristiğin eğitim olduğu görülmüştür, mısır yağı ve zeytinyağı için pozitif, margarin ve ayçiçek yağı için negatif etki gösterdiği belirtilmiştir. Hanehalkı reisinin cinsiyeti yemeklik yağ talebinde istatistiksel olarak anlamlı çıkan bir diğer karakteristiktir. Yüksek gelirli ve yüksek eğitilmiş tüketicilerin, alt-gelir ve daha az eğitilmiş tüketicilere göre zeytinyağı gibi daha kaliteli ve sağlıklı bir yağ tükettiği görülmüştür. (Akbaş, 2007)

Beyaz (2007); yapmış olduğu yüksek lisans tezinde talep teorisi, teorinin kısıtlarıyla tutarlı tam talep modellerinin (sistemleri) evrimi ve bu modellerin ampirik çalışmalarda kullanılmasını incelemiştir. Çalışmada TÜİK “2003 Bütçe Anketi” verileri ile ampirik bir analiz yapılmıştır. Çalışmanın içeriğinde, TÜİK 2003 Bütçe Anketi verileri ile ana harcama grupları ve gıda-alt grupları için yapılan talep çalışmasında Working-Leser fonksiyonu kullanılmıştır. Çalışmada kullanılan açıklayıcı değişkenler; hane genişliği ve lokasyon, hanehalkı reisinin eğitim düzeyi hane nüfusunun yaş kompozisyonu lokasyon (kır ve kent) ve cinsiyet gibi sosyo-demografik değişkenlerdir. Hanehalkı nüfusu içinde 18 altı yaş gruplarının oranları toplamı gıda-alkolsüz içecek harcamalarını pozitif etkilemektedir. Hanehalkı nüfusu içinde gelir getiren fert sayısı arttıkça konut harcama payları azalış gösterirken, lokanta ve restoran harcama payları artmaktadır. Hane genişlikleri konut harcama paylarını negatif etkilerken, mobilya ve ev bakımı harcamalarını pozitif etkilemektedir. Hanehalkı nüfusu içinde 18 ve altı yaş gruplarının payları toplamı et, balık, meyve ve sebze harcama paylarını azaltırken şekerleme, hazır gıda ve çay-kahve harcama paylarını artırmaktadır. Hanehalkı nüfusu içinde 0-6 ve 14- 18 yaş grupları diğer alkolsüz içecek harcamalarını negatif etkilemektedir. Yetişkin nüfus (18 ve üstü yaş grupları) et, balık, meyve harcamalarını pozitif etkilemektedir. Yağ harcamalarını hane nüfusu içinde 14-18 yaş grubu negatif etkilerken, 7-13 yaş grubu yağ harcamalarını pozitif etkilemektedir. Elde edilen sonuçlara göre; gıda-alkolsüz içecekler ve gıda alt grupları için yaş, eğitim düzeylerinin, hanehalkı reisinin cinsiyetinin ve yerleşim yerlerinin önemli bir açıklayıcı değişken olduğu ortaya çıkmıştır. Model sonuçlarına göre ilköğretim ve altı eğitim düzeyleri gıda-alkolsüz içecek harcama paylarını pozitif yönde etkilerken lise ve üstü eğitim düzeyleri ise negatif etkilemektedir. Hanehalkı reisi erkek olan hanelerde ekmek ve tahıl, et ve çikolata ve benzeri ürünlerin bütçe payı hanehalkı reisi bayan olanlardan daha yüksektir. Süt ve süt ürünleri harcama payı 7-13,

14-18 ve 18 üzeri yaş gruplarından negatif etkilenirken eğitim seviyeleri yükseldikçe pozitif etkilenmektedir. Eğitim seviyesi düşük olan kesimden yüksek eğitilmiş kesime doğru süt ve süt ürünlerine ayrılan bütçe payı artış göstermektedir. Kırsal kesimde yağ, sebze, çikolata ve benzeri ürünler, çay, süt ve süt ürünleri harcamaları kente göre daha yüksektir. Lise ve üstü eğitim düzeyleri meyve harcamalarını pozitif etkilerken sebze harcamalarını negatif etkilemektedir. Türkiye genelinde olduğu gibi Türkiye-kent için gıda-alkolsüz içecek harcamalarında yaş grupları ve lise üstü eğitim düzeyleri negatif yönde etki gösterirken ilköğretim ve altı eğitim düzeylerinde pozitif yönde etki göstermektedir. Düşük eğitim düzeyi yağ ve sebze harcamalarını artırırken yüksek eğitim düzeyleri süt ve süt ürünleri harcamalarını ve meyve harcamalarını artırmaktadır. Çay ve kahve harcamalarını 0-6, 7-13, 14-18 yaş grupları negatif etkilerken diğer alkolsüz içecekler harcamasını pozitif etkilemektedir. Hanehalkı reisi bayan olan hanelerde yumurta, süt ve süt ürünleri, sebze, çay-kahve harcamaları bütçe payı hanehalkı reisi erkek olan hanelerden daha yüksektir. Türkiye-kent geneli için gıda-alkolsüz içeceklere ayrılan bütçe payı %30'dur. Genel olarak gıda-alkolsüz içecekler içinde gıda alt grupları harcamalarında en yüksek bütçe payı ekmek ve tahıllara (%25) ayrılırken en düşük bütçe payını ise yumurtaya (%3) ayrılmıştır. (Beyaz, 2007)

Bittencourt vd. (2007); yapmış oldukları çalışmalarında Japonya'da gıda tüketim kalıplarının belirleyicilerini incelemişlerdir. Çalışmada AIDS modeline dayalı bir kesitsel çalışma ile tahmin yapılmıştır. Japonya'da yapılan hanehalkı anketine ait 1.281 gözlemden derlenen veriler çalışmanın analizinde kullanılmıştır. Çalışmanın sonucuna göre ekonomik ve ekonomik olmayan her faktörün yaşam boyunca gıda tüketimi üzerinde farklı etkilerinin olduğu görülmüştür. Bazı gıda gruplarının tüketimindeki değişikliklerin fiyat ve gelirin etkileri ile açıklanabilirken, diğerleri demografik özellikler ile açıklanabileceği sonucuna ulaşılmıştır. Mali kısıtlamanın bağlayıcı olmadığı, yaşamlarının farklı dönemlerinde tüketicilerin gıda seçeneklerini tahmin etmede yerleşim yerinin az veya hiç etkisinin olmadığı sonucuna ulaşılmıştır. Tüketim modelini etkileyen diğer önemli faktörlerin aile büyüklüğü, çocuk sayısı, yaşam tarzı ve sağlık kaygısının olması sayılabilir. Farklı yaş grupları için farklı et tüketim kalıplarının olduğu belirtilmiştir. 41'den az yaşı bulunan genç bireyler ile 41 ile 60 yaş arasındaki bireylerin süt tüketim kalıpları birbirine benzemektedir. Süt tüketiminde anlamlı fark emeklilik yaşı (sonra 60 yaşında) öncesi ve sonrasındaki insanlar arasında görülmektedir. Yaşlı insanlarda yumurta tüketme eğilimi gençlere göre daha azdır. Çocuk sayısının balık, et ve süt tüketiminde önemli etkilerinin olduğu görülmüştür. Yerleşim yerinin farklı yaş gruplarında balık, et, yumurta ve sebze tüketimini etkilediği gözlenmiştir. Yaşlı

bireylerin sağlık endişelerinden dolayı daha seçici oldukları ve diyet ürünlerini tercih ettikleri görülmüş, buna göre daha az et ürünleri yerken, balık, sebze ve meyve gibi yüksek vitamin ve mineral içeren, az kolesterolü daha sağlıklı gıdaları tükettikleri görülmüştür. Hanehalkı olgunlaştığında gıda tüketim kalıpları değişmektedir, gıda tüketim kalıpları ekonomik, demografik, yaşam tarzı ve sağlık gibi diğer faktörlerden etkilenmektedir. (Bittencourt vd., 2007)

Bopape ve Myers (2007); yapmış oldukları çalışmalarında Güney Afrika'da kırsal ve kentsel haneler arasında talep davranışları ve gelir grupları farklılıklarını dikkate alarak gıda harcaması desenlerini analiz etmiştir. Çalışmada kullanılan veriler KwaZulu-Natal Gelir Dinamikleri Çalışmasından elde edilmiştir. Üç anket içeren panel veri 1993, 1998 ve 2004 yıllarına aittir. Toplamda 2.181 gözlem analizde kullanılmıştır. Çalışmada QUAIDS modeli kullanılmış ve Maksimum Olabilirlik (Maximum Likelihood) yöntemi kullanılarak tahmin edilmiştir. Modelde; gruplara ayrılmış gelir, kır-kent durumu, gıda grupları, hanehalkı toplam gıda harcamaları, hanehalkı toplam geliri, hanehalkı büyüklüğü, hanehalkı reisinin eğitimi, yaşı gibi karakteristikler kullanılmıştır. Çalışmanın sonucuna göre balık ve et tüm hanehalkı gruplarında lüks bir gıda ürünü olarak görülmüştür. Kırsal ve düşük gelirli hanehalklarında et ve balık için yapılan harcamalar kentsel ve yüksek gelirli hane halklarına göre daha esnektir (Bopape ve Myers, 2007).

Colón-Ramos vd. (2007); yapmış oldukları çalışmalarında Kosta Rika'da 1994-2004 yılları içerisinde yemek pişirmede kullanılan yağın seçiminde sosyoekonomik ve yaşam standardı belirleyicilerini incelemişlerdir. 2.247 kayıt ile yapılan bu araştırmada, yemek pişirmede kullanılan yağ tipi, sosyoekonomik ve demografik veriler toplanmış ve kullanılmıştır. Verilerin analizinde Student-T testi, Ki-Kare ve Lojistik Regresyon yöntemleri kullanılmıştır. Soya yağı hem kırsal hem de kentsel alanda yaygın olarak kullanılan bir yağ çeşidi olarak görülmüştür. Kırsal kesimde yaşayan ailelerin %49'u soya yağını kullanırken ikinci en çok kullanılan yağın %28 ile palm yağı olarak görülmüştür. Kentte yaşayan ailelerin %54'ü Soya yağı, %20'si palm yağı, %26'sı ise diğer doymamış yağları (mısır, ayçiçeği ve zeytinyağı) kullandığı sonucuna ulaşılmıştır. Doymamış yağlar içinde, popülasyonun %20'sinin mısır ve ayçiçeği yağını, %2,4'ünün zeytinyağını, %1'inin ise diğer yağları tükettiği görülmüştür. Sonuçlara göre palm yağı kullanıcılarının daha düşük sosyoekonomik statüye, eğitime ve gelire sahip oldukları görülmüştür, soya yağı veya diğer doymamış yağ kullanıcılarına göre işsiz, tarım işçisi ya da yarı vasıflı bir işe sahip olmalarının yüksek olasılıklı olduğu sonucuna varılmıştır. Soya yağı tüketenlerin orta ile düşük sosyoekonomik statüye sahip oldukları, eğitim seviyesinin, gelir düzeyinin ve hanedeki ev eşyası sayısının

palm yağı kullanıcılarına göre yüksek, diğer doymamış yağ tüketicilerine göre düşük olduğu görülmüştür. Diğer doymamış yağ tüketicilerinin daha yüksek eğitim düzeyine ve gelire sahip oldukları ve daha çok ev eşyasına sahip oldukları görülmüştür. Zaman içerisinde palm yağı tüketimi düşerken, soya yağı ve diğer doymamış yağ tüketimi artmıştır. Yüksek sosyoekonomik statü, sağlık bilinci soya yağı ve diğer doymamış yağ seçiminde etkili olan belirleyiciler olarak görülmüştür (Colòn-Ramos vd., 2007).

Aktaş (2008); yapmış olduğu doktora tezinde, Batı Akdeniz Bölgesi kent merkezlerinde 2006 yılına ilişkin hanehalklarının aylık et talebini (dana-sığır, koyun-keçi, tavuk ve balık) gelir, ilgili ürünün fiyatı ve sosyodemografik özellikleri de dikkate alarak tahmin etmek, ilgili ürünlerin fiyat-talep, gelir-talep esnekliklerinin hesaplamak ve sosyodemografik özelliklerine ilişkin değişkenlerin talep üzerine olan etkilerinin belirlenmesi amaçlanmıştır. Çalışmada Doğrusal Formda Yaklaşık İdeal Talep Sistem (LA/AIDS) modeli kullanılmıştır. Araştırma bölgesi için incelenen ürünlere ilişkin harcama, Marshalcı ve Hicksci esneklikleri hesaplanmıştır. Tahmin edilen modele fiyatlar, hanehalkı büyüklüğü, hanehalkı bireylerin ikamet ettiği bölge, hanehalkı reisinin öğrenim düzeyi, hanehalkı reisinin yaşı gibi demografik değişkenler ilave edilmiş ve talep teorisinin kısıtları altında tahminler yapılmıştır. Anket sonuçlarına göre Batı Akdeniz bölgesi kentsel alanda, ortalama hanehalkı genişliğinin 3.58 kişi olduğu belirlenmiştir. Araştırma bölgesinde gelir grupları arttıkça hanehalkı birey sayısı ve hanehalkından çalışan birey sayısı artmaktadır. Araştırmada, kullanılabilir gelire göre oluşturulan gelir grupları esas alınarak değerlendirmeler yapılmıştır. Araştırma bölgesindeki hanehalklarının ortalama gelirlerinin 2 049,612 TL/Ay olduğu görülmüştür. Hanehalklarının gelirleri arttıkça, gelirlerinden gıda tüketimi ve kira harcamaları için ayırmış oldukları oranın azaldığı, giyim harcamaları, eğitim harcamaları, ev dışı gıda harcamaları ve alkollü içecek ve sigara-tütün harcamaları için ise arttığı görülmektedir. Gelir grupları itibarıyla toplam et, gıda ve toplam harcamaların oransal dağılımı incelendiğinden gelir grupları arttıkça hanehalklarının toplam gıda harcaması içerisinde et harcamasına ayırdıkları oran düşmektedir. İncelenen hanehalklarının toplam gıda harcamaları ve toplam harcama miktarı gelir grupları arttıkça artmaktadır. Toplam gıda harcaması ve toplam harcamanın, toplam gelir içerisindeki dağılımında ise gelir grupları arttıkça toplam gelir içerisindeki oranı azalmaktadır. Araştırma sonucunda hanehalklarının gelir seviyeleri ile dana-sığır eti tüketim miktarı ve harcamaları arasında doğrusal bir ilişki olduğu tespit edilmiştir. Başka bir deyişle hanehalklarının gelirleri yükseldikçe dana-sığır tüketim miktarı ve harcamaları da artmaktadır. Nitekim 1. gelir grubunda 1,092 kg/ay olan dana-sığır eti tüketim miktarı 5. gelir grubunda ise 1,839 kg/ay'a yükselmiştir. Tavuk ve balık tüketim

miktarında ise gelir arttıkça azalma olduğu yani gelir arttıkça tüketim miktarının azaldığı tespit edilmiştir. Koyun-keçi etinde ise gelir gruplarına göre tüketim miktarının farklılık gösterdiği görülmüştür (Aktaş, 2008).

Kim (2008); yapmış olduğu çalışmasında ABD margarin sanayisinde talep ve fiyatlandırma ilişkisini tahmin etmiştir. Çalışmada Karma Logit Modeli kullanılmıştır. Talep fonksiyonunun tahmini Genelleştirilmiş Momentler Yöntemiyle yapılmıştır. Çalışmada kullanılan veriler IRI'den elde edilmiştir. (Information Resources Incorporated) Veriler, Aralık 1999 ile Kasım 2003 arasında ABD'nin 32 şehrinden derlenmiştir. Çalışmanın sonuçlarına göre, margarinin fiyat ve kalori bilgisi tüketici tercihlerini olumsuz etkilemiştir. Tereyağı katkılı margarin, margarinin tadını iyileştirdiği için pozitif etkisi olduğu görülmüştür. Yağ ve gelir etkileşimi düşünüldüğünde tüketicilerin gelirleri arttıkça sağlık konusunda daha bilinçli hale geldikleri görülmüştür. Etkileşim negatiftir ve anlamlıdır. Margarin talebinin belirleyicileri: paket boyutu, ürün formlarına besin gerçeklerinin eklenmesi (kalori değeri, içerdiği yağların belirtilmesi gibi), hanehalkı geliri ve yaş kompozisyonu olarak görülmüştür. Özellikle yüksek gelirli hanelerin diyet yağ seviyelerine karşı daha duyarlı olduğu görülmüştür. (Kim, 2008)

Pan vd. (2008); yapmış oldukları çalışmalarında, Hindistan'da yemeklik yağ için hanehalkı harcamalarını incelemiştir. Veriler, "National Institute of Extension Management" tarafından, tabakalı örnekleme tekniği ile Hindistan'ın kırsal ve kentsel kesiminden 2000-Ağustos/2001-Ağustos ayları arasından toplanan 1.192 gözlemden oluşmaktadır. Çalışmanın sonuçlarına göre 2004/2005 yıllarında Hindistan, Çin, Endonezya, ABD ve Avrupa Birliğinde kişi başı yağ tüketim değerleri, 11,2 kg, 15,8 kg, 16,3 kg, 29,6 kg, 18,8 kg olduğu belirtilmiştir. Palm yağı düşük fiyat, lojistik avantajları, tüketici kabulü gibi nedenlerden dolayı Hindistan bitkisel yağ ithalatının çoğunluğunu oluşturmaktadır. Ancak diğer bitkisel yağlar, büyüyen orta sınıfın giderek sağlıklı beslenme alışkanlıkları kazanmasından dolayı Hint pazarına yavaş yavaş girmektedir. Örneklenmiş her ev için ankete hanehalkı tüketim miktarı, çeşitli mallar üzerinde toplam harcama verileri ve demografik özellikleri dahil edilmiştir. Linquid Eksik Talep Sistemi ve Shonkwiler ve Yen yaklaşımı, hanehalkı araştırması verilerine dayanarak daha verimli talep analizini geliştirmek için kullanılmıştır. Hanehalkı reisinin yaşı, kırsal ya da kentsel konumda ikamet, eğitime devam edilen yılın sayısı, din (Müslüman ve diğerleri), cinsiyet, evlilik durumu, vejetaryen olup olmama durumu, aylık toplam harcama, aylık gelir, gıda harcaması, toplulaştırılmış gıda harcaması (aggregated), aylık yenilebilir yağ, fıstık yağı, sıvı tereyağı, kolza yağı ve palm yağı harcamaları çalışmaya dâhil edilen değişkenlerdir. Gıda alışkanlığı, konum, hanehalkı

reisinin eğitimi ve diğer demografik değişkenlerin yemeklik yağların seçimi üzerinde önemli etkilerinin olduğu görülmüştür. Ortalama olarak palm yağı fiyatının, fıstık ve kanola gibi diğer yağlara göre %20-30 daha ucuz olduğu belirtilmiştir. Hindistan da hane halkının kullandığı yağları fıstık yağı, sıvı tereyağ, kolza tohumu yağı, palm yağı ve diğer yağlar olarak belirlemişlerdir. Hindistan'da diğer yağlar için harcama esneklikleri nispeten düşük iken, yenilebilir fıstık yağı için büyük gelir ve harcama esneklikleri olduğunu göstermiştir. Sonuçlar yemeklik fıstık yağının bir lüks olduğunu göstermiştir. Yemeklik yağların seçiminde ev (hanehalkı) konumunu ve yemek alışkanlıkları değişkenlerinin en güçlü önemli etkileri olduğu gösterilmiştir. Kentte yaşayanların gelirleri, kırsalda yaşayanlara göre 2,26 kat daha büyük olduğu, kişi başına düşen gıda harcamasının kentsel haneler için kırsal gıda harcamalarına kıyasla %70 daha yüksek olduğu belirtilmiştir. Kentsel kişi başına düşen yemeklik yağ harcamaları kırsal yemeklik yağ harcamalarına göre %31 daha yüksek olduğu görülmüştür. Kentsel ve kırsal nüfus arasında yemeklik yağ tüketim alışkanlıklarında önemli farklılıklar görülmüştür. Kırsal kesimde kişi başına yenilebilir fıstık yağı tüketimi kentsel kesime göre %28 daha yüksek olduğu bulunmuştur. Kentsel kesimde kişi başına düşen sıvı yağ, kolza yağı ve palm yağı tüketimi ise sırasıyla %52, %54 ve %102 oranla kırsal kesimde kişi başına tüketimine göre daha yüksek bulunmuştur. Genel olarak, kentsel alanlarda kişi başına bitkisel yağ tüketimi 7,01 kg iken kırsal alandaki kişiler için 5,9 kg olarak bulunmuştur. 1.192 haneden sadece 6'sı (%0,5) dört tip yağı tüketirken, %75'den fazlası iki ya da daha fazla çeşit yağ kullanmışlardır. Kırsal alanlarda yaşayan hanelerin kentsel alanda yaşayan hanelere göre sıvı tereyağı tüketmesi daha az olası iken, palm yağı tüketmesi daha yüksek olasıdır. Gelir değişkeni sıvı tereyağı tüketiminde pozitif bir rol oynarken, kolza ve palm yağı tüketiminde negatif bir rol oynamaktadır. Din ve gıda alışkanlıkları, yemeklik yağ tüketiminin seçiminde önemli rol oynamaktadır. Vejetaryenlerin sıvı tereyağı ve kolza yağını, fıstık yağı ve palm yağına göre tüketme olasılıkları daha fazladır, bunun nedeni ise vejetaryenlerin protein alımı ile kısmen ilgili olduğu söylenebilir. Yapılan çalışmada tereyağı yağ tüketimi ile eğitim arasında önemli ve pozitif bir ilişki olduğunu görülmüştür. Ayrıca hanenin konumunun yemeklik yağ seçiminde önemli olduğu görülmüştür. (Pan, 2008)

Tiryaki (2008); yapmış olduğu çalışmada hanehalkının zeytinyağı tüketim tercihlerini belirleyen sosyoekonomik ve demografik farklılıkları incelemiştir. TÜİK tarafından 2003-2004 yılları arasında 12 ay boyunca 12 bölge 25.764 haneden derlenen veriler kullanılarak analiz yapılmıştır. Verilerin analizinde Ki Kare Analiz yöntemi kullanılmıştır. Çalışmanın bulgularına göre hanenin geliri, hanehalkı büyüklüğü, hanehalkı reisinin eğitimi, yaşı ve cinsiyeti, eşin çalışıp çalışmama durumu, yerleşim alanı (kır-kent) ve bölgesel

farklılıklar gibi hanehalkı demografik ve sosyoekonomik karakteristikleri istatistiksel olarak anlamlı ve Türk hanehalkları için zeytinyağı tüketim seçiminde önemli rol oynadığı görülmüştür. Kentte yaşayan, hanehalkı büyüklüğünün küçük olduğu, yüksek eğitim seviyesine sahip hanehalkı reisinin olduğu, hanehalkı reisinin eşinin çalıştığı, yüksek gelire sahip hanelerin diğer hanelere göre daha çok zeytinyağı tüketme olasılığının yüksek olduğu görülmüştür. Düşük gelire sahip hanelerin %3'ü zeytinyağı tüketirken yüksek gelirde bu oran %14,3 olarak görülmüştür. Benzer sonuçlar hanehalkı reisinin eğitim seviyesinde de görülmüştür. Hanedeki birey sayısının az olduğu hanelerin zeytinyağı tüketme olasılığı, birey sayısının fazla olduğu hanelere göre daha yüksektir. Kentsel alanlardaki hanelerin zeytinyağı tüketme olasılığı, kırsal alanlardaki hanelere göre daha fazladır. Yaşlı tüketicilerin zeytinyağı tüketim tercihlerinin genç tüketicilere göre daha yüksek olduğu, hanede bulunan yaşlı birey sayısının zeytinyağı seçiminde önemli bir faktör olduğu görülmüştür. Zeytinyağı tüketen hanelerin hiç zeytinyağı tüketmeyen hanelere göre daha yüksek gelir ve eğitim seviyesine sahip olduğu görülürken, zeytinyağı tüketiminde gelirin istatistiksel olarak önemli olduğu vurgulanmıştır. Gelir seviyesi ve zeytinyağı tüketimi arasında pozitif bir ilişki olduğu, benzer bir pozitif ilişkinin de eğitimde olduğu görülmüştür. Eğitim seviyesi üniversite düzeyine doğru yükseldikçe zeytinyağı tüketim yüzdesi artmaktadır. Hanehalkı reisinin erkek olduğu hanelerde zeytinyağı tüketme olasılığı kadın olan hanelere göre daha yüksektir. Hanehalkı reisinin cinsiyeti, medeni durumu zeytinyağı tüketim tercihinde istatistiksel olarak anlamlı bulunmamıştır. Zeytinyağı tüketim tercihlerini bölgelere göre baktığımızda Marmara Bölgesinde %8,1, Ege %11,85, Akdeniz %7,36, Karadeniz %3,22, İç Anadolu %8,72, Doğu Anadolu %5,37, Güney Doğu Anadolu Bölgesinde ise %5,97 olarak görülmüştür. (Tiryaki, 2008)

Mullie vd. (2009); yapmış oldukları çalışmalarında fonksiyonel gıda tüketimi ile ilişkili kültürel, sosyoekonomik ve beslenme belirleyicilerini tanımlamaya çalışmışlardır. Kesitsel çalışma (cross sectional) yöntemiyle 5.000 askere gönderilen anketle gıda tüketim sıklığı kaydedilmiştir. Çalışmanın analizinde betimleyici istatistikler ve Lojistik Regresyon Modeli kullanılmıştır. Yanıt verenlerin %26,3'ü günlük bazda, fitosteroller ve steroller ile güçlendirilmiş margarinleri kullanılmıştır. %14'ü bir yâda daha fazla porsiyon fındık tüketmiştir. %12,3'ü haftada bir kez yağlı balık tüketmiştir. Yaş, vücut kitle indeksi ve gelirin artması güçlendirilmiş margarinin günlük tüketimini artırdığı görülmüştür. Ayrıca evli olmanın ve Flaman kültüre sahip olmanın güçlendirilmiş margarin tüketimi ile ilişkisi olduğu görülmüştür. Artan yaş, evli olma ve kültürel arka plan yüksek margarin tüketimi ile ilişkili olduğu görülmüştür. Güçlendirilmiş margarinler, fındık, çay ve yağlı balık tüketimi, kültürel

birikimden güçlü bir şekilde etkilenmiştir. Flamanca konuşan erkeklerin güçlendirilmiş margarin tüketiminin Fransızca konuşan erkeklere göre daha yüksek olduğu görülmüştür. Fonksiyonel gıda tüketiminin, sağlıklı beslenme şekli ile ilişkili olduğu görülmüştür. (Mullie vd., 2009)

Begum vd. (2010); yapmış oldukları çalışmalarında Nowshera kent bölgesinde sosyoekonomik faktörlerin gıda tüketim kalıplarına etkisini incelemişlerdir. Analiz, önceden tasarlanmış bir anket ve rastgele seçilmiş 375 haneden derlenen veriler kullanılarak yapılmıştır. 2005 yılında Pakistan'da nüfusun %23,9'u yoksulluk sınırının altında yaşadığı, Pakistan nüfusunun %60'ının günlük kişi başı 2\$'ın (Dolar) altında bir para ile yaşamını sürdürmekte olduğunu belirtmişlerdir. Bu nedenle çalışma sosyoekonomik faktörlerin gıda tüketim kalıplarına etkisini incelemek üzere yapılmıştır. Veri seti, aile büyüklüğü, cinsiyet, yaş, okuma yazma oranı, eğitim düzeyi gibi sosyal özellikleri, gelir durumu (maaş ve diğer kaynaklar) gibi ekonomik özellikleri, buğday unu, süt, sebze, bakliyat, pirinç, yumurta, şeker, çay ve meyve gibi gıda maddelerinin tüketimi bilgilerini içermektedir. Çalışmada tanımlayıcı istatistikler ve Çoklu Regresyon Modeli kullanılırken, verilerin analizinde En Küçük Kareler Yöntemi kullanılmıştır. Hane büyüklüğü ve ortalama aylık toplam gelir tüketilen gıda maddelerinin çoğu üzerinde önemli etkisi olan sosyoekonomik faktörlerdir. Bu çalışmanın sonucuna göre gelir ile buğday unu, et, süt, meyve, sebze, çay ve yumurta tüketimi arasında önemli ve pozitif bir ilişki olmasına karşın, gelir ile şeker, yağ ve bakliyat ürünlerinin tüketimi arasında bir ilişki görülmemiştir. Hane büyüklüğü ile buğday unu, süt, pirinç, meyve, çay, yağ ve yumurta gibi gıda maddelerinin tüketimi arasındaki korelasyon anlamlı görülmüştür. (Begum vd., 2010)

Davis vd. (2010); yapmış oldukları çalışmalarında (kitap olarak yayınlanmış) 12 ayrı süt ve süt ürünleri (mandıra ürünleri) ile margarinin perakende alım verilerini incelemişlerdir. Çalışmanın verileri, Nielsen 2007 (2007 Nielsen Homescan) hane tarama verileridir. Nielsen verilerine seçilmiş sosyoekonomik ve sosyodemografik veriler eklenmiştir. Analizde kullanılan toplam 63.061 gözlem bulunmaktadır. Elde edilen veri seti ile seçilmiş olan ürünler için sosyoekonomik ve sosyodemografik değişkenlerin toplam talep ve harcama esneklikleri üzerindeki etkileri araştırılmıştır. Harcama esnekliklerini elde etmek için Sansürlü Talep Sistemi (Censored Demand System) kullanılmıştır. Seçilmiş gıda ürünleri dondurma, buzlu süt ve şerbet, soğutulmuş yoğurt (refrigerated yogurt), dondurulmuş yoğurt (frozen yogurt), içilebilir yoğurt (drinkable yogurt), tam yağlı süt, az yağlı süt, konserve süt, doğal peynir, işlenmiş peynir, çökelek (cottage cheese: ekşimik, çökelek, süzme peynir), tereyağı ve margarindir. Sosyoekonomik ve sosyodemografik veriler, hanehalkı büyüklüğü, Average age

of female head of household, altı yaş altındaki çocuk sahipliği, 13-17 yaş arası genç birey sahipliği, hanehalkının tek bireyden oluşup oluşmaması, bölge, ırk, gelir (kategorize edilmiş), üniversite eğitimi almış kadın hanehalkı reisi (Female head of household w/some college)'dir. Her gıda ürünü için, her hanenin ortalama olarak 12 ay boyunca tüketmiş olduğu ürün miktarları ve harcadıkları paralar detaylı bir şekilde belirtilmiştir. Ortalama olarak her hanenin 12 aylık tüketim miktarları verilerinden bazıları tereyağı 82,54 Oz veya Floz, margarin için 153,80 Oz veya Floz'dur. Sansörlü Talep Sisteminin analiz sonuçlarına göre hanehalkı büyüklüğü, şerbet ve buzlu süt, içilebilir yoğurt, tam yağlı süt, az yağlı süt, işlenmiş peynir ve margarinin satın alımlarında istatistiksel olarak anlamlı ve pozitif etkilidir. Kadın hanehalkı reisinin yaşı, dondurma, şerbet ve buzlu süt, dondurulmuş yoğurt, konserve süt, peynir ve tereyağı gibi süt ve süt ürünlerinin satın alımlarında istatistiksel olarak anlamlı ve pozitif etkilidir. Tek kişilik ve üniversite eğitilmiş kadın hanehalkı reisinin bulunduğu hanelerde soğutulmuş yoğurt, dondurulmuş yoğurt ve az yağlı süt satın alımı olumlu, istatistiksel olarak anlamlı bir etkiye sahiptir. Siyahî hane halkı reisleri dondurma, şerbet ve buzlu süt, soğutulmuş ve içilebilir yoğurt, yağlı süt, konserve süt ve tereyağının satın alımlarında istatistiksel olarak anlamlı ve pozitif etkilidir. Asyalı hane halkı reisleri dondurma, soğutulmuş ve içilebilir yoğurt ve tüm sıvı sütlerin satın alımlarında istatistiksel olarak anlamlı ve pozitif etkilidir. ABD'nin Doğu, Orta ve Batı kesimlerinde yaşayan hanelerde soğutulmuş yoğurt, doğal ve çökelek peyniri ve tereyağı üzerinde diğer süt ürünleri ve margarine göre daha büyük etkiye sahiptir. Buradan anlaşılabileceği üzere tereyağı, margarin, soğutulmuş yoğurt gibi süt/mandıra ürünlerinin tüketiminde yaşanan bölgenin (ülkenin kuzey, güney, doğu, batı ve orta kesimi gibi) etkisi görülmektedir. Gelir kategorisi incelendiğinde en düşük iki gelir grubundaki (19.999 \$ ve altı ile 20.000 \$ - 34.999 \$ arası) hanehalklarının tam yağlı süt ve az yağlı süt alımları istatistiksel olarak anlamlı ve pozitif etkilidir. 34.999 \$ ve altında gelire sahip hanehalklarının, şerbet ve sütlü buz, işlenmiş ve çökelek peyniri, tereyağı, yoğurt gibi süt/mandıra ürünleri ile margarin satın alımları istatistiksel olarak anlamlı ve pozitif etkilidir. (Davis vd., 2010)

Garcia ve Grande (2010); yapmış oldukları çalışmalarında farklı yıllardan toplanan verileri kullanarak ve literatürde bildirilen anahtar değişkenleri karşılaştırarak, İspanya'da yaşlı tüketicilerin gıda harcama kalıplarının belirleyicilerinden ampirik bir çalışma yapmışlardır. Çalışmanın verileri İspanya Ulusal İstatistik Enstitüsü tarafından 1998 ve 2004 yıllarında derlenen hanehalkı bütçe anketine ait verilerdir. Hanehalkı reisinin yaşının 65 ve üzerinde olduğu, emekli kişiler seçilmiştir, 1998 yılına ait 1.974, 2004 yılına ait 2.056 anket analizde kullanılmıştır. Modelde kullanılan açıklayıcı değişkenler, demografik ve coğrafi

değişkenler olmak üzere ikiye ayrılmıştır. Demografik değişkenler: cinsiyet, medeni durum, hanehalkı tipi, eğitim düzeyi ve yaş olarak belirlenmiştir. Coğrafi değişkenler; bölge, belediye tipi (nüfus büyüklüğüne göre kategorize edilmiştir), kır-kent durumu ve nüfus yoğunluğu olarak belirlenmiştir. Araştırmanın analizinde Genel Doğrusal Model, Varyan Analizi ve Ki Kare yöntemleri kullanılmıştır. Her iki yıl için yapılan analiz sonuçlarına göre hanehalkı tipi, gelir, yaş ve eğitim durumu gıda harcamalarını etkileyen karakteristikler olarak görülmüştür. (Garcia ve Grande, 2010)

Jacobson vd. (2010); yapmış oldukları çalışmalarında hanehalkı gıda harcamalarına etki eden değişkenleri belirlemek istemişlerdir. Çalışmada kullanılan veriler Kıbrıs Ulusal İstatistik Servisi tarafından 2004 yılında yapılan 625 anketten derlenmiştir. Anketteki yıllık kazanç verileri 2003'e aittir. Gıda harcamalarına etki edebilecek dört değişken; aile kazancı, hanehalkı büyüklüğü, annenin yaşı ve eğitimi olarak belirlenmiştir. Çalışmada Çok Değişkenli Regresyon Modeli kullanılmıştır. Regresyonun sonucuna göre ev gıda harcamalarına etki eden değişkenler; ailenin kazancı, hanehalkı büyüklüğü ve annenin yaşı olarak görülmüştür. Hanehalkı büyüklüğünün hane gıda harcamalarına etkisi, gelirin etkisinden daha büyüktür. Düşük gelire sahip ve emekli hanehalkı için yapılan regresyon analizinde, hanehalkı gıda harcamalarına gelirin etkisi hanehalkı büyüklüğünden daha önemli olduğu görülmüştür. (Jacobson vd., 2010)

Oktay (2010); yapmış olduğu doktora tez çalışmasında zeytinyağının iç tüketimine yönelik farklı bir bakış açısı getirmek ve tüketimi doğrudan ve dolaylı olarak etkileyen değişkenlerin belirlenmesini hedeflemiştir. Çalışmada TÜİK tarafından 2006 yılında 8.640 haneden derlenen Hanehalkı Bütçe Anketi verileri kullanılmıştır. Veriler, hanehalklarının gelir düzeylerine göre beş gruba ayrılarak zeytinyağı satın alma ve tüketimleri analize tabi tutulmuştur. 775 ailenin zeytinyağı tüketimleri araştırılmıştır. Bu çalışmada hanehalkı reisinin yaşı, eğitimi, cinsiyeti, medeni durumu, çalışıp çalışmama durumu ve benzeri demografik değişkenlerle birlikte, hanehalklarının toplam kullanılabilir geliri, gıda harcamaları, toplam yağ harcamaları ve zeytinyağı harcama değerlerini de kapsamaktadır. Çalışmada kullanılan verilerin analizinde, çapraz tablolardan yararlanılmış ve sosyo-demografik gruplar itibarıyla zeytinyağı harcama ve tüketim miktarlarının karşılaştırılmasında F Testi, T Testi ve Ki Kare (Chi Square) analizlerinden yararlanılmıştır. Zeytinyağı talep tahmininde Working-Leser ve Tobit modelleri kullanılmıştır. Araştırmanın sonucuna göre 2006 yılında Türkiye'de kişi başına yıllık zeytinyağı tüketiminin hiç tüketmeyen kişiler ve haneler dâhil edildiğinde 1 ila 1,5 kg civarında olduğu belirtilirken, Türkiye'de ailelerin %9'unun zeytinyağı tükettiği dikkate alınarak yapılan değerlendirmede, zeytinyağı tüketen ailelerde kişi başına yıllık

zeytinyağı tüketiminin 11,15 kg olduğu görülmüştür. Son yıllarda genellikle tüketici eğiliminin bitkisel yağ ağırlıklı olduğu ve zeytinyağının bitkisel yağlar içindeki tüketim oranının %5 civarında olduğu belirtilmiştir. Hanehalkı Bütçe Anketi verileri ışığında zeytinyağı tüketicilerinin sosyoekonomik özellikleri, aylık ve yıllık zeytinyağı tüketimleri, gıda harcamaları içinde zeytinyağı harcama payları hakkında bulgular elde edilmiştir. Çalışmada Türkiye’de zeytinyağı talebini etkileyen faktörler “Ekonomik Değişkenler” ve “Ekonomik Olmayan Değişkenler” olmak üzere ikiye ayrılmıştır. Ekonomik değişkenler sırasıyla; nüfusun gelir düzeyi, zeytinyağı fiyatları, Türkiye’de tüketilen yemeklik sıvı yağlar ve zeytinyağı, pazarlama yapısı, vergiler olarak belirlenirken, ekonomik olmayan değişkenler; nüfus artışı, tüketici davranışları, beslenme şekli olarak belirlenmiştir. Çalışmada gelir ve eğitim düzeyi arttıkça zeytinyağı tüketiminin arttığına dair bulgular elde edilmiştir. Ayrıca, gelişmiş ülkelerde son yıllardaki zeytinyağı tüketim rakamlarına bakıldığında da bu ürünün belirli bir tüketici kitlesine hitap ettiği belirtilmiştir. Gelir gruplarına göre aylık zeytinyağı tüketim ortalamaları arasındaki farkın istatistiksel olarak anlamlı olduğu sonucuna varılmış, yerleşim yerlerine göre bir değerlendirme yapıldığında ise kentsel alanda zeytinyağı tüketiminin kırsal alandan daha yüksek olduğu hesaplanmıştır. Hanehalkı genişliği olarak Türkiye ortalaması yaklaşık 4 olarak hesaplanmış, buna göre aileler 4 ve 4’den küçük ile 4’den büyük aileler olarak iki gruba ayrılmıştır. 4 veya 4 kişiden az üyesi bulunan hanelerde zeytinyağı tüketimi, 4 kişiden fazla üyesi bulunanlara oranla yaklaşık 1,7 kat daha fazla olduğu görülmüştür. Hanehalkı reisinin cinsiyeti değişkenine göre aile reisi erkek olduğunda tüketimin arttığı görülmüş ve sonuçların anlamlı olduğu tespit edilmiştir. Hanehalkı reisinin medeni durumunun zeytinyağı tüketiminde istatistiki açıdan önemli bir farka sebep olmadığı görülmüş, hanehalkı reisinin eğitim durumu zeytinyağı tüketiminde önemli bir değişken olarak öne çıkmıştır. Hanehalkı reisinin yaşı değişkeni olarak alındığında hanehalkı reisinin yaşının 30 yaşından küçük olduğu hanelerin aylık zeytinyağı tüketim miktarının 0,17 lt olduğu ve aile reisinin yaşı yükseldikçe hanehalklarının aylık zeytinyağı tüketim miktarlarının arttığı görülmüştür. Tobit modeli kullanılarak elde edilen tahminler ışığında toplam harcamalar ile zeytinyağı talebi arasında pozitif bir ilişki olduğu görülmüştür. (Oktay, 2010)

Polat (2010); yapmış olduğu doktora tezinde tüketicilerin yaş, eğitim, meslek, medeni durumu, cinsiyeti, geliri ve sosyal sınıfı gibi demografik ve sosyoekonomik özelliklerinin tüketici davranışını ciddi şekilde etkilediği belirtmiştir. Çalışmanın verilerinin derlenmesinde belirli bir sistematığe göre geliştirmiş oldukları anket formu kullanılmıştır. Anket, İstanbul ilinde rastsal seçimli küme örnekleme yöntemiyle, küme büyüklüğü optimal maliyet kriteri esas alınarak belirlenen on üç ilçe, küme noktaları için ise mahalleler baz alınarak ve İstanbul

içindeki mahalleler seçmen yoğunluğuna göre nüfusa oranlı bir şekilde bilgisayar yardımı ile son seçim kayıtları esas alınarak yapılmıştır. Verinin analizinde Faktör Analizi, Parametrik Olmayan Wilcoxon Sıra Testi, Lojistik Regresyon Analizi, Analitik Hiyerarşi Prosedürü (AHP) yöntemleri kullanılmıştır. Cinsiyet, yaş, eğitim düzeyi, evlilik durumu, öğrencilik durumu, vejetaryenlik durumu, çalışma durumu, işi, sigortalılık durumu, hanede yaşayan birey sayısı, hane gelirine katkı yapan birey sayısı ve hanedeki çocuk sayısı gibi hanehalkı karakteristiklerini gösteren veriler anketle derlenmiştir. Araştırmanın bulgularına göre margarini tüketicilerin %31'i her gün, %25'i haftada bir kaç kere, %18'i haftada bir kere, %12'si iki haftada bir kere, %7'si ayda bir kere kullandığı, geri kalan %8'inin ise hiç kullanmadığı görülmüştür. Her gün ve haftada bir kaç kere margarin kullananları sık kullananlar kategorisine dâhil edildiğinde, tüketicilerin %56'sının margarini günlük yaşamlarında yoğun olarak kullandığı görülmüştür. Araştırma bulgularına göre tereyağını tüketicilerin %39'u her gün, %26'sı haftada bir kaç kere, %9'u haftada bir kere, %3'ü iki haftada bir kere, %6'sı ayda bir kere kullandığı, geri kalan %17'si ise hiç kullanmadığı görülmüştür. Bu durumun tereyağının Türk tüketicilerinin günlük mutfağında önemli yer tuttuğunu ve tüketicilerin tereyağı tüketiminin, tüketim sıklığına göre daha dengeli dağılım gösterdiğini göstermektedir. Araştırma bulgularına göre Türk tüketicilerinin en yoğun kullandığı yemeklik yağın ayçiçeği yağı olduğu görülmüştür. Tüketicilerin %82'si ayçiçeği yağını her gün kullanmaktadır. Ayçiçeği yağını hiç kullanmayanların oranı sadece %8 düzeyindedir. Mısırözü yağında da zeytinyağına benzer olarak kullanıcıların hiç kullanmayanlar ve her gün kullananlar olarak ikiye ayrıldığı görülmektedir. Tüketicilerin %23'ü mısırözü yağını her gün kullanmakta, %62'si ise hiç kullanmamaktadır. Diğer üç yemeklik yağ türü olan soya yağı, kanola yağı ve fındık yağının tüketiciler tarafından tercih edilmediği araştırma bulguları arasındadır. Soya yağını hiç kullanmayanların oranı %97, kanola yağını hiç kullanmayanların oranı %97, fındık yağını hiç kullanmayanların oranı ise %98 seviyesinde olduğu görülmüştür. Araştırma bulgularına göre zeytinyağı kullanımını incelediğimizde tüketicilerin esasen yoğun günlük kullanıcılar ve hiç kullanmayanlar olarak iki gruba ayrıldığı görülmektedir. Araştırma bulgularına göre zeytinyağını tüketicilerin %44'ü her gün kullanmakta, %34'ü ise hiç kullanmamaktadır. Diğer kategorilere bakıldığında, tüketicilerin %16'sı zeytinyağını haftada bir kaç kere, %2'si haftada bir kere, %1'i iki haftada bir kere, %2'si ise ayda bir kere kullanmaktadır. Zeytinyağını her gün ve haftada bir kaç kere kullananların oranı %60 seviyesindedir. Çalışmada zeytinyağı kullanıcılarının kullanmayanlara oranla daha fazla sağlık riski bilinci taşıdığını göstermektedir. Tüketiciler zeytinyağını fiyat konusunda diğer yağlara (tereyağı, bitkisel yağlar ve margarin) göre 3.,

besin değeri ve lezzet konusunda 1., sağlık konusunda ise 2. olarak tercih etmişlerdir. Tüketicilerin gelir durumunun zeytinyağına öncelik verme derecesi üzerinde anlamlı pozitif etkiye sahip olduğu bulunmuştur. Tüketicilerin gelir düzeyi arttıkça zeytinyağını tercih etme olasılığı artmaktadır. Yapılan lojistik regresyon ve parametrik olmayan ortanca testi sonuçları zeytinyağı kullanıcılarının sağlık ve tüketim konusunda bilinçli tüketiciler olduğunu göstermiştir. Tüketicilerin yemeklik yağ tercihinde Tobit model kullanmıştır ve kurulan bu modelde tüketicilerin gelir durumunun ise zeytinyağı öncelik verme derecesi üzerinde anlamlı pozitif etkiye sahip olduğu bulunmuştur. (Polat, 2010)

Wang vd. (2010); yapmış oldukları çalışmalarında küçük çocuğa sahip Tibetli annelerin diyet kalıpları, besin alımları ve sosyodemografik gruplar arasındaki farklılıkları araştırmışlardır. Çalışmanın verileri Lhasa, Tibet kırsal kesiminde yaşayan, 24 aydan küçük çocuğu olan 386 kadın ile yapılan görüşmeler sonucunda derlenmiştir. Sosyodemografik değişkenlere göre besin tüketimlerini karşılaştırmak için parametrik olmayan testler kullanılmıştır. Sonuçlar, 2002 yılı Çin Ulusal Beslenme-Sağlık Anketi sonuçları ve Besin Referans Değerleri ile karşılaştırılmıştır. Çalışmanın analizinde kadınların yaşı, eğitim yılı, annenin mesleği (maternal occupation) ve hanehalkı büyüklüğü açıklayıcı değişkenler olarak kullanılmıştır. Tahıllar, et, sebze, meyve, yumurta ürünleri fındık, balık, içecekler, alkollü içecekler ve aperiatifleri kapsayan gıda ürünlerine ilişkin veriler katılımcılardan derlenmiştir. Derlenen verilerden annelerin beslenme durumları belirlenmiştir. Protein, yağ, karbonhidrat, A-B1-C ve E vitamini, kalsiyum, demir ve çinko alımları hesaplanmıştır. Shapiro-Wilk, Wilcoxon ve Kruskal-Wallis testleri verilerin analizinde kullanılmıştır. Besin referans değerleri ulusal standartlara dayanmaktadır. Orta derece fiziksel aktivite düzeyi ile Çin diyet referans alımları kullanılmıştır. Enerji, yağ, karbonhidrat, vitamin E ve çinko alımının besin referans değerlerine yakın olduğu, bununla birlikte, demir alımının referans değerlerinden daha yüksek olduğu görülmüştür. Diğer besinlerin alımlarının referans değerinden düşük olduğu (referans değerin yaklaşık %70'i) görülmüştür. Besin alımları ve sosyodemografik kategoriler arasında karşılaştırıldığında, öğrenim görülen yıla göre besin alımlarında önemli farklılıklar görülmüştür. Eğitim seviyesi yüksek olan bireylerin, düşük olan bireylere göre vitamin A ve C'yi anlamlı derecede yüksek aldıkları görülmüştür. Tibet beslenme alışkanlıklarının genelde tahıl ürünlerinin tüketimini yansıttığı görülmüştür. Sebze ve meyvenin daha az sıklıkla yendiği görülmüştür. Yağ ve protein alımlarının, Ulusal Beslenme ve Sağlık Anketi sonucunun değerlerinden önemli derecede daha düşük olduğu görülmüştür. Yağ alımının yaş ve eğitim ile ilişkili olduğu, yaş arttıkça yağ alım miktarının azaldığı, eğitim yılı arttıkça yağ alım miktarının arttığı görülmüştür. Cinsiyet, yaş, eğitim düzeyi ve fiziksel

aktivite gibi faktörlerin yeme alışkanlıklarını etkilediği görülmüştür. Varlıklı ve/veya eğitilmiş kadınların daha sağlıklı diyetler yaptığı görülmüştür. Eğitim gibi sosyo-demografik faktörlerin küçük çocuklu Tibetli annelerin besin alımı ile ilişkili olduğu görülmüştür. İyi eğitilmiş annelerin, eğitim düzeyi düşük olanlara göre daha dengeli diyet yemek ve gıda alımı yaptıkları görülmüştür. (Wang vd., 2010)

Yimer (2011); yapmış olduğu çalışmada Heterastatik Çift Engel modeli kullanarak tahıl harcamalarının belirleyicilerini analiz etmiştir. Çalışmada, 2004-2005 Hanehalkı Gelir ve Tüketim Harcamaları anketinden elde edilen toplam 21.600 hanehalkına ait veriler kullanılmıştır. Hanehalkının gıda tüketiminin yanı sıra sosyoekonomik ve demografik özellikleri hakkında ayrıntılı bilgi elde edilmiştir. Hanehalkı reisinin eğitimi, yaşı, medeni durumu ve cinsiyeti gibi değişkenler kurulan modelde kullanılmıştır. Hanehalkı reisinin yaşı, tahıl tüketim seviyesinin belirlenmesinde etkili bir karakteristik iken, hanehalkının tahıl tüketme kararını etkileyecek bir karakteristik olarak görülmemiştir. Hane büyüklükleri, bağımlılık oranları, eğitim, istihdam durumu, cinsiyet gibi sosyo-demografik karakteristikler tahıl tüketim kararı ve tüketim düzeyini açıklayan ve anlamlı karakteristikler olarak görülmüştür. Ayrıca gelir ve fiyatlar gibi ekonomik faktörler tahıl tüketim talebinin önemli belirleyicileri olarak görülmüştür. Gelir ve eğitim mısır tüketimi ile negatif ilişkili iken tahıl harcamalarını pozitif ve anlamlı etkileyen karakteristikler olarak görülmüştür. Eğitilmiş hanehalkı reislerinin mısır ve süpürge darısı tüketimi son derece düşüktür (Yimer, 2011).

Bamiro (2012); yapmış olduğu çalışmada Nijerya'da hane dışındaki gıda tüketiminin belirleyicilerini araştırmıştır. Çalışmada kullanılan veriler Lagos kent merkezinde ve Lagos kenti çevresindeki yerlerden derlenmiştir. Verilerin analizinde tanımlayıcı istatistikler, Tobit ve Probit Regresyon Modelleri kullanılmıştır. Analizde kullanılan açıklayıcı değişkenler; aylık hanehalkı geliri, hanehalkı reisinin cinsiyeti, eğitim seviyesi ve medeni durumu, hanehalkı büyüklüğü, hanedeki 10 yaşın altındaki çocuk yüzdesi, hanedeki 10-18 (18 dahil değil) yaş arası çocuk yüzdesi, hanedeki kadın yüzdesidir. Çalışmanın sonuçlarına göre erkeklerin, kadın meslektaşlarına göre daha fazla ev dışı gıda tüketiminde bulundukları görülmüştür. Ev dışı gıda tüketimini yapan bireylerin büyük çoğunluğunu memurlar ve iş adamlarının oluşturduğu görülmüştür. Probit Regresyon sonuçlarına göre ev dışı yemek seçimine sosyoekonomik değişkenlerin etkisinin olduğu görülmüştür. Ev dışı gıda harcamalarının düzeyi genellikle cinsiyet, hanehalkı büyüklüğü ve evdeki yetişkinlerin oranından etkilenmiştir. Hanehalkı büyüklüğü arttıkça ev dışı gıda tüketimi azalmaktadır, ayrıca ev dışı gıda tüketim oranı hanedeki yetişkin yüzdesi ile birlikte artmaktadır. (Bamiro, 2012)

Çağlayan ve Astar (2012); yapmış oldukları çalışmalarında Türkiye’de hanehalkı toplam tüketim harcamalarının belirleyicilerini incelemişlerdir. Çalışmanın verileri TÜİK tarafından derlenen 2009 yılına ait 12 aylık hanehalkı tüketim harcaması verileridir. Toplam 5.658 kayıt analize dâhil edilmiştir. Çalışmada Quantile Regresyon modeli, kırsal ve kentsel alanlar için farklı noktada tüketim bağlantılarını incelemek için kullanılmıştır. Analizde kullanılan bağımsız değişkenler; gelir, hanehalkı büyüklüğü, kır-kent durumu, cinsiyet, hanehalkı reisinin eğitim durumu, yaşı, medeni durumu ve kariyer durumu (sosyal statü), sigorta durumu, hanehalkı tipi, mülkiyet durumudur. Modeller kırsal ve kentsel alanlar için ayrı ayrı tahmin edildiğinde, tüketim harcamalarını etkileyen faktörlerin farklılıklarını görmek mümkündür. Bulgulara göre yaş karakteristiği genel ve kentsel tahminlerdeki tüketim harcamalarını artırırken, kırsal tahminlerdeki tüketim harcamalarını azalttığı görülmüştür. Kırsal tahminlerde yaş, gelir, medeni durum, sigorta ve hanehalkı büyüklüğü önemli olarak bulunmuştur. Kırsal kentsel ayırımı gözetmeksizin yapılan tüm gözlemler yoluyla yapılan tahminlerde, kentsel tahminlerde aynı değişkenler için erkeklerin tüketim harcamalarının alt değeri, kadınların tüketim harcamalarına oldukça yakındır. Tahminlerinin sonuçları harcamalar ile gelir arasında pozitif bir bağlantı olduğunu göstermektedir, gelir arttıkça harcamaların yükseldiği görülmüştür. Bu artış özellikler üst kuantillerde daha yüksektir. Alt kuantillerde kentsel sakinlerin tüketim harcamaları kırsal sakinlerin neredeyse iki katı kadar yüksek iken, üst kuantillerde azaldığı görülmüştür. Bunun nedeninin kentsel alanlarda pahalı ve zor yaşam koşulları olabileceği belirtilmiştir. Kırsal ve kentsel sakinlerin arasındaki tüketim harcamaları farkları üst kuantiller de azalmaktadır. Sigorta durumunun önemi üst kuantillerde azalmaktadır. Alt kuantillerde genişletilmiş ailelerin tüketim harcamaları tek bir yetişkinden oluşan hanelere göre daha yüksektir (hanehalkı tipi). Yaş artışı tüketim harcamalarını yükseltmektedir. Üst kuantillerde mülkiyet durumunun tüketim harcamaları üzerine etkisi olduğu görülmüştür. Cinsiyet değişkeni incelendiğinde erkeklerin tüketim harcamalarının, kadınların tüketim harcamalarından daha düşük olduğu görülmüştür. Modeller kırsal ve kentsel olarak ayrı ayrı incelendiğinde tüketim harcamalarını etkileyen farklılıkların daha rahat görülebileceği belirtilmiştir. Kentsel tahminlerde elde edilen sonuçlar tüm verilerin kullanımı yoluyla tahmin edilen sonuçlara yakındır. Sigortalı olma durumu bir hayat garantisi olarak kabul edilmekte ve tüketim harcamalarını etkileyen bir faktör olarak dikkati çekmektedir. Kentsel tahminlerden elde edilen sonuçların tüm verilerin kullanımı yoluyla tahmin edilen sonuçlara yakın olduğu görülmüştür. (Çağlayan ve Astar, 2012)

Meng vd. (2012); yapmış oldukları çalışmalarında Gana’nın Kuzey Bölgesindeki kırsal hanelerin tarım gelirini belirleyen hanehalkı faktörlerinin ve tarımsal özelliklerin

hangileri olduğunu belirlemişlerdir. Ayrıca çalışmanın bir diğer amacı tarım gelirinin, tarım dışı gelirin ve diğer sosyo-demografik faktörlerin, hanehalkı taze sebze harcamalarını nasıl etkilediğini incelemektir. Gana Ulusal İstatistik Servisinin 27 Temmuz-2 Ağustos 2010 arasında derlediği ve toplam 204 gözlemden oluşan veriler analizde kullanılmıştır. Hanelerden taze sebze harcamaları, tarım gelirleri, tarım dışı gelirler, yaş, cinsiyet, medeni durum, eğitim durumu gibi sosyo-demografik bilgileri, ekilen temel gıda bitkileri (staple crops: pirinç, mısır, tatlı patates, manyok (cassava), süpürge darısı (sorgum), darı) ve endüstri bitkileri (cash crops: biber, patlıcan, bamya, domates, pamuk, tütün) ve yer fıstığı ekilen dönümlük alanların toplam sayısı gibi tarımsal özelliklerin verileri derlenmiştir. Tahmin yaklaşımı için Eşanlı Denklem Metodu ampirik modelde uygulanmıştır. Model En Küçük Kareler Yöntemi ile çözülmüştür. Sebze harcamaları denkleminde, hanehalkı haftalık taze sebze harcamalarını etkileyen altı sosyo-demografik hanehalkı karakteristiği istatistiksel olarak anlamlıdır. Tarım dışı gelire sahip olup olmama ve eğitim düzeyi haftalık taze sebze harcamalarını pozitif olarak etkilediği görülmüştür. Cinsiyet, yaş, çocuk ve yaşlıların sayısı gibi demografik faktörler haftalık taze sebze harcamalarını belirlemede önemli olduğu görülmüştür (Meng vd., 2012).

Sekhampu (2012); yapmış olduğu çalışmasında Güney Afrika'da düşük gelirli bir yerleşim yerinde seçilen sosyoekonomik özelliklerinin gıda harcama kalıpları üzerine etkisini analiz etmiştir. Çalışmanın anketi Mart 2012 yılında, rastgele seçilen 585 kişiye yüz yüze görüşme yoluyla uygulanmıştır. Çalışmanın analizinde Çoklu Regresyon Modeli kullanılmıştır. Hanenin aylık geliri, hanehalkı büyüklüğü, hanehalkı reisinin örgün eğitim yılı sayısı, hanehalkı reisinin cinsiyeti, yaşı, medeni durumu ve istihdam durumu regresyon modeline açıklayıcı değişken olarak dâhil edilmiştir. Hanenin geliri, hanehalkı büyüklüğü, hanehalkı reisinin yaşı, istihdam durumu, medeni durumu ve eğitim düzeyinin diğer şartlar sabitken toplam aylık gıda harcamasını etkilediği görülmüştür. Hanehalkı reisinin cinsiyetinin bir etkisi olmadığı gözlemlenmiştir. Katılımcıların medeni durumu ile gıda harcamaları arasında negatif bir ilişki olduğu görülmüştür. Hanenin geliri arttıkça, hanede yapılan toplam gıda harcaması artmaktadır. Hanehalkı büyüklüğü arttıkça gıda için yapılan harcamalarda artmaktadır. Hanehalkı reisinin eğitim yılı arttıkça gıdaya harcanan para daha eğitimsiz hanehalkı reislerine göre artmaktadır. Diğer değişkenler sabitken evli katılımcılar, evlenmemiş katılımcılara göre gıda üzerine göreceli olarak daha az harcama yapmaktadırlar. Hanehalkı büyüklüğü, istihdam durumu ve hanehalkı reisinin medeni durumu gıda harcamalarının en önemli belirleyicileri olarak bulunmuştur. (Sekhampu, 2012)

Ali vd. (2013); yapmış oldukları çalışmalarında Pakistan'da yemeklik yağ tüketimini etkileyen başlıca faktörleri belirlemişlerdir. Yemeklik yağın Pakistan'da gıda ana bileşeni

olduğu ve hanehalkı yiyecek tüketiminin önemli bir maddesi olduğu belirtilmiştir. Pakistan yemeklik yağ tüketimini etkileyen birçok faktörün olduğu ve bunların; gelirdeki artış, nüfus artışı, kentleşme, markayı sevmeye ve antipati duymaya olduğu belirtilmiştir. Fiyat, gelir, aile büyüklüğü ve aylık gelirin önemli ölçüde sadeyağ (tereyağı) ve yağ tüketimi fonksiyonunu etkilediğini öne sürmüşlerdir. Veriler 2012 yılında önceden belirlenmiş 150 haneden toplanmıştır. Ürünlerin satın alma ve tüketici tercihlerini incelemek için regresyon (Double log form) ve faktör analizi yöntemleri kullanılmıştır. Sıvı yağ ve sadeyağ talep fonksiyonu tahmini ardındaki ana fikir yemeklik yağ tüketimini etkileyen belli başlı faktörleri belirlemektir. Modeli tahmin etmede kullanılan bağımsız değişkenler; tüketilen aylık yağ miktarı, hanenin aylık geliri, yağ fiyatları, yemeklik yağ giderlerinin toplam gıda harcaması içerisindeki payı, aile büyüklüğü, hanehalkı reisinin eğitim düzeyi ve mesleği olarak belirlenmiştir. Modelin sonuçları ayrı ayrı belirtilmiştir. Sıvı yağ talep fonksiyonu analiz sonuçlarına göre yağ tüketimi büyük ölçüde gelire bağlıdır. Yağ giderlerinin payı, toplam gıda harcamalarında önemli bir değişkendir. Aile büyüklüğü hanehalkının yağ tüketimini etkileyen bir faktör olarak belirlenmiştir. Eğitim katsayısı yağ tüketiminde olumlu ilişkili ama analizde anlamlı olarak çıkmamıştır. Hanehalkı reisinin mesleği değişkeni pozitif elastikiyet katsayısına sahip olmasına karşın önemsiz bir değişken olarak bulunmuştur. Sadeyağ Talep Fonksiyonu analizi sonuçlarına göre sadeyağ tüketiminin gelire bağlı olduğu görülmüştür. Sadeyağ harcamalarının payı toplam gıda harcamalarında önemli bir değişken olduğu görülmüştür. Aile büyüklüğü ile sadeyağ tüketimini etkileyen bir faktör olarak görülmüştür. Eğitim, sadeyağ tüketiminde pozitif ilişkili fakat önemsiz bir değişken olarak çıkmıştır. Hanehalkı reisinin mesleği değişkeni önemsiz bir değişken olarak görülmüştür. (Ali vd., 2013)

Sayılı vd. (2013); yapmış oldukları çalışmalarında Ordu ili merkez ilçedeki ailelerin yemeklik yağ tüketim durumları incelenmiştir. Ayrıca fındık yağı tüketimine etki eden faktörlere ilişkin faktör analizi yapılmıştır. Araştırmada kullanılan veriler Ocak-2013 yılında 272 adet kişi ile yapılan anket çalışması sonucu elde edilmiştir. Araştırma sonuçlarına göre, en fazla tüketilen yağlar tereyağı (%71,32) ve fındık yağıdır (%61,76). Toplam yağ tüketim miktarı aile başına 6,71 kg/ay ve kişi başına ise 1,89 kg/ay'dır. Tüketicilerin eğitim durumları incelendiğinde; fakülte mezunu olanların oranının (%35,66) yüksek olduğu dikkat çekmektedir. Bunu yüksekokul (%20,22), lise (%20,22), ilkokul (%12,13), ortaokul (%5,14), yüksek lisans (%5,14) ve doktora (%0,73) mezunları ile okur-yazar (%0,36) ve okur-yazar olmayanlar (%0,36) izlemektedir. Ankete katılanların %49,26'sını memurlar, %22,79'unu ev hanımları, %12,13'ünü öğrenciler, %5,88'ini emekliler, %2,94'ünü esnaf, %2,57'sini işçi, %0,73'ünü

özel sektörde çalışanlar ve %0,36'sını ise çiftçiler oluşturmaktadır. Araştırma sonuçlarına göre; tüketicilerin %73,89'unun evli olduğu tespit edilmiştir. Anket uygulanan ailelerde ortalama aile genişliğinin 3,55 kişi ve çalışan birey sayısının ise 1,52 kişi olduğu belirlenmiştir. Ailelerin aylık gelirleri ortalama 3149,75 TL olup bunun da %21,19'unun (667,48 TL) gıda harcaması için kullanıldığı tespit edilmiştir. (Sayılı vd., 2013)

Tümer (2013); yapmış olduğu çalışmasında hanehalkının sofralık zeytin ve zeytinyağı tüketimine ilişkin sosyoekonomik ve demografik özelliklerin etkisini belirlemeyi amaçlamıştır. Orantılı Örneklem Yöntemine dayalı olarak 2012 yılında, yüz yüze yapılan 384 anketten elde edilen veriler kullanılmıştır. Hanehalkı reisinin cinsiyet ve eğitim düzeyi, hane geliri ve diyet durumunun yapılan Ki kare analizi sonucunda sofralı zeytini tüketim tüketicilerin kararları üzerinde istatistiksel olarak anlamlı etkilerinin olduğu görülmüştür. Hanehalkı reisinin yaş, cinsiyet ve eğitim düzeyi, hane geliri, büyüklüğü ve diyet durumunun tüketicilerin zeytinyağı tüketim kararlarında istatistiksel olarak anlamlı etkilere sahip olduğu görülmüştür. (Tümer, 2013)

Akaichi ve Revoredo-Giha (2014); yapmış oldukları çalışmalarında, altı farklı süt ve süt ürünü (taze süt, süt tozu, margarin ve tereyağı, chambiko (bir tür mayhoş pıhtılaşmış süt ürünü), yoğurt ve peynir) için Malavi'li hanelerin taleplerini değerlendirmişlerdir. Analizde, Malavi Ulusal İstatistik Ofisi tarafından Mart 2010-Mart 2011 arasında derlenmiş Üçüncü Bütünleşik Hanehalkı Anketi verileri kullanılmıştır. Analizde ulusal, kırsal ve kentsel olmak üzere toplam 12.271 gözlem kullanılmıştır. Çalışmada Sansürlü Üç Aşamalı AIDS modeli kullanılmıştır. Kentsel alanlarda yaşayan hanehalkının taze süt, süt tozu, tereyağı ve margarin ve yoğurt ürünlerini kırsal alanlarda yaşayan hanehalklarına göre tüketme olasılığının daha yüksek oranda olduğu görülmüştür. Kuzeyde yaşayan Malavi'lilerin, ülkenin geri kalanında yaşayan Malavili'lere göre taze süt ve chambiko ürünlerini tüketme olasılıklarının daha yüksek olduğu görülürken, yoğurt tüketme olasılıklarının ise daha düşük olduğu görülmüştür. Güney'de yaşayan Malavi'lilerin, ülkenin kuzey ve merkezinde yaşayan Malavi'lilere göre yoğurt tüketme olasılıklarının daha yüksek olduğu, taze süt ve süt tozu gibi gıda ürünlerini tüketme olasılıklarının daha düşük olduğu görülmüştür. Yoksul olmayan hanelerin, yoksul hanelere göre taze süt ve süt tozu tüketme olasılıklarının daha yüksek olduğu görülmüştür. Fakat tereyağı, margarin, chambiko, peynir ve yoğurt ürünlerinin tüketimi açısından yoksul ve yoksul olmayan haneler arasında anlamlı bir farklılığın olmadığı görülmüştür. Malavi'lilerin süt ve süt ürünlerini tüketme farklılıklarının nedeni yerleşim alanı, yoksulluk düzeyi olarak görülmektedir. Süt ve süt ürünlerinin fiyatları arttıkça tüketimlerinin düştüğü görülmüştür.

Ucuz süt tozunun, taze süt talebi üzerinde olumsuz etkisinin olabileceği belirtilmiştir. (Akaichi ve Revoredo-Giha, 2014)

Khan vd. (2014); çalışmalarında 1975-2012 yılları arasında Pakistan’da yıllık olarak derlenen verileri zaman serileri kullanarak tüketim ile gelir arasındaki ilişkiyi araştırmışlardır. Gelirin tüketime etkisini analiz etmek için Doğrusal Regresyon modeli kullanılmıştır. En küçük karalar yöntemi ile modelin parametreleri tahmin edilmiştir. Ampirik sonuçlar gelir ve tüketim arasındaki pozitif bir ilişkinin varlığını desteklemektedir (Khan vd., 2014).

Kostakis (2014); yapmış olduğu çalışmasında finansal kriz sırasında 800 Yunan hanenin sosyo-demografik ve davranış belirleyicilerinin hane harcamalarına etkisini analiz etmiştir. Seçilen 800 örnek, tabakalı rastgele örnekleme yoluyla seçilmiştir ve veriler Ekim 2011-Şubat 2012 yılları arasında derlenmiştir. Makalede Çok Değişkenli Regresyon ve Logit Model kullanılmıştır. Gelir, cinsiyet, eğitim, yaş, medeni durum, ikamet, ucuz etiketli malları alıp almadığı, hazır yemek siparişlerini azaltıp azaltmama durumu, emeklilik durumu (emekli/emekli değil), kısa vadede gıda fiyatlarında bir artış olup olmadığının farkına varması, kiralık bir evde yaşayıp yaşamadığı, tüketici mevcut ekonomik durumu göz önüne aldığı ebeveynliğin artan bir maliyet olduğu göz önüne alınıp alınmadığı gibi birçok değişken kullanılmıştır. Kişi başı aylık gıda harcamasının 173,88 Euro olduğu, gıda harcamasının toplam harcamanın %16,82’sine karşılık geldiği görülmüştür. Kırsal tüketicilerin ortalama aylık gelirinin 853,34 Euro iken kentsel tüketicilerin ortalama aylık gelirinin 918,15 Euro olduğu görülmüştür. Demografik, sosyoekonomik karakteristikler ve tüketici tutumları ile gıda harcamaları arasında güçlü ilişkiler bulunmuştur ve gıda harcamalarının ana belirleyicileri olarak görülmüştür. Gelir, cinsiyet, yaş, eğitim düzeyi, medeni durum, yerleşim yeri ve istihdam durumu gibi değişkenlerin gıda harcamaları üzerinde etkisi olduğu sonucuna ulaşılmıştır. Yüksek eğitilmiş bireylerin daha sağlıklı ürünleri tercih ettikleri, evli bireylerin daha fazla tükettikleri görülmüştür. Emekli bireylerin sağlıklı ürünler tüketme konusunda diğer yaş gruplarındaki bireylere göre daha dikkatli oldukları gözlenmiştir. Kadınların erkeklere göre beslenme konusuna daha fazla önem verdikleri ve gıdaya daha fazla para harcadıkları görülmüştür. (Kostakis, 2014)

Mayén vd. (2014); yapmış oldukları çalışmalarında sosyoekonomik statü ile düşük ve orta gelir grubunda yer alan bireylerin diyet alımları arasındaki ilişkiyi belirlemek, değerlendirmek için araştırmaların sistematik bir değerlendirmesini yapmışlardır. Yüksek gelirli ülkelerde, yüksek sosyoekonomik statü genellikle sağlıklı bir diyet ile ilişkilidir, fakat düşük ve orta gelirli ülkelerde, sosyal farklılıkların diyet alımları ile ilişkili olup olmadığının bir araştırma konusu olarak ortaya çıktığı belirtilmiştir. Çalışmada, 1996-2013 arasında

yayınlanan, düşük ve orta gelir grubunda yer alan erişkinlerde yapılan kohort ve kesitsel çalışmalar sistematik olarak incelenmiştir. Düşük ve orta gelire sahip on yedi ülkeden toplam 33 çalışma (5 düşük gelirli ülke, 12 orta gelirli ülke, 31 kesitsel, 2 dikey (longitudinal) çalışma) incelenmiştir. Çalışmaların çoğu Brezilya (8), Çin (6) ve İran'da (4) gerçekleştirilmiştir. Çalışmada sosyoekonomik statünün, yüksek gelirli, düşük ve orta gelirli ülkelerde sağlığın önemli bir belirleyicisi olduğu belirtilmiştir. Yüksek gelirli ülkelerde, yüksek sosyoekonomik statüye sahip bireylerin kepekli tahıllar (tam tahıllı ürünler), yağsız et, balık, düşük yağlı süt ve süt ürünleri (mandıra ürünleri), meyve ve sebzeler gibi sağlıklı gıdalar tüketme eğiliminde iken, düşük sosyoekonomik statüye sahip bireylerin fazla yağ ve daha az lif tüketme eğiliminde olduğu belirtilmiştir. Sosyoekonomik değişkenler ile enerji, protein, yağ, karbonhidrat, lif alım ilişkileri, taranan bütün makalelerden elde edilen sonuçlar doğrultusunda ayrı ayrı tablo haline getirilmiştir. Oluşturulan tablolarda yayını yapan kişi, ilgili gıda alımı ile ilişkili gördüğü sosyoekonomik değişkenleri detaylı bir şekilde belirtmiştir. Düşük gelir grubundaki ülkelerde, sosyoekonomik statü, diyet yağ alımı, eğitim, gelir, konum ve diğer değişkenler arasındaki ilişki belirtilmiştir. Orta gelir grubundaki ülkelerde yüksek sosyoekonomik statü veya kent konumu, genellikle yüksek yağ (toplam yağ, kolesterol, çoklu doymamış yağ, doymuş yağ, tekli doymamış yağ) tüketimi ile ilişkili olduğu görülmüştür. Kent lokasyonu ile kır lokasyonu karşılaştırıldığında toplam yağ alımında %6 farklılık olduğu, tekli doymamış yağ için ise farklılığın %29 olduğu görülmüştür. Düşük ve orta alt gelir grubundaki ülkelerde yüksek sosyoekonomik statü ile birlikte, sosyoekonomik faktörler ve coğrafi konumun gıda alım kalıpları ve miktarları üzerinde etkisi olduğu görülmüştür. Kentte yaşayan bireylerin genellikle sağlıklı beslendikleri görülmüştür. Orta gelir grubundaki ülkelerde yüksek sosyoekonomik sınıfın, yüksek enerji ve doymuş yağ tüketimi gibi sağlıksız diyet kalıpları ile ilgili olduğu görülmüştür. Yüksek sosyoekonomik statüye sahip ve kentte yaşayan bireylerin fazla yağ tükettikleri görülmüştür. Bunun nedeninin, yağ açısından zengin bir diyetin artan refah ile ilişkili olmasının sonucu olduğu belirtilmiştir. Zengin bireylerde, yüksek yağ alımının hayvan kökenli gıdaların büyük miktarda tüketiminin bir sonucu olabileceği belirtilmiştir. Genel popülasyonda doymamış yağ alımı, doymuş yağ alımından fazladır, bunun nedeninin ithal edilen et ve süt ürünlerinden geldiği görülmüştür. Düşük ve orta alt gelire sahip ülkelerde yaşayan, yüksek sosyoekonomik statüye sahip ve kentsel lokasyonda yaşayan bireylerin sağlıklı bir diyet takip etme eğiliminde olduğu, bu bireylerin enerji, kolesterol ve doymuş yağı daha fazla tüketirken, daha az lif tükettikleri görülmüştür. (Mayén vd., 2014)

Uluskar (2014); yapmış olduđu yüksek lisans tezinde, tüketicilerin eğitim durumları ile gelir düzeylerine göre hanehalkı ortalama zeytinyağı tüketimleri arasındaki fark istatistikî açıdan anlamlı çıkmıştır. (Uluskar, 2014)

Terin vd. (2015); yapmış oldukları çalışmalarında Türkiye’de ekonomik büyüme ile sağlanan gelir artışının, diğer sosyo-demografik ve ekonomik faktörlerin hane halkı süt ve süt ürünleri tüketim harcamalarına etkileri analiz edilmiştir. Çalışmada kullanılan veriler, Türkiye İstatistik Kurumu (TÜİK) tarafından 2010 yılında yapılan Hane Halkı Tüketim Harcamaları anketlerinden elde edilmiştir. Örneklem büyüklüğü, eksik ve aykırı gözlemler çıkarıldıktan sonra 9.442 kişi olarak belirlenmiştir. Çalışmada Çoklu Heckman Seçicilik Modeli (HSM) kullanılmıştır. Analizde kullanılan, hanehalkı reisine ait sosyo-demografik karakteristikler cinsiyet, medeni durum, çalışıp çalışmama durumu, yaş grubu, eğitim yılı ve zorunlu sağlık sigortasıdır. Hanehalkına ait karakteristikler ise teknoloji sayısı (evde kullanılan bilgisayar, cep telefonu, internet, uydu ve TV vb) gayrimenkul varlığı, gelir grubu, ev sahiplik durumu, çocukların yaş gruplarına göre dağılımı ve eş değeri aile büyüklüğüdür. Araştırma sonuçları, kırsal ve kentsel alandaki hanelerin sahip olduđu sosyo-demografik ve ekonomik karakteristiklerin süt ve süt ürünleri tüketim harcamalarını farklı şekilde etkilediğini göstermiştir. Hanehalkı ve hane reisi özelliklerinin gerek kırsal gerekse kentsel alanda hanelerin süt ve süt ürünleri tüketim olasılığı ile harcama düzeylerini belirlemede önemli bir role sahip olduđu araştırma sonuçları tarafından desteklenmiştir. Kentsel alanda yaşayan hanelerin, hem süt hem de peynir ile diğer süt ürünlerine, kırsal alanda yaşanan hanelere göre daha fazla harcama yaptığı tespit edilmiştir. Buna karşın kırsal alanda yaşayan hanelerin yoğurt harcamasının kentsel alandaki hanelere göre daha yüksek olduđu görülmüştür. Bütün süt ve süt ürünlerinde, kentteki hanelerde tüketim olasılıklarının kırsaldaki hanelere göre daha yüksek olduđu bulunmuştur. Kentsel alanda hane reisinin erkek olması, özellikle genç olması (20-29 yaş), yeşil karta sahip olması ve ailenin kirada oturması hem süt ve süt ürünleri tüketim olasılığını hem de süt ve süt ürünleri tüketim harcamasını negatif yönde etkilemektedir. Hane reisinin evli olması, ailenin apartmanda yaşaması, çocuklu aile olması, 0-5 yaş grubunda çocuğa sahip olması ve hane gelirinin yıllık 15 bin TL’den fazla olması süt ve süt ürünleri tüketim olasılığı ve tüketim harcamasını pozitif yönde etkilemiştir. Kırsal alanda hane reisinin erkek olması, genç (20-29 yaş) olması, yeşil karta sahip olması kentsel alanda olduđu gibi hem süt ve süt ürünleri tüketim olasılığını hem de süt ve süt ürünleri tüketim harcamasını negatif yönde etkilemiştir. Buna karşın, hane reisin evli olması, çalışıyor olması, zorunlu sigortaya sahip olması, ailenin 0-5 yaş grubunda çocuğa sahip olması, gayrimenkul sayısı ve hane gelirinin yıllık 15 bin TL’den fazla olması süt ve süt ürünleri

tüketim harcamasını pozitif yönde etkilemiştir. Araştırma sonuçlarından da anlaşılacağı üzere gerek kentsel gerekse kırsal alanda hanelerin sahip oldukları sosyo-demografik ve ekonomik faktörlerin hanelerin süt ve süt ürünleri tüketim olasılığı ve tüketim harcama düzeylerini önemli ölçüde etkilemektedir. (Terin vd., 2015)



DÖRDÜNCÜ BÖLÜM

HANEHALKI YAĞ TÜKETİMİNİN MODELLENMESİ

4.1 Uygulamanın Amacı

Tez çalışmasında gerçekleştirilen uygulamanın amacı; hanehalkının kullanmış olduğu dört farklı yağ türü olan zeytinyağı, margarin ve diğer bitkisel yağlar, yenilebilir sıvı yağ ve tereyağı tüketimini etkileyen faktörlerin, tüketim düzeylerinin ve hanehalkının karakteristik özelliklerine göre tüketim durumlarının Karar Ağacı yöntemi ile belirlenmesidir.

4.2 Uygulamada Kullanılan Veriler

Tezde kullanılan veriler 2009-2012 yılları arasında TÜİK tarafından derlenen Hanehalkı Bütçe Anketine ait verilerdir. TÜİK tarafından bu yıllara ait ayrı ayrı gönderilen veride üç farklı veri dosyası bulunmaktadır. Bunlar;

- Fert Karakteristikleri (cinsiyet, eğitim durumu, sigorta durumu, ...)
- Hane Karakteristikleri (konut tipi, hanehalkı tipi, otomobil sahipliği, ...)
- Tüketim Karakteristikleridir (tütilen mal ve hizmetlere ilişkin değerler)

Her yıla ait, üç farklı veri setinde yer alan veriler “Bülten Numarası” adı verilen tekil bir alan ile birleştirilmiştir ve dört yıla ait birer veri dosyası oluşturulmuştur. Daha sonra gerekli ön işleme ve temizleme aşamaları gerçekleştirilmiş ve veri setleri birleştirilerek tek bir veri dosyası oluşturulmuştur.

4.2.1 Veri Ön Hazırlama Süreci

Veri madenciliği yapan analistin en büyük sorunlarından biri çalışmış olduğu veri seti, veri tabanı ya da veri ambarındaki verilerin eksik, yanlış veya hatalı olmasıdır. Bu durum normal şartlar altında veride olmaması gereken bir durumdur. Fakat bu gibi durumlarla uygulamalarda sıklıkla karşılaşmaktadır. Bu durumlar genelde veri girişinin yapıldığı esnada gerçekleşen, veri girişini yapan kullanıcının hatasından kaynaklanmaktadır. Veri tabanlarında bilgi keşfi sürecinin en önemli aşamalarından birisi, hatta bu sürecin en uzun zaman alan kısmı “Veri Ön İşleme ve Temizleme” aşamasıdır. Veri madencilerinin en önemli görevlerinden birisi çalışmış olduğu veri setine hâkim olmasıdır. Veri setindeki tüm alanları, bu alanların içerdiği değerleri bilmesi son derece önemlidir. Veri madencileri, analiz yapacağı veri setine hâkim olduktan sonra, veri setindeki eksik, yanlış veya hatalı verileri yapacağı işleme göre ön işleme tabii tutmalı ve temizlemelidir. Çünkü veri madenciliği teknikleri hatalı, eksik veya yanlış verilerin olması gibi durumlarda düzgün sonuçlar vermeyebilir. Veri

Önişleme ve Temizleme aşaması gibi veri tabanlarında bilgi keşfi sürecinde yer alan tüm aşamalar tezin Veri Madenciliği kısmında detaylı bir şekilde anlatılmıştır.

4.2.1.1 Fert Karakteristiği Önişleme ve Temizleme Aşaması

Bu kısımda 2009, 2010, 2011 ve 2012 yıllarına ait Türkiye İstatistik Kurumu tarafından derlenmiş Hane Halkı Bütçe Anketi verilerinden Fert Karakteristiklerinin veri ön hazırlama işlemine yer verilmiştir. Veri setinin fert karakteristiklerini içeren alanların birkaçında işlem yapılmasına karşın sadece analizde kullanmış olduğumuz değişkenler ile ilgili kısma yer verilmiştir. Düzenleme yapılan karakteristikler yaş, eğitim durumu ve sağlık sigortası karakteristikleridir.

- Tablo 4.1 2009, 2010, 2011 ve 2012 yıllarında hanede bulunan bireylerin yaşlarının ve/veya bitirdiği yaşın değerini göstermektedir.

Tablo 4.1 Hanehalkı Bireylerinin Yaş Aralıkları

	Yıllar (2009, 2010, 2011)		Yıllar (2012)		
Karakteristik ve Açıklaması	Aldığı Değerler	Değer Açıklaması	Karakteristik ve Açıklaması	Aldığı Değerler	Değer Açıklaması
YAS, Bitirilen yaş grubu	1	0-5 yaş	YAS, Hanehalkı fertlerinin anket ayında bitirmiş olduğu yaş verilmiştir	0...99	
	2	6-14 yaş			
	3	15-19 yaş			
	4	20-24 yaş			
	5	25-29 yaş			
	6	30-34 yaş			
	7	35-39 yaş			
	8	40-44 yaş			
	9	45-49 yaş			
	10	50-54 yaş			
	11	55-59 yaş			
	12	60-64 yaş			
	13	65+yaş			

Kaynak: TÜİK, 2009; TÜİK, 2010; TÜİK, 2011a; TÜİK, 2012

Tablo 4.1 incelendiğinde 2009, 2010 ve 2011 yıllarında veri setinde kullanılan YAS karakteristikleri ortak değerler içerirken 2012 yılında bu değerler farklılık göstermektedir. TÜİK tarafından 2009, 2010 ve 2011 yılında derlenen veri setlerinde YAS karakteristiği kategorik hale getirilmiş ve 13 farklı kategori kullanılmıştır, 2012'de derlenen veri setinde ise kategorik bir değişken kullanılmamış ferdin anket ayında bitirdiği yaş yazılmıştır. Analizin düzgün bir şekilde yapılabilmesi için 2012 yılında kullanılan “YAS” karakteristiğinin diğer

yıllarda ortak olarak düzenlenen duruma çevrilmesi daha uygundur, çünkü diğer 3 yılda ferdin yaşına ilişkin kesin bir değer bulunmamakta bunun yerine yaş kategorisine ait bir değer ait bir değer bulunmamaktadır.

- 2009, 2010, 2011 ve 2012 yıllarına ait derlen verilerde sağlık sigortası içerikleri birbirinden farklıdır. Tablo 4.2 hanehalkı reisinin sağlık sigorta durumunu göstermektedir.

Tablo 4.2 Hanehalkı Reisinin Sağlık Sigorta Durumu

	Yıllar (2009, 2010, 2011)		Yıllar (2012)		
Karakteristik ve Açıklaması	Aldığı Değerler	Değer Açıklaması	Karakteristik ve Açıklaması	Aldığı Değerler	Değer Açıklaması
HHRSIG, Sağlık sigortasının olup olmadığı	1	Evet-Zorunlu Sigortalı (Emekli Sandığı, Bağ-Kur, SSK	HHRSIG, Sağlık sigortasının olup olmadığı	1	Evet-Sosyal Güvenlik Kurumu
	2	Evet-Diğer (Banka, vakıf vb.)		2	Evet-Özel Sağlık Sigortası
	3	Evet-İsteğe Bağlı Sigortalı		3	Evet- Banka, vakıf vb.
	4	Evet-Yeşil Kart		4	Evet-İsteğe Bağlı Sigortalı
	5	Hayır		5	Evet-Yeşil Kart
				6	Hayır

Kaynak: TÜİK, 2009; TÜİK, 2010; TÜİK, 2011a; TÜİK, 2012

Tablo 4.2 incelendiğinde 2009, 2010 ve 2011 yıllarında veri setinde hanehalkı reisinin sağlık sigortası (HHRSIG) karakteristiği ortak değerler içerirken, 2012 yılında bu değerler farklılık göstermektedir. TÜİK tarafından 2009, 2010 ve 2011 yılında derlenen veri setlerinde bu hane karakteristiği için beş farklı sağlık sigortası çeşidi belirlenirken bu sayı 2012'de derlenen veri setinde altı farklı sağlık sigortası çeşidi belirlenmiştir. Analizin düzgün bir şekilde yapılabilmesi için 2012 yılında kullanılan "sağlık sigortası" karakteristiğinin 2009, 2010 ve 2011 yılına çevrilmesi daha uygundur. Bu işlem şu şekilde gerçekleştirilmiştir; 2012 veri setlerinde bulunan HHRSIG karakteristiğinde bulunan 2 numaralı sigorta türü "Özel Sağlık Sigortası" ve 3 numaralı sigorta türü "Banka, vakıf vb", 2009, 2010 ve 2011 yıllarında bulunan 2 numaralı sigorta türü "Diğer (Banka, vakıf vb)" seçeneğine çevrilmiştir.

- Önişleme yapılan bir diğer değişken ise eğitim karakteristiğidir. 4 yıla ait verisetlerinin tamamında eğitim karakteristiği aynı şekilde derlenmiştir, fakat eğitim düzeyleri açısından birliktelik sağlaması için yeni bir gruplandırma yapılmıştır. Tablo 4.3 Hanehalkı reisinin eğitim durumunu göstermektedir.

Tablo 4.3 Hanehalkı Reisinin Eğitim Durumu

Yıllar (2009, 2010, 2011 ve 2012)		
Karakteristik ve Açıklaması	Aldığı Değerler	Değer Açıklaması
HHREgit, Hanehalkı reisinin eğitim durumu	1	Okuryazar değil
	2	Okuryazar olup bir okulu bitirmedi
	3	İlkokul
	4	İlköğretim
	5	Ortaokul
	6	Orta dengi meslek
	7	Lise
	8	Lise dengi meslek
	9	2-3 yıllık yüksekokul
	10	4 yıllık yüksekokul, fakülte
	11	Yüksek lisans, doktora (lisansüstü)

Kaynak: TÜİK, 2009; TÜİK, 2010; TÜİK, 2011a; TÜİK, 2012

Tablo 4.3'deki "İlköğretim, Ortaokul ile Orta Dengi Meslek" eğitim durumları "İlköğretim" olarak birleştirildi, benzer şekilde "Lise ile Lise Dengi Meslek" eğitim durumları "Lise" olarak birleştirildi ve daha sonra yeniden kodlama yapılmıştır.

4.2.1.2 Tüketim Karakteristiği Önişleme ve Temizleme Aşaması

Bu kısımda 2009, 2010, 2011 ve 2012 yıllarına ait Türkiye İstatistik Kurumu tarafından derlenmiş Hane Halkı Bütçe Anketi verilerinden Tüketim Karakteristiklerinin veri ön hazırlama işlemine yer verilmiştir.

1. Yeni değişkenlerin oluşturulması
2. Veri setinin tüketim kalemlerini içeren alanların tümündeki karakteristiklerin ve hanehalkı yıllık harcanabilir gelir karakteristiğinin enflasyon değerlerine göre düzenlenmesi yapılmıştır.

4.2.1.2.1 Yeni Değişkenlerin Oluşturulması

Bu işlem için aşağıdaki sıra izlenmiştir.

1. Herbir yağ grubunu tüketen haneler ve tüketmeyen haneler için yeni bir değişken oluşturulmuştur. (ZeytinyagiC, YeniliryagC, MargarinC, TereyagC) Eğer hanehalkı ilgili yağ grubunu tüketmişse “1”, tüketmemişse “0” değeri verilmiştir.

$$\text{Eğer } \begin{cases} \text{Zeytinyağı Tüketimi} = 0, & \text{ZeytinyagiC} = 0 \\ \text{Zeytinyağı Tüketimi} \neq 0, & \text{ZeytinyagiC} = 1 \end{cases}$$

$$\text{Eğer } \begin{cases} \text{Yenilir Yağ Tüketimi} = 0, & \text{YeniliryagC} = 0 \\ \text{Yenilir Yağ Tüketimi} \neq 0, & \text{YeniliryagC} = 1 \end{cases}$$

$$\text{Eğer } \begin{cases} \text{Margarin Tüketimi} = 0, & \text{MargarinC} = 0 \\ \text{Margarin Tüketimi} \neq 0, & \text{MargarinC} = 1 \end{cases}$$

$$\text{Eğer } \begin{cases} \text{Tereyağı Tüketimi} = 0, & \text{TereyagiC} = 0 \\ \text{Tereyağı Tüketimi} \neq 0, & \text{TereyagiC} = 1 \end{cases}$$

2. 40.333 kayıtlı bulunduğu veri setinde, hanehalkı yıllık harcanabilir geliri “Two-Step Kümeleme Analizi” ile 4 gelir grubuna ayrılmış ve yeni bir değişken (GelirC) olarak veri setine dâhil edilmiştir. Kümeleme işlemi sonucunda meydana gelen kümeler ve bu kümelere ait detaylı bilgiler bu bölümde yer alan “Gelir Karakteristiğine Ait İstatistikler” kısmında detaylı bir şekilde anlatılmıştır.

4.2.1.2.2 Enflasyondan Arındırma İşlemi

Bu kısımda tüketim kalemlerine ait veriler ile yıllık harcanabilir gelir verisinin enflasyondan arındırılma işlemi gerçekleştirilmiştir. Analizin düzgün bir şekilde yapılabilmesi için 2009, 2010, 2011 ve 2012 yılında derlenen hanehalkı tüketim verileri ile gelir verilerinin referans seçilen bir tarihe göre enflasyon düzenlenmesinin yapılması gerekmektedir. Bu işlem için “2012-Aralık” ayı referans olarak alınmıştır, daha sonra geriye kalan aylardaki tüketim verileri ile gelir verileri referans alınan ay'a göre belirlenen enflasyon oranları ile yeniden düzenlenmesi ile yapılmıştır. Tüketim ve gelir verilerini enflasyondan arındırmak için bir program yazılmıştır, yazılan programın kodu ve arayüzü tez çalışmasının “Ek 1” kısmında yer almaktadır.

4.2.1.3 Hane Karakteristiklerine Ait Veri Önişleme ve Temizleme Aşaması

Bu kısımda birçok değişken üzerinde önişlem yapmamıza karşın, analizde kullanmış olduğumuz iki değişkene ait veri önişleme kısmına yer verilmiştir.

- Tablo 4.4 2009, 2010, 2011 ve 2012 yıllarında hanede kullanılan yakıt türlerini göstermektedir. Tablo 4.4 incelendiğinde 2009, 2010 ve 2011 yıllarında veri setinde kullanılan YAKIT_1 karakteristiği ortak değerler içerirken 2012 yılında bu değerler farklılık göstermektedir. TÜİK tarafından 2009, 2010 ve 2011 yılında derlenen veri setlerinde bu üç hane karakteristiği için 10 farklı yakıt türü belirlenirken bu sayı 2012'de derlenen veri setinde 6 farklı yakıt türü belirlenmiştir.

Tablo 4.4 Hanede Kullanılan Birinci Yakıt Türü

Karakteristik ve Açıklaması	Yıllar (2009, 2010, 2011)		Yıllar (2012)	
	Aldığı Değerler	Değer Açıklaması	Aldığı Değerler	Değer Açıklaması
YAKIT_1, Yakıt Türü	1	Odun	1	Odun
	2	Kömür	2	Kömür
	3	Doğalgaz	3	Doğalgaz
	4	Fuel-oil	4	Elektrik
	5	Motorin-mazot	5	Tezek
	6	Gazyağı	6	Diğer
	7	LPG (Tüpgaz)		
	8	Elektrik		
	9	Tezek		
	10	Diğer		

Kaynak: TÜİK, 2009; TÜİK, 2010; TÜİK, 2011a; TÜİK, 2012

Analizin düzgün ve sağlıklı olarak yapılabilmesi açısından 2009, 2010 ve 2011 yıllarında kullanılan YAKIT_1, karakteristiğinin 2012 yılına çevrilmesi daha uygundur, çünkü 2012 yılında 4 yakıt türüne (Fuel-oil, Motorin-mazot, Gaz Yağı, LPG (Tüpgaz)) ait herhangi bir değer bulunmamaktadır. 2009, 2010 ve 2011 yıllarına ait veri setinde “Fuel-oil, Motorin-mazot, Gaz Yağı, LPG (Tüpgaz)” yakıt türleri “Diğer” seçeneğine dönüştürülmüştür. Yapılan işlemden sonra yeniden kodlama yapılmıştır.

- 2009, 2010 ve 2011 yıllarına ait veri setinde haneye ait otomobil olup olmadığına ait bir hane karakteristiği bulunmaktadır (OTOMOBİL, 0 Yok, 1 Var).

$$\text{Eğer } \begin{cases} OTOMOBİL = 0, \text{ Haneye ait otomobil yok} \\ OTOMOBİL = 1, \text{ Haneye ait otomobil var} \end{cases}$$

Bu sorunun cevabına bağılı olarak eğer haneye ait otomobil var ise bir sonraki karakteristik olan OTOADET karakteristiğinde adet olarak sayısı belirtilmektedir. Diğer yandan bu 2 hane halkı karakteristiği TÜİK tarafından 2012 yılında birleştirilmiş ve OTOADET olarak belirlenmiştir, eğer hanede otomobil yok ise 0 (Sıfır) değerini, var ise adet olarak miktar belirtilmiştir. Bu sorunun çözümünde 2012 yılına OTOMOBİL adı ile yeni bir sütun eklenerek çözülmüştür. Yeni oluşan sütunun diğer yıllar ile uyumlu olması için şu işlem yapılmıştır,

$$\text{Eğer } \begin{cases} \text{OTOADET} = 0 \text{ ise } \text{OTOMOBİL} = 0 \\ \text{OTOADET} \neq 0 \text{ ise } \text{OTOMOBİL} = 1 \end{cases}$$

olacak şekilde OTOMOBİL sütununa yeni kodlama yapılmıştır.

Fert, hane ve tüketim karakteristiklerinden Tablo 4.5'deki değişkenler çalışmanın analizinde kullanılmıştır. Bu tabloda yer alan karakteristikler ve aldıkları değerler önişlenmiş ve temizlenmiştir.

Tablo 4.5 Çalışmanın Analizinde Kullanılan Karakteristikler

S. No	Karakteristiğin				Tipi
	İsmi	Açıklaması	Aldığı Değerler		
1	HHRCins	Hanehalkı reisinin cinsiyeti	1	Erkek	Flag
			2	Bayan	
2	HHRYas	Hanehalkı reisinin	1	0-5 yaş	Continious
			2	6-14 yaş	
			3	15-19 yaş	
			4	20-24 yaş	
			5	25-29 yaş	
			6	30-34 yaş	
			7	35-39 yaş	
			8	40-44 yaş	
			9	45-49 yaş	
			10	50-54 yaş	
			11	55-59 yaş	
			12	60-64 yaş	
			13	65+yaş	

S. No	Karakteristiğın				Tipi
	İsmi	Açıklaması	Aldığı Değerler		
3	HHREgit	Hanehalkı reisinin eğitim durumu	1	Okuryazar değil	Ordinal
			2	Okuryazar olup bir okulu bitirmedi	
			3	İlkokul	
			4	İlköğretim	
			5	Lise	
			6	2-3 yıllık yüksekokul	
			7	4 yıllık yüksekokul, fakülte	
			8	Yüksek lisans, doktora (lisansüstü)	
4	HHTipi	Hanehalkı tipi	1	Tek çocuklu çekirdek aile	Nominal
			2	İki çocuklu çekirdek aile	
			3	Üç veya daha fazla çocuklu çekirdek aile	
			4	Çocuksuz çift	
			5	Ataerkil veya geniş aile	
			6	Tek yetişkinli aile	
			7	Birarada yaşayan kişiler	
5	HHRMedDur	Hanehalkı reisinin medeni durumu	1	Hiç evlenmedi	Nominal
			2	Evli	
			3	Eşi öldü	
			4	Boşandı	
6	HHRSig	Hanehalkı reisinin sigorta durumu	1	Evet-Zorunlu Sigortalı (Emekli Sandığı, Bağ-Kur, SSK)	Nominal
			2	Evet-Diğer (Banka, vakıf vb.)	
			3	Evet-İsteğe Bağlı Sigortalı	
			4	Evet-Yeşil Kart	
			5	Hayır	
7	Mulkiyet	Konutta mülkiyet durumu	1	Ev sahibi	Nominal
			2	Kiracı	
			3	Lojman	
			4	Diğer	
8	Konutiki	İkinci konut sahipliği	1	Evet	Flag
			2	Hayır	

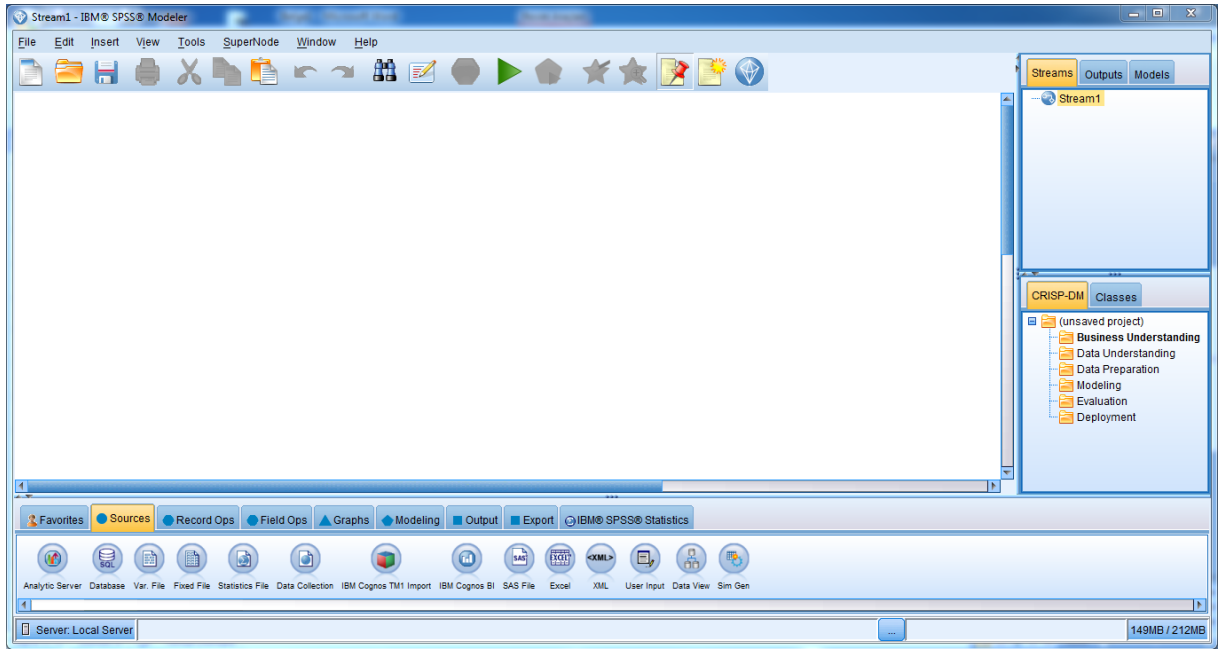
S. No	Karakteristiğın				Tipi
	İsmi	Açıklaması	Aldığı Değerler		
9	Otomobil	Otomobil sahipliği	0	Yok	Flag
			1	Var	
10	HHB	Hanehalkı büyüklüğü	0-99		Continious
11	Kırkntkd	Kır kent durumu	1	Kır	Nominal
			2	Kent	
12	GelirC	Hanehalkı yıllık harcanabilir geliri	Düşük		Ordinal
			Orta Alt		
			Orta Üst		
			Yüksel		
13	ZorBanka	Oturduğu yer itibari ile bankaya ulaşma zorluğu	1	Çok zor	Ordinal
			2	Zor	
			3	Kolay	
			4	Çok kolay	
14	MusKonut	Müstakil konut sahipliği	0	Yok	Flag
			1	Var	
15	YAKIT_1	Konutta en çok kullanılan birinci yakıt türü	1	Odun	Nominal
			2	Kömür	
			3	Doğalgaz	
			4	Elektrik	
			5	Tezek	
			6	Diğer	
16	Elliyasuzeribirey	Hanede Elli yaş üzeri birey olup olmaması	0	Yok	Flag
			1	Var	
17	SıfırBesyas	Hanede Sıfır Beş yaş arası birey olup olmadığı	0	Yok	Flag
			1	Var	
18	ZeytinyagiC	Hanehalkının zeytinyağı tüketip tüketmemesi	0	Tüketmiyor	Flag
			1	Tüketiyor	

S. No	Karakteristiğın				Tipi
	İsmi	Açıklaması	Aldığı Değerler		
19	YeniliryagC	Hanehalkının yenilebilir yağ (mısır, ayçiçeği, kanola vb.) tüketip tüketmemesi	0	Tüketmiyor	Flag
			1	Tüketiyor	
20	TereyagC	Hanehalkının tereyağı tüketip tüketmemesi	0	Tüketmiyor	Flag
			1	Tüketiyor	
21	MargarinC	Hanehalkının margarin ve diğer bitkisel yağları tüketip tüketmemesi	0	Tüketmiyor	Flag
			1	Tüketiyor	

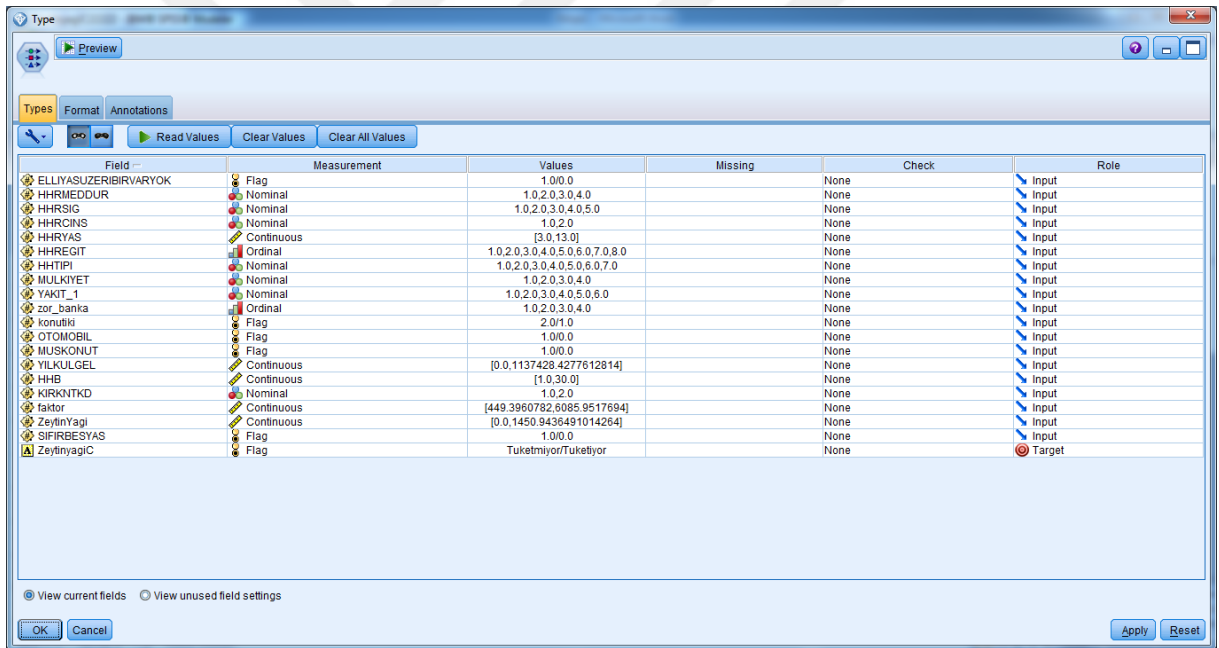
Kaynak: TÜİK, 2009; TÜİK, 2010; TÜİK, 2011a; TÜİK, 2012

4.3 Uygulamada Kullanılan Yöntem ve Veri Madenciliği Yazılımı

Veri madenciliği modelleme aşamasında kullanılan yöntem “Karar Ağaçları”, verilerin analizinde ise IBM SPSS Modeler veri madenciliği programı kullanılmıştır. IBM SPSS Modeler, hızlı bir şekilde öngörü modelleri geliştirilebilen, bireylerin karar alma yeteneğini yükselten ve bunları iş dünyasının içerisine entegre edebilen veri madenciliği araçları bütünüdür. CRISP-DM modeli etrafında tasarlanmış, daha iyi iş sonuçlarına götüren IBM SPSS Modeler tüm veri madenciliği sürecini desteklemektedir. YSA, Birliktelik Kuralları, Sınıflandırma modelleme seçeneklerini içermektedir. Her bir model, ayrı güce sahip ve problemlerin ayrı türleri için uygundur. Şekil 4.1 "IBM SPSS Modeler" programının ana ekranını göstermektedir. Şekil 4.1'in alt kısmında görülen Sources, Record Options, Modeling ve diğer sekmeler veri madenciliği uygulamasının gerçekleştirildiği ve yürütüldüğü sekmelerdir. Şeklin ortasında yer alan büyük alan ise veri madenciliği sürecinin ve veri akışının tanımlandığı alandır. Bütün eklenen nesneler bu alanda görülecektir.



Şekil 4.1 IBM SPSS Modeler Ana Ekranı



Şekil 4.2 IBM SPSS Modeler Type Modülü

Şekil 4.2, kullanılan değişkenlerin türlerinin belirlendiği “Type” modülünü göstermektedir. Burada her değişkenin türleri ve modelleme aşamasında modele ne şekilde dâhil olacağı belirlenebilmektedir. Ayrıca “Values” kısmında ilgili değişkenin hangi aralıkta değerler aldığı görülebilmektedir.

4.4 Analiz Sonucunda Elde Edilen Bulgular

4.4.1 Hedef Değişkenlere Ait Temel İstatistikler

Bu bölümde “zeytinyağı, yenilebilir sıvı yağ, margarin ve diğer bitkisel yağlar, tereyağı” tüketim değerinden elde edilen ve “yıllık harcanabilir gelir” karakteristiğinden “Two-Step Kümeleme Analizi” ile oluşturulan GelirC değişkenine ait tanımlayıcı istatistikler hesaplanmıştır.

Zeytinyağı Tüketimine Ait Tanımlayıcı İstatistikler:

TÜİK tarafından derlenen veri setinde yer alan zeytinyağı tüketim karakteristiği yalnızca zeytinyağı verisini içermektedir. Tablo 4.6, zeytinyağı tüketen hanelere, zeytinyağı tüketen hanelerde OECD Eş Değer Fert Büyüklüğüne (kişi başı) göre ve hanehalkı genel zeytinyağı tüketimine ait temel istatistiklerini göstermektedir.

Tablo 4.6 Zeytinyağı Tüketimine Ait Tanımlayıcı İstatistikler

	Eş Değer Fert Büyüklüğüne (OECD) Göre Tüketim	Zeytinyağı Tüketen Hanelerde Tüketim	Genel Tüketim
Toplam Hane Sayısı	4.212	4.212	40.033
Ortalama	21,708	40,947	4,308
Minimum	0,442	1,016	0,000
Maksimum	594,546	1450,944	1450,944
Standart Sapma	33,816	65,488	24,678
Medyan	13,426	26,085	0,000

Tablo 4.6’deki tanımlayıcı istatistiklere baktığımızda, zeytinyağı tüketen 4.212 hane görülmektedir. OECD eş değer büyüklüğüne göre zeytinyağı tüketen hanelerde kişi başı ortalama tüketim 21,708 TL olarak görülmektedir, en az tüketim 0,442 TL, en yüksek tüketim ise 594,546 TL olarak görülmektedir. Zeytinyağı tüketen hanelerde ortalama tüketim 40,947 TL, en düşük tüketim 1,016 TL, en yüksek tüketim 1450,944 TL olarak görülmektedir. Zeytinyağı tüketen ve tüketmeyen bütün haneler dâhil edildiği zaman ortalama tüketim 4,308 TL, en yüksek tüketim 1450,644 TL olarak görülmektedir.

Diğer Yenilebilir Sıvı Yağ (Zeytinyağı hariç bitkisel yağlar: Mısır, ayçiçeği, pamuk, kolza, aspir, vb) Tüketimine Ait İstatistikler:

TÜİK tarafından derlenen veri setinde yer alan diğer yenilebilir sıvı yağ tüketim karakteristiği mısır yağı, ayçiçeği yağı, pamuk yağı, soya yağı, yer fıstığı yağı, ceviz yağı, vb. verilerini içerirken morina veya trança balığı karaciğer yağını içermemektedir. Tablo 4.7,

diğer yenilebilir sıvı yağ tüketen hanelere, yenilebilir sıvı yağ tüketen hanelerde OECD Eş Değer Fert Büyüklüğüne (kişi başı) göre ve hanehalkı genel yenilebilir sıvı yağ tüketimine ait temel istatistiklerini göstermektedir.

Tablo 4.7 Diğer Yenilebilir Sıvı Yağ Tüketimine Ait Temel İstatistikler

	Eş Değer Fert Büyüklüğüne (OECD) Göre Tüketim	Diğer Yenilebilir Sıvı Yağ Tüketen Hanelerde Tüketim	Genel Tüketim
Toplam Hane Sayısı	21.897	21.897	40.033
Ortalama	11,506	25,117	13,738
Minimum	0,147	0,339	0,000
Maksimum	205,099	568,940	568,940
Standart Sapma	8,677	19,593	19,139
Medyan	9,396	20,078	7,527

Tablo 4.7’deki tanımlayıcı istatistiklere baktığımızda, diğer yenilebilir sıvı yağ tüketen 21.897 hane görülmektedir. OECD eş değer büyüklüğüne göre diğer yenilebilir sıvı yağ tüketen hanelerde kişi başı ortalama tüketim 11.506 TL olarak görülmektedir, en az tüketim 0,147 TL, en yüksek tüketim ise 205,099 TL olarak görülmektedir. Diğer yenilebilir sıvı yağ tüketen hanelerde ortalama tüketim 25,117 TL, en düşük tüketim 0,339 TL, en yüksek tüketim 568,940 TL olarak görülmektedir. Diğer yenilebilir sıvı yağ tüketen ve tüketmeyen bütün haneler dâhil edildiği zaman ortalama tüketim 13,738 TL, en yüksek tüketim 568,940 TL olarak görülmektedir.

Margarin ve Diğer Bitkisel Yağ Tüketimine Ait Tanımlayıcı İstatistikler:

TÜİK tarafından derlenen veri setinde yer alan margarin ve diğer bitkisel yağ tüketim karakteristiği margarin (diyet margarin dahil) ve diğer bitkisel katı yağlar (yer fıstığı ezmesi dahil) verilerini içermektedir. Tablo 4.8, margarin ve diğer bitkisel yağ tüketen hanelere, margarin ve diğer bitkisel yağ tüketen hanelerde OECD Eş Değer Fert Büyüklüğüne (kişi başı) göre ve hanehalkı genel margarin ve diğer bitkisel yağ tüketimine ait temel istatistiklerini göstermektedir.

Tablo 4.8 Margarin ve Diğer Bitkisel Yağ Tüketimine Ait Tanımlayıcı İstatistikler

	Eş Değer Fert Büyükliğüne (OECD) Göre Tüketim	Margarin ve Diğer Bitkisel Yağ Tüketen Hanelerde Tüketim	Genel Tüketim
Toplam Hane Sayısı	17.679	17.679	40.033
Ortalama	2,687	6,095	2,692
Minimum	0,147	0,546	0,000
Maksimum	58,206	144,734	144,734
Standart Sapma	3,268	9,011	6,709
Medyan	1,817	3,892	0,000

Tablo 4.8'deki tanımlayıcı istatistiklere baktığımızda, margarin ve diğer bitkisel yağ tüketen 17.679 hane görülmektedir. OECD eş değer büyüklüğüne göre margarin ve diğer bitkisel yağ tüketen hanelerde kişi başı ortalama tüketim 2,687 TL olarak görülmektedir, en az tüketim 0,147 TL, en yüksek tüketim ise 58,206 TL olarak görülmektedir. Margarin ve diğer bitkisel yağ tüketen hanelerde ortalama tüketim 6,095 TL, en düşük tüketim 0,546 TL, en yüksek tüketim 144,734 TL olarak görülmektedir. Margarin ve diğer bitkisel yağ tüketen ve tüketmeyen bütün haneler dâhil edildiği zaman ortalama tüketim 2,692 TL, en yüksek tüketim 144,734 TL olarak görülmektedir.

Tereyağı Tüketimine Ait Tanımlayıcı İstatistikler:

TÜİK tarafından derlenen veri setinde yer alan tereyağı karakteristiği tereyağı ve tereyağı ürünleri (butter oil-sade yağ (tereyağı içerisindeki su ve katı maddelerin (yağ hariç) ısıtılarak arındırılmasıyla oluşturuluyor), saf yağ (ghee), vb.) verilerini içermektedir. Tablo 4.9, tereyağı tüketen hanelere, tereyağı tüketen hanelerde OECD Eş Değer Fert Büyükliğüne (kişi başı) göre ve hanehalkı genel tereyağı tüketimine ait temel istatistiklerini göstermektedir.

Tablo 4.9 Tereyağı Tüketimine Ait Tanımlayıcı İstatistikler

	Eş Değer Fert Büyükliğüne (OECD) Göre Tüketim	Tereyağı Tüketen Hanelerde Tüketim	Genel Tüketim
Toplam Hane Sayısı	9.912	9.912	40.033
Ortalama	9,032	18,753	4,643
Minimum	0,105	0,252	0,000
Maksimum	265,029	582,734	582,734
Standart Sapma	11,829	26,335	15,402
Medyan	5,669	12,077	0,000

Tablo 4.9'deki tanımlayıcı istatistiklere baktığımızda, tereyağı tüketen 9.912 hane görülmektedir. OECD eş değer büyüklüğüne göre tereyağı tüketen hanelerde kişi başı ortalama tüketim 9,032 TL olarak görülmektedir, en az tüketim 0,105 TL, en yüksek tüketim ise 265,029 TL olarak görülmektedir. Tereyağı tüketen hanelerde ortalama tüketim 18,753 TL, en düşük tüketim 0,252 TL, en yüksek tüketim 582,734 TL olarak görülmektedir. Tereyağı tüketen ve tüketmeyen bütün haneler dâhil edildiği zaman ortalama tüketim 4,643 TL, en yüksek tüketim 582,744 TL olarak görülmektedir.

Gelir Karakteristiğine Ait İstatistikler:

Veri setinde bulunan haneye ait "Yıllık Harcanabilir Gelir" verisi "Two Step Kümeleme Analizi" ile 4 gruba ayrılmıştır. Tablo 4.10, Kümeleme sonucu ortaya çıkan kümelere ait bilgileri göstermektedir.

Tablo 4.10 Gelir Grupları

Gelir Grubu	Hane Sayısı	Hane Yüzdesi	Ortalama Yıllık Harcanabilir Gelir	Ortalama Aylık Harcanabilir Gelir
Düşük Gelir	19.807	%49,5	14.316,49 TL	1.193,04 TL
Orta Alt Gelir	15.836	%39,6	33.044,24 TL	2.753,68 TL
Orta Üst Gelir	3.973	%9,9	69.647,31 TL	5.803,94 TL
Yüksek Gelir	417	%1	214.634,12 TL	17.886,17 TL

Yüksek gelire sahip 417 hane bulunmaktadır, tüm hanelerin %1'ini oluşturmakta ve ortalama yıllık harcanabilir gelir 214.634,12 TL (17.886,17 TL/Ay) olarak bulunmuştur. Orta üst gelire sahip 3973 hane bulunmaktadır, tüm hanelerin %9,9'unu oluşturmakta ve ortalama yıllık harcanabilir gelir 69.647,31 TL (5.803,94 TL/Ay) olarak bulunmuştur. Orta alt gelire sahip 15.836 hane bulunmaktadır, tüm hanelerin %39,6'sını oluşturmakta ve ortalama yıllık harcanabilir gelir 33.044,24 TL (2.753,68 TL/Ay) olarak bulunmuştur. Düşük sahip 19.807 hane bulunmaktadır, tüm hanelerin %49,5'ini oluşturmakta ve ortalama yıllık harcanabilir gelir 14.316,49 TL (1.193,04 TL/Ay) olarak bulunmuştur.

4.4.2 Zeytinyağı Tüketimini Etkileyen Faktörler

Kurulan bu modelde amaç, Hanehalkı Bütçe Anketi verileri kullanılarak, hanehalkının tüketmiş olduğu zeytinyağı tüketimini etkileyen faktörlerin ve hanehalklarının tüketim durumlarının Karar Ağacı yöntemi ile belirlenmesidir. Hanehalkı zeytinyağı tüketim değişkeni bağımlı değişkendir, yani hedef değişkendir. Zeytinyağı tüketimine etki eden

karakteristikler, Karar Ağacı yöntemlerinden CHAID yöntemi kullanılarak bulunmuştur. Kurulan model sonucunda CHAID yönteminin ürettiği sonuca göre tüketimi etkileyen karakteristiker Tablo 4.11’da gösterilmiştir.

Tablo 4.11 Zeytinyağı Tüketimini Etkileyen Karakteristikler

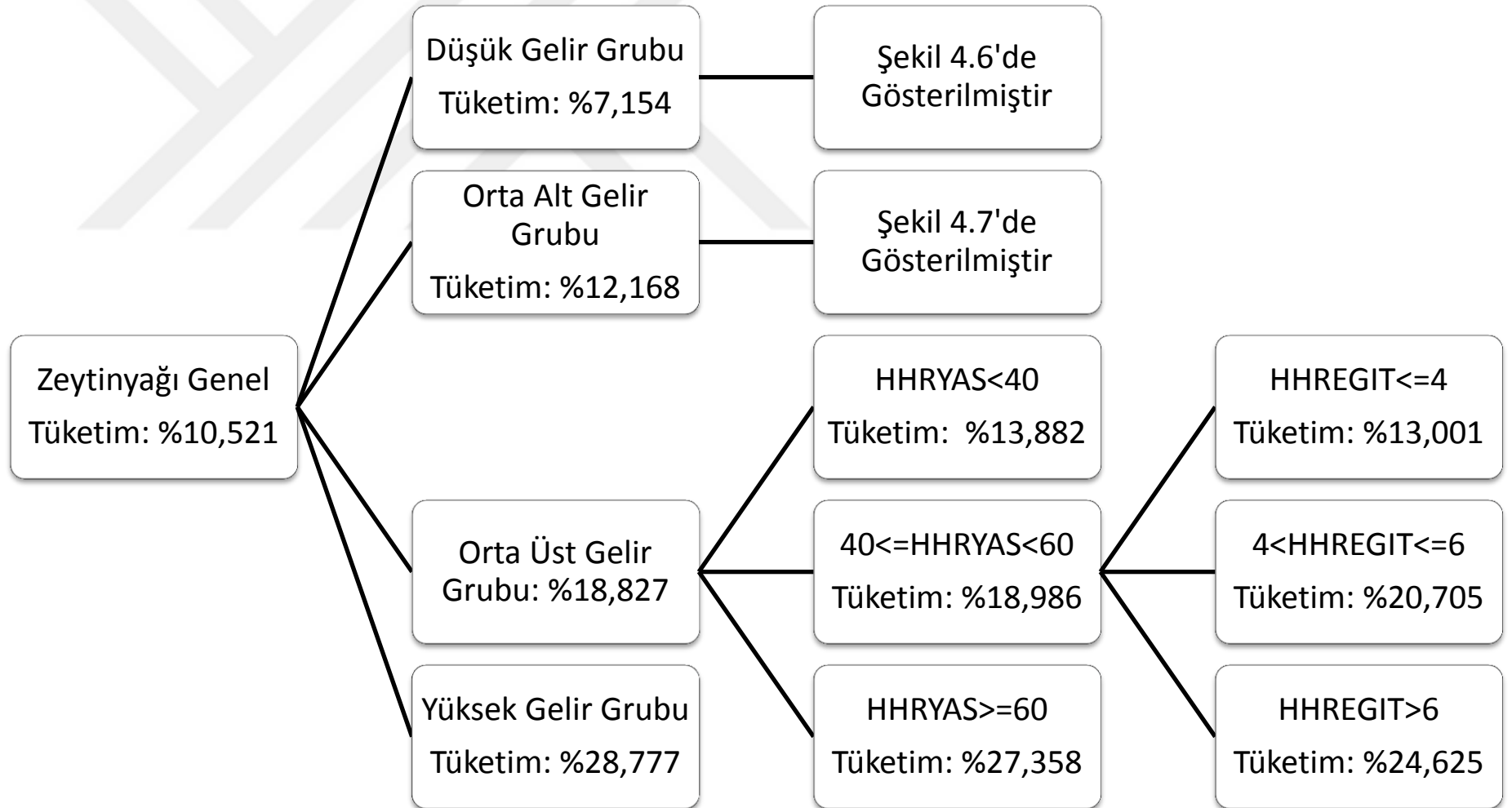
S. No	Belirleyicinin İsmi	Açıklama
1	GelirC	Gelir
2	HHB	Hanehalkı büyüklüğü
3	Elliyasuzeribirey	Hanede Elli yaş üzeri birey olup olmaması
4	HHREgit	Hanehalkı reisinin eğitim durumu
5	HHRSig	Hanehalkı reisinin sigorta durumu
6	HHRYas	Hanehalkı reisinin yaşı
7	Otomobil	Otomobil sahipliği
8	Kırkntkd	Kır kent durumu
9	SifirBesyas	Hanede Sıfır Beş yaş arası birey olup olmadığı
10	Yakıt_1	Konutta en çok kullanılan birinci yakıt türü

Meydana gelen karar ağacını yorumladığımızda aşağıdaki sonuçlara ulaşılmıştır. Tablo 4.12 Hanehalklarının zeytinyağı tüketim durumlarını göstermektedir.

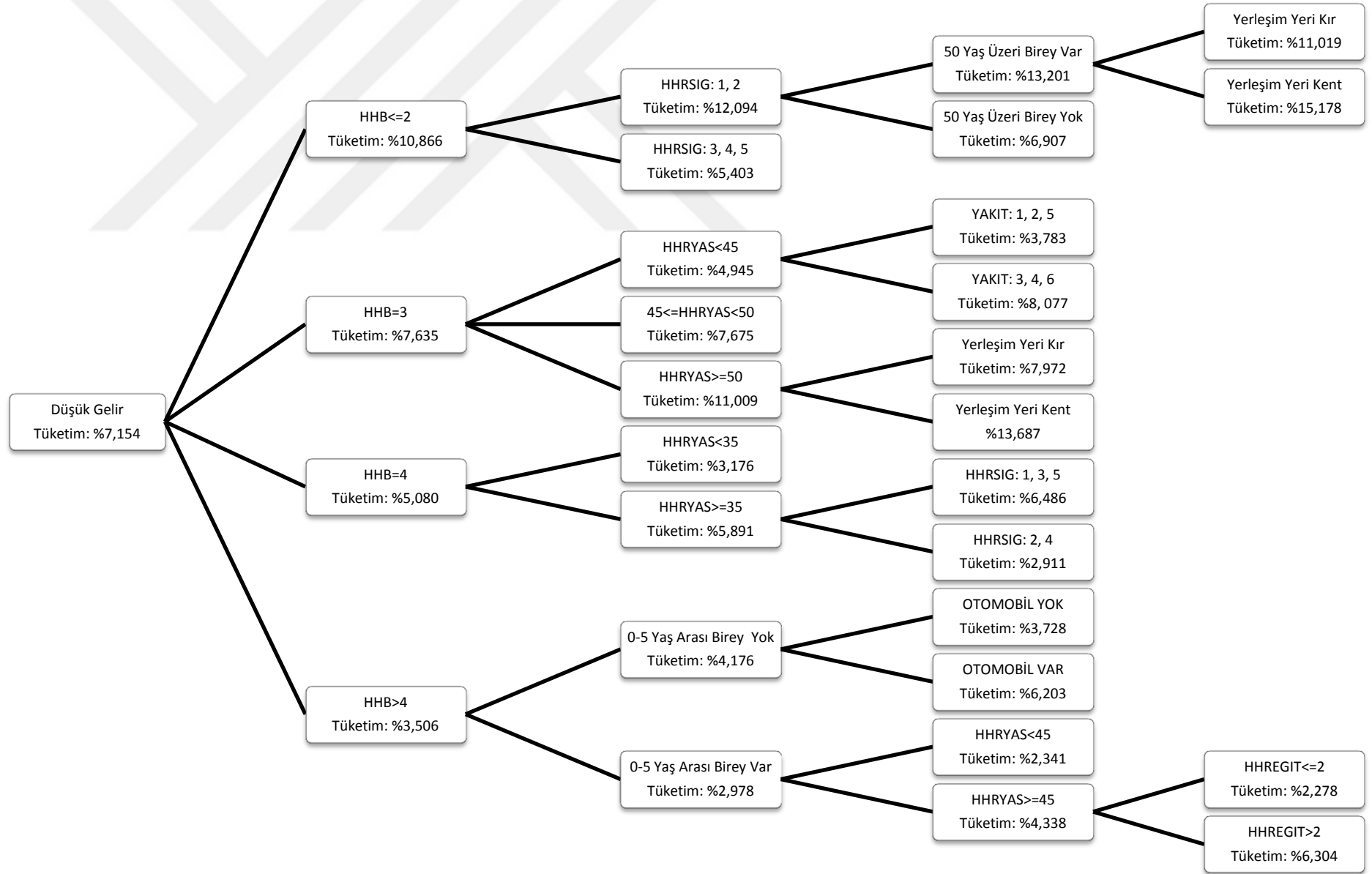
Tablo 4.12 Hanehalkı Zeytinyağı Tüketim Durumu

Zeytinyağı Tüketim Durumu	Yüzde	Sayı
Zeytinyağı Tüketmeyen Hane	89,479	35.821
Zeytinyağı Tüketen Hane	10,521	4.212
Toplam	100	40.033

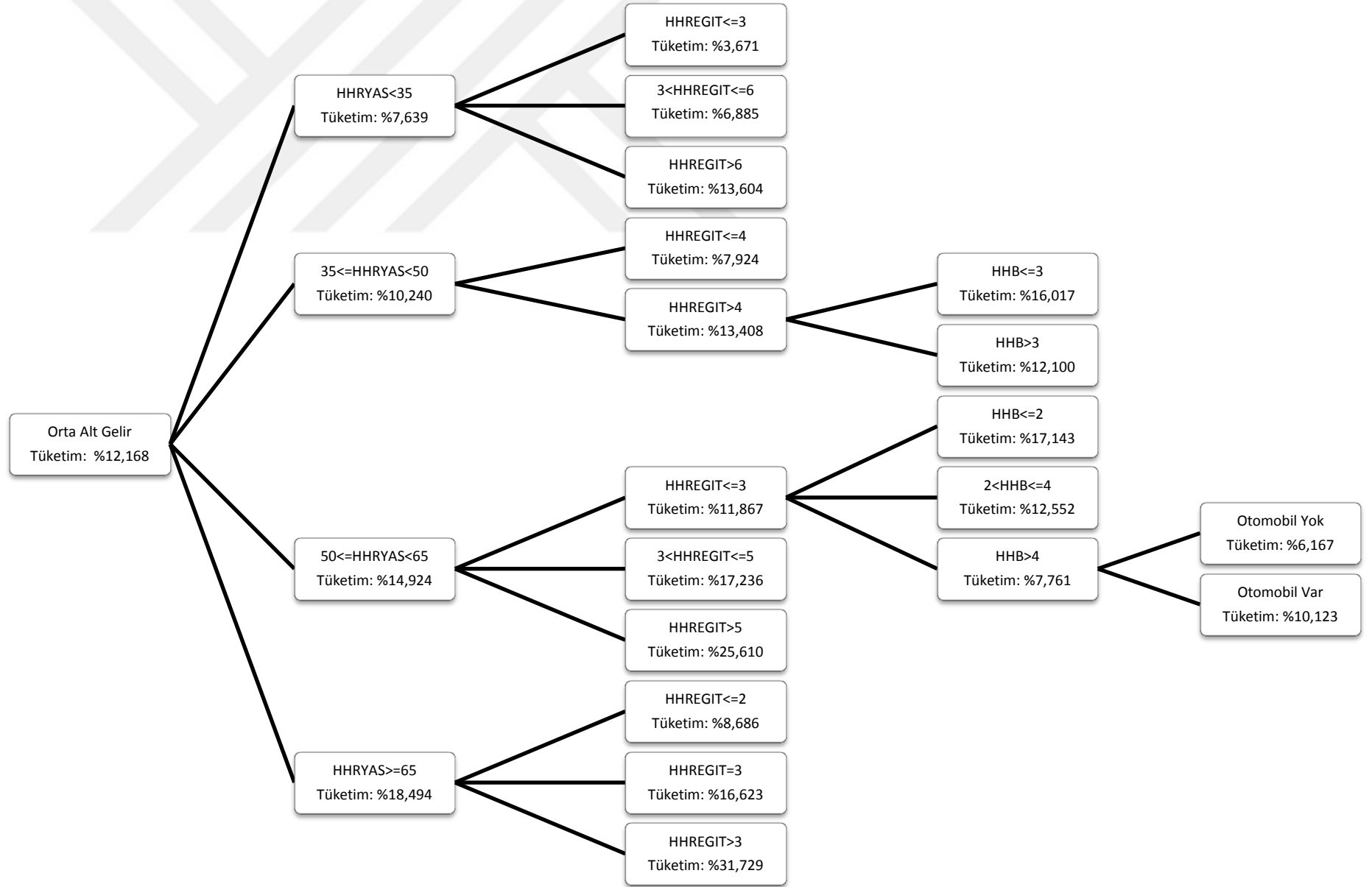
Tablo 4.12’ye göre toplam kayıt sayısı 40.333’dür. Zeytinyağı tüketen hane sayısı 4.212 olup toplam hanelerin yaklaşık %10,5’luk kısmına gelmektedir. Zeytinyağı tüketmeyen hane sayısı 35.821 olup toplam hanelerin yaklaşık %89,5’ine karşılık gelmektedir.



Şekil 4.5 Zeytinyağı Karar Ağacı-1



Şekil 4.6 Zeytinyağı Karar Ağacı-2



Şekil 4.7 Zeytinyağı Karar Ağacı-3

Modelleme sonucunda meydana gelen Karar Ağacı'nı yorumladığımızda aşağıdaki çıkarımlara ulaşılmıştır.

- 1) Karar ağacının açık görünümü Şekil 4.5'de gösterilmektedir. Meydana gelen ağaç görünümüne baktığımızda ağacı ilk bölen değişkenin Gelir olduğunu görüyoruz. Kök düğümde zeytinyağı tüketen hanelerin oranı %10,52 olarak görülmüştür. Zeytinyağı tüketimine ait karar ağacının açık görünümü Şekil 4.5, 4.6 ve 4.7'de gösterilmektedir.
- 2) Şekil 4.5'i incelediğimizde, düşük gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı %7,154 olarak görülmektedir. Bu sonuca göre düşük gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı genel tüketim oranına göre azalmıştır.
- 3) Orta Alt gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı %12,168 olarak görülmektedir. Bu sonuca göre orta alt gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı genel tüketim oranına göre artmıştır.
- 4) Orta Üst gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı %18,827 olarak görülmektedir, bu durum genel tüketim oranının (kök düğüm) yaklaşık 1,5 katı kadardır. Bu sonuca göre orta üst gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı genel tüketim oranına göre artmıştır.
- 5) Yüksek gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı %28,777 olarak görülmektedir, bu durum genel tüketim oranının (Kök düğüm) yaklaşık 2,5 katı kadardır. Bu sonuca göre yüksek gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı genel tüketim oranına göre artmıştır.
- 6) Yüksek gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı, orta üst gelire sahip hanelerin bulunduğu düğümdeki zeytinyağı tüketen hanelerin oranının yaklaşık 1,5 katı, orta alt gelire sahip hanelerin bulunduğu düğümdeki zeytinyağı tüketen hanelerin oranının yaklaşık 2 katı, düşük gelire sahip hanelerin bulunduğu düğümdeki zeytinyağı tüketen hanelerin oranının yaklaşık 4 katıdır.
- 7) Şekil 4.5'e göre, hanelerin geliri arttıkça zeytinyağı tüketen hanelerin oranı artmaktadır sonucuna ulaşılmaktadır.

Zeytinyağı tüketiminin orta üst ve yüksek gelir grubu düğümü altındaki karar ağacı yapısı Şekil 4.5'de gösterilmiştir. Bu düğümde sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Orta üst gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketimi konusunda hane halkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmüştür. Orta

üst gelire sahip hanelerin bulunduğu düğümün ve tüm alt düğümlerin detaylı özellikleri Şekil 4.5’de belirtilmiştir.

- 2) Orta üst gelire sahip, hanehalkı reisinin yaşının 40'ın altında olduğu düğümde zeytinyağı tüketen hanelerin oranının %13,882 olduğu görülmektedir, bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 3) Orta üst gelire sahip, hanehalkı reisinin yaşının 40 ile 59 arasında olduğu düğümde zeytinyağı tüketen hanelerin oranının %18,986 olduğu görülmektedir, bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranının yaklaşık 1,5 katıdır.
- 4) Orta üst gelire sahip, hanehalkı reisinin yaşının 60 ve 60'ın üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranının %27,358 olduğu görülmektedir, bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranının yaklaşık 2,5 katından fazladır.
- 5) Orta üst gelir grubunda yer alan hanelerde hanehalkı reisinin yaşı arttıkça zeytinyağı tüketen hanelerin oranı atmaktadır.
- 6) Orta üst gelire sahip, hanehalkı reisinin yaşının 40 ile 59 arasında olduğu düğümde zeytinyağı tüketimi konusunda hanehalkı reisinin eğitim durumu karakteristiğinin ayırt edici bir özellik olduğu görülmüştür.
- 7) Orta üst gelire sahip, hanehalkı reisinin yaşının 40 ile 59 arasında olduğu, hanehalkı reisinin eğitim durumunun ilköğretim veya altında olduğu düğümde zeytinyağı tüketen hanelerin oranının %13,001 olduğu görülmüştür. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 8) Orta üst gelire sahip, hanehalkı reisinin yaşının 40 ile 59 arasında olduğu, hanehalkı reisinin eğitim durumunun 2-3 yıllık yüksek okul veya lise olduğu düğümde zeytinyağı tüketen hanelerin oranının %20,705 olduğu görülmüştür. Bu düğümde, zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır ve zeytinyağı tüketen hanelerin oranı, genel tüketim oranının yaklaşık 2 katıdır.
- 9) Orta üst gelire sahip, hanehalkı reisinin yaşının 40 ile 59 arasında olduğu, hanehalkı reisinin eğitim durumunun fakülte/yüksekokul veya lisansüstü olduğu düğümde zeytinyağı tüketen hanelerin oranının %24,625 olduğu görülmüştür, bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranının yaklaşık 2,5 katıdır.
- 10) Orta üst gelire sahip, hanehalkı reisinin yaşının 40 ile 59 arasında olduğu hanelerde hanehalkı reisinin eğitim düzeyi arttıkça zeytinyağı tüketen hanelerin oranı atmaktadır.

- 11) Yüksek gelire sahip hanelerin bulunduğu düğümde alt düğüm oluşturacak herhangi bir karakteristik gözlenmemiştir. Bu durum Şekil 4.5’de görülmektedir.

Zeytinyağı tüketiminin düşük gelir grubu düğümü altındaki karar ağacı yapısı Şekil 4.6’da gösterilmiştir. Bu düğümünden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Düşük gelire sahip olan hanelerin bulunduğu düğümde zeytinyağı tüketimi konusunda hanehalkı büyüklüğü karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Bu yorumu ve Düşük gelir grubunda yer alan hanelerin bulunduğu düğümlerdeki tüm dallanmaları ve detaylı bilgiler için Şekil 4.6 incelenebilir.
- 2) Düşük gelire sahip, hanehalkı büyüklüğünün 1 veya 2 olduğu düğümde zeytinyağı tüketen hanelerin oranı %10,866 olarak görülmektedir. Bu durum genel tüketim oranı ile yakınlık göstermektedir.
- 3) Düşük gelire sahip, hanehalkı büyüklüğünün 3 olduğu düğümde zeytinyağı tüketen hanelerin oranı %7,635 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Düşük gelire sahip, hanehalkı büyüklüğünün 4 olduğu düğümde zeytinyağı tüketen hanelerin oranı %5,080 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 5) Düşük gelire sahip, hanehalkı büyüklüğünün 5 ve üzeri olduğu hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı %3,506 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 6) Düşük gelire sahip hanelerin bulunduğu tüm düğümleri incelediğimizde, hanehalkı büyüklüğünün 1 veya 2 olduğu hanelerin bulunduğu düğümde zeytinyağı tüketen hanelerin oranı genel tüketim oranına göre küçük bir miktar artış gösterirken, diğer bütün düğümlerde zeytinyağı tüketen hanelerin oranı azalmıştır.
- 7) Düşük gelir grubunda yer alan hanelerde, hanehalkı büyüklüğü arttıkça zeytinyağı tüketen hanelerin oranı azalmaktadır.
- 8) Düşük gelire sahip, hanehalkı büyüklüğünün 1 veya 2 olduğu düğümde zeytinyağı tüketimi konusunda hanehalkı reisinin sigorta durumu karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 9) Düşük gelire sahip, hanehalkı büyüklüğünün 3 olduğu düğümde zeytinyağı tüketimi konusunda hanehalkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmüştür.

- 10) Düşük gelire sahip, hanehalkı büyüklüğünün 4 olduğu düğümde de zeytinyağı tüketimi konusunda hanehalkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmüştür.
- 11) Düşük gelire sahip, hanehalkı büyüklüğünün 5 ve 5'in üzerinde olduğu düğümde zeytinyağı tüketimi konusunda hanede Sıfır-Beş yaş birey sahipliği karakteristiğinin ayırt edici bir özellik olduğu görülmüştür.
- 12) Düşük gelire sahip hanelerin bulunduğu düğümde bir sonraki dallanmayı belirleyici karakteristiğin hanehalkı büyüklüğü olduğunu daha önce belirtmiştik. Farklı hanehalkı büyüklüklerinde dallanmayı gerçekleştiren karakteristiklerin değiştiği görülmüştür. Bu durum 8-11 arasındaki çıkarımları okuyarak görülebilir.
- 13) Düşük gelire sahip, hanehalkı büyüklüğünün 1 veya 2 olduğu, hanehalkı reisinin sigorta durumu zorunlu sigortalı veya diğer sağlık sigortası (banka, vakıf, özel sağlık sigortası vb) olduğu düğümde zeytinyağı tüketen hanelerin oranı %12,094 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 14) Düşük gelire sahip, hanehalkı büyüklüğünün 1 veya 2 olduğu, hanehalkı reisinin sigorta durumu isteğe bağlı sigortalı veya yeşil kartlı veya sigortasız olduğu düğümde zeytinyağı tüketen hanelerin oranı %5,403 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 15) Düşük gelire sahip, hanehalkı büyüklüğünün 1 veya 2 olduğu, hanehalkı reisinin sigorta durumu zorunlu sigortalı veya diğer sağlık sigortası (banka, vakıf, özel sağlık sigortası vb) olduğu düğümde zeytinyağı tüketimi konusunda hane de Elli yaş ve üzeri birey sahipliği karakteristiğinin ayırt edici bir özellik olarak görülmektedir.
- 16) Düşük gelire sahip, hanehalkı büyüklüğünün 1 veya 2 olduğu, hanehalkı reisinin sigorta durumu zorunlu sigortalı veya diğer sağlık sigortası (banka, vakıf, özel sağlık sigortası vb) olduğu, hanede Elli yaş ve üzeri bireyin olmadığı düğümde zeytinyağı tüketen hanelerin oranı %6,907 olarak görülmektedir, bu durum genel tüketim oranına göre azalmıştır.
- 17) Düşük gelire sahip, hanehalkı büyüklüğünün 1 veya 2 olduğu, hanehalkı reisinin sigorta durumu zorunlu sigortalı veya diğer sağlık sigortası (banka, vakıf, özel sağlık sigortası vb) olduğu, hanede Elli yaş ve üzeri bireyin olduğu düğümde zeytinyağı tüketen hanelerin oranı %13,201 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Elli yaş ve üzeri bireyin var olduğu düğümde zeytinyağı tüketen hanelerin oranı, Elli yaş ve üzeri bireyin olmadığı düğümüne göre daha fazladır ve neredeyse 2 katına yakındır.

- 18) Gelir grubunun düşük olduğu, hanehalkı büyüklüğünün 3, hanehalkı reisinin yaşının 45'in altında olduğu düğümde zeytinyağı tüketen hanelerin oranı %4,945 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 19) Gelir grubunun düşük olduğu, hanehalkı büyüklüğünün 3, hanehalkı reisinin yaşının 45 ile 49 arasında olduğu düğümde zeytinyağı tüketen hanelerin oranı %7,675 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 20) Gelir grubunun düşük olduğu, hanehalkı büyüklüğünün 3, hanehalkı reisinin yaşının 50 ve 50'nin üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranı %11,009 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 21) Düşük gelir grubunda, hanehalkı büyüklüğünün 3 olduğu hanelerde hanehalkı reisinin yaşı arttıkça zeytinyağı tüketen hanelerin oranı artmaktadır. Bu durum 19-21 arasındaki çıkarımlar ile Şekil 4.6 incelenerek görülebilir.
- 22) Düşük gelire sahip, hanehalkı büyüklüğünün 4 olduğu düğümde zeytinyağı tüketimi konusunda hanehalkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmüştür.
- 23) Düşük gelire sahip, hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 35'in altında olduğu düğümde zeytinyağı tüketen hanelerin oranı %3,176 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 24) Düşük gelire sahip, hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 35 ve 35'in üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranı %5,891 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 25) Düşük gelire sahip, hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 35 ve 35'in üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranı, hanehalkı reisinin yaşının 35'in altında olduğu düğüme daha fazladır ve yaklaşık 1,5 katıdır.
- 26) Düşük gelire sahip, hanehalkı büyüklüğünün 5 ve 5'in üzerinde olduğu düğümde zeytinyağı tüketimi konusunda hanede Sıfır-Beş yaş birey sahipliği karakteristiğinin ayırt edici bir özellik olduğu görülmüştür.
- 27) Düşük gelire sahip, hanehalkı büyüklüğünün 5 ve 5'in üzerinde olduğu, hanede Sıfır-Beş yaş bireyin olmadığı düğümde zeytinyağı tüketen hanelerin oranı %4,176 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.

- 28) Düşük gelire sahip, hanehalkı büyüklüğünün 5 ve 5'in üzerinde olduğu, hanede Sıfır-Beş yaş bireyin olduğu düğümde zeytinyağı tüketen hanelerin oranı %2,978 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 29) Düşük gelire sahip, hanehalkı büyüklüğünün 5 ve 5'in üzerinde olduğu, Sıfır-Beş yaş bireyin olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun okuma yazma biliyor (okula gitmedi) veya okuma yazma bilmiyor olduğu düğümde zeytinyağı tüketen hanelerin oranı %2,278 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır. Bu düğüm, ağaçtaki tüm düğümler arasında zeytinyağı tüketen hanelerin en az olduğu düğümdür. Genel tüketim oranının yaklaşık dörtte biri kadardır.

Zeytinyağı tüketiminin orta alt gelir grubu düğümü altındaki karar ağacı yapısı Şekil 4.7'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Orta Alt gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketimi konusunda hanehalkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmüştür. Orta alt gelire sahip hanelerin bulunduğu düğümün ve tüm alt düğümlerin detaylı özellikleri Şekil 4.7'de belirtilmiştir.
- 2) Orta alt gelire sahip, hanehalkı reisinin yaşı 35'in altında olduğu düğümde zeytinyağı tüketen hanelerin oranı %7,639 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Orta alt gelire sahip, hanehalkı reisinin yaşı 35 ile 49 arasında olduğu düğümde zeytinyağı tüketen hanelerin oranı %10,240 olarak görülmektedir, bu durum genel tüketim oranı ile yakınlık göstermektedir.
- 4) Orta alt gelire sahip, hanehalkı reisinin yaşı 50 ile 64 arasında olduğu düğümde zeytinyağı tüketen hanelerin oranı %14,927 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 5) Orta alt gelire sahip, hanehalkı reisinin yaşı 65 ve 65'in üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranı %18,494 olarak görülmektedir. Bu düğümde, zeytinyağı tüketen hanelerin oranı genel tüketim oranına göre artmıştır ve genel tüketim oranının yaklaşık 2 katıdır.
- 6) Orta alt gelir grubunda yer alan hanelerde, hanehalkı reisinin yaşı arttıkça zeytinyağı tüketen hanelerin oranı artmaktadır. Bu durum madde 2-5 ile Şekil 4.7 detaylı incelenerek görülebilir.

- 7) Orta Alt gelire sahip, hanehalkı reisinin yaşının 35'in altında olduğu düğümde hanehalkı reisinin eğitim durumu karakteristiğinin ayırt edici bir özellik olduğu görülmüştür.
- 8) Orta alt gelire sahip, hanehalkı reisinin yaşı 35'in altında olduğu, hanehalkı reisinin eğitim durumunun ilköğretim veya okuma yazma biliyor (okula gitmedi) veya okuma yazma bilmiyor olduğu düğümde zeytinyağı tüketen hanelerin oranı %3,671 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 9) Orta alt gelire sahip, hanehalkı reisinin yaşı 35'in altında olduğu, hanehalkı reisinin eğitim durumunun ilköğretim veya lise veya 2-3 yıllık yüksek okul olduğu düğümde zeytinyağı tüketen hanelerin oranı %6,885 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 10) Orta alt gelire sahip, hanehalkı reisinin yaşı 35'in altında olduğu, hanehalkı reisinin eğitim durumunun fakülte/yüksek okul veya lisansüstü (yüksek lisans, doktora) olduğu düğümde zeytinyağı tüketen hanelerin oranı %13,604 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Ayrıca bu düğümde zeytinyağı tüketen hanelerin oranı, hanehalkı reisinin eğitim durumunun ilköğretim veya lise veya 2-3 yıllık yüksek okul olduğu düğümdeki zeytinyağı tüketen hanelerin oranının yaklaşık 2 katı, eğitim durumunun ilköğretim veya okuma yazma biliyor (okula gitmedi) veya okuma yazma bilmiyor olduğu düğümdeki hanelerin oranının yaklaşık 4 katıdır.
- 11) Orta alt gelire sahip, hanehalkı reisinin yaşı 35'in altında olduğu hanelerde, hanehalkı reisinin eğitim seviyesi arttıkça zeytinyağı tüketen hanelerin oranı artmaktadır.
- 12) Orta Alt gelire sahip, hanehalkı reisinin yaşının 35 ile 49 arasında olduğu düğümde hanehalkı reisinin eğitim durumu karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 13) Orta alt gelire sahip, hanehalkı reisinin yaşının 35 ile 49 arasında olduğu, hanehalkı reisinin eğitim durumunun ilköğretim veya ilköğretim veya okuma yazma biliyor (okula gitmedi) veya okuma yazma bilmiyor olduğu düğümde zeytinyağı tüketen hanelerin oranı %7,924 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 14) Orta alt gelire sahip, hanehalkı reisinin yaşının 35 ile 49 arasında olduğu, hanehalkı reisinin eğitim durumunun lise veya üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranı %13,408 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

- 15) Orta alt gelire sahip, hanehalkı reisinin yaşının 35 ile 49 arasında olduğu, hanehalkı reisinin eğitim durumunun lise veya üstünde olduğu düğümde zeytinyağı tüketen hanelerin oranı, hanehalkı reisinin eğitim durumunun ilköğretim veya altında olduğu düğüme göre daha fazladır ve yaklaşık 2 katıdır.
- 16) Orta alt gelire sahip, hanehalkı reisinin yaşı 35'in altında olduğu, hanehalkı reisinin eğitim durumunun fakülte/yüksekokul veya üzeri olduğu düğümde zeytinyağı tüketen hanelerin oranı %13,604 olarak görülürken, orta alt gelire sahip, hanehalkı reisinin yaşının 35 ile 49 arasında olduğu, hanehalkı reisinin eğitim durumunun lise veya üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranı %13,408 olarak görülmektedir. Bu iki düğümde zeytinyağı tüketen hanelerin oranı neredeyse birine eşittir.
- 17) Orta alt gelire sahip, hanehalkı reisinin yaşının 50 ile 64 arasında olduğu düğümde hanehalkı reisinin eğitim durumu karakteristiğinin ayırt edici bir özellik olduğu görülmüştür. Bu düğümden meydana gelen tüm alt düğümlerde zeytinyağı tüketen hanelerin oranının genel tüketim oranının üzerindedir.
- 18) Orta alt gelire sahip, hanehalkı reisinin yaşının 50 ile 64 arasında olduğu, hanehalkı reisinin eğitim durumunun ilkokul veya okuma yazma bilmiyor (okula gitmedi) veya okuma yazma bilmiyor olduğu düğümde zeytinyağı tüketen hanelerin oranı %11,867 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 19) Orta alt gelire sahip, hanehalkı reisinin yaşının 50 ile 64 arasında olduğu, hanehalkı reisinin eğitim durumunun ilköğretim veya lise olduğu düğümde zeytinyağı tüketen hanelerin oranı %17,236 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 20) Orta alt gelire sahip, hanehalkı reisinin yaşının 50 ile 64 arasında olduğu, hanehalkı reisinin eğitim durumunun 2-3 yıllık yüksek okul veya fakülte/yüksekokul veya lisansüstü olduğu düğümde zeytinyağı tüketen hanelerin oranı %25,610 olarak görülmektedir. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 21) Orta alt gelire sahip, hanehalkı reisinin yaşının 50 ile 64 arasında olduğu hanehalkı reisinin eğitim durumunun 2-3 yıllık yüksek okul veya üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranı, hanehalkı reisinin eğitim durumunun ilkokul veya altında olduğu düğüme göre daha fazladır ve yaklaşık 2,5 katıdır. Ayrıca hanehalkı reisinin eğitim durumunun 2-3 yıllık yüksek okul ve üstü olduğu düğümde zeytinyağı tüketen hanelerin oranı genel tüketim oranının yaklaşık 2,5 katıdır.

- 22) Orta alt gelire sahip, hanehalkı reisinin yaşının 50 ile 64 arasında olduğu hanelerde, hanehalkı reisinin eğitim seviyesi arttıkça zeytinyağı tüketen hanelerin oranı artmaktadır.
- 23) Orta alt gelire sahip, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu düğümde hanehalkı reisinin eğitim durumu karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 24) Orta alt gelire sahip, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun okuma yazma biliyor (okula gitmedi) veya okuma yazma bilmiyor olduğu düğümde zeytinyağı tüketen hanelerin oranının %8,686 olarak görülmüştür. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 25) Orta alt gelire sahip, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun ilkokul olduğu düğümde zeytinyağı tüketen hanelerin oranının %16,623 olarak görülmüştür. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 26) Orta alt gelire sahip, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun ilköğretim veya üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranının %31,729 olarak görülmüştür. Bu düğümde zeytinyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Ayrıca genel tüketim oranının yaklaşık 3 katıdır.
- 27) Orta alt gelire sahip, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun ilköğretim veya üzerinde olduğu düğümde zeytinyağı tüketen hanelerin oranı, hanehalkı reisinin eğitim durumunun ilkokul olduğu düğüme göre daha fazladır ve yaklaşık 2 katı, hanehalkı reisinin okuryazar (okul bitirmedi) veya okuma yazma bilmiyor olduğu düğüme göre daha fazladır ve yaklaşık 4 katıdır. Bu düğüm, diğer tüm düğümler içinde zeytinyağı tüketen hanelerin oranının en yüksek olduğu düğümdür.
- 28) Orta alt gelire sahip, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu hanelerde hanehalkı reisinin eğitim seviyesi arttıkça zeytinyağı tüketen hanelerin oranı artmaktadır.

4.4.3 Yenilebilir Sıvı Yağ Tüketimini Etkileyen Faktörler

TÜİK tarafından zeytinyağı haricindeki sıvı yağlar, diğer yenilebilir sıvı yağlar olarak olarak isimlendirilmektedir. Kullanım kolaylığı olması bakımından “Diğer Yenilebilir Sıvı Yağlar” yerine “Yenilebilir Sıvı Yağ” olarak isimlendirme yapılmıştır. Kurulan bu modelde amaç, Hanehalkı Bütçe Anketi verileri kullanılarak, hanehalkının tüketmiş olduğu yenilebilir sıvı yağ tüketimini etkileyen faktörlerin ve hanehalklarının tüketim durumlarının Karar Ağacı

yöntemi ile belirlenmesidir. Hanehalkı yenilebilir sıvı yağ tüketim değişkeni bağımlı değişkendir yani hedef değişkendir. Yenilebilir sıvı yağ tüketimine etki eden karakteristikler Karar Ağacı yöntemlerinden CHAID yöntemi kullanılarak bulunmuştur. Kurulan model sonucunda CHAID yönteminin ürettiği sonuca göre tüketimi etkileyen karakteristikler Tablo 4.13’de gösterilmiştir.

Tablo 4.13 Yenilebilir Sıvı Yağ Tüketimini Etkileyen Karakteristikler

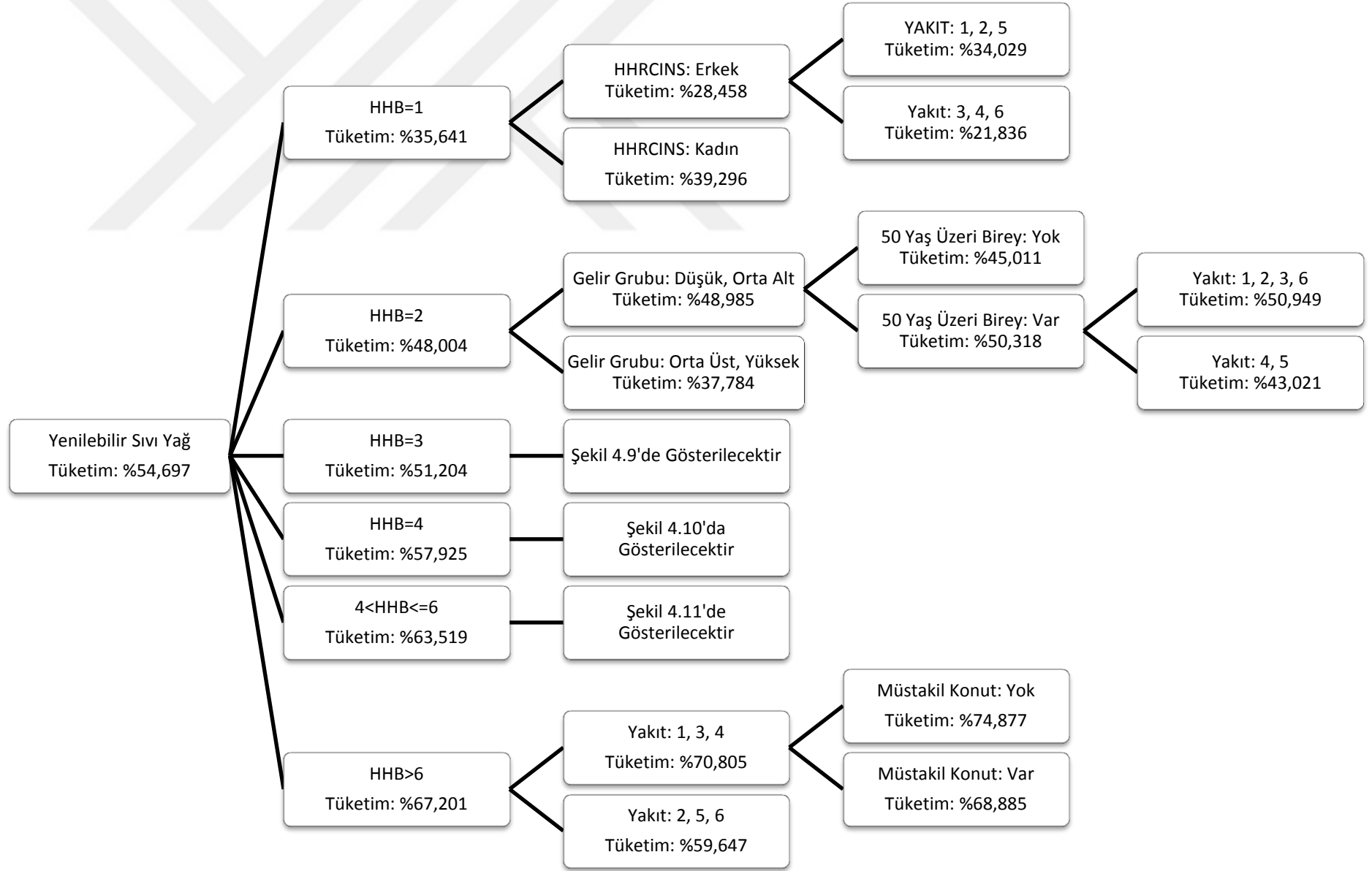
S. No	Belirleyicinin İsmi	Açıklama
1	HHB	Hanehalkı büyüklüğü
2	HHRyas	Hanehalkı reisinin yaşı
3	HHRCins	Hanehalkı reisinin cinsiyeti
4	Kırkntkd	Kır kent durumu
5	Elliyasuzerbirey	Hanede Elli yaş üzeri birey olup olmaması
6	Mulkiyet	Konutta Mülkiyet Durumu
7	Yakıt_1	Konutta en çok kullanılan birinci yakıt türü
8	SıfırBesyas	Hanede Sıfır Beş yaş arası birey olup olmadığı
9	Otomobil	Otomobil sahipliği
10	MusKonut	Müstakil konut sahipliği
11	GelirC	Gelir
12	ZorBanka	Oturduğu yer itibari ile bankaya ulaşma zorluğu

Tablo 4.14 Hanehalklarının yenilebilir sıvı yağ tüketim durumlarını göstermektedir.

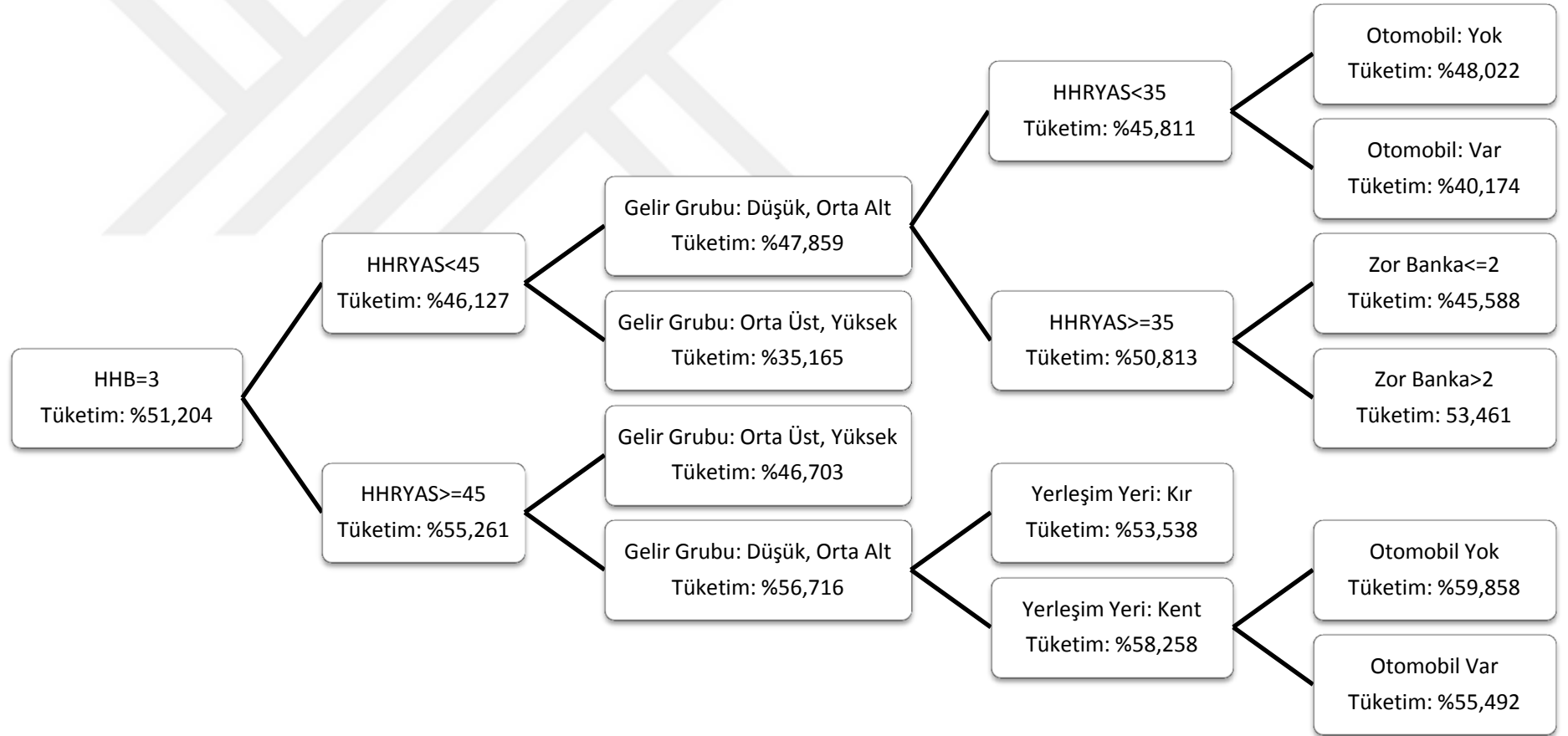
Tablo 4.14 Hanehalkı Yenilebilir Sıvı Yağ Tüketim Durumu

Yenilebilir Sıvı Yağ Tüketim Durumu	Yüzde	Sayı
Yenilebilir Sıvı Yağ Tüketen Hane	%54,697	21.897
Yenilebilir Sıvı Yağ Tüketmeyen Hane	%45,303	18.136
Toplam	100	40.033

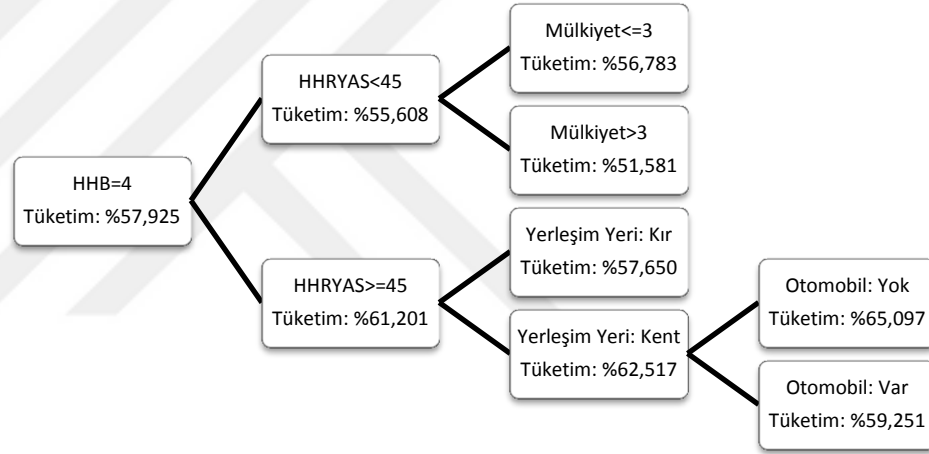
Tablo 4.14’e göre toplam kayıt sayısı 40.333’dür. Yenilebilir sıvı yağ tüketen hane sayısı 21.897 olup toplam hanelerin yaklaşık %54,7’lik kısmına gelmektedir. Yenilebilir sıvı yağ tüketmeyen hane sayısı 18.136 olup toplam hanelerin yaklaşık %45,3’üne karşılık gelmektedir.



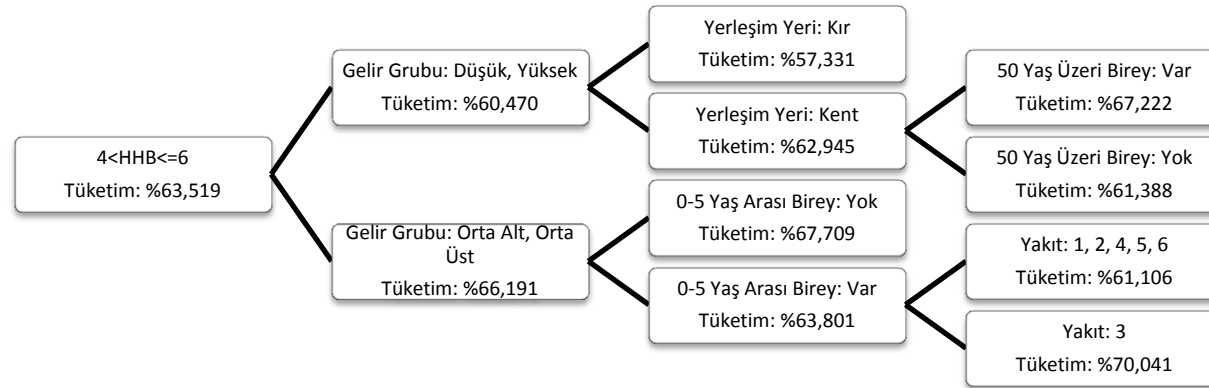
Şekil 4.8 Yenilebilir Sıvı Yağ Karar Ağacı-1



Şekil 4.9 Yenilebilir Sıvı Yağ Karar Ağacı-2



Şekil 4.10 Yenilebilir Sıvı Yağ Karar Ağacı-3



Şekil 4.11 Yenilebilir Sıvı Yağ Karar Ağacı-4

Modelleme sonucunda meydana gelen Karar Ağacını yorumladığımızda aşağıdaki çıkarımlara ulaşılmıştır.

- 1) Karar ağacının açık görünümüne baktığımızda ağacı ilk bölen değişkenin Hanehalkı Büyüklüğü olduğunu görüyoruz. Kök düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %54,697 olarak görülmüştür. Bu yorumla ilgili detaya Şekil 4.8'dan ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 1 olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %35,641 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Hanehalkı büyüklüğünün 2 olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %48,004 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Hanehalkı büyüklüğünün 3 olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %51,204 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 5) Hanehalkı büyüklüğünün 4 olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %57,925 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 6) Hanehalkı büyüklüğünün 5 veya 6 olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %63,519 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 7) Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %67,201 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 8) Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, hanehalkı büyüklüğünün 1 olduğu düğüme göre daha fazladır ve yaklaşık 2 katıdır.
- 9) Hanehalkı büyüklüğü arttıkça yenilebilir sıvı yağ tüketen hanelerin oranı artmaktadır. Bu yoruma, 2-7 arasındaki maddeleri inceleyerek ve Şekil 4.8'den ulaşılabilir.

Yenilebilir sıvı yağ tüketiminin, hanehalkı büyüklüğünün 1'e eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.8'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 1 olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda hanehalkı reisinin cinsiyeti karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.

Hanehalkı büyüklüğünün 1 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.8'den ulaşılabilir.

- 2) Hanehalkı büyüklüğünün 1 olduğu, hanehalkı reisinin cinsiyetinin erkek olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %28,458 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Hanehalkı büyüklüğünün 1 olduğu, hanehalkı reisinin cinsiyetinin kadın olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %39,296 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Hanede bir bireyin yaşadığı ve bireyin cinsiyetinin bayan olduğu düğümde yenilebilir sıvı yağ tüketim oranı, bireyin erkek olduğu düğüme göre daha fazladır.
- 5) Hanede yaşayan birey sayısının 1 kişi olduğu, bireyin cinsiyetinin erkek olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda yakıt karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 6) Hanede yaşayan birey sayısının 1 kişi olduğu, bireyin cinsiyetinin erkek olduğu, hanede en çok kullanılan yakıt türünün odun veya kömür veya tezek olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %34,029 olarak görülmüştür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 7) Hanede yaşayan birey sayısının 1 kişi olduğu, bireyin cinsiyetinin erkek olduğu, hanede en çok kullanılan yakıt türünün doğalgaz veya elektrik veya diğer (Fuel-oil, Motorin-mazot, Gazyağı, LPG (Tüpgaz) ve Diğer) olduğu düğümde tüketim oranı %21,836 olarak görülmüştür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır. Ayrıca bu düğüm, diğer tüm düğümler içerisinde yenilebilir sıvı yağ oranının en düşük olduğu düğümdür.

Yenilebilir sıvı yağ tüketiminin, hanehalkı büyüklüğünün 2'ye eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.8'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 2 olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda gelir karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 2 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.8'dan ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 2 olduğu, gelir grubunun düşük veya orta alt olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %48,985 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır..

- 3) Hanehalkı büyüklüğünün 2 olduğu, gelir grubunun orta üst veya yüksek olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %37,784 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Hanehalkı büyüklüğünün 2 olduğu hanelerde hanenin geliri arttıkça yenilebilir sıvı yağ tüketen hanelerin oranı azalmaktadır. Düşük veya orta alt gelire sahip hanelerin bulunduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı orta üst veya yüksek gelire sahip hanelerin bulunduğu düğüme göre nispeten daha ucuz olan bu yağ grubunu çokca tüketmektedir.

Yenilebilir sıvı yağ tüketiminin, hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu düğüm altındaki karar ağacı yapısı Şekil 4.8'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda yakıt karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.8'den ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu, hanede en çok kullanılan yakıt türünün odun veya doğalgaz veya elektrik olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %70,805 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 3) Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu, hanede en çok kullanılan yakıt türünün kömür veya tezek veya diğer (Fuel-oil, Motorin-mazot, Gazyağı, LPG (Tüpgaz) ve Diğer) olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %59,647 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 4) Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu, hanede en çok kullanılan yakıt türünün odun veya doğalgaz veya elektrik olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda müstakil konut sahipliği karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 5) Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu, hanede en çok kullanılan yakıt türünün odun veya doğalgaz veya elektrik olduğu, müstakil konuta sahip olmayan hanelerin bulunduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %74,877 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Bu düğüm, diğer tüm düğümler içerisinde yenilebilir sıvı yağ tüketen hanelerin oranının en yüksek olduğu düğümdür.

- 6) Hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu, hanede en çok kullanılan yakıt türünün odun veya doğalgaz veya elektrik olduğu, müstakil konuta sahip hanelerin bulunduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %68,885 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

Yenilebilir sıvı yağ tüketiminin, hanehalkı büyüklüğünün 3'e eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.9'da gösterilmiştir. Bu düğümde sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 3 olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda hanehalkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 3 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.9'den ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %46,127 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %55,261 olarak görülmektedir, bu durum genel tüketim oranı ile benzerlik göstermektedir.
- 4) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda gelir karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 5) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu, hanenin gelir durumunun düşük veya orta alt olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %47,859 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 6) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu, hanenin gelir durumunun orta üst veya yüksek olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %35,165 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 7) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda gelir karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.

- 8) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu, hanenin gelir durumunun düşük veya orta alt olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %56,716 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 9) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu, hanenin gelir durumunun orta üst veya yüksek olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranının %46,703 olduğu görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 10) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu hanelerin bulunduğu düğüm ile, Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu hanelerin bulunduğu düğümün her ikisinde de bölünmeyi gerçekleştiren karakteristik gelir'dir. Her iki düğümde de meydana gelen alt düğümler aynı şekilde bölünmüştür. Bununla ilgili detaylı bilgi Şekil 4.9'den görülebilir.
- 11) Hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu, hanenin gelir durumunun düşük veya orta alt olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, hanehalkı büyüklüğünün 3 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu, hanenin gelir durumunun düşük veya orta alt olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranından daha yüksektir. Benzer şekilde aynı durum gelir grubunun orta üst veya yüksek olduğu durumlar için de geçerlidir.

Yenilebilir sıvı yağ tüketiminin, hanehalkı büyüklüğünün 4'e eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.10'da gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 4 olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda hanehalkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 4 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.10'dan ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %55,608 olarak görülmektedir, bu durum genel tüketim oranı ile yakınlık göstermektedir.
- 3) Hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %61,201 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

- 4) Hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda mülkiyet karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 5) Hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu, konutta mülkiye durumunun ev sahibi veya kiracı veya lojman olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %56,783 olarak görülmüştür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 6) Hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 45'in altında olduğu, konutta mülkiye durumunun diğer olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %51,581 olarak görülmüştür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 7) Hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda kır-kent karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 8) Hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu, hanenin bulunduğu konum itibariyle kır olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %57,650 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 9) Hanehalkı büyüklüğünün 4 olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu, hanenin bulunduğu konum itibariyle kent olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %62,517 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

Yenilebilir sıvı yağ tüketiminin, hanehalkı büyüklüğünün 5 veya 6'ya eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.11'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 5 veya 6 olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda gelir karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 5 veya 6 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.11'dan ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 5 veya 6 olduğu, düşük veya yüksek gelire sahip hanelerin bulunduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %60,470 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

- 3) Hanehalkı büyüklüğünün 5 veya 6 olduğu, orta alt veya orta üst gelire sahip hanelerin bulunduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %66,191 olarak görülmektedir. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 4) Hanehalkı büyüklüğünün 5 veya 6 olduğu, hanenin gelir durumunun düşük veya yüksek olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda kır-kent karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 5) Hanehalkı büyüklüğünün 5 veya 6 olduğu, hanenin gelir durumunun düşük veya yüksek olduğu, hanenin bulunduğu konum itibariyle kır olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %57,331 olarak görülmüştür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 6) Hanehalkı büyüklüğünün 5 veya 6 olduğu, hanenin gelir durumunun düşük veya yüksek olduğu, hanenin bulunduğu konum itibariyle kent olduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %62,945 olarak görülmüştür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 7) Hanehalkı büyüklüğünün 5 veya 6 olduğu, hanenin gelir durumunun orta alt veya orta üst olduğu düğümde yenilebilir sıvı yağ tüketimi konusunda Sıfır-Beş yaş birey sahipliği karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 8) Hanehalkı büyüklüğünün 5 veya 6 olduğu, hanenin gelir durumunun orta alt veya orta üst olduğu, hanede Sıfır-Beş yaş bireyin bulunmadığı düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %67,709 olarak görülmüştür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 9) Hanehalkı büyüklüğünün 5 veya 6 olduğu, hanenin gelir durumunun orta alt veya orta üst olduğu, hanede Sıfır-Beş yaş bireyin bulunduğu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %63,801 olarak görülmüştür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

4.4.4 Margarin ve Diğer Bitkisel Yağ Tüketimini Etkileyen Faktörler

Kurulan bu modelde amaç, Hanehalkı Bütçe Anketi verileri kullanılarak, hanehalkının tüketmiş olduğu margarin ve diğer bitkisel yağ tüketimini etkileyen faktörlerin ve hanehalklarının tüketim durumlarının Karar Ağacı yöntemi ile belirlenmesidir. Hanehalkı margarin ve diğer bitkisel yağ tüketim değişkeni bağımlı değişkendir yani hedef değişkendir. Margarin ve Diğer Bitkisel Yağ tüketimine etki eden karakteristikler Karar Ağacı yöntemlerinden CHAID yöntemi kullanılarak bulunmuştur. Kurulan model sonucunda

CHAID yönteminin ürettiği sonuca göre tüketimi etkileyen karakteristikler Tablo 4.15’de gösterilmiştir.

Tablo 4.15 Margarin ve Diğer Bitkisel Yağ Tüketimini Etkileyen Karakteristikler

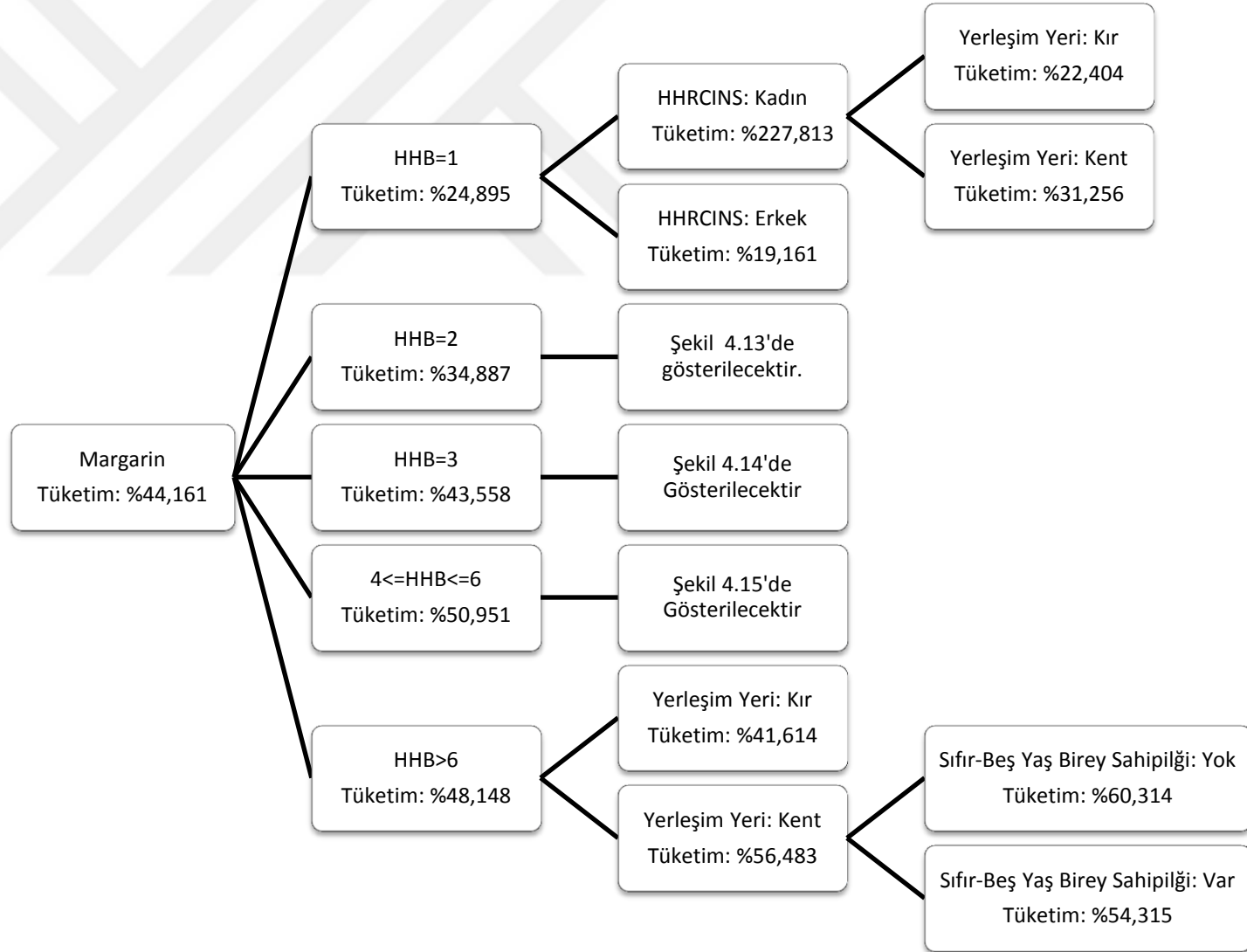
S. No	Belirleyicinin İsmi	Açıklama
1	HHB	Hanehalkı büyüklüğü
2	Kırkntkd	Kır kent durumu
3	MusKonut	Müstakil konut sahipliği
4	HHREgit	Hanehalkı reisinin eğitim durumu
5	HHRYas	Hanehalkı reisinin yaşı
6	HHRCins	Hanehalkı reisinin cinsiyeti
7	HHRSIG	Hanehalkı reisinin sigorta durumu
8	SifirBesyas	Hanede Sıfır Beş yaş arası birey olup olmadığı
9	ZorBanka	Oturduğu yer itibari ile bankaya ulaşma zorluğu
10	GelirC	Gelir

Tablo 4.16 Hanehalklarının Margarin ve Diğer Bitkisel Yağ tüketim durumlarını göstermektedir.

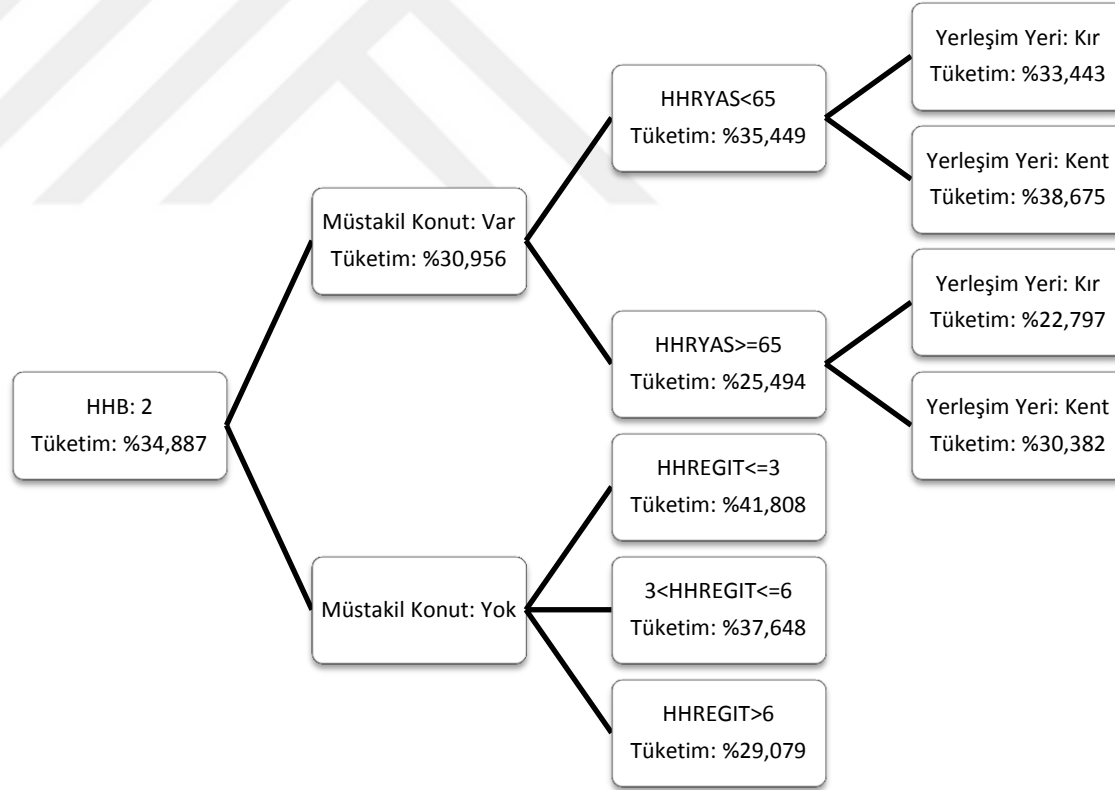
Tablo 4.16 Hanehalkı Margarin ve Diğer Bitkisel Yağ Tüketim Durumu

Margarin ve Diğer Bitkisel Yağ Tüketim Durumu	Yüzde	Sayı
Margarin ve Diğer Bitkisel Yağ Tüketmeyen Hane	44,161	17.679
Margarin ve Diğer Bitkisel Yağ Tüketen Hane	55,839	22.354
Toplam	100	40.033

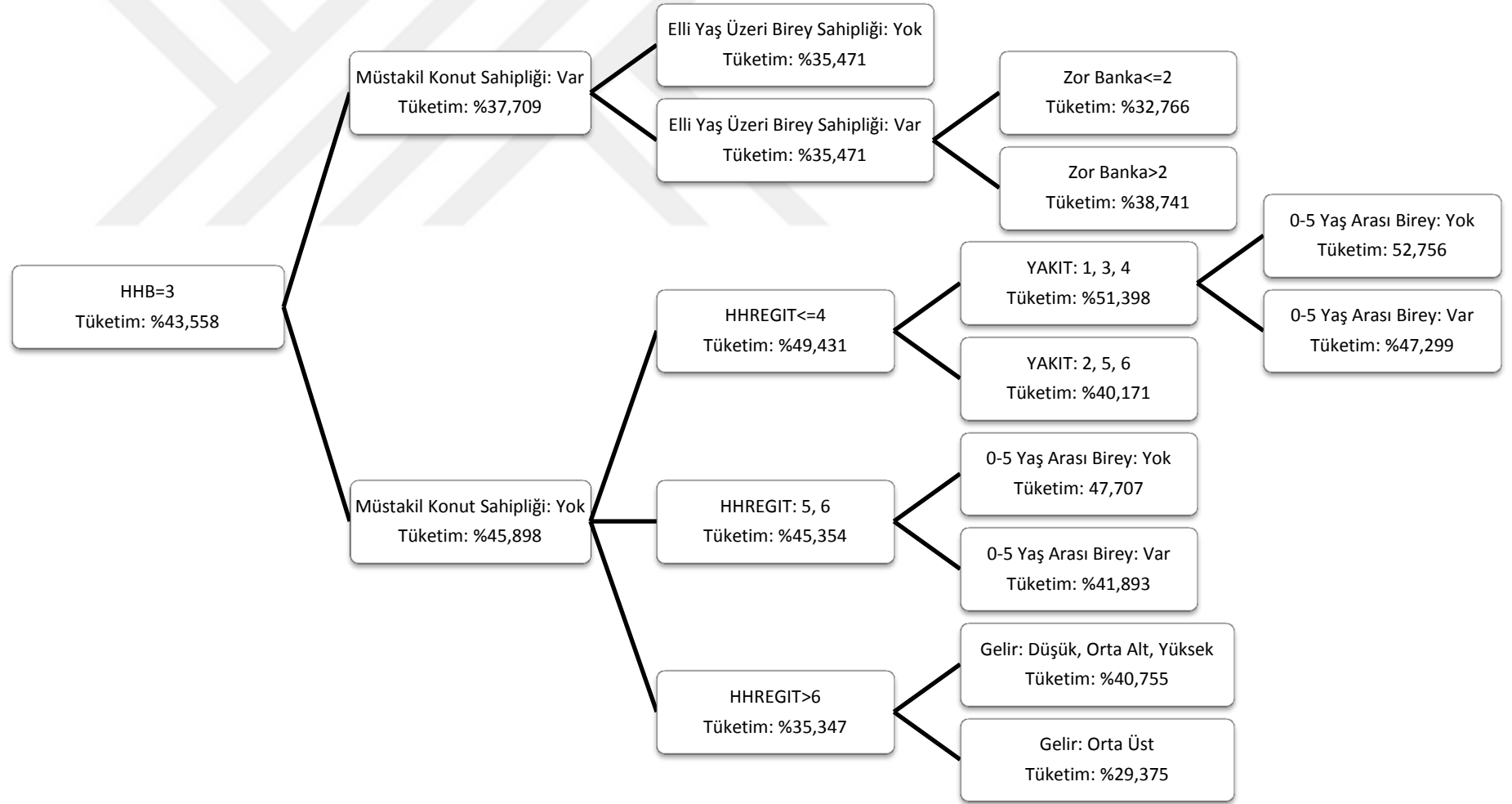
Bu bölümde, TÜİK tarafından “Margarin ve Diğer Bitkisel Yağlar” olarak isimlendirilen karakteristiği anlaşılma kolaylığı açısından “Margarin” olarak isimlendireceğiz. Tablo 4.16’ya göre toplam kayıt sayısı 40.333’dür. Margarin tüketen hane sayısı 17.679 olup toplam hanelerin yaklaşık %44,2’lik kısmına karşılık gelmektedir. Margarin tüketmeyen hane sayısı 22.354 olup toplam hanelerin yaklaşık %55,8’ine karşılık gelmektedir.



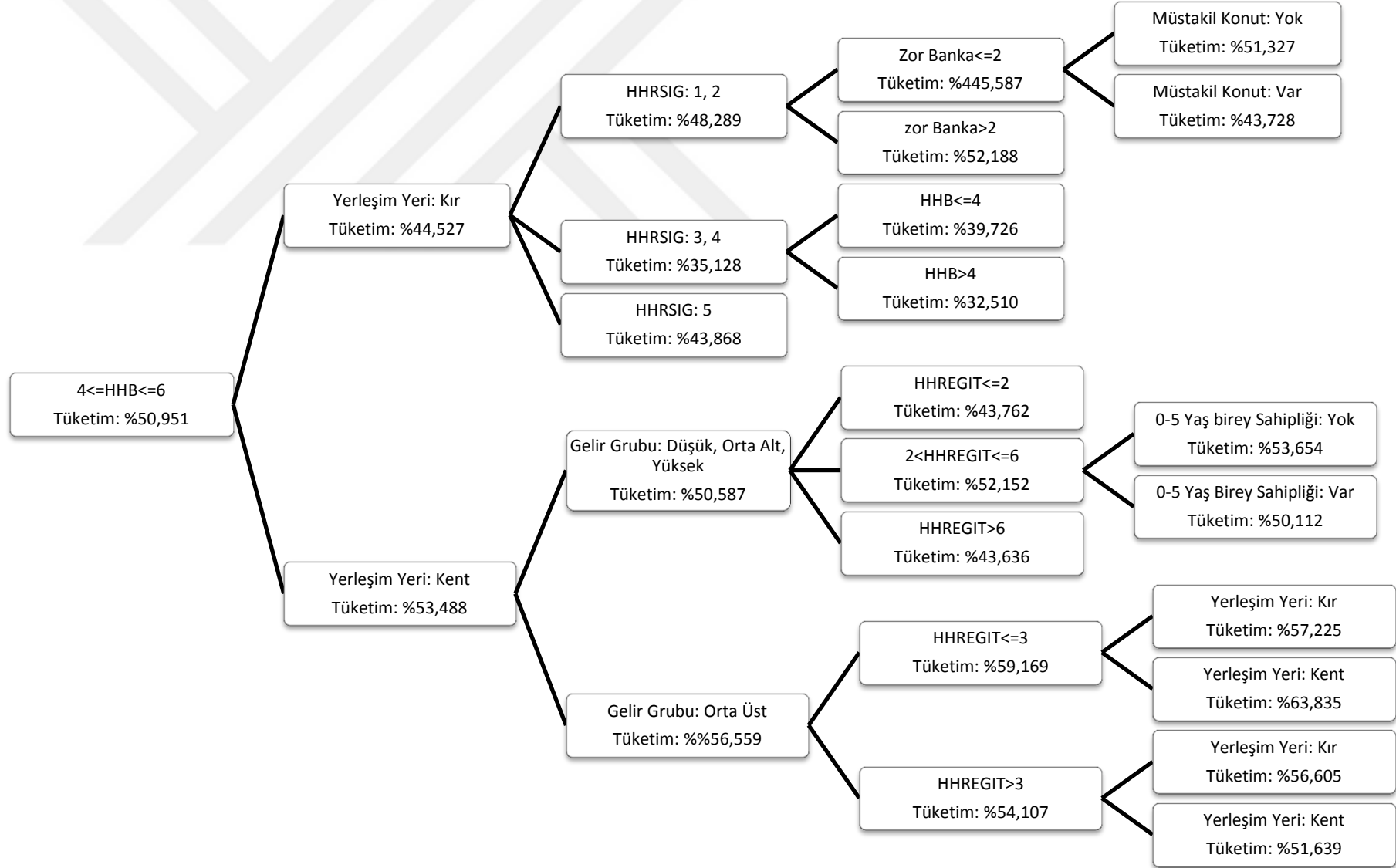
Şekil 4.12 Margarin ve Diğer Bitkisel Yağlar Karar Ağacı-1



Şekil 4.13 Margarin ve Diğer Bitkisel Yağlar Karar Ağacı-2



Şekil 4.14 Margarin ve Diğer Bitkisel Yağlar Karar Ağacı-3



Şekil 4.15 Margarin ve Diğer Bitkisel Yağlar Karar Ağacı-4

Modelleme sonucunda meydana gelen Karar Ağacı'nı yorumladığımızda aşağıdaki çıkarımlara ulaşılmıştır.

- 1) Karar ağacının açık görünümüne baktığımızda ağacı ilk bölen değişkenin Hanehalkı Büyüklüğü olduğunu görüyoruz. Kök düğümde Margarin tüketen hanelerin oranı yaklaşık %44,2 olarak görülmüştür. Bu yorumla ilgili detaya Şekil 4.12'den ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 1 olduğu düğümde Margarin tüketen hanelerin oranının %24,895 olduğu görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Hanehalkı büyüklüğünün 2 olduğu düğümde Margarin tüketen hanelerin oranının %34,887 olduğu görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Hanehalkı büyüklüğünün 3 olduğu düğümde Margarin tüketen hanelerin oranının %43,588 olduğu görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 5) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu düğümde Margarin tüketen hanelerin oranının %50,951 olduğu görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 6) Hanehalkı büyüklüğünün 6'nın üzerinde olduğu düğümde Margarin tüketen hanelerin oranının %48,148 olduğu görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 7) Hanehalkı büyüklüğünün 4-6 arasında olduğu veya 6'nın üzerinde olduğu düğümde Margarin tüketen hanelerin oranı, hanehalkı büyüklüğünün 1 olduğu düğüme göre daha fazladır ve yaklaşık 2 katıdır.
- 8) Hanehalkı büyüklüğü arttıkça Margarin tüketen hanelerin oranı artmaktadır. Bu yoruma, 2-6 arasındaki maddeleri inceleyerek ve Şekil 4.12'den ulaşılabilir.

Margarin tüketiminin, hanehalkı büyüklüğünün 1'e eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.12'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 1 olduğu düğümde Margarin tüketimi konusunda hanehalkı reisinin cinsiyeti karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 1 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.12'den ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 1 olduğu, hanehalkı reisinin cinsiyetinin erkek olduğu düğümde margarin tüketen hanelerin oranı %19,161 olarak görülmektedir. Bu düğümde margarin

tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır. Bu düğüm, diğer tüm düğümler içerisinde margarin tüketiminin en düşük olduğu düğümdür.

- 3) Hanehalkı büyüklüğünün 1 olduğu, hanehalkı reisinin cinsiyetinin kadın olduğu düğümde margarin tüketen hanelerin oranı %27,813 olarak görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Hanede bir bireyin yaşadığı ve bireyin cinsiyetinin bayan olduğu düğümde Margarin tüketen hanelerin oranı, bireyin erkek olduğu düğüme göre daha fazladır.
- 5) Hanede yaşayan birey sayısının 1 kişi olduğu, bireyin cinsiyetinin bayan olduğu düğümde margarin tüketimi konusunda kır-kent karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 6) Hanede yaşayan birey sayısının 1 kişi olduğu, bireyin cinsiyetinin bayan olduğu, hanenin bulunduğu yerin kır olduğu düğümde margarin tüketen hanelerin oranı %22,404 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 7) Hanede yaşayan birey sayısının 1 kişi olduğu, bireyin cinsiyetinin bayan olduğu, hanenin bulunduğu yerin kent olduğu düğümde margarin tüketen hanelerin oranı %31,256 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.

Margarin tüketiminin, hanehalkı büyüklüğünün 6'nın üzerinde olduğu düğüm altındaki karar ağacı yapısı Şekil 4.12'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 6'nın üzerinde olduğu düğümde margarin tüketimi konusunda kır-kent karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 6'nın üzerinde olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.12'den ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 6'nın üzerinde olduğu, hanenin bulunduğu konum itibariyle kır olduğu düğümde margarin tüketen hanelerin oranı %41,614 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Hanehalkı büyüklüğünün 6'nın üzerinde olduğu, hanenin bulunduğu konum itibariyle kent olduğu düğümde margarin tüketen hanelerin oranı %56,483 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 4) Hanehalkı büyüklüğünün 6'nın üzerinde olduğu, hanenin bulunduğu konum itibariyle kent olduğu düğümde margarin tüketimi konusunda 0-5 yaş arası bireye sahip olup olmama karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.

- 5) Hanehalkı büyüklüğünün 6'nın üzerinde olduğu, hanenin bulunduğu konum itibariyle kent olduğu, hanede 0-5 yaş arası bireyin bulunmadığı düğümde margarin tüketen hanelerin oranı %60,314 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 6) Hanehalkı büyüklüğünün 6'nın üzerinde olduğu, hanenin bulunduğu konum itibariyle kent olduğu, hanede 0-5 yaş arası bireyin bulunduğu düğümde margarin tüketen hanelerin oranı %54,315 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

Margarin tüketiminin, hanehalkı büyüklüğünün 2'ye eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.13'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 2 olduğu düğümde margarin tüketimi konusunda müstakil konut karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 2 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.13'den ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olmayan hanelerin olduğu düğümde margarin tüketen hanelerin oranı %38,056 olarak görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olan hanelerin olduğu düğümde Margarin tüketen hanelerin oranı %30,956 olarak görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olmayan hanelerin olduğu düğümde margarin tüketimi konusunda hanehalkı reisinin eğitim durumu karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 5) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olmayan, hanehalkı reisinin eğitim durumunun ilkokul veya altında olduğu düğümde margarin tüketen hanelerin oranı %41,808 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 6) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olmayan, hanehalkı reisinin eğitim durumunun ilköğretim veya lise veya 2-3 yıllık meslek yüksek okulu olduğu düğümde margarin tüketen hanelerin oranı %37,648 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 7) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olmayan, hanehalkı reisinin eğitim durumunun fakülte/yüksek okul veya lisansüstü olduğu düğümde margarin tüketen

hanelerin oranı %29,079 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.

- 8) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olmayan hanelerde hanehalkı reisinin eğitim düzeyi arttıkça margarin tüketen hanelerin oranı azalmaktadır.
- 9) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olan hanelerin olduğu düğümde margarin tüketimi konusunda hanehalkı reisinin yaş durumu karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 10) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olan, hanehalkı reisinin yaşının 65'in altında olduğu düğümde margarin tüketen hanelerin oranı %35,449 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 11) Hanehalkı büyüklüğünün 2 olduğu, müstakil konuta sahip olan, hanehalkı reisinin yaşının 65 ve üzerinde olduğu düğümde margarin tüketen hanelerin oranı %25,494 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.

Margarin tüketiminin, hanehalkı büyüklüğünün 3'e eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.14'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 3 olduğu düğümde margarin tüketimi konusunda müstakil konut karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 3 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.14'den ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olmayan hanelerin olduğu düğümde margarin tüketen hanelerin oranı %45,898 olarak görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 3) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olan hanelerin olduğu düğümde Margarin tüketen hanelerin oranı %37,709 olarak görülmektedir. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olmayan hanelerin olduğu düğümde margarin tüketimi konusunda hanehalkı reisinin eğitim durumu karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 5) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olmayan, hanehalkı reisinin eğitim durumunun ilköğretim veya altında olduğu düğümde margarin tüketen hanelerin oranı %49,431 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

- 6) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olmayan, hanehalkı reisinin eğitim durumunun lise veya 2-3 yıllık meslek yüksek okulu olduğu düğümde margarin tüketen hanelerin oranı %45,354 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 7) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olmayan, hanehalkı reisinin eğitim durumunun fakülte/yüksek okul veya lisansüstü olduğu düğümde margarin tüketen hanelerin oranı %35,347 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 8) Hanehalkı büyüklüğünün 3 olduğu hanelerde hanehalkı reisinin eğitim düzeyi arttıkça margarin tüketen hanelerin oranı azalmaktadır.
- 9) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olan hanelerin olduğu düğümde margarin tüketimi konusunda elli yaş ve üzeri bireye sahip olup olmama karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 10) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olan, hanede elli yaş ve üzeri bireyin olmadığı düğümde margarin tüketen hanelerin oranı %43,142 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 11) Hanehalkı büyüklüğünün 3 olduğu, müstakil konuta sahip olan, hanede elli yaş ve üzeri bireyin olduğu düğümde margarin tüketen hanelerin oranı %35,471 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.

Margarin tüketiminin, hanehalkı büyüklüğünün 4 veya 5 veya 6'ya eşit olduğu düğüm altındaki karar ağacı yapısı Şekil 4.15'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu düğümde margarin tüketimi konusunda kır-kent karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu düğüm ve ilgili olduğu diğer tüm alt düğümlere Şekil 4.15'den ulaşılabilir.
- 2) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kır olduğu düğümde margarin tüketen hanelerin oranı %44,527 olarak görülmüştür. Bu durum, genel tüketim oranına göre benzerlik göstermektedir.
- 3) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kır olduğu düğümde margarin tüketimi konusunda hanehalkı reisinin sigorta durumu karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 4) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kır olduğu, hanehalkı reisinin sigorta durumu emekli sandığı, SSK, SGK, Bağkur, özel sağlık

sigortası, banka, vakıf olduğu düğümde margarin tüketen hanelerin oranı %48,289 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

- 5) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kır olduğu, hanehalkı reisinin sigorta durumu “isteğe bağlı sigortalı veya yeşil kartlı” olduğu düğümde margarin tüketen hanelerin oranı %35,128 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 6) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kır olduğu, hanehalkı reisinin sigortasız olduğu düğümde margarin tüketen hanelerin oranı %43,868 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre çok küçük bir miktar azalmış olmasına karşın genel tüketim oranına çok yakındır.
- 7) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kent olduğu düğümde margarin tüketen hanelerin oranı %53,488 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 8) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kent olduğu düğümde margarin tüketimi konusunda gelir karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.
- 9) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kent olduğu, hanenin gelir durumunun düşük veya orta üst veya yüksek olduğu düğümde margarin tüketen hanelerin oranı %50,587 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 10) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kent olduğu, hanenin gelir durumunun orta alt olduğu düğümde margarin tüketen hanelerin oranı %56,559 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 11) Hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibariyle kent olduğu, hanenin gelir durumunun orta alt olduğu, hanehalkı reisinin eğitim durumunun ilkökul veya altında olduğu, hanehalkı tipinin ataerkil veya geniş aile olduğu düğümde margarin tüketen hanelerin oranı %63,835 olarak görülmüştür. Bu düğümde margarin tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Ayrıca bu düğüm, diğer tüm düğümler içerisinde margarin tüketen hanelerin oranının en yüksek olduğu düğümdür.

4.4.5 Tereyağı Tüketimini Etkileyen Faktörler

Kurulan bu modelde amaç, Hanehalkı Bütçe Anketi verileri kullanılarak, hanehalkının tüketmiş olduğu tereyağı tüketimini etkileyen faktörlerin ve hanehalklarının tüketim durumlarının Karar Ağacı yöntemi ile belirlenmesidir. Hanehalkı tereyağı tüketim değişkeni bağımlı değişkendir yani hedef değişkendir. Tereyağı tüketimine etki eden karakteristikler Karar Ağacı yöntemlerinden CHAID yöntemi kullanılarak bulunmuştur. Kurulan model sonucunda CHAID yönteminin ürettiği sonuca göre tüketimi etkileyen karakteristikler Tablo 4.17’de gösterilmiştir.

Tablo 4.17 Tereyağı Tüketimini Etkileyen Karakteristikler

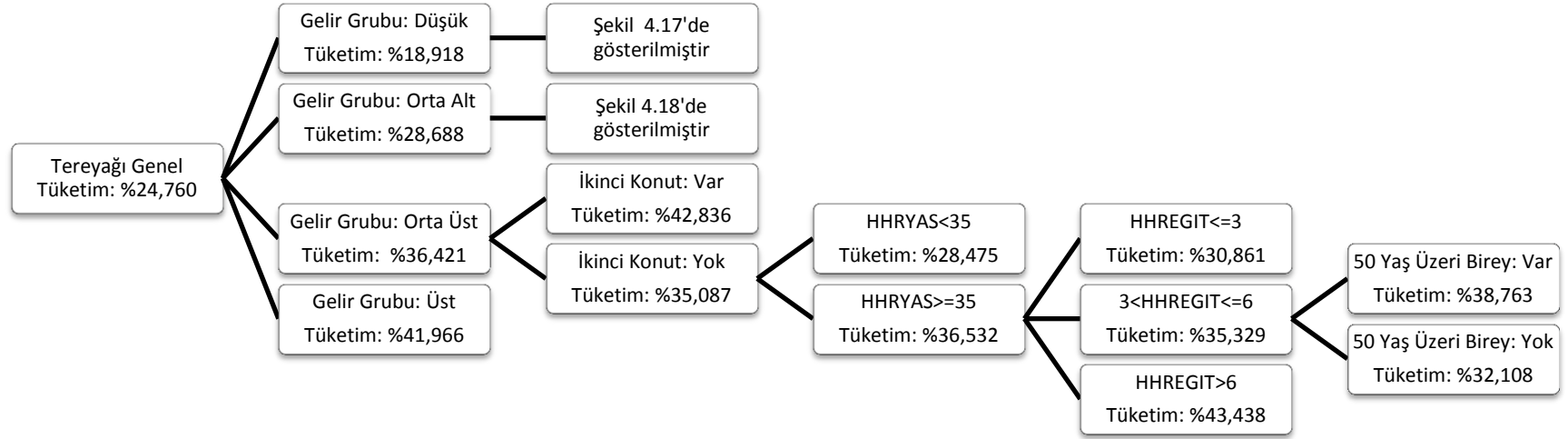
S. No	Belirleyicinin İsmi	Açıklama
1	GelirC	Gelir
2	HHRyas	Hanehalkı reisinin yaşı
3	Otomobil	Otomobil sahipliği
4	HHRSig	Hanehalkı reisinin sigorta durumu
5	HHREgit	Hanehalkı reisinin eğitim durumu
6	Konutiki	İkinci konut sahipliği
7	HHB	Hanehalkı büyüklüğü
8	Kırkntkd	Kır kent durumu
9	ZorBanka	Oturduğu yer itibari ile bankaya ulaşma zorluğu
10	Elliyasuzeribirey	Hanede Elli yaş üzeri birey olup olmaması

Tablo 4.18 Hanehalklarının tereyağı tüketim durumlarını göstermektedir.

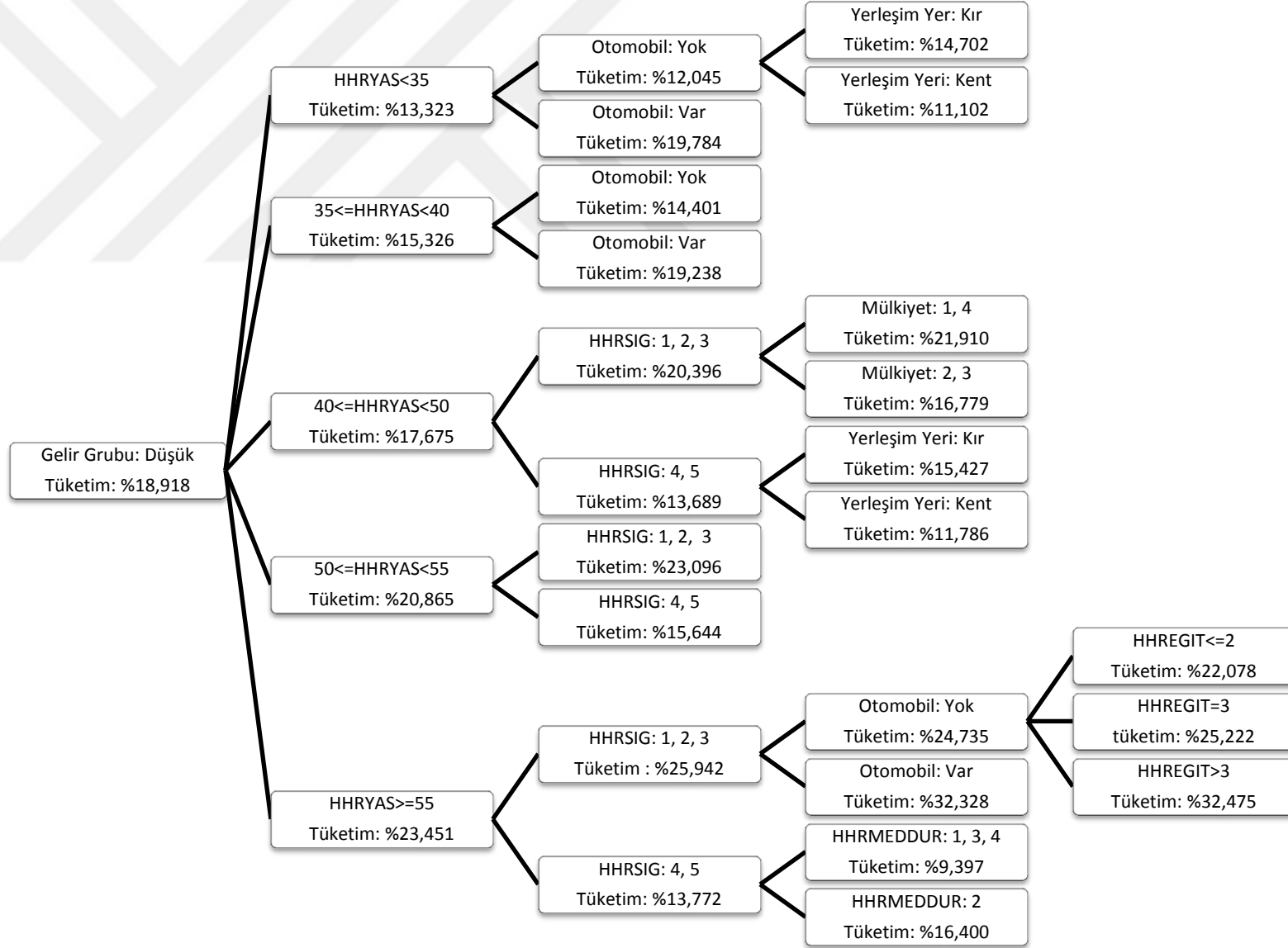
Tablo 4.18 Hanehalkı Tereyağı Tüketim Durumu

Tereyağı ve Diğer Bitkisel Yağ Tüketim Durumu	Yüzde	Sayı
Tereyağı Tüketen Hane	24,76	9.912
Tereyağı Tüketmeyen Hane	75,24	30.121
Toplam	100	40.033

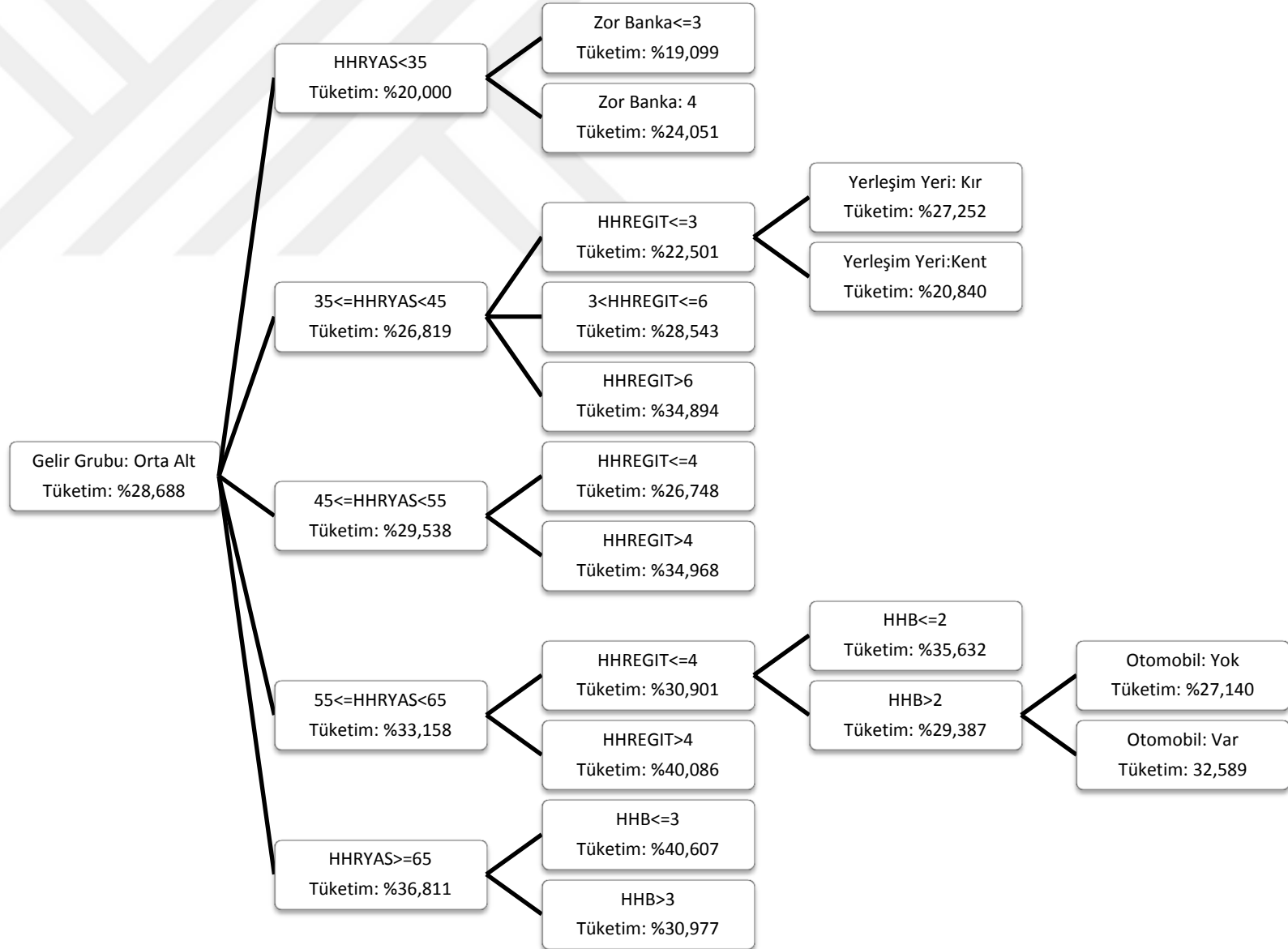
Tablo 4.18’e göre toplam kayıt sayısı 40.333’dür. Tereyağı tüketen hane sayısı 9.912 olup toplam hanelerin yaklaşık %24,76’lık kısmına karşılık gelmektedir. Tereyağı tüketmeyen hane sayısı 30.121 olup toplam hanelerin yaklaşık %75,24’ine karşılık gelmektedir.



Şekil 4.16 Tereyağı Karar Ağacı-1



Şekil 4.17 Tereyağı Karar Ağacı-2



Şekil 4.18 Tereyağı Karar Ağacı-3

Modelleme sonucunda meydana gelen Karar Ağacı'nı yorumladığımızda aşağıdaki çıkarımlara ulaşılmıştır.

- 1) Kök düğümde tereyağı tüketen hanelerin oranı yaklaşık %24,76 olarak görülmüştür. Karar ağacının açık görünümüne baktığımızda ağacı ilk bölen değişkenin gelir olduğunu görüyoruz. Bu yorumla ilgili detaya Şekil 4.16'den ulaşılabilir.
- 2) Düşük gelire sahip hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %18,918 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Orta alt gelire sahip hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %28,688 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 4) Orta üst gelire sahip hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %36,421 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 5) Yüksek gelire sahip hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %41,966 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 6) Şekil 4.16'e göre, hanelerin geliri arttıkça tereyağı tüketen hanelerin oranı artmaktadır.

Tereyağı tüketiminin, orta üst gelir grubu düğümü altındaki karar ağacı yapısı Şekil 4.16'da gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Orta üst gelir grubunda yer alan hanelerin bulunduğu düğümde tereyağı tüketimi konusunda ikinci konut sahipliği karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Orta üst gelire sahip hanelerin bulunduğu düğüm ve tüm alt düğümlere ait detaylı özellikler Şekil 4.16'de belirtilmiştir.
- 2) Orta üst gelir grubunda yer alan, ikinci konuta sahip olan hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %42,836 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 3) Orta üst gelir grubunda yer alan, ikinci konuta sahip olmayan hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %35,087 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 4) Orta üst gelir grubunda yer alan, ikinci konuta sahip olmayan hanelerin bulunduğu düğümde tereyağı tüketimi konusunda hane halkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmektedir.

- 5) Orta üst gelir grubundan yer alan, ikinci konuya sahip olmayan, hanehalkı reisinin yaşı 35'in altında olduğu düğümde tereyağı tüketen hanelerin oranı %28,475 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 6) Orta üst gelir grubundan yer alan, ikinci konuya sahip olmayan, hanehalkı reisinin yaşı 35 ve 35'in üzerinde olduğu düğümde tereyağı tüketen hanelerin oranı %36,532 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 7) Orta üst gelir grubundan yer alan, ikinci konuya sahip olmayan, hanehalkı reisinin yaşı 35 ve 35'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun fakülte/yüksek okul veya lisansüstü eğitim düzeyinde olduğu düğümde tereyağı tüketen hanelerin oranı %43,438 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır. Bu düğümde tereyağı tüketen hanelerin oranı, meydana gelen karar ağacındaki tüm düğümler içerisinde en yüksek olanıdır.

Tereyağı tüketiminin, düşük gelir grubu düğümü altındaki karar ağacı yapısı Şekil 4.17'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Düşük gelir grubunda yer alan hanelerin bulunduğu düğümde tereyağı tüketimi konusunda hanehalkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Düşük gelire sahip hanelerin bulunduğu düğüm ve tüm alt düğümlere ait detaylı özellikler Şekil 4.17'de belirtilmiştir.
- 2) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu düğümde tereyağı tüketen hanelerin oranı %13,323 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 39 arasında olduğu düğümde tereyağı tüketen hanelerin oranı %15,326 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 4) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 40 ile 49 arasında olduğu düğümde tereyağı tüketen hanelerin oranı %17,675 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 5) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 50 ile 54 arasında olduğu düğümde tereyağı tüketen hanelerin oranı %20,865 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.

- 6) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ve üzerinde olduğu düğümde tereyağı tüketen hanelerin oranı %23,451 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 7) Düşük gelire sahip hanelerin bulunduğu düğümde, hanehalkı reisinin yaşı arttıkça tereyağı tüketen hanelerin oranı artmaktadır.
- 8) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu düğümde tereyağı tüketimi konusunda otomobil sahipliği karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 9) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu, otomobile sahip olmayan hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %12,045 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 10) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu, otomobile sahip olan hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %19,784 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 11) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu, otomobile sahip olmayan hanelerin bulunduğu düğümde tereyağı tüketimi konusunda kır-kent karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 12) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu, otomobile sahip olmayan, hanenin bulunduğu yer itibariyle kır olduğu düğümde tereyağı tüketen hanelerin oranı %14,702 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 13) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu, otomobile sahip olmayan, hanenin bulunduğu yer itibariyle kent olduğu düğümde tereyağı tüketen hanelerin oranı %11,102 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 14) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 39 arasında olduğu düğümde tereyağı tüketimi konusunda otomobil sahipliği karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 15) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 39 arasında olduğu, otomobile sahip olmayan hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %14,401 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.

- 16) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 39 arasında olduğu, otomobile sahip olan hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı %19,238 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır. Ayrıca bu düğüm ile 10. çıkarımda yer alan düğümlerin oranları neredeyse birbirine eşittir.
- 17) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 40 ile 49 arasında olduğu düğümde tereyağı tüketimi konusunda hanehalkı reisinin sigorta durumu karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 18) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 40 ile 49 arasında olduğu, hanehalkı reisinin sigorta durumunun zorunlu sigortalı veya diğer (banka, vakıf vb) veya isteğe bağlı sigortalı olduğu düğümde tereyağı tüketen hanelerin oranı %20,396 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 19) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 40 ile 49 arasında olduğu, hanehalkı reisinin sigorta durumunun yeşil kart veya sigortasız olduğu düğümde tereyağı tüketen hanelerin oranı %13,689 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 20) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 50 ile 54 arasında olduğu düğümde tereyağı tüketimi konusunda hanehalkı reisinin sigorta durumu karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 21) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 50 ile 54 arasında olduğu, hanehalkı reisinin sigorta durumunun zorunlu sigortalı veya diğer (banka, vakıf vb) veya isteğe bağlı sigortalı olduğu düğümde tereyağı tüketen hanelerin oranı %23,096 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 22) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 50 ile 54 arasında olduğu, hanehalkı reisinin sigorta durumunun yeşil kart veya sigortasız olduğu düğümde tereyağı tüketen hanelerin oranı %15,644 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 23) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ve üzerinde olduğu düğümde tereyağı tüketimi konusunda hanehalkı reisinin sigorta durumu karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 24) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ve üzerinde olduğu, hanehalkı reisinin sigorta durumunun zorunlu sigortalı veya diğer (banka, vakıf vb) veya

isteğe bağlı sigortalı olduğu düğümde tereyağı tüketen hanelerin oranı %25,942 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

- 25) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ve üzerinde olduğu, hanehalkı reisinin sigorta durumunun yeşil kart veya sigortasız olduğu düğümde tereyağı tüketen hanelerin oranı %13,772 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 26) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ve üzerinde olduğu, hanehalkı reisinin sigorta durumunun yeşil kart veya sigortasız olduğu, hanehalkı reisinin medeni durumunun hiç evlenmedi veya eşi öldü veya boşanmış olduğu düğümde tereyağı tüketen hanelerin oranı %9,397 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır. Ayrıca bu düğümde tereyağı tüketen hanelerin oranı, meydana gelen karar ağacındaki tüm düğümler içerisinde en düşük olanıdır.
- 27) Düşük gelir grubunda yer alan, hanehalkı reisinin yaşı 35'in altında olduğu düğüm ile 35-39 arasında olduğu düğümde bölünmeyi gerçekleştiren karakteristik otomobil sahipliği iken, hanehalkı reisinin yaşının 40-49 arası, 50-54 arası ile 55 ve üzerinde olduğu üç düğümde bölünmeyi gerçekleştiren karakteristik hanehalkı reisinin sigorta durumudur.

Tereyağı tüketiminin, orta alt gelir grubu düğümü altındaki karar ağacı yapısı Şekil 4.18'de gösterilmiştir. Bu düğümden sonraki ağaç yapısı yorumlandığında aşağıdaki bulgular elde edilmiştir.

- 1) Orta alt gelir grubunda yer alan hanelerin bulunduğu düğümde tereyağı tüketimi konusunda hanehalkı reisinin yaşı karakteristiğinin ayırt edici bir özellik olduğu görülmektedir. Orta alt gelire sahip hanelerin bulunduğu düğüm ve tüm alt düğümlere ait detaylı özellikler Şekil 4.18'de belirtilmiştir.
- 2) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu düğümde tereyağı tüketen hanelerin oranı %20,000 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 3) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 44 arasında olduğu düğümde tereyağı tüketen hanelerin oranı %26,819 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 4) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 45 ile 54 arasında olduğu düğümde tereyağı tüketen hanelerin oranı %29,538 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

- 5) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ile 64 arasında olduğu düğümde tereyağı tüketen hanelerin oranı %33,158 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 6) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 65 ve üzerinde olduğu düğümde tereyağı tüketen hanelerin oranı %36,811 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 7) Orta alt gelire sahip hanelerin bulunduğu düğümde, hanehalkı reisinin yaşı arttıkça tereyağı tüketen hanelerin oranı artmaktadır.
- 8) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu düğümde tereyağı tüketimi konusunda zor banka karakteristiğinin ayırt edici bir özellik olarak görülmüştür.
- 9) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu, hanenin ikamet ettiği yer itibariyle bankaya ulaşma durumunun çok zor veya zor veya kolay olduğu düğümde tereyağı tüketen hanelerin oranı %19,099 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 10) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35'in altında olduğu, hanenin ikamet ettiği yer itibariyle bankaya ulaşma durumunun çok kolay olduğu düğümde tereyağı tüketen hanelerin oranı %24,051 olarak görülmüştür. Bu durum genel tüketim oranına göre azalmış olmasına karşın, kök düğüm ile neredeyse eşit orana sahiptir.
- 11) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 44 arasında olduğu düğümde tereyağı tüketimi konusunda hanehalkı reisinin eğitim durumu karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 12) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 44 arasında olduğu, hanehalkı reisinin eğitim düzeyinin ilkokul veya altında olduğu düğümde tereyağı tüketen hanelerin oranı %22,501 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre azalmıştır.
- 13) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 44 arasında olduğu, hanehalkı reisinin eğitim düzeyinin ilköğretim veya lise veya 2-3 yıllık yüksek okul olduğu düğümde tereyağı tüketen hanelerin oranı %28,543 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 14) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 35 ile 44 arasında olduğu, hanehalkı reisinin eğitim düzeyinin fakülte/yüksek okul veya lisansüstü olduğu düğümde tereyağı tüketen hanelerin oranı %34,894 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

- 15) Orta alt gelire sahip, hanehalkı reisinin yaşının 35 ile 44 arasında olduğu düğümde, hanehalkı reisinin eğitim düzeyi arttıkça tereyağı tüketen hanelerin oranı artmaktadır.
- 16) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 45 ile 54 arasında olduğu düğümde tereyağı tüketimi konusunda hanehalkı reisinin eğitim durumu karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 17) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 45 ile 54 arasında olduğu, hanehalkı reisinin eğitim düzeyinin ilköğretim veya altında olduğu düğümde tereyağı tüketen hanelerin oranı %26,748 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 18) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 45 ile 54 arasında olduğu, hanehalkı reisinin eğitim düzeyinin lise veya üzerinde olduğu düğümde tereyağı tüketen hanelerin oranı %34,968 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 19) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ile 64 arasında olduğu düğümde tereyağı tüketimi konusunda hanehalkı reisinin eğitim durumu karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 20) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ile 64 arasında olduğu, hanehalkı reisinin eğitim düzeyinin ilköğretim veya altında olduğu düğümde tereyağı tüketen hanelerin oranı %30,901 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 21) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 45 ile 54 arasında olduğu, hanehalkı reisinin eğitim düzeyinin lise veya üzerinde olduğu düğümde tereyağı tüketen hanelerin oranı %40,086 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 22) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu düğümde tereyağı tüketimi konusunda hanehalkı büyüklüğü karakteristiği ayırt edici bir özellik olarak görülmüştür.
- 23) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu, hanehalkı büyüklüğünün 3 veya altında olduğu düğümde tereyağı tüketen hanelerin oranı %40,607 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.
- 24) Orta alt gelir grubunda yer alan, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu, hanehalkı büyüklüğünün 3'ün üzerinde olduğu düğümde tereyağı tüketen hanelerin oranı %30,977 olarak görülmüştür. Bu düğümde tereyağı tüketen hanelerin oranı, genel tüketim oranına göre artmıştır.

SONUÇ

Bu tez çalışmasında Türkiye İstatistik Kurumu tarafından her yıl düzenli olarak derlenen “Hanehalkı Bütçe Anketi Araştırması” veri seti kullanılarak kurumsal olmayan hanehalkının tüketmiş olduğu zeytinyağı, tereyağı, margarin ve diğer bitkisel yağlar ile yenilebilir yağ tüketimlerinin hanehalkı karakteristiklerine göre nasıl farklılaştığının Karar Ağaçları ile belirlenmesi amaçlanmıştır.

Çalışmada 2009-2012 yılları arasındaki, 40.333 kayıta sahip dört yıllık veri seti kullanılmıştır. Veri seti içerisinde hanehalkına ait tüketim, fert ve hanehalkı karakteristikleri bulunmaktadır. Dört yıla ait veri setleri genel olarak birbiri ile aynı yapıda oluşturulmuş olmasına karşın bazı farklılıklar gözlenmiştir. Bu farklılıklar veri ön işleme aşamasında giderilerek nihai veri tabanı oluşturulmuştur.

Tez çalışmasının ilk bölümünde veri madenciliği kavramı ve gelişimi, bilgi keşfi sürecinde kullanılan metodolojiler ve aşamaları detaylı olarak anlatılmıştır. İkinci bölümünde ise veri madenciliği yöntemlerinden Karar Ağaçları, Birliktelik Kuralları, Kümeleme Analizi, Yapay Sinir Ağları ile ilgili bilgiler verilmiştir. Üçüncü bölümde, genel olarak yağlarla ilgili bilgilere ve istatistikî verilere, hanehalkı genel tüketimini, gıda ve yağ tüketimini etkileyen faktörlerle ilgili bilgilere yer verilmiştir. Üçüncü bölümde ayrıca hanehalkı genel tüketim, gıda tüketimi ve yağ tüketim harcamaları konusunda yapılmış çalışmalar özetlenmiştir. Dördüncü bölümde ise analizde kullanılan verinin ön işleme aşaması, kurulan modeller ve sonuçlarına ait bilgilere yer verilmiştir.

Tezin üçüncü bölümünde hanehalkı genel tüketim, gıda tüketim ve yağ tüketimi alanında daha önce yapılmış olan çalışmalar detaylı olarak incelendiğinde çalışmaların iktisadi ve ekonometrik modeller ile istatistikî yöntemler kullanılarak yapıldığı görülmüştür. Gerçekleştirilen bu tez çalışmasında veri madenciliği süreci izlenmiştir ve Türkiye hanehalkı yağ tüketimini etkileyen faktörlerin belirlenmesinde veri madenciliği ve yöntemleri kullanılmıştır.

Tezin dördüncü bölümünde veri ön işleme ve birleştirme aşamasından sonra veri madenciliği hedefleri doğrultusunda nihai veri setleri elde edilmiştir. Hanehalkı gıda ve yağ tüketimi alanında yapılmış olan çalışmalardan elde edilen bilgiler kullanılarak veri tabanı içerisinden 21 değişken belirlenmiş ve analizde kullanılmıştır. Bu değişkenler ve detayları tezin dördüncü bölümünde yer almaktadır. Dört farklı yağ türü için birçok karar ağacı algoritmaları kullanılarak model denemeleri gerçekleştirilmiştir. CHAID, CART, QUEST ve

C5.0 karar ağacı algoritmaları kullanılarak oluşturulan model denemelerinde en uygun sonucu CHAID Algoritması vermiştir.

Yağ türlerini tüketen hanelere ilişkin kurulan modellerin sonuçları, meydana gelen karar ağacı yapıları, istatistikî bilgiler ve görseller dördüncü bölümde detaylı olarak anlatılmıştır. Zeytinyağı tüketen hanelere ilişkin kurulan modelin sonuçlarına göre tüketimi etkileyen faktörlerin; gelir, hanehalkı büyüklüğü, hanehalkı reisinin yaşı, eğitim düzeyi ve sigorta durumu, hanede Elli yaş üzeri ve Sıfır-Beş yaş arası birey sahipliği, otomobil sahipliği, kır/kent durumu ve hanede en çok kullanılan birinci yakıt türü olduğu görülmüştür. Karar ağacının açık görünümüne baktığımız zaman, ağacı ilk bölen değişkenin gelir olduğu görülürken, alt düğümlerde hanehalkı reisinin yaşı ve eğitim düzeyi, hanehalkı büyüklüğü olduğu görülmüştür. Meydana gelen karar ağacı incelendiğinde yapılmış olan çalışmanın, önceki çalışmaları desteklerken bazı farklılıkları da beraberinde getirdiği görülmüştür. Kurulan modelin sonucuna göre hanehalkı zeytinyağı tüketimini etkileyen en önemli karakteristiğin gelir olduğu görülmüştür. Düşük gelire sahip hanelerin bulunduğu düğümde zeytinyağı tüketimini etkileyen karakteristik hanehalkı büyüklüğü olurken, orta alt ve orta üst gelire sahip hanelerin bulunduğu düğümde hanehalkı reisinin yaşı karakteristiğinin zeytinyağı tüketimini etkilediği görülmüştür. Düşük gelire sahip hanelerin bulunduğu düğümde, hanehalkı büyüklüğü arttıkça zeytinyağı tüketen hanelerin oranı azalırken, orta alt ve orta üst gelire sahip hanelerin bulunduğu düğümde hanehalkı reisinin yaşı arttıkça zeytinyağı tüketen hanelerin oranı artmaktadır.

Yenilebilir sıvı yağ tüketen hanelere ilişkin kurulan modelin sonuçlarına göre tüketimi etkileyen faktörlerin; gelir, hanehalkı büyüklüğü, hanehalkı reisinin yaşı, eğitim düzeyi ve sigorta durumu, hanede elli yaş üzeri ve sıfır-beş yaş arası birey sahipliği, otomobil sahipliği, kır/kent durumu, hanede en çok kullanılan birinci yakıt türü, mülkiyet durum, müstakil konut sahipliği ve hanehalkının bulunduğu yer itibarıyla bankaya ulaşım durumu olduğu görülmüştür. Karar ağacının açık görünümüne baktığımız zaman, ağacı ilk bölen değişkenin hanehalkı büyüklüğü olduğu görülürken, alt düğümlerde hanehalkı reisinin yaşı ve cinsiyeti, gelir ve hanede en çok kullanılan birinci yakıt türü olduğu görülmüştür. Meydana gelen karar ağacı incelendiğinde yapılmış olan çalışmanın, önceki çalışmaları desteklerken bazı farklılıkları da beraberinde getirdiği görülmüştür. Kurulan modelin sonucuna göre hanehalkı diğer yenilebilir sıvı yağ tüketimini etkileyen en önemli karakteristiğin hanehalkı büyüklüğü olduğu görülmüştür. Hanehalkı büyüklüğünün bir olduğu düğümde diğer yenilebilir sıvı yağ tüketimini etkileyen karakteristik hanehalkı reisinin cinsiyeti olurken, iki veya beş veya altı (HHB=2 veya 5 veya 6) olduğu düğümde gelir, üç ve dört (HHB=3 veya 4) olduğu düğümde

hanehalkı reisinin yaşı, altıdan büyük ($HHB > 6$) olduğu düğümde ise konutta en çok kullanılan birinci yakıt türü karakteristiğinin diğer yenilebilir sıvı yağ tüketimini etkilediği görülmüştür. Hanehalkı büyüklüğünün bir olduğu ve hanehalkı reisinin cinsiyetinin erkek olduğu düğümde diğer yenilebilir sıvı yağ tüketen hanelerin oranı, cinsiyetin kadın olduğu düğümdeki orana göre daha düşüktür. Hanehalkı büyüklüğünün iki olduğu hanelerin bulunduğu düğümde, hanenin geliri arttıkça diğer yenilebilir sıvı yağ tüketen hanelerin oranı azalmaktadır. Hanehalkı büyüklüğünün üç veya dört olduğu düğümde hanehalkı reisinin yaşı arttıkça diğer yenilebilir sıvı yağ tüketen hanelerin oranının artmaktadır.

Margarin ve diğer bitkisel yağları tüketen hanelere ilişkin kurulan modelin sonuçlarına göre tüketimi etkileyen faktörlerin; hanehalkı büyüklüğü, kır/kent durumu, müstakil konut sahipliği, hanehalkı reisinin eğitim düzeyi, yaşı, cinsiyeti ve sigorta durumu, hanede sıfır-beş yaş arası birey sahipliği, hanenin bulunduğu konum itibarıyla bankaya ulaşma durumu ve gelir olduğu görülmüştür. Karar ağacının açık görünümüne baktığımız zaman, ağacı ilk bölen değişkenin hanehalkı büyüklüğü olduğu görülürken, alt düğümlerde hanehalkı reisinin cinsiyeti, kır/kent durumu ve müstakil konut sahipliği olduğu görülmüştür. Meydana gelen karar ağacı incelendiğinde yapılmış olan çalışmanın, önceki çalışmaları desteklerken bazı farklılıkları da beraberinde getirdiği görülmüştür. Kurulan modelin sonucuna göre hanehalkı margarin ve diğer bitkisel yağ tüketimini etkileyen en önemli karakteristiğin hanehalkı büyüklüğü olduğu görülmüştür. Hanehalkı büyüklüğünün bir olduğu düğümde margarin ve diğer bitkisel yağ tüketimini etkileyen karakteristik hanehalkı reisinin cinsiyeti olurken, iki veya üç ($HHB = 2$ veya 3) olduğu düğümde müstakil konut sahipliği, üçten büyük ($HHB > 3$) olduğu düğümde ise kır-kent karakteristiğinin margarin ve diğer bitkisel yağ tüketimini etkilediği görülmüştür. Hanehalkı büyüklüğünün bir olduğu ve hanehalkı reisinin cinsiyetinin erkek olduğu düğümde margarin ve diğer bitkisel yağ tüketen hanelerin oranı, cinsiyetin kadın olduğu düğümdeki orana göre daha düşüktür. Hanede yaşayan birey sayısının iki veya üç olduğu ve müstakil konuta sahip olmayan hanelerin bulunduğu düğümde margarin ve diğer bitkisel yağ tüketen hanelerin oranı, müstakil konuta sahip olan hanelerin bulunduğu düğüme göre daha fazladır. Hanede yaşayan birey sayısının üçten ($HHB > 3$) fazla olduğu ve kırdaki yaşayan hanelerin bulunduğu düğümde margarin ve diğer bitkisel yağ tüketen hanelerin oranı, kentte yaşayan hanelerin bulunduğu düğüme göre daha düşüktür.

Tereyağı tüketen hanelere ilişkin kurulan modelin sonuçlarına göre tüketimi etkileyen faktörlerin; gelir, hanehalkı reisinin yaşı, otomobil sahipliği, hanehalkı reisinin sigorta durumu, hanehalkı reisinin eğitimi, ikinci konut sahipliği, hanehalkı büyüklüğü, kır/kent durumu, konutun bulunduğu yer itibarıyla bankaya ulaşma durumu ve hanede elli yaş üzeri

birey sahipliği olduğu görülmüştür. Karar ağacının açık görünümüne baktığımız zaman, ağacı ilk bölen değişkenin gelir olduğu görülürken, alt düğümlerde ikinci konut sahipliği, hanehalkı reisinin yaşı ve eğitim düzeyi, hanehalkı büyüklüğü olduğu görülmüştür. Meydana gelen karar ağacı incelendiğinde yapılmış olan çalışmanın, önceki çalışmaları desteklerken bazı farklılıkları da beraberinde getirdiği görülmüştür. Kurulan modelin sonucuna göre hanehalkı tereyağı tüketimini etkileyen en önemli karakteristiğin gelir olduğu görülmüştür. Düşük veya orta alt gelire sahip hanelerin bulunduğu düğümde tereyağı tüketimini etkileyen karakteristik hanehalkı reisinin yaşı olurken, orta üst gelire sahip hanelerin bulunduğu düğümde ikinci konut sahipliği karakteristiğinin tereyağı tüketimini etkilediği görülmüştür. Düşük veya orta alt gelire sahip hanelerin bulunduğu düğümde, hanehalkı reisinin yaşı arttıkça tereyağı tüketen hanelerin oranının arttığı görülmüştür. Orta üst gelire sahip ve ikinci konutu bulanan hanelerin bulunduğu düğümde tereyağı tüketen hanelerin oranı, ikinci konuta sahip olmayan hanelerin bulunduğu düğüme göre daha yüksektir.

Gelir, hanehalkı büyüklüğü, hanehalkı reisinin yaşı, kır/kent durumu karakteristikleri dört farklı yağ türünde de tüketimi etkileyen ortak faktörlerdir. İkinci konut sahipliği sadece tereyağı tüketimini etkileyen bir faktörken, müstakil konut sahipliği margarin ve diğer bitkisel yağlar ile yenilebilir sıvı yağ tüketimini etkileyen, konutta en çok kullanılan birinci yakıt türü ise sadece zeytinyağı ve yenilebilir sıvı yağ tüketimini etkileyen bir faktör olarak görülmüştür.

Farklı yağ türleri için kurulmuş olan dört farklı modelde, analiz sonucu meydana gelen Karar Ağacının her bir düğümü bir çıkarım (kural) olarak yazılmış, her bir düğümde ilgili yağ türlerini tüketen hanelerin oranı belirtilmiştir.

Oluşturulan birinci modelde, zeytinyağı tüketen hanelerin oranı %10,521 olarak görülmüştür. Zeytinyağı tüketen haneler için oluşturulan karar ağacında en çok tüketim yapılan hanelerin olduğu düğüm; orta alt gelire sahip, hanehalkı reisinin yaşının 65 ve 65'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun ilköğretim veya üzerinde olduğu düğümdür. Bu düğümde zeytinyağı tüketen hanelerin oranı %31,729 olarak görülürken, genel tüketim oranının yaklaşık 3 katıdır. En az tüketimin yapıldığı düğüm; düşük gelire sahip, hanehalkı büyüklüğünün 5 ve 5'in üzerinde olduğu, hanede sıfır-beş yaş bireyin olduğu, hanehalkı reisinin yaşının 45 ve 45'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun okuma yazma biliyor (okula gitmedi) veya okuma yazma bilmiyor olduğu düğümdür. Bu düğümde zeytinyağı tüketen hanelerin oranı %2,278 olarak görülmektedir ve genel tüketim oranının yaklaşık dörtte biri kadardır.

Oluşturulan ikinci modelde, yenilebilir sıvı yağ tüketen hanelerin oranı %54,697 olarak görülmüştür. Yenilebilir sıvı yağ tüketen haneler için oluşturulan karar ağacında en çok

tüketim yapılan hanelerin olduğu düğüm; hanehalkı büyüklüğünün 7 ve 7'nin üzerinde olduğu, hanede en çok kullanılan yakıt türünün odun veya doğalgaz veya elektrik olduğu, müstakil konuta sahip olmayan hanelerin bulunduğu düğümdür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %74,877 olarak görülmektedir. En az tüketimin yapıldığı düğüm; hanede yaşayan birey sayısının bir kişi olduğu, bireyin cinsiyetinin erkek olduğu, hanede en çok kullanılan yakıt türünün doğalgaz veya elektrik veya diğer (Fuel-oil, Motorin-mazot, Gazyağı, LPG (Tüpgaz) ve Diğer) olduğu düğümdür. Bu düğümde yenilebilir sıvı yağ tüketen hanelerin oranı %21,836 olarak görülmüştür ve genel tüketim oranının yaklaşık beşte ikisi kadardır.

Oluşturulan üçüncü modelde, margarin ve diğer bitkisel yağları tüketen hanelerin oranı %44,161 olarak görülmüştür. Margarin ve diğer bitkisel yağları tüketen haneler için oluşturulan karar ağacında en çok tüketim yapılan hanelerin olduğu düğüm; hanehalkı büyüklüğünün 4 veya 5 veya 6 olduğu, hanenin bulunduğu konum itibarıyla kent olduğu, hanenin gelir durumunun orta alt olduğu, hanehalkı reisinin eğitim durumunun ilkökul veya altında olduğu, hanehalkı tipinin ataerkil veya geniş aile olduğu düğümdür. Bu düğümde margarin tüketen hanelerin oranı %63,835 olarak görülmüştür. En az tüketimin yapıldığı düğüm; hanehalkı büyüklüğünün 1 olduğu, hanehalkı reisinin cinsiyetinin erkek olduğu düğümdür. Bu düğümde margarin tüketen hanelerin oranı %19,161 olarak görülmektedir ve genel tüketim oranının yaklaşık yarısına eşittir.

Oluşturulan dördüncü modelde, tereyağı tüketen hanelerin oranı %24,760 olarak görülmüştür. Tereyağını tüketen haneler için oluşturulan karar ağacında en çok tüketim yapılan hanelerin olduğu düğüm; orta üst gelir grubundan yer alan, ikinci konuya sahip olmayan, hanehalkı reisinin yaşı 35 ve 35'in üzerinde olduğu, hanehalkı reisinin eğitim durumunun fakülte/yüksek okul veya lisansüstü eğitim düzeyinde olduğu düğümdür. Bu düğümde tereyağı tüketen hanelerin oranı %43,438 olarak görülmüştür. En az tüketimin yapıldığı düğüm; düşük gelir grubunda yer alan, hanehalkı reisinin yaşının 55 ve üzerinde olduğu, hanehalkı reisinin sigorta durumunun yeşil kart veya sigortasız olduğu, hanehalkı reisinin medeni durumunun hiç evlenmedi veya eşi öldü veya boşanmış olduğu düğümdür. Bu düğümde tereyağı tüketen hanelerin oranı %9,397 olarak görülmüştür ve genel tüketim oranının yaklaşık beşte ikisine eşittir.

Sonuç olarak zeytinyağı ve tereyağı gibi zengin besin öğelerine sahip, sağlıklı ve pahalı yağ türlerinin tüketimini etkileyen faktörler genel olarak birbirine benzerdir. İki yağ türünde de en önemli faktörler gelir, hanehalkı reisinin yaşı ve eğitim düzeyi, hanehalkı büyüklüğü ve otomobil sahipliği olduğu görülmüştür. Margarin ve diğer bitkisel yağlar ile

yenilebilir yağ (mısır, ayçiçeği, kolza, vb) gibi yağ türlerinin tüketimini etkileyen faktörlerin de birbirine benzer olduğu görülmüştür. İki yağ türünde de en önemli faktörlerin hanehalkı büyüklüğü, kır/kent durumu, hanehalkı reisinin yaşı ve cinsiyeti, gelir, sıfır-beş yaş birey sahipliği ve müstakil konut sahipliği olduğu görülmüştür.

Hanehalkı Bütçe Anketi, tezde anlatıldığı üzere birçok önemli bilginin elde edilmesinde ve planlamasında kullanılmaktadır. Her bir yıla ait ortalama 10.000 kayıt içeren bir veri seti, veri madenciliği yöntemlerinin uygulanması için yeterli sayıdadır. Anket verilerinin tüketim kısmını içeren yaklaşık 200 farklı tüketim kalemi mevcuttur. Bu kalemlerden her biri ayrı ayrı ya da toplulaştırılarak veri madenciliği ve yöntemleri uygulanabilir.



KAYNAKÇA

- Abdulai, A. (2002). "Household demand for food in Switzerland. A quadratic almost ideal demand system". Swiss Journal of Economics and Statistics (SJES), 138(I): 1-18.
- Aggarwal, C. C. (ed.). (2015). Data classification: algorithms and applications, CRC Press, Boca Raton
- Aggarwal, C. C. & Reddy, C. K. (ed.). (2014). Data clustering: algorithms and applications, CRC Press, Boca Raton
- Akaichi, F. & Revoredo-Giha, C. (2014). "The demand for dairy products in Malawi". African Journal of Agricultural and Resource Economics Volume, 9(3): 214-225.
- Akbay, C., (2007). "Urban Households' Cooking Oil And Fat Consumption Patterns In Turkey: Quality vs. Quantity". Quality & Quantity, 41(6): 851-867.
- Akın, Y. K. (2008). Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi. Yayınlanmamış Doktora Tezi. Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul
- Akpınar, H. (2000). "Veri tabanlarında bilgi keşfi ve veri madenciliği". İÜ İşletme Fakültesi Dergisi, 29(1): 1-22.
- Aktaş, A. R. (2008) Kentsel Alanda Et Talep Analizi: Batı Akdeniz Bölgesi Örneği. Yayınlanmamış Doktora Tezi. Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü, Isparta
- Ali, Z., Aslam, M. & Rasool, S. (2013). "Factors Affecting Consumption of Edible Oil in Pakistan". International Journal of Business and Management, 15(1): 87-92.
- Almuallim, H., (1996). "An efficient algorithm for optimal pruning of decision trees". Artificial Intelligence 83(2): 347-362.
- Altunışık, R., Özdemir S. ve Torlak Ö. (2001). Modern Pazarlama. Değişim Yayınları, Adapazarı
- Ambalajlı Süt ve Süt Ürünleri Sanayicileri Derneği. (2010). Dünya ve Türkiye Süt Endüstrisi Raporu, Elma Matbaa, Ankara
- Anderson, D. & Mcneill, G. (1992). Artificial Neural Networks Technology, Kaman Sciences Corporation, New York
- Bamiro, O. M. (2012). "Determinants of consumption of food away from home in Lagos Metropolis, Nigeria". Research Journal of Social Sciences and Management, 2(5): 15-23.

- Barros, R. C., de Carvalho, A. C. & Freitas, A. A. (2015). Automatic design of decision-tree induction algorithms, Springer, Springer Cham Heidelberg New York Dordrecht London
- Baş, N. (2006). Yapay Sinir Ağları Yaklaşımı ve Bir Uygulama. Yayımlanmamış Yüksek Lisans Tezi. Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstanbul
- Başaran, İ., “Margarin”. <http://w3.balikesir.edu.tr/~ismet/zeytin/Margarin.pdf>. (erişim tarihi: 03.10.2015)
- Begum, S., Farooq, M., Khan, M., Begum, N. & Khan, A. (2010). “Influence of socio-economic factors on food consumption pattern in district Nowshera”. *Sarhad Journal of Agriculture*, 26(3): 405-408.
- Berg, L. (1996). “Age distribution, saving and consumption in Sweden”. 10th Anniversary Meeting of the ESPE Uppsala University, 13-15 June 1996, Sweden, s. 1-19.
- Berry, M. J. A. & Linoff, G. S. (2004). Data Mining Techniques for Marketing, Sales, and Customer Relationship Management (Second Edition), Wiley Publishing Inc., Indianapolis, Indiana
- Beyaz, F. B. (2007). Türkiye’de Hanehalkı Tüketim Harcamaları ve Talep Tahmini. Yayımlanmamış Yüksek Lisans Tezi, Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Antalya
- Birant, D., Kut, A., Ventura, M., Altınok, H., Altınok, B., Altınok, E. ve Ihlamur, M. (2010). “İş Zekası Çözümleri için Çok Boyutlu Birliktelik Kuralları Analizi”. *Akademik Bilişim*, 10, 256.
- Bittencourt, M. V., Teratanavat, R. P. & Chern, W. S. (2007). “Food consumption and demographics in Japan: Implications for an aging population”. *Agribusiness*, 23(4): 529-551.
- Biyodoc, “Sinir Hücrelerinin Yapısı”. <http://www.biyodoc.com/Sinir-sistemi-sinir-hucrelerinin-yapisi-ve-noron-cesitleri.html>. (erişim tarihi: 13.05.2014)
- Bopape, L. & Myers, R. (2007). “Analysis of household demand for food in South Africa: Model selection, expenditure endogeneity, and the influence of socio-demographic effects”. In African econometrics society annual conference, 4-6 July 2007 Cape Town South Africa
- Bramer, M. (2007). Principles of Data Mining, Springer, London
- Brusilovsky, P., “Data Mining vs. Statistics“. <http://www.bisolutions.us/Data-Mining-vs-Statistics.php>. (erişim tarihi: 16.07.2015)

- BYSD Bitkisel Yağ Sanayicileri Derneği, “İstatistikler”. <http://www.bysd.org.tr/Istatistikler.aspx>, (erişim tarihi: 13.05.2016a)
- BYSD Bitkisel Yağ Sanayicileri Derneği, “Margarin”. <http://www.bysd.org.tr/margarin.pdf>, (erişim tarihi: 25.06.2016b)
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. & Zanasi, A. (1998). *Discovering data mining: from concept to implementation*. Prentice-Hall, Inc., Upper Saddle River, NJ
- Cai, J. (2006). *Decision Tree Pruning Using Expert Knowledge*. Doctor of Philosophy, The Graduate Faculty Of The University Of Akron, ABD
- Case, A. & Deaton, A. (2003). *Consumption, health, gender, and poverty* (Vol. 3020). World Bank Publications. Paper No: 3020
- Cebesoy, S., “OLAP”. <http://www.suleymancebesoy.com/Posts.aspx?PostID=145>. (erişim tarihi: 15.02.2014)
- Cieslak, D. A. & Chawla, N. V., (2008). "Learning decision trees for unbalanced data". *Machine learning and knowledge discovery in databases*. Springer Berlin Heidelberg, Proceedings, Part I: 241-256.
- Cios, K. J., Swiniarski, R. W., Pedrycz, W. & Kurgan, L. A. (2007). *Data Mining A Knowledge Approach*. Springer US, New York
- Colón-Ramos, U., Kabagambe, E. K., Baylin, A., Ascherio, A., Campos, H. & Peterson, K. E. (2007). “Socio-economic status and health awareness are associated with choice of cooking oil in Costa Rica”. *Public Health Nutrition*, 10(11): 1214-1222.
- Çağlayan, E. & Astar, M. (2012). “A Microeconomic Analysis of Household Consumption Expenditure Determinants for Both Rural and Urban Areas in Turkey”. *American International Journal of Contemporary Research*, 2(2): 27-34
- Çalışkan Ş. (2003). *Kocaeli İlinde (Kır-Kent Ayrımında) Hanehalkı Tüketim Harcamalarının Analizi*. Yayınlanmamış Doktora Tezi. Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü, Kocaeli
- Çayıroğlu, İ., “İleri Algoritma Analizi-5 Yapay Sinir Ağları”. <http://www.ibrahimcayiroglu.com/Dokumanlar/IleriAlgoritmaAnalizi/IleriAlgoritmaAnalizi-5.Hafta-YapaySinirAglari.pdf>. (erişim tarihi, 17.08.2015)
- Çelik, Ş. (2013). “Kümeleme Analizi ile Sağlık Göstergelerine Göre Türkiye’deki İllerin Sınıflandırılması”. *Doğuş Üniversitesi Dergisi*, 14(2): 175-194.
- Data Mining Articles., “Association Rules Algorithms”. <http://www.dataminingarticles.com/association-analysis/association-rules-algorithms/>. (erişim tarihi: 02.03.2014)

- Davis C. G., Dong D., Blayney D. P. ve Owens A. (2010). An analysis of US household dairy demand. US Department of Agriculture, Economic Research Service, USA
- Davis, C. G., Moussie, M., Dinning, J. S. ve Christakis, G. J. (1983). "Socioeconomic determinants of food expenditure patterns among racially different low-income households: an empirical analysis". *Western Journal of Agricultural Economics*, 8(2): 183-196.
- Deaton, A. & Paxson, C. (1998). "Economies of scale, household size, and the demand for food". *Journal of political economy*, 106(5): 897-930.
- Demiralay, M. ve Çamurcu, A. Y. (2005). "Cure, agnes ve k-means algoritmalarındaki kümeleme yeteneklerinin karşılaştırılması". *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 4(8): 1-18.
- Devedzic, V. (2002). Knowledge discovery and data mining in databases. Fon-School of Business Administration, University of Belgrade, Yugoslavia.
- Dinçer, E. (2006). Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması. Yayımlanmamış Yüksek Lisans Tezi. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli
- Ding, Q. & Sundarraj, G. (2006). "Association Rule Mining from XML Data". *Conference on Data Mining In DMIN06*, 26-29 June 2006, Las Vegas, Nevada, USA, s. 144-152.
- Diri, B., "Birliktelik Kuralları".
<https://www.ce.yildiz.edu.tr/personal/banud/file/22082/BirliktelikKurallari.ppt>. (erişim tarihi: 19.06.2016)
- Donskoy, V. I. (2013). "Splitting criteria, binary decision tree synthesis, and algorithm LISTBB". *Taurida Journal of Computer Science Theory and Mathematics*, 1(22): 11-34.
- Dönek, E. (1996). "Tüketim Modellerinde Gelir Dağılımı ve Yaş Yapısı: İktisat Politikası Açısından Bir Değerlendirme". *Hacettepe Üniversitesi İ.İ.B.F. Dergisi*, 14(2): 81-105.
- Dreyfus, G. (ed.). (2005). *Neural networks: methodology and applications*. Springer Science & Business Media, Germany
- Eke, B., (1987). "Bir Sosyal Sınıf Belirleyicisi Olarak Meslek Faktörü". *İktisat Fakültesi Mecmuası*, 43(1-4): 377-401
- Erkan, H. (1991). *Ekonomi Sosyolojisi*. Alper Matbaası, İzmir
- Ertek, T., (2009). *Temel Ekonomi (Basından Örneklerle)*. Beta Basın Yayın Dağıtım, İstanbul

- Esenkar, T., “Veri Madenciliği ve İstatistik”. <http://blog.visilabs.com/veri-madenciligi-ve-istatistik/>. (erişim tarihi: 02.03.2014).
- Esposito, F., Malerba, D., Semeraro, G. & Kay, J. A. (1997). “A comparative analysis of methods for pruning decision trees”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5): 476-491.
- Esposito, F., Malerba, D., Semeraro, G. & Tamma, V. (1999). “The effects of pruning methods on the predictive accuracy of induced decision trees”. *Applied Stochastic Models in Business and Industry*, 15(4): 277-299.
- Fang, C., & Beghin, J. C. (2002). “Urban demand for edible oils and fats in China: Evidence from household survey data”. *Journal of Comparative Economics*, 30(4): 732-753.
- Fayyad, U. M., & Irani, K. B. (1992). “The attribute selection problem in decision tree generation”. In *AAAI*, 12–16 July, 1992, San Jose, California, s. 104-110.
- Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. (1996). “From Data Mining to Knowledge Discovery in Databases”. *Artificial Intelligence Magazine*, 17(3): 37-54.
- Ferri, C., Flach, P. & Hernández-Orallo, J. (2002). “Learning decision trees using the area under the ROC curve”. In *ICML*, 8-12 July 2002, San Francisco, CA, USA, Vol. 2, s. 139-146.
- Firat, A. F. (1991). “Consumption and gender: a common history. In *Gender and Consumer Behavior*”, Association for Consumer Research Volume 1, Salt Lake City, s. 378-386.
- Fisunoğlu, H. M. ve Şengül, S. (2011). “Adana Kentsel Alanda Hanehalkı Tüketimi”. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 20(1): 251-266.
- Friel, S., Kelleher, C. C., Nolan, G. & Harrington, J. (2003). “Social diversity of Irish adults nutritional intake”. *European Journal of Clinical Nutrition*, 57(7): 865-875.
- Ganesh, S. (2002). “Data mining: Should it be included in the statistics curriculum?”. In *The 6th international conference on teaching statistics (ICOTS 6)*, 7-12 July 2002, Cape Town, South Africa.
- García, T. & Grande, I. (2010). “Determinants of food expenditure patterns among older consumers. The Spanish case”. *Appetite*, 54(1): 62-70.
- Gartner Group., “Data Mining (Gartner IT Glossary)”. <http://www.gartner.com/it-glossary/data-mining>. (erişim tarihi: 02.03.2014)
- Gimeno, E., Fito, M., Lamuela-Raventos, R. M., Castellote, A. I., Covas, M., Farre, M., De La Torre-Boronat M.C. & Lopez-Sabater, M. C. (2002). "Original Communications-Effect Of Ingestion Of Virgin Olive Oil On Human Low-Density Lipoprotein Composition", *European Journal Of Clinical Nutrition*, 56(2): 114-120

- Gould, B. W. (2002). "Household composition and food expenditures in China". *Agribusiness*, 18(3): 387-407.
- Gould, B. W., Cox, T. L. & Perali, F. (1991). "Demand for food fats and oils: the role of demographic variables and government donations". *American Journal of Agricultural Economics*, 73(1): 212-221.
- Grąbczewski, K. (2014). *Meta-learning in decision tree induction*. Springer, Springer Cham Heidelberg New York Dordrecht London
- Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques 2nd Edition*, Morgan Kaufmann Publishers, San Francisco, CA
- Han, J., Pei, H. & Yin, Y. (2000). "Mining Frequent Patterns without Candidate Generation". *SIGMOD'00*, 29(2): 1-12.
- Hand, D., Mannila, H. & Smyth, P. (2001). *Principles of Data Mining*, MIT Press, Cambridge, MA
- Haykin, S. (2009). *Neural Networks and Learning Machines (Vol. 3)*, Pearson Education, New Jersey
- Heart Org, "Know Your Fats".
http://www.heart.org/HEARTORG/Conditions/Cholesterol/PreventionTreatmentofHighCholesterol/Know-Your-Fats_UCM_305628_Article.jsp. (erişim tarihi: 03.10.2015)
- Heaton Research, "Jordan Neural Network".
http://www.heatonresearch.com/wiki/Jordan_Neural_Network. (erişim tarihi: 10.03.2015)
- Houthakker, H. S. (1957). "An international comparison of household expenditure patterns, commemorating the centenary of Engel's law". *Econometrica, Journal of the Econometric Society*, 25(4): 532-551.
- Huang, K. S. & Lin, B. H. (2000). *Estimation of food demand and nutrient elasticities from household survey data*, US Department of Agriculture Economic Research Service, Washington DC, USA
- Hulshof, K. F. A. M., Brussaard, J. H., Kruizinga, A. G., Telman, J. & Löwik, M. R. H. (2003). "Socio-economic status, dietary intake and 10y trends: the Dutch National Food Consumption Survey". *European Journal of Clinical Nutrition*, 57(1): 128-137.
- Hupkens C. L., Knibbe, R. A. & Drop M. J. (1997). "Social class differences in women's fat and fibre consumption: a cross-national study". *Appetite*, 28(2): 131-149.
- IBM, (1996). *Parallel Mining of Association Rules Design, Implementation and Experience*. IBM Research Division Almaden Research Center, San Jose, CA

International Olive Council, “World Olive Oil Figures”.

<http://www.internationaloliveoil.Org/Estaticos/View/131-World-Olive-Oil-Figures>.

(erişim tarihi: 12.07.2015)

Ippolito, P. M. & Mathios, A. D. (1995). “Information and advertising: The case of fat consumption in the United States”. *The American Economic Review*, 85(2): 91-95.

Irmak, S. (2009) Veri Madenciliği Yöntemleri ile Sağlık Sektörü Veritabanlarında Bilgi Keşfi: Tanımlayıcı ve Kestirimci Model Uygulamaları. Yayınlanmamış Doktora Tezi. Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Antalya

Jacobson, D., Mavrikiou, P. M. & Minas, C. (2010). “Household size, income and expenditure on food: The case of Cyprus”. *The Journal of Socio-Economics*, 39(2): 319-328.

Jae, M. K., Ryu, J. S. & Abdel-Ghany, M. (2000). “Family characteristics and convenience food expenditure in urban Korea”. *Journal of Consumer Studies & Home Economics*, 24(4): 252-256.

Januszewska, R. & Viaene, J. (2005). “Food purchase behaviour of different consumer groups in Belgium”. *Journal of International Food & Agribusiness Marketing*, 17(2): 213-227.

Johnson, A. R. & Wichern, D. W. (1992). *Applied multivariate statistical analysis*. Sixth Edition, Pearson Prentice Hall, New Jersey

Kaderli, Y. C. (2006). İstanbul Menkul Kıymetler Borsası'nda (İMKB) İşlem Gören Gıda Sektörü Şirketlerinin Finansal Performansının Ölçülmesi ve Özsermaye Kârlılığını Etkileyen Faktörlerin Belirlenmesi. Yayınlanmamış Yüksek Lisans Tezi, Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü, Aydın

Kalınlı, A. (2002). Elman Ağının Simulated Annealing Algoritması Kullanarak Sistem Kimliklendirme İçin Eğitilmesi”. *Osmangazi Üniversitesi Müh. Mim. Fak. Dergisi*, 16(2): 25-42

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons, Hoboken, New Jersey

Karafakıoğlu, M. (2006). *Pazarlama ilkeleri*. Literatür Yayıncılık, İstanbul

Karakaş, M., (2001). “Tüketim Kültürü ve Tüketimin Yeniden Üretimi”. *Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 3(1): 11-27

Karakaya, E. ve Akbay C. (2013). “İstanbul İlinde Tüketicilerin Süt ve Süt Ürünleri Tüketim Alışkanlıkları”. *Uludağ Üniv. Ziraat Fakültesi Dergisi*, 27(1): 65-77

Kaufman, L. & Rousseeuw, P. J. (1990). *Finding Groups in Data*, John Wiley & Sons, Inc., Hoboken, New Jersey

Kdnuggets, “Data Mining Methodology”.

http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm (erişim tarihi: 16.02.2014)

Kdnuggets, Gregory Piatetsky-Shapiro, “Difference between Data Mining and Statistics”.

<http://www.kdnuggets.com/faq/difference-data-mining-statistics.html>. (erişim tarihi: 16.07.2015)

Khan A. Q., Azam M. & Qureshi M. E. (2014). “Analysis of consumption behaviour concerning current income and lags consumption: Empirical evidence from Pakistan”. *Ekonomski Vjesnik/Econviews: Review of contemporary business, entrepreneurship and economic issues*, 27(1): 59-70.

Kızılgöl, Ö. A. (2012). “Türkiye’de eğitimde cinsiyet eşitsizliğinin yoksulluk üzerindeki etkisi”. *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(1): 179-191.

Kim, D. (2008). “Demand and pricing in the US margarine industry”. *Journal of Agricultural & Food Industrial Organization*, 6(1): 1-19

Kogan, J., Nicholas C., & Teboulle M. (ed.). (2006). *Grouping multidimensional data*. Springer, The Netherlands

Kolsarıcı, Ö., Gür, A., Başalma, D., Kaya, M.D. ve İşler, N. (2006). " Yağlı Tohumlu Bitkiler Üretimi". *Tarım ve Mühendislik*, 78-79: 65-78.

Kostakis, I. (2014). “The Determinants Of Households’ food Consumption In Greece”. *International Journal Of Food And Agricultural Economics (Ijfaec)*, 2(2): 17-28

Koyuncugil, A. S. (2006). *Bulanık Veri Madenciliği ve Sermaye Piyasalarına Uygulanması*. Yayınlanmamış Doktora Tezi. Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara

Koyuncugil, A. S. ve Özgülbaş, N. (2009). “Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları”. *International Journal Of Informatics Technologies*, 2(2): 21-32

Kröse, B. & Van Der Smagt, P. (1996). *An Introduction to Neural Networks*, The University of Amsterdam, Amsterdam

Kulluk, S. (2009). *Karınca Koloni Optimizasyonu ile Yapay Sinir Ağlarından Kural Çıkarımı*. Yayınlanmamış Doktora Tezi. Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Kayseri

Kuonen, D., “Data Mining and Statistics: What is the Connection?”. <http://www.tdan.com/view-articles/5226/>. (erişim tarihi: 02.03.2014)

- Kurenkov, A., "A Brief History of Neural Nets and Deep Learning, Part 3".
<http://www.andreykurenkov.com/writing/a-brief-history-of-neural-nets-and-deep-learning-part-3/>. (erişim tarihi: 12.05.2016)
- Kuzey, C. (2012). Veri Madenciliğinde Destek Vektör Makinaları ve Karar Ağaçları Yöntemlerini Kullanarak Bilgi Çalışanlarının Kurum Performansı Üzerine Etkisinin Ölçülmesi ve Bir Uygulama. Yayımlanmamış Doktora Tezi. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul
- Külekcî, M., Topaloğlu, A. ve Aksoy, A. (2006). "Dondurulmuş gıda tüketimini etkileyen sosyo-ekonomik özelliklerin belirlenmesi; Erzurum ili örneği". Atatürk Üniversitesi Ziraat Fakültesi Dergisi, 39(1): 91-101.
- Lanjouw, P. & Ravallion, M. (1995). Poverty and household size. The economic journal, 105(433): 1415-1434.
- Larose, D. T. (2005). Discovering Knowledge In Data An Introduction To Data Mining, John Wiley & Sons, Inc., Hoboken, New Jersey
- Lazaridis, P., (2004). "Olive Oil Consumption In Greece: A Microeconometric Analysis". Journal Of Family And Economic Issues, 25(3): 411-430.
- Liu K. E. & Chern W. S. (2001). "Effects of Model Specification and Demographic Variables on Food Consumption: Microdata Evidence from Jiangsu, China". 11th Annual World Food and Agribusiness Forum, World Food and Agribusiness Symposium of the International Food and Agribusiness Management Association, June 2001, Sydney, Australia, s. 27-28.
- Magrabi, F. M., Chung, Y. S., Cha, S. S. & Yang S. J. (1991). The economics of household consumption. Greenwood Publishing Group, New York
- Maimon, O. & Rokach, L. (ed). (2005a). Data mining and knowledge discovery handbook (Vol. 2). Springer, New York
- Maimon, O. & Rokach, L. (2005b). Decomposition Methodology For Knowledge Discovery And Data Mining Theory And Applications. World Scientific Publishing Co. Pte. Ltd., Singapore
- Mariscal, G., Marban, O. & Fernandez, C. (2010). "A survey of data mining and knowledge discovery process models and methodologies". The Knowledge Engineering Review, 25(02): 137-166.
- Math Works., "Dendogram".
<http://www.mathworks.com/help/stats/dendrogram.html?requestedDomain=www.mathworks.com>. (erişim tarihi: 19.05.2016)

- Mattisson, I., Wirfält, E., Gullberg, B. & Berglund, G. (2001). "Fat intake is more strongly associated with lifestyle factors than with socio-economic characteristics, regardless of energy adjustment approach". *European journal of clinical nutrition*, 55(6): 452-461.
- Mayén, A. L., Marques-Vidal, P., Paccaud, F., Bovet, P. & Stringhini, S. (2014). "Socioeconomic determinants of dietary patterns in low-and middle-income countries: a systematic review". *The American journal of clinical nutrition*, 100(6): 1520-1531
- Mcgee, D., (2003). "ABD'de Perakende Zeytinyağı Sektöründe Pazarlama ve Türkiye'nin Pazar Payının Arttırılmasına Yönelik Görüşler", Türkiye I. Zeytinyağı ve Sofralık Zeytin Sempozyumu, 02-03 Ekim 2003, İzmir, s. 70-82
- MEB Milli Eğitim Bakanlığı. (2013). Gıda Teknolojisi Tereyağı, Ankara
- Mehrotra, K., Mohan, C. K. & Ranka, S. (1996). *Elements Of Artificial Neural Networks*. MIT pres, USA
- Meng, T., Florkowski, W. J., Kolavalli, S. & Ibrahim, M. (2012). "Food expenditures and income in rural households in the northern region of Ghana". In *Agricultural & Applied Economics Association, Annual Meeting*. 12-14 August 2012, Seattle, Washington,
- Michael, R. T. (1975). *Education, Income, and Human Behavior, The Impact of Education Behavior*, NBER. s. 233-252.
- Moss, L. T. & Atre, S. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*, Addison Wesley Professional, Boston
- Mullie, P., Guelinckx, I., Clarys, P., Degrave, E., Hulens, M., & Vansant, G. (2009). "Cultural, socioeconomic and nutritional determinants of functional food consumption patterns". *European journal of clinical nutrition*, 63(11): 1290-1296.
- Mümsad Mutfak Ürünleri ve Margarin Sanayicileri Derneği, "Margarin/Margarinin Tarihçesi". <http://www.mumsad.org.tr/margarin/margarinin-tarihcesi>. (erişim tarihi: 14.03.2015)
- Mümsad Mutfak Ürünleri ve Margarin Sanayicileri Derneği, "İstatistikler". <http://www.mumsad.org.tr/istatistikler/istatistikler>. (erişim tarihi: 26.05.2016)
- Nayga, R. M. (1994). "Effects of socioeconomic and demographic factors on consumption of selected food nutrients". *Agricultural and Resource Economics Review*, 23(2): 171-182.
- Nisbet, R., Elder, J. & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Academic Press Elsevier, Canada

- Odabaşı, Y. (1996). Tüketici Davranışı ve Pazarlama Stratejisi. Anadolu Üniversitesi İşletme Fakültesi Yayınları, Eskişehir
- Odabaşı, Y. (1999). Tüketim Kültürü. Sistem Yayıncılık, İstanbul
- Oğuzlar, A. (2003). “Veri ön işleme”. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 2(21): 67-76
- Oğuzlar, A. (2004). “CART Analizi ile Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi”. Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 18(3-4): 79-90
- Oktay, D. (2010). Ege Bölgesi’nde Zeytinyağı Tüketiminin Artırılmasında Arz Zinciri Yönetiminin Olası Katkıları Üzerine Bir Araştırma. Yayımlanmamış Doktora Tezi, Ege Üniversitesi Fen Bilimleri Enstitüsü, İzmir
- Olive Oil Market, “Trends In World Olive Oil Consumption - Ioc Report“. <http://www.oliveoilmarket.eu/Trends-in-World-Olive-Oil-Consumption-ioc-Report/>. (erişim tarihi: 10.07.2015)
- Olson, D. L. & Delen, D. (2008). Advanced Data Mining Techniques. Springer Science & Business Media, Berlin
- Örmeci Kart, M. Ç. ve Demircan, V. (2014). “Dünyada ve Türkiye’de Süt ve Süt Ürünleri Üretimi, Tüketimi ve Ticaretindeki Gelişmeler”. Academic Food Journal/Akademik Gıda, 12(1): 78-96
- Özdoğan, G. (2010). Öbek Bilgisayarlarda Paralel Fp-Growth Gerçekleştirimi. Yayımlanmamış Yüksek Lisans Tezi. TOBB Ekonomi ve Teknoloji Üniversitesi Fen Bilimleri Enstitüsü, Ankara
- Özer, H., (1999). Türkiye’de Hanehalkı Tüketim Harcamalarının Doğrusal Harcama Sistemi Yaklaşımıyla Analizi. Yayımlanmamış Doktora Tezi. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum
- Özmen, Ş. (2001). “İş Hayatı Veri Madenciliği ile İstatistik Uygulamalarını Yeniden Keşfediyor”. V. Ulusal Ekonometri ve İstatistik Sempozyumu, 19–22 Eylül 2001, Adana
- Öztemel E. (2012). Yapay Sinir Ağları. Papatya Yayıncılık, Ankara
- Öztürk, F., Yalçın, M., & Dıraman, H. (2009). “Türkiye zeytinyağı ekonomisine genel bir bakış”. Gıda Teknolojileri Elektronik Dergisi, 4(2): 35-51.
- Palmer, A., Gervilla, E., & Jiménez, R. (2011). Data mining: machine learning and statistical techniques. INTECH Open Access Publisher.

- Pan, S., Mohanty, S., & Welch, M. (2008). "India edible oil consumption: A censored incomplete demand approach". *Journal of Agricultural and Applied Economics*, 40(3): 821-835.
- Pandre, A., "Cluster Analysis". <https://apandre.wordpress.com/visible-data/cluster-analysis/>. (erişim tarihi: 19.05.2016)
- Paola, T., "Recurrent Neural Prediction Model for Digits Recognition". <http://www.ijser.org/paper/Recurrent-Neural-Prediction-Model-for-Digits-Recognition.html>. (erişim tarihi: 10.03.2015)
- Pedrycz, W. (2005). *Knowledge-based clustering: from data to information granules*. John Wiley & Sons, Hoboken, New Jersey
- Polat, F. (2010). *Yemeklik Yağ Sektöründe Tüketici Davranışlarını Etkileyen Faktörlerin Analizi*. Yayımlanmamış Doktora Tezi. Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara
- Prabhu, S. & Venatesan, N. (2007). *Data mining and warehousing*, New Age International (P) Limited, Publishers, New Delhi
- Rao, C. R., Wegman, E.J. & Solka, J.L. (ed.). (2005). *Classification and Regression Trees, Bagging and Boosting*, *Handbook of Statistics 24: Data Mining and Data Visualization*, Elsevier B.V., The Netherlands
- Raper, K. C., Wanzala, M. N. & Nayga Jr, R. M. (2002). "Food expenditures and household demographic composition in the US: a demand systems approach". *Applied Economics*, 34(8): 981-992.
- Ray, R. (1980). "Analysis of a time series of household expenditure surveys for India". *The Review of Economics and Statistics*, 62(4): 595-602.
- Ricciuto, L., Tarasuk, V. & Yatchew, A. (2006). "Socio-demographic influences on food purchasing among Canadian households". *European Journal of Clinical Nutrition*, 60(6): 778-790.
- Rokach, L. & Maimon, O. (2014). *Data mining with decision trees: theory and applications* (2nd edition). World Scientific Publishing Co. Pte. Ltd., Singapore
- Roos, E., Lahelma, E., Virtanen, M., Prättälä, R. & Pietinen, P. (1998). "Gender, socioeconomic status and family status as determinants of food behaviour". *Social science & medicine*, 46(12): 1519-1529.
- Sachs, J. D. & Larrain, F. B. (1993). *Macroeconomics in the global economy*, Prentice Hall, New Jersey

- Salford Systems. (2001). CART Tree Structured Non-Parametric Data Analysis. Salford Systems, San Diego, USA
- San, O. M., Huynh, V. N. & Nakamori, Y. (2004). "An alternative extension of the k-means algorithm for clustering categorical data". International Journal of Applied Mathematics and Computer Science, 14(2): 241-248.
- Sayad, S., "Hierarchical Clustering". http://www.saedsayad.com/clustering_hierarchical.htm. (erişim tarihi: 15.09.2015)
- Sayılı, M., Gözener, B., Kalpaklıoğlu, G. (2013). "Ordu İli Merkez İlçede Ailelerin Fındık Yağı Tüketim Durumları". Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi, 30(1): 84-90
- Sayili, M., Akca, H. & Vuray, I. (2005). "Edible Oil and Fat Consumption and Income-Expenditure Elasticity: A Cross Section Study". Journal of Applied Sciences, 5(4): 716-719.
- Sekhampu, T. J. (2012). "Socio-economic determinants of household food expenditure in a low income township in South Africa". Mediterranean Journal of Social Sciences, 3(3): 449-453
- Sumathi, S. & Sivanandam, S. N. (2006). Introduction to data mining and its applications, Springer, Berlin
- Sunay, S. A. (2010). Kalpteki Fokal Ventriküler Aritmi Kaynağının Yerinin Tespitinde Kümeleme ve Sınıflandırma Uygulamaları. Yayımlanmamış Doktora Tezi. Başkent Üniversitesi Fen Bilimleri Enstitüsü, Ankara
- T.C. Gümrük ve Ticaret Bakanlığı Kooperatifçilik Genel Müdürlüğü. (2014). 2013 Yılı Ayçiçeği Raporu, Ankara
- T.C. Gümrük ve Ticaret Bakanlığı Kooperatifçilik Genel Müdürlüğü. (2015). 2014 Yılı Ayçiçeği Raporu, Ankara
- Tan, P. N., Steinbach, M. ve Kumar, V. (2006). Introduction to data mining (Vol. 1). Pearson Addison Wesley, Boston
- Taniar, D. (ed.). (2008). Data Mining and Knowledge Discovery Technologies, IGI Publishing, Hershey, New York
- Tapkan, P., Özbakır, L. ve Baykasoğlu, A. "Weka ile Veri Madenciliği Süreci ve Örnek Uygulama". Endüstri Mühendisliği Yazılımları ve Uygulamaları Kongresi, 30 Eylül-02 Ekim 2011, İzmir, s. 247-262

- Tarapanoff, K., Quoniam, L., Araújo Júnior, R. H. D. and Alvares, L. M. A. D. R. (2001). "Intelligence obtained by applying data mining to a database of French theses on the subject of Brazil". *Information Research*, 7(1)
- Tarı, R. ve Pehlivanoglu F. (2007). "Kocaeli İlinde Tüketici Davranışlarının Gelir-Harcama Grupları İlişkisi Açısından Analizi (Tüketim Harcamaları Profili)". *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 1(13): 192-210
- Tariş, "Zetinyayağının İnsan Sağlığı ve Beslenmesindeki Önemi".
http://www.tariszeytinyagi.com/index.php?option=com_content&view=article&id=16&Itemid=15. (erişim tarihi: 09.07.2015)
- Taylor, B. J. (ed.). (2006). *Methods and procedures for the verification and validation of artificial neural networks*. Springer Science & Business Media, USA
- TEAE Tarımsal Ekonomi Araştırma Enstitüsü. (2005). *Yağlı Tohumlar & Bitkisel Yağlar Durum ve Tahmin 2005/2006*, Yayın No: 136, Ankara
- Tekçe, A., Türkiye’de Ayçiçeği Tohumu ve Ayçiçeği Yağı Üretimi, Arz Talep Dengesi, Yağlı Tohumlu Bitkiler ve Bitkisel Yağlar Konferansı 2014, İstanbul (www.bysd.org.tr, Raporlar)
- TEPGE Tarımsal Ekonomi ve Politika Geliştirme Enstitüsü. (2014). *Durum ve Tahmin Süt ve Süt Ürünleri 2014*, Yayın No: 233, Ankara
- Terin, M., Bilgiç A., Güler, İ. O., Yavuz, F. (2015). "Türkiye’de Süt Ürünleri Tüketim Harcamalarına Etki Eden". *Tarım Bilimleri Dergisi*, 21(4): 500-515.
- Thearling, K., "An Introduction to Data Mining".
www.thearling.com/text/dmwhite/dmwhite.htm. (erişim tarihi: 16.02.2014)
- Timofeev, R., (2004). *Classification and Regression Trees (CART) Theory and Applications*. Yayımlanmamış Yüksek Lisans Tezi. Humboldt University CASE - Center of Applied Statistics and Economics, Berlin,
- Tiryaki, Y. G. (2008), "Household's Olive Oil Consumption Preferences Socioeconomic And Demographic Differences". *Agro Food Industry Hi-Tech*, 19(5): 26-29.
- Tiryaki, Y. G. ve Akbay, C., (2005). "Türkiye’de Ailelerin Sosyo-Ekonomik Gruplar İtibariyle Zeytinyağı Tüketimi". *Zeytinyağı ve Pirina Yağı Sempozyum ve Sergisi*, TMMOB Kimya Mühendisleri Odası 50. Yıl Etkinliği, 10-12 Kasım 2005, İzmir, s. 381-391.
- Top, B. T. ve Uçum, İ. (2012). "Türkiye’de bitkisel yağ açığı". *Tarımsal Ekonomi ve Politika Geliştirme Enstitüsü Tepge Bakış*, 14(2): 1-8

- Tumer, E. I. (2013). "Socioeconomic and demographic characteristics affecting household's table olive and olive oil consumption". *Agro FOOD Industry Hi Tech*, 24(3): 53-55
- Tunalıoğlu, R. Tiryaki, G. Y. Tan, S. ve Taşkaya, B., (2003). "Dünya Zeytinyağı Tüketimindeki Gelişmeler; Bu Gelişmeyi Destekleyen Çalışmalar ve Türkiye Zeytinyağı Tüketimindeki Değişimler". *Türkiye I. Zeytinyağı ve Sofralık Zeytin Sempozyumu*, 02-03 EKİM 2003, İzmir, s. 49-58.
- TÜİK. (2008). *Tüketim Harcamaları, Yoksulluk ve Gelir Dağılımı. Sorularla Resmi İstatistikler Dizisi-6. Türkiye İstatistik Kurumu Matbaası, Ankara*
- TÜİK. (2009). *2009 TÜİK Hane Halkı Bütçe Anketi Verileri CD*
- TÜİK. (2010). *2010 TÜİK Hane Halkı Bütçe Anketi Verileri CD*
- TÜİK. (2011a). *2011 TÜİK Hane Halkı Bütçe Anketi Verileri CD*
- TÜİK. (2011b). *Hanehalkı Bütçe Araştırması Tüketim Harcamaları 2009. Türkiye İstatistik Kurumu Matbaası, Ankara*
- TÜİK. (2012). *2012 TÜİK Hane Halkı Bütçe Anketi Verileri CD*
- TÜİK, "Eş Değerlilik Ölçeği".
http://www.tuik.gov.tr/MicroVeri/GYKA_2011/turkce/metaveri/tanim/essdeggerlik-oelccegi/index.html (erişim tarihi: 12.07.2015)
- Türk Gıda Kodeksi, "Tereyağı, Diğer Süt Yağı Esaslı Sürülebilir Ürünler ve Sadeyağ Tebliği (Tebliğ No: 2005/19)".
<http://mevzuat.basbakanlik.gov.tr/Metin.Aspx?MevzuatKod=9.5.8208&MevzuatIliski=0&sourceXmlSearch=sürülebilir>. (erişim tarihi: 15.12.2015)
- Türkiye Kalkınma Bankası A.Ş. (1999). *Sektörel Araştırmalar Margarin Sektörü, TKB Matbaası, Ankara*
- Tüzüntürk, S. (2010). "Veri Madenciliği ve İstatistik". *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi*, 29(1): 65-90
- Two Crows Corporation (1999). *Introduction to Data Mining and Knowledge Discovery Third Edition, Two Crows Corporation, Potomac*
- TZOB. (2010). "Zirai ve İktisadi Rapor 2007-2010", *Türkiye Ziraat Odaları Birliği, Ankara*
- Uluskar, F. (2014). *Muğla İlinde Zeytinyağı Kullanımı. Yayımlanmamış Yüksek Lisans Tezi. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara*
- USK Ulusal Süt Konseyi. (2014). *Dünya ve Türkiye'de Süt Sektör İstatistikleri 2013, Ankara*
- Üstünel, B., (1997), *Ekonominin Temelleri, Alfa Yayıncılık, İstanbul*
- van Diepen, M. & Franses, P. H. (2006). "Evaluating chi-squared automatic interaction detection". *Information Systems*, 31(8): 814-831.

- Vatansever, M. (2008). Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizlerinde Kullanımı ve Uygulanması. Yayınlanmamış Doktora Tezi. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul
- Vural, B. B. (2007). Yapay Sinir Ağları ile Finansal Tahmin. Yayınlanmamış Yüksek Lisans Tezi. Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Ankara
- Wang, Z., Dang, S. & Yan, H. (2010). “Nutrient intakes of rural Tibetan mothers: A cross-sectional survey”. BMC public health, 10(1): 801-809
- Wong, M. L. & Leung, K. S. (2002). Data mining using grammar based genetic programming and applications. Kluwer Academic Publishers, USA
- Wu, X. & Kumar, V. (ed.). (2009). Top Ten Algorithms In Data Mining. CRC Press, Boca Raton
- Xu, Z., Wang, L., Luo, J. & Zhang, J. (2005). “A modified clustering algorithm for data mining”. In Geoscience and Remote Sensing Symposium 2005 (IGARSS’05), 2005 IEEE International, 25-29 Jul 2005, Seoul, South Korea, s.741-744
- Yakut, E., (2012). Veri Madenciliği Tekniklerinden C5.0 Algoritması ve Destek Vektör Makineleri ile Yapay Sinir Ağlarının Sınıflandırma Başarılarının Karşılaştırılması: İmalat Sektöründe Bir Uygulama. Yayınlanmamış Doktora Tezi. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum
- Yapay Sinir Ağları, “Biyolojik Sinir Hücresi - Yapay Sinir Ağları Karşılaştırması”. <http://yapaysiniraglari.com/Dersler/IcerikId/29>. (erişim tarihi: 14.05.2014)
- Yavuz, S. ve Deveci, M. (2012). “İstatiksel Normalizasyon Tekniklerinin Yapay Sinir Ağın Performansına Etkisi”. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, (40): 167-187.
- Ye, N. (ed.). (2003). The handbook of data mining (Vol. 24). Lawrence Erlbaum Associates Publishers, New Jersey
- Yen, S. T., Kan, K. & Su, S. J. (2002). “Household demand for fats and oils: two-step estimation of a censored demand system”. Applied Economics, 34(14): 1799-1806.
- Yıldırım, P., Uludağ, M. ve Görür, A. (2008). “Hastane Bilgi Sistemlerinde Veri Madenciliği”. Akademik Bilişim 2008, 30 Ocak-1 Şubat 2008, Çanakkale, s. 429-434
- Yimer, S. (2011). “Determinants of food consumption expenditure in Ethiopia.Int”. J. Econ. Res, 2(5): 151-165.
- Zhang, C. & Zhang, S. (2002). Association rule mining: models and algorithms. Springer-Verlag, Berlin

- Zhao, C. M. & Luan, J. (2006). "Data mining: Going beyond traditional statistics". *New Directions for Institutional Research*, 131: 7-16.
- Zhou, Z. H. (2003). "Three Perspectives of Data Mining Artificial Intelligence". *Artificial Intelligence*, 143(1): 139-146.
- Zurada, M. J. (1992). *Introduction to Artificial Neural Systems First Edition Edition*, By West Publishing Company, St. Paul, MN, USA



EK 1-ENFLASYONDAN ARINDIRMA İŞLEMİNDE KULLANILAN PROGRAMIN ARAYÜZÜ VE KODLARI

Satır Sütun Dosya Aç ve Oku

Satırlar Sütunlar Hesapla

Seçilen Katsayı AY YIL

Kaydet Excel Ver

			GIDA VE ALKOLSÜZ İÇECEKLER	ALKOLLO İÇECEKLER VE TÜTÜN	
2009	Ocak	164,15	1,343466341760...	192,63	1,80
2009	Şubat	164,18	1,343220855158...	192,82	1,80
2009	Mart	170,6	1,292672919109...	206,12	1,68
2009	Nisan	170	1,297235294117...	206,16	1,68
2009	Mayıs	170,32	1,294798027242...	206,24	1,68
2009	Haziran	167,66	1,3153405702016	206,27	1,68
2009	Temmuz	166,59	1,323788942913...	228,73	1,51
2009	Ağustos	163,97	1,344941147770...	232,69	1,49
2009	Eylül	164,32	1,342076436222...	232,7	1,49
2009	Ekim	169,18	1,303522875044...	233,02	1,49
2009	Kasım	172,92	1,275329632199...	232,99	1,49
2009	Aralık	176,76	1,247623896809...	232,88	1,49
2010	Ocak	179,55	1,228237259816...	284,2	1,22

Program Arayüzü

Program Kodları

```
Imports Microsoft.Office.Interop
```

```
Public Class Form1
```

```
Private Sub Form1_Load(sender As Object, e As EventArgs) Handles MyBase.Load
```

```
Dim dosya As New Excel.Application
```

```
dosya.Workbooks.Open(Application.StartupPath & "/katsayi.xlsx")
```

```
dgKatsayi.RowCount = 48
```

```
dgKatsayi.ColumnCount = 28
```

```
For i = 0 To 27
```

```
dgKatsayi.Columns(i).HeaderText = dosya.Cells(1, i + 1).value
```

```
Next
```

```
For i = 2 To 49
```

```
For j = 1 To 28
```

```
dgKatsayi.Rows(i - 2).Cells(j - 1).Value = dosya.Cells(i, j).value
```

```
Next
```

```
Next
```

```
dosya.Quit()
```

```

End Sub

Private Sub dgKatsayi_CellClick(sender As Object, e As
DataGridViewCellEventArgs) Handles dgKatsayi.CellClick
    txtKatsayi.Text = dgKatsayi.CurrentCell.Value
    If e.ColumnIndex <> 0 Then
        lbAy.Text = dgKatsayi.CurrentRow.Cells(0).Value & " " &
dgKatsayi.CurrentRow.Cells(1).Value & " " & dgKatsayi.Columns(e.ColumnIndex -
1).HeaderText
    End If
End Sub

Private Sub btnAc_Click(sender As Object, e As EventArgs)
End Sub

Private Sub btnOku_Click(sender As Object, e As EventArgs) Handles btnOku.Click
    If txtSat.Text = "" Or txtSut.Text = "" Then
        MsgBox("Önce Satır Sutun Giriniz", MsgBoxStyle.Critical, "Hata")
        Return
    End If
    Dim sat As Integer = Convert.ToInt32(txtSat.Text)
    Dim sut As Integer = Convert.ToInt32(txtSut.Text)
    OpenFileDialog1.Filter = "Excel 2007 Dosyaları|*.xlsx|Excel 2003
Dosyaları|*.xls"
    If OpenFileDialog1.ShowDialog = Windows.Forms.DialogResult.Cancel Then
        Return
    End If
    Dim dosya As New Excel.Application
    dosya.Workbooks.Open(OpenFileDialog1.FileName)
    dgData.ColumnCount = sut + 1
    dgData.RowCount = sat
    For i = 0 To dgData.RowCount - 1
        dgData.Rows(i).Cells(0).Value = i
    Next
    For i = 1 To sat
        For j = 1 To sut
            dgData.Rows(i - 1).Cells(j).Value = dosya.Cells(i, j).value
        Next
        lbSonuc.Text = i & ". Kayıt Okunuyor..."
        Application.DoEvents()
    Next
    dosya.Quit()
    lbSonuc.Text = "İşlem Tamam"
End Sub

```

```

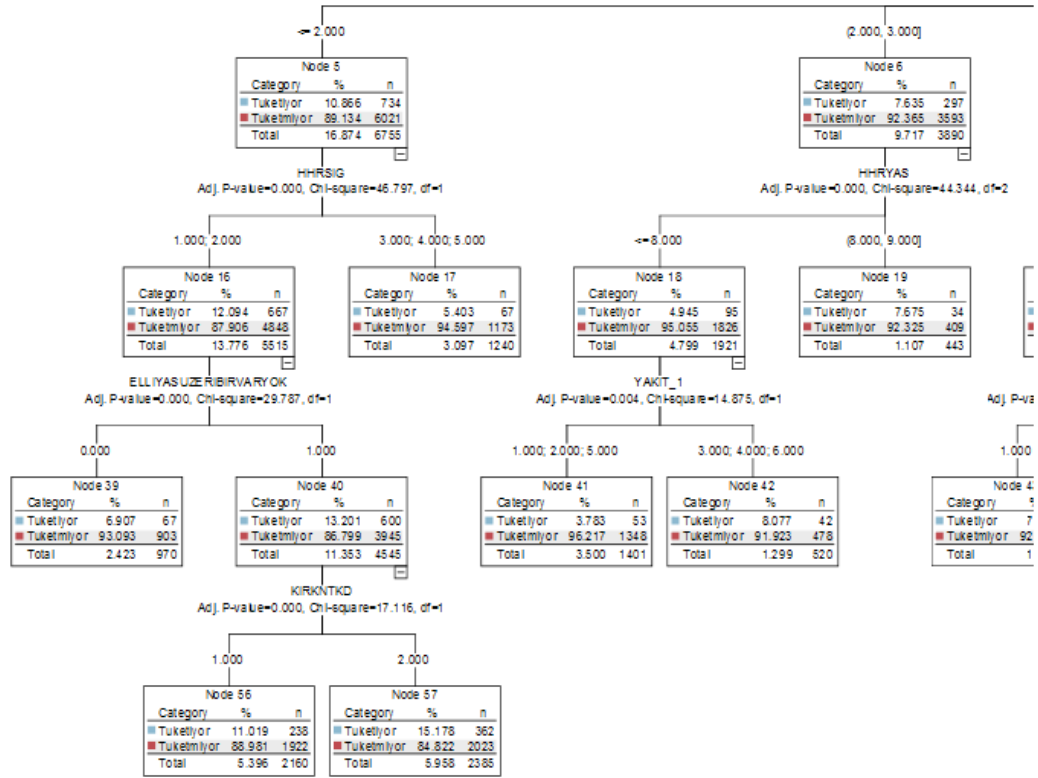
Private Sub Button1_Click(sender As Object, e As EventArgs) Handles
Button1.Click
    If txtKatsayi.Text = "" Then
        MsgBox("Önce Katsayı Seçiniz", MsgBoxStyle.Critical, "Hata")
        Return
    End If
    Dim sat1 As Integer = Convert.ToInt32(txtSat1.Text)
    Dim sat2 As Integer = Convert.ToInt32(txtSat2.Text)
    Dim sut1 As Integer = Convert.ToInt32(txtSut1.Text)
    Dim sut2 As Integer = Convert.ToInt32(txtSut2.Text)
    Dim katsayi As Double = Convert.ToDouble(txtKatsayi.Text)
    Dim deger As Double = 0
    For i = sat1 To sat2
        For j = sut1 To sut2
            deger = dgData.Rows(i).Cells(j + 2).Value
            dgData.Rows(i).Cells(j + 2).Value = deger * katsayi
        Next
    Next
    MsgBox("İşlem Tamam")
End Sub

Private Sub btnKaydet_Click(sender As Object, e As EventArgs) Handles
btnKaydet.Click
    Randomize()
    SaveFileDialog1.FileName = "veri_" & Int(Rnd() * 1000)
    If SaveFileDialog1.ShowDialog = Windows.Forms.DialogResult.Cancel Then
        Return
    End If
    Dim dosya As New Excel.Application
    dosya.Workbooks.Add()
    For i = 0 To dgData.RowCount - 1
        For j = 0 To dgData.ColumnCount - 1
            dosya.Cells(i + 1, j + 1).value = dgData.Rows(i).Cells(j).Value
        Next
        lbSonuc.Text = i & ". Kayıt Okunuyor..."
        Application.DoEvents()
    Next
    dosya.ActiveWorkbook.SaveAs(SaveFileDialog1.FileName)
    MsgBox("İşlem Tamam")
    lbSonuc.Text = "İşlem Tamam"
End Sub
End Class

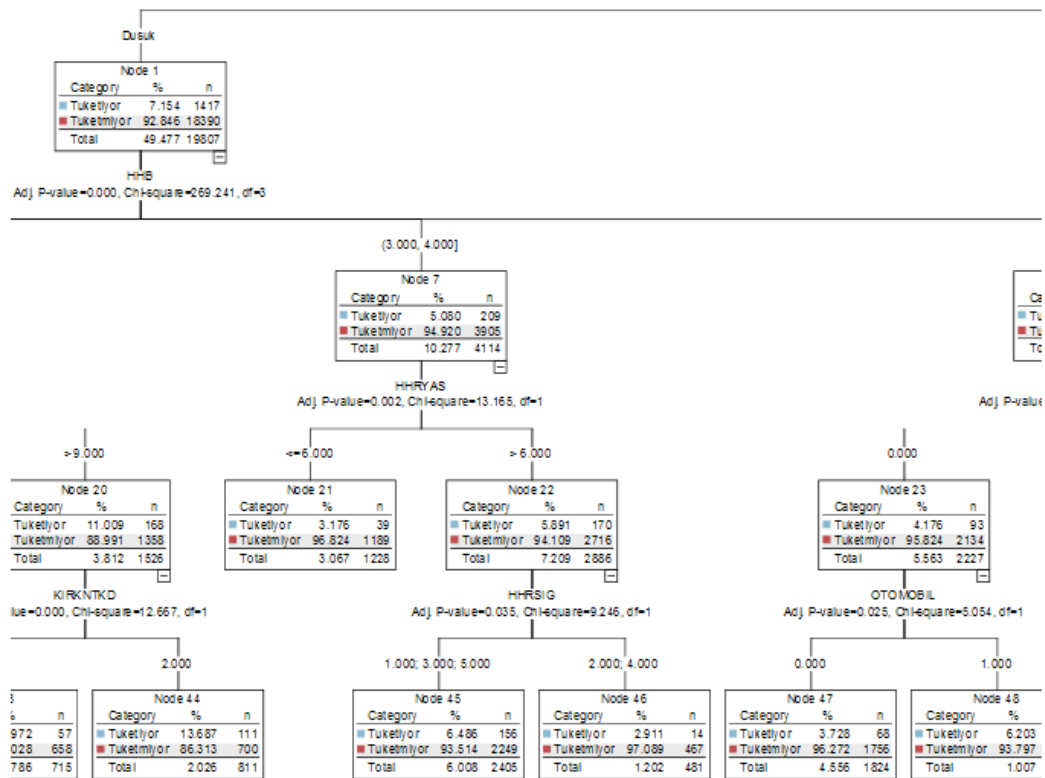
```

EK 2-ZEYTİNYAĞI KARAR AĞACI

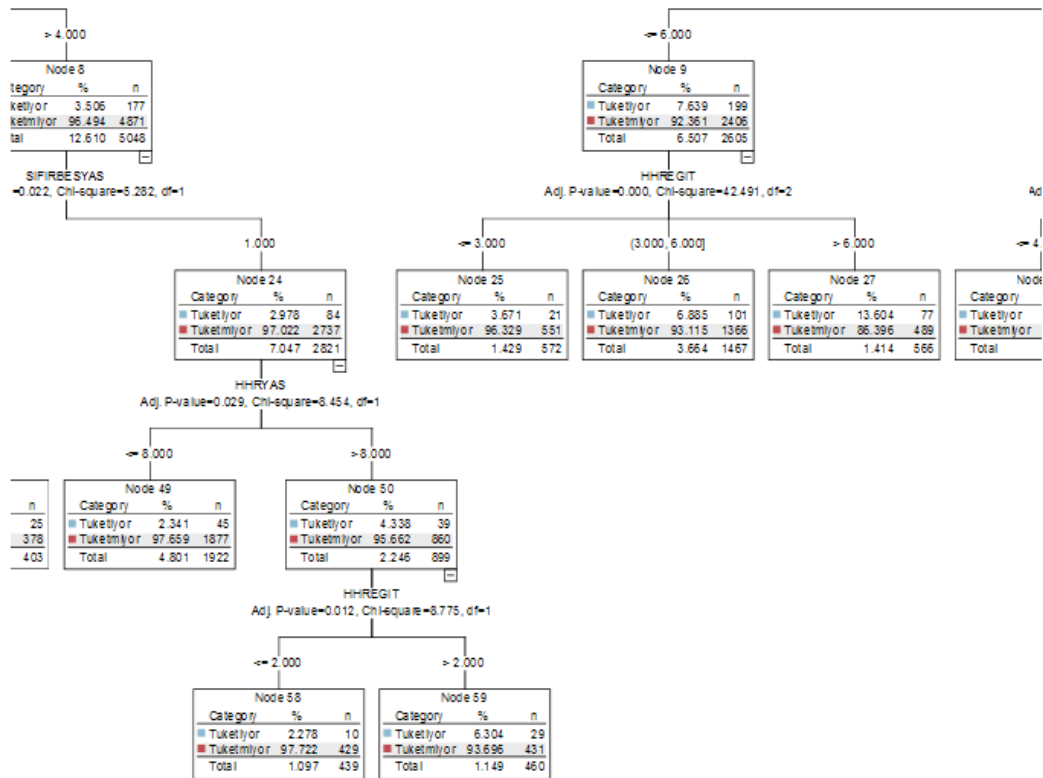
ZeytinyagiC



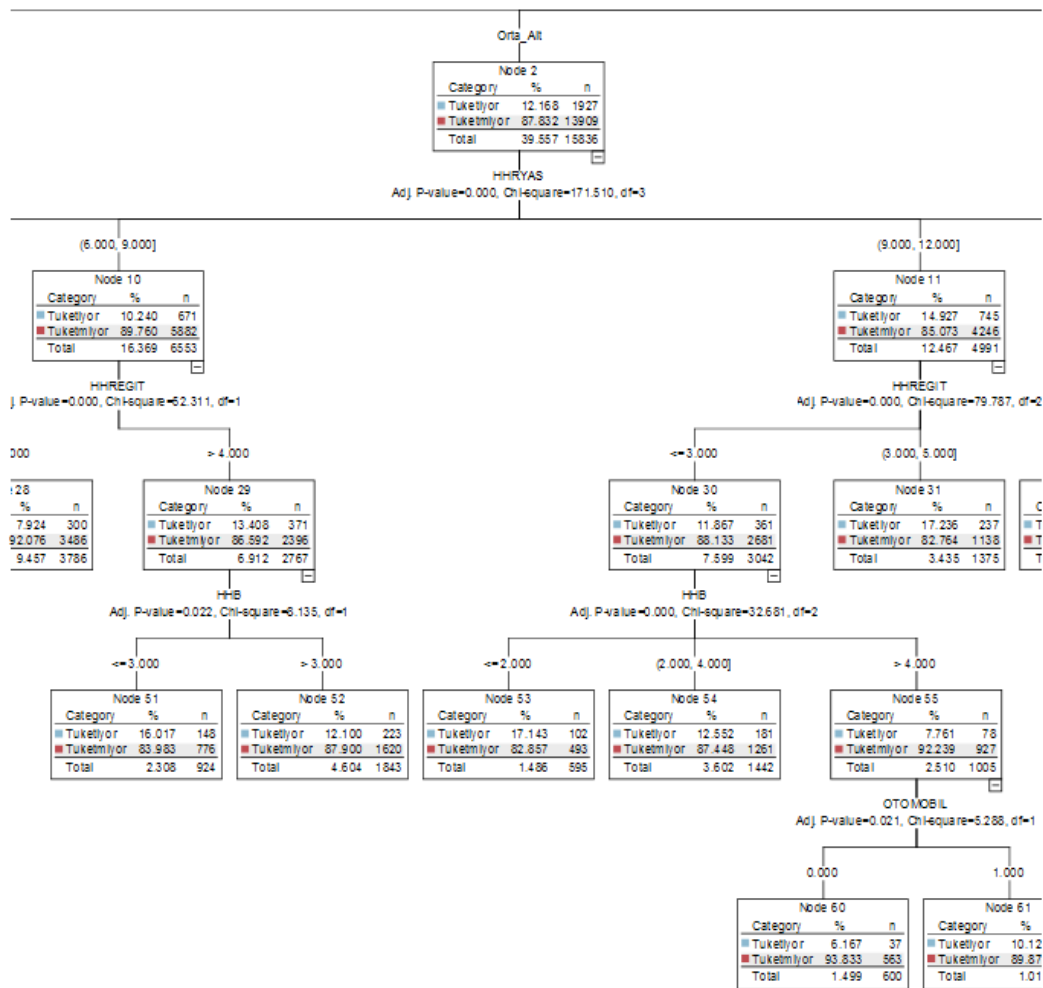
ZeytinyagiC



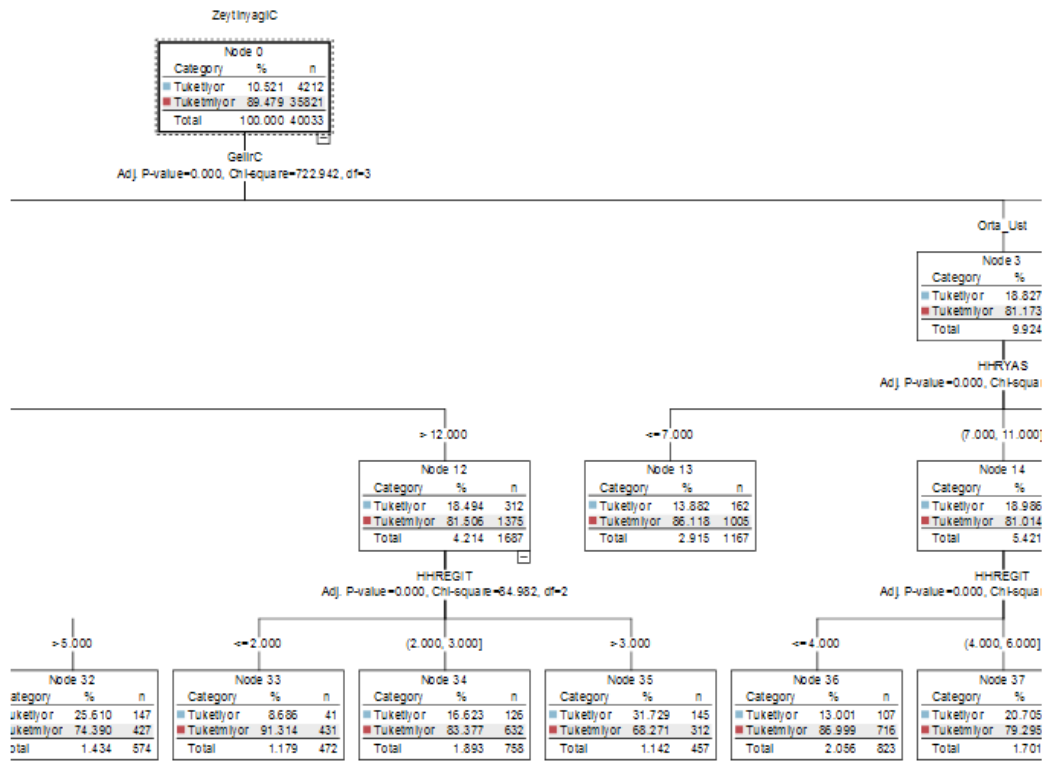
ZeytinyagiC



ZeytinyagiC



ZeytinyagiC



	n
3	41
7	384
2	405

ZeytinyagiC

Yuksekk		
Node 4		
Category	%	n
Tuketiyor	28.777	120
Tuketmiyor	71.223	297
Total	1.042	417

e=49.002, df=2

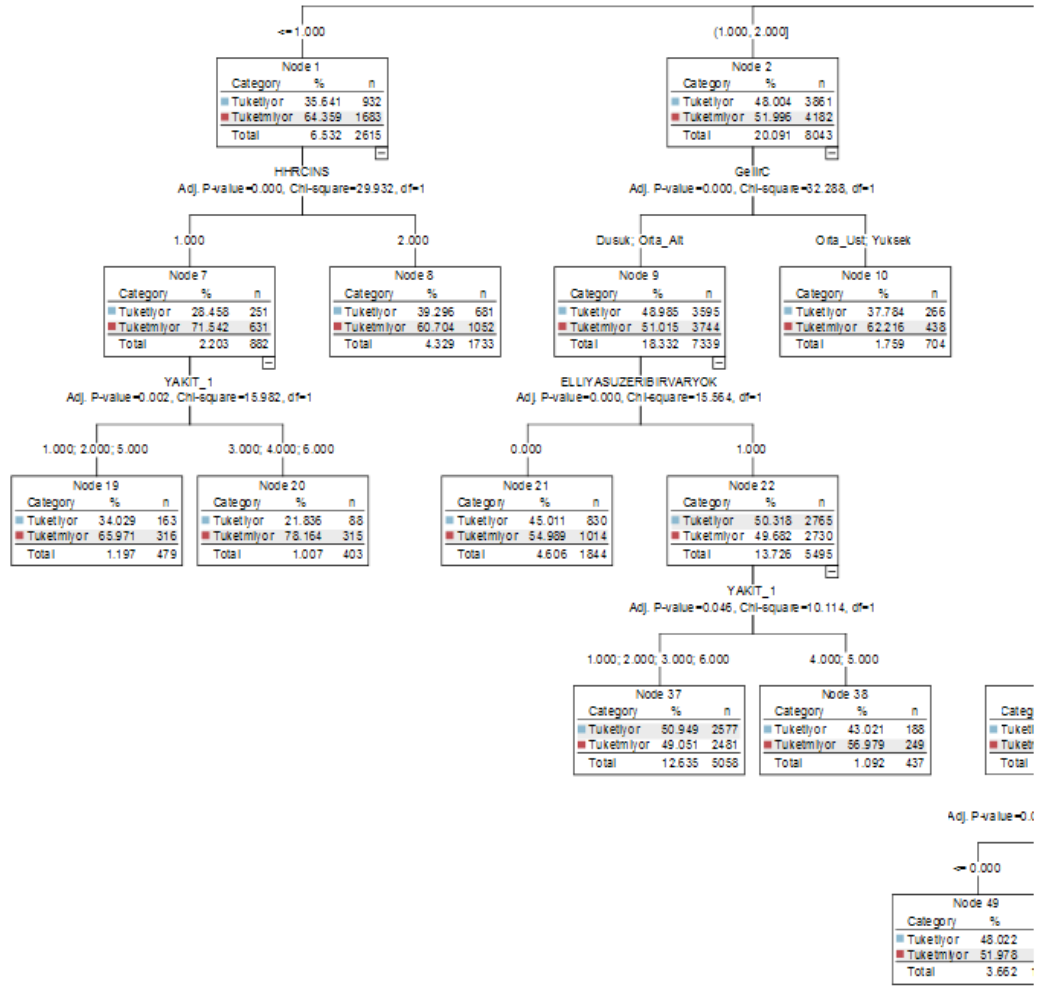
> 11.000		
Node 15		
Category	%	n
Tuketiyor	27.358	174
Tuketmiyor	72.642	462
Total	1.589	636

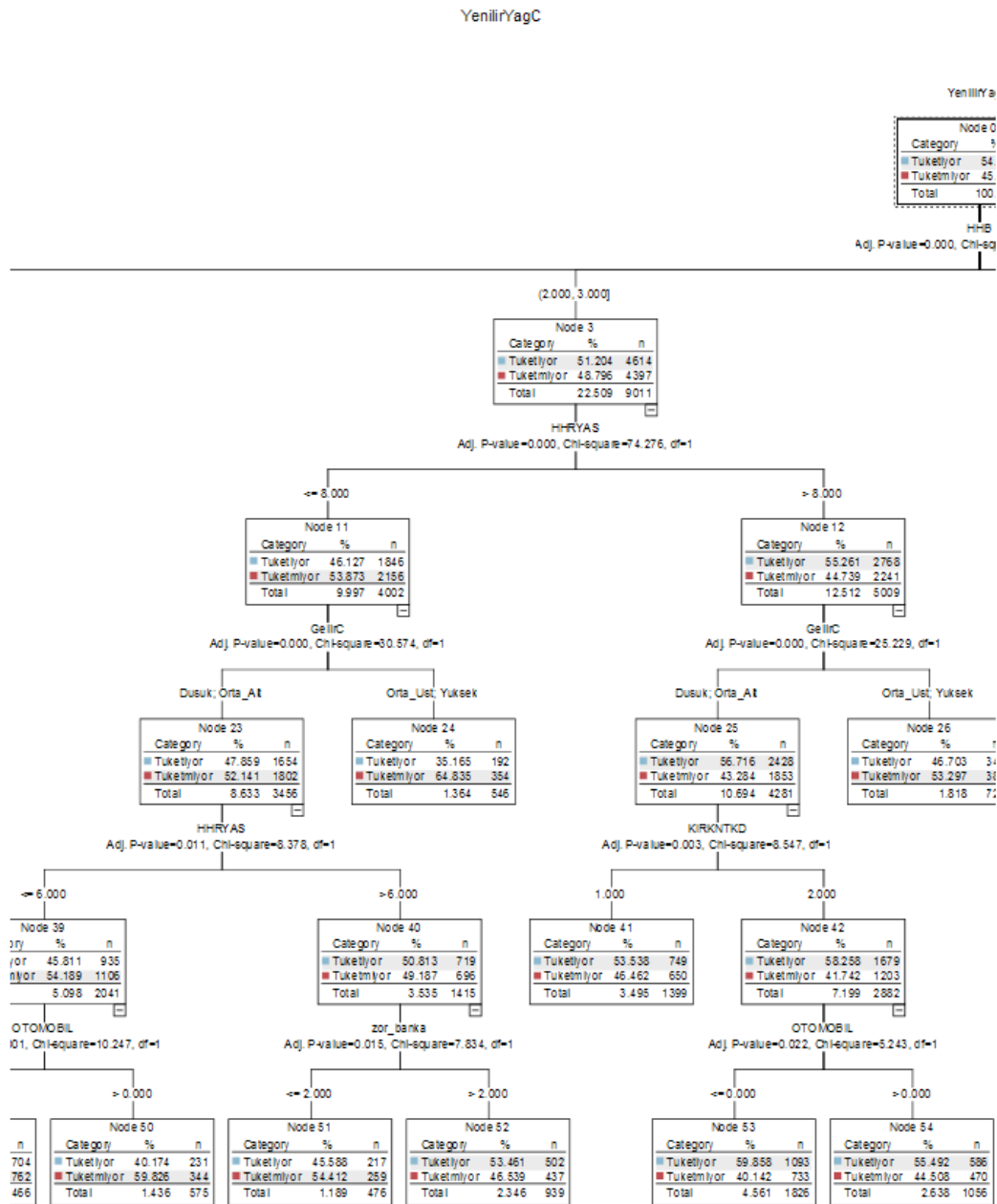
e=34.239, df=2

> 6.000		
Node 38		
Category	%	n
Tuketiyor	24.625	164
Tuketmiyor	75.375	502
Total	1.664	666

EK 3-YENİLEBİLİR SIVI YAĞ KARAR AĞACI

YenilirYagC



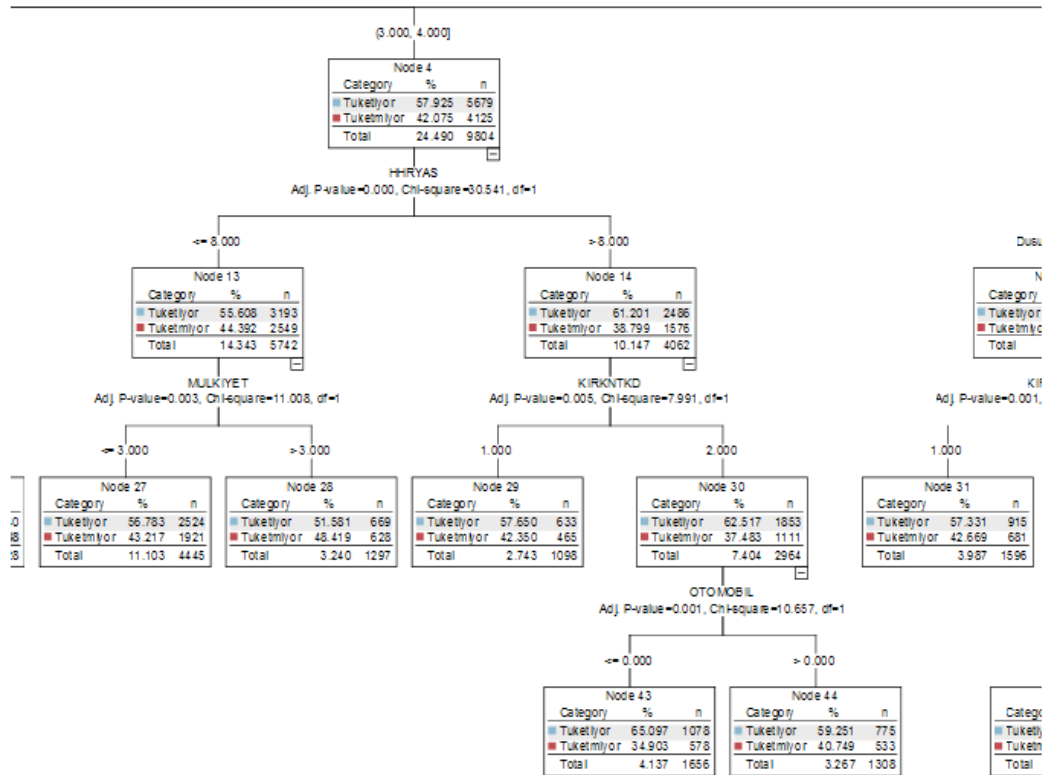


YeniliryagC

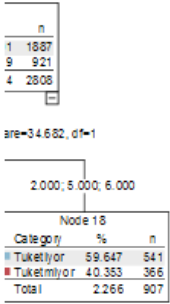
jC

	n
697	21897
303	18136
000	40033

jare=1034.874, df=5



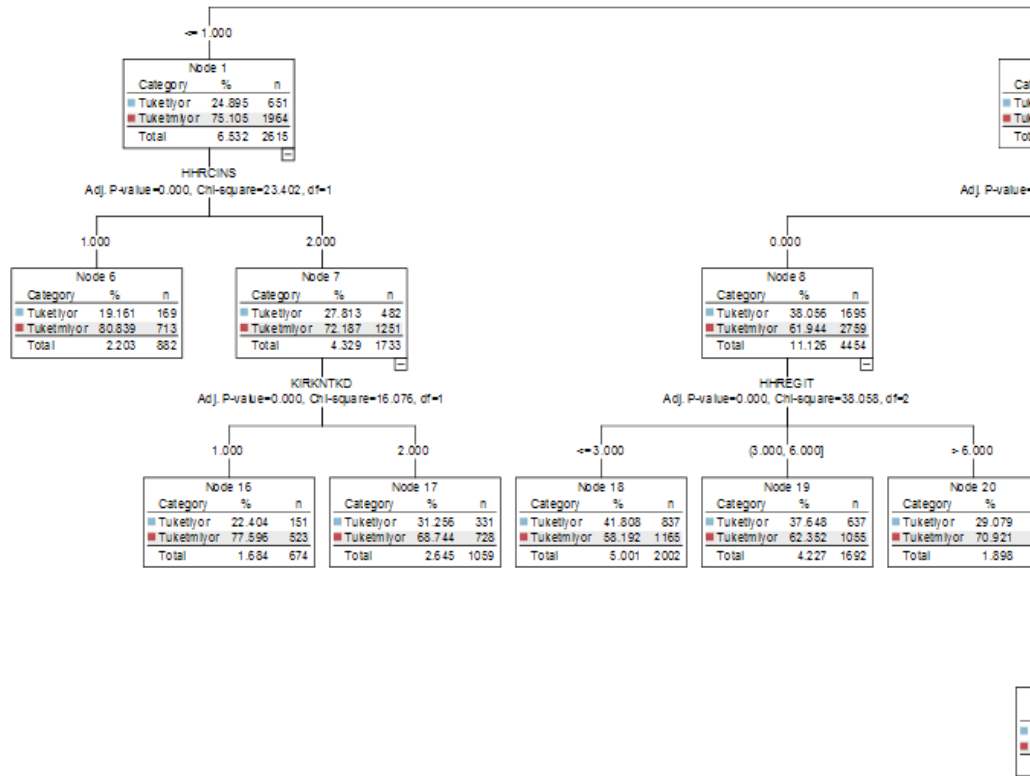
YeniliriYagC



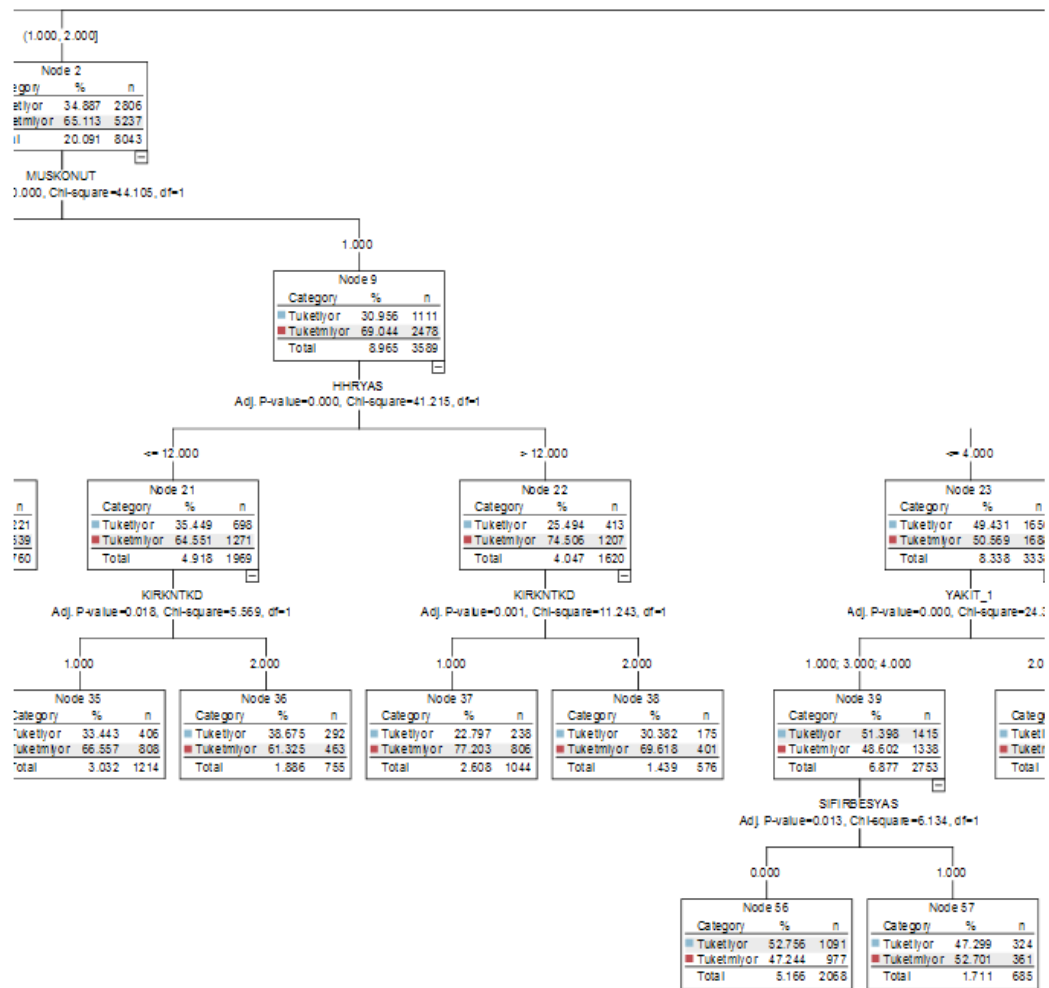
n
890
402
1292

EK 4-MARGARİN VE DİĞER BİTKİSEL YAĞLAR KARAR AĞACI

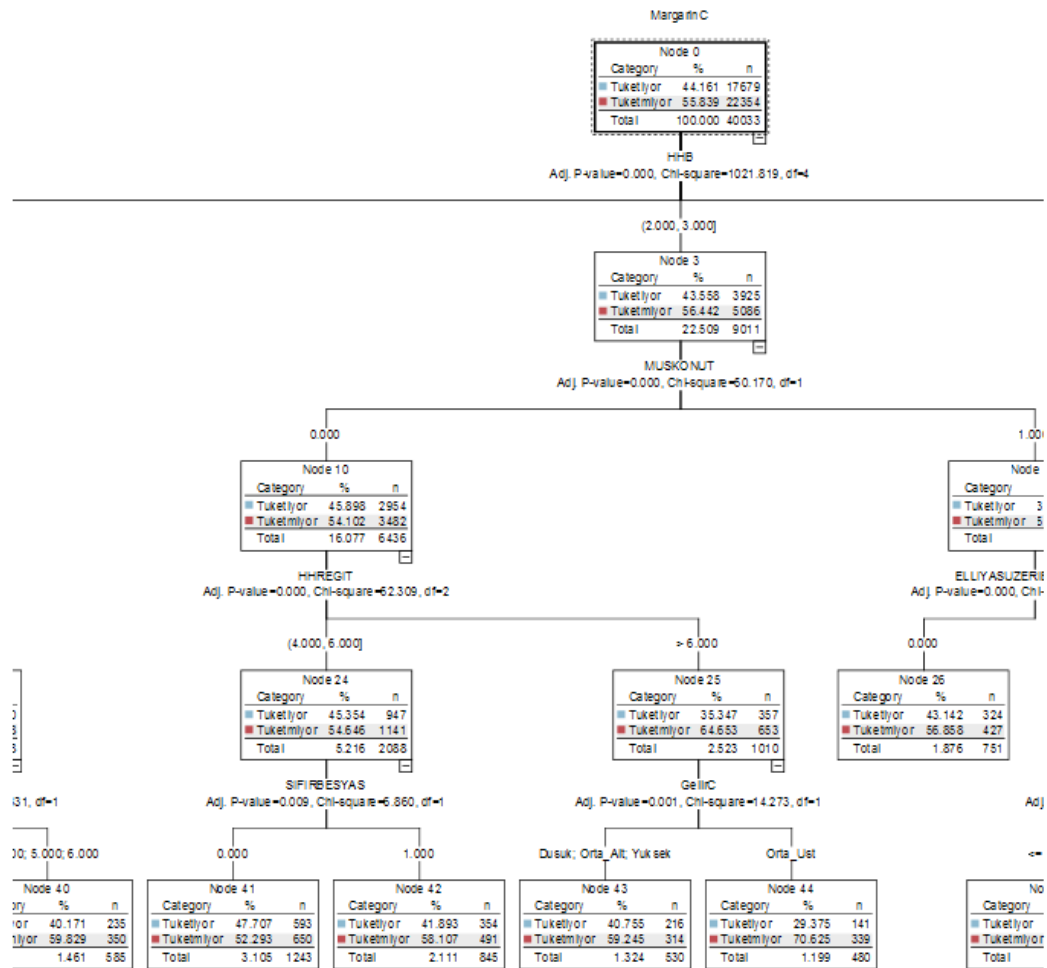
MargarinC



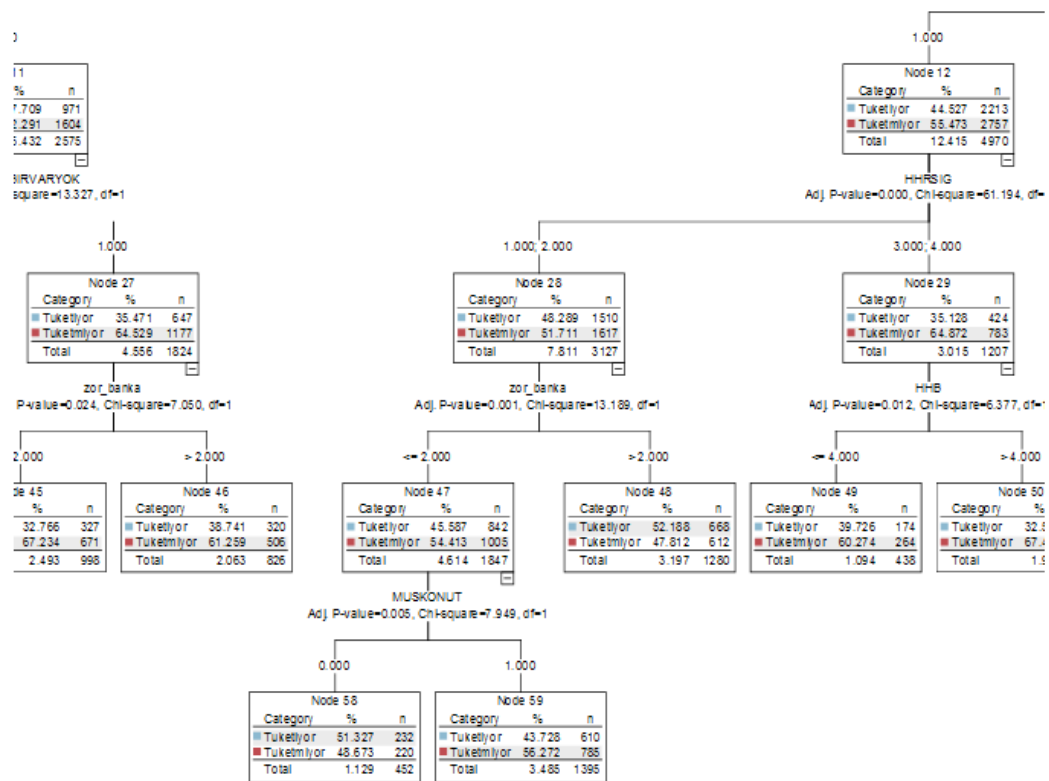
MargarinC



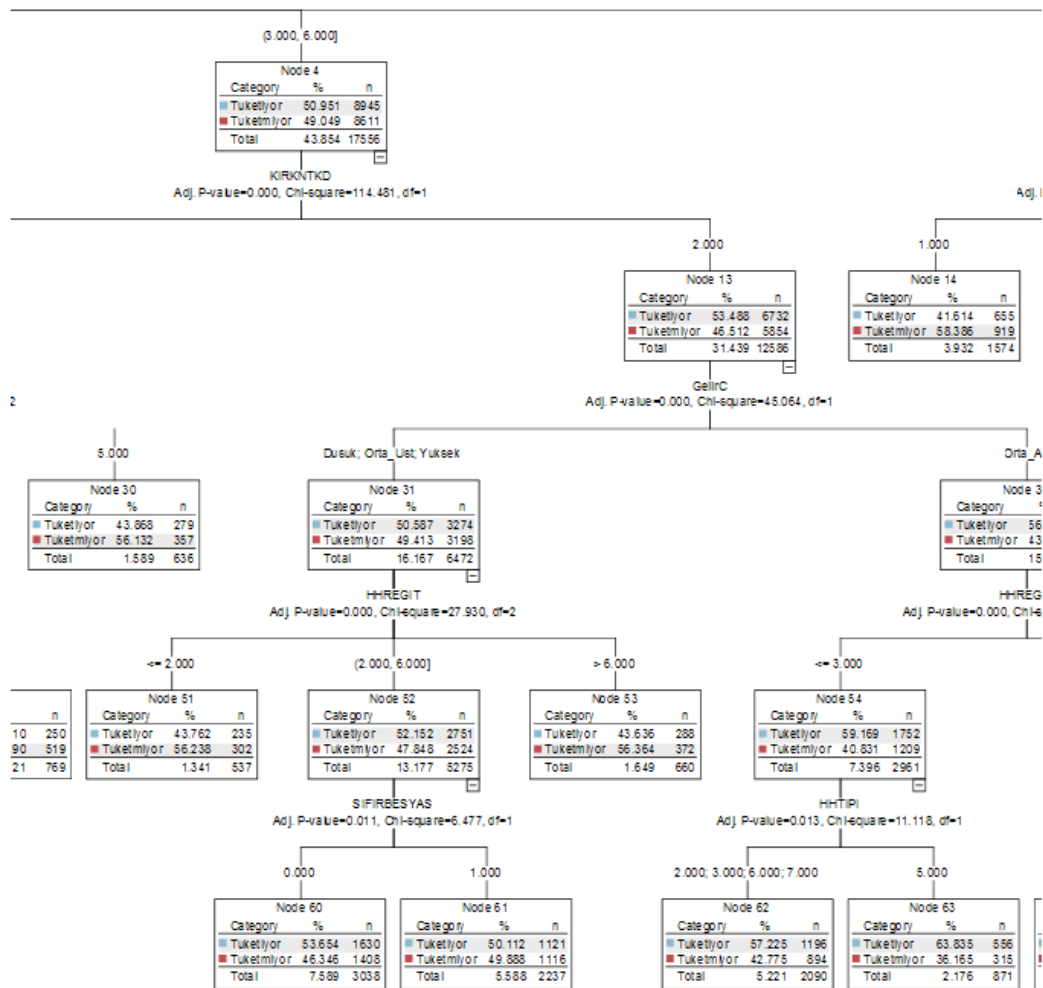
MargarinC



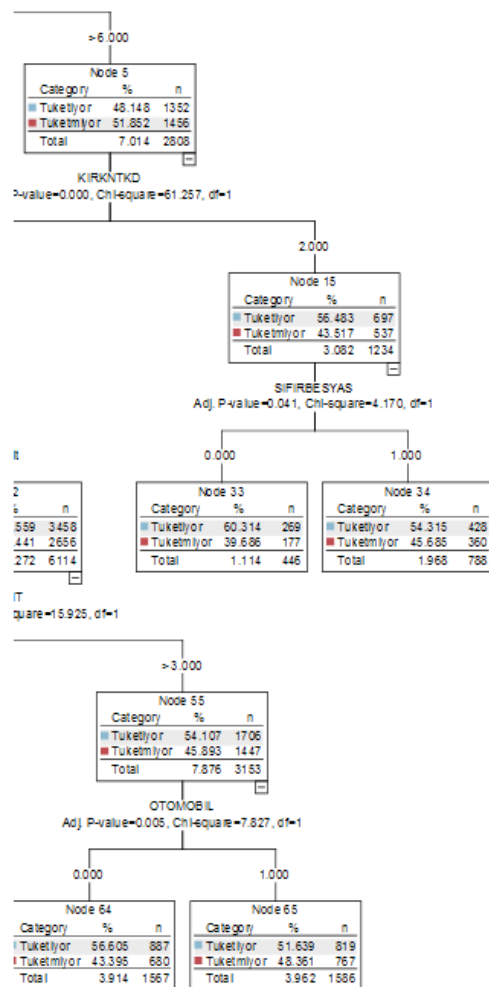
MargarinC



MargarinC

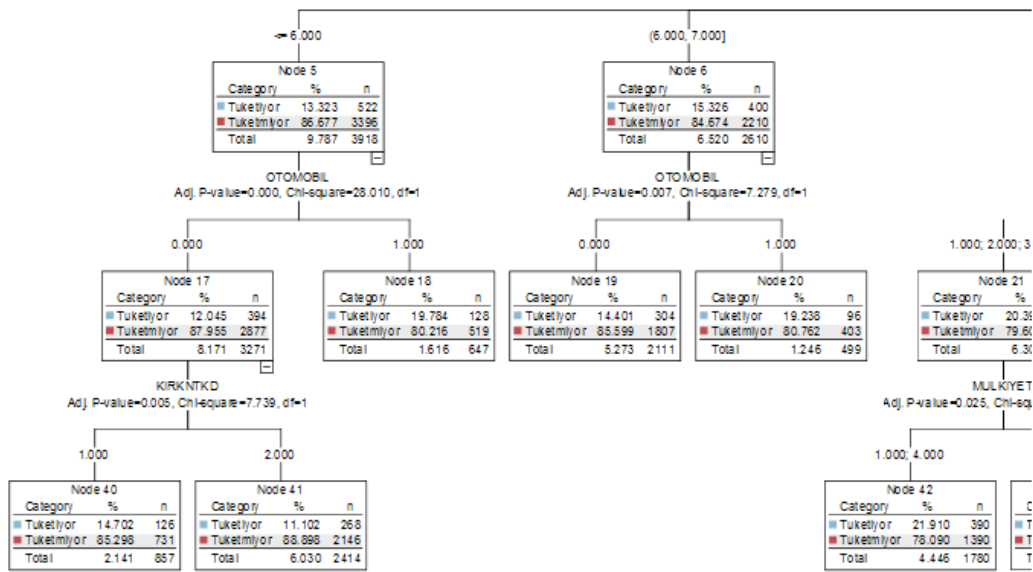


MargarinC

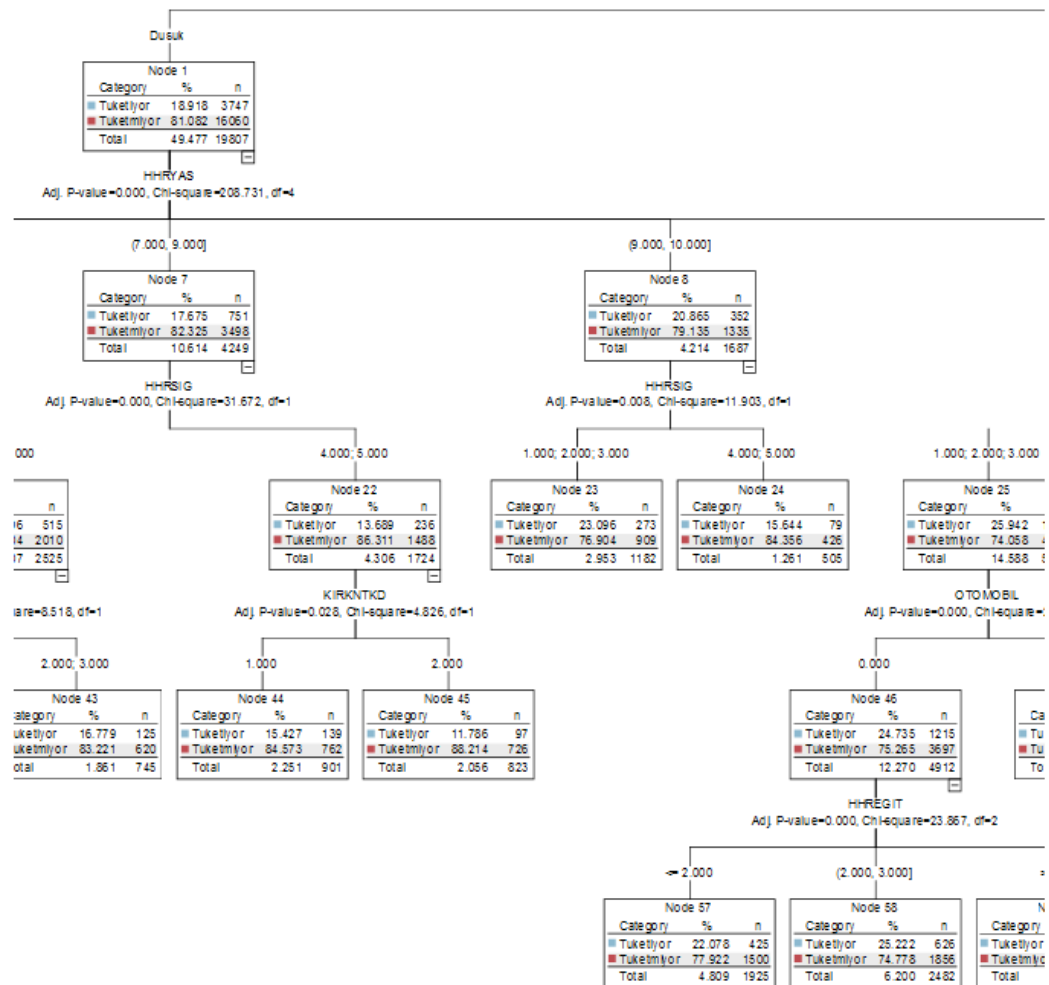


EK 5-TEREYAĞI KARAR AĞACI

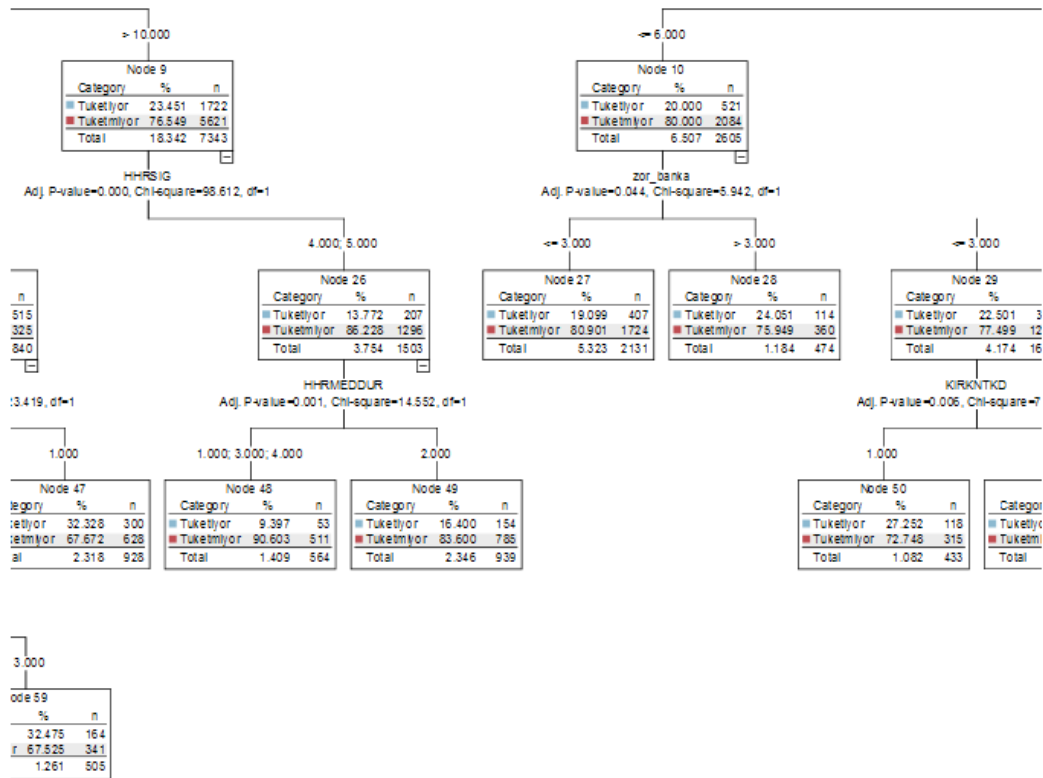
TereyagıC



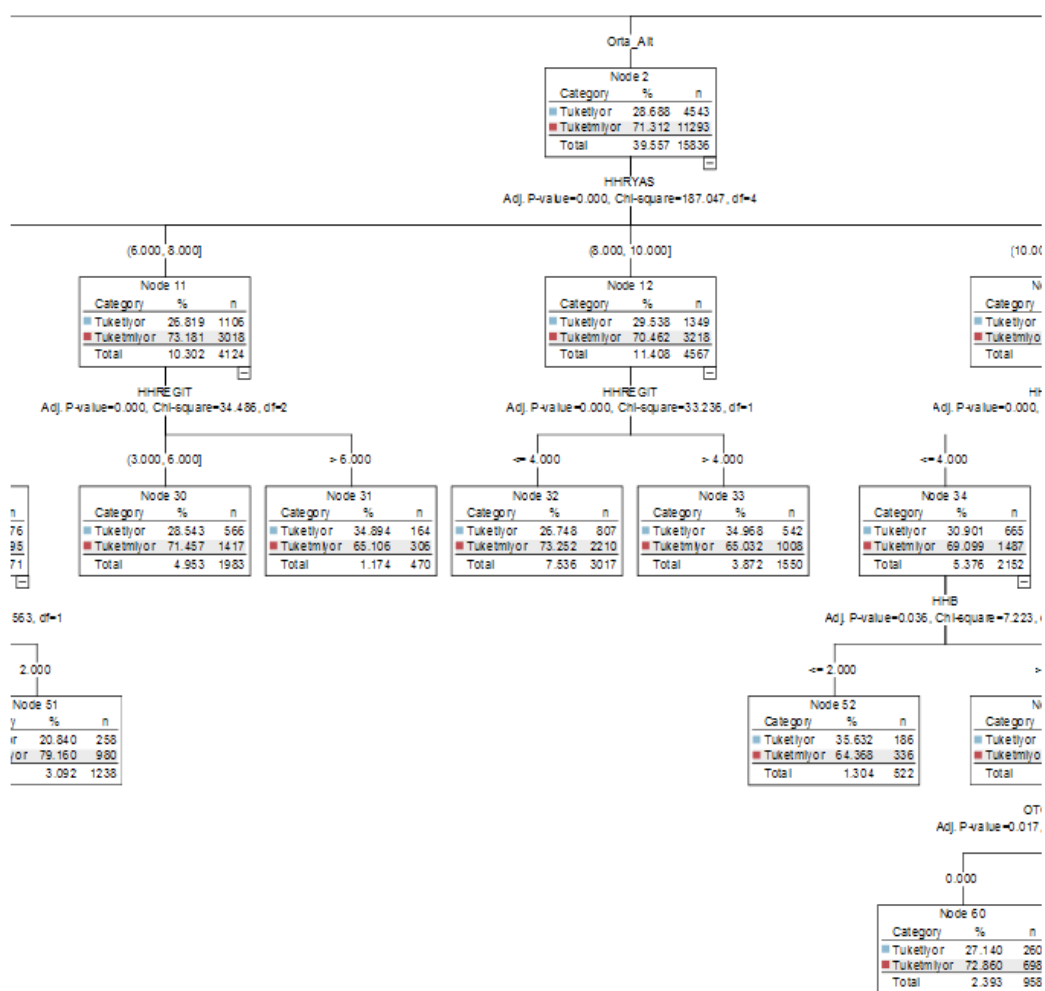
TereyagıC



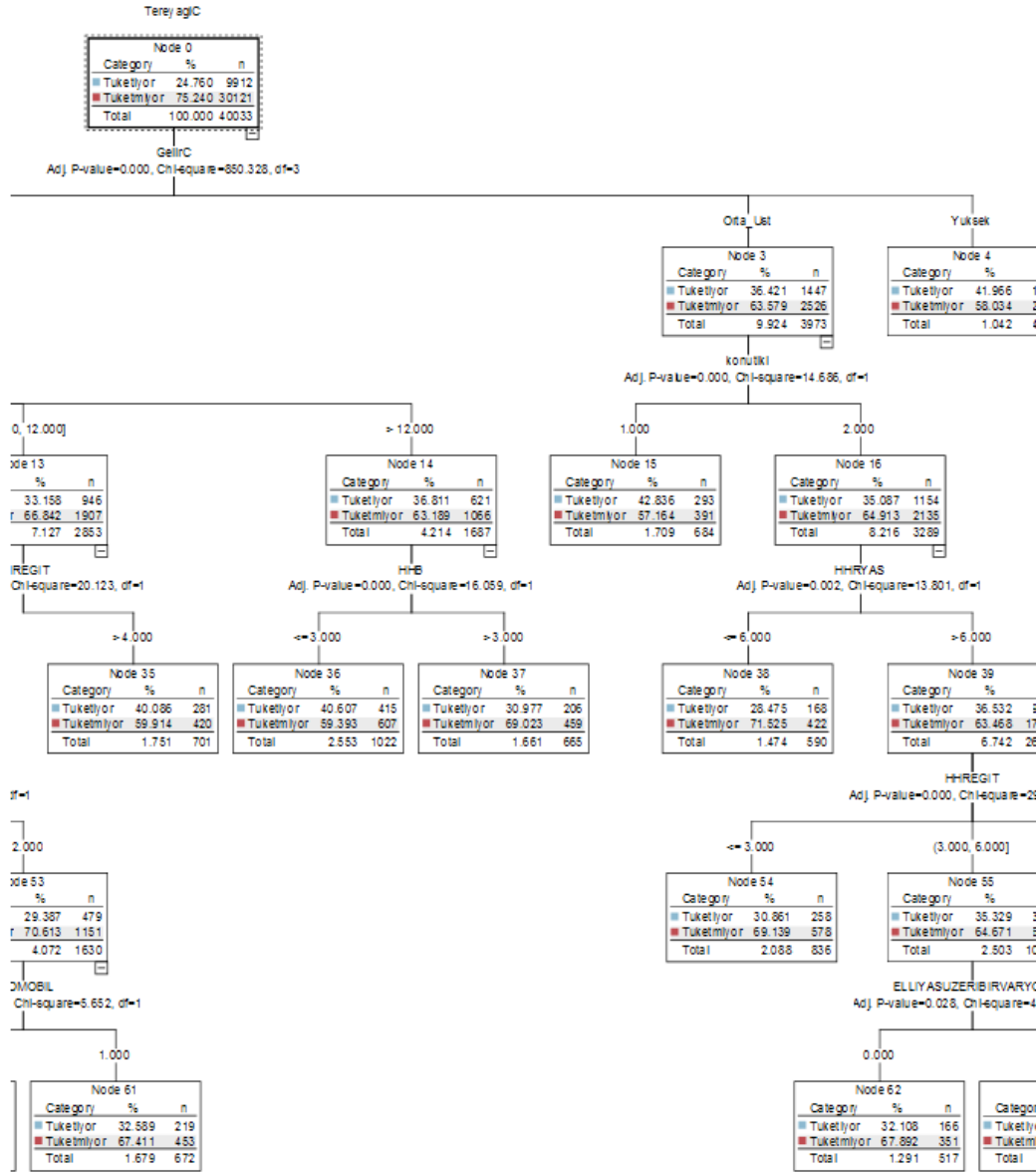
TereyagıC



TereyagiC



TereyagiC



ÖZGEÇMİŞ

Adı ve SOYADI : Uğur ERCAN

Eğitim Durumu

Mezun Olduğu Lise : Çağlayan Lisesi, Antalya, 1999

Lisans Diploması : Mersin Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, 2003

Yüksek Lisans Diploması : Akdeniz Üniversitesi, Fen Bilimler Enstitüsü, Matematik Ana Bilim Dalı, Antalya, 2008

Tez Konusu : Laplace Operatörünün Doğurduğu Gauss-Weierstrass İntegrali ve Bazı Özelliklerinin İncelenmesi

Doktora Diploması : Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Ana Bilim Dalı, Antalya, 2016

Tez Konusu : Veri Madenciliği ile Hanehalkı Yağ Tüketiminin Modellenmesi

Yabancı Dil : İngilizce

İş Denevimi

Çalıştığı Kurumlar :

- Akdeniz Üniversitesi Enformatik Böl. Bşk. Öğretim Görevlisi, 01.07.2004 - Halen
- Akdeniz Üniversitesi Tıp Fakültesi Hastanesi Bilgi İşlem (Staj), Haziran-Temmuz 2002
- Dedeman Otel Antalya, Bilgi İşlem (Staj), Temmuz-Ağustos 2001

Burs ve Ödüller :

- Mühendislik Fakültesi ve Bilgisayar Mühendisliği Bölüm Birincisi (2003)
- TÜBİTAK Yüksek Lisans Bursu, (2005, 2 Yıl)

E-Posta : ugurercan@akdeniz.edu.tr