

T.C.
İSTANBUL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANA BİLİM DALI
SAYISAL YÖNTEMLER BİLİM DALI

DOKTORA TEZİ

**KAMU HASTANELERİNDEKİ TİG (TEŞHİSLE
İLİŞKİLİ GRUPLAR) VERİLERİNİN SAĞLIK
ANALİTİĞİ YAKLAŞIMI İLE İNCELENMESİ**

PAKİZE YİĞİT
2502110338

Danışman: Prof. Dr. EYÜP ÇETİN

İSTANBUL,2015



DOKTORA
TEZ ONAYI

ÖĞRENCİNİN

Adı ve Soyadı : PAKİZE YİĞİT

Numarası : 2502110338

Anabilim/Bilim Dalı : SAYISAL YÖNTEMLER

Danışman : PROF.DR EYÜP ÇETİN

Tez Savunma Tarihi : 29.04.2015

Tez Savunma Saati : 11:00

Tez Başlığı : KAMU HASTANELERİNDEKİ TİG (TEŞHİSLE İLİŞKİLİ GRUPLAR)
VERİLERİNİN SAĞLIK ANALİTİĞİ YAKLAŞIMI İLE İNCELENMESİ

TEZ SAVUNMA SINAVI, Lisansüstü Öğretim Yönetmeliği'nin 50. Maddesi uyarınca yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin KABULÜNE OYBİRLİĞİ / ~~OYÇOKLUĞUYLA~~ karar verilmiştir.

JÜRİ ÜYESİ	İMZA	KANAATİ (KABUL / RED / DÜZELTME)
1- PROF.DR EYÜP ÇETİN		Kabul
2- PROF.DR MEHPARE TİMOR		Kabul
3- DOÇ.DR UMMAN TUĞBA ŞİMŞEK GÜRSOY		Kabul
4-DOÇ.DR SONA MARDİKYAN		Kabul
5- DOÇ.DR EYLEM DENİZ HOWE		Kabul

YEDEK JÜRİ ÜYESİ	İMZA	KANAATİ (KABUL / RED / DÜZELTME)
1- PROF.DR ERGÜN EROĞLU		
2- PROF.DR TUNCAY CAN		

KAMU HASTANELERİNDEKİ TİG (TEŞHİSLE İLİŞKİLİ GRUPLAR) VERİLERİNİN SAĞLIK ANALİTİĞİ YAKLAŞIMI İLE İNCELENMESİ

Pakize Yiğit

ÖZ

Sağlık hizmetlerindeki büyük veri analitiği ve uygulamaları sağlık hizmetlerinin kalitesini arttırmak aynı zamanda maliyetleri azaltmak için kullanılmaktadır. Veri madenciliği teknikleri ile büyük veri tabanlarından yararlı bilgiler elde edilir. Bu çalışmanın amacı, bu veri madenciliği tekniklerini bir yıllık Teşhis İlişkili Gruplar (TİG) verilerine uygulayarak sağlık politikası oluşturan karar vericilere destek olmaktır. Öncelikle kümeleme teknikleri ile oluşan TİG’ler profil bilgisine göre kümelere ayrılmıştır. Ardından, hastanede aşırı yatış süresine (AYS) etki eden değişkenler lojistik regresyon ve karar ağacı uygulayarak tahmin edilmeye çalışılmıştır. Son olarak ise oluşan diabet hasta TİG’leri birliktelik analizi kullanılarak sonuçlar yorumlanmıştır.

Anahtar Kelimeler: Sağlık Analitiği, Veri madenciliği, Teşhis İlişkili Gruplar

ANALYSING DRG (DIAGNOSIS-RELATED GROUP) DATA IN TRAINING HOSPITALS WITH HEALTH ANALYTICAL APPROACH

ABSTRACT

Big data analytics and applications in healthcare are utilized to improve quality of care and also reduce costs. Data Mining is a technique, which enables valuable information in the large databases. The purpose of this study is to conduct data mining techniques to Diagnosis-Related Groups warehouse to support health politicians. Firstly, DRG's are segmented based on their demographic characteristics, by means of clustering analysis. Secondly, It is used to Logistic Regression and Classification and Regression Trees to study variables associated with high length of stay outliers. Finally, diabete DRG's are determined by using Association Rules and, the application results are presented within this order.

Keywords: Health Analytics, Data Mining, Diagnosis-Related Groups.

ÖNSÖZ

Tez çalışması boyunca değerli katkıları ve yönlendirmeleri ile çok yardımcı olan danışman Hocam Prof. Dr. Eyüp Çetin'e ve yine değerli katkılarıyla beni cesaretlendiren ve bilgi birikimi ile aydınlatan Hocam Doç. Dr. Umman Tuğba Şimşek Gürsoy'a, tez izleme komitemde bulunan Doç. Dr. Sona Mardikyan'a çok teşekkür ederim.

Beni TİG konusunda çalışmaya cesaretlendiren ve değerli katkıları ile beni her zaman aydınlatan Prof. Dr. Sabahattin Aydın'a, yine kıymetli tecrübelerinden faydalandığım Hocalarım Prof. Dr. Osman Hayran ve Prof. Dr. Hacer Özgen Narcı 'ya teşekkür ederim.

Katkıları ve manevi destekleri için Hocalarım Prof. Dr. Huriye Çatalca ve Prof. Dr. Hasan Çatalca'ya minnetlerimi sunarım.

Çalışmada kullanılan TİG verilerinin sağlanmasında katkıda bulunan, Sağlık Bakanlığından görev yapan, Sağlık Hizmetleri Genel Müdür Yardımcısı Dr. Hasan Güler'e, Sağlık Politikaları Kurulu Üyesi Dr. Mehmet Demir'e teşekkürlerimi sunarım.

Ayrıca, uygulama bölümünde kullandığım SAS Enterprise Miner 9.4 programını sağlayan SAS Türkiye Firması'na da teşekkürü bir borç bilirim.

Tez çalışmam boyunca sonsuz destekleri ve sabırları ile hep yanımda olan Aileme ve yakın arkadaşlarıma sonsuz teşekkür ederim.

İÇİNDEKİLER

ÖZ	iii
ABSTRACT.....	iv
ÖNSÖZ	v
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ	x
KISALTMA LİSTESİ.....	xii
GİRİŞ	1
1. SAĞLIK ANALİTİĞİ	3
1.1. Veri Madenciliği.....	6
1.1.1. Veri Madenciliği Süreci	9
1.1.2. Veri Tabanındaki Gelişmeler	12
1.2. Veri Madenciliği Teknikleri	16
1.2.1. Kümeleme Analizi	16
1.2.2. Birliktelik Analizi.....	25
1.2.3. Karar Ağaçları	28
1.2.4. Lojistik Regresyon	33
1.2.5. Bayes Sınıflandırıcılar.....	35
1.2.6. Yapay Sinir Ağları	38
1.2.7. Destek Vektör Makineleri	40
2. SAĞLIK HİZMETLERİ FİNANSMANI.....	41
2.1. Sağlık Hizmetlerinin Finansman Yöntemi	45
2.1.1. Özel Finansman Yöntemleri	46
2.1.2. Kamusal Nitelikli Finansman Kaynakları.....	49
2.2. Hizmet Sunucularına Ödeme.....	51
2.3. Sağlık Hizmetlerinde Geri Ödeme Yöntemleri	51
2.3.1. Harcama Kalemli Bütçeler ve Genel (Global) Bütçeler	52

2.3.2.	Kişi Başı Geri Ödeme:	52
2.3.3.	Hizmet Başına Geri Ödeme	53
2.3.4.	Gün Başına Ödeme	54
2.3.5.	Vakaya Dayalı Ödeme	55
2.4.	Türkiye’de Sağlık Hizmetlerinin Finansmanı	56
2.4.1.	Özel Nitelikli Sağlık Finansmanı	56
2.4.2.	Türkiye’de Kamusal Nitelikli Sağlık Finansmanı.....	57
2.5.	Teşhis İlişkili Gruplar ile Klinik Kodlama.....	60
2.5.1.	Sınıflandırılma	61
2.5.2.	Kodlanma	62
2.5.3.	TİG Oluşum Süreci	63
2.5.4.	TİG’e Dayalı Finansmanın Bileşenleri	69
2.6.	Teşhis İlişkili Gruplarla Geri Ödeme	71
2.6.1.	Türkiye’de Teşhisle İlişkili Gruplarla Geri Ödeme	76
3.	UYGULAMA	79
3.1.	Problemin Tanımlanması	79
3.2.	Verilerin Hazırlanma Süreci.....	81
3.3.	Kümeleme Analizi.....	85
3.4.	Lojistik Regresyon ve Karar Ağacı Analizleri	93
3.4.1.	Lojistik Regresyon Modeli.....	95
3.4.2.	Karar Ağacı Modeli.....	96
3.4.3.	Lojistik Regresyon ve Karar Ağacı Modellerinin Karşılaştırılması ..	100
3.5.	Birliktelik Analizi.....	106
4.	SONUÇLAR.....	111
	KAYNAKLAR	115
	ÖZGEÇMİŞ	128

TABLÖLAR LİSTESİ

Tablo 1-1: Kümeleme Analizinde Kullanılan Uzaklık Ölçütleri	18
Tablo 1-2: Bölümleyici Kümeleme Algoritmaları Özellikleri.....	20
Tablo 1-3: Hiyerarşik Kümeleme Analizinin özellikleri	22
Tablo 1-4: Yoğunluk Temelli Algoritmaların Özellikleri.....	23
Tablo 1-5: Izgara Tabanlı Kümeleme Tekniklerinin Özellikleri	24
Tablo 2-1: Sağlık Sunucuları Ödeme Sistemleri Karşılaştırması	56
Tablo 2-2: TİG Alan Listesi	64
Tablo 2-3: 25 Majör Tanı Sınıfı	65
Tablo 2-4: TİG'in Atanması ve Oluşumu	66
Tablo 2-5: TİG için temel veriler	68
Tablo 2-6: TİG Swot Analizi	75
Tablo 2-7: Tipik bir Avustralya Eğitim Hastanesinde DRG Finansmanından Önce ve Sonraki Etkinlik İstatistikleri	76
Tablo 3-1: Hastaların Cinsiyet Dağılımı	81
Tablo 3-2: Hastaların Hastaneden Çıkış Nedenleri.....	82
Tablo 3-3: Hastaların Sosyal Güvence Dağılımları	82
Tablo 3-4: Hastaların Yaş Dağılımları	83
Tablo 3-5: Hastalara Uygulanan Tedavilerin Bağlı Değer Dağılımları	83
Tablo 3-6: Hastaların Hastanede Yatış Günü Dağılımları	84
Tablo 3-7: Hastalardaki Komorbidite Görülme Sıklığı	84
Tablo 3-8: Kohonen Ağ ile Oluşan 12 Kümedeki İstatistikler	87
Tablo 3-9: Kümelerdeki En Çok Tekrar Eden 5 TİG	88
Tablo 3-10: AYS'yi etkileyen faktörlerin lojistik regresyon ile incelenmesi	95
Tablo 3-11: Lojistik Regresyon Sınıflandırma Tabloları.....	96
Tablo 3-12: Karar Ağacı Analizinde Değişkenlerin Önem Tabloları	96
Tablo 3-13: Karar Ağacı İstatistikleri	98
Tablo 3-14: Karar Ağacı Sınıflandırma Durumları.....	99
Tablo 3-15: Eğitim Veri Kümesinin Performansının Kıyaslanması	100
Tablo 3-16: Test Veri Kümesinin Performansının Kıyaslanması	100

Tablo 3-17: Eğitim Kümelerinin Lift Değerleri.....	102
Tablo 3-18: Eğitim Kümesinin Response Değerleri	104
Tablo 3-19: Diyabet Kodları	107
Tablo 3-20: Diyabet Hastaları Birliktelik Analizi Sonuçları	109
Tablo 3-21: GSS Olmayan Diyabet Hastalarının Birliktelik Kuralları	110

ŞEKİLLER LİSTESİ

Şekil 1-1: Veri Madenciliğinin Etkilendiği Disiplinler.....	8
Şekil 1-2: CRISP-DM'e göre Veri Madenciliği Süreci	10
Şekil 1-3: Veri Ambarı Mimarisi	13
Şekil 1-4: Veri Tabanı ve Veri Madenciliği Gelişim Süreci.....	15
Şekil 1-5: Cerrahi Komplikasyon Riski Taşıyan Hastalar için Karar Ağacı	30
Şekil 1-6: İkili bağımlı değişkenin S ve ters S olasılık fonksiyonu grafikleri	34
Şekil 1-7: Yapay Sinir Hücresinin Yapısı	39
Şekil 2-1: Sağlık Hizmetleri Üçgeni	42
Şekil 2-2: Sağlık Hizmetleri Finansmanı İşlevleri ve Toplum İlişkisi.....	44
Şekil 2-3: Sağlık Hizmetlerinin Alternatif Finansman Yöntemleri	45
Şekil 2-4: Türk Sağlık Sistemindeki Temel Mali Akışlar, 2008 (GSS Yasasının Kabulünden Sonra).....	59
Şekil 2-5: TİG Hipertansiyon Alt Grupları	62
Şekil 2-6: TİG Oluşum Süreci	63
Şekil 2-7: TİG Algoritması Oluşumu.....	63
Şekil 2-8: ICD-10 Tanı Kodlarının Yapısı.....	67
Şekil 2-9: TİG Finansman Bileşenleri.....	69
Şekil 2-10: Dünyada TİG kullanımı.....	73
Şekil 2-11: TİG'in Avrupa'da temel bir finansal altyapı olarak devreye girmesine yönelik kronoloji	74
Şekil 2-12: Sağlık Bakanlığı TİG Süreci	77
Şekil 3-1: Kümeleme Analizi SAS Enterprise Miner Uygulaması.....	86
Şekil 3-2: Kohonen Geri Bildirim Haritaları	89
Şekil 3-3: Kohonen Ağları Kümeleme Analizi Sonuçlarına Uygulanan Segment Analizi ile Her bir Kümedeki Önemli Değişkenler.....	90
Şekil 3-4: Lojistik Regresyon ve Karar Ağacı SAS Enterprise Miner Uygulaması ..	94
Şekil 3-5:C&R Karar Ağacı Modelinin Gösterimi	97
Şekil 3-6: Modellerin Lift Grafikleri.....	102
Şekil 3-7: Mdellerin Cumulative Lift Grafikleri.....	103
Şekil 3-8: Modellerin Response Grafikleri	105

Şekil 3-9: Modellerin Cumulative Response Grafikleri.....	105
Şekil 3-10: Eğitim ve Onay Kümelerinin Sınıflandırma Grafikleri.....	106
Şekil 3-11: Birliktelik Analizi SPSS Clementine	107

KISALTMA LİSTESİ

Kısaltma	Açıklama
TİG	Teşhis (Tanı) İlişkili Gruplar
DRG	Diagnosis-Related Groups
ZB	Zettabayt
TB	Terabayt
CRISP–DM	The Cross Industry Standard Process for Data Mining
SEMMA	Sample (Örnekleme), Explore (İncele), Modify(Düzeltil), Model (Modelle) ve Assess (Değerlendir)
ETL	Veri Dönüştürme
WHO	Dünya Sağlık Örgütü
AYS	Aşırı Yatış Süresi
NYS	Normal Yatış Süresi

GİRİŞ

1960'lı yıllardan sonra veri tabanı teknolojisinde meydana gelen gelişmeler sonucunda, 1990'lı yıllarda bu verilerden yararlı bilgi çıkarabilmek için veri tabanlarında bilgi keşfi ve veri madenciliği kavramları ilk defa ortaya çıkmıştır. Veri madenciliği büyük gelişmeler göstererek pazarlama, risk yönetimi, biyoloji, tıp, sağlık, finans, haberleşme gibi bir çok farklı alanda kullanılmaya başlanmıştır.

Sağlık verilerinde geçmişten bu yana veriler toplanmakta idi. Fakat bu verilerin dijital ortamlarda toplanma kolaylığı ve maliyetinin azalması ile ileri analiz teknikleri sağlık alanında da kullanılmış ve böylece sağlık analitiği ortaya çıkmıştır. Veri madenciliği önemli sağlık analitiği tekniklerinden bir tanesidir.

Dünya genelinde 20. yüzyılda halk sağlığı kavramı öne çıkmış ve 1970'lerden sonra Dünya Sağlık Örgütünün de çalışmalarıyla sağlık hizmetlerine erişim önem kazanmıştır. Bununla birlikte ülkeler, yaşlanan nüfus ve kronik hastalıkların artmasıyla sağlık hizmetlerinin maliyetine katlanmakta zorlanmaya başlamışlar ve bu duruma çareler üretme ihtiyacı duymuşlardır. Bu çalışmaların sonucunda Amerika'da 1970'li yılların sonunda ilk defa sağlık harcamalarının geri ödenmesinde kullanılmak üzere yatan hastalar için hastalıkların sınıflandırılması amacıyla "Diagnosis-Related Group (DRG)", (Teşhis İlişkili Gruplar (TİG)) ortaya çıkmıştır. Bu sisteme göre, aynı TİG'e sahip hastaların klinik açıdan benzer hastalık tipleri gösterdikleri ve aynı miktarda kaynak tükettikleri kabul edilmektedir.

Dünyada bir çok ülkede yaygın olarak kullanılan TİG'ler ile sağlık istatistiği, yatırım, planlama, insan kaynaklarının yönetimi, hastane performanslarının değerlendirilmesi, analiz ve verimlilik çalışmaları, klinik kalite, uluslararası karşılaştırmalar gibi daha pek çok alanda çalışmalar yürütülebilmektedir.

TİG başta olmak üzere geri ödeme modellerinin sağlık hizmetlerinde ülkelerin daha kaliteli ve daha az maliyetli sağlık hizmetleri verebilmelerindeki en büyük etken olmasının ortaya çıkması ile bu konu uluslararası bir boyut kazanmış, sağlık analitiği çalışmalarının en önemli alanlarından biri haline gelmiştir.

Ülkemizde, Sağlıkta Dönüşüm Programı çerçevesinde TİG çalışmaları 2005-2009 yıllarında Avustralya'dan örnek alınarak, pilot çalışmalarla başlamıştır. 2011 yılından itibaren ülke genelinde yaygınlaşarak veriler düzenli olarak toplanmaya başlanmış ve önemli bir veri havuzu oluşturulmuştur.

Bu çalışmada, TİG verilerinin toplandığı 2012 yılına ait 15 eğitim araştırma hastanesinin mevcut durumu veri madenciliği teknikleri ile analiz edilmiştir.

İlk bölümde sağlık analitiği ve veri madenciliği kavramları, veri tabanı teknolojileri ve veri madenciliği teknikleri incelenmiştir.

İkinci bölümde sağlık hizmetleri finansmanı konusu açıklanarak, TİG'lerin sağlık finansmanındaki rolü ve TİG'lerin nasıl oluştuğu anlatılmıştır.

Üçüncü bölümde ise, TİG veri ambarından elde edilen veriler, veri madenciliği tekniklerinden kümeleme analizi, lojistik regresyon, karar ağaçları ve birliktelik analizi kullanılarak yorumlanmıştır.

1. SAĞLIK ANALİTİĞİ

Sağlık hizmetleri tarih boyunca hastanın bakımı, yönetsel gereklilikler gibi nedenlerle sürekli olarak kayıt tutmuştur. Bu veriler önceden kağıt üzerinde toplanırken artık dijital ortamda saklanmaktadır¹. Son yıllarda ise hem sağlık bilgi teknolojileri ulaşılabilirliğinin kolaylığı hem de onların sağlık hizmetleri çalışanları tarafından kullanılması sayısında üstel bir artış görülmektedir².

Sağlık hizmetlerinde elektronik kayıt tutma ve bilgi teknolojilerinin kullanımı her yönden destekleniyor: devletlerin yasalar ve yönetmeliklerle zorunlu hale getirmesi (örneğin elektronik sağlık kayıtlarının benimsenmesi ve anlamlı kullanımı), çeşitli uygulamaların çoğalması ve ulaşılabilir olması, yapılandırılmış ve yapılandırılmamış sağlık kayıtlarının saklanması ve erişilmesinin maliyetinin hızla azalması, verileri ve sunucuları birbirine bağlamanın benimsenmesi ve standartlaşmasıdır.

Sağlık hizmetlerindeki “büyük veri”, geleneksel yazılım/donanımlarla bilinen veri yönetimi metotları ve araçları ile yönetilmesi çok zor olan, büyük ve karmaşık sağlık elektronik veri setleridir. Bu veriler, bilgisayarlı hekim veri bilgi sistemi ve klinik bilgi sistemindeki klinik bilgiler (hekimin yazdığı notlar, ilaçlar, tıbbi görüntüler, laboratuvar sonuçları, sigorta bilgileri...vb. ve diğer yönetsel veriler), hastanın elektronik bilgi sistemindeki kaydı, sensör verileri, twitter tweet’leri, bloglar, facebook gibi platformlardaki durum güncellemeleri web siteleri gibi sosyal medya gönderileri, acil bakım hizmetleri bilgisi gibi direk hasta odaklı olmayan bilgiler, haber akışı ve medikal yayınlardaki makalelerdir.

Toplanan sağlık verilerinden anlamlı bilgiler çıkarabilmek için ortaya çıkan sağlık analitiği, hasta seviyesinde toplanan verinin gelişmiş istatistiksel tekniklerle analizine

1 Raghupathi Wullianallur, Viju Raghupathi. "Big data analytics in healthcare: promise and potential." **Health Information Science and Systems** 2.1 (2014) 3, 2014

2 Raghupathi, Wullianallur, and V. Raghupathi. "An overview of health analytics." **J Health Med Informat** 4.132 (2013): 2, 2013

dayanan ve hızla gelişen bir disiplindir. Sağlık analitiği, sağlık hizmetlerini geliştirmek ve daha iyi bir risk yönetimi sağlamak için kullanılan gelişmiş teknikler bütünüdür³.

İstatistiksel modeller, veri madenciliği, ve klinik karar destek kullanımı gibi sağlık analitiği yöntemleri yıllardır var olmasına rağmen, son zamanlarda daha gelişmiş ve sofistike bilgi teknolojileri ve araçları entegre bir şekilde karar vermeyi desteklemek için kolayca bulunabilir hale gelmiştir.

Sağlık analitiği “analiz” i, veri tabanı ve araçları teknolojileri, elektronik sağlık kayıtları sistemleri, veri ambarı, web uygulamaları, klinik karar destek sistemleri, sağlık verilerinin kendi içinde sorunsuz ve entegre işlenmesi gibi bir çok boyutta birleştirir. Bu entegre uygulamalar doktorlara, hizmet verenlere, teşhis koyanlara, tedavi edenlere, sağlık yöneticilerine yardım eder.

Zorunlu gereksinimler nedeniyle ortaya çıkan sağlık analitiği, sağlık hizmetlerinin kalitesindeki potansiyel kaliteyi arttırmak ve bunu yaparken de maliyeti azaltmak en temel amacıdır.

Amerika Sağlık sistemlerinin raporuna göre 2011 yılında toplanan sağlık hizmetleri verisi 150 Eksabayt (EB) (1 EB=1024 Terabayt (TB))’dir. Bu büyüme hızı ile Amerika’daki sağlık hizmetleri büyük verisi yakında Zetabayt (1021 gigabayt) ile ve yine kısa bir zaman sonra da Yotabayt (1024 zetabayt) ile ifade edilecektir.

Sağlık analitikçileri bu büyük verilerdeki ilişkileri bularak, veri içindeki örüntüleri ve trendleri anlayarak, bakım hizmetleri kalitesini arttırmayı, hayat kurtarmayı ve harcamaları azaltmayı iyileştirme potansiyeline sahiptirler. Büyük veriler analiz ve sentez edildiğinde , bahsedilen ilişkiler, örüntüler ve trendler ile, sağlık hizmeti verenler ve diğer sağlık hizmetleri sunan sistemdeki paydaşlar, beklendiği gibi daha az maliyetle yüksek kalitede bakım ve toplamda daha iyi sonuçlar elde ederek, kapsamlı ve derinlemesine tanı ve tedaviler oluşturlar.

³ Eyüp Çetin,”**Medikal Karar Verme Yöntemleri: Sağlık Yönetimi ve Tıpta Medikal Uygulamalar**”, Beta Yayıncılık, 2013, (Hamparsum Bozdoğan teşekkür yazısı).

Ayrıca sađlık hizmetleri geri ödeme modelleri deđiřmekte; anlamlı kullanım ve performansa dayalı ödeme bugünün sađlık hizmetlerinde kritik yeni faktörler olarak ortaya çıkmaktadır. Kar, ilk motivasyon olmamasına (olmamalıdır da) rağmen, büyük veri ile etkili bir şekilde başa çıkabilmek için, sađlık kuruluşları için gerekli araçlara altyapıya ve tekniklere sahip olmak son derece önemlidir. Aksi takdirde, milyonlarca dolar gelir ve kar kaybetme riski göze alınmalıdır⁴.

Sađlık Hizmetlerindeki Avantajları:

Potansiyel yararları içinde hastalıkların erken teşhisiyle, daha kolay ve etkili bir şekilde tedavi edilmesi, bireysel sađlığın ve toplum sađlığının yönetilmesi ve sađlık hizmetlerindeki dolandırıcılığın çabuk ve etkili bir şekilde tespit edilebilmesi sayılabilir.

Bu muazzam miktardaki veri ile belirli gelişmeler ve çıktılar kestirilebilir ve/veya tahmin edilebilir: hastanın yatış süresi, elektif cerrahi müdahale gerektiren hastalar, ameliyattan yarar görmeyecek hastalar, komplikasyonlar, tıbbi komplikasyonlar ve kan zehirlenmesi için riskli hastalar, hastanede kazanılmış hastalıklar, hastalıkların ilerlemesi, hastalığın ilerlemesine neden olan faktörler, olası komplikasyonlar...vb. Amerikan Sađlık Hizmetleri Şirketi Mc Kinsey'nin tahminlerine göre Amerika'da sađlık analitiđi her yıl 300 milyon dolardan daha fazla tasarruf sađlamakta ve bu tasarrufların üçte ikisi ulusal sađlık harcamalarının yaklaşık % 8'ini oluşturmaktadır.

Yine McKinsey'in arařtırmalarına göre sađlık analitikçileri üç ana başlıkta, israf ve verimsizliđin azaltılmasına katkıda bulunuyor ve bunlar: klinik işlemler, arařtırma ve geliştirme ve halk sađlığıdır.

Sađlık analitiđinin kavramsal çerçevesi "büyük veri" ile çalışan "veri madenciliđi" ya da Veri Tabanlarında Bilgi Keşfi ile aynıdır. Bu çalışmada, veri madenciliđi kullanılacağı için veri madenciliđi ayrıntılı olarak ele alınmıştır.

⁴ Raghupathi, 2014, a.g.e.s.1-3.

1.1. Veri Madenciliği

Son yıllarda bilgi teknolojilerindeki gelişmeler sonucunda verilerin saklanması kolaylaşmış ve ucuzlamıştır. Toplanan bu verilerden yararlı bilgi üretme ihtiyacı doğduğunda geleneksel yöntemlerin bu ihtiyaca karşılık vermedikleri görülmüştür. İşte bu büyük verilerden yararlı bilgi elde edilmesi amacıyla veri madenciliği ve veri tabanlarında bilgi keşfi kavramları ortaya çıkmıştır. 1980’li yıllarda ortaya çıkan veri tabanlarında bilgi keşfi kavramı, veri madenciliğinin de içinde bulunduğu veri temizleme, birleştirme, dönüştürme, sunum gibi süreçlerin tamamına verilen isimdir. Fakat son yıllarda, bu süreçte bulunan en önemli aşamanın veri madenciliği olmasından dolayı⁵ veri tabanlarında bilgi keşfi yerine veri madenciliği adı ile anılmaktadır.

Teknoloji hiç durmadan gelişmekte ve bilişim uzmanları oluşan verileri yönetmekte ve bilgiye dönüştürmekte zorlanmaktadır. Bir uluslararası pazarlama firması olan IDC’nin 2013 yılında yayınladığı rapora göre 2012 yılında dünya genelinde toplanan ve çoğaltılmış veri boyutunun 2.8 zettabayt (ZB) olduğu tahmin edilmektedir (1 ZB yaklaşık 1 milyar terabayt (TB)). Rapora göre 2020 yılında, toplanan verinin 40 ZB olacağı tahmin edilmekte ve bu rakamın 2010 yılında toplanan verinin 50 katı olacağı düşünülmektedir⁶.

Veri madenciliği için farklı tanımlar yapılmaktadır. Bazıları şöyledir:

Veri madenciliği, geniş veri tabanlarından bilgi çıkarımını hedeflemek için makine öğrenimi, örüntü tanıma, istatistik, veri tabanı ve görselleştirme tekniklerini bir araya getiren disiplinler arası bir alandır⁷.

⁵ Haldun Akpınar, “Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği”. **İ.Ü. İşletme Fakültesi Dergisi**, Cilt: 29, S:1, s. 1-22, 2000, s. 2.

⁶ <http://searchdatamanagement.techtarget.com/feature/Big-data-growth-increases-data-integration-degree-of-difficulty> (Mart 2015).

⁷ Peter Cabena, Pablo Hadjinian, Rolf Stadler, Jaap Verhees ve Alessandro Zanasi, “**Discovering Data Mining: From Concept to Implementation**” (1 edition), Prentice Hall PTR, NJ Upper Saddle River USA 1998, s.12.

Bilgi keşfi, veri içerisinde gizli kalmış, önceden bilinemeyen ve potansiyel olarak kullanışlı olan anlamlı bilginin çıkarılmasıdır⁸.

Veri madenciliği, içerisinde var olan anlamlı örüntü ve kuralları ortaya çıkarmak amacıyla, büyük miktarlardaki verinin otomatik ve yarı otomatik araçlar yardımıyla incelenmesi ve analiz edilmesi sürecidir⁹.

Veri madenciliği, büyük ölçekli veriler arasından bilgiye ulaşma, bilgiyi madenleme işidir. Veri madenciliği, veri tabanı, veri ambarı ve bunun gibi yerlerde depolanmış olan büyük miktarlardaki verilerden ilginç bilginin keşfedilmesi sürecidir¹⁰.

Veri madenciliği, kuralları ve ilişkileri keşfederek, önceden bilinemeyen açık ve yararlı sonuçlar elde etmek amacıyla, büyük miktardaki verinin seçilmesi, incelenmesi ve modellenmesi sürecidir¹¹.

Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanılarak aranmasıdır¹².

Veri madenciliğine ilişkin çözümlerin kökleri klasik veri analizinden gelmektedir. Veri madenciliği birçok disiplinden etkilenecek geliştirilmiştir ve bunlardan en önemli ikisi istatistik ve makine öğrenimidir. İstatistiğin kökenleri matematiğe dayanmaktadır, bunun için veri madenciliğinde matematiksel bir dayanağın olması, yani bir şeyin uygulanmasından önce teorik temelinde hassas olarak kurgulanması isteğine bir vurgu vardır. Buna tezat olarak, makine öğrenimi köklerini büyük oranda bilgisayar uygulamalarından almaktadır. Bu da, uygulamacı bir yönelime, yani

⁸ William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus, "Knowledge Discovery in Databases : An Overview", **AI Magazine**, Vol. 13, No. 3, 57-70, 1992, s.58.

⁹ Michael Berry, Gordon Linoff, "**Mastering Data Mining : The Art and Science of CRM**", John Wiley & Sons, 2000, s.7.

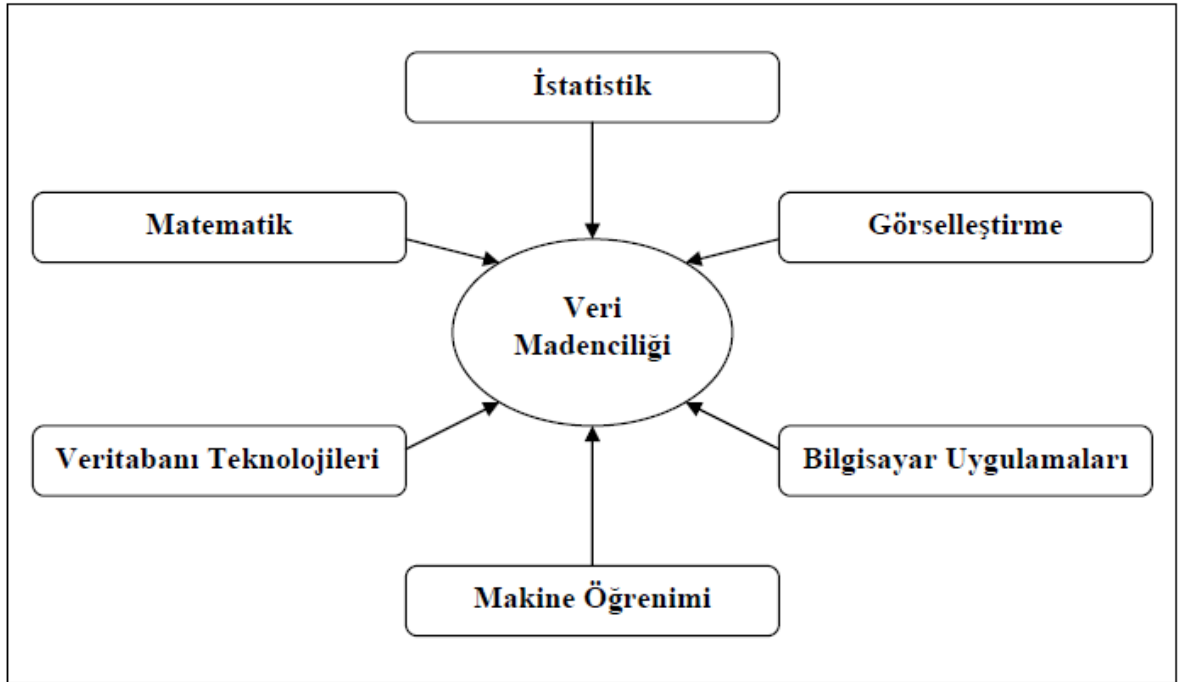
¹⁰ Jiawei Han ve Micheline Kanber, "**Data Mining Concepts and Techniques**", ABD, Morgan Kaufman Publishers, 2006, s.5-7.

¹¹ Paolo Giudici, **Applied Data Mining: Statistical Methods for Business and Industry**, John Wiley & Sons, 2003, s.2.

¹² Ethem Alpaydın, "ZEKİ VERİ MADENCİLİĞİ: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri", **Bilişim 2000 Eğitim Semineri**, (Çevrimiçi) www.cmpe.boun.edu.tr/~ethem/files/papers/veri-maden_2k-notlar.doc, Mart 2015.

etkinliğinin biçimsel bir kanıtına gerek olmaksızın, ne kadar iyi performans gösterdiğini görmek amacıyla bir şeyi test etme isteğine neden olmaktadır¹³.

Veri madenciliğinin multidisipliner yapısından dolayı, bir çok disiplinden etkilendiği görülmektedir: veri tabanı, istatistik, makine öğrenimi, veri tabanı teknolojileri, veri görselleştirme, bilgisayar teknolojisi...vb. Bunların içinden ilk üçü, veri madenciliğine en çok katkı yapan disiplinlerdir: veri tabanı, makine öğrenimi ve istatistik. Veri madenciliğinde istatistik geçerliliği, veri tabanı teknolojileri verimliliği ve yapay öğrenme ise faydalılığı ön plana çıkarmaktadır¹⁴.



Şekil 1-1: Veri Madenciliğinin Etkilendiği Disiplinler

¹³ Mehmed Kantardzic, “**Data Mining: Concepts, Models, Methods, and Algorithms**”, IEEE Press, Hoes Lane, Piscataway, NJ, ABD, 2003, s.4.

¹⁴ Zhi-Hua Zhou, “Three Perspectives of Data Mining”, **Artificial Intelligence**, No.143, 139-146, 2003, s.139.

1.1.1. Veri Madenciliği Süreci

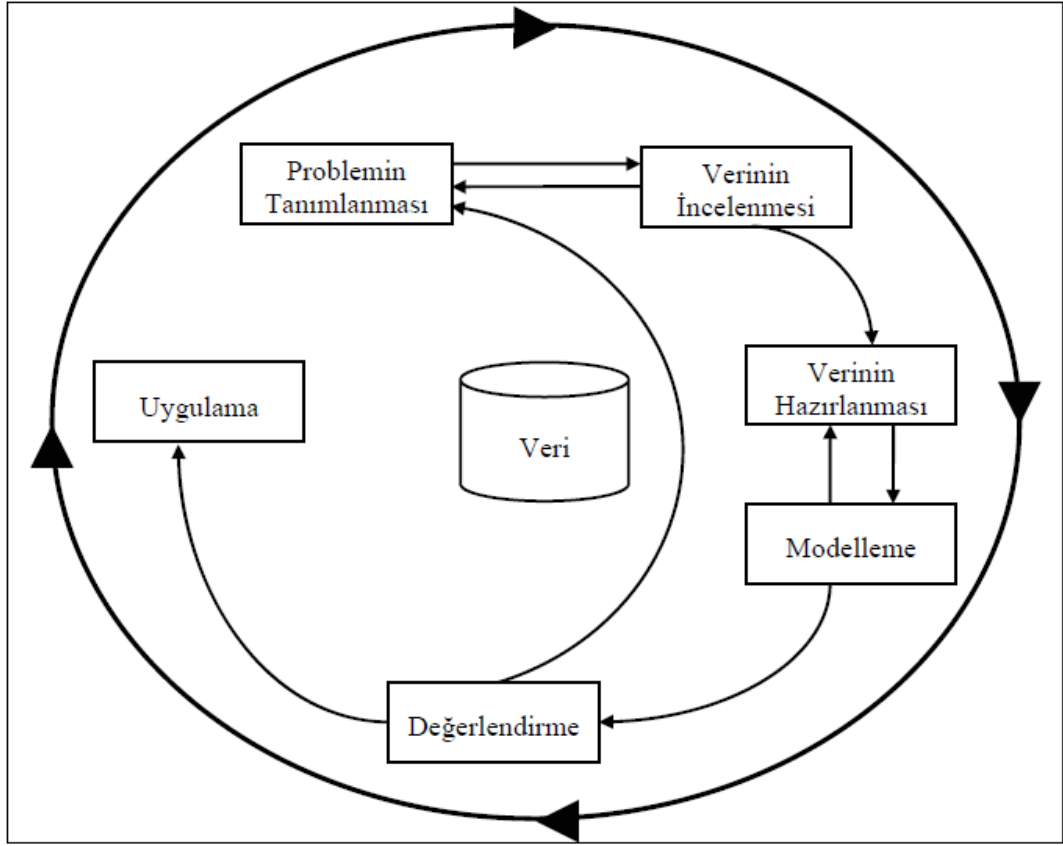
Literatürde veri madenciliği süreci veri tabanlarında bilgi keşfi süreci olarak da adlandırılmaktadır. Han ve Kamber'in bilgi keşfi süreci şu aşamalardan oluşmaktadır¹⁵:

- Verileri temizleme
- Verileri birleştirme
- Verilerin seçimi
- Verileri dönüştürme
- Veri Madenciliği
- Örüntü Değerlendirme
- Bilginin Sunumu

CRISP–DM (The Cross Industry Standard Process for Data Mining) konsorsiyumu, veri madenciliği süreçlerini iki yüzden fazla araştırmacıyı bir araya getirerek 2000 yılında modellemiştir. CRISP–DM modeli, veri madenciliği sürecini altı temel aşamadan oluşan bir yaşam çemberi olarak tanımlanmıştır. Model, her bir aşamanın tamamlanması ile yeni bir aşamaya geçilmesini ya da gerektiğinde önceki aşamalara geri dönülmesini önermektedir¹⁶.

¹⁵ Han vd.a.g.e.s.7.

¹⁶ IBM SPSS Modeler CRISP-DM Guide, 2011, s.1. (Erişim: Şubat 2015)
ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/CRISP_DM.pdf



Şekil 1-2: CRISP-DM'e göre Veri Madenciliği Süreci

CRISP–DM veri madenciliği şu şekilde özetlenebilir:

- **Problemin Tanımlanması:** Veri madenciliği sürecinin en önemli aşamasıdır. Veri madenciliği çalışmalarında başarılı olmanın ilk şartı, uygulamanın hangi amaç için yapılacağını açık bir şekilde tanımlanmasıdır¹⁷.
- **Verinin İncelenmesi:** Verinin incelenmesi verinin toplanması, gerekli araçlara yüklenmesi, çeşitli istatistiksel analizler yapılarak tanımlanması, grafiklerinin çizilmesi, amaca yönelik hipotezlerin yazılması, verinin kalitesinin (eksik, yanlış, uç değer içeren veriler...vb.) ortaya çıkması gibi durumları içeren aşamadır.
- **Verinin Hazırlanması:** Veriyi hazırlamak veri madenciliğinin en önemli ve en çok zaman alan aşamalarından biridir ve değişken seçimi, veri birleştirme

¹⁷ Umman Tuğba Şimşek Gürsoy,” **Veri Madenciliği ve Bilgi Keşfi**”, Pegem Akademi yayınları, Ankara, 2009, s.31.

temizleme ve dönüştürme işlemleri gibi faaliyetler bu aşamada gerçekleştirilmektedir.

- **Modelleme:** Bu aşamada hazırlanan veri ile modelleme teknikleri gerçekleştirilmekte, sonuçlar yorumlanmaktadır. Bu süreç tekrarlandıkça sonuçlar daha güvenilir hale gelmektedir. Aynı problem için bir çok veri madenciliği tekniği olması dolayısıyla, verinin hazırlanması ve incelenmesi aşamalarına geri dönülmesi gerekebilmektedir.
- **Değerlendirme:** Bu aşamada kurulan modellerin çalışmanın hedefini karşılamasının sınanması, başarı kriterlerine göre değerlendirilmesi yapılır. Sonuç olarak bu aşamada modelleme aşamasına geri dönülmesi ya da hangi modelin en iyi olduğuna karar verilmesi olmak üzere iki ayrı karar çıkabilir.
- **Uygulama:** Kurulan ve geçerlilik ve güvenilirliği kabul edilen modelin probleme uygulanmasıdır.

1.1.1.1.SAS SEMMA Süreci

SEMMA, SAS Enterprise Miner'ın veri madenciliği sürecinde kullandığı metodolojidir. Sample (Örnekle), Explore (İncele), Modify (Düzeltil), Model (Modelle) ve Assess (Değerlendir) kelimelerinin kısaltmasıdır¹⁸.

Bu aşamalar şöyle özetlenebilir:

- **Sample:** Veri madenciliği çalışması tüm veriye uygulanmak yerine, tüm veriyi temsil eden az birime uygulanır.
- **Explore:** Bu aşamada verinin tanımsal istatistikleri ve grafikleri çizilerek veri incelenir. Bu aşamada temel bileşenler analizi gibi analizler de uygulanabilir.
- **Modify:** Veri analize hazır hale getirilir. Veri, uç değerler, kayıp değerlerden arındırılır, dönüşüm uygulanır ve kümeleme analizi de uygulanabilir.

¹⁸ Data Mining Using SAS® Enterprise Miner™ Case Study Approach Third Edition, Amerika, 2013, s.5-6. <http://support.sas.com/documentation/cdl/en/emcs/66392/PDF/default/emcs.pdf> (Ocak 2015)

- **Model:** Veri madenciliği modelinin kurulduğu aşamadır. Yapay sinir ağları, karar ağaçları, lojistik regresyon, destek vektör makineleri...vb. uygulanır.
- **Asses:** Uygulanan modellerden en iyi olanının seçildiği aşamadır. Modellerin grafikleri ve istatistiklerini birlikte vererek karşılaştırma olanağı sağlar.

1.1.2. Veri Tabanındaki Gelişmeler

Veri tabanı ve dolayısıyla veri madenciliğinde sıklıkla kullanılan bazı terimlerin açıklamaları şöyledir:

Veri Dönüştürme (ETL): Veri ambarı üzerine ETL (Extract-Transform-Load (Çekme-Dönüştürme-Yüklem) süreci ile heterojen kaynaklardan alınan veriler üzerinde veri madenciliği algoritmalarının uygulanması sağlanarak, normal sorgular ile elde edilemeyen bilgilerin su yüzüne çıkması sağlanmaktadır¹⁹.

Sürecin ilk aşaması olan “extract” da veri kaynaktan çekilir. Transform (dönüştürme) aşamasında ise istenen kurallar extract edilen veriye uygulanır. Bu sayede hedeflenen veri kaynaktan elde edilmiş olur. ETL sürecinin son ayağı olan load (yüklem) aşamasında ise veri veri ambarına yüklenir.

Veri Ambarı: Temel olarak veri madenciliği çalışmaları için veri ve veri tabanı gerekmektedir; ancak işletmelerde kullanılan işlemsel veri tabanları (transactional database) doğrudan veri madenciliği amacıyla kullanılabilmesi için uygun hale getirilmelidir. İşte, belirli bir döneme ait, yapılacak çalışmaya göre konu odaklı olarak düzenlenmiş, birleştirilmiş ve sabitlenmiş işletmelere ait veri tabanlarına veri ambarları denilir²⁰.

Veri ambarları amaca yöneliktir. Örneğin bir sağlık veri tabanında hastanın demografik, hastalık geçmişi ve ilaç... vb. farklı başlıklarda saklanmaktadır. Veri ambarları bütünleşik yapıdadır, yani bütün bu farklı başlıktaki veriler birbirine

¹⁹ Selda Düzgünoğlu, Adnan Yazıcı, Ünal Yarımağan, “Tıp Bilişiminde Veri Ambarı ve Veri Madenciliği Uygulaması”, **3. Ulusal Tıp Bilişimi Kongresi**, s.110-115, 2006, s.111.

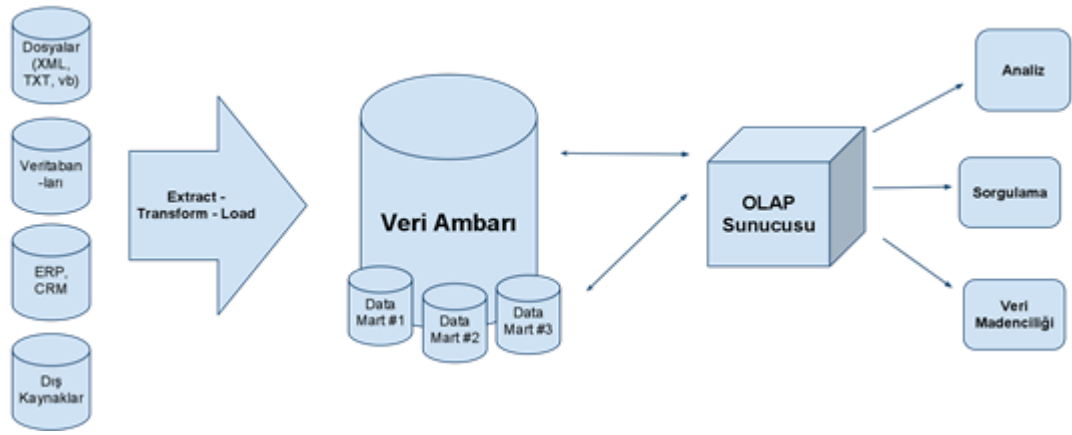
²⁰ Gökhan Silahtaroglu, “**Kavram ve Algoritmalarıyla Temel Veri Madenciliği**”, Papatya Yayıncılık, İstanbul, 2008, s. 15.

bağlantılıdır. Veriler zaman bilgisi ile birlikte saklanmaktadır ve veriler silinemez, yeni veri eklenemez, üzerinde herhangi bir değişiklik yapılamaz.

Data Mart: Veri ambarı tüm işletmeyi ilgilendiren bir yapıdır. Veri ambarlarının tek bir konuya odaklanan parçasına data mart denir.

OLTP (Online Transaction Processing): Bir kurumun günlük verilerinin işlendiği ortamlara OLTP sistemler adı verilmektedir. Bu sistemlere ilişkin veri tabanlarına veri kaydedilebilir, veriye erişilerek raporlanabilir ve istendiğinde silinebilir²¹.

OLAP (Online Analytical Processing): Veri ambarları üzerinde karar vermeye yardımcı olacak çok boyutlu analizler yapma işlemlerine OLAP denir.



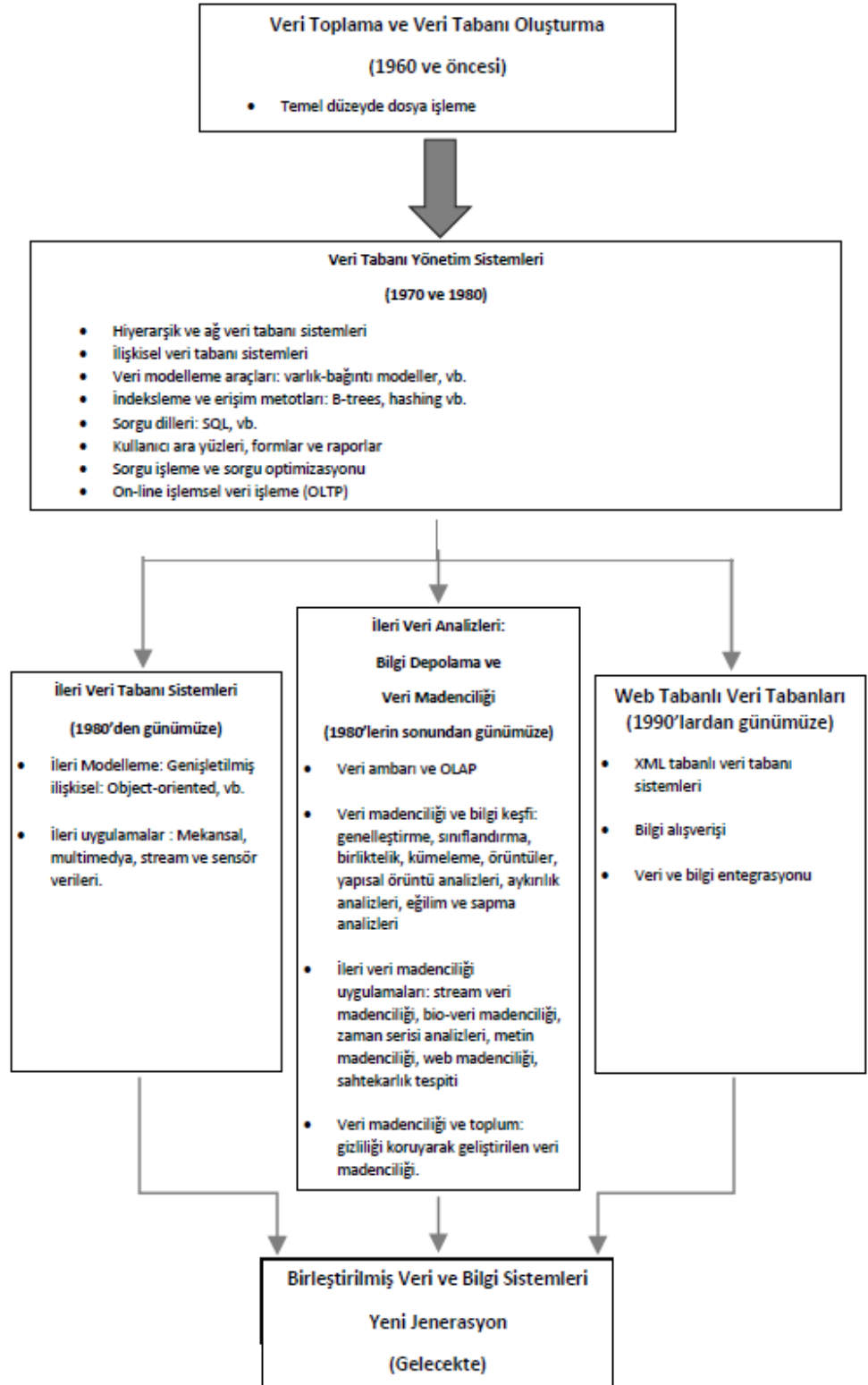
Şekil 1-3: Veri Ambarı Mimarisi

1960'larda veri tabanı ve verilerin depolanması kavramları bilişim dünyasında yerini almıştır. 1960'ların sonunda veri tabanları sadece depolama amacı ile değil aynı zamanda işlem yapma kapasitelerinden dolayı da kullanılmaya başlanmıştır. 1970'lerde E.F. Codd'un yayınladığı ilişkisel veri tabanları hakkındaki makalesi sonucunda insanlar ilişkisel veri tabanları ile tanışmış oldular. Aynı yıllarda, sistem R, MS SQL Server, Sybase geliştirilmiştir. 1980'lerde SQL standart dil olarak kabul edilmiştir. Veri tabanı yönetim sistemleri yaygınlaşmış, bilimsel ve kamusal alanlarda, özel sektörde sıklıkla kullanılmaya başlanmıştır. Bu yıllarda şirketler,

²¹ Yalçın Özkan, “Veri Madenciliği Yöntemleri”, Papatya Yayınları, İstanbul, 2008, s.20.

ürünleri, müşterileri ve rakipleri ile ilgili verilerin toplandığı ve bugünkü adıyla A-CRM (analitik müşteri ilişkileri yönetimi) projelerinin temelini oluşturan veri tabanlarını oluşturmuştur. 1990'larda Visual Basic, ODBC, Excel ve Access kavramları bilişim dünyasında yerini almıştır. Artık veri miktarı katlanarak artan veri tabanlarından, faydalı bilgilerin sadece sorgu ve raporlama sistemleri ile elde edilemeyeceği kabul edilmiş ve nasıl bulunabileceği düşünülmeye başlanmıştır. Bunun üzerine çalışmalara ve uluslararası çalıştaylara başlanmıştır. 1989 yılında Gregory Piatetsky-Shapiro'nun organize ettiği ilk VTBK çalıştay gerçekleştirilmiştir. 1995 yılında bu çalıştay, "ACM SIGKDD Conference on Knowledge Discovery and Data Mining" adı altında düzenli hale getirilmiştir. 1990'lı yılların sonu ve 2000'li yıllarda VM sürekli gelişmiş ve hemen hemen tüm alanlara uygulanmaya başlanmıştır. Google, Facebook, Twitter, LinkedIn gibi popüler ticari şirketlerin ve kamusal güvenlik ile ilgilenen kurumların ciddi yatırım yapması alana olan ilgiyi daha da artmıştır²².

²² Muhsin Özgür Dolgun, "Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının Bağımlı Değişken Prevelansı Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması", **Hacettepe -Üniversitesi Sağlık Bilimler Enstitüsü, Biyoistatistik Programı Doktora Tezi**, Ankara, 2014. s.7-9.



Şekil 1-4: Veri Tabanı ve Veri Madenciliği Gelişim Süreci

1.2. Veri Madenciliği Teknikleri

1.2.1. Kümeleme Analizi

Küme analizi (Cluster analysis) veya kümeleme (clustering) en basit tanımı ile veri dizisinde yer alan benzer nesnelerin aynı gruplarda yer alacak şekilde ayrıştırılmasıdır. Kümeleme analizinde, aynı küme içerisinde yer alan nesnelerin olabildiğince benzer, farklı kümelerde yer alan nesnelerin olabildiğince birbirleri ile ayrışık olması hedeflenir²³.

Kümeleme analizinin veri madenciliğinde etkin bir şekilde kullanılabilmesi için şu koşulları sağlaması gerekmektedir²⁴.

- **Ölçeklenebilirlik:** Pek çok kümeleme teknikleri yalnız küçük veri setleri için iyi sonuçlar vermekle yetinmemeli aynı zamanda milyonlarca veriyi içeren veri tabanlarında da etkili olmalıdır.
- **Farklı ölçek türleri ile ölçülmüş veri setlerine uygulanabilme:** Etkin bir kümeleme algoritmasının nominal , ordinal ve karma verileri de kümeleyebilmesi gerekmektedir. (Kümeleme algoritmaları çoğunlukla oran ölçekli değişkenleri etkin bir şekilde kümelerler.)
- **Değişkenlerin belirlenmesinde alan bilgisini minimize edebilme:** Kümeleme analizi sonuçları kullanılan değişkenlere duyarlı olduğu için çok boyutlu uzayda değişken seçimi yapmak zordur.
- **Rasgele Modellerle ilgili kümelerin keşfi:** Kümeleme algoritmalarının büyük bir kısmı Öklid ve Manhattan uzaklık ölçülerinden yararlanarak benzer büyüklük ve yoğunlukta küresel kümeleri belirlemektedir. Ancak bir küme herhangi bir yapıda olabileceğinden rassal yapıdaki kümeleri ortaya koyabilecek kümeleme algoritmaları geliştirmek gerekmektedir.

²³ Haldun Akpınar, “**DATA: Veri Madenciliği Veri Analizi**”, Papatya Yayınları, İstanbul, 2014, s. 11.

²⁴ Zhijie Xu, Laisheng Wang, Jiangin Luo ve Jiangin Zhang, “A Modified Clustering Algorithm for Data Mining”, **IGRSS’05 Proceedings 2005 IEEE Int.**, Vol.2, 2005, s.741.

- **Gürültülü veri ile çalışabilme:** İyi bir kümeleme algoritmasının gürültülü ve eksik verilerin etkisinden bağımsız olması gerekmektedir.
- **Veri giriş düzenine duyarsızlık:** Bazı kümeleme algoritmaları farklı düzende girilen aynı veri setini farklı kümeleyebilmektedir. Bu nedenle veri girişine duyarlı olmayan kümeleme algoritmaları geliştirmek önem kazanmaktadır.
- **Yüksek Boyutluluk:** Çok boyutlu verileri kümeleyebilmek iyi bir kümeleme algoritmasının en önemli özelliklerindendir.

Kümeleme analizinde sınıflandırma analizindeki gibi küme sayısı önceden bilinmez. Kümeleme analizi, elemanlar arasındaki uzaklıkları ölçebilmek için benzerlik ve uzaklık ölçütlerini kullanır.

Uzaklık, iki nesne arasındaki konum farklılığının sayısal olarak tanımlanması olup, bir kümenin elemanları arasındaki uzaklığı tanımlayan fonksiyon metrik olarak isimlendirilmektedir. Benzerlik ölçüleri ise nesneler arasındaki benzerlik derecesini hesaplamada kullanılan fonksiyonlardır. Benzerlik derecesi $[0,1]$ veya $[-1,+1]$ aralığında yer alır. Benzerlik derecesinin 1 olması tam benzerlik olduğunu, 0 $[0,1]$ veya -1 $[-1,+1]$ tam benzeşmezlik durumunu gösterir²⁵.

Etkin bir kümeleme analizi, en uygun benzerlik ve uzaklık ölçütlerinin kullanılması ile mümkündür.

Uzaklıklar birçok yol ile tanımlanabilir ancak genellikle aşağıdaki koşulların sağlanması gerekmektedir²⁶:

Negatif olmama özelliği:	$d(x,y) \geq 0$
Kendine yakınlık özelliği:	$d(x,x) = 0$
Simetri özelliği:	$d(x,y) = d(y,x)$
Üçgen eşitsizliği özelliği:	$d(x,y) \leq d(x,z) + d(z,y)$

²⁵ Akpınar, a.g.e. s. 161-175.

²⁶ Galit Shmueli, Nitin R. Patel ve Peter C. Bruce., “**Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner**”, John Wiley & Sons, Hoboken, NJ, USA, 2007, s.222-223.

Tablo 1-1: Kümeleme Analizinde Kullanılan Uzaklık Ölçütleri²⁷

Uzaklık Fonksiyonu	Formülü
Öklit Uzaklığı	$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Manhattan Uzaklığı	$d(x, y) = \sum_{i=1}^n x_i - y_i $
Tchebyshev Uzaklığı	$d(x, y) = \max_{i=1,2,\dots,n} x_i - y_i $
Minowski Uzaklığı	$d(x, y) = \sqrt[p]{\sum_{i=1}^n x_i - y_i ^p}, \quad p > 0$
Canberra Uzaklığı	$d(x, y) = \sum_{i=1}^n \frac{ x_i - y_i }{x_i + y_i}, \quad x_i, y_i > 0$
Korelasyon Uzaklığı	$d(x, y) = 1 - r_{(x,y)}^2$
Mahalanobis Uzaklığı	$d(x, y) = \sqrt{ x_i - y_i \sum x_i - y_i }$
Açısal Ayrım	$d(x, y) = \sum_{i=1}^n x_i y_i / \left[\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \right]^{1/2}$

Öklid uzaklığı en yaygın kullanılan uzaklık ölçüsüdür. Ancak Öklid uzaklığının üç temel özelliği kullanılırken dikkate alınmalıdır. Birincisi, yüksek oranda ölçeğe bağımlı olmasıdır. Bir değişkenin biriminin değiştirilmesinin sonuçlar üzerinde büyük etkisi olabilir. Standardizasyon bu sorun için bir çözüm olabilir. Öklid uzaklığının dikkat edilmesi gereken ikinci özelliği de ölçümler arasındaki ilişkileri dikkate almamasıdır. Bu yüzden eğer ölçümler güçlü bir şekilde ilişkili (ör. yüksek

²⁷ Levent Terlemez, “Eş İşlem Stratejisi Yöntemiyle İMKB’de Portföy Oluşturmada Veri Madenciliği Uygulaması”, **Anadolu Üniversitesi Sosyal Bilimler Enstitüsü Doktora Tezi**, Eskişehir 2008, s.43.

korelasyon) ise Mahalanobis uzaklığı gibi başka uzaklık fonksiyonları kullanılmalıdır. Üçüncü olarak ise eğer veri aykırı değerler içeriyor ve bunların giderilmesi mümkün değilse, Manhattan uzaklığı gibi daha sağlam uzaklık fonksiyonları tercih edilebilir²⁸.

1.2.1.1. Kümeleme Algoritmaları

Literatürde bir çok kümeleme algoritması bulunmaktadır. Algoritmaları net bir şekilde sınıflandırmak çok güçtür, çünkü algoritma başka bir çok algoritmanın ortak özelliğini gösterebilir²⁹.

1.2.1.2. Model Esaslı Kümeleme Teknikleri

Model bazlı yöntemler, veriye uygun olan en uygun modeli seçtiği varsayımı ile çalışıyor olmasıdır. İstatistiksel kümeleme yöntemleri ve Yapay Sinir Ağları (En çok kullanılan SOM-Self Organizing Maps- kendini düzenleyen haritalar) olmak üzere ikiye ayrılır.

1.2.1.3. Bölümleyici Kümeleme Teknikleri

Merkeze dayalı bölümleyici kümeleme teknikleri, n adet nesneden oluşan veri setini başlangıçta belirlenen k ($k < n$) adet kümeye ayırmaktadır. Kümeleme sonucunda elde edilen kümelerde, küme içi benzerlik maksimum iken kümeler arası benzerlik minimumdur³⁰. Küme sayısının başlangıçta tanımlanması zorunluluğu, yöntemin en çok eleştirilen darboğazlarının başında gelmektedir³¹.

En çok kullanılan bölümleyici kümeleme algoritmaları K-Means, k-Medoid, PAM, CLARA ve CLARANS algoritmalarıdır (Tablo 1-2.).

²⁸ Shmueli vd. a.g.e.s.223-225.

²⁹ Jiawei Han ve Micheline Kanber, "Data Mining Concepts and Techniques", ABD, Morgan Kaufman Publishers, 2006, s.398.

³⁰ A.e.s.401.

³¹ Akpınar,a.g.e.s. 311.

Tablo 1-2: Bölümleyici Kümeleme Algoritmaları Özellikleri³²

Kümeleme Tekniği	Kullanılan Veri türü	Hesaplama Karmaşıklığı	Geometrik Şekli	Aykırı Ve Gürültülü Değer Durumu	Girdi Parametresi	Sonuçlar	Kümeleme Kriteri (Ölçütü)
K-Means	Nümerik	$O(tkn)$	Konveks olmayan şekilli kümeler için uygun	Duyarlı	Küme Sayısı	Kümelerin Merkezleri	$\min_{v_1, v_2, \dots, v_k} (E_k)$ $E = \sum_{i=1}^k \sum_{p \in C_i} p - m_i ^2$
K-Mode	Kategorik	$O(n)$	Konveks olmayan şekilli kümeler için uygun	Duyarlı	Küme Sayısı	Kümelerin Modları	$\min_{Q_1, Q_2, \dots, Q_k} (E_k)$ $E = \sum_{i=1}^k \sum_{j=1}^n d(X_i, Q_i)$ $d(X_i, Q_i) \rightarrow X_i$ kategorik veriler arasındaki uzaklık, Q_i modu temsil etmektedir
PAM (Medoids)	Nümerik	$O[k(n-k)^2]$	Konveks olmayan şekilli kümeler için uygun	Duyarlı	Küme Sayısı	Kümelerin Medoidleri	$\min(TC_{ih})$ $TC_{ih} = \sum_{j=1}^n C_{jih}$
CLARA (Medoids)	Nümerik	$O[ks^2 + k(n-k)]$	Konveks olmayan şekilli kümeler için uygun	Duyarlı	Küme Sayısı	Kümelerin Medoidleri	$\min(TC_{ih})$ $TC_{ih} = \sum_{j=1}^n C_{jih}$ C_{jih} = j-nci gözlemin i-nci küme yerine h-nci kümeye atanma maliyeti
CLARANS (Medoids)	Nümerik	$O(kn)^2$	Konveks olmayan şekilli kümeler için uygun	Duyarlı	Küme Sayısı	Kümelerin Medoidleri	$\min(TC_{ih})$ $TC_{ih} = \sum_{j=1}^n C_{jih}$

1.2.1.4. Hiyerarşik Kümeleme Teknikleri

Hiyerarşik kümeleme analizinin ana çalışma prensibi, benzer öznitelik değerlerine sahip olan nesnelerin adım adım bir araya getirilmesi veya tam tersine bütünden adım adım ayrılmasıdır. Ayrıca başka analizlere zemin hazırlamak üzere sadece özniteliklerin de kümelenmesi için kullanılabilir. Bu çalışma prensibine uygun olarak hiyerarşik kümeleme analizinde birleştirici ve ayrıştırıcı olmak üzere iki temel yaklaşım bulunmaktadır.

Tüme varım olarak da isimlendirilen birleştirici yaklaşımda, başlangıçta mevcut tüm nesnelerin her biri birbirinden ayrıdır. Nesnelerin öznitelik değerlerinin birbirlerine

³² Yasemin Koldere Akın, “Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi”, **Marmara Üniversitesi, Sosyal Bilimler Enstitüsü Ekonometri ABD İstatistik BD Doktora Tezi**, 2008, s. 88.

olan uzaklık/benzerliklerine göre, birbirlerine en yakın/benzer nesneler birleşerek daha az sayıdaki yeni kümelerde toplanır.

Buna karşılık, tümden gelen olarak da isimlendirilen bölücü yaklaşımda, birleştirici yaklaşımın tam tersi olarak başlangıçta tüm nesneler tek bir kümede toplanmıştır. Her aşamada uzaklık/benzerliklerine göre, nesneler başlangıçtaki bu ana kümeden ayrılarak, farklı kümeler oluşturur³³.

Hiyerarşik kümeleme tekniklerinin avantajları, veri nesnelerinin yoğunluğu konusunda esnek olması, uzaklık ölçülerinin tümünü değerlendirebiliyor olması ve her ölçek türü ile uygulanabilir olmasıdır. Tekniğin dezavantajları ise, iterasyonların bitiş kriterlerin de belirsizliklerin bulunması ve oluşturulan kümeye iyileştirme amacı ile bile geri dönülememesidir³⁴.

Bazı önemli hiyerarşik yöntemler ve özellikleri Tablo 1-3'deki gibi özetlenebilir.

³³ Akpınar,a.g.e.s.299.

³⁴ Pavel Berkhin, "Survey of Clustering Data Mining Techniques", **Technical Report**, Accrue Software Inc., San Jose, California, USA, 2002, s.7.

Tablo 1-3: Hiyerarşik Kümeleme Analizinin özellikleri³⁵

Kümeleme Tekniği	Kullanılan Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekli	Aykırı Ve Gürültülü Değer Durumu	Girdi Parametresi	Sonuçlar	Kümeleme Kriteri (Ölçütü)
BIRCH	Nümerik	$O(n)$	Konveks olmayan şekilli kümeler için uygun değil	Elverişli	Başlangıç (eşik) Değeri, Dal Faktörü	CF=(N: Kümedeki nesne sayısı, LS: Kümedeki nesne sayısının toplamı, SS: nesnelerin kareleri toplamı	Seçilen bir uzaklık kriterine göre nesnelerin kümelere atanması
CURE	Nümerik	$O(n^2 \log n), O(n)$	Keyfi şekiller için uygun	Elverişli	Küme ve temsili kümelerin sayısı	Kümelere nesnelerin atanması	Kümeler ile temsili en yakın nokta çifti Birleştirilmesi
ROCK	Kategorik	$O(n^2 + nm_m m_a + n^2 \log n)$ m_m : bir noktaya komşu olan noktaların maksimum sayısı, m_a : bir noktaya komşu olan komşu noktaların ortalaması	Keyfi şekiller için uygun	Elverişli	Küme sayısı	Kümelere nesnelerin atanması	$E_i = \sum_{i=1}^k n_i \sum_{p_q, p_r \in v_i} \frac{link(p_q, p_r)}{n_i^{1+2f(\theta)}}$ v_i : küme merkezleri $link(p_q, p_r)$: p_q ile p_r arasındaki komşu noktaların sayısı
CHAMELEON	Benzerlik fonksiyonu belirtildiği durumlarda tüm veri türleri için uygun	$O(n^2)$	Keyfi şekiller için uygun	Elverişli	Küme sayısı	Doğal v homojen kümelerin bulunması	$RI(C_i, C_j) = \frac{ EC_{(C_i, C_j)} }{\frac{1}{2}(EC_{C_i} + EC_{C_j})}$ $RC(C_i, C_j) = \frac{\bar{S}_{EC[C_i, C_j]}}{\frac{ C_i }{ C_i + C_j } \bar{S}_{EC_{C_i}} + \frac{ C_j }{ C_i + C_j } \bar{S}_{EC_{C_j}}}$

1.2.1.5. Yoğunluğa Temelli Kümeleme Teknikleri

Hiyerarşik ve bölümleyici küme analizi algoritmaları, veri noktaları arasındaki uzaklıkları temel alarak ilgili hesapları gerçekleştirir. Ancak günümüzde özellikle X-Ray, uydu ve diğer otomatik cihazlardan elde edilen yığın ve mekânsal veri klasik algoritmaların üstesinden gelemeyeceği büyüklükte ve dağılımdadır. İşte bu verileri analiz etmek için geliştirilen algoritmalarından bir kısmı veri noktalarının belirli mekanlarda yoğunlaşması üzerine odaklanmıştır. Bu yaklaşımı geliştiren algoritmalar yoğunluk temelli kümeleme teknikleri olarak isimlendirilir (Tablo 1-4) ³⁶.

³⁵ Akın, a.g.e.s.104.

³⁶ Akpınar, a.g.e.s.357.

Tablo 1-4: Yoğunluk Temelli Algoritmaların Özellikleri³⁷

Kümeleme Tekniği	Kullanılan Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekli	Aykırı Ve Gürültülü Değer Durumu	Girdi Parametresi	Sonuçlar	Kümeleme Kriteri (Ölçütü)
DBSCAN	Nümerik	$O(n \log n)$	Keyfi şekiller için uygun	Elverişli	Küme yarıçapları(ϵ)ve minimum nesne sayısı (MinPts)	Kümelere nesnelerin atanması	Küme içinde yoğunluğa katılan nesneler birleşir
OPTICS	Nümerik	$O(n \log n)$	Keyfi şekiller için uygun	Elverişli	Küme yarıçapları(ϵ)ve minimum nesne sayısı (MinPts)	Kümeleme yapısını ortaya koyan küme sıralamasının oluşturulması	En az belirlenen nesne sayısına sahip olacak ve nesnelerin belirlenen ulaşılabilirlik değeri dahilinde belirli bir yützdenin kalmasını sağlar.
DENCLUE	Nümerik	$O(n \log n)$	Keyfi şekiller için uygun	Elverişli	Küme yarıçapı (σ), Minimum nesne sayısı(ζ)	Kümelere nesnelerin atanması	$f_{Gauss}^D(x) = \sum_{i=1}^n e^{-\frac{d(x,x_i)^2}{2\sigma^2}}$

1.2.1.6. Izgara Tabanlı Kümeleme Teknikleri

Kümelemenin yapılabilmesi için çok yüksek miktarda bellek gerektiren büyük veri tabanlarının, kümelere ayrılması için numaralandırılmış ızgaralardan oluşan hücreler yardımıyla yapılan kümelere ızgara tabanlı kümeleme teknikleri adı verilir³⁸. Bu teknikler etkili algoritmalar olmasına rağmen, dezavantajları algoritmalar etkisi önceden belirlenmiş ızgaraların büyüklüğünden ve belirli hücrelerin eşik değerinden ciddi şekilde etkilenmeleridir.

³⁷ Akın, a.g.e.s.115.

³⁸ Gökhan Silahtaroglu, “**Kavram ve Algoritmalarıyla Temel Veri Madenciliği**”, Papatya Yayıncılık, İstanbul, 2008, s. 129.

Tablo 1-5: Izgara Tabanlı Kümeleme Tekniklerinin Özellikleri³⁹

Kümeleme Tekniği	Kullanılan Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekli	Aykırı Ve Gürültülü Değer Durumu	Girdi Parametresi	Sonuçlar	Kümeleme Kriteri (Ölçütü)
DBSCAN	Nümerik	$O(n \log n)$	Keyfi şekiller için uygun	Elverişli	Küme yarıçapları(ϵ)ve minimum nesne sayısı (MinPts)	Kümelere nesnelerin atanması	Küme içinde yoğunluğa katılan nesneler birleşir
OPTICS	Nümerik	$O(n \log n)$	Keyfi şekiller için uygun	Elverişli	Küme yarıçapları(ϵ)ve minimum nesne sayısı (MinPts)	Kümeleme yapısını ortaya koyan küme sıralamasının oluşturulması	En az belirlenen nesne sayısına sahip olacak ve nesnelerin belirlenen ulaşılabilirlik değeri dahilinde belirli bir yüzdenin kalmasını sağlar.
DENCLUE	Nümerik	$O(n \log n)$	Keyfi şekiller için uygun	Elverişli	Küme yarıçapı (σ), Minimum nesne sayısı (ζ)	Kümelere nesnelerin atanması	$f_{Gauss}^D(x) = \sum_{i=1}^n e^{-\frac{d(x,x_i)^2}{2\sigma^2}}$

1.2.1.7. Kendini Düzenleyen Haritalar-Kohonen Ağları

Kohonen ağları diğer bir adıyla SOM(Self Organizing Map) kendi kendini organize eden harita yöntemi olarak adlandırılmaktadır. Kohonen Ağları, yapay sinir ağları tekniklerinden biri olan standart ileri besleme ağlarından sonra en yaygın kullanılan teknik olup, kümeleme ve boyut indirgeme için kullanılan denetimsiz iki katmanlı bir yapay sinir ağı olarak tanımlanabilir⁴⁰.

Kohonen ağları giriş ve çıkış nöronlarından oluşur. Her bir sinir (nöron), girdi değişkenleri veya aynı anlamda girdi alanlarının her biri ile bağlantılıdır ve tekrar ağırlıkları (önemleri) bu bağlantıların her birinin üzerinde yer almaktadır. Bir sinir için ağırlıklar, analizde kullanılan girdi alanlarının oluşturduğu küme için bir profili temsil etmektedir. Genellikle bir Kohonen ağı, az sayıda birim çok sayıda gözlemi özetlediğinde (güçlü birimler) veya çok sayıda birim gözlemlerin her hangi birine karşılık gelemediğinde (zayıf birimler)son bulacaktır. Kohonen veya kendini düzenleyen harita, bir çıktı olarak düşünülen sinirleri içermesine karşılık, bu ağlarda

³⁹ Akın, a.g.e.s.115.

⁴⁰ Dmitri G. ROUSSINOV ve Hsinchun CHEN, “Document Clustering For Electronic Meetings: An Experimental Comparison Of Two Techniques”, **Decision Support Systems**, Vol: 27, Issue: 1-2, 1999, s. 70.

gerçek çıktı katmanı bulunmamaktadır⁴¹. Kohonen ağırları eğitilirken, bu topolojik komşuluk referans vektörlerinin yenilenmesinde kullanılır. Bu ağlarda kullanılan öğrenme algoritması denetimsizdir. Yani, ağ eğitilirken bağımlı değişken kullanılmaz⁴².

Bir yapay sinir ağı tekniği olan Kohonen Ağları diğer istatistiksel kümeleme yöntemlerine pek çok yönden üstünlük sağlamaktadır. Büyük veri setleri ile çalışıldığı için istatistiksel yöntemlerin varsayımlarını yerine getirememekte ve bu yöntemlerin gücünü azaltmaktadır. Ayrıca SOM yüksek boyuttaki verileri kümeleyebilme özelliğinin yanında görselleştirebilmektedir. Bu yöntemin dezavantajı ise en iyi kümeleme sonucuna karar verilmesine dair net bir ölçünün bulunmamasıdır.

1.2.2. Birliktelik Analizi

Olayların birlikte gerçekleşme durumlarını çözümleyen veri madenciliği yöntemleridir. Bu yöntemler birlikte olma kurallarını belirli olasılıklarla ortaya koyar⁴³. Bir birliktelik kuralı, veri tabanı içerisinde belirli nesnelerin birlikte görülme olasılığı ifadesidir⁴⁴.

Birliktelik analizinin en yaygın uygulaması perakende satışlarda müşterilerin satın alma eğilimlerini belirlemek amacıyla yapılmaktadır. Müşterilerin bir anda satın aldığı tüm ürünleri ele alarak satın alma eğilimini ortaya koyan uygulamalara “pazar sepet analizi” adı verilmektedir⁴⁵.

Birliktelik kuralları yöntemlerinin uygulanmasında dikkat edilmesi gereken iki konu vardır. Birincisi, büyük bir işlemsel veri setinden örüntülerin çıkartılmasının bilgisayar kaynakları ve hesaplamalar açısından oldukça maliyetli ve zaman alıcı olmasıdır.

⁴¹ Ayşe Oğuzlar, "Kümeleme Analizinde Yeni Bir Yaklaşım: Kendini Düzenleyen Ağlar (Kohonen Ağları)." *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 19.2,2005, s. 94-95

⁴² Hüdaverdi Bircan, Metin Zontul ve A.Gürkan Yüksek, “SOM Tipinde Yapay Sinir Ağlarını Kullanarak Türkiye’nin İhracat Yaptığı Ülkelerin Kümelenmesi Üzerine Bir Çalışma”, *Atatürk Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt:20,Sayı:2, 2006, s.222-223

⁴³ Yalçın Özkan, “**Veri Madenciliği Yöntemleri**”, Papatya Yayınları, İstanbul, 2008, s.157.

⁴⁴ Hand,a.g.e.s.158.

⁴⁵ Özkan, s.157.

İkincisi ise bulunan bazı örüntülerin şans eseri ortaya çıkma ve sahte olma ihtimalidir. Bunun için keşfedilen örüntüler ayrıca değerlendirme süreçlerinden geçirilmelidir⁴⁶.

Birliktelik kurallarında üretilen kuralların bağıllık gücünü ölçmek için destek (support), güven (confidence) ve kaldırma oranı (lift ratio) gibi ölçütler kullanılır. Bu ölçütler, kuralın ilginçliğini ve ilgililiğini ifade eden iki ölçüdür. Kuralların matematiksel gösterimi şöyledir⁴⁷:

$$support(A \Rightarrow B) = P(A \cup B)$$

$$confidence(A \Rightarrow B) = P(B | A) = \frac{P(A \cup B)}{P(A)}$$

$$lift ratio(A \Rightarrow B) = \frac{confidence(A \Rightarrow B)}{expected\ confidence(A \Rightarrow B)} = \frac{P(B | A)}{P(B)}$$

A öncül (antecedent), B ise ardıl (consequent) öge setidir. Destek, öncül ve ardıl öge setlerini birlikte içeren işlem sayısının veri tabanındaki toplam işlem sayısına oranıdır. Güven ölçüsü, ardıl öge setinin öncül öge setine koşullu olasılığıdır. Yani, işlemlerin öncül ve ardıl öge setlerini birlikte içermeye olasılığının, öncül öge setini içermeye olasılığına oranıdır. Kaldırma oranı ise, kuralın güven değerinin beklenen güven değerine oranıdır. Beklenen güven değeri (expected confidence) ardıl öge setini içeren işlem sayısının veri tabanındaki toplam işlem sayısına oranıdır.

⁴⁶ Pang-Ning Tan, Michael Steinbach, Vipin Kumar, “**Introduction to Data Mining**”, Pearson, Addison- Wesley, Boston, MA, USA, 2006, s.328.

⁴⁷ Han vd.a.g.e.s.246 ve Shmueli vd.a.g.e.s.206-207.

1.2.2.1. Apriori Algoritması

Apriori algoritması birliktelik kurallarında en çok kullanılan algoritmadır. Agrawal ve Srikant (1994) tarafından önerilen ve Boolean birliktelik kuralları için sık öge setleri madenciliğinde kullanılan bir algoritmadır. Algoritmanın ismi, sık öge setlerinin özelliklerine ait öncül bilgi (prior) kullanması prensibine dayanmaktadır⁴⁸.

Apriori algoritmasının aşamaları şunlardır⁴⁹:

- a) Birliktelik çözümlemesinin yapılabilmesi için öncelikle destek ve güven ölçülerini karşılaştırmak üzere eşik değerler belirlenir. Elde edilen sonuçların bu eşik değerlerden eşit veya büyük olması istenir.
- b) Veri tabanı taranarak çözümlemeye dahil edilecek her bir ürün için tekrar sayıları, yani destek sayıları hesaplanır. Bu destek sayıları eşik destek sayısı ile karşılaştırılır. Eşik destek sayısından küçük değere sahip satırlar çözümlemeden çıkarılır ve koşula uygun kayıtlar göz önüne alınır.
- c) Yukarıdaki adımda seçilen ürünler bu kez ikişerli gruplandırılarak, bu grupların tekrar sayıları, yani destek sayıları elde edilir. Bu sayılar eşik destek sayıları ile karşılaştırılır. Eşik değerden küçük değerlere sahip satırlar çözümlemeden çıkarılır.
- d) Bu kez üçerli, dörderli vb. gruplandırmalar yapılarak bu grupların destek sayıları elde edilir ve eşik değer ile karşılaştırılır, eşik değerlere uygun olduğu sürece işlemlere devam edilir.
- e) Ürün grubu belirlendikten sonra kural destek ölçütüne bakılarak birliktelik kuralları türetilir ve bu kuralların her birisiyle ilgili olarak güven ölçütleri hesaplanır.

⁴⁸ Han vd.a.g.e.234-235.

⁴⁹ Özkan, a.g.e.s.159.

1.2.3. Karar Ağaçları

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, veri madenciliğinde hem kural çıkarmanın hem yorumlanmalarının hem de sonuçlarının görsel olarak sunumunun kolay olması, veri tabanı sistemlerine kolayca entegre edilebilmeleri, güvenilirliklerinin iyi olması nedenleri ile veri madenciliği modelleri içerisinde en yaygın kullanıma sahip olan tekniktir⁵⁰⁵¹. Büyük veri tabanlarının kullanıldığı pek çok sınıflama probleminde ve karmaşık ya da hata içeren bilgilerde karar ağaçları yararlı bir çözüm olmaktadır⁵².

Karar ağaçları ve karar kuralları birçok gerçek yaşam uygulamasında sınıflandırma problemlerine güçlü çözümler üreten veri madenciliği yöntemleridir. Başlıcaları makine öğrenimi ve uygulamalı istatistik alanları olmak üzere literatürde tanımlanmış çok sayıda karar ağacı endüksiyon algoritması vardır. Bu algoritmalar bir girdi-çıkı örneklem setinden karar ağaçları yapılandıran denetimli öğrenme metotlarıdır. Tipik bir karar ağacı öğrenme sistemi, arama uzayının bir bölümünde bir çözüm arayan yukarıdan-aşağıya bir strateji kullanır. Karar ağacı, özniteliklerin test edildiği düğümlerden (node) oluşur. Bir düğümün altında kalan dallar, düğümde test edilen tüm olası sonuçlara karşılık gelir⁵³.

Karar ağaçları akış şemasına benzeyen yapılardır. Her bir nitelik bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının elemanlarıdır. En son yapı yaprak, en üst yapı kök ve bunların arasında kalan yapılar ise dal olarak isimlendirilir⁵⁴. Her bir düğüm, bir değişken üzerinde gerçekleştirilen sınamayı, her bir dal ilgili değişken üzerinde gerçekleştirilen bölümlmeleri ve her bir yaprak da öngörünün sonucunu ifade etmektedir⁵⁵.

⁵⁰ Jie Sun, Hui Li,” Data mining method for listed companies’ financial distress prediction”, **Knowledge-Based Systems**, 21 (1–5), 2008, s.2.

⁵¹ Sigurdur Olafsson , Xiaonan Li, Shuning Wu,” Operations research and data mining”, **European Journal of Operational Research**, 187 1429–1448, 2008, s.1436.

⁵² M.J. Aitkenhead,” A co-evolving decision tree classification method”, **Expert Systems with Applications**, 34, 18–25, 2008, s.18.

⁵³ Mehmed Kantardzic, “**Data Mining: Concepts, Models, Methods, and Algorithms**”, IEEE Press, Hoes Lane, Piscataway, USA, 2003, s.139-140.

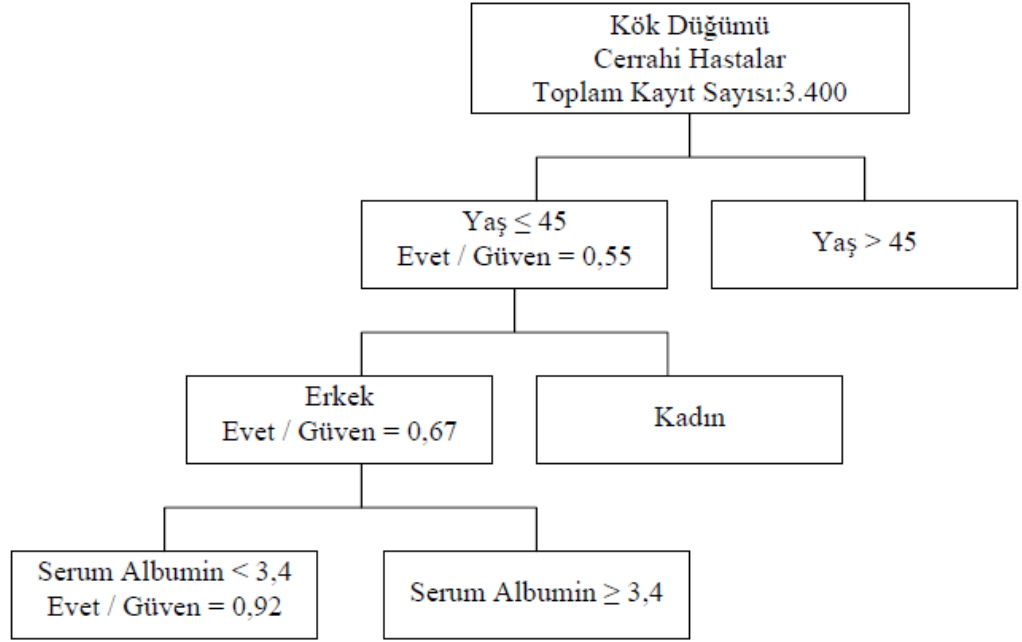
⁵⁴ Özkan, a.g.e.s.53.

⁵⁵ Han vd.a.g.e.s.291.

Şekil 1.5’de bir karar ağacı örneği görülmektedir. Bu karar ağacı ameliyat edilmiş hastaların hangi özelliklere göre ve hangi olasılıklarla belirli bir cerrahi komplikasyon geliştirebileceğini bulmak için oluşturulmuştur. Bağımlı değişken ikili yapıdadır (binary) ve bu yapıda 1, Evet seçeneğini gösterirken 0 da Hayır’ı göstermektedir. Modelin en tepesinde ağacın kök düğümü bulunmaktadır ve 3400 gözlemin tamamını barındırmaktadır. Karar ağacı yapılandırıldıkça benzer davranış gösteren homojen hastalar grubunun daha rafine alt bölümlerini (subsegments) oluşturmaktadır. Herhangi bir terminal düğümden (son düğüm) yukarı doğru bakıldığında o hastalar grubunu tanımlayan bir eğer-ise (if-then) kurallar zinciri elde edilir. Karar ağacında yukarıdan aşağı doğru inildikçe ise kurala yeni karakteristikler eklenir ve tahminin doğruluğu konusundaki güven değeri de artar. Örnekte verilen sol alt tarafta yer alan terminal düğümünün kuralı şu şekilde tanımlanabilir: 45 yaşından küçük, erkek ve serum albümin değerleri 3,4 g/dL’den daha aşağı olan hastaların belirli bir cerrahi komplikasyon geliştirme riski güven seviyesi %92’dir^{56 57}.

⁵⁶ Anne M. Berger ve Charles R. Berger, “Data Mining as a Tool for Research and Knowledge Development in Nursing”, **Computers, Informatics, Nursing**, Volume 22(3) (s.123-131), 2004, s.128.

⁵⁷ Sezgin Irmak, “Veri madenciliği Yöntemleri ile Sağlık Sektörü Veri tabanlarında Bilgi Keşfi: Tanımlayıcı ve Kestirimci Model Uygulamaları”, **Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü İşletme Doktora Tezi**, 2009, s. 37.



Şekil 1-5: Cerrahi Komplikasyon Riski Taşıyan Hastalar için Karar Ağacı

Karar Ağaçlarında bağımlı değişken kesikli ise sınıflandırma ağacı, sürekli ise tahmin ağacıdır.

Karar ağaçları oluşturulması için bazı gerekli kavramlar ve tanımları şöyle özetlenebilir⁵⁸:

- Ayırma özelliklerinin seçilmesi: Ayırma özellikleri için hangi özelliklerin kullanılacağı inşa edilmiş karar ağacının uygulanmasındaki performansını etkiler. Bazı özellikler diğerlerinden daha iyidir. Örneğin, bir veri tabanında yer alan kaydın isim özelliği kesinlikle kullanılmamalıdır. Özelliğin seçimi, eğitim setindeki verinin incelenmesinin yanında alan uzmanlarının bilgilendirilmiş girdisini de gerektirir.
- Ayırma özelliklerinin sıralanması: Özelliklerin seçildiği sıra da önemlidir. Bazı durumlarda birinci olarak seçilen bir özellik, gereksiz karşılaştırmaları içerebileceğinden, ikinci bir sefer incelenmeyi gerektirebilir.

⁵⁸ Terlemez, a.g.e. s.49.

- Ayırmalar: Özelliklerin sıralanması ile ilişkilendirilenler alınacak ayırmaların sayısıdır. Bazı özelliklerle birlikte, tanım kümesi küçüktür, dolayısıyla ayırmaların sayısı açıkça tanım kümesine dayanır. Buna rağmen, eğer tanım kümesi sürekli ise veya çok sayıda değere sahipse, kullanılacak ayırmaların sayısı kolaylıkla belirlenemez.
- Ağaç yapısı: Sınıflandırma için ağacın uygulanma performansını iyileştirmek için, en az seviyeye sahip dengelenmiş bir ağaç arzu edilir. Buna rağmen, bazı durumlarda çok yollu dallanmalı daha karmaşık karşılaştırmalara ihtiyaç duyulabilir. Bazı algoritmalar sadece ikili ağaçlar inşa eder.
- Durma ölçütü: Ağacın oluşturulması, veri seti mükemmel bir şekilde sınıflandırıldığında kesinlikle durur. Daha büyük ağaçların yaratılmasını önlemek için daha önce durulmasının arzu edilebileceği durumlar olabilir. Bu sınıflandırmanın doğruluğu ve performansı arasındaki bir alış verıştır. Buna ilave olarak daha erken durmak, gereğinden fazla uyumu bertaraf etmek için de uygulanabilir.
- Eğitim verisi: Oluşturulan karar ağacının yapısı eğitim verisine bağlıdır. Eğer eğitim verisi çok küçük ise, o zaman oluşturulmuş ağaç daha genel veriyle düzgün çalışacak kadar belirli olmayabilir. Eğer eğitim verisi çok büyük ise, o zaman oluşturulan ağaç gereğinden fazla uyum gösterebilir.
- Budama: Ağaç oluşturulduktan sonra, sınıflandırma aşaması esnasında ağacın performansını iyileştirmek için ağaçta bazı değişikliklere ihtiyaç duyulabilir. Budama aşaması daha iyi performansa ulaşabilmek için, gereksiz karşılaştırmaları veya alt ağaçları kaldırabilir.

Karar Ağacı algoritmalarından en çok kullanılanlar ID3, C4.5, C5.0 ve CART'dır.

1.2.3.1. CART

CART veya C&RT (Classification and Regression Trees) Beriman ve arkadaşları tarafından 1984 yılında geliştirilmiş, ikili (binary) ağaç olarak büyüyen bir algoritmadır. CART veriyi iki alt kümeye ayırır. Böylece bir sonraki adımda oluşacak

alt küme, bir öncekinden daha homojen olacaktır. Bu sonuç bulunana kadar devam eden kendini tekrar eden bir süreçtir⁵⁹. CART analizi, cevap değişkeninin kesikli olması durumunda sınıflandırma ağacı, (Classification Tree, CT) sürekli olması durumunda ise regresyon ağacı (Regression Tree, RT) olarak adlandırılır.

1.2.1.1. ID3

Karar ağaçları yardımıyla sınıflandırma işlemlerini yerine getirmek üzere Quinlan (1979) tarafından gerçekleştirilmiştir. Karmaşık bir karar verme sürecini basit kararların bir koleksiyonuna parçalaması ve kolay yorumlanabilir çözüm sağlaması, karar ağaçlarının en önemli özellikleridir. ID'ler, sembolik verinin sınıflandırması için popüler ve etkili karar verme yöntemleridir. Nümerik değerlerin olduğu durumlara genellikle uygun değildirler. Çoğu gerçek hayat problemleri sembolik olmayan (sürekli) değerlerle çalıştığı için, özellik seçiminden önce kesiklendirme uygulanmalıdır⁶⁰.

ID3 algoritması diğer değişkenler içerisinde sınıflamada en ayırıcı özelliğe sahip değişkeni bulurken entropi kavramından yararlanır. Entropi kavramı eldeki bilginin sayısallaştırılmasıdır⁶¹. Bir sistemdeki belirsizliğin ölçüsüne entropi denir⁶². ID3, düğümleri ayırmak için entropiyi kullanmaktadır.

1.2.1.2. CHAID

CHAID (Chi-Squared Automatic Interaction Detector) algoritması 1980 yılında Kaas tarafından geliştirilmiş oldukça başarılı bir karar ağacı tekniğidir. CHAID adından da anlaşılacağı gibi ayırma kriteri olarak Ki-kareyi kullanır. CHAID algoritması, tahmin edici değişkenin tüm değerlerini dikkate alarak analiz yapar. Hedef değişkeni dikkate

⁵⁹ Umman Tuğba Şimşek Gürsoy, "Veri Madenciliği ve Bilgi Keşfi", Pegem Akademi yayınları, Ankara, 2009, s.75.

⁶⁰ Esra Kahya Özyirmidokuz, "Veri Madenciliği Tekniklerini Kullanarak İmalat Verilerinin Modellenmesi ve Analizi", Erciyes Üniversitesi Sosyal Bilimler Enstitüsü, Kayseri, 2009, s.56.

⁶¹ Silahtaroglu, a.g.e.s.51.

⁶² Özkan, a.g.e.s.55.

olarak istatistik olarak benzer olan değişkenleri birleştirir ve farklı olan değişkenlerle işlemi sürdürür. Daha sonra karar ağacının ilk dalını oluşturmak için en iyi tahmin edici değişkeni seçer. Her bir düğüm, seçilen değişkenin benzer değerlerinden oluşur. Bu süreç ağaç tamamıyla büyüyene kadar tekrarlanarak devam eder. Yapılacak testler hedef değişkenin türüne göre değişmektedir. Eğer değişken sürekli bir değişken ise F testi, kategorik (nominal/ordinal) bir değişken ise Ki- kare testi kullanılır⁶³.

1.2.1.3. C4.5 VE C5

C4.5 ve C5.0 1993 yılında Quinlan tarafından geliştirilmiş olup, ID3 algoritmasının geliştirilmiş algoritmalarıdır. İlk geliştirildiklerinde kullanımları kategorik veri ile sınırlıdır, şu andaki versiyonları sürekli veri türleri üzerinde de uygulama olanağı sağlamaktadır. Bu algoritmalarda karar ağacı oluşturulurken kayıp veriler hesba katılmaz. Kazanım oranı hesaplanırken, sadece verileri eksik olmayan kayıtlar kullanılır. Bu algoritmalar, kayıp verileri diğer veri ve değişkenler yardımıyla öngörerek kazanım oranının hesaplamasında kullanır. Böylece ID3 algoritmasından daha duyarlı ve daha anlamlı kurallar çıkartılabilen ağaç üretilir.

1.2.4. Lojistik Regresyon

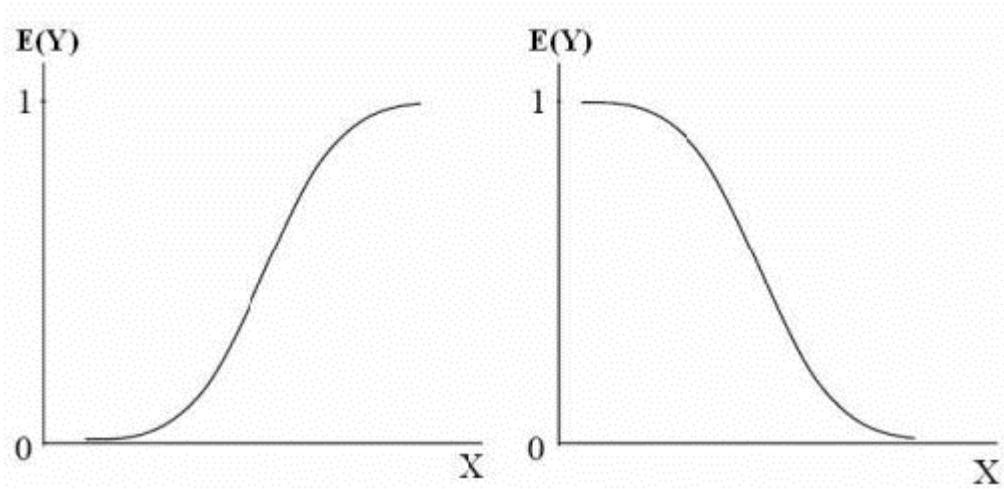
Lojistik regresyon analizi, bağımsız değişkenlerde çoklu normallik ve gruplar arası varyans kovaryans matrislerinin eşitliği varsayımları sağlanamadığında diskriminant analizine alternatif olarak geliştirilmiştir. Lojistik regresyon analizi, çoklu regresyon analizine benzediği, basit, kolay ve açık istatistik testlere sahip olduğu, metrik ve metrik olmayan değişkenleri birlikte analiz edebilme ve detaylı sonuç verme özelliklerine sahip olduğu için tercih edilmektedir⁶⁴. Ancak diskriminant analizi tüm öngörü değişkenlerinin normal dağıldığı varsayımı, nominal öngörü değişkenlerinin kullanılamaması ve sınıfları ayıran sınırların doğrusal olması gibi sebeplerle veri madenciliğinde oldukça yaygın olarak kullanılmamaktadır⁶⁵.

⁶³ Gürsoy, a.g.e.s.76.

⁶⁴ Gürsoy, a.g.e.s.101.

⁶⁵ Introduction to Data Mining and Knowledge Discovery, Third Edition, 2005, <http://www.twocrows.com/intro-dm.pdf> (Şubat, 2015)

Hem teorik hem de deneysel incelemeler bağımlı değişken iki sonuçlu iken cevap fonksiyonunun $p/(1-p)$ şeklinin S veya ters S şeklinde olacağını göstermiştir. Şekil 1-6’da görüldüğü gibi bağımlı değişken bitiş noktaları dışında yaklaşık olarak doğrusaldır yani cevap fonksiyonları 0 ile 1 değerlerinde X ve Y eksenlerine asimptotiktir ⁶⁶.



Şekil 1-6: İkili bağımlı değişkenin S ve ters S olasılık fonksiyonu grafikleri

Lojistik regresyon modeli şu şekildedir⁶⁷:

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}$$

⁶⁶ Hüdaverdi Bircan, “Lojistik regresyon analizi: tıp verileri üzerine bir uygulama”, **Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi**, 2, 185-208, 2004, s.(189-190).

⁶⁷Kazım Özdamar, “**Paket Programlar İle İstatistiksel Veri Analizi**”, Kaan Kitabevi, Eskişehir, 2002, s. 625 – 627.

- $\pi(x)$:İncelenen olayın gözlenme olasılığını,
- β_0 :Bağımsız değişkenler sıfır değerini aldığı anda bağımlı değişkenin değerini başka bir ifadeyle sabiti,
- $\beta_1\beta_2....\beta_p$:Bağımsız değişkenlerin regresyon katsayılarını,
- $X_1X_2....X_p$:Bağımsız değişkenleri,
- p :Bağımsız değişken sayısını,
- e :2,718 sayısını göstermektedir.

1.2.5. Bayes Sınıflandırıcılar

Bayesyen sınıflandırma tekniği, elde var olan, hali hazırda sınıflanmış verileri kullanarak yeni bir verinin mevcut sınıflardan herhangi birine girme olasılığını hesaplayan bir yöntemdir. Bayesyen kuralına dayalı geliştirilmiş algoritma ve sınıflandırma teknikleri bu adla anılır⁶⁸.

Bir sınıflandırma probleminde amaç belirli kestirimci değişkenler kümesi verildiği durumda, her sınıfa ilişkin üyelik olasılıklarını tahmin etmektir. Bu tür bir olasılık koşullu olasılık olarak adlandırılır. Y olayının verilen X olayına göre koşullu olasılığı $P(Y | X)$ sadece X olayının gerçekleştiği durumlarda Y olayının da gerçekleşmesi olasılığını temsil eder⁶⁹. Bayes teoremi şöyle formüle edilir:

$$P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}$$

Naïve Bayes sınıflandırıcılar Bayes teoremine dayalı daha sade ve bazı veri madenciliği tekniklerine oranla daha hızlı ve yüksek performanslı bir veri madenciliği yöntemidir. Naïve Bayes sınıflandırıcılar belirli bir sınıfta yer alan değişkenin

⁶⁸ Silahtaroglu, a.g.e., s.60.

⁶⁹ Shmueli vd.s.94.

değerinin diğer değişkenlerinin değerlerinden bağımsız olduğunu varsayar. Bu varsayım koşullu bağımsızlık kavramı ile ifade edilir⁷⁰.

Naïve Bayes sınıflandırıcı aşağıdaki şekilde çalışmaktadır⁷¹

- 1) D değişken grupları ve ilgili sınıf etiketlerinden oluşan bir eğitim seti olduğu varsayılın. Her değişkenler grubu bir n -boyutlu öznitelik vektörü $X = (x_1, x_2, \dots, x_n)$ ile temsil edilmektedir ve $A_1, A_2, \dots, A_n$ n adet öznitelikten oluşan bir değişkenler grubunda yapılan n adet ölçümü belirtmektedir.
- 2) $C_1, C_2, \dots, C_m$ dan oluşan m tane sınıf olduğunu varsayılın. Bir bilinmeyen veri örnekleme olan X'de, sınıflandırıcı X'in X'e koşullu olmak üzere en yüksek ardıl olasılığa sahip olduğunu öngörür. Bu durum, Navie Bayesyen sınıflandırıcının bilinmeyen örneklem X'i sadece aşağıda belirtilen durumda sınıf C_i 'ye atamaktadır.

$$P(C_i | X) > P(C_j | X) \quad ; \quad 1 \leq j \leq m, j \neq i$$

Dolayısıyla $P(C_i | X)$ maksimize edilmiş olur. $P(C_i | X)$ 'in maksimize edildiği C_i sınıfı maksimum ardıl hipotez olarak adlandırılır Bayes teoremi ise aşağıdaki gibidir:

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{P(X)}$$

- 3) $P(X)$ bütün sınıflar için sabit olduğundan, sadece $P(X | C_i)P(C_i)$ 'nin maksimize edilmesi gerekmektedir. Eğer sınıfların öncül olasılıkları bilinmiyorsa, çoğunlukla bütün sınıfların eşit olasılığa sahip olduğu kabul edilir: $P(C_1) = P(C_2) = \dots = P(C_m)$. Böyle olduğu durumlarda $P(X | C_i)$ maksimize edilir. Aksi halde $P(X | C_i)P(C_i)$ maksimize edilir. Sınıf öncül olasılıkları $P(C_i) = \frac{|C_{i,D}|}{|D|}$ ile tahmin edilebilir. $|C_{i,D}|$ C_i sınıfının D eğitim seti içerisindeki eğitim değişkenler grubu sayısıdır.

⁷⁰ Han vd.a.g.e.s.310.

⁷¹ Han vd.a.g.e.s.311-313.

- 4) Birçok öznitelik (attribute) içeren veri setlerinde $P(X|C_i)$ 'nin hesaplanması oldukça maliyetli olabilir. $P(X|C_i)$ 'nin değerlendirilmesinde hesaplamaların azaltılabilmesi için, Naïve sınıf koşullu bağımsızlık varsayımı yapılır. Bu varsayım, özniteliklerin değerlerinin birbirinden koşullu bağımsızlığını öngörür. Yani öznitelikler arasında bir bağımlılık ilişkisi yoktur. Bu nedenle,

$$\begin{aligned} P(X | C_i) &= \prod_{k=1}^n P(x_k | C_i) \\ &= P(x_1 | C_i) \times P(x_2 | C_i) \times \cdots \times P(x_n | C_i) \end{aligned}$$

$P(x_1|C_i)$, $P(x_2|C_i)$,....., $P(x_n|C_i)$ olasılıkları eğitim değişken gruplarından kolaylıkla hesaplanabilir. Burada x_k , X değişkenler grubu için A_k özniteliğinin aldığı değeri temsil etmektedir. Her öznitelik için, özniteliğin kategorik veya sürekli bir değişken olup olmadığına bakılmalıdır. $P(X|C_i)$ 'yi hesaplamak için aşağıdaki durumlar dikkate alınır:

- Eğer A_k kategorik değişken ise, bu durumda $P(X_k|C_i)$, C_i sınıfının D içindeki A_k 'nın x_k değerini aldığı değişken grubu sayısının, C_i sınıfının D içindeki toplam değişken grubu sayısına ($|C_{i,D}|$) oranıdır.
- Eğer A_k sürekli bir değişken ise biraz daha fazla ancak daha kolay bir hesaplama yapmak gereklidir. Sürekli değer alan bir değişkenin tipik olarak aşağıda tanımlandığı şekilde ortalama ve standart sapmaya sahip normal dağılıma sahip olduğu varsayılır. Böylece;

$$P(x_k | C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i})$$

μ_{C_i} ve σ_{C_i} 'yi, C_i sınıfının eğitim değişken grupları için A_k özniteliğinin değerlerine ait ortalama ve standart sapmayı hesaplamak gerekmektedir. Sonra bu iki değer normal dağılım denkleminde $P(X_k|C_i)$ 'yi tahmin etmek üzere x_k ile birlikte yerine konulur.

- X 'in etiketini kestirmek için her sınıf için $P(X_k|C_i)$, $P(C_i)$ değerlendirilir. Sınıflandırıcı aşağıdaki durum gerçekleştiğinde X değişkenler grubunun sınıf etiketini C_i olarak kestirir.

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \quad ; \quad 1 \leq j \leq m, j \neq i$$

Naïve Bayes sınıflandırıcıların iyi tarafı basitliğine, bilgisayar kaynakları ve hesaplamalar açısından etkinliğine ve iyi sınıflandırma performansına dayandırılabilir. Amacın veri setindeki kayıtların belirli bir sınıfın olasılık değerlerine göre sınıflandırılması veya derecelendirilmesi olduğu durumlarda Naïve Bayes sınıflandırıcılar iyi performans göstermektedir⁷².

1.2.6. Yapay Sinir Ağları

Yapay sinir ağları (YSA), insan beyninin özelliklerinden olan, öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacıyla geliştirilen bilgisayar sistemleridir⁷³.

YSA kavramı beynin çalışma ilkelerinin sayısal bilgisayarlar üzerinde taklit edilmesi fikri olarak ortaya çıkmış ve ilk çalışmalar beyni oluşturan biyolojik hücrelerin, ya da literatürdeki ismi ile nöronların matematiksel olarak modellenmesi üzerinde yoğunlaşmıştır⁷⁴.

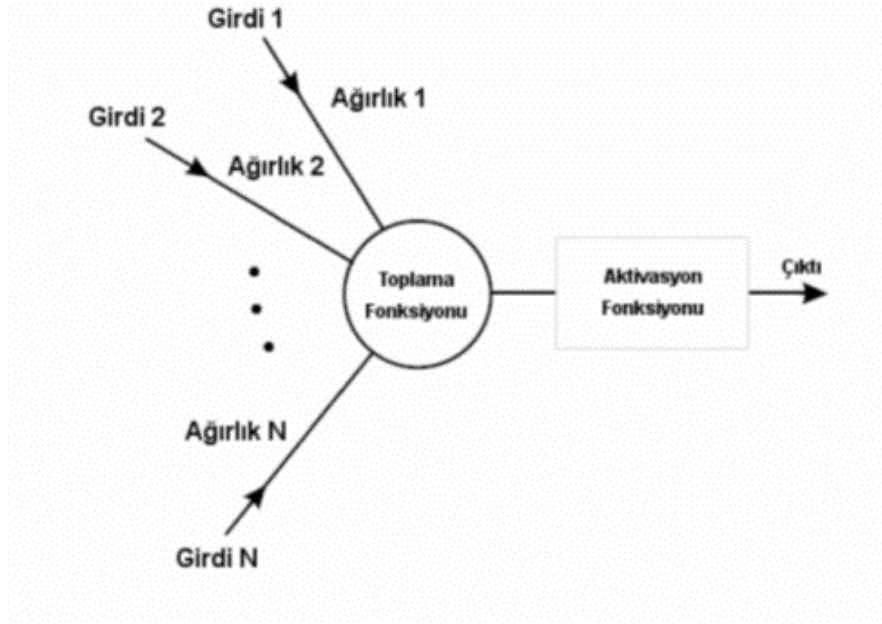
Sinir sisteminin temel elemanı nöronlar olduğu gibi, YSA'nın temel elemanı da yapay nöronlardır. Yapay nöronların 5 temel elemanı vardır: Girdiler, Ağırlıklar, Toplama Fonksiyonu, Aktivasyon Fonksiyonu ve Çıktılar.

Yapay sinir hücreleri bir araya gelerek YSA'yı oluştururlar. Bu sinir ağı yapısı oluşurken hücreler birbirine paralel 3 katman halinde dizilir: Girdi Katmanı, Ara Katman (Gizli Katman) ve Çıkış Katmanı (Şekil 1-7) .

⁷² Shmueli vd.a.g.e.s.100-103.

⁷³ Ercan Öztemel, "Yapay Sinir Ağları", 1. Baskı, Papatya Yayıncılık, İstanbul 2003, s.29.

⁷⁴ Zekai Şen, "Yapay Sinir Ağları İlkeleri", Su Vakfı Yayınları, İstanbul 2004, s.7.



Şekil 1-7: Yapay Sinir Hücresinin Yapısı⁷⁵

YSA'nın çalışma prensibi, ağa gösterilen girdilere göre, ağın onu çeşitli işlemlerden geçirerek çıktı seti oluşturmaktır. Bunun için ağa girdiler göstererek ağın doğru çıktı oluşturmaya (ağın eğitilmesi) gerçekleştirilmelidir. Ağa gösterilen girdiler, vektör halinde ağa sunulmalıdır. Ağ buna göre çıktı vektörü oluşturur.

YSA'nın, girdi vektörlerini, çıktı vektörlerine nasıl dönüştürdüğü bilinmemektedir. Bu nedenle YSA'nın çalışma şekli "kara kutu" olarak nitelendirilmektedir. Bu da ağa olan güveni azaltmaktadır.

Yapay sinir ağlarında ağın eğitilmesi öğretmenli, öğretmensiz, destekleyici ve karma yöntemler olmak üzere 4 tanedir.

Yapay sinir ağı mimarileri, sinirler arasındaki bağlantıların veya ağ içindeki işaretlerin akış yönlerine göre ileri beslemeli ve geri beslemeli ağlar olmak üzere ikiye ayrılmaktadır.

⁷⁵ Kevin Gurney, Computers and Symbols versus Nets and Neurons, UCL Draft Papers, U.K, 1996.s.4.

1.2.7. Destek Vektör Makineleri

Veri madenciliği sınıflama problemlerinde kullanılan bir diğer yöntem Destek Vektör Makinesi (DVM) (Support Vector Machines) adını taşımaktadır. Bu yöntem sınıflandırmayı doğrusal ya da doğrusal olmayan bir fonksiyon yardımıyla yerine getirir. Destek vektör makinesi yöntemi, veriyi birbirinden ayırmak için en uygun fonksiyon tahmin edilmesi esasına dayanır. Daha çok makine öğrenmesi yöntemleri arasında yer alan bu yöntem günümüzde veri madenciliği alanında da tercih edilmeye başlanmıştır⁷⁶.

1992 yılında Vlademir Vapnik, Bernhard Boser ve Isabelle Guyon tarafından geliştirilen DVM el yazısı tanıma, zaman serisi analizi, konuşma tanıma, meme kanseri teşhisi gibi bir çok alanda başarı ile uygulanmıştır.

Diğer sınıflandırma yöntemleri ile karşılaştırıldığında, eğitim süresinin oldukça uzun olmasına rağmen, yüksek güvenilirliği, ezbere öğrenmeye olan mukavemeti ve lineer olmayan sınıflandırmadaki başarı düzeyi ile DVM tercih edilen bir yöntem olmuştur⁷⁷.

⁷⁶ Silahtaroglu, a.g.e.s.83.

⁷⁷ Akpınar, a.g.e.s.268.

2. SAĞLIK HİZMETLERİ FİNANSMANI

Dünya sağlık örgütü (WHO) 1946 yılında sağlığı “yalnızca hastalık ve sakatlığın olmayışı değil, bedensel, ruhsal ve sosyal yönden tam bir iyilik hali” olarak tanımlamış ve sağlığı sosyal yaşam ile de ilişkilendirmiştir. Böylelikle sağlık için sadece “hasta” olmama hali değil, korunması ve geliştirilmesi gereken ve sadece bireyi değil toplumu merkeze alması gereken bir tanım yapılmıştır.

Çağdaş hukuk sistemlerinde kişinin sahip olduğu temel hak ve özgürlüklerin başında yaşama hakkı gelmektedir. Bu hak gerçekleşmediği takdirde diğer hakların gerçekleşme ve kullanılma şansı yoktur. Gerek uluslararası belgelerde gerekse birçok ülke anayasasında temel hakların odak noktasını yaşama, maddi ve manevi varlığını koruma ve geliştirme hakkı oluşturmaktadır. Yaşama hakkının ayrılmaz bir parçası da yaşamın sağlıklı bir şekilde sürdürülmesidir. Bu nedenle devlet bireylerin yaşama hakkını güvenceye kavuştururken, tıbbın ulaştığı son gelişmeleri de göz önüne alarak gerekli sağlık olanaklarını seferber etmek durumundadır. Çünkü sosyal devlet, bireylerin sağlıklı bir yaşam sürdürmeleri için ne gerekiyorsa onu yapmak veya yaptırmakla yükümlüdür⁷⁸⁷⁹.

Sağlığın gerek birey ve gerekse toplum açısından önem arz etmesi, devletin özellikle sağlık hizmetlerinin sunumu sürecinde aktif bir rol almasını zorunlu kılmaktadır. Bu bakımdan çağdaş devlet, sağlık hizmetleri sunumunun etkin şekilde gerçekleştirilmesini sağlayarak, özelde bireyin genelde ise toplumun refah seviyesini yükseltmeyi amaç edinmektedir⁸⁰. Dolayısıyla ülkelerin sağlık sistemleri bireylerin sağlık hakkını güvence altına almalı ve bunu yeterli, ulaşılabilir, sürdürülebilir ve

⁷⁸ Recep Makas , Türk Sosyal Güvenlik Sisteminde Norm ve **Standart Birliği**, **Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Hukuk Anabilim Dalı Özel Hukuk Bilim Dalı Doktora Tezi**, 2003 s.191.

⁷⁹ Gülbiye Yenimahalleli Yaşar, Sağlığın Finansmanı ve Türkiye için Sağlık Finansman Modeli Önerisi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü Çalışma Ekonomisi ve Endüstri İlişkileri ABD Doktora Tezi, Ankara, 2006, s. 1-2.

⁸⁰ Ferdi Çelikay ve Erdal Gümüş, "Türkiye'de Sağlık Hizmetleri ve Finansmanı", **Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi**, 2009, s.4.

kaliteli olarak sunulmalıdır. Ayrıca, sadece sağlığı geliştirmeye çalışmayıp, bireyleri hastalığın finansal riskine karşı da korumalıdır⁸¹.

Sağlık hizmetlerinin finansmanı ve sağlık harcamalarının ülkelerin ekonomik gelişmişlik düzeyinden aldığı pay her ülkede tartışılmakta ve sağlık politikalarını belirleyenlerin gündemlerindeki en önemli konular arasında yer almaktadır. Modern sağlık sistemlerinde sağlık hizmetlerinin finansmanında, hizmeti veren ile alan arasında doğrudan bir finansal akışın sağlanması yerine ağırlıklı olarak üçüncü taraf ödeyicinin finansman rolünü üstlendiği bir model benimsenmiştir. Bunun en önemli nedenlerinden biri, bireyin sağlık hizmetine ne zaman ihtiyacı olacağının belirsiz olması ve bu ihtiyaç ortaya çıktığında katlanması gereken maliyetlerin belirsizliğidir. Sağlık ve sağlık hizmetlerinin niteliği gereği çoğu zaman bu maliyet bireysel olarak katlanılabilecek boyutların çok üzerinde olup üçüncü tarafın müdahalesi olmadığı takdirde bireyin gerçek sağlık hizmeti ihtiyacından daha az hizmet alması ile sonuçlanabilir. Sağlık hizmetlerinin sunumu ve finansmanı, en basit haliyle hizmeti sunanlar ile finanse edenler arasında bir kaynak değişimi ya da transferi süreci olarak tanımlanabilir⁸².



Şekil 2-1: Sağlık Hizmetleri Üçgeni⁸³

⁸¹ WHO, “The World Health Report 2000. Health Systems: Improving Performance”, Geneva, 2000. s.8-24.

⁸² Mehtap Tatar, “Sağlık Hizmetlerinin Finansman Modelleri: Sosyal Sağlık Sigortasının Türkiye’de Gelişimi”, Sosyal Güvenlik Dergisi, 2011 (1), s.104-105.

⁸³ Elias Mossialos, Anna Dixon, “Funding health care: An introduction”, Berkshire: Open University Press, 2002, s.2.

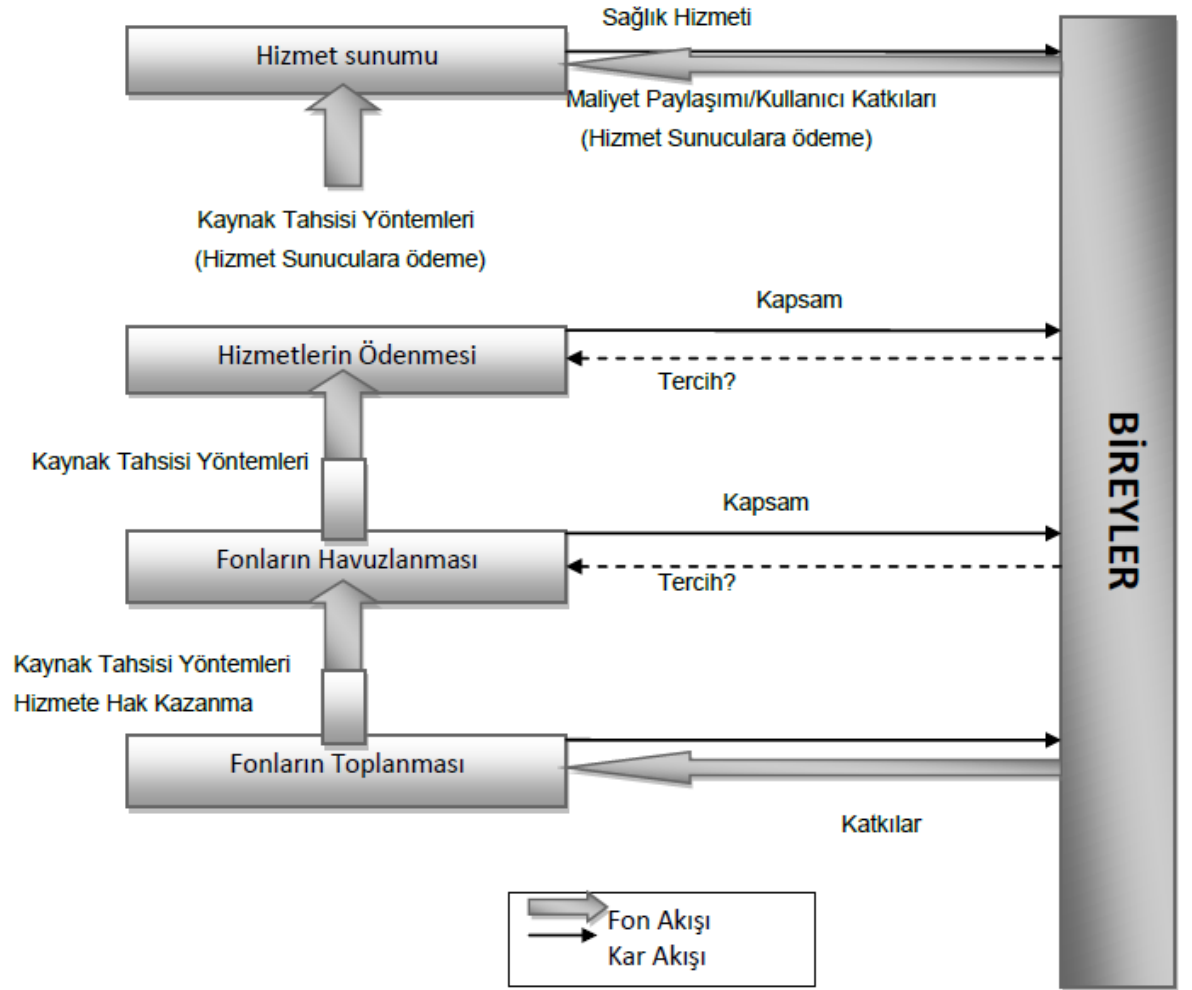
Sağlık Hizmetleri üçgeninde görüldüğü gibi bu süreçte, hizmet sunucuları, kendilerine başvuran bireylerin sağlık ihtiyaçlarını gidermeye yönelik hizmetleri sunarlar ve hastalar ya da ülkenin benimsediği finansman modeline göre üçüncü taraf ödeyici hizmetin karşılığını hizmet sunana aktarır. Bazı durumlarda kaynak akışı doğrudan doğruya hasta ile hizmet sunucusu arasında gerçekleşebilir. Örneğin, sunulan hizmetin maliyetinin bireysel olarak katlanılabilecek düzeyde olması, bireyin ya da sunulan sağlık hizmetinin üçüncü taraf ödeyicinin kapsamı içinde olmaması gibi. Bu veya benzeri durumlarda sağlık hizmetlerinin önemli bir bölümünde üçüncü taraf ödeyici önemli bir rol oynar. Ayrıca, üçüncü taraf ödeyicinin arada olduğu durumlarda bile, örneğin, katkı payları gibi düzenlemelerle hastalar halen finansmana katkıda bulunabilirler⁸⁴.

Sağlık hizmetlerinin sunumu ve finansmanı arasındaki ilişki temelde kaynakların transferi olarak açıklanabilir: hizmet sunucularının mal veya hizmet olarak sağlık hizmeti kaynaklarını bireylere transfer etmesi ve bireylerin veya üçüncü şahısların da finansal kaynakları hizmet sunuculara transfer etmesi. Bir mal veya hizmet alındığında en basit ödeme şekli doğrudan ödemedir. Tüketici (birinci taraf) bir mal veya hizmet aldığı anda hizmet sunana (ikinci taraf) ödemektedir. Sağlık sistemleri, hastalanmanın finansal riskine karşı güvence sağlamak için üçüncü taraf ödeyicileri geliştirmiştir. Özel ve kamu kuruluşu olabilen üçüncü taraf ödeyici, sağlık hizmetlerini finanse etmek için, güvence altına aldığı nüfustan doğrudan veya dolaylı olarak gelir toplamaktadır. Toplanan gelir ya doğrudan ödedikleri hizmet bedelini geri ödemek için hastalara (geri ödeme) ya da sundukları hizmetin karşılığı olarak hizmet sunuculara aktarılmaktadır. Sağlık sistemleri sadece hastalanmanın finansal riskine karşı güvence sağlamakla kalmamış, zamanla kişilerarası yeniden dağıtım aracı haline de gelmişlerdir⁸⁵.

⁸⁴ Tatar, a.g.e. s.105-106.

⁸⁵ Mossialos, a.g.e. s.2-3.

Bireylerin sağlık hizmetlerine ulaştığından emin olunması için sağlık hizmetleri finansmanının birbiriyle ilişkili üç temel işlevi vardır: gelir toplama, fon havuzlama ve hizmet sunucularına ödeme yapılması⁸⁶.



Şekil 2-2: Sağlık Hizmetleri Finansmanı İşlevleri ve Toplum İlişkisi⁸⁷

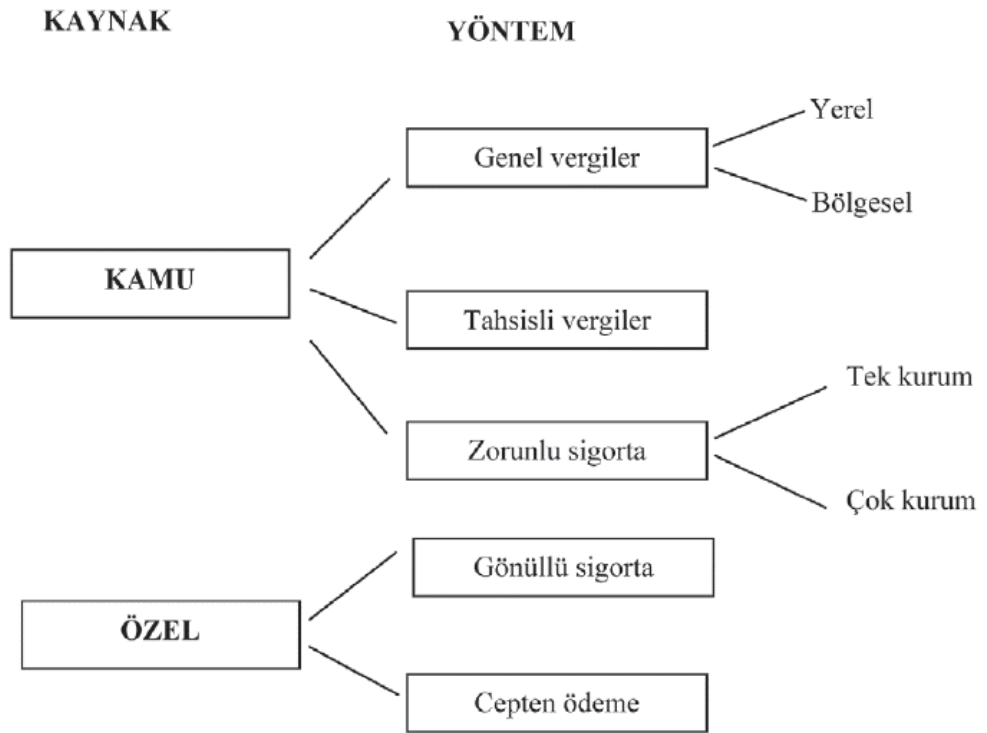
⁸⁶ Who, a.g.e., s.95.

⁸⁷ Joseph Kutzin, "A descriptive framework for country-level analysis of health care financing arrangements", Health Policy 56, 2001, s.174.

Ülkelerin gelir toplama, risk havuzlama ve hizmet sunucularına ödeme yapabilmesi için pek çok yol vardır. Risk havuzlama, finansal kaynakların toplanması ve yönetimidir. Böylece sağlık harcamalarının yaratacağı finansal risk, hizmetleri kimin kullandığına bakmaksızın, havuzdaki tüm bireyler tarafından paylaşılır⁸⁸.

2.1. Sağlık Hizmetlerinin Finansman Yöntemi

Sağlık hizmetlerinin finansman yöntemleri kamu ve özel olmak üzere iki temel grupta toplanmaktadır. Genellikle ülkeler finansmanlarını bu temel grupların değişik şekilleri ile gerçekleştirmektedirler.



Şekil 2-3: Sağlık Hizmetlerinin Alternatif Finansman Yöntemleri⁸⁹

⁸⁸ Pablo Gottret, George Schieber, "Health Financing Revisited", The World Bank, 2006, s. 4-5.

⁸⁹ Tatar, a.g.e.,s.111.

2.1.1. Özel Finansman Yöntemleri

2.1.1.1. Cepten Ödemeler

Cepten ödemeler, bireylerin sağlık hizmetleri kullanımı için doğrudan cepten ödeyerek yaptıkları tüm harcamalardır. Hastalık riski için herhangi bir finansal koruma sağlamamaktadır. Cepten ödemeler; doğrudan ödeme, kullanıcı katkısı ve resmi olmayan ödeme olarak üç başlık altında incelenebilir⁹⁰.

2.1.1.1.1. Doğrudan Ödeme

Doğrudan ödemeler; sosyal güvence kapsamı dışında kalan ve hizmet arzının sınırlı olduğu veya uzun bekleme süreleri nedeniyle hizmete ulaşımın aksadığı durumlarda yapılan ödemelerdir. Bu ödemeler genellikle dışçiler, güvence kapsamı dışındaki ilaçlar için eczaneler, özel hekim muayeneleri, özel hastaneler, özel laboratuvarlar veya test klinikleri gibi özel sektör hizmet sunucularına yapılmaktadır. Bazı ülkelerde özel sağlık hizmetleri için yapılan harcamalar vergilerle desteklenerek özel sektörden hizmet alımı teşvik edilmektedir⁹¹.

2.1.1.1.2. Kullanıcı Katkıları

1978 yılında yapılan Alma-Ata Konferansı'nda, temel sağlık hizmetlerini bütün vatandaşlara yaygınlaştırma yaklaşımı kabul görmüş ve en maliyet-etkin strateji olarak değerlendirilmiştir. Ancak, kaynakların yetersizliği nedeni ile anılan hedefe ulaşmak için ek kaynak yaratılması gerektiği tartışılmaya başlanmıştır⁹².

Sağlık harcamalarının genelde düşük olduğu durumlarda, sağlık hizmetleri için var olandan daha fazla kaynak yaratmak veya sağlık harcamalarının yüksek olduğu

⁹⁰ Mossialos, a.g.e. s. 22.

⁹¹ A.e.

⁹² Lee K, Goodman H. Global policy networks: the propagation of health care financing reform since the 1980s. In: Lee K, Buse K, Fustukian S, eds. Isted. Health Policy in a Globalising World. Cambridge, Cambridge University Press; 2002. s.99.

durumlarda, talebi azaltarak verimliliği arttırmak şeklinde iki farklı durumda kullanıcı katkıları kullanılabilir⁹³.

Bu yolla toplanacak ek gelir, gelir düzeyi düşük olan vatandaşların desteklenmesi veya finansman sistemindeki adaletsizliklerin azaltılması için kullanılabilir. Bununla birlikte kullanıcı katkıları; finansman yükünün topluma yayılmasını ve riskin paylaşılmasını engellemekte, dayanışmayı azaltmakta ve sonuçta da finansmanda adalet konusunu olumsuz etkilemektedir. Bunun nedeni; kullanıcı katkıları ile finansman yükünün gelir düzeyi düşük ve sağlık durumu kötü olan bireylere kaymasıdır⁹⁴.

2.1.1.1.3. Resmi Olmayan Ödemeler

Kayda geçmez ve birçok ülkede kanun dışıdır. Tedavi sonrası "teşekkür hediyeleri" veya tedavi öncesi doktorlara sunulan para veya hediyeler resmi olmayan ödemeler olarak isimlendirilmektedir. Temel nedenleri; hizmet sunucularını etkileme arzusudur⁹⁵.

Resmi olmayan ödemelerin temel nedenleri; kamu sistemindeki finansal kaynakların yetersizliği, özel sağlık hizmetlerinin yetersizliği, hizmet sunucularını etkileme arzusu ve kültürel geleneklerdir⁹⁶.

2.1.1.1.4. Özel Sağlık Sigortaları

Günümüzde özel sağlık sigortaları; işlevlerine, primlerin hesaplanma şekillerine, yardımların belirlenmesine ve sigorta şirketlerinin statüsüne göre farklılaşmaktadır. İşlevleri açısından iki tür özel sağlık sigortası bulunmaktadır. Birincisi, zorunlu sağlık sigortası yerine geçen "ikame edici özel sağlık sigortası", diğeri ise, zorunlu sağlık

⁹³ Andrew Creese, "User fees", British Medical Journal 1997, s.202.

⁹⁴ Hakan İstanbulluoğlu, Mahir Güleç, Recai Oğur, "Sağlık Hizmetlerinin Finansman Yöntemleri", Dirim Tıp Gazetesi, 2010, s.91.

⁹⁵ A.e.s.91.

⁹⁶ Mossialos, a.g.e. s.24.

sigortası ile birlikte görülen “tamamlayıcı özel sağlık sigortası”dır. Primlerin hesaplanma şekilleri açısından; bireysel, grupsal ve toplumsal riske göre belirlenen özel sağlık sigortaları mevcuttur. Statü açısından sigorta şirketleri, kar amaçlı olanlar ve olmayanlar olarak farklılaşmaktadır.

İkame edici özel sağlık sigortaları zorunlu kamu sigortalarının alternatifidir. Bu sigorta, zorunlu kamu sigortaları kapsamı dışında kalan bireyler ile zorunlu kamu sigortaları kapsamından çıkma hakkı olan sigortalılar tarafından tercih edilebilmektedir. Tamamlayıcı özel sağlık sigortası genellikle; zorunlu kamu sigortası dışında kalan hizmetlerin tamamı veya bir kısmı için, zorunlu sistem tarafından yalnızca bir miktarı ödenen hizmetler için ve sağlık hizmetlerine ulaşımı hızlandırmak için tercih edilmektedir⁹⁷.

Özel sağlık sigortasının teorideki en büyük avantajı; kısıtlı olan kamu kaynaklarının düşük gelirli, dezavantajlı ve özel sigortaya ulaşamayacak gruplar için harcanmasına imkân verebilmesidir. Bunu, gelir düzeyi göreceli olarak daha yüksek olan bireylerin özel sağlık sigortasını seçmelerine izin vererek sağlamak amaçlanmaktadır. Ayrıca, özel sağlık sigortalarının kâr güdüsü ile yenilikleri ve verimliliği desteklemesi, tüketicilerin tercihini arttırması söz konusudur⁹⁸.

Bu potansiyel avantajların gerçekleşmesi özel sağlık sigortalarının işleyiş ve denetimi ile yakından ilişkilidir. Düzenlemenin yokluğu veya yetersizliği aşağıda sıralanan bazı sorunların yaşanmasına neden olabilir⁹⁹:

- Primlerin miktarının doğru olarak belirlenememesi,
- Kişilerin sigorta şirketinden var olan hastalıklarını gizleyebilmeleri,
- Özel sigorta şirketlerinin **yüksek** sağlık riskli grupları sigortalamaktan kaçınmaları,
- Bireylerin, sigortalandıktan **sonra**, riskli davranışlarda bulunabilmeleri,
- Şirketlerin sigorta tekelleri yaratma eğilimleri,

⁹⁷ A.e.19-20.

⁹⁸ Alan Maynard, Anna Dixon., “Private health insurance and medical savings accounts: theory and experience”, In “Funding health care: options for Europe”, Open University Press, 2002, s.111.

⁹⁹ İstanbulluoğlu vd, a.e.s. 91-92.

- Adalet ve dayanışma kavramlarının zedelenebilmesi.

2.1.1.1.5. Tıbbi Tasarruf Hesapları

Tıbbi tasarruf hesapları; bireyler, hane halkları ve firmaların gelecekte karşılaşılabilecekleri sağlık risklerine karşı, gönüllü veya zorunlu olarak, kendilerine ait banka hesaplarına, önceden para yatırımları ve bu parayı yalnızca sağlık harcamaları için kullanmaları esaslarına dayanmaktadır¹⁰⁰.

2.1.2. Kamusal Nitelikli Finansman Kaynakları

Kamusal nitelikli finansman kaynakları, 19. yüzyılın sonunda sosyal sigortacılık yöntemi ile başlamış ve 20. yüzyılın ilk aylarında da vergilerle finansman ile devam etmiştir.

2.1.2.1. Sosyal Sağlık Sigortacılığı

Zorunlu sağlık sigortacılığı yasal olarak ilk kez 1883 yılında, Otto Van Bismarck önderliğinde Almanya’da başlatılmıştır. Bu uygulama belirli endüstrilerde, belli bir ücret düzeyi altında çalışanların hastalık fonuna üye olma zorunluluğunu ve fonun finansmanının işçi ve işverenlerden alınan zorunlu katkılarla gerçekleştirilmesini gerektirmektedir¹⁰¹.

Sağlık hizmetlerinin sosyal sigorta ile finansmanında hem çalışanın hem de işverenin, belirlenmiş kurallar çerçevesinde bireyin geliri esas alınarak bir prim ödemesi, bu primlerin bir havuzda toplanması ve daha sonra sağlık ihtiyacı olanların bu havuzda biriken kaynakları kullanarak sağlık hizmetlerini alması söz konusudur. Sosyal sağlık sigortasında primlerin miktarı ve toplanma şekli, gelirlerin toplanması ve harcamaların yapılmasından sorumlu kurumların sayısı ve örgütlenme şekli ve sağlık hizmetlerinin

¹⁰⁰ Anna Dixon, “Are Medical Savings Accounts a Viable Option for Funding Health Care”, Croatian Medical Journal, Vol.43, n.4, s.409.

¹⁰¹ Aviva Ron, Brian Abel-Smith, Giovanni Tamburi, “Health Insurance in Developing Countries: The Social Security Approach”, Ceneva: ILO; 1990. s.8.

sunum şekli ülkeden ülkeye farklılık göstermekle birlikte bazı temel özellikleri bulunmaktadır¹⁰²:

- Sigortalanan bireyler gelirleri temel alınarak düzenli olarak katkıda bulunurlar
- Sistem genellikle bağımsız ya da yarı özerk kurumlar tarafından yönetilir.
- Sosyal sağlık sigortası toplumun tamamı ya da büyük bir bölümü için zorunludur. Bazı ülkelerde bu zorunluluk belirli bir gelir düzeyinin üzerindeki bireyler için kalkabilmekte (Hollanda) ya da bireyler belirli bir gelir grubunun üzerindeyse sosyal sigorta ile özel sigorta arasında tercih yapabilmektedir (Almanya).

2.1.2.2. Vergilerle Finansman

Günümüzde genel vergi gelirleri neredeyse bütün ülkelerde sağlık hizmetlerinin finansmanı için kullanılmaktadır. Bu gelirler özellikle gelişmekte olan ülkeler için önemli bir finansman kaynağıdır. Bugün 191 DSÖ üyesinin 106'sı ağırlıklı olarak vergilerle finanse edilmektedir. Sağlık hizmetlerinin vergilerle finansmanı vergi kaynakları, toplanma düzeyleri ve vergi türleri açısından farklılaşarak heterojen bir yapı göstermektedir. Vergi kaynakları dolaysız ve dolaylı vergiler olarak ikiye ayrılmaktadır. Dolaysız vergiler, faktör gelirleri (rant, ücret, faiz, kâr) ve servet üzerinden alınmaktadır. Dolaylı vergiler, mallar ve işlemler üzerinden (satış vergisi, katma değer vergisi, özel tüketim vergisi, ithalat ve ihracat vergileri gibi) alınan vergilerdir¹⁰³.

2.1.2.3. Fonların Havuzlanması

Sağlık finansmanında tek bir havuzun bulunması, hem riskin hem de gelirin ülkedeki bütün üyeler arasında yeniden dağıtımını en üst düzeyde sağlamaktadır. Büyük fonlar

¹⁰² Tatar, a.e.s. 116.

¹⁰³ Yaşar, a.g.e. 41-43, 170.

yoluyla katkı oranları azaltılmakta, dayanışma daha büyük ölçüde gerçekleşmektedir¹⁰⁴.

2.2. Hizmet Sunucularına Ödeme

Hizmet sunucularına ödeme konusunda hizmet başına ödeme, gün başına ödeme, vaka başına ödeme, kişi başına ödeme, bütçe ve ücret ödeme gibi farklı yöntemler bulunmaktadır (WHO, 2009). Bu başlık sağlık hizmetlerine geri ödeme yöntemleri başlığı altında incelenecektir.

2.3. Sağlık Hizmetlerinde Geri Ödeme Yöntemleri

Sağlık hizmetlerinde geri ödeme yöntemleri temel olarak, sağlık hizmetlerinin maliyetlerinin karşılaştırılması, üçüncü taraf ödeyici ve sağlık sigortası kavramlarına dayanır.

Sağlık hizmetlerinde maliyetlerin karşılanması, üçüncü tarafın belirlediği, geri ödeme ile gerçekleşir. Üçüncü taraf ödeyici, en basit tanımıyla hizmet sunucularına yapılacak ödemeleri belirleyen ve finansman temin eden kamuya ait ya da özel organizasyonlardır. Buna paralel biçimde sağlık hizmetleri için yapılacak geri ödemelerde de hastaların sigortalı olup olmadığı temel belirleyici faktörler arasındadır. Bu durumun temelinde esasen sigorta kavramının tanımı yatmaktadır. Sigorta, kişilerin karşı karşıya kalabilecekleri kayıp riskinin başka bir kesim tarafından (sigorta kuruluşu) üstlenilerek azaltılmasını amaçlayan sistemdir¹⁰⁵.

¹⁰⁴ Kutzin J. "Towards Universal Health Care Coverage: A Goal-oriented Framework for Policy Analysis", Geneva: The World Bank; 2004. s.9.

¹⁰⁵ Çağdaş Erkan Akyürek, "Sağlıkta Bir Geri Ödeme Yöntemi Olarak Global Bütçe ve Türkiye.", *Sosyal Güvenlik Dergisi*, 2002, s.126-127.

2.3.1. Harcama Kalemli Bütçeler ve Genel (Global) Bütçeler

Global bütçe kavramını “sunulan hizmetlerin ücret ve kalitesini tayin eden genel bir harcama sınırı” şeklinde tanımlamak mümkündür. Global bütçe pek çok farklı şekilde mümkün olabilmektedir ve tüm harcama programlarına veya bu programların belirli kısımlarına uygulanması mümkündür. Aynı şekilde ödeyici tarafından kapsanan nüfusun, bir sağlık kuruluşunca sunulan hizmetlerin ya da bu hizmetleri sunan kuruluşların tamamına ya da belirli kısımlarına da uygulanabilir¹⁰⁶.

2.3.2. Kişi Başı Geri Ödeme:

Kişi başına ödeme yöntemi, sağlık hizmetlerinde üçüncü taraf ödeyicilerin hizmet sunucularına belirli bir zaman dilimi içerisinde, kişi başına belirlenmiş sabit ücretler üzerinden ödeme yaptığı bir yöntemdir. Kişinin belli bir süre zarfında kullanacağı tüm hizmetleri karşılamak üzere bir ödeme miktarının tayin edilmesi esasına dayanmaktadır¹⁰⁷.

Bu yöntemde her bir hastaya sunulan hizmetlerin hacminin ya da yoğunluğunun ödeme üzerinde herhangi bir etkisi yoktur. Daha fazla hizmet sunumu ödemeyi artırmadığı gibi, daha az hizmet sunulmasının da ödemeyi azaltıcı bir etkisi yoktur. Bir hizmet sunucusu üçüncü taraf ödeyici ile belirli bir gruba kişi başına ödeme esasıyla sağlık hizmeti sunma konusunda mutabakata varırsa, hizmet sunucusu o grupta yer alan her bir kişi için sağlık hizmetinden faydalanıp faydalanmadığına bakılmaksızın ödeme alır¹⁰⁸.

Özellikle aile hekimliği gibi uygulaması bölge esasında yürütülen hizmetlerin ödemesinde sıklıkla kullanılan bir ödeme yöntemidir. Hekim, önceden belirlenmiş tüm hizmetleri sunmayı kabul ettiği kişileri kendi listesine kaydeder ve kaydettiği her bir

¹⁰⁶ A.e.s.139.

¹⁰⁷ Yusuf Çelik, “Sağlık Ekonomisi”, Siyasal Kitabevi, 2011, s.220.

¹⁰⁸ Akyürek, a.e.s.132.

kiři için (sıklıkla aylık) sabit bir ücret alır. Yani kendisine kayıtlı bir kiřide gebelik takibi de yapsa apse drenajı da yapsa o kiři için sabit ücreti alır¹⁰⁹.

Kiři başına ödeme saęlık hizmeti sunucularına ödeme yapanlara göre daha yüksek finansal risk getirmektedir. Bu durum saęlık hizmeti sunucuları üzerinde maliyet kontrolü ve saęlık hizmetlerinin sunulmasında maliyet-etkililięi saęlama konusunda baskı yaratmaktadır. Kiři başına ödeme yöntemi saęlık hizmetleri için ödeme yapan kurum veya mekanizmaların maliyetlerini daha iyi analiz, kontrol ve tahmin etmelerine olanak saęlamaktadır¹¹⁰.

2.3.3. Hizmet Başına Geri Ödeme

Hizmet başına geri ödemedede, hizmet sunucusuna ürettięi her birim hizmet için ödeme yapılır. Geri ödemelerin tayininde sıkça başvurulanan bir yöntemdir. Sunulan her bir kalem hizmet için belirlenmiř bir fiyat söz konusudur. Saęlık hizmeti sunucuları bu belirlenmiř fiyatlar üzerinden sunmuř olduęu hizmetleri fatura eder. Bu faturalardan hareketle geri ödeyici konumundaki saęlık sigortası kuruluřu ödenecek hizmetler kapsamında yer alan hizmetler için hizmet sunucusuna gerekli ödemeleri yapar. Saęlık hizmetlerinde geri ödemelere esas olan miktarlar hizmet sunucusunun çıkarmıř olduęu faturalarda yer alanlardır. Geri ödeyici konumundaki kuruluř gerekli incelemeleri ve lüzumlu gördüęü kesintileri yaptıktan sonra geri ödemeyi gerçekleştirir.

Hizmet başına ödeme prensibi ile geri ödeme yapan bir saęlık sigortası bünyesinde yer alan kiřilerin önemli bir hareket serbestlięi söz konusu dur. Hangi hekime görünecekleri ve hangi kořullar altında tedavi alacakları konusunda mensubu oldukları sigorta planı onlara karar süreçlerinin hemen hemen hepsinde yer alma imkanı yaratmaktadır.

¹⁰⁹ Dilaver Tengilimoęlu, Oęuz Iřık, ve Mahmut Akbolat, "Saęlık iřletmeleri yönetimi" Nobel Akademik Yayıncılık, 2009, s.239.

¹¹⁰ Howard Barnum, Joseph Kutzin, ve Helen Saxenian. "Incentives and provider payment methods." The International Journal of Health Planning and Management 10.1, 1995, s.6-7.

Hasta açısından hizmet başına ödeme yönteminin dezavantajı, bu ödeme yönteminin diğerlerine kıyasla daha fazla kesinti veya tamamlayıcı ödeme içermesidir. Sağlık sigortası kuruluşları açısından bakıldığında yöntemin en belli başlı olumsuz yönü belirsizliktir. Hizmet sunucularına, sunmuş oldukları sağlık hizmeti karşılığında yapılacak ödemelerin yarattığı maliyet bilinmemektedir. Çünkü hastalara hangi sağlık hizmetlerinin sunulacağından önceden bilinmesi gibi bir durum söz konusu değildir. Bunlara ek olarak, hizmet sunucularının her bir hizmet için ücretleri yükseltmeleri, hastaların beklenenden daha fazla hizmet almaları ve ucuz olan hizmetlerin yerine göreceli olarak daha pahalılarının tercih edilmesi halinde maliyetler artacaktır¹¹¹.

2.3.4. Gün Başına Ödeme

Gün başına ödeme; ödemenin hastanın yattığı gün sayısına göre yapılmasına dayanan, kolay ve yönetim maliyeti düşük bir yöntemdir. Ancak bu yöntemde de hizmet sunucular daha fazla kar elde etmek için hastanın gereğinden daha uzun süre yatırılmasını veya hastanın gerekmediği halde hastaneye yatırılmasını sağlayabilmektedirler. Gereksiz yere uzun süre yatırmayı engellemek için, belli bir süreden sonra günlük ödeme bedeli azaltılabilmektedir¹¹².

Günlük ödeme hastaya verilen günlük hizmetlerin hepsini kapsamaktadır. Günlük ödemeler hastaya verilen tıbbi tedavi, ilaçlar, bandajlar, protezler, konaklama gibi unsurlara ait harcamaları kapsamaktadır¹¹³. Günlük ödeme ücretleri sunucu (hastane) ile sağlık fonu arasındaki müzakerelerden sonra belirlenmektedir. Teorik olarak bir üniversite hastanesi ile bir kırsal alandaki küçük hastaneye yapılacak günlük ödeme arasında fark olması beklenmektedir¹¹⁴ (Normand ve Weber, 1994: 67).

¹¹¹ Akyürek, a.e., s.131.

¹¹² WHO, "Reaching Universal Coverage via Social Health Insurance: Key Design Features in the Transition Period", by Guy Carrin and Chris James, Discussion Paper, n.2, Department "Health System Financing, Expenditure and Resource Allocation (FER), Cluster "Evidence and Information for Policy (EIP)", Geneva, 2004, s.37-38.

¹¹³ Matilde Pinto ve Bernt Anderson, "Paying Health Care Providers in the Caribbean", Pan American Health Organization, Regional Office of WHO, USA, 2001, s.12-13.

¹¹⁴ Charles Normand ve Axel Weber, Sociat Health Insurance: "A Guidebook for Planning" WHO, 1994, s.67.

2.3.5. Vakaya Dayalı Ödeme

Vaka başına ödeme; hem ayakta hem de yatarak tedavilerde, hastaya konulacak teşhise göre yapılan ödemedir. En önemli şekli, hastanın teşhis grubuna göre, belirlenen sabit bir fiyat üzerinden, o teşhis doğrultusunda yapılan her hizmetin kapsama dahil olduğu Teşhise Dayalı Grup (Örnek: TİG) ödemesidir. Bu yöntemde teşhis gruplarının çok net bir şekilde belirlenmesi önemlidir. Maliyetleri kontrol etmek açısından önemli olan bu yöntem, gereksiz hizmet sunumunu arttırmaktadır. Ayrıca hizmet sunucuların daha ciddi ve dolayısıyla daha pahalı olan vakaları tercih edip, daha karmaşık, teşhis ve tedavisi daha ucuz olan hastaları başka yerlere göndermeleri mümkündür. Bu nedenle hastalığın doğru teşhis edilip edilmediği ve karmaşık vakaların diğer hizmet sunucularına gönderilip gönderilmediğini denetlemek çok önemlidir. Bu denetimler de oldukça maliyetlidir¹¹⁵.

Aynı teşhis (grup) içerisinde sınıflanan hastalar, benzer teşhis ve tedavilere, kaynak tüketimine ve kalış sürelerine sahip olmaktadır. Her bir TİG, ağırlık adı verilen bir ödeme oranına sahiptir. Ağırlıklar kendi içerisinde görecelidirler. Daha yüksek düzeydeki ağırlıklar, bakım ve tedavi için daha fazla kaynağa gereksinimi olan hastaların yer aldığı gruplarla ilgilidir. Daha yüksek düzeyde kaynak tüketimi, hastalığın şiddetine ya da pahalı ekipmanla tıbbi uygulamalar gibi bakım ve tedavi için ihtiyaç duyulan hizmetlerin türlerine bağlı olarak sunulacak hizmetlerin yoğunluğu ile alakalıdır. Daha yüksek ağırlıklar, daha yüksek ödemelere dönüşmektedir¹¹⁶.

Aşağıdaki tablo ödeme modellerinin karşılaştırılmasını göstermektedir¹¹⁷:

¹¹⁵ Who 2004, a.e.s.38.

¹¹⁶ Anne B. Casto ve Elizabeth Layman, “**Principles of Healthcare Reimbursement**”, Michigan: American Health Information Management Association, 2006, s.9.

¹¹⁷ Hugh Waters, Peter Hussey; “**Pricing Health Services for Purchasers: A Review of Methods and Experiences**”, Washington, The World Bank, 2004, s.3.

Tablo 2-1: Sağlık Sunucuları Ödeme Sistemleri Karşılaştırması

Ödeme Mekanizmaları	Sağlıkla İlgili Problemlerin Önlenmesi	Yasal Beklentilerin Karşılanması	Hizmet Tahsisi	Maliyetler
Kalem Bütçeleme	+/-	+/-	--	+++
Global Bütçe	++	+/-	--	+++
Kişi Başına Ödeme	+++	++	--	+++
Tanıya Dayalı Ödeme	+/-	++	++	++
Hizmet Başı Ödeme	+/-	+++	+++	---
+++ = Etkilerin çok olumlu olması ++ = Bazı olumlu etkilerin olması +/- = Az etki yada etkinin olmaması -- = Bazı olumsuz etkiler olması --- = Etkilerin çok olumsuz olması				

2.4. Türkiye’de Sağlık Hizmetlerinin Finansmanı

2.4.1. Özel Nitelikli Sağlık Finansmanı

2.4.1.1. Türkiye’de Cepten Ödemeler

Türkiye 2003 yılından beri sağlıkta bir takım reformlara başlamış ve bunlara da “Sağlıkta Dönüşüm” adını vermiştir. Bu proje kapsamında, Yeşil Kartlıların ayakta tedavi hizmetleri kapsam içine alınmıştır. SSK hastaneleri Sağlık Bakanlığına devredilmiş ve SSK’lıların kamu sağlık tesislerine ulaşımı sağlanmıştır. Ayrıca, SSK ve Yeşil Kartlıların tüm serbest eczanelerden ilaç teminine izin verilmiştir. Sosyal Güvenlik Kurumu kurularak SSK, BAĞ-KUR ve ES aynı çatı altında toplanmıştır. Aynı zamanda, sağlık sigortası alanında yaşanan karmaşanın önüne geçilmesi, uygulamada standartların ve norm birliğinin oluşturulması amacıyla Ekim 2008’de Genel Sağlık Sigortası sistemine geçilmiştir. Bu düzenlemelerin sonucunda TÜİK Yaşam Memnuniyet Araştırması sonuçlarına göre; ilaç ve tedavi masraflarını cepten

harcamalarla karşılayanların oranı 2003’de %32,1 iken, 2007’de %16,5’a gerilemiştir. 1999 yılında hane halkı sağlık harcamalarının GSYH’deki payı %1,4 olarak gerçekleşirken, 2006 yılında %1,1 seviyesine gerilemiştir¹¹⁸.

2.4.1.2. Türkiye’de Özel Sağlık Sigortacılığı

Ülkemizde 1980’li yılların başından itibaren, özel sağlık sigortalarına olan talep artmıştır. Ancak özel sağlık sigortacılığı sektörü 2000-2001 kriz döneminde durağanlık dönemine girmiştir. Söz konusu durağanlığı ortadan kaldırmak ve özel sağlık sigortacılığını teşvik etmek için özel sağlık sigortalarından alınan Banka ve Sigorta Muamele Vergisi 4697 sayılı Kanun ile 07.10.2001 tarihinde yürürlükten kaldırılmıştır. Bu durum sigortalılar tarafından özel sağlık sigortaları için ödenen primlerin %5 azalması anlamına gelmiştir. Ülkemizde 1997-2002 yılları arasında sigorta şirketi sayısının azalmasına karşılık sigortalı sayısında az da olsa artış görülmüştür¹¹⁹.

Türkiye’de özel sağlık sigortacılığı; 1990’lı yıllarda 12 milyon ABD Doları prim üretimi ile yaygınlaşmaya başlamış, 2000 yılına kadar büyüyen bir kapasite ile devam etmiş, 2000/2001 ekonomik krizlerinden etkilenecek küçülmüş ancak 2002 yılından itibaren tekrar artışa geçerek, 2005 yılı sonunda 596 milyon ABD Doları olan prim üretimine ulaşmıştır. Sigortalı sayısı, 2000-2005 yılları arasında % 41’lik artışla 974.251’e ulaşmıştır. 2005 yılı sağlık sigortası priminin, % 70’i ilk 5, % 43’ü ilk 2 şirket tarafından üretilmiştir¹²⁰.

2.4.2. Türkiye’de Kamusal Nitelikli Sağlık Finansmanı

Bağımlı çalışanlar için, sigortacılık tekniği öngörülerek 1950 yılında yürürlüğe sokulan kanun, ülkemizde kamusal nitelikli sağlık finansmanı olarak

¹¹⁸ Seher Nur Sülkü, “Türkiye’de Sağlık Dönüşüm Programı Öncesi ve Sonrasında Sağlık Hizmetlerinin Sunumu ve Finansmanının Sunumu Finansmanı ve Sağlık Harcamaları”, T.C. Maliye Bakanlığı Strateji Geliştirme Başkanlığı, 2011, Ankara, s.94-95.

¹¹⁹ İstanbulluoğlu vd.a.e.s.94.

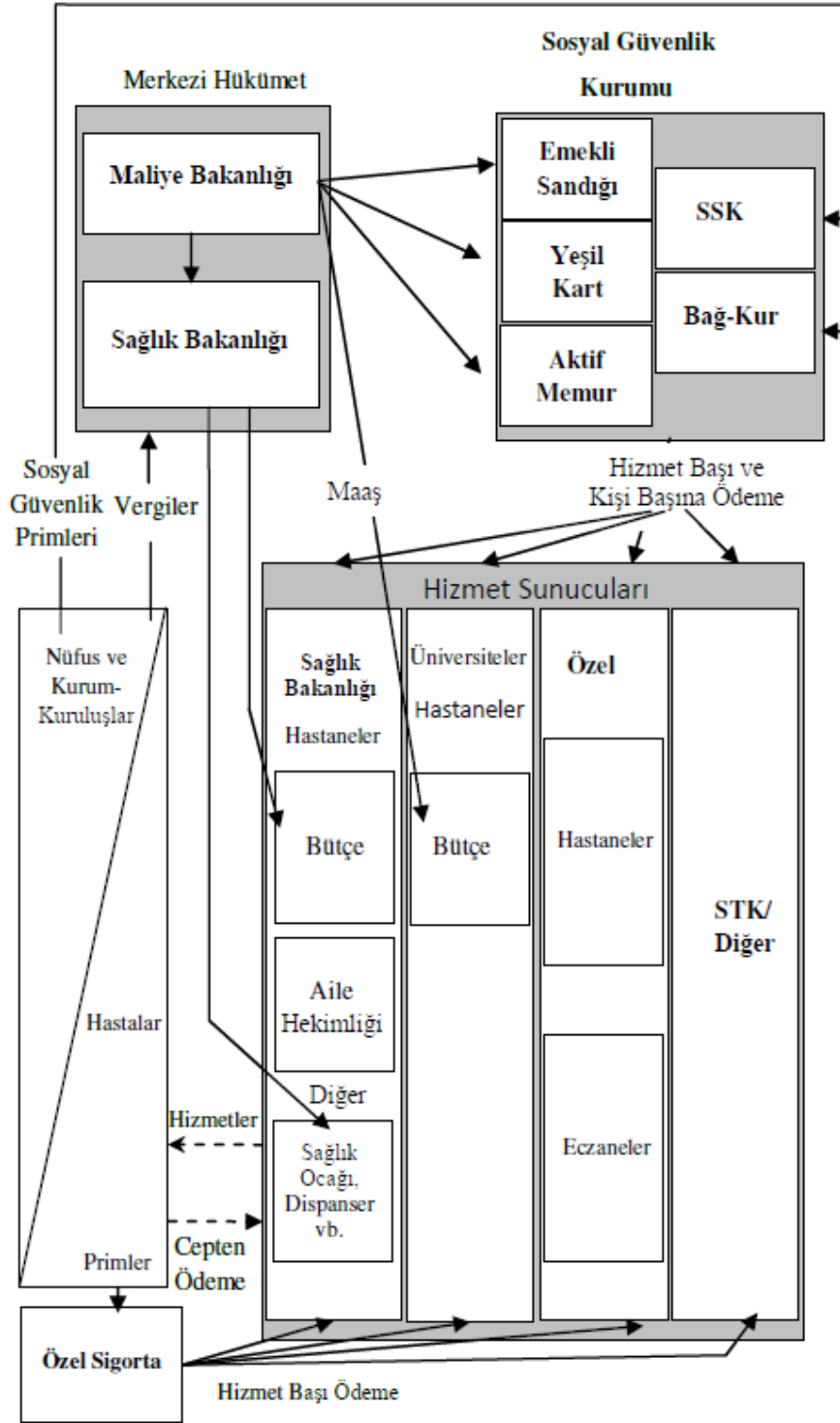
¹²⁰ Haluk Özseri, “Genel sağlık sigortası özel sağlık sigortacılığı ilişkisi”, SD Dergisi, Aralık, Ocak, Şubat, 2006-2007-1, <http://www.sdplatform.com/Dergi/9/Genel-saglik-sigortasi-ozel-saglik-sigortaciligi-iliskisi.aspx>

değerlendirilebilecek ilk uygulamadır. Bunu 1965 yılında sağlık harcamaları kurumlarınca karşılanmak üzere kamu çalışanlarının, 1971 yılında emekli memurların, 1985 yılında bağımsız çalışanların güvence altına alınması ve 1992 yılında çıkarılan Yeşil Kart Kanunu izlemiştir. 2006 yılında kabul edilen Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu (SSGSS), tek bir finansman kurumu ile bütün vatandaşların sosyal sigortacılık yöntemiyle sağlık güvencesine kavuşturulmasını hedeflemektedir¹²¹.

Genel sağlık sigortası sisteminin uygulanmaya başlanması ile birlikte sağlık hizmetlerinin finansmanında Sosyal Güvenlik Kurumu öncelikli bir konuma gelmiş, özellikle devlet bütçesinden Sağlık Bakanlığı ya da Maliye Bakanlığı aracılığıyla direkt olarak sağlık hizmetleri finansmanının sağlanması yerine, Sosyal Güvenlik Kurumu aracılığıyla dolaylı bir finansman yöntemi benimsenmiştir. Temel sağlık hizmetleri, salgın hastalıklar ile mücadele, ana ve çocuk sağlığının korunması ile aile planlaması hizmetlerinin sağlanması, gerekli aşı, serum ve benzeri ürünlerin tedariki ve sunumu gibi görevler Sağlık Bakanlığı tarafından yerine getirilmektedir. Bu bağlamda Sağlık Bakanlığı, her yıl Maliye Bakanlığı Merkezi Yönetim Bütçe Kanunu'na göre aldığı ödeneği kendine bağlı hastanelere tahsis etmekte ya da bizzat alt birimleri vasıtasıyla hizmet sunumunun gerçekleştirilmesini sağlamaktadır¹²².

¹²¹ İstanbulluoğlu, a.e.s.94-95.

¹²² Çelikay, vd.a.g.e., s 207.



Şekil 2-4: Türk Sağlık Sistemindeki Temel Mali Akışlar, 2008 (GSS Yasasının Kabulünden Sonra)¹²³

2.5. Teşhis İlişkili Gruplar ile Klinik Kodlama

Hastanelerde, klinik veriyi oluşturan hekimler, bu verileri kodlara çevirmekten sorumlu klinik kodlayıcılar ve TİG'den sorumlu yöneticiler, TİG'den sorumlu kişilerdir. Bu kişilerden beklenen görevler şu şekilde özetlenebilir:

- Klinik kodlayıcılar, genellikle sağlık alanında eğitim almış, hastalık süreçlerini bilen, tedavi süreçlerine hakim, tıbbi kayıtların içeriğini çözümleyebilen klinik kodlama eğitimi almış kişilerdir.
- Tam ve doğru tıbbi bilgilerin dökümanite ve klinik kodlamacı ile ortak çalışması hekimlerden beklenen görevlerdir.
- TİG'den sorumlu hastane yöneticileri ise klinik kodlamacıların görevlerini eksiksiz yapmalarını sağlayacak düzenlemeleri yapmaktan sorumludurlar.

Klinik kodlama süreci hastanın taburcu olması ve hasta dosyasının klinik kodlama birimine gelmesiyle başlar ve iki aşamada gerçekleşir¹²⁴:

Birinci aşamada tıbbi kayıtlar hangi durumların kodlanacağını belirlemek için analiz edilir. Kayıtlardan hastanın yaşı, cinsiyeti, taburcu tarihi vb. bilgileri kontrol edilir. Hastanın epikrizi (çıkış özeti), ameliyat raporları, gözlem notları, radyoloji raporları, laboratuvar sonuçları, konsültasyon raporları, patoloji raporu vb. temel kayıtları incelenerek hastaneye yatış nedeni, hastanın öyküsü, fiziksel muayene bulguları, tanıları, hastaya uygulanan işlemler, komorbiditeleri (eş tanıları), komplikasyonları¹²⁵ belirlenir. ICD-10-AM kodlama standartları çerçevesinde hangi durumun/durumların kodlanacağına karar verilir. Ana tanı¹²⁶ ve ek tanıların¹²⁷ belirlenmesi için bu son

¹²⁴ Recep Akdağ, TİG Uygulama Rehberi, Ankara, T.C. Sağlık Bakanlığı Tedavi Hizmetleri Genel Müdürlüğü Performans Yönetimi Kalite Geliştirme Daire Başkanlığı, 2011, s.12-14.

¹²⁵ Hastalık sırasında ortaya çıkan ve hastalığın temel özellikleriyle ilişkili olmayan her türlü olumsuz sağlık olayı veya süreci (TDK)

¹²⁶ Ana tanı, hastanın hastanede olmasına neden olan, inceleme sonrasında konulan hastalık tanısıdır.

¹²⁷ Hasta bir semptom ile yatırılır, yatışı sırasında semptomun esas nedeni bilinir ve bu durum tedavi edilirse ana tanı semptom olur semptomun nedeni ise ek tanı olarak kodlanır.

derece önemlidir. Ana tanı ve ek tanı ayırımı doğru kodlamanın yapılması ve kodların uygun TİG'lere yerleşmesi açısından son derece önemlidir.

İkinci aşama kodlanacak durumların sistemden bulunarak kodların atanmasıdır. Bu aşamada şu adımlar izlenir:

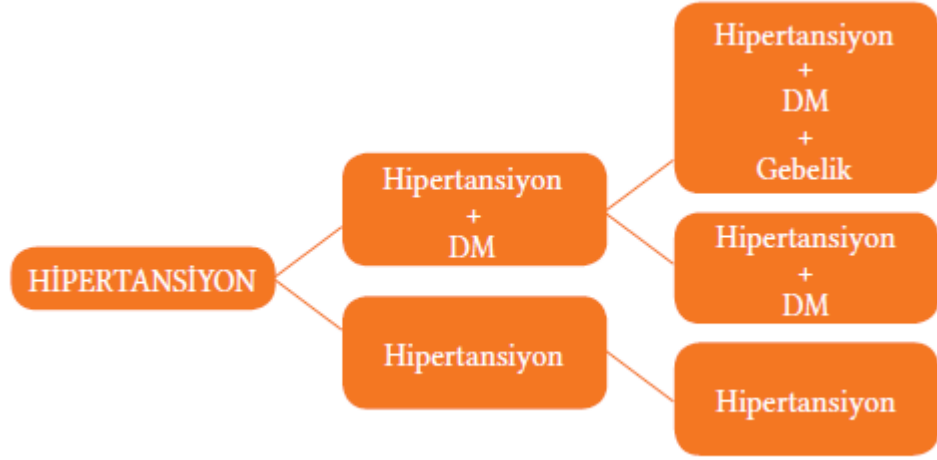
- Tanı ve işlemler için alfabetik indeksin uygun kısmına başvurulur.
- Tüm açıklayıcı, yönlendirici notlar okunur ve en uygun kod/kodlar belirlenir.
- Kontrol etmek amacıyla tablo listesinden kod/ kodlar kontrol edilir.
- Tablo listesinde bulunan açıklayıcı notlar okunur, kod/kodlar atanır.
- Kodlayıcı, atanan bu kodların hastanın tıbbi durumunu tam olarak ifade ettiğinden emin olmalıdır.

Bu iki aşama sonucunda hastalar, hastalıklarına göre 16.013 tanı, 6055 işlem, 2408 morfoloji kodları içerisinde uygun olanları ile kodlanmakta ve grouper programı ile 25 Major Tanı Sınıfına ve 665 TİG'e atanmaktadır.

2.5.1. Sınıflandırılma

Her hasta benzersizdir ve her hastanın bir dizi farklı klinik tanısı, risk faktörü ve aile durumu vardır. Bu çeşitliliğin anlamlandırılması için, bunları benzer koşullarla tanımlayan gruplara ayrılması gereklidir. Hastayı ana tanıya göre gruplandırmak ve daha sonra ikincil hastalıkların varlığına ya da yokluğuna göre alt gruplara ayırmak gerekmektedir¹²⁸.

¹²⁸ İrfan Şencan, Muzaffer Şeker ve Mehmet Demir, Teşhis İlişkili Gruplar İleri Klinik Kodlama Standartları, Sağlık Hizmetleri Genel Müdürlüğü, 2013, s.2.



Şekil 2-5: TİG Hipertansiyon Alt Grupları

Sınıflandırma, benzer hastalık veya işlemlerin aynı grupta toplanmasıdır. Hastalıkları, etkilenen vücut kısmına (topografi), nedene (etyoloji), dokudaki patolojik değişikliğin tipine (morfoloji), ya da sonuçta ortaya çıkan fonksiyonel anormalliğe göre sınıflamak mümkündür.

2.5.2. Kodlanma¹²⁹

Kodlama, hastalıkların, yaralanmaların ve prosedürlerin nümerik veya alfanümerik yapıda ifade edilmesidir. Nümerik yapı işlemlerde olduğu gibi toplam yedi rakamdan oluşan işlemlerin kodlanmasında kullanılmaktadır. Örnek: 92514-19. Alfanümerik yapı ise tanılarda, morfolojide ve sağlık durumunu etkileyen faktörlerde olduğu gibi bir harf ve arkasından gelen rakamlardan oluşan toplam üç ila beş karakter arasında değişen kod yapısına sahiptir. Örnek: J45.0 veya M8140/3.

Sağlık verilerindeki kodlamaların kullanım alanları şöyledir:

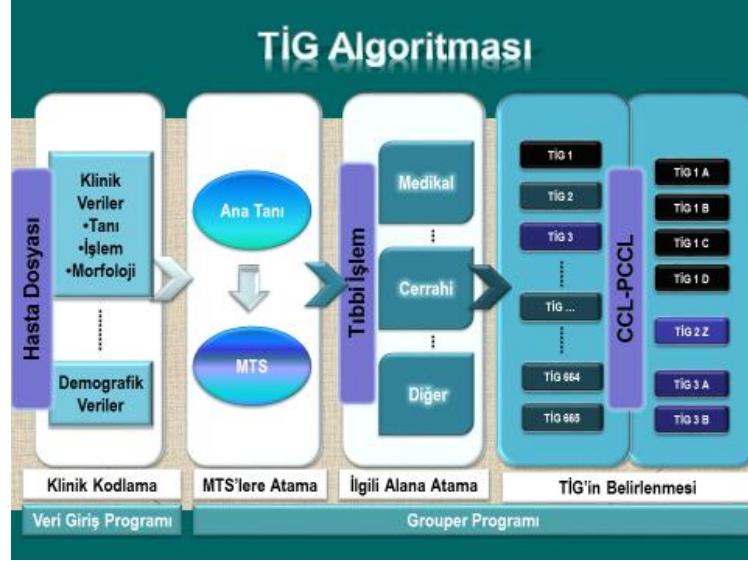
- Sağlık hizmetlerinin karşılaştırılması
- Sağlık politikalarının geliştirilmesi
- Sağlık hizmetlerinin değerlendirilmesi
- Maliyet hesaplamaları
- Ödeme sistemleri

¹²⁹ A.e s.14.

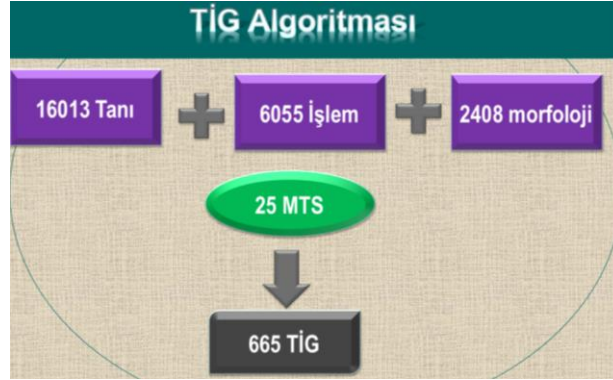
- Araştırma ve planlamalar
-

2.5.3. TİG Oluşum Süreci

TİG oluşum süreci 4 ana başlık altında toplanmaktadır¹³⁰.



Şekil 2-6: TİG Oluşum Süreci



Şekil 2-7: TİG Algoritması Oluşumu

2.5.3.1. Ön Değerlendirme

Ön değerlendirme sürecinde tüm hasta dosyası bir bütün olarak dikkatle incelenmelidir. Hastanın demografik bilgilerinin (hastanın cinsiyeti, yaşı, hastanede

¹³⁰ A.e.s.14-19

kalış süresi, yatış şekli (sevk, doğrudan başvuru, hasta türü/servis değişimi vs.), yoğun bakımda kaldığı gün sayısı, izinli gün sayısı, bebekse yatış ağırlığı...vb.) eksiksiz olarak girilmesi TİG oluşum sürecinde çok önemlidir. Ayrıca, klinik veriler kodlanırken, ana tanı ek tanı dikkatli bir şekilde kodlanmalıdır.

2.5.3.2. Majör Tanı Sınıfı (MTS) Ataması

Hastalıklar ve sağlık durumunu etkileyen diğer faktörlerin gerekli kodlamaları sonucunda ana tanıya uygun olarak 25 farklı MTS'ye atama olmaktadır.

MTS'ler genel olarak ilgili ana tanıların sistemlere göre sınıflandırılmasının yapıldığı bir algoritmaya sahiptir. (Dolaşım sistemi hastalıkları, kan hastalıkları, solunum sistemi hastalıkları vs.) Ayrıca Hatalı TİG'ler olarak sınıflandırılan, kapsamlı olan ve olmayan ameliyathane işlemleri ana tanı ile ilişkili olmayan, neonatal tanı ve yaş/kilo ile tutarsız gibi durumlarda oluşan TİG'ler de bu MTS yapısı içerisinde yer almaktadır.

2.5.3.3. Alan Ataması

Mevcut tanılara ya da hastada tespit edilen diğer durumlara yönelik yapılan işlem türüne göre alan ataması gerçekleştirilmektedir. TİG kodlama yapısında 3 farklı alan bulunmaktadır.

Tablo 2-2: TİG Alan Listesi

Alanı	Kod Aralığı	Sayısı
Dahili	60-99	347
Cerrahi	01-39	280
Diğer	40-59	38

- a) Medikal Alan:** Çoğunlukla vücudun iç organlarının tedavisine yönelik dâhili branşlarda yapılan yatışlarda, cerrahi müdahale içermeyen hasta gruplarının oluşturduğu TİG'lerdir.

- b) Cerrahi Alan:** Vücuttaki yapı bozukluklarının ameliyatla onarılmasına ya da hastalıklı organı kesip çıkararak iyileştirilmesine dayanan cerrahi nitelikteki ameliyathane işlemlerinin yer aldığı TİG gruplarıdır.
- c) Diğer Alan:** Çoğunlukla aynı gün içinde yapılan, ameliyathanede gerçekleştirilen işlemlerin yer almadığı, basit müdahaleleri içeren işlem gruplarıdır.

Tablo 2-3: 25 Majör Tanı Sınıfı¹³¹

MTS	TİG SAYISI	MTS KISALTMASI
MTS - 99 Hatalı TİG'ler	6	
MTS - 00 Öncü - Majör Tanı Sınıfları (MTS)	12	A
MTS - 01 Sinir Sistemi Hastalıkları	53	B
MTS - 02 Göz Hastalıkları	20	C
MTS - 03 KBB & Ağız Hastalıkları	28	D
MTS - 04 Solunum Sistemi Hastalıkları	42	E
MTS - 05 Dolaşım Sistemi Hastalıkları	67	F
MTS - 06 Sindirim Sistemi Hastalıkları	52	G
MTS - 07 Hepatobiliyer Sistem Hastalıkları	29	H
MTS- 08 Kas-iskelet ve Bağ Dokusu Hastalıkları	79	I
MTS - 09 Deri ve Meme Hastalıkları	29	J
MTS - 10 Endokrin ve Metabolik Hastalıklar	19	K
MTS - 11 Böbrek ve İdrar Yolları Hastalıkları	37	L
MTS - 12 Erkek Üreme Organları Hastalıkları	19	M
MTS -13 Kadın Üreme Organları Hastalıkları	20	N
MTS - 14 Gebelik, Doğum ve Lohusalık	17	O
MTS - 15 Yenidoğan (ve Diğer Neonatlar)	25	P
MTS- 16 Kan ve Kan Yapıcı Organların Hastalıklar	10	Q
MTS - 17 Neoplastik Hastalıklar	18	R
MTS - 18 Enfeksiyöz ve Parazitik Hastalıklar	17	T
MTS - 19 Akıl Sağlığı Bozuklukları	13	U
MTS - 20 Alkol/İlaç Kullanımı/ bunlara bağlı Organik Akıl Sağlığı Bozukluklar	8	V
MTS - 21 Yaralanma, Zehirlenme ve Toksik İlaç Etkileri	24	W,X
MTS - 22 Yanıklar	8	Y
MTS - 23 Sağlığı Etkileyen Faktörler ve Sağlık Hizmetleri ile Kurulan Diğer Tıp Temaslar	13	Z

¹³¹ A.e. s.17.

2.5.3.4. TİG'in Oluşumu

Cerrahi, medikal ve diğer gibi ilgili alanlara atanan ana tanımlar daha sonra eşlik eden ek tanımlar, komplikasyon ve komorbidite durumlarının varlığı dikkate alınarak ilgili TİG gruplarına atanırlar.

Sistemin, alan ataması ve TİG oluşumu basamakları gruplayıcı program tarafından verilerin bulunmasına göre ilgili bölümlere sistem tarafından otomatik olarak atanmaktadır.

TİG Kodunun yapısı şu şekildedir:

İlk bölüm, bir harften oluşur. TİG'in ait olduğu MTS grubunu gösterir. İkinci bölüm, TİG'in hangi alana ait olduğunu gösterir. 01-99 arasında değişen rakamlardan oluşur. Üçüncü Bölüm: Kaynak Kullanım derecesini göstermektedir. A, B, C, D ve Z olarak harflerden oluşmaktadır.

Bunlardan A, B, C, D olarak ifade edilenler kaynak kullanımının daha çok olandan az olana doğru değiştiğini gösterirken; Z olarak yer alan TİG'lerin ise tek olduğu ve kaynak kullanımına göre ayrımın olmadığını göstermektedir.

Tablo 2-4: TİG'in Atanması ve Oluşumu



2.5.3.5. ICD-10 ve ACHI kodlarının yapısı

TİG'ler oluşturulmadan önce yapılan kodlar için ICD-10 AM (International Statistical Classification of Diseases and Related Health Problems, Tenth Revision,

Australian Modification- Hastalıkların ve İlgili Sağlık Sorunlarının Uluslararası İstatistiksel Sınıflandırması (Avustralya modifikasyonu) ve ACHI (Australian Classification of Health Intervention- Avusturya Sağlık Girişimleri Sınıflaması) kodlarından yararlanılır.

ICD-10-AM¹³²: ICD-10-AM ile; hastalıklara ilişkin tanımlar, hastalıkların tanı ve tedavisinde uygulanan işlemler (ACHI), sağlık durumunu etkileyen faktörler, semptomlar, belirtiler ve anormal klinik bulgular sınıflandırılmış ve kodlanmıştır. Beş ciltten oluşur.

Cilt 1: Hastalıkların tablo listesi

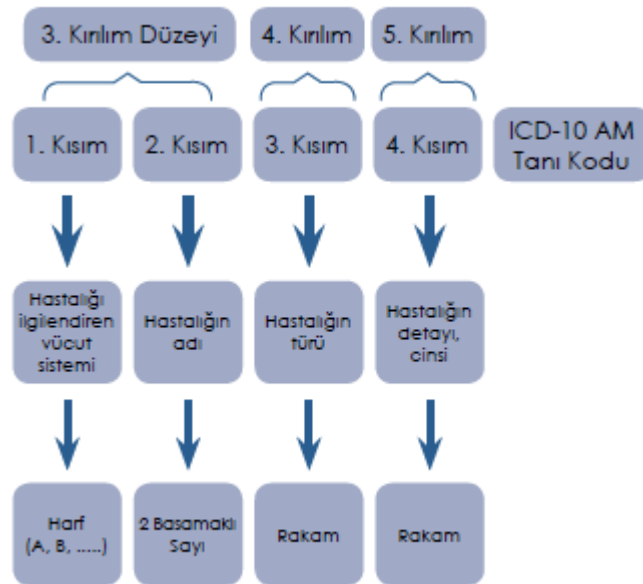
Cilt 2: Hastalıkların alfabetik dizini (indeks)

Cilt 3: İşlemlerin (ACHIin) tablo listesi

Cilt 4: İşlemlerin (ACHIin) alfabetik dizini

Cilt 5: Avustralya kodlama standartları

ICD-10 AM tanı kodları alfa nümerik ve 3 – 4 veya 5 kırımlı yapıdadır.



Şekil 2-8: ICD-10 Tanı Kodlarının Yapısı¹³³

¹³² Ali Yılmaz; "Klinik Kodlama Süreci", Tanı İlişkili Gruplar e-Bülteni, Sayı:1, Ekim-Kasım-Aralık 2007, s.11.

¹³³ Akdağ, a.e.s.16.

ACHI Avustralya Sağlık Girişimleri Sınıflaması olup cerrahi, tanısal ve tetkik amaçlı yapılan işlemler sınıflandırılmıştır. İşlem kodları 7 basamaklı ve numerik yapıdadır. Cerrahi işlemler, medikal girişimler, yardımcı sağlık hizmetleri girişimlerini ve dış işlemlerini kapsar. Anatomik bölgeye göre düzenlenmiştir.

Tablo 2-5: TİG için temel veriler¹³⁴

Ana tanı	Araştırma sonunda, hastanın hastanedeki yatışının esas nedeni olarak ortaya konulan tanıdır veya ana yatış nedenidir.
İşlem(ler)	Genellikle, TİG ataması üzerinde yalnızca bir işlem etkilidir. Birden fazla işlemin kaydedildiği durumda, kaynak kullanımının yüksek olduğu TİG'lere doğru bir geçiş olur.
Ek tanıları (önemli komorbidite veya komplikasyonlar)	Gerek ana tanı ile birlikte seyreden (komorbidite) gerek hasta yatışında ortaya çıkan (komplikasyon) bir durum veya şikayet.
Yatışta hastanın yaşı	Yeni doğanlar dışında yıl olarak alınması yeterlidir.
Cinsiyet	Erkek veya kadın
Taburcu şekli (taburcu edildikten sonra gideceği yer)	Hastanın taburcu edildiği sıradaki durumu ve ilgili olduğu hallerde gideceği yeri ifade eder.
Yeni doğanın doğum ağırlığı	Doğum ağırlığı TİG gruplaması için kullanılan verilerden biridir.

¹³⁴ Mehmet Demir, Umut Beylik, Mine Öztürk, İbrahim Doluküp, "Sağlık Hizmetlerinde Teşhisle İlişkili Gruplar (TİG): Diyabet Vakaları Üzerine Karşılaştırma", "Sağlıkta Performans ve Kalite Dergisi", Sayı 6, Ankara, 2013, s. 7.

2.5.4. TİG’e Dayalı Finansmanın Bileşenleri¹³⁵



Şekil 2-9: TİG Finansman Bileşenleri

- a) **Bağıl Değer:** Bağıl değer hesaplanırken öncelikle bir TİG grubundaki hastaların toplam maliyeti, toplam hasta sayısına bölünerek ortalama maliyet hesaplanır. Bir TİG grubunun ortalama maliyeti hesaplandıktan sonra tüm ülke veya pilot hastaneler grubu için ortalama maliyet; maliyetlerin toplamı, toplam hasta sayısına bölünerek hesaplanır. Bir TİG grubunun ortalama maliyeti, genel ortalama maliyete bölünerek bağıl değeri hesaplanır¹³⁶.

$$B. D. = \frac{\text{İlgili TİG için Ortalama Maliyet}}{\text{Tüm Vakalar için Ortalama Maliyet}}$$

Bir hastanın maliyeti için ortalama bir değer belirlenir, buna taban fiyat denir. Her bir TİG’in bağıl değeri taban fiyat ile çarpılarak TİG başına ödeme, o hastanedeki vaka sayısı ile çarpılarak ise o TİG için hastanenin alacağı geri ödeme miktarı hesaplanır.

¹³⁵ Şencan ve vd., a.g.e., s.3.

¹³⁶ Yılmaz, Ali, “Uluslararası TİG Çalışmaları: TİG Ortamında Ürün Maliyetlendirmesi”, Tanı İlişkili Gruplar e-Bülteni, Sayı 7, s.16.

b) Vaka Karma İndeksi (VKİ): Bir hastanenin vaka üretimini bir başka hastane ile karşılaştırılmasını sağlayan orandır. Hastanenin tedavi etmiş olduğu hastalıkların kompleksliğinin ölçülmesini sağlayan bir kavramdır¹³⁷.

$$\text{Hastane VKİ} = \frac{\sum(\text{TİG Bağlı Değer} \times \text{Vaka Sayısı})}{\text{A Hastanesi için Toplam Vaka Sayısı}}$$

VKİ'nin kullanım alanları şöyledir:

- Klinik faaliyetlerin ölçülmesi
- Hastaneler arası performans değerlendirmesi
- Finansman
- Hastane içi yönetim aracı
- Kalite ölçümlerine başlamak için bir araç
- Hastanede klinik ve finansal kararların verilmesi
- Hekimler arası karşılaştırma istatistikleri ve hizmet sunucu profilleri tespitinde
- Hastane içi ve hastaneler arası bakım kalitesi karşılaştırmaları
- Klinik rehberler, protokoller ve sürekli kalite gelişimi projelerini destekleme
- Veri ve tanımlama standartlarının oluşturulması

c) Taban Fiyat: Bir hastanın maliyeti için ortalama bir değer belirlenir, buna taban fiyat denir. Her bir TİG'in bağlı değeri taban fiyat ile çarpılarak TİG başına ödeme, o hastanedeki vaka sayısı ile çarpılarak ise o TİG için hastanenin alacağı geri ödeme miktarı hesaplanır.

Taban Fiyat¹³⁸,

- Ortalama hasta maliyetini ifade eder.
- Ortalama vaka için taban fiyat bağlı değeri 1.0 dir.

¹³⁷ A.e. Şencan vd., s.6.

¹³⁸ Reinhard Busse, Alexander Geisler, Wilm Quentin ve Miriam Wiley; "Diagnoses-Related Groups in Europa", WHO-EuroDRG-7.Framework Programme, The McGraw Hill Companies Open University Press, USA, 2011 ve Umut Beylik, "Sağlık Kurumlarında Tanı İlişkili Gruplara Dayalı Geri Ödeme Modeli: Türkiye-Avustralya Karşılaştırması", Gazi Üniversitesi Sosyal Bilimler Enstitüsü İşletme ABD Sağlık Kurumları Yönetimi Doktora Tezi, Ankara, 2014, s.130.

- Taban fiyat olmadan DRG fiyatlarını hesaplamak mümkün değildir.
- DRG fiyatları olmadan ödeme ya da bütçeleme sistemi oluşturmak mümkün değildir.
- Taban fiyat ulusal, eş - grup veya hastane bazında yeni bir finansal sistem oluşturulmasına bağlı olarak hesaplanabilir.

Taban fiyatı hesaplamak için, şunlar belirlenmelidir:

- Hesaplamaların hangi düzeyde yapılacağı,
- Dağıtılacak olan para miktarı,
- DRG'ye göre vakaların sayısı,
- Vaka karması indeksi (VKİ/CMI).

Taban fiyat mevcut parayı vaka karmasına göre ayarlanmış hasta sayısına (vaka sayısı) bölerek ortalama hasta için fiyatı vermektedir.

2.6. Teşhis İlişkili Gruplarla Geri Ödeme

Günümüzde sağlık hizmetlerinin yeterli, kaliteli ve hakkaniyetli bir biçimde sunulması, ayrılan kaynak miktarı kadar bu kaynakların akılcı ve verimli kullanılmasına bağlıdır. Sağlık sektörüne ayrılan kaynakların kullanımını ve kullanım sonuçlarını sorgulayan yaklaşımlar, sağlık harcamalarının hızlı artışı ile birlikte önem kazanmıştır. Hem sağlıklı bir toplum yaratma hem de kaynakların daha iyi kullanımı yönünde artan baskılar, birçok ülkede sağlık sisteminin performansını geliştirmeye ve ölçmeye yönelik girişimlerin artmasına neden olmuş sağlık hizmetlerinin nasıl fiyatlandırılacağı ve geri ödemelerinin nasıl gerçekleştirileceği üzerine yapılan çalışmalar ülkelerin kendi çabaları ile sınırlı kalmayıp uluslar arası bir boyut kazanmıştır¹³⁹.

Sürdürülen çalışmalarda kalite ve maliyet etkenlerini dengeleyebilecek stratejik araçlardan en önemlisi sağlıkta geri ödeme sistemleridir. Uygulanacak geri ödeme

¹³⁹ Serap Durukan, "Farklı DRG Sistemlerinde Uygulanan Maliyet Muhasebesi Metodolojilerinin Karşılaştırılması ve Bunların Sağlık Hizmetleri Kalitesi Üzerine Etkileri", Sağlık Hizmetleri Finansman Yapısının Güçlendirilmesi ve Yeniden Yapılandırılması İçin Altyapı Geliştirilmesi Projesi, Tanı İlişkili Gruplar e-Bülteni-4, s.7.

sisteminin, bakım kalitesini artırmayı ödüllendirirken, gereksiz maliyetleri de cezalandıracak bir yapıya sahip olması gerekmektedir. Geri ödeme sistemleri, bu önemli misyonu nedeniyle sağlık sisteminin en önemli araştırma alanlarından biri haline gelmiştir¹⁴⁰.

En uzun süredir uygulamada bulunan ve en çok tutulan ödeme mekanizması, Tanı ilişkili gruplar (DRGs) olarak bilinen ve 1983'te ilk kez uygulanan bir sistemdir¹⁴¹.

TİG, Hastalıkların klinik ve maliyet verilerini kullanılarak gruplandırılmasını ve benzer hastalıkların benzer gruplara atanmasını içeren “Yatan Hasta Sınıflandırma Sistemi”dir¹⁴².

Amerika’da kamu sigortacılığı olan Medicare programının yaşadığı sıkıntıları gidermek için yeni bir geri ödeme modeli arayışları 1970’li yılların sonunda Dr. Robert Fetter ve ekibinin Yale Üniversitesi’ndeki çalışmalarına kaynaklık teşkil etmiştir¹⁴³. Bu dönemde Dr. Robert Fetter ile Yale Üniversitesi’ndeki meslektaşları kalite, hastanenin “üretimi”, hizmet kullanımı ve kaynak tüketimi incelenebilmesi amacıyla; yatarak tedavi gören hastaların tıbbi kayıtlarından elde edilecek tanısal veriler, prosedürlerle ilgili veriler ve demografik verilerini hastane vakalarının klinik açıdan anlamlı ve kaynak kullanımı bazında homojen gruplar şeklinde düzenlenip düzenlenemeyeceğini belirleme çalışmaları başlamıştır. Böylelikle ortaya çıkan TİG geri ödeme modeli 1983 yılında Amerika’da zorunlu hale gelmiştir¹⁴⁴.

Dünya’da TİG kullanımı açısından önde gelen ülkeler, Avustralya, Kanada ve Norveç olarak görülmektedir¹⁴⁵.

¹⁴⁰ Murat Narmanlı, Günnur Ertong, Ahmet Dikici, Ergin Soysal, Ahmet Tümay, Murat Kahraman Güngör,” Ulusal Teşhis İlişkili Gruplar Sistemine Geçiş İçin Yol Haritası”, **IX.Ulusal Tıp Bilişimi Kongresi**, 2012, s.30.

¹⁴¹ Durukan, a.g.e.,s.7.

¹⁴² Abdullah Öztürk, “Geçmişten Günümüze Türkiye’de TİG Uygulamaları Sunumu”, **2.Ulusal TİG Sempozyumu Eylül 2014**.

¹⁴³ Sabahattin Aydın, Pakize Yiğit, Mehmet Demir, Hasan Güler,” Tanı ilişkili gruplama verileri çerçevesinde Türkiye’de ürogenital kanserlere bakış”, **Yeni Üroloji Dergisi**, 8 (2): 72-78, 2013 ,s.73.

¹⁴⁴ Jugna SHAH, “Uluslararası TİG Çalışmaları: Amerika Örneği”, Sağlık Hizmetleri Finansman Yapısının Güçlendirilmesi ve Yeniden Yapılandırılması İçin Altyapı Geliştirilmesi Projesi, **Tanı İlişkili Gruplar e-Bülteni-3**, 2008, s.5-6.

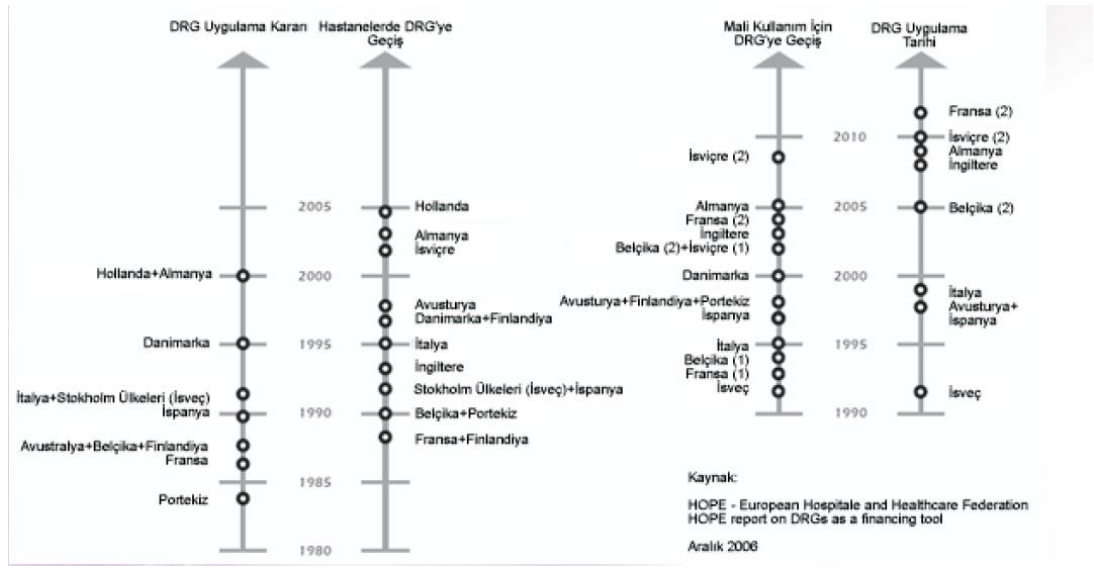
¹⁴⁵ <http://www.tig.saglik.gov.tr/> (Aralık 2012)

Aşağıdaki haritada TİG (DRG) kullanan ülkeler ve Avrupa'daki TİG'in görülmektedir. Şekilde ise, Avrupa'da ülkelerin TİG' e geçiş yılları ayrıntılı olarak verilmiştir.



Şekil 2-10: Dünyada TİG kullanımı¹⁴⁶

¹⁴⁶ Mehtap Tatar, "Sağlık Sisteminin Gelişiminde TİG'in Önemi ve Etkisi", **2.Ulusal TİG Sempozyumu Eylül 2014.**



Şekil 2-11: TİG'in Avrupa'da temel bir finansal altyapı olarak devreye girmesine yönelik kronoloji¹⁴⁷

TİG, diğer sistemlere göre önemli avantajlar sunmaktadır. Gruplandırmalarda vakaların karmaşıklığının göz önüne alınması, hastanelerin zor vakalardan kaçmasının önüne geçer. On binlerce hastalık için ayrı ayrı uğraşmak yerine yüzlerle ifade edilen gruplar ile çalışmak sağlık sisteminin yönetilebilirliği açısından ciddi bir avantaj sağlar. Sonuç odaklı bir sistem olması nedeniyle, hastaların gereksiz yere ve uzun süreli yatırılmasının önüne geçer. Kullanılan kaynaklar ile ilgili doğrudan bir tasarrufu olmadığı için, maliyetlerin kontrolünün sorumluluğunu hastaneye bırakır. Bu da kaynakların etkin kullanımı için çok önemli bir etken oluşturur¹⁴⁸.

¹⁴⁷ Richard Paul Marshall, "DRG sisteminde Yönetim Araçları: DRG Uygulamalarında Ayakta Hasta ve Diğer Servis Yönetim Araçları", Sağlık Hizmetleri Finansman Yapısının Güçlendirilmesi ve Yeniden Yapılandırılması İçin Altyapı Geliştirilmesi Projesi, **Tanı İlişkili Gruplar e-Bülteni-8**, 2009, s.5.

¹⁴⁸ Narmanlı, s.31.

Tablo 2-6: TİG Swot Analizi¹⁴⁹

Güçlü Yanlar	Zayıf Yanlar	Fırsatlar	Tehditler
Hastane kalış sürelerinde kısalma	Hastane başvurularının ve tekrar başvurularının artışı	Hastanelerde toplanan veri kalitesinde artış	Kaynakların gereğinden daha az kullanılması
Hastane kaynaklarının daha etkin kullanılması	Fazla kodlama eğilimi	Kurulması daha zor olsa da, ulusal versiyonun geliştirilmesi	Hastaların gereğinden daha az tedavi edilmesi
Performans artışı	Kurulmasının karmaşık olması		Komplike vakalardan kaçınılması

ABD'de uygulamanın ilk yıllarında, TİG'e dayalı bir ödeme sisteminin yapısında bulunan maliyeti azaltma ve verimliliği artırma yönündeki teşvikin, hastalara zarar verip vermeyeceğini sorgulayan pek çok eleştiri dile getirilmiştir. Örneğin, hastaneler gereksiz hizmetlerin yanı sıra gerekli olan hizmetleri de ortadan kaldırarak masraflarını haddinden fazla kısma çalışacak mıdır? Hastaneler, sabit TİG ödeme oranını kendi yararına kullanmak için “hastaları daha hızlı ve eskisinden daha da hasta olarak” taburcu edecek midir? Hastaneler, yüksek yeniden başvuru oranları yüzünden aynı hastaları tekrar mı muayene edecektir? Medicare hastalarının ölüm oranları artacak mıdır? Hastaneler yalnızca belli tipteki hastaları seçip tedavi ederken, diğerlerini geri mi çevirecektir? Hastaneler sistemi “aldatmak” için hastaları kabul edip, sonra da onları anında sevk mi edecektir? gibi sorular yöneltilmiştir. Eleştiriler çoğunlukla, Medicare programının hastanın yüksek kalitede sağlık hizmeti alabilmesini etkileyebilecek konuları hesaba katıp katmadığı ile ilgili olmuştur. Birleşik Devletler ‘de bu gibi kaygılar o kadar yaygınlaşmıştır ki, hastaneler oldukça dikkatli ve titiz bir şekilde izlenmiştir. Kodlama uygulamaları, doktorun hazırladığı dokümantasyon, vaka karması indeksleri, yeniden başvuru oranları, ölümler ve diğer göstergeler dahil, yukarıda belirtilen hususlar hasta taburcu etme bilgilerini incelenmiştir. Uygulamanın ilk yıllarında hastanede kalış süresi, ölüm oranları, yeniden yatma oranları vb. hususlarda TİG'lerin etkisini inceleyen pek çok araştırma yapılmıştır. Bu araştırmaların bazı genel sonuçlar şöyledir¹⁵⁰:

- Kalış süresinde azalma (1983'ten 1989 yılına kadar 10 günden 8,5 güne)

¹⁴⁹ A.e.s.32.

¹⁵⁰ Durukan, a.g.e., s.5.

- Özellikle ilk yıllarda artan, ileriki yıllarda ise istikrara kavuşan VKİ,
- Hastaneye yatma oranında anlamlı artış görülmemesi; bazı araştırmalar bu oranda düşüş olduğunu göstermiştir.
- Ölüm oranlarında büyük değişiklik görülmemesi,
- 1983'ten 1986'ya kadar lens işlemleri gibi uygulamaların günlük cerrahi merkezlerinde yapılmaya başlanmasıyla günlük cerrahi merkezlerinde (ayakta hasta cerrahi hizmetlerinde) %500 artış,
- Ayakta hasta bakım hizmetlerinde ve harcamalarında artış,
- Hastaneye yeniden yatma oranlarında herhangi bir anormal değişikliğe rastlanmamıştır.
- Evde bakım hizmetlerinde ve hemşirelik hizmetlerinden faydalanma oranında artış,
- Hastaların bir hastaneden diğerine sevk oranında çok az değişiklik olmuştur veya hiç değişme olmamıştır.

Tablo 2-7: Tipik bir Avustralya Eğitim Hastanesinde DRG Finansmanından Önce ve Sonraki Etkinlik İstatistikleri¹⁵¹

1990 (DRG'lerden önce)	2000 (DRG'lerden sonra)	
Yatak sayısı: 500	Yatak sayısı: 380	%24
Yataklı hasta sayısı : 26.000	Yataklı hasta sayısı : 42.800	%64
Ortalama hastanede kalış süresi : 6,0 gün	Ortalama hastanede kalış süresi : 3,2 gün	%46
Yatak dönüşüm oranı: 52 hasta	Yatak dönüşüm oranı: 113 hasta	%117

2.6.1. Türkiye’de Teşhisle İlişkili Gruplarla Geri Ödeme

¹⁵¹ Hacettepe Üniversitesi, Sağlık Hizmetleri Finansman Yönetiminin Güçlendirilmesi ve Yeniden Yapılandırılması için Alt Yapı Geliştirilmesi Projesi, Karar Alıcılar için DRG Sisteminin Benzerlik-Farklılık Raporu, 2005, s.29.

TİG çalışmaları Türkiye’de sağlıkta dönüşüm ile HÜAP (Sağlık Hizmetleri Finansman Yönetiminin Güçlendirilmesi ve Yeniden Yapılandırılması için Alt Yapı Geliştirilmesi Projesi)’nin alt başlıklarından biri olarak 2005-2009 yıllarında ilk defa başlamıştır. 2009 yılında Sağlık Bakanlığı’na (SB) bağlı TİG Şube başkanlığı, 2012’de TİG Daire Başkanlığı kurulmuştur. Toplanan veriler ilk defa Bakanlık Global Bütçesi devlet hastanelerine dağıtılırken kullanılmıştır. Ülkemizde TİG çalışmaları halen devam etmekte olup, halen TİG uygulamasına geçiş süreci devam etmektedir ¹⁵².

2004-2006 (HUAP)	7 Hastane
2006-2008 (HUAP)	48 Hastane
2009 SB	TİG Şube <u>Müd.</u> Kuruldu 50 Hastane
2010	260 Hastane
2011	550 Hastane
2012	TİG Daire Başkanlığı kuruldu
2013-2014	523 Hastane Özel ve Üniversiteler dahil sisteme entegre oldu

Şekil 2-12: Sağlık Bakanlığı TİG Süreci¹⁵³

TİG çalışmalarına temel olacak yapı için Avustralya örneği ele alınmıştır. Çalışmanın başlangıcından itibaren ICD10-AM (International Classification of Diseases 10th Revision, Australian Modification: Uluslararası Hastalık Sınıflandırması 10. Versiyon, Avustralya Modifikasyonu) 4. Güncellemesi ve TİG algoritması olarak da AR-DRG (Australian Refined Diagnosis Related Groups: Grupları tayin eden algoritma, grouper) 5.1 versiyonu kullanılmaya başlanmıştır¹⁵⁴.

¹⁵² Yıldız Ayanoglu, Umut Beylik, Fatih Orhan, “Tanı İlişkili Gruplara (DRG) Göre Hastaneler ve Ülkeler Arası Karşılaştırma: Bir Vaka Örneği”, **Elektronik Sosyal Bilimler Dergisi**, Güz-2014 Cilt:13 Sayı:51 (273-290), 2014, s.277-280.

¹⁵³ Öztürk, a.g.e.

¹⁵⁴ Akdağ, a.g.e.s.9.

2010-2012 yıllarında geri ödemelerin % 10'u TİG üzerinden yapılmıştır¹⁵⁵.

Ülkemizde, başta sağlık finansmanının geri ödeme boyutunda olmak üzere, performans ölçümü, insan kaynakları planlaması, klinik faaliyetle ilişkili sağlık istatistiklerinin oluşturulması, sağlık politikalarının geliştirilmesi, araştırma ve planlamalarda kullanılmak üzere, Sağlık Bakanlığınca yürütülen Sağlıkta Dönüşüm Programı çerçevesinde Sağlık Hizmetleri Genel Müdürlüğü tarafından Teşhis İlişkili Gruplar (TİG) çalışmaları yürütülmektedir. 2013 yılından beri özel hastanelere yapılacak uzman tabip kadro dağıtımında kullanılacak parametrelerden birinin de TİG verileri olması planlandığından, özel hastanelerden de titizlikle veri toplanmaktadır¹⁵⁶.

¹⁵⁵ Öztürk, a.g.e.

¹⁵⁶ <http://www.tig.saglik.gov.tr/> Haziran 2013 ve Kasım 2013 Resmi Yazıları.

3. UYGULAMA

Uygulamada İstanbul, Ankara ve İzmir’de bulunan 15 Eğitim-Araştırma hastanesinde 1 Ocak - 31 Aralık 2012 tarihinde yatan hastaların TİG verileri kullanılacaktır. Veriler Sağlık Bakanlığı izni ile alınmıştır.

15 eğitim araştırma hastanesinde 15 Majör Tanı Sınıfında (MTS) 664 TİG kodunda 693.289 hasta kaydı bulunmaktadır. Veri setinde, hastaların cinsiyeti, yaşı, yatış tarihi, hastanede kalış süreleri, sosyal güvenceleri, hastaneden taburcu olma nedenleri, TİG’in bağıl değeri ve hastaya uygulanan işlemin türü bulunmaktadır.

3.1. Problemin Tanımlanması

Sağlık hizmetlerinde kaynaklar sınırlıdır ve çok maliyetlidir. Bu nedenle, sağlık hizmetlerinde bu kaynakları en etkili ve verimli şekilde dağıtmak bütün ülkelerin en temel sorunlarından bir tanesidir. İşte bu nedenle ülkeler bu kaynakların etkili ve verimli şekilde dağıtılması için hastalıkların sınıflandırılmasına geçmiştir. Bu sınıflandırmanın en çok kullanılanı ise bu tezin konusu olan TİG’dir.

Bu çalışmada, ülkemizde 2005 yılından beri pilot çalışmalarla başlamış, 2011’de yaygın olarak kullanılmaya başlanan TİG’in veri madenciliği yöntemleri ile incelenmesi amaçlanmıştır. Dünyanın pek çok ülkesinde kullanılan TİG’in başarıya ulaşması için hala çalışmalar sürmektedir. Her ülke, başlangıçta başka bir ülkeyi model alarak bu sisteme geçse de, her ülkenn sistemi kendisine özgüdür ve bazı aksaklıklar mevcuttur ve her ülke bunu geliştirmek için çalışmaktadır. Türkiye TİG’e henüz yeni geçmiş bir ülkedir ve TİG algoritmasını Avustralya’dan almaktadır. Ulusal TİG oluşturabilmek için titizlikle oluşan TİG’ler üzerine çalışılması gerekmektedir.

Bu çalışmada ülkemizde 2012 yılında Eğitim-Araştırma hastanelerinde oluşan TİG’lerin etkililiği araştırılacaktır. Oluşan TİG verilerinin, hastanede kalış süreleri ile

TİG bağıl değerleri arasında bir ilişki olup olmadığı, bir hastanın ne kadar yaşı yüksek ise ona uygulanan tedavinin o kadar uzun süreceği gibi, bu sistemi kullanan bütün ülkelerin araştırdığı ve cevap aradığı sorular olacaktır. Çünkü hastanede yatış süresi, hastanın maliyetini arttıran en önemli nedenlerden bir tanesidir. Fakat TİG ile hastanelere yapılan ödemeler bağıl değerine göre ağırlıklandırılarak yapılır, yani hastanede yatış süresinden bağımsızdır. Peki oluşan TİG’lerin bağıl değerleri ile kalış süresi arasında bir ilişki var mıdır? Bu soru araştırmanın temelini oluşturmaktadır.

Ayrıca, veri setimizde bulunan her bir kayıt ayrı bir vaka olarak değerlendirilmiştir. Yani bir hastanın mükerrer yatışları buna dahildir. TİG frekansları aynı tanı grubunda hastanede bir yılda yatarak tedavi gören hasta sayısını yansıtmaktadır. Bu açıdan, TİG çerçevesinde elde edilen veriler, toplumun yansıması değil, hastanelerdeki faaliyetin yansımasıdır. Dolayısıyla bu verilerden hareketle geliştirilecek politikalar doğrudan klinik uygulamalar ve hastane planlamaları ile ilgili olacaktır.

Çalışmada önce kümeleme analizi uygulanarak bu 15 eğitim araştırma hastanesinde yatan hastaların çalışmada kullanılan TİG veri seti çerçevesinde profilleri incelenmiştir. Buradaki amaç, kümelerde hasta yatış süresi ve yaş ile bağıl değer arasında bir ilişki olup olmadığını görebilmektir.

İkinci olarak da, hastanelerin en büyük sıkıntısı olan hasta yatış süresi fazla olan hastalara etki eden değişkenler TİG çerçevesinde lojistik regresyon analizi ve karar ağaçları karşılaştırmalı olarak incelenmiştir.

Son olarak da, diyabet hastalığına sahip hastaların durumları birliktelik analizi ile incelenmiştir. Diyabet vakalarının incelenmesinin nedeni ise son yıllarda ülkemizde diyabet vakalarının endişe verici şekilde artmasıdır.

2013 yılında Uluslararası Diyabet Federasyonu tarafından yayınlanan “Altıncı Diyabet Atlası” dünyada diyabetin mevcut durumu ve geleceği 2035 yılı tahminlerine göre Türkiye, diyabetli nüfus itibarı ile diyabetin dünyada en yüksek olacağı ilk 10 ülke arasına girecektir. Mevcut duruma göre ise ülkemiz Avrupa ülkeleri arasında diyabet prevalansının en yüksek olduğu ülkedir. Türkiye’de yakın zamanda SGK tarafından yayımlanan çalışmaya göre 2008 yılından 2012 yılına kadar diyabetli kişi sayısı

2.514.842 kişiden 5.217.709 kişiye ulaşarak yaklaşık olarak %107 artmıştır. Bu hastalara sağlık bütçesinden ayrılan pay ise % 16.4'den % 23'e çıkmıştır¹⁵⁷.

Bu yüzden bu hastaların TİG verisi birliktelik kurallarını incelemek geliştirilecek sağlık politikaları için önem arz etmektedir.

Analizler, SAS Enterprise Miner 9.4 ve SPSS Clementine 12.0 Programı ile gerçekleştirilmiştir.

3.2. Verilerin Hazırlanma Süreci

Veriler, Sağlık Bakanlığı'ndan temin edilmiştir. Veri setinde herhangi bir boş veri bulunmamaktadır ve verilerin hepsinin doğru olarak girildiği varsayılmıştır.

Veride herhangi bir birleştirme ve temizleme işlemi yapılmamıştır. Veride sadece dönüştürme işlemi uygulanmıştır.

Cinsiyet, şehir, hastane, hastanın taburculuk nedeni, TİG kodu, hastaya uygulanan işlem kodlanmıştır. Bu değişkenler çalışmanın nominal ölçekli değişkenleridir.

Bu 15 eğitim araştırma hastanesinde 2012 yılında yatan hastaların % 49'u erkek, % 51'i ise erkektir.

Tablo 3-1: Hastaların Cinsiyet Dağılımı

Cinsiyet	Oran (%)
Erkek	49%
Kadın	51%

Hastaların %1,6'sı tedavileri sonucunda vefat etmiş, geri kalan hastalar şifa, başka hastaneye ya da servise sevk gibi nedenlerle hastaneden ayrılmışlardır. Hastaların

¹⁵⁷ T.C. Sağlık Bakanlığı Halk Sağlığı Kurumu, "Türkiye Diyabet Programı", Ankara, 2014, s. 32-37.

hastaneden çıkış nedenleri, analize vefat eden ve diğer hastalar olarak ayrılarak alınmıştır.

Tablo 3-2: Hastaların Hastaneden Çıkış Nedenleri

Taburcu Nedeni	Oran %	Taburcu Nedeni	Oran (%)
Bakımevine Transfer	1%	Vefat	2%
Başka Bir Hastaneye Sevk	1%	Diğer Nedenler	98%
İstatistiksel Taburcu-Hasta Türü /Servis Değişimi	0.05		
İstatistiksel Taburcu-İzinli Ayrılış	2%		
Ölüm	2%		
Psikiyatri	0.03%		
Rehabilitasyon İçin Taburcu/Transfer	0.10%		
Şifa İle Taburcu/Eve	88%		
Tavsiyeye Karşı Ayrılış	0.021		

Hastaların % 90'ı Genel Sağlık Sigortalı (GSS), % 3,6'sı Yeşil Kartlı, % 2,4'ü ücretli, % 0,4'ü özel sigortalıdır. Hastaların sosyal güvenceleri GSS ve diğer olarak ayrılarak analize alınmıştır.

Tablo 3-3: Hastaların Sosyal Güvence Dağılımları

Sosyal Güvence	Oran (%)	Sosyal Güvence	Oran (%)
GSS	90%	Diğer	10%
Diğer	4%	GSS	90%
Yeşil Kart	4%		
Ücretli	2%		
Özel Sigortalı	0.4%		

Hastanın yaşı bu konuda daha önce yapılan çalışmalar incelenerek 8 kategoriye ayrılmıştır. Yaş değişkeni, kümeleme analizi ve birliktelik analizinde ordinal, karar ağacı ve lojistik regresyon analizlerinde oran ölçekli olarak analize alınmıştır.

Tablo 3-4: Hastaların Yaş Dağılımları

Yaş	% Oran
0-10	12%
11-15	2%
16-20	4%
21-30	12%
31-40	13.0%
41-50	14%
51-65	22%
65+	21%

Vakaların bağıl değerleri ise kartillerine göre 3'a ayrılmıştır. 1. Kartilden az olan bağıl değerler düşük, 1. Kartil ve 3. Kartil arasındaki bağıl değerler orta, 3. Kartilden fazla olanlar ise yüksek bağıl değerler olarak adlandırılmıştır. Çalışmadaki eğitim araştırma hastanelerinde en çok uygulanan tedaviler orta seviyede bağıl değere sahip tedavilerdir.

Tablo 3-5: Hastalara Uygulanan Tedavilerin Bağıl Değer Dağılımları

Bağıl Değer Aralığı	İsim	Oran(%)
≤ 0.48	düşük	26%
0.49-1.33	orta	49%
≥ 1.34	yüksek	25%

Hastanede kalış süresi kümeleme analizi için, daha önce yapılan çalışmalara göre 4 kategoriye ayrılmıştır. Çalışmadaki eğitim araştırma hastanelerinde yapılan tedavilerin büyük bir çoğunluğu gününbirlik tedavilerden oluşmaktadır.

Tablo 3-6: Hastaların Hastanede Yatış Günü Dağılımları

Yatış Günü Grup	Kod	Oran (%)
Gününbirlik (0 gün)	1	32%
1-2 gün	2	27%
2-5 gün	3	19%
>5 gün	4	22%

Lojistik regresyon ve karar ağacı analizlerinde, kümeleme analizinden farklı olarak hastaların komorbidite ve komplikasyona (Gerek ana tanı ile birlikte seyreden (komorbidite) gerek hasta yatışında ortaya çıkan (komplikasyon) bir durum veya şikayet) sahip olup olmadıkları da incelenmiştir.

Tablo 3-7: Hastalardaki Komorbidite Görülme Sıklığı

Komorbidite	Oran (%)
Komorbidite Var	8%
Komorbidite Yok	92%

Lojistik regresyon ve karar ağacı analizleri için ise hastanede kalış süresi aşırı yatış süresini (length of stay outliers) göstermek için ikiye ayrılmıştır. Literatürde aşırı yatış süresi geometrik ortalama+2 standart sapmadan fazla olan yatışlar ve kartillere dayanan uç değer analizi olmak üzere iki türdedir. Bu çalışmada ise geometrik

ortalama +2 standart sapma yaklaşımı kullanılarak aşırı yatış süreleri belirlenmiştir ve bu oran % 6.8 bulunmuştur.

TİG kodları nominal değişken olarak kümeleme analizinde kullanılmıştır.

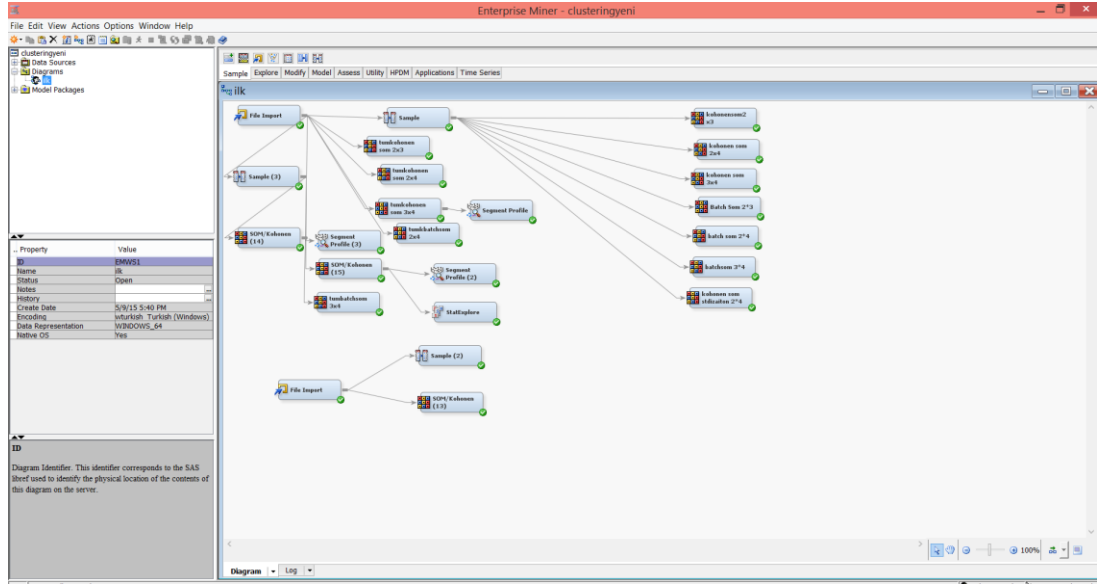
Kümeleme, Lojistik Regresyon ve Karar Ağacı Analizleri SAS Enterprise Miner 13.1’de gerçekleştirilmiştir.

Birliktelik analizi için veri ikili (binary) hale getirilmiştir ve sonuçların kolay yorumlanabilmesi için sadece diabet vakalarına uygulanmıştır. Birliktelik Analizleri SPSS Clementine 12.0 programında gerçekleştirilmiştir.

3.3. Kümeleme Analizi

Kümeleme analizinde en doğru sonucu bulabilmek için K-Means, Hiyerarşik, Kendini düzenleyen haritalar (Self-Organizing Maps-SOM, Kohonen Ağları) kümeleme yöntemleri denenmiştir. Ayrıca veri seti çok büyük olduğu için tabakalı örnekleme yöntemi ile veriden çeşitli örneklemler alınarak da analizler tekrarlanmıştır. Fakat, istatistiksel ve klinik anlamlılık olarak en uygun küme bütün veri seti ile uygulanan sonuçlardır. Bu nedenle sadece bu sonucun bulgularına yer verilmiştir.

Klinik anlamlılık sağlanması için uzman görüşü de alınarak yapılan inceleme sonucunda en uygun kümeleme sonucunun SOM kümeleme yöntemi ile olduğu görülmüştür. Kümeleme sonuçlarına göre 12 küme elde edilmiştir.



Şekil 3-1: Kümeleme Analizi SAS Enterprise Miner Uygulaması

Kümeleme analizi sonrasında uygulanan Segment Analizine göre, her değişken her kümede anlamlı bulunmamıştır. Fakat bütün kümelerdeki bütün değişkenler yorumlanmıştır.

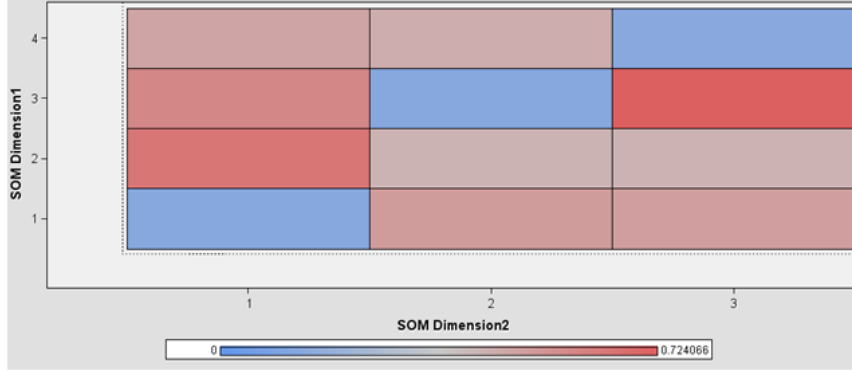
Tablo 3-8: Kohonen Ağ ile Oluşan 12 Kümedeki İstatistikler

SOM Segment ID	Gözlem Sayısı	Gözlem Sayısı %	SOM Boyut1	SOM Boyut2	Ort. Bağlı Değer	Ort. Yaş	Ort. Yatış Günü	Cinsiyet		Sosyal Güvence		Hastaneden Çıkış Nedeni		Uygulanan İşlemin Türü		
								Kadın	Erkek	SGK	Diğer	Vefat	Diğer	Cerrahi	Dahili	Diğer
1	8,323	1.2%	1	1	0.130	0.697	0.279	0%	100%	95%	5%	0%	100%	100%	0%	0%
2	35,578	5.1%	1	2	0.516	0.389	0.526	0%	100%	0%	100%	2%	98%	37%	56%	6%
3	7,076	1.0%	1	3	0.506	0.058	0.472	100%	0%	100%	0%	0%	100%	0%	100%	0%
4	122,696	17.7%	2	1	0.645	0.508	0.572	0%	100%	100%	0%	2%	98%	100%	0%	0%
5	143,666	20.7%	2	2	0.433	0.486	0.506	0%	100%	100%	0%	2%	98%	0%	100%	0%
6	147,206	21.2%	2	3	0.436	0.502	0.495	100%	0%	100%	0%	2%	98%	0%	100%	0%
7	139,443	20.1%	3	1	0.585	0.519	0.510	100%	0%	100%	0%	1%	99%	100%	0%	0%
8	8,799	1.3%	3	2	0.130	0.513	0.161	100%	0%	93%	7%	0%	100%	0%	0%	100%
9	12,215	1.8%	3	3	0.724	0.658	0.595	0%	100%	100%	0%	2%	98%	0%	0%	100%
10	33,411	4.8%	4	1	0.487	0.410	0.501	100%	0%	0%	100%	1%	99%	39%	56%	5%
11	17,817	2.6%	4	2	0.453	0.637	0.393	100%	0%	100%	0%	1%	99%	0%	0%	100%
12	17,053	2.5%	4	3	0.133	0.621	0.166	0%	100%	99%	1%	0%	100%	0%	0%	100%

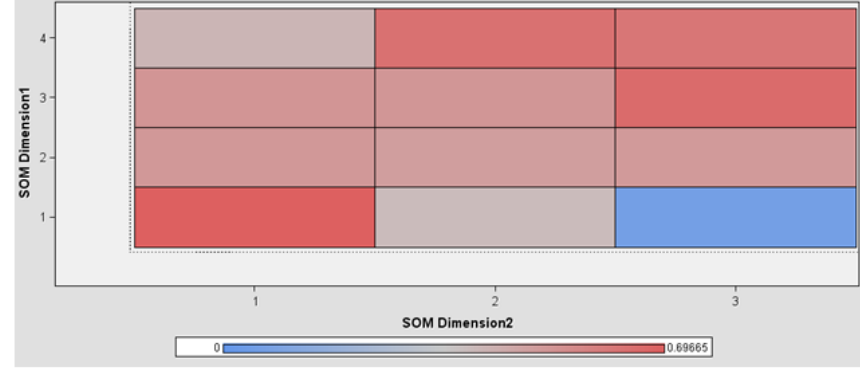
Tablo 3-9: Kümelerdeki En Çok Tekrar Eden 5 TİG

Kümelerdeki En Çok Tekrar Eden 5 TİG Kodu

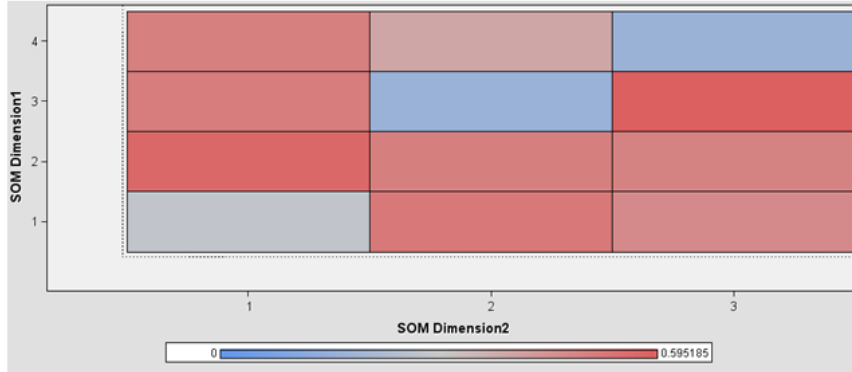
TİG	TİG Açıklama	küme1	TİG	TİG Açıklama	küme7
C03Z	Göz, Retinaya Yönelik Tanı ve Tedavi İşlemleri	100%	O01C	Doğum, Sezeryan, Katastrofik/Şiddetli KK Bulunmayan	6%
			C03Z	Göz, Retinaya Yönelik Tanı ve Tedavi İşlemleri	6%
			J11Z	Cilt, Subkütan ve Meme İşlemleri, Diğer	5%
			C16A	Göz, Lens Tanı ve Tedavi İşlemleri	4%
			H08B	Kolesistektomi, Laparoskopik, Kapalı Koledok Eksplorasyonu Yapılmayan, Katastrofik/Şiddetli KK Bulunmayan	4%
TİG	TİG Açıklama	küme2	TİG	TİG Açıklama	küme8
P67D	Yenidoğan, Başvurudaki Ağırlık 2499 g Üzeri, Önemli Ameliyathane İşlemi Olmaksızın, Problemin Eşlik Etmediği	7%	G45B	Gastroskopi, Diğer, Majör Olmayan Sindirim Sistemi Hastalıkları İçin, Aynı Gün İçerisinde	100%
X64B	Yaralanmalar, Diğer, Zehirlenmeler ve Toksik Etki Tanısı, 60 Yaş Altı, KK Bulunmayan	3%			
X62B	Zehirlenmeler/Toksik Etki, İlaçlar ve Diğer Maddeler, 60 Yaş Altı, KK Bulunmayan	3%			
G07B	Appendektomi, Katastrofik/Şiddetli KK Bulunmayan	2%			
D10Z	Nazal Tanı ve Tedavi İşlemleri	2%			
TİG	TİG Açıklama	küme3	TİG	TİG Açıklama	küme9
P67D	Yenidoğan, Başvurudaki Ağırlık 2499 g Üzeri, Önemli Ameliyathane İşlemi Olmaksızın, Problemin Eşlik Etmediği	100%	F42B	Dolaşım Sistemi Bozuklukları Akut Myokard Enfarktı Bulunmayan, İnvaziv Kardiyak İnceleme İşlemleri İle, Kompleks Tanı/İşlem Bulunmayan	27%
			G42A	Gastroskopi, Diğer, Majör Sindirim Sistemi Hastalıkları İçin	10%
			E41Z	Solunum Sistemi Tanısı, İnvaziv Olmayan Ventilasyon Desteği Bulunan	10%
			H42C	ERCP, Diğer Terapötik İşlemler, KK Bulunmayan	8%
			G45A	Gastroskopi, Diğer, Majör Olmayan Sindirim Sistemi Hastalıkları İçin	8%
			F42A	Dolaşım Sistemi Bozuklukları Akut Myokard Enfarktı Bulunmayan, İnvaziv Kardiyak İnceleme İşlemleri İle, Kompleks Tanı/İşlem Bulunan	8%
			G46B	Gastroskopi, Komplike, Katastrofik/Şiddetli KK Bulunmayan	6%
			G44B	Kolonoskopi, Diğer, Katastrofik/Şiddetli KK Bulunmayan	6%
TİG	TİG Açıklama	küme4	TİG	TİG Açıklama	küme10
G09Z	Herni (Fıtık) İşlemleri, İnguinal ve Femoral, 0 Yaş Üzeri	6%	P67D	Yenidoğan, Başvurudaki Ağırlık 2499 g Üzeri, Önemli Ameliyathane İşlemi Olmaksızın, Problemin Eşlik Etmediği	7%
J11Z	Cilt, Subkütan ve Meme İşlemleri, Diğer	5%	O60C	Doğum, Vajinal Yolla, Tekil, Komplike Olmayan, Diğer Olumsuzlukların Eşlik Etmediği	6%
C16A	Göz, Lens Tanı ve Tedavi İşlemleri	5%	X62B	Zehirlenmeler/Toksik Etki, İlaçlar ve Diğer Maddeler, 60 Yaş Altı, KK Bulunmayan	5%
D10Z	Nazal Tanı ve Tedavi İşlemleri	4%	O01C	Doğum, Sezeryan, Katastrofik/Şiddetli KK Bulunmayan	5%
G07B	Appendektomi, Katastrofik/Şiddetli KK Bulunmayan	4%	C16A	Göz, Lens Tanı ve Tedavi İşlemleri	2%
TİG	TİG Açıklama	küme5	TİG	TİG Açıklama	küme11
P67D	Yenidoğan, Başvurudaki Ağırlık 2499 g Üzeri, Önemli Ameliyathane İşlemi Olmaksızın, Problemin Eşlik Etmediği	5%	G44C	Kolonoskopi, Diğer, Aynı Gün İçerisinde	22%
D66B	Kulak, Burun, Boğaz ve Ağzı, Diğer Tanılar, KK Bulunmayan	3%	F42B	Dolaşım Sistemi Bozuklukları Akut Myokard Enfarktı Bulunmayan, İnvaziv Kardiyak İnceleme İşlemleri İle, Kompleks Tanı/İşlem Bulunmayan	14%
I74C	Önköl, El Bileği, El, Ayak Yaralanmaları, 75 Yaş Altı, KK Bulunmayan	3%	G46C	Kolonoskopi, Komplike, Aynı Gün İçerisinde	9%
C63B	Göz, Diğer Hastalıkları, KK Bulunmayan	2%	H42C	ERCP, Diğer Terapötik İşlemler, KK Bulunmayan	8%
I75C	Omuz, Kol, Dirsek, Diz, Bacak veya Ayak Bileği Yaralanmaları, 65 Yaş Altı, KK Bulunan	2%	L41Z	Sistoüretroskopi, aynı Gün İçerisinde	6%
TİG	TİG Açıklama	küme6	TİG	TİG Açıklama	küme12
O60C	Doğum, Vajinal Yolla, Tekil, Komplike Olmayan, Diğer Olumsuzlukların Eşlik Etmediği	6%	G45B	Gastroskopi, Diğer, Majör Olmayan Sindirim Sistemi Hastalıkları İçin, Aynı Gün İçerisinde	31%
X64B	Endokrin Hastalıkları, Katastrofik/Şiddetli KK Bulunmayan	4%	G44C	Kolonoskopi, Diğer, Aynı Gün İçerisinde	26%
D66B	Kulak, Burun, Boğaz ve Ağzı, Diğer Tanılar, KK Bulunmayan	3%	L41Z	Sistoüretroskopi, aynı Gün İçerisinde	19%
Q61C	Eritrosit (Kırmızı Kan Hücre) Hastalıkları, Katastrofik/Şiddetli KK Bulunmayan	3%	Z40Z	Hasta Takibi, Endoskopik İşlemin Eşlik Ettiği	8%
O66A	Başvuru, Antenatal ve Diğer Obstetrik Nedenler	2%	G46C	Kolonoskopi, Komplike, Aynı Gün İçerisinde	7%



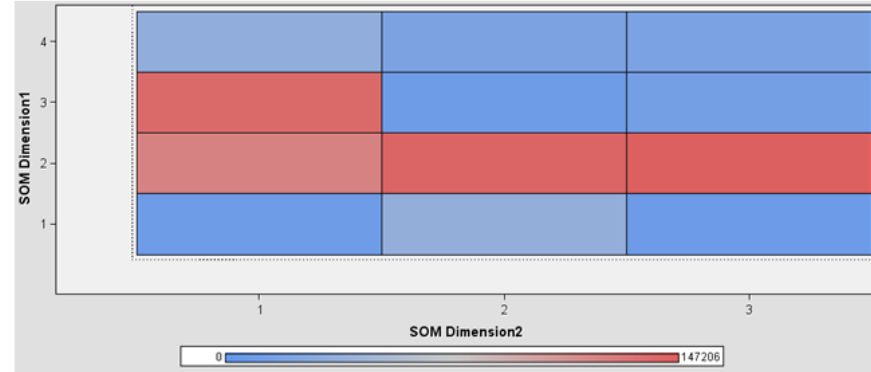
Kümelerdeki Bağıl Değer Dağılımı



Kümelerdeki Yaş Dağılımları

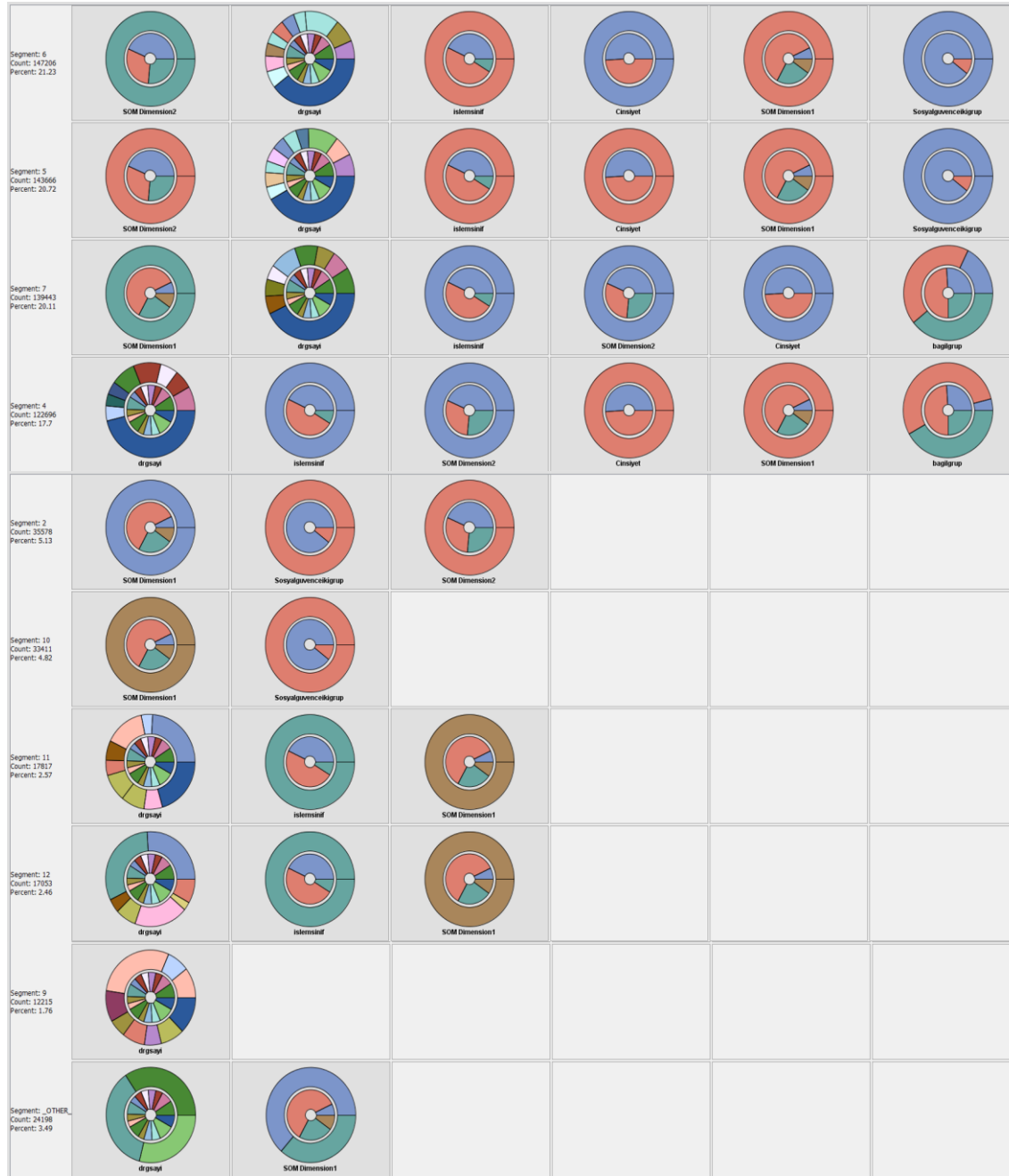


Kümelerdeki Yatış Günü Dağılımı



Kümelerdeki Gözlem Sayısı

Şekil 3-2: Kohonen Geri Bildirim Haritaları



Şekil 3-3: Kohonen Ağları Kümeleme Analizi Sonuçlarına Uygulanan Segment Analizi ile Her bir Kümedeki Önemli Değişkenler

SOM kümeleme analizinde nominal ve ordinal ölçekli değişkenler kullanılmıştır. Hastanın yatış süresi, hastaya uygulanan işlemin bağıl değeri ve hastanın yaşı ordinal ölçekli değişkenlerdir ve bu değişkenlerin kümelerdeki ortalamaları yorumlanmıştır. Tablo 3.8-9 ve Şekil 3.2 ve 3’de sonuçlar görülmektedir.

Şekil 3.2’de Kohonen Ağ haritaları görülmektedir. Renkler maviden kırmızıya göre değişmektedir. Mavi renk düşük değerleri, kırmızı ise yüksek değerleri temsil etmektedir. Bağıl değer ve yatış günü haritaları incelendiğinde, kümelerdeki bağıl değer ve yatış günleri dağılımlarının paralellik gösterdiği görülmektedir.

Birinci kümedeki hastalar, ortalama en düşük TİG vaka bağıl değer, en yüksek yaş ve düşük yatış gününe sahip % 95’i SGK’lı erkek hastalardır. Hastalara sadece “Göz, Retinaya Yönelik Tanı ve Tedavi İşlemleri” uygulanmıştır. Küme tek TİG’den oluşmaktadır.

İkinci kümedeki hastalar sosyal güvencesi GSS dışındaki sosyal güvencelere sahip erkek hastalardır. Vakalara ortalama orta dereceli bağıl değerli % 37 cerrahi, % 56 dahili tedavi uygulanmış ve % 2’si vefatla sonuçlanmıştır. Hastalar, orta derecede yatış gününe ve düşük yaş ortalamasına sahiptir. Oluşan TİG kodlarının ise genelde acil müdahaleler olduğu görülmektedir

Üçüncü kümedeki vakaların hepsi “Yenidoğan, Başvurudaki Ağırlık 2499 g Üzeri, Önemli Ameliyathane İşlemi Olmaksızın, Problemin Eşlik Etmediği” yeni doğan kız bebeklerdir. Hepsi GSS’li ailelerin bebekleridir ve dahili vakalardır.

Dördüncü kümedeki vakaların ortalama yüksek TİG vaka bağıl değerli, orta yaşlı ve ortalamanın biraz üstünde yatış gününe sahip GSS’li cerrahi tedavi gören erkekler olduğu görülmektedir. Oluşan TİG’ler sonucunda hastaların % 2’si vefat etmiştir. Oluşan TİG’ler genellikle hastanelerde sık yapılan ameliyatlardan oluşmaktadır.

Beşinci kümedeki hastalar dördüncü kümeye göre ortalama daha düşük bağıl değerli dahili vakalardır. Yine bütün hastalar GSS’li ve uygulanan tedavi sonrasında hastaların % 2’si vefat etmiştir. Oluşan TİG’ler vakaların kronik vakalar olduğunu göstermektedir.

Altıncı kümedeki vakalar, ortalamanın altında TİG vaka bağıl değerine, ortalama yaşa ve yatış süresine sahip GSS’li kadınlardır. Bu kümede oluşan TİG’ler komorbiditesi bulunmayan ve hafif yararlanan dahili işlem gören vakalardır.

Yedinci kümeyi, cerrahi tedavi görmüş, ortalamanın üstünde TİG vaka bağıl değere, ortalama yaşa ve yatış gününe sahip GSS’li kadınlar oluşturmaktadır. Oluşan TİG’ler cerrahi basit işlemlerdir.

Tanı amaçlı kullanılan gününbirlik basit bir işlem olan” Gastroskopi, Diğer, Majör Olmayan Sindirim Sistemi Hastalıkları İçin, Aynı Gün İçerisinde “ TİG kodu tek başına sekizinci kümeyi oluşturmaktadır. Bu hastalar, orta yaşın üzerinde ve % 93’ü GSS’li olan kadın vakalardır.

Dokuzuncu küme, ortalama TİG vaka bağıl değeri ve yatış günü en yüksek olan kümedir. Hastalar, % 2’si vefat eden GSS’li, diğer işlemler görmüş yaş ortalaması yüksek erkeklerdir. Oluşan TİG’ler ciddi vakalardır.

Onuncu kümede ortalama TİG vaka bağıl değerleri ve yatış süreleri ortalama seviyesinde, yaşları ise ortalamanın altında bulunan % 39’u cerrahi, % 56’sı dahili tedavi görmüş, GSS’si bulunmayan kadınlar oluşturmaktadır. Oluşan TİG’ler incelendiğinde ise, vakaların acil ve öncelikli olduğu görülmektedir.

Onbirinci kümede ortalama yatış günü düşük, yaşları yüksek, TİG vaka bağıl değerleri ise ortalamanın biraz altında kalan diğer işlemlerin uygulandığı kadın vakalardan oluşmaktadır. Vakaların hepsi GSS’lidir. Oluşan TİG’ler ise tanı amaçlı müdahalelerdir.

Onikinci küme ise ortalama yatış günü düşük, yaşları yüksek, TİG vaka bağıl değerleri ise ortalamanın oldukça altında kalan diğer işlemlerin uygulandığı erkek vakalardan

oluşmaktadır. Vakaların %99'u GSS'lidir. Oluşan TİG'ler ise tanı amaçlı müdahalelerdir. Bu küme gastroskopi TİG'inin ve Göz Retinaya yapılan işlemler TİG'lerinin oluşturduğu kümeler sayılmazsa (1. ve 8. Kümeler) en düşük ortalama TİG vaka bağıl değerine ve yatış gününe sahip kümedir.

3.4. Lojistik Regresyon ve Karar Ağacı Analizleri

Bu çalışmada hastanelerin finansmanında ve yönetiminde önemli bir sorun olan hastanın aşırı yatış süresi (AYS) (length of stay outliers), TİG verileri çerçevesinde incelenecektir. Bu amaçla hastanın hastanede kalış süresini etkileyen faktörler olan cinsiyet, TİG vaka bağıl değerleri, hastanın hastaneden çıkış nedeni (iki grup), hastaya uygulanan işlemin türü, hastanın komorbiditesi olup olmadığı ve yaşı bağımsız değişken olarak alınmıştır. Bazı çalışmalar sosyal güvencenin de hastanın kalış süresini etkilediğini bulmuştur. Fakat bu çalışmada sosyal güvence göz önüne alınmamıştır.

Hastaların hastanede kalış süreleri sağa çarpık bir dağılım gösterdiği için, literatürde geometrik ortalamanın 2 standart sapma fazlasının üstünde yatış süresine sahip hastalar AYS olarak adlandırılmaktadır. Bazı kaynaklar, klasik bilinen anlamda uç değer analizi ile yani 3. kartilden 1.5 kartillerarası aralık fazlası AYS olarak adlandırılmaktadır. Bu çalışmada, aşırı yatış süresi için geometrik ortalama yöntemi kullanılmıştır.

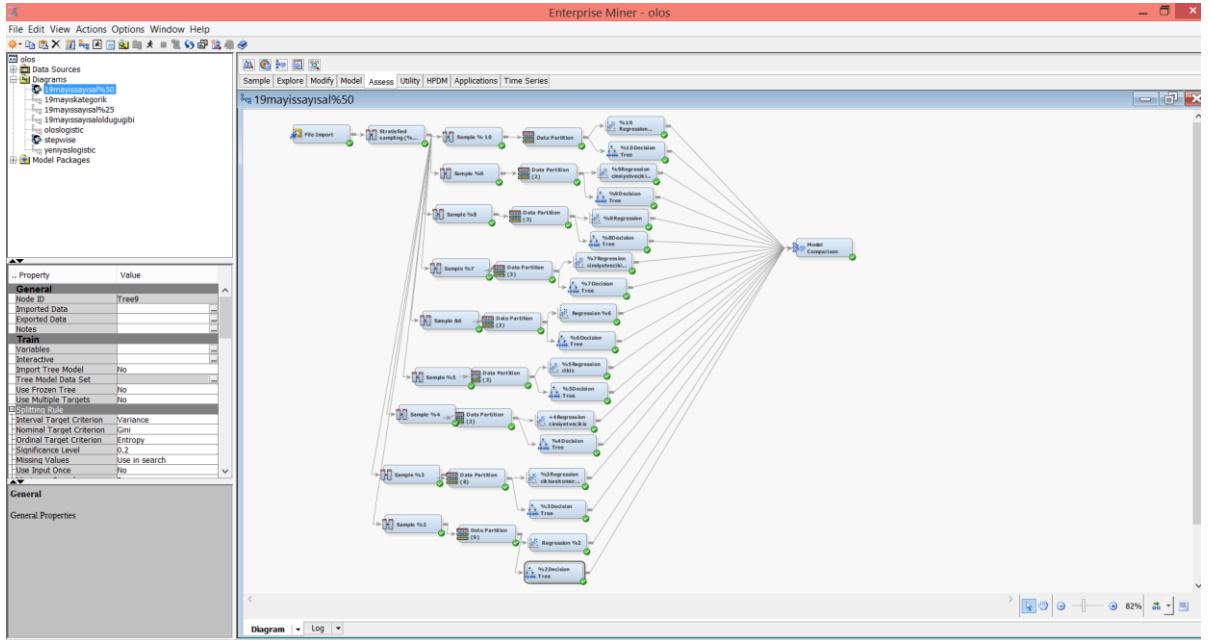
Çalışmada bulunan 15 eğitim araştırma hastanesinden toplanan veriler 693,284 hasta kaydından aşırı yatış süresi (AYS) oranı % 6.8'dir. Hasta yatış günlerinin ortalaması 4 gün bulunmuş ve 13 günden fazla kalma süresi aşırı yatış süresi olarak kabul edilmiştir. Yatış süresi normal (NYS) kabul edilen hastaların (ortalama/medyan) yatış gün sayısı 2.5/2, aşırı yatış günlerinin ise 23.5/18 olarak bulunmuştur.

Lojistik regresyon analizinde örneklem sayısı 10.000¹⁵⁸'in üzerine çıktığında etki büyüklüğünün neredeyse 0 olmasından dolayı p değerleri anlamsız hale gelir. Bu

¹⁵⁸ Mingfeng Lin, Henry C. Lucas Jr, and Galit Shmueli. "Research commentary-too big to fail: large samples and the p-value problem." Information Systems Research 24.4 : 906-917,2013,s.1.

nedenle, analiz yapılmadan önce, her iki model içinde veriden tabakalı örnekleme yöntemi ile 10.000'den az örneklemler alınarak analiz gerçekleştirilmiştir. Fakat, veride var olan AYS oranının % 6,8 olmasından dolayı kurulan modeller NYS'yi tahmin etmekte başarılı fakat, AYS'yi tahmin etmekte başarısız olmuştur. Modelin tahmin gücünü arttırmak için yine tabakalı örnekleme yöntemi ile 10.000'den az örneklemler seçilmiş fakat örneklemdaki AYS ve NYS oranları % 50-%50 eşit alınmıştır. Bu şekilde kurulan modeller, her iki oranı tahmin etmekte başarılı olmuşlar ve bu sonuçlar yorumlanmıştır.

Analizler, tabakalı örnekleme yöntemi ile AYS-NYS oranı %50-%50 olan, 9436 örneklemden % 40 eğitim, % 30 onay ve % 30 test kümesi seçilerek gerçekleştirilmiştir. AYS 1, NYS ise 0 olarak kodlanmıştır. Yaş ve bağıl değer değişkenleri sayısal olarak modele alınmış, diğer değişkenler ise kategorik olarak modele alınmıştır.



Şekil 3-4: Lojistik Regresyon ve Karar Ağacı SAS Enterprise Miner Uygulaması

3.4.1. Lojistik Regresyon Modeli

Lojistik Regresyon analizi sonucunda oluşan model aşağıdaki tabloda gösterilmiştir. Model sonucunda, hastaların hastaneden çıkış nedenleri anlamlı bulunmamıştır.

Tablo 3-10: AYS'yi etkileyen faktörlerin lojistik regresyon ile incelenmesi

Değişken	Grup	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Odds	95% Güven Aralığı	
Intercept	1	-1.933	0.1343	207.14	<.0001	0.15	0.11	0.19
Cinsiyet	Kadın	-0.1262	0.0377	11.2	0.0008	0.88	0.82	0.95
Yaş	-	0.0163	0.00167	95.2	<.0001	1.02	1.01	1.02
Bağıl Değer	-	0.5247	0.0301	304.77	<.0001	1.69	1.59	1.79
İşlem Sınıfları	Cerrahi	-0.4239	0.0724	34.33	<.0001	0.65	0.57	0.75
	Dahili	0.8041	0.0677	140.9	<.0001	2.24	1.96	2.55
Komorbidite	Var	0.1544	0.0609	6.42	0.0113	1.17	1.04	1.31

Lojistik regresyon ile kurulan modele göre, kadın hastaların AYS kalma olasılığı, erkek hastaların AYS kalma olasılığına göre 0,88 kat daha azdır. Hastaların yaşı ve uygulanan tedavinin bağıl değeri arttıkça, AYS kalma olasılığı da artmaktadır. Hastaya cerrahi tedavi uygulandığında, diğer tedavi yöntemlerine göre AYS kalma olasılığı 0,65 kat azalmaktadır. Hastaya dahili tedaviler uygulandığında ise, hastanın AYS kalma olasılığı diğer yöntemler uygulandığındaki AYS kalma olasılığından 2,24 kat daha fazladır. Komorbiditesi olan bir hastanın, olmayan bir hastaya göre AYS kalma olasılığı 1.17 kat daha fazladır.

Kurulan model ile eğitim kümesi % 73.7, onay kümesi ise % 72.8 oranında doğru tahmin yapmıştır.

Tablo 3-11: Lojistik Regresyon Sınıflandırma Tabloları

Eğitim Kümesi	AYS (Tahmin)	
AYS (Gerçek)	EVET	HAYIR
EVET	76.4%	28.4%
HAYIR	23.6%	71.6%
Doğru Sınıflandırma Yüzdesi: %73.7		

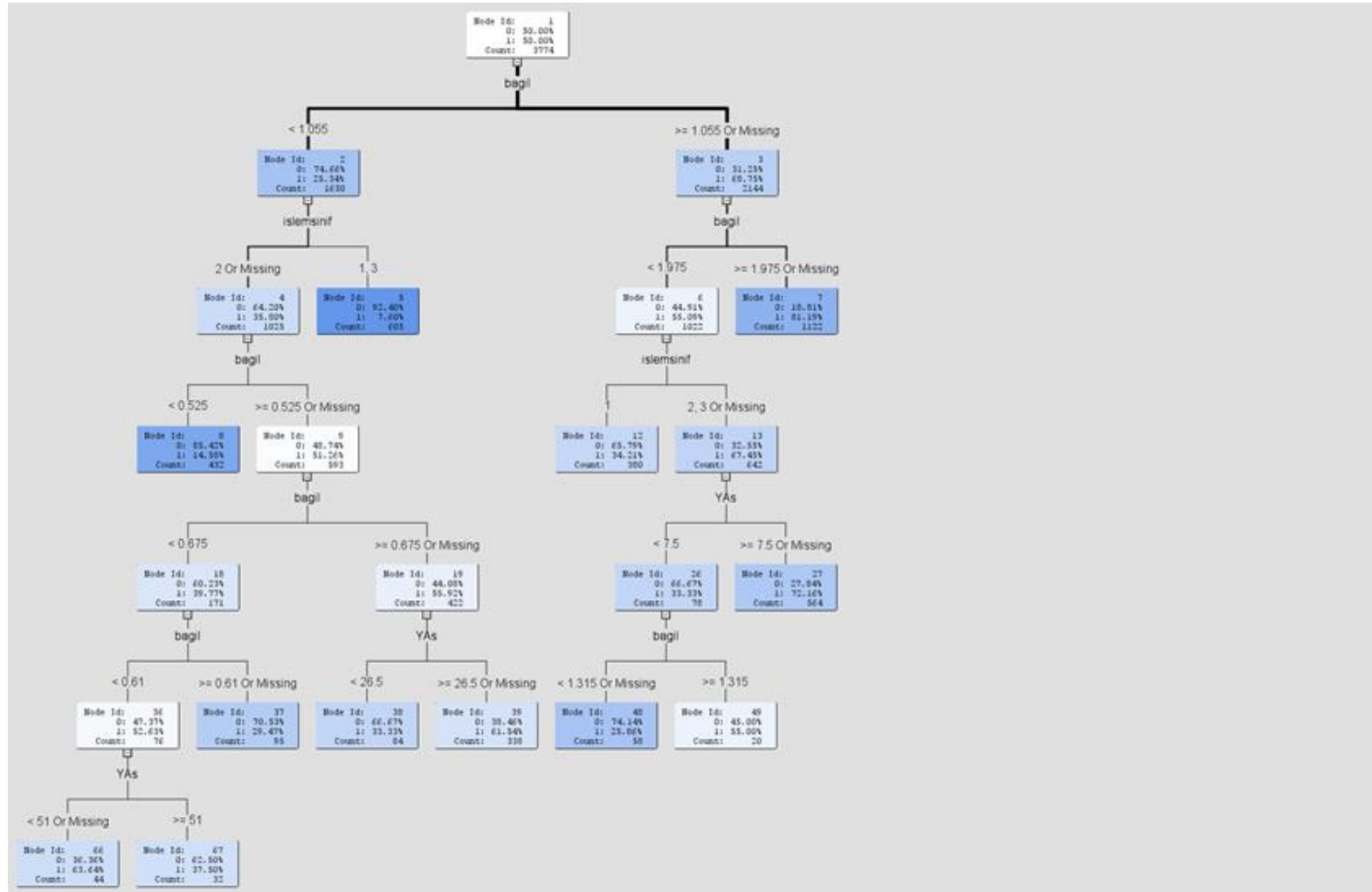
Onay Kümesi	AYS (Tahmin)	
AYS (Gerçek)	EVET	HAYIR
EVET	75.4%	29.3%
HAYIR	24.6%	70.7%
Doğru Sınıflandırma Yüzdesi: %72.8		

3.4.2. Karar Ağacı Modeli

Aynı değişkenlerle, AYS'yi tahmin etmek için CR&T Karar Ağacı ile model kurulduğunda çıkış nedeni ve cinsiyet modele alınmamıştır. Bağlı değer, modeldeki en önemli ayırıcı değişken olarak bulunmuştur. Kurulan model ile değişkenlerin önemi şöyledir:

Tablo 3-12: Karar Ağacı Analizinde Değişkenlerin Önem Tabloları

Değişkenler	Training Importance	Validation Importance
Bağlı Değer	1.00	1.00
İşlemler	0.79	0.49
Yaş	0.79	0.27
Komorbidite	0.26	0.07
Çıkış Nedeni	0.00	0.00
Cinsiyet	0.00	0.00



Şekil 3-5:C&R Karar Ağacı Modelinin Gösterimi

Eğitim ve doğrulama verisi ile elde edilen karar ağaçları 6 seviyeden ve 12 uç düğümden oluşmaktadır.

Tablo 3-13: Karar Ağacı İstatistikleri

Node	Seviye (Depth)	Eğitim Gözlem Sayısı	Eğitim % AYS
7	2	1122	0.81
5	2	605	0.08
27	4	564	0.72
8	3	432	0.15
12	3	380	0.34
39	5	338	0.62
37	5	95	0.29
38	5	84	0.33
48	5	58	0.26
66	6	44	0.64
67	6	32	0.38
49	5	20	0.55

SAS Enterprise Miner karar ağacı oluştururken düğümden bulunan örneklem sayısına göre düğümün rengini değiştirmektedir. Rengin koyu olması fazla örneklem sayısına, açık olması ise düşük örneklem sayısına işaret eder.

Daha önce de belirtildiği gibi, en başarılı model AYS-NYS oranının %50-%50 olduğu durumda elde edilmiştir ve ağaç bu şekilde oluşmuştur. Elde edilen kurallar şöyledir:

- Hastaya uygulanan işlemin bağıl değeri 1.975'in üzerinde olan hastaların AYS kalma oranı % 81'dir.
- Hastaya bağıl değeri 1.055'den küçük olan cerrahi ve diğer işlemlerim uygulandığı durumlarda AYS kalma oranı % 8'dir.

- Yaşı 7.5'den büyük ve bağıl değeri 1.975'den küçük olan dahili ve diğer işlemlerin uygulandığı hastaların AYS kalma oranı %72 'dir.
- Bağıl değeri 0.525'den küçük dahili işlemlerin uygulandığı hastaların AYS kalma oranı % 15'dir.
- Bağıl değeri 1.975'den küçük cerrahi işlemlerin uygulandığı hastaların AYS kalma oranı % 34'dür.
- Yaşı 26.5'den büyük olan bağıl değeri 0.675'den büyük dahili işlemlerin uygulandığı hastaların AYS kalma oranı % 62' dir.
- Bağıl değeri 0.61'den büyük dahili işlemlerin uygulandığı hastaların AYS kalma oranı % 29 'dur.
- Yaşı 26.5'den küçük olan bağıl değeri 0.675'den büyük dahili işlemlerin uygulandığı hastaların AYS kalma oranı % 33' dür.
- Yaşı 7.5'den küçük ve bağıl değeri 1.32'den küçük olan dahili ve diğer işlemlerin uygulandığı hastaların AYS kalma oranı % 26 ' dir.
- Yaşı 51'den küçük olan bağıl değeri 0.61'den küçük dahili işlemlerin uygulandığı hastaların AYS kalma oranı % 64' dür.
- Yaşı 51'den büyük olan bağıl değeri 0.61'den küçük dahili işlemlerin uygulandığı hastaların AYS kalma oranı % 38 'dir.
- Yaşı 7.5'den küçük ve bağıl değeri 1.32'den büyük olan dahili ve diğer işlemlerin uygulandığı hastaların AYS kalma oranı % 55 ' dir.

Tablo 3-14: Karar Ağacı Sınıflandırma Durumları

Eğitim Kümesi	AYS (tahmin)	
AYS (gerçek)	EVET	HAYIR
EVET	74.95%	19.10%
HAYIR	25.05%	80.90%
Doğru Sınıflandırma Yüzdesi: %77.6		

Onay Kümesi	AYS (tahmin)	
AYS (gerçek)	EVET	HAYIR
EVET	73.7%	20.9%
HAYIR	26.3%	79.1%
Doğru Sınıflandırma Yüzdesi: %76.1		

3.4.3. Lojistik Regresyon ve Karar Ağacı Modellerinin Karşılaştırılması

Modellerin etkinliğinin karşılaştırmasında ROC Analizinin yanında lift (yükselme) ve response (yanıt) grafiklerinden yararlanılmıştır. ROC analizinin yanında lift ve response grafiklerinin kullanılmasının nedeni, ROC analizinin her zaman yeterli olmamasıdır.

ROC analizinde, grafikler modellerin başarısının göstermesine rağmen bazı rakamlara da bakmak gerekmektedir. Bu çalışmada, ROC analizinden elde edilen, standart sapma, yanlış sınıflandırma hatası, ROC eğrisinin altında kalan alan ve Kolmogorov Smirnov istatistiği (1- özgüllük (specificity)) incelenecektir.

Tablo 3-15: Eğitim Veri Kümesinin Performansının Kıyaslanması

Model	Standart Sapma	Yanlış Sınıflandırma Hatası	ROC Alan	K-S İstatistik
Lojistik Regresyon	0.18	0.26	0.81	0.48
Karar Ağacı	0.16	0.22	0.83	0.55

Tablo 3-16: Test Veri Kümesinin Performansının Kıyaslanması

Model	Standart Sapma	Yanlış Sınıflandırma Hatası	ROC Alan	K-S İstatistik
Lojistik Regresyon	0.18	0.27	0.81	0.47
Karar Ağacı	0.17	0.25	0.81	0.51

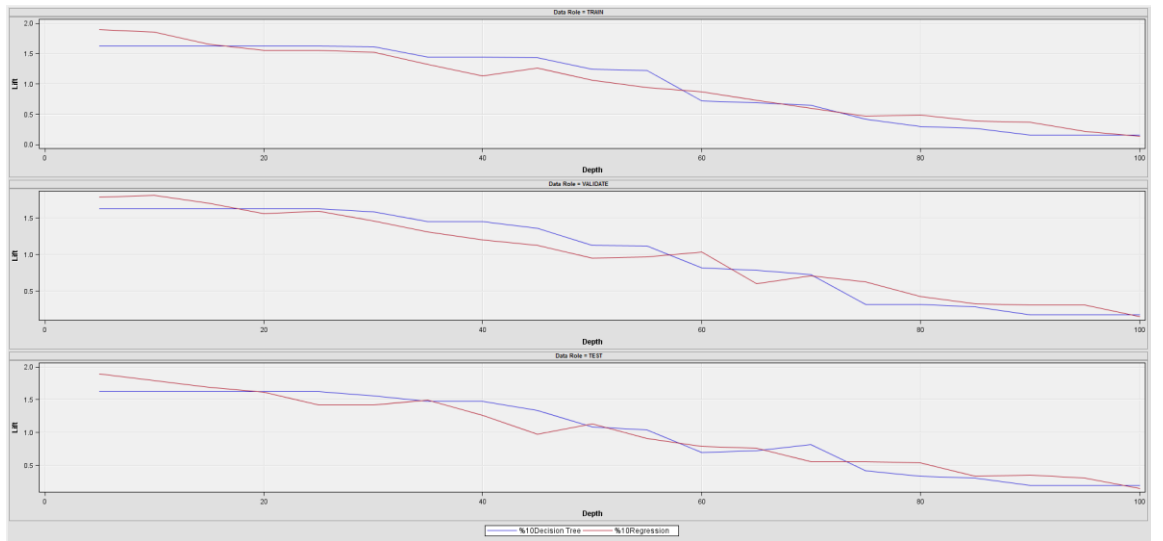
Lift (Kaldıraç, yükselme) değeri ve lift grafikleri modelin etkinliğinin önemli bir ölçüsüdür. Modelden ve model olmadan elde edilen sonuçların oranıdır. Tabloda görüldüğü gibi, veri seti desil olarak ifade edilen 20 eşit gruba ayrılır ve ‘Lift’ tablosunda başlangıç desilleri ne kadar yüksek yanıtlama oranına sahip ise, modelin tahmin performansı o kadar güçlüdür.

Örneğin modelin lift değerinin 3 olması bu aralıkta kalan hedef değerlerini, veriden alınacak rastgele bir gruba göre 3 kat daha iyi ve doğru tahmin ettiğinin göstergesidir. Bir modelin iyi olabilmesi için lift grafiği sol tarafta % 100’ün üzerinde başlar, ve yüksek bir düz çizgi halinde devam eder ve sağ taraftaki % 100’de hızlı bir şekilde düşer.

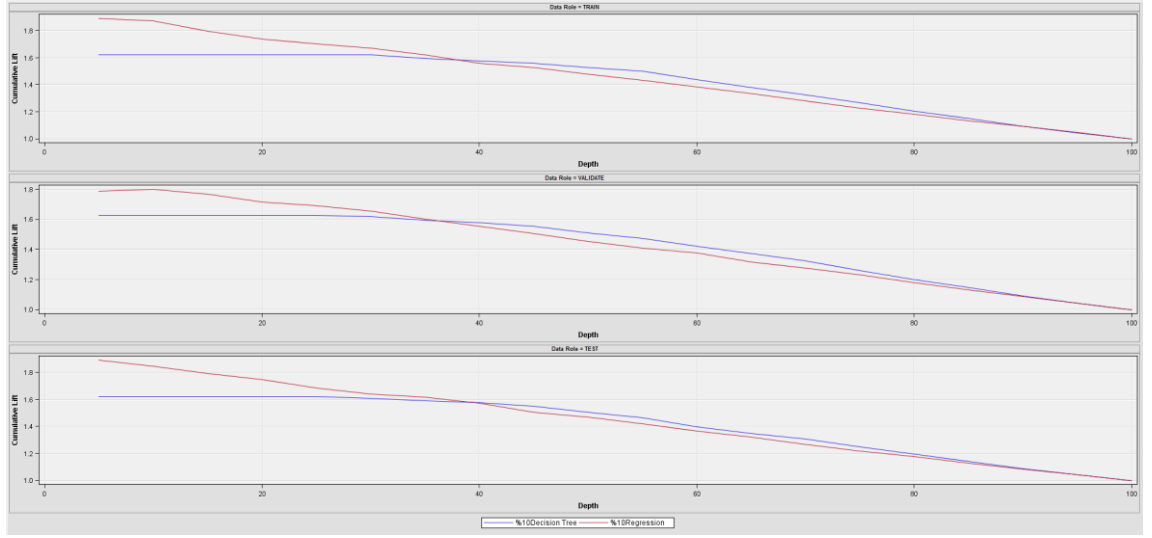
Burada, lojistik regresyon ile elde edilen en iyi modelin 1.89 kat, karar ağaçları ile 1.62 kat daha iyi tahmin ettiği bulunmuştur. Lift grafiğinde lojistik regresyonun, kümülatif Lift grafiğinde ise karar ağacının daha iyi sonuç verdiği görülmektedir.

Tablo 3-17: Eğitim Kümelerinin Lift Değerleri

Desil	Lojistik Regresyon		Karar Ağaçları	
	Lift	Cumulative Lift	Lift	Cumulative Lift
5	1.89	1.89	1.62	1.62
10	1.85	1.87	1.62	1.62
15	1.65	1.80	1.62	1.62
20	1.55	1.74	1.62	1.62
25	1.56	1.70	1.62	1.62
30	1.52	1.67	1.61	1.62
35	1.32	1.62	1.44	1.60
40	1.13	1.56	1.44	1.58
45	1.26	1.53	1.43	1.56
50	1.06	1.48	1.24	1.53
55	0.94	1.43	1.23	1.50
60	0.87	1.38	0.72	1.44
65	0.73	1.33	0.68	1.38
70	0.60	1.28	0.65	1.33
75	0.47	1.23	0.42	1.27
80	0.49	1.18	0.29	1.20
85	0.38	1.13	0.26	1.15
90	0.37	1.09	0.15	1.09
95	0.21	1.05	0.15	1.04
100	0.14	1.00	0.15	1.00



Şekil 3-6: Modellerin Lift Grafikleri



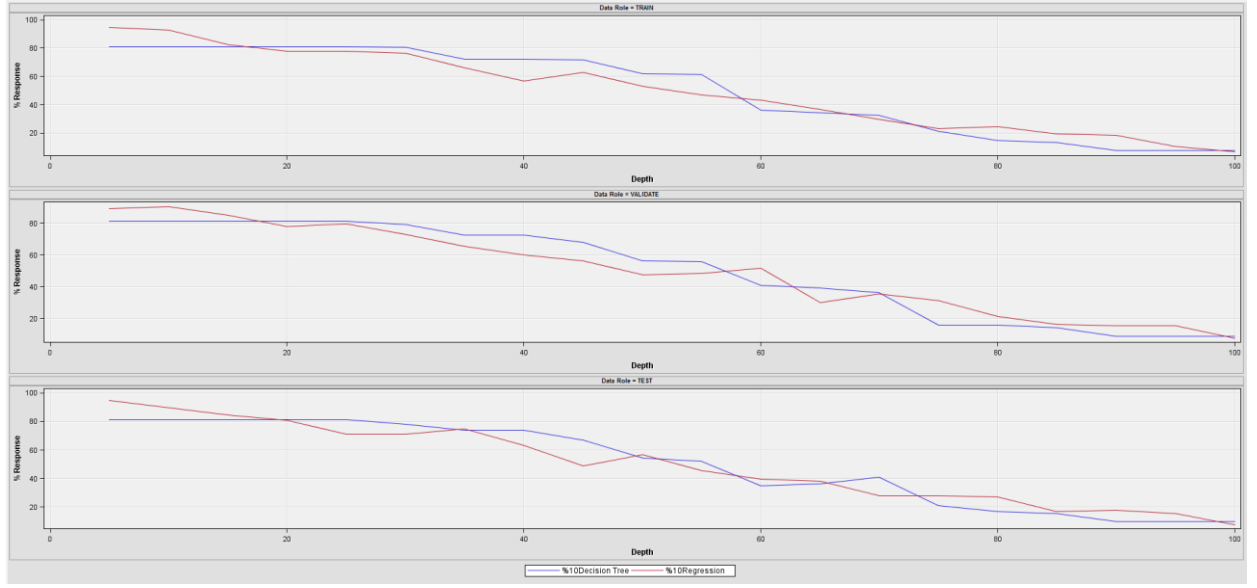
Şekil 3-7: Mdellerin Kümülatif Lift Grafikleri

Yanıt (response ve cumulative response): Her bir yüzdelik dilimdeki doğru tahmin edilen kayıtların oranıdır. %100 oranına yakın bir noktadan başlaması beklenir ve dereceli olarak azalır. İyi bir model için, doğru % 100 noktasına yakın bir yerden başlar ve grafiğin ilerleyen düzeylerinde düşüş gösterir.

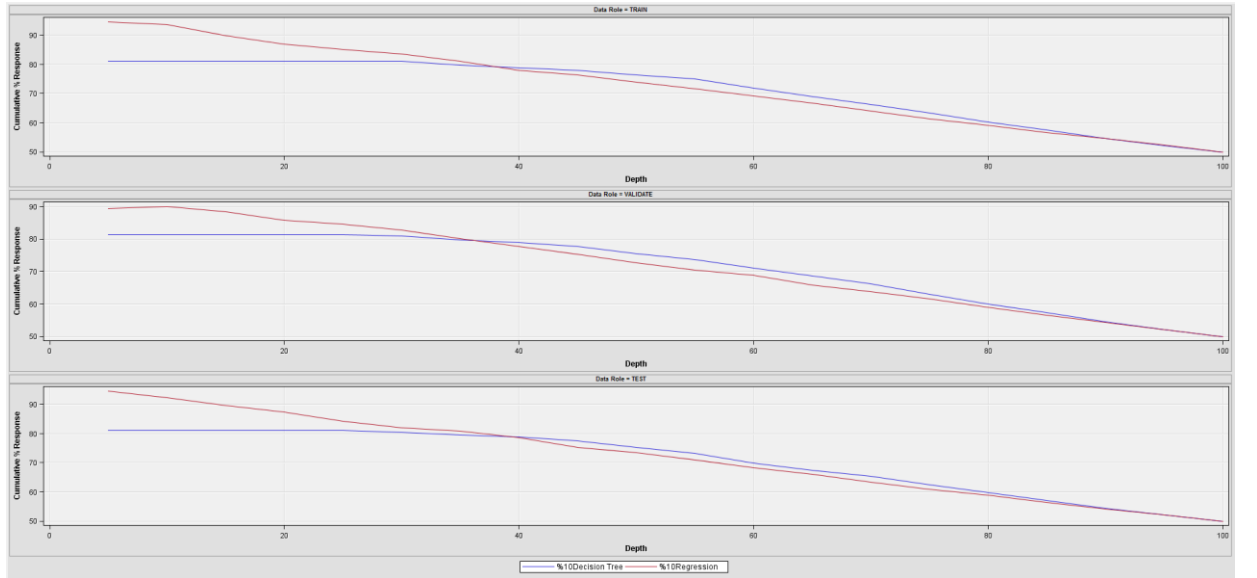
Modellerin response grafikleri incelendiğinde, response grafiğine göre lojistik regresyonun, kümülatif response grafiğine göre ise karar ağacının daha iyi sonuç verdiği görülmektedir.

Tablo 3-18: Eğitim Kümesinin Response Değerleri

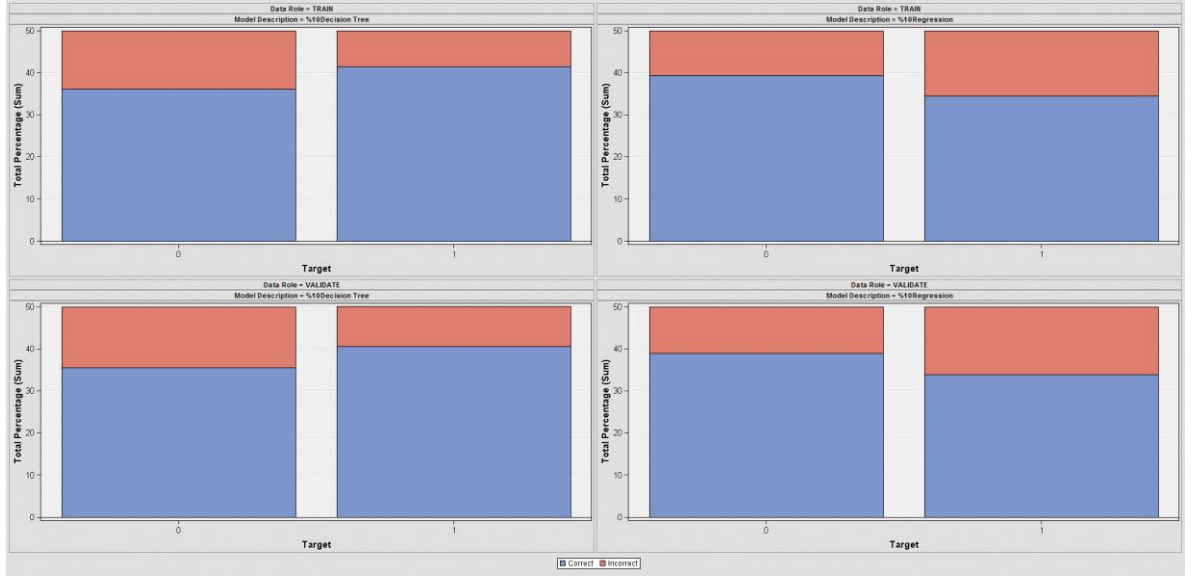
Desil	Lojistik Regresyon		Karar Ağaçları	
	%Response	Cumulative % Response	%Response	Cumulative % Response
5	94.71	94.71	81.19	81.19
10	92.59	93.65	81.19	81.19
15	82.54	89.95	81.19	81.19
20	77.66	86.89	81.19	81.19
25	77.78	85.06	81.19	81.19
30	76.19	83.58	80.67	81.11
35	65.96	81.07	72.16	79.83
40	56.61	78.01	72.16	78.87
45	62.96	76.34	71.58	78.06
50	53.19	74.03	61.88	76.45
55	47.09	71.58	61.26	75.07
60	43.39	69.23	36.09	71.81
65	36.51	66.71	34.21	68.92
70	29.79	64.08	32.36	66.32
75	23.28	61.36	20.96	63.29
80	24.34	59.04	14.58	60.24
85	19.15	56.70	13.14	57.48
90	18.52	54.58	7.60	54.71
95	10.58	52.26	7.60	52.22
100	6.91	50.00	7.60	50.00



Şekil 3-8: Modellerin Response Grafikleri



Şekil 3-9: Modellerin Kümülatif Response Grafikleri



Şekil 3-10: Eğitim ve Onay Kümelerinin Sınıflandırma Grafikleri

Lojistik regresyon ve karar ağacı modellerinin sınıflandırma yüzdeleri de incelendiğinde, karar ağaçlarının AYS'yi tahmin etmekte, lojistik regresyondan daha başarılı olduğu görülmektedir.

Sonuç olarak, karar ağacı modelinin, lojistik regresyon modelinden daha başarılı olduğu görülmektedir.

3.5. Birliktelik Analizi

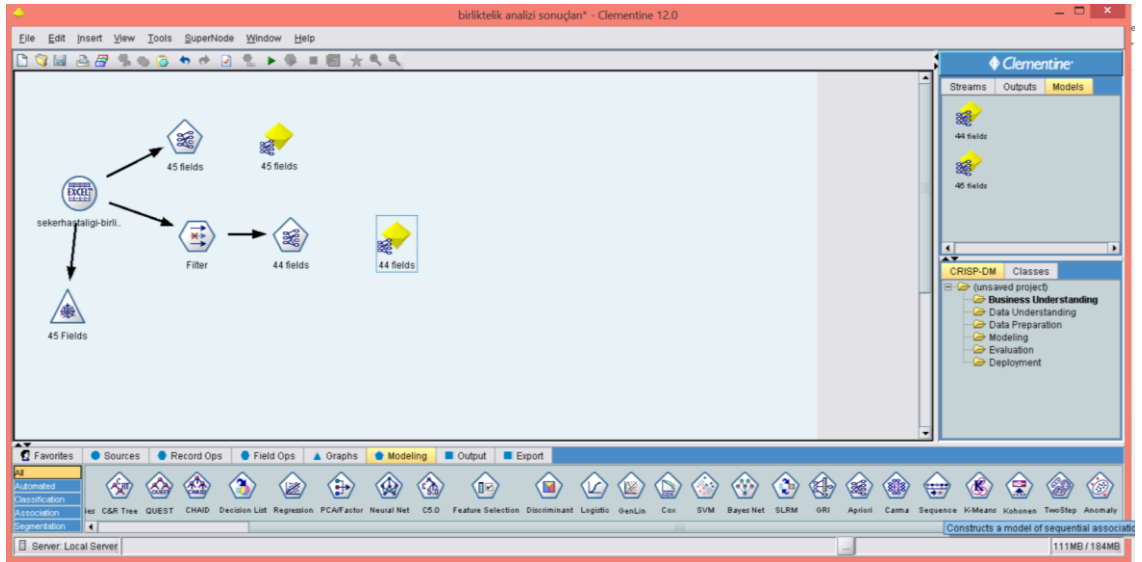
15 eğitim araştırma hastanesinde yatan diyabet hastalarının TİG çerçevesinde birliktelik kurallarının belirlenebilmesi apriori algoritmasından yararlanılmıştır. Kural oluşturulurken, mümkün tüm kuralları görebilmek için minimum destek değeri %10, minimum güven değeri ise %55 olarak belirlenmiştir. Kurallar lift değerlerine göre sıralanmış ve 917 kural elde edilmiştir.

Oluşan TİG Diyabet kodu sayısı 5347'dir. Diyabet kodları K01Z, K60A, K60B olmak üzere 3 tanedir. Analize kodların kendisi alınmıştır. Fakat işlem türü ve bağlı değerleri alınmamıştır.

Tablo 3-19: Diyabet Kodları

Kodlar	İşlem	Bağıl Değer	TİG Kod Açıklama
K01Z	Cerrahi	Yüksek	Ayak, Diyabetik Ayak İşlemleri
K60A	Dahiliye	Yüksek	Diyabet, Katastrofik/Şiddetli KK (Komplikasyon / Komorbidite) Bulunan,
K60B	Dahiliye	Orta	Diyabet, Katastrofik/Şiddetli KK Bulunmayan

TİG kodlarını yorumlarken, K01Z vakalarının en yüksek bağıl değere sahip, K60B vakalarının ise yeni teşhis konulan komorbiditesi bulunmayan, daha genç insanlarda görülmesi olası TİG vakasıdır.



Şekil 3-11: Birliktelik Analizi SPSS Clementine

15 Eğitim Araştırma Hastanesinde bulunan diyabet hastaları ile gerçekleştirilen birliktelik analizi sonucunda 317 kural bulunmuştur. Burada sadece en yüksek lift (kaldırma oranı) değerine sahip ilk 50 kural, destek, güven ve kaldırma oranı değerleriyle birlikte gösterilmiştir.

Bu 15 eğitim araştırma hastanesinde, K01Z Diyabet vakası olup şifa ile taburcu olanların % 72.3'ü erkektir. Bu vakalar, 15 eğitim araştırma hastanelerinde diyabet tedavisi gören yatan hastaların %11.5'ini oluşturmaktadır. Bu birliktelik kuralının lift oranı ise 1.52'dir. Bu değer öncül ögenin olduğu durumların ardıl ögenin gerçekleşme

olasılığını ne kadar arttırdığının göstergesi olarak kabul edilmektedir. Kaldırma oranı bir birliktelik kuralının önem derecesinin göstergelerinden biri olarak kabul edilebilir. Değerin 1'e yakın olduğu durumlarda kuralın önem derecesinin düşük olduğu, 1'den küçük olduğu durumlarda ardıl ögenin gerçekleşme olasılığına negatif katkısı olduğu, 1'den büyük olduğu durumlarda ise bu olasılığa pozitif katkısı olduğu söylenebilir. Dolayısıyla bu kural için 1.52 kaldırma oranı değeri bu kuralın önemini göstermesinin yanında K01Z şifa ile taburcu olan diyabet vakalarının erkek olma olasılığını yükselttiğini göstermektedir.

Hastanede, K60A diyabet vakalarının % 70'i hastanede 5 günden fazla sürede tedavi olmuşlardır. Bu hastalar tüm hastaların % 12.8'ini oluşturmaktadır. 1-5 gün yatan ve şifa ile taburcu olan kadın hastaların %88.5'i K60B vakalarıdır. Bu vakalar, tüm vakaların % 13'ünü oluşturmaktadır. Kuralların lift oranları yine 1'in üstündedir.

Tablo 3-20: Diyabet Hastaları Birliktelik Analizi Sonuçları

Kural	Ardıl Öge (Consequent)	Öncül Öge (Antecedent)	Destek % (Support)	Güven (%) (Confidence)	Lift (Kaldırma Oranı)
1	Erkek	K01Z veŞifa ile Taburcu	11.49	72.31	1.52
2	Erkek	K01Z veŞifa ile Taburcu ve sgk	10.74	72.30	1.52
3	Erkek	K01Z ve sgk	11.95	71.83	1.51
4	Erkek	K01Z	12.87	71.80	1.51
5	Şgunustu	K60A ve sgk	12.16	70.46	1.26
6	Şgunustu	K60A	12.76	70.23	1.25
7	K60B	1-Şgun ve Kadın veŞifa ile Taburcu	12.96	88.46	1.19
8	K60B	1-Şgun ve Kadın veŞifa ile Taburcu ve sgk	12.16	88.15	1.19
9	Kadın	K60A ve sgk	12.16	61.85	1.18
10	Kadın	K60A	12.76	61.73	1.18
11	Kadın	yas51-65 ve Şgunustu ve K60B veŞifa ile Taburcu ve sgk	11.07	61.49	1.17
12	K60B	1-Şgun ve Kadın	16.39	86.19	1.16
13	K60B	1-Şgun ve Kadın ve sgk	15.34	86.10	1.16
14	Kadın	yas51-65 ve Şgunustu ve K60B ve sgk	13.28	60.56	1.15
15	Kadın	yas51-65 ve Şgunustu ve K60B veŞifa ile Taburcu	11.77	60.41	1.15
16	Kadın	yas65+ ve K60B veŞifa ile Taburcu	15.32	60.20	1.15
17	Kadın	yas65+ ve K60B veŞifa ile Taburcu ve sgk	14.63	60.10	1.14
18	K60B	1-Şgun veŞifa ile Taburcu ve sgk	23.94	84.77	1.14
19	K60B	1-Şgun veŞifa ile Taburcu	25.74	84.74	1.14
20	K60B	1-Şgun ve yas51-65 ve sgk	10.03	84.51	1.14
21	Kadın	yas51-65 ve K60B veŞifa ile Taburcu ve sgk	19.81	59.68	1.14
22	Kadın	yas51-65 ve Şgunustu ve K60B	14.09	59.50	1.13
23	Kadın	yas65+ ve K60B	19.12	59.39	1.13
24	Kadın	yas65+ ve Şgunustu veŞifa ile Taburcu	14.18	59.37	1.13
25	Kadın	yas51-65 ve K60B ve sgk	24.02	59.35	1.13
26	K60B	1-Şgun ve sgk	29.29	83.84	1.13
27	Kadın	yas51-65 ve K60B veŞifa ile Taburcu	20.82	59.21	1.13
28	Kadın	yas65+ ve Şgunustu ve K60B	10.72	59.16	1.13
29	Kadın	yas65+ ve Şgunustu veŞifa ile Taburcu ve sgk	13.67	59.10	1.13
30	K60B	1-Şgun	31.44	83.64	1.12
31	K60B	1-Şgun ve yas51-65	10.49	83.60	1.12
32	Kadın	yas65+ ve K60B ve sgk	18.18	58.85	1.12
33	Kadın	yas51-65 ve K60B	25.20	58.80	1.12
34	Kadın	yas65+ ve Şgunustu ve K60B ve sgk	10.19	58.72	1.12
35	K60B	yas41-50 veŞifa ile Taburcu ve sgk	10.27	82.51	1.11
36	Şgunustu	yas65+ ve Kadın veŞifa ile Taburcu ve sgk	13.06	61.89	1.10
37	Şifailetaburcu_	K01Z ve sgk	11.95	89.83	1.10
38	Şgunustu	yas65+ ve Kadın veŞifa ile Taburcu	13.66	61.64	1.10
39	Kadın	yas65+ ve Şgunustu	17.49	57.75	1.10
40	K60B	yas41-50 veŞifa ile Taburcu	11.28	81.76	1.10
41	K60B	yas41-50 ve sgk	11.97	81.56	1.10
42	Şifailetaburcu_	K01Z	12.87	89.24	1.10
43	K60B	1-Şgun ve Erkek ve sgk	13.95	81.37	1.09
44	K60B	1-Şgun ve Erkek veŞifa ile Taburcu ve sgk	11.78	81.27	1.09
45	Kadın	Şgunustu ve K60B veŞifa ile Taburcu ve sgk	30.96	57.22	1.09
46	K60B	1-Şgun ve Erkek veŞifa ile Taburcu	12.78	80.97	1.09
47	K60B	yas41-50	13.24	80.93	1.09
48	K60B	1-Şgun ve Erkek	15.06	80.87	1.09
49	Kadın	yas65+ ve Şgunustu ve sgk	16.78	57.08	1.09
50	Kadın	yas51-65 ve Şgunustu veŞifa ile Taburcu ve sgk	15.77	56.82	1.08

Eğitim ve araştırma hastanelerinde tedavi gören hastaların %90'ı Genel sağlık sigortalı hastalardır. Kümeleme analizi sonuçlarına göre sosyal güvencenin TİG vakalarına göre farklılık gösterdiği ortaya çıkmıştır. Bu nedenle, genel sağlık sigortalı hastaları dışarıda tutularak, diyabet hastalarına tekrar birliktelik kuralı uygulanmıştır ve en yüksek lift oranına göre ilk 10 kural gösterilmiştir. Genel sağlık sigortalı olmayan hastalar için, 2012 yılı çalışmada bulunan eğitim araştırma hastanelerinde 359 TİG kodu oluşmuştur.

Hastane 3'de tedavi gören K60B vakalarının % 80'inin hastanedeki servisleri değişmiştir ve bu hastalar tüm vakaların % 5'ini oluşturmaktadır. Lift oranı ise 8.54 ile çok yüksektir. Hastane 2'de tedavi gören diyabet vakalarının % 83'ü hastanede 5 günden fazla sürede tedavi görmüştür. Bu vakalar tüm vakaların % 6'sını oluşturmaktadır. Yaşları 21-30 arasında değişen diyabet vakaları % 95 K60B vakasıdır. Bu vakalar tüm vakaların % 6 'sını oluşturmaktadır.

Tablo 3-21: GSS Olmayan Diyabet Hastalarının Birliktelik Kuralları

Kural	Ardıl Öge (Consequent)	Öncül Öge (Antecedent)	Destek % (Support)	Güven (%) (Confidence)	Lift (Kaldırma Oranı)
1	Servis Değişimi	hastane3 and K60B	5.24	80.00	8.54
2	Servis Değişimi	hastane3	5.85	78.59	8.39
3	Şgunustu	hastane2	6.19	83.08	1.48
4	Şgunustu	hastane2 and Şifa	6.04	82.97	1.48
5	Şgunustu	K60A and Kadın and Şifa	5.85	78.91	1.41
6	Şgunustu	hastane3	5.85	77.00	1.37
7	Şgunustu	hastane3 and K60B	5.24	76.43	1.36
8	Şgunustu	K60A and Şifa	9.41	75.94	1.36
9	K60B	yas21-30	5.74	95.44	1.28
10	K60B	hastane9 and Şifa	5.24	92.50	1.24

4. SONUÇLAR

Bu çalışmada, son yıllarda toplanan yığınla veriden anlamlı bilgiler üretmek için kullanılan veri madenciliği tekniği sağlık hizmetleri alanında kullanılmıştır. Son yıllarda, veri tabanları teknolojisinin gelişimi ve geleneksel yöntemlerden farklı yöntemlerin kullanım ihtiyacı ile veri madenciliği gibi gelişmiş tekniklerin sağlık alanında kullanımı sağlık analitiği olarak adlandırılmaktadır.

Veri madenciliği, sağlık analitiği adı altında sağlık hizmetlerinde kullanılarak, sağlık politikalarına ve sağlık alanındaki karar vericilerine çok önemli katkılar sağlamaktadır.

Uluslararası boyutta sağlık hizmetlerinin geri ödemesi en önemli sağlık analitiği çalışma alanlarından bir tanesi olmuştur. Ülkemizde ve dünyada geri ödeme modelleri ile sağlıkta azalan kaynakların verimli kullanılması ve kalitenin artırılması amaçlanmaktadır. Ülkemizde 2005 yılında pilot çalışmalarla başlayan ve bir yatan hasta sınıflandırma sistemi ve geri ödeme modeli olan TİG bu çalışmanın konusunu oluşturmaktadır.

Çalışmada kullanılan veriler, 2012 yılında İstanbul, Ankara ve İzmir'deki 15 eğitim araştırma hastanesinde oluşan 693,289 TİG verileridir. Çalışmada 693,289 hasta kaydı bulunmaktadır. Veri setinde, hastaların cinsiyeti, yaşı, yatış tarihi, hastanede kalış süreleri, sosyal güvenceleri, hastaneden taburcu olma nedenleri, TİG'in bağlı değeri, hastanın komorbiditeye sahip olup olmadığı ve hastaya uygulanan işlemin türü bulunmaktadır.

Aynı TİG'e sahip olan hastaların klinik açıdan benzer hastalık tipleri göstermeleri ve aynı miktarda kaynak tükettikleri kabul edilmektedir. Bu kaynak kullanımı da hastanın hastanede kalış süresi ile ilişkilendirilmektedir. Çünkü TİG uygulamasında, aynı TİG vaka koduna sahip hastalar için aynı geri ödeme yapılır, ödemeler hastanın hastanede kalış süresinden bağımsızdır. Ayrıca, hastanelerin yapısal değişikliğine göre (hastaların kompleksliği, insan kaynakları, alt yapı...vb.) ek ödemeler eklenmektedir.

Çalışmada sadece belirtilen verilerle çalışılmış, fakat TİG'lerin oluşmasında çok büyük öneme sahip ICD kodları, ACHI işlem kodları ile çalışılamamıştır. Bunun yanında, hastanın bulunduğu servis ile hastanın tedavisinin acil serviste başlayıp başlamadığı gibi finansal açıdan çok önemli veriler çalışmada kullanılamamıştır. Ayrıca çalışmada sadece eğitim araştırma hastaneleri ile çalışıldığından dolayı, devlet ve özel hastanelerdeki durum karşılaştırılmaları yapılamamıştır. Bunlar araştırmanın kısıtlarını oluşturmaktadır.

Oluşan TİG'lerin profillerini araştırmak amacıyla kümeleme analizi, TİG'lere göre AYS'ni etkileyen faktörleri görebilmek ve değişkenlerin AYS'yi nasıl etkilediğini görebilmek için karar ağacı ve lojistik regresyon uygulanmıştır. Ayrıca, yine sağlık politikalarına katkı sağlayacağına inanılan birliktelik analizi diyabet hastalarına uygulanarak, bu hastalarda oluşan benzer durumlar incelenmeye çalışılmıştır.

Ayrıca literatürde TİG verilerinin veri madenciliği tekniklerine çok uygun olmayabildiği ve bu nedenle yorumlanmasının oldukça zor olduğu görülmektedir. Bu nedenle çalışmada elde edilen bulgular dikkatle incelenmiştir.

Kümeleme analizinde kümeler oluşturulurken pek çok yöntem denenmiştir ve en doğru sonuçlar bir yapay sinir ağı uygulaması olan kohonen ağları ile elde edilmiş ve 12 küme oluşmuştur. Sonuç olarak, TİG'ler cinsiyete göre net bir şekilde ayrılmıştır. Kadınlarda ve erkeklerde en yüksek yaş ortalaması ile en yüksek yatış süresi ortalaması ve TİG değerlerinin maliyetlerini gösteren en yüksek bağıl değer ortalamasına sahip kümeler aynı çıkmıştır. Diğer oluşan kümelerde yüksek yaş ortalamasına sahip olup, düşük bağıl değere sahip hastaların olduğu kümelere bakıldığında, oluşan TİG'lerin günü birlik basit işlemlerden oluşan kümeler olduğu görülmüştür. Ayrıca, hastaların çoğunluğu genel sağlık sigortasına sahipken, bazı kümelerde genel sağlık sigortasına sahip olmayan hastaların olduğu görülmüştür. Bu hastalara ait oluşan TİG'ler incelendiğinde, oluşan TİG'lerin çoğunlukla acil durumlarda oluşan TİG'ler olduğu görülmüştür.

Kümeleme analizi sonucunda, kümelerdeki hastanın yatış süresi ile TİG vaka bağıl değeri paralellik göstermiştir. Buna göre, 2012 yılında eğitim-araştırma hastanelerinde toplanan TİG verilerine göre, TİG oluşumu uygundur. Ek olarak, oluşan TİG'lerin

cinsiyete göre ayrışması ve GSS'li hastaların olmadığı kümelerdeki vakaların, hastane seçme şansı olmayan acil vakalar olması dikkat çekicidir. Erkek hastalardaki Göz ve Retinaya yapılan cerrahi işlemlerin, yeni doğan kız bebeklerdeki dahili müdahalelerin ve kadınlardaki Gastroskopi TİG'lerinin başlı başına birer küme oluşturması incelenmelidir.

Hastanelerin en yüksek maliyet nedenlerinden biri olan ve TİG vaka sisteminin uzun dönemde etkisinin en net olarak görülebildiği, hastanelerdeki kalış sürelerinin durumu normal ve aşırı olmak üzere 2'ye ayrılarak, değişkenlerin aşırı yatış süresine olan etkisi görülmeye çalışılmıştır. Kurulan modellerin normal ve aşırı yatış sürelerini etkili bir şekilde tahmin edebilmesi için %6.8 olan AYS oranı, tabakalı örnekleme yöntemi ile %50 olarak alınarak analizler gerçekleştirilmiştir. Ayrıca, p değerlerinin anlamlılık seviyesinin yanlı çıkmaması için veri setinden tabakalı örnekleme yöntemi ile 9438 örneklem çekilerek sonuçlar elde edilmiştir.

Hastanede aşırı yatış süresini etkileyen en önemli faktör TİG'lerin bağıl değerleri bulunmuştur. Cinsiyet değişkeni ise lojistik regresyon analizinde hastanede kalış süresini anlamlı derecede etkilemiş fakat karar ağacı analizinde anlamlı bulunmamış ve modele alınmamıştır. Hastaneden çıkış nedeni ise her iki modelde de anlamlı bulunamamıştır. Her iki modelin de AYS'yi tahmin performansları karşılaştırıldığında karar ağacı modelinin daha başarılı olduğu görülmüştür. AYS'yi etkileyen en önemli değişkenler ise, TİG bağıl değerleri, komorbidite, hastaya uygulanan işlemin türü olarak bulunmuştur.

Birliktelik analizi sonucuna göre diyabet hastalarında en yüksek bağıl değere sahip K01Z TİG'inin daha çok erkek hastalarda görüldüğü, düşük bağıl değere sahip K60B vakalarının ise genç yaştaki hastalarda görüldüğü, kadın hastaların çoğunlukla hastanede 1-5 gün tedavi görerek şifa ile hastaneden ayrıldığı görülmüştür.

Bu çalışmadaki veriler, sadece 2012 yılı İstanbul, Ankara ve İzmir'deki eğitim araştırma hastanelerini kapsamaktadır. Bu analizler, özellikle AYS'deki değişim,

yıllara, hastane tipine, hastanın acil serviste kalıp kalmadığı gibi durumlar karşılaştırmalı olarak incelenerek çok önemli bilgiler elde edilebilir.

Çalışmada daha yararlı bilgi elde etmek için, hastanın belirleyici kodu, birincil ve ikincil ICD kodları ile hastanın acil serviste tedavi görüp görmediği verileri de gelecekteki çalışmalara eklenmelidir. Ayrıca, veride hastalara ait özel bir numara da bulunmadığı için hastaneye kısa sürede tekrar yatış yapmış vakalarda görülememektedir.

Birliktelik analizi, KOAH, meme kanseri, iskemik kalp rahatsızlıkları gibi ülkemizde hastalık yükü ve dolayısıyla maliyeti yüksek olan hastalıklar için uygulanarak sağlık politikalarına katkı sağlanabilir.

Daha sonraki çalışmalarda analizler hastanelerdeki ayrımlara göre tekrar uygulanıp, vaka karma indeksleri ile ilişkileri yorumlanabilir.

Ülkemizde geri ödemelerde TİG'e geçiş sürecimiz devam etmekte ve üzerinde çalışılması gerekmektedir. Bu nedenle, bu veri havuzunda sağlık analitiği yöntemlerinin titizlikle uygulanarak kendi TİG'imizin oluşturulması sağlanmalıdır.

KAYNAKLAR

- Akdağ, R. "TİG Uygulama Rehberi", T.C. Sağlık Bakanlığı Tedavi Hizmetleri Genel Müdürlüğü Performans Yönetimi Kalite Geliştirme Daire Başkanlığı, Ankara, 2011.
- Akın, Y.K. : "Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi", **Marmara Üniversitesi, Sosyal Bilimler Enstitüsü Ekonometri ABD İstatistik BD Doktora Tezi**, 2008.
- Akpınar, H. : "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği". **İ.Ü. İşletme Fakültesi Dergisi**, Cilt: 29, s.1-22, 2000.
- Akpınar, H. "DATA: Veri Madenciliği Veri Analizi", Papatya Yayınları, İstanbul, 2014.
- Akyürek, Ç.E. "Sağlıkta Bir Geri Ödeme Yöntemi Olarak Global Bütçe ve Türkiye", **Sosyal Güvenlik Dergisi**, 2002.
- Alpaydın E. : "ZEKİ VERİ MADENCİLİĞİ: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri", Bilişim 2000 Eğitim Semineri, (Çevrimiçi) www.cmpe.boun.edu.tr/~ethem/files/papers/veri-maden_2k-notlar.doc, Ocak 2015.
- Aitkenhead M.J. : "A co-evolving decision tree classification method", **Expert Systems with Applications**, 34,18–25,2008,.
- Averill, R. F., Goldfield, N., Steinbeck, B., Grant, T., Muldoon, J., Brough, A. J., & Gay, J. Development of the all patient refined DRGs (APR-DRGs). 3M Health Information Systems Research Report, 8, 1-22,1997.
- Barker, A., Brand, C., Evans, S., Cameron, P., & Jolley, D. "Death in low-mortality diagnosis-related groups: frequency, and the impact of patient and hospital characteristics." **Med J Aust** 195.2 (2011): 89-94, 2011.

- Barnum H., Kutzin J.,
Saxenian H. "Incentives and provider payment methods." **The International journal of health planning and management**, 10.1,1995.
- Bei A. , De Luca S., Member,
IEEE, Ruscitti G., Salamon
D., "Health-Mining: a Disease Management Support
Service based on Data Mining and Rule Extraction",
**Engineering in Medicine and Biology 27th Annual
Conference, Shanghai, China**, 2005.
- Berger A.M., Berger C.R. : " Data Mining as a Tool for Research and Knowledge
Development in Nursing", **Computers, Informatics,
Nursing**, Volume 22(3), 123-131, 2004.
- Berkhin P. : "Survey of Clustering Data Mining Techniques",
Technical Report, Accrue Software Inc., San Jose,
California, USA, 2002.
- Berry M., Linoff G. : "**Mastering Data Mining : The Art and Science of
CRM**", John Wiley & Sons, 2000.
- Beylik U. "Sağlık Kurumlarında Tanı İlişkili Gruplara Dayalı Geri
Ödeme Modeli: Türkiye-Avustralya Karşılaştırması",
**Gazi Üniversitesi Sosyal Bilimler Enstitüsü İşletme
ABD Sağlık Kurumları Yönetimi Doktora Tezi**,
Ankara, 2014.
- Bradley, Paul S. "Implications of big data analytics on population health
management." **Big Data 1.3**, (2013): 152-159,2013.
- Busse R., Geisler A., Q. : "**Diagnoses-Related Groups in Europa**", WHO-
EuroDRG-7.Framework Programme, The McGraw Hill
Companies Open University Press, USA, 2011.
- Bircan H., Zontul M. ve
Yüksek A.G. "SOM Tipinde Yapay Sinir Ağlarını Kullanarak
Türkiye'nin İhracat Yaptığı Ülkelerin Kümelenmesi
Üzerine Bir Çalışma", **Atatürk Üniversitesi İktisadi ve
İdari Bilimler Fakültesi Dergisi**, Cilt:20,Sayı:2, 2006,
s.222-223

- Bircan, H. : “Lojistik regresyon analizi: tıp verileri üzerine bir uygulama”, **Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi**, 2, 185-208, 2004.
- Cabena, P.,Hadjinian P.,Stadler R., Verhees J. Zanası A. : “**Discovering Data Mining: From Concept to Implementation**” (1 edition), Prentice Hall PTR, NJ Upper Saddle River USA 1998.
- Cao, S.,Li1 J. ,Zhang X., Tian1 Y., Zhao J., Xue, W. : “Decision-Tree-Based Analysis of the Factors Influencing Single Disease Costs”, **2012 International Conference on Systems and Informatics**, s.2251-2254, 2012.
- Casto, A.B., Layman, E. : “Principles of Healthcare Reimbursement”, **Michigan: American Health Information Management Association**, 2006.
- Cerrito, Patricia B. : “Data Mining Episode Groupers.” , SAS Institute.
- Collica, R. : “**Customer Segmentation and Clustering Using SAS Enterprise Miner, Second Edition**” Sas Institue, 2011.
- Concaro S., Sacchi L., Cerra C., Fratino P., Bellazzi, R. : “Mining Healthcare Data with Temporal Association Rules: Improvements and Assessment for a Practical Use”, **Artificial Intelligence in Medicine**, Springer Berlin Heidelberg,16-25, 2009.
- Creese A. : “**User fees**”, British Medical Journal 1997.
- Çelik Y. : “Sağlık Ekonomisi”, Siyasal Kitabevi, 2011.
- Çelikay F., Gümüş E. : “Türkiye’de Sağlık Hizmetleri ve Finansmanı”, **Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi** , 2009.
- De Vries, G. G., J. M. H. Vissers, and Guus De Vries. : “The use of patient classification systems for production control of hospitals.” **Casemix Quart 2.2**, 65-70,2000.
- Demir M., Beylik U., Öztürk M.,Doluküp İ. : ” Sağlık Hizmetlerinde Teşhisle İlişkili Gruplar (TİG): Diyabet Vakaları Üzerine Karşılaştırma”, “**Sağlıkta Performans ve Kalite Dergisi**”, Sayı 6,Ankara, 2013.

- Dixon A. “Are Medical Savings Accounts a Viable Option for Funding Health Care”, **Croatian Medical Journal**, Vol.43, 2002.
- Dogaru, R., Zaharie, D., Lungeanu, D., Bernad, E., & Bari, M. “A framework for mining association rules in data on perinatal care”. **In The 8th International Conference on Technical Informatics (CONTI'2008)**, Timisoara, Romania, 2008.
- Dolgun, M.Ö. : “Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının Bağımlı Değişken Prevelansı Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması”, **Hacettepe - Üniversitesi Sağlık Bilimler Enstitüsü, Biyoistatistik Programı Doktora Tezi**, Ankara, 2014.
- Düzgünoğlu S., Yazıcı A. , : “Tıp Bilişiminde Veri Ambarı ve Veri Madenciliği Uygulaması”, 3. Ulusal Tıp Bilişimi Kongresi, s.110-115, 2006.
- Elghazel, H., Deslandres, V., : “A new clustering approach for symbolic data and its validation: Application to the healthcare data”, **In Foundations of Intelligent Systems**, s. 473-482. Springer Berlin Heidelberg, 2006.
- Elghazel, H., Deslandres, V., : “Clinical pathway analysis using graph-based approach and Markov models”, **In Digital Information Management, ICDIM'07**. 2nd International Conference on (Vol. 1, pp. 279-284). IEEE, 2007.
- El-Sappagh, Shaker H., Samir : “Data Mining and Knowledge Discovery: Applications, Techniques, Challenges and Process Models in Healthcare”, **International Journal of Engineering Research and Applications (IJERA)**, Vol. 3, Issue 3, 900-906, 2013.

- Freitas, A., Silva-Costa, T., Lopes, F., Garcia-Lema, I., Teixeira-Pinto, A., Brazdil, P., & Costa-Pereira, A.
- Freitas, A., Gaspar, J., Rocha, N., Marreiros, G., & Costa-Pereira, A.
- Frawley W.J., Piatetsky-Shapiro G., and Matheus C..
- Yaşar, G.Y.
- Giudici P.
- Häkkinen, Unto, et al.
- Häkkinena, U., Iversenb T., Peltolaa M., Seppälää T.T., Malmivaaraa A., Belicza, E., Fattored G., Numeratod D., Heijink, R., Medinf E., Rehnbergf, C., Hickey, Stephanie J.
- Gottret P. ve Schieber G.
- Gürsoy Şimşek, U.T
- “Factors influencing hospital high length of stay outliers”, **BMC health services research**, 12(1), 265, 2012.
- “Quality in Hospital Administrative Databases”, **Applied Mathematics & Information Sciences**, 8, No. 1L, 1-6, 2014.
- “Knowledge Discovery in Databases : An Overview”, **AI Magazine**, Vol. 13, No. 3, 57-70, 1992.
- “Sağlığın Finansmanı ve Türkiye için Sağlık Finansman Modeli Önerisi”, **Ankara Üniversitesi Sosyal Bilimler Enstitüsü Çalışma Ekonomisi ve Endüstri İlişkileri ABD Doktora Tezi**, Ankara, 2006.
- Applied Data Mining: Statistical Methods for Business and Industry**, John Wiley & Sons, 2003.
- "Patient classification and hospital costs of care for acute myocardial infarction in nine European countries." **Health economics** 21.S2, 19-29, 2012.
- “Health care performance comparison using a disease-based approach: The EuroHOPE project”, **Health Policy**, 112, 100– 109, 2003.
- "Naive Bayes Classification of Public Health Data with Greedy Feature Selection." **Communications of the IIMA** 13.2,7,2014.
- ” **Health Financing Revisited**”, The World Bank, 2006.
- “**Veri Madenciliği ve Bilgi Keşfi**”, Pegem Akademi yayınları, Ankara, 2009.

- HanJ., Kanber M. : "Data Mining Concepts and Techniques", ABD, Morgan Kaufman Publishers, 2006, s.5-7.
- Homayounfar, Payam, and Mieczyslaw L. Owoc. "Data mining research trends in computerized patient records." **Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on. IEEE, 2011.**
- Irmak S. : "Veri madenciliği Yöntemleri ile Sağlık Sektörü Veritabanlarında Bilgi Keşfi: Tanımlayıcı ve Kestirimci Model Uygulamaları", **Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü İşletme Doktora Tezi**, 2009.
- Isken, Mark W., and Balaji Rajagopalan. "Data mining to support simulation modeling of patient flow in hospitals." **Journal of medical systems** 26.2 (2002): 179-197,2012.
- İstanbulluoğlu H., Güleç M., Oğur R. : "Sağlık Hizmetlerinin Finansman Yöntemleri", **Dirim Tıp Gazetesi**, 2010.
- Jackson, T. "Using computerised patient-level costing data for setting DRG weights: the Victorian (Australia) cost weight studies." **Health Policy** 56.2 (2001): 149-163, 2001.
- Jain, D., and Sumanlata G.. "Predicting the Effect of Diabetes on Kidney using Classification in Tanagra." **International Journal of Computer Science and Mobile Computing** 3.4 ,2014.
- Jay, N.,Nuemi G., Gadreau M., Quantin, C. "A data mining approach for grouping and analyzing trajectories of care using claim data: the example of breast cancer", **BMC Medical Informatics and DecisionMaking**, 13:130, 2013.
- Jay, N., Napoli, A., & Kohler, F. "Cancer patient flows discovery in DRG databases" **In MIE**, 2006.
- Kantardzic M. : "Data Mining: Concepts, Models, Methods, and Algorithms", IEEE Press, Hoes Lane, Piscataway, NJ,ABD, 2003.

- Jay, N., Kohler, F., & Napoli, A.
- Kim, Y. J., Oh, Y., Park, S., Cho, S., & Park, H.
- Ko, M., and Osei-Bryson, K.M.
- Ko, M., and Osei-Bryson, K.M.
- Ko, M., and Osei-Bryson, K.M.
- Kraft, Margaret R., K. Desouza, and Ida Androwich.
- Kudyba, Stephan, and Thomas Gregorio.
- "Using formal concept analysis for mining and interpreting patient flows within a healthcare network", **Fourth International Conference on Concept Lattices and Their Applications**, 263-268, 2006.
- Stratified Sampling Design Based on Data Mining. **Healthcare informatics research**, 19(3), 186-195, 2013.
- "Reexamining the impact of information technology investment on productivity using regression tree and multivariate adaptive regression splines (MARS)." **Information Technology and Management 9.4** (2008): 285-299, 2008.
- "Using regression splines to assess the impact of information technology investments on productivity in the health care industry." **Information Systems Journal**, 14.1 (2004): 43-63, 2014.
- "The productivity impact of information technology in the healthcare industry: an empirical study using a regression spline-based approach", **Information and Software Technology 46**, 65–73, 2004.
- "Data mining in healthcare information systems: case study of a veterans' administration spinal cord injury population." **System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on**. IEEE, 2003.
- "Identifying factors that impact patient length of stay metrics for healthcare providers with advanced analytics." **Health informatics journal 16.4**, 235-245, 2010.

- Kutzin J. : "A descriptive framework for country-level analysis of health care financing arrangements", **Health Policy** 56, 2001.
- Kutzin J. "Towards Universal Health Care Coverage: A Goal-oriented Framework for Policy Analysis", **Geneva: The World Bank**; 2004.
- Langenbrunner J. ve Wiley M., "Hospital Payment Mechanisms: Theory And Practice in Transition Countries", **Hospital in a Changing Europe, Open University Press**, Buchingham, 2002.
- Lee K, Goodman H. "Global policy networks: the propagation of health care financing reform since the 1980s." In: Lee K, Buse K, Fustukian S, eds. Isted. Health Policy in a Globalising World". Cambridge, Cambridge University Press; 2002.
- Lee, M. T., Su, Z. Y., Hou, Y. H., Liao, H. C., & Lian, J. D. A decision support system for diagnosis related groups coding. **Expert Systems with Applications**, 38(4), 3626-3631, 2011.
- Lin M., Lucas Jr H.C., and Shmueli G. "Research commentary-too big to fail: large samples and the p-value problem." ,**Information Systems Research**, 24.4 : 906-917, 2013.
- Mach, M.A., Salem, A.B.M. "Intelligent Techniques for Business Intelligence in Healthcare", **10th International Conference on Intelligent Systems Design and Applications**, 2010.
- Makas, R. : "Türk Sosyal Güvenlik Sisteminde Norm ve Standart Birliği", **Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Hukuk Anabilim Dalı Özel Hukuk Bilim Dalı Doktora Tezi**, 2003.

- Maynard A., Dixon A. : "Private health insurance and medical savings accounts: theory and experience", In "**Funding health care: options for Europe**", Open University Press, 2002.
- Mossialos E., Dixon A. : Elias Mossialos, Anna Dixon., "**Funding health care: An introduction**", Berkshire: Open University Press, 2002.
- Obenshain, Mary K. "Application of data mining techniques to healthcare data." **Infection Control** 25.08: 690-695,2004.
- Oğuzlar A. "Kümeleme Analizinde Yeni Bir Yaklaşım: Kendini Düzenleyen Ağlar (Kohonen Ağları)." **Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi**, 19.2,2005.
- Okamoto, K., Uchiyama, T., Takemura, T., Adachi, T., Kume, N., Kuroda, T., ... & Yoshihara, H. "Automatic Selection of Diagnosis Procedure Combination Codes Based on Partial Treatment Data Relative to the Number of Hospitalization Days". Proc. APAMI 2012, (4), 1031,2012.
- Olafsson S. , Li X., Wu S. "Operations research and data mining", **European Journal of Operational Research**, 187 1429–1448, 2008.
- Özkan, Y. : "Veri Madenciliği Yöntemleri", Papatya Yayınları, İstanbul,2008.
- Özdamar, K. : "Paket Programlar İle İstatistiksel Veri Analizi", Kaan Kitabevi, Eskişehir, 2002.
- Özsarı H. "Genel sağlık sigortası özel sağlık sigortacılığı ilişkisi",**SD dergisi**, Aralık, Ocak, Şubat 2006-2007-1
- Öztemel E. : "Yapay Sinir Ağları", 1. Baskı, Papatya Yayıncılık, İstanbul,2003.
- Özyirmidokuz,E.K. "Veri Madenciliği Tekniklerini Kullanarak İmalat Verilerinin Modellenmesi ve Analizi", **Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Doktora Tezi**, Kayseri, 2009.

- Parr-Rud, Olivia. **“Business Analytics Using SAS Enterprise Guide and SAS Enterprise Miner: A Beginner's Guide”**. SAS Institute, 2014.
- Raghupathi W., Viju R.. "Big data analytics in healthcare: promise and potential." *Health Information Science and Systems* 2.13, 2014.
- Rajagopalan, Balaji, and Mark W. Isken "Exploiting data preparation to enhance mining and knowledge discovery." **Systems, Man, and Cybernetics, Part C: Applications and Reviews**, IEEE Transactions on 31.4 (2001): 460-467, 2001.
- Rajagopalan, Balaji, and Mark W. Isken "An overview of health analytics." *J Health Med Informat* 4.132 (2013): 2, 2013.
- Rajesh, K., and V. Sangeetha. "Application of data mining methods and techniques for diabetes diagnosis." **International Journal of Engineering and Innovative Technology (IJEIT)** Volume 2, 2012.
- Ron A., Abel-Smith B. ve : "Health Insurance in Developing Countries: The Social Tamburi G. Security Approach", Ceneva: ILO; 1990.
- Roussinov, D.F. ve Chen H., "Document Clustering For Electronic Meetings: An Experimental Comparison Of Two Techniques", **Decision Support Systems**, Vol: 27, Issue: 1-2, 1999.
- Shmueli G. , Patel N.R. ve : **“Data Mining for Business Intelligence: Concepts, Bruce P.C. Techniques, and Applications in Microsoft Office Excel with XLMiner”**, John Wiley & Sons, Hoboken, NJ, USA, 2007.
- Sarma, K. S. **“Predictive Modeling with SAS Enterprise Miner: Practical Solutions for Business Applications, Second Edition”**, SAS Institute, 2013.
- Serdén, L., Rikard L., and Måns R.. "Have DRG-based prospective payment systems influenced the number of secondary diagnoses in health

- care administrative data?" **Health Policy** 65.2 2003: 101-107, 2013.
- Silahtaroğlu G. : ”**Kavram ve Algoritmalarıyla Temel Veri Madenciliği**”, Papatya Yayıncılık, İstanbul, 2008.
- Siew, E. G., Smith, K. A., Churilov, L., & Ibrahim, M. “A neural clustering approach to iso-resource grouping for acute healthcare in Australia". In System Sciences, 2002. HICSS. **Proceedings of the 35th Annual Hawaii International Conference on System Sciences** , IEEE, 2002.
- Srinivasan, U., Arunasalam,B. “Leveraging Big Data Analytics to Reduce Healthcare Costs”, **IEEE Computer Society**, 2013.
- Steinwald B. , Dummit, L.A. “Hospital case-mix change: sicker patients or DRG Creep”, **Health Affairs**, 8, no.2, 35-47, 1989.
- Sun J., Li H. : ” Data mining method for listed companies’ financial distress prediction”, **Knowledge-Based Systems**, 21 (1–5), 2008.
- Sülkü, Seher Nur “**Türkiye’de Sağlık Dönüşüm Programı Öncesi ve Sonrasında Sağlık Hizmetlerinin Sunumu ve Finansmanının Sunumu Finansmanı ve Sağlık Harcamaları**”, T.C. Maliye Bakanlığı Strateji Geliştirme Başkanlığı,2011.
- Şen Z. : “**Yapay Sinir Ağları İlkeleri**”, Su Vakfı Yayınları, İstanbul, 2004.
- Şencan İ., Şeker M. ve Demir M. : Teşhis İlişkili Gruplar İleri Klinik Kodlama Standartları, Sağlık Hizmetleri Genel Müdürlüğü, 2013.
- Tan P.P., Steinbach M., Kumar V., : “**Introduction to Data Mining**”, Pearson, Addison-Wesley, Boston, MA, USA, 2006.
- Tan, A.C., and Gilbert D.. "An empirical comparison of supervised machine learning techniques in bioinformatics." **Proceedings of the First Asia-Pacific bioinformatics conference on**

- Bioinformatics 2003**-Volume 19. Australian Computer Society, Inc., 2003.
- Tatar M. : “Sağlık Hizmetlerinin Finansman Modelleri: Sosyal Sağlık Sigortasının Türkiye’de Gelişimi”, **Sosyal Güvenlik Dergisi**, 2011.
- Tengilimoğlu D., Işık O., ve Akbolat M. : ”Sağlık işletmeleri yönetimi” Nobel Akademik Yayıncılık, 2009.
- Terlemez L. : “Eş İşlem Stratejisi Yöntemiyle İMKB’de Portföy Oluşturmada Veri Madenciliği Uygulaması”, **Anadolu Üniversitesi Sosyal Bilimler Enstitüsü Doktora Tezi**, Eskişehir, 2008.
- Wang, Z., Liu, R., Li, P., Jiang, C., Hao, M. “How to Make Diagnosis Related Groups Payment More Feasible in Developing Countries- A Case Study in Shanghai, China”, **Iranian J Publ Health**, Vol. 43, No.5, 572-578, 2014.
- Wang, Z., Liu, R., Li, P., Jiang, C. "Exploring the transition to DRGs in Developing Countries: A case study in Shanghai, China." **Pakistan journal of medical sciences** 30.2, 2014
- Williams, N. “Conclusions from a NAIVE Bayes Operator Predicting the Medicare 2011 Transaction Data Set”. arXiv preprint arXiv:1403.7087,2014.
- WHO : ”**The World Health Report 2000, Health Systems: Improving Performance**”, Geneva,2000.
- Xu Z., Wang L., Luo J. ve Zhang J., : “A Modified Clustering Algorithm for Data Mining”, IGRSS’05 Proceedings 2005 IEEE Int., Vol.2, 2005.
- Yılmaz A. : “Klinik Kodlama Süreci”, **Tanı İlişkili Gruplar e-Bülteni**, Sayı:1, Ekim-Kasım-Aralık 2007.
- Yılmaz A. “Uluslararası TİG Çalışmaları: TİG Ortamında Ürün Maliyetlendirmesi”, **Tanı İlişkili Gruplar e-Bülteni**, Sayı 7, 2009.

- Zhou Z.H. : “Three Perspectives of Data Mining”, **Artificial Intelligence**, No.143, 139-146, 2003.
- IBM SPSS Modeler CRISP-DM Guide ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/CRISP_DM.pdf (Şubat 2015)
- OECD Sağlık Sistemi 2008
İncelemeleri Türkiye <http://sbu.saglik.gov.tr/Ekutuphane/kitaplar/OECDKITAP.pdf> (Ekim 2014)
- SAS Enterprise Miner : Data Mining Using SAS® Enterprise Miner™A Case Study Approach Third Edition, Amerika, 2013, s.5-6.
<http://support.sas.com/documentation/cdl/en/emcs/66392/PDF/default/emcs.pdf> Ocak 2015).
- Teşhis İlişkili Gruplar Daire Başkanlığı 2.Ulusal TİG Sempozyumu Sunumları (Eylül 2014) <http://www.sb.gov.tr/testtig/belge/1-38688/2ulusal-tig-sempozyumu-gerceklestirildi-30-eylul-2014-a-.html> (Eylül 2014-Mart 2015).
- ... : <http://searchdatamanagement.techtarget.com/feature/Big-data-growth-increases-data-integration-degree-of-difficulty> (Mart 2015).
- ... : Introduction to Data Mining and Knowledge Discovery, Third Edition, 2005,
<http://www.twocrows.com/intro-dm.pdf> (Şubat, 2015)

ÖZGEÇMİŞ

Pakize Yiğit, 2007 yılında Yıldız Teknik Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünden mezun olmuştur. 2011 yılında İstanbul Üniversitesi Sosyal Bilimler Enstitüsü İşletme ABD Sayısal Yöntemler yüksek lisans programından mezun olmuştur. Aynı yıl aynı bölümde doktora başlamış ve İstanbul Medipol Üniversitesi Sağlık Bilimleri Fakültesi, Sağlık Yönetimi Bölümünde araştırma görevlisi olarak çalışmaya başlamıştır.

Yiğit, 2014 yılından beri İstanbul Medipol Üniversitesinde Öğretim Görevlisi olarak görev yapmaya devam etmektedir.