

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ PROGRAMI
YÜKSEK LİSANS TEZİ

VERİ MADENCİLİĞİ TEKNİKLERİ İLE ÜNİVERSİTE
ÖĞRENCİLERİNİN SEÇMELİ DERS EĞİLİMLERİNE İLİŞKİN
VERİ DESENLERİNİN BELİRLENMESİ: EKONOMETRİ
BÖLÜMÜ ÖĞRENCİLERİ ÜZERİNE BİR UYGULAMA

Dilayla BAYYURT

Danışman

Doç. Dr. Mehmet AKSARAYLI

İZMİR- 2016

YÜKSEK LİSANS
TEZ/ PROJE ONAY SAYFASI

2013800204

Üniversite : Dokuz Eylül Üniversitesi
Enstitü : Sosyal Bilimler Enstitüsü
Adı ve Soyadı : Dilayla BAYYURT
Tez Başlığı : Veri Madenciliği Teknikleri İle Üniversite Öğrencilerinin Seçmeli Ders Eğilimlerine İlişkin Veri Desenlerinin Belirlenmesi: Ekonometri Bölümü Öğrencileri Üzerine Bir Uygulama
Savunma Tarihi : 28.07.2016
Danışmanı : Doç.Dr.Mehmet AKSARAYLI

JÜRİ ÜYELERİ

<u>Ünvanı, Adı, Soyadı</u>	<u>Üniversitesi</u>	<u>İmza</u>
Doç.Dr.Mehmet AKSARAYLI	DOKUZ EYLÜL ÜNİVERSİTESİ	
Prof.Dr.İpek DEVECİ KOCAKOÇ	DOKUZ EYLÜL ÜNİVERSİTESİ	
Doç.Dr.Mehmet Emre GÜLER	İZMİR KATİP ÇELEBİ ÜNİVERSİTESİ	

Oybirliği
Oy Çokluğu ()

Dilayla BAYYURT tarafından hazırlanmış ve sunulmuş "**Veri Madenciliği Teknikleri İle Üniversite Öğrencilerinin Seçmeli Ders Eğilimlerine İlişkin Veri Desenlerinin Belirlenmesi: Ekonometri Bölümü Öğrencileri Üzerine Bir Uygulama**" başlıklı Tezi() / Projesi() kabul edilmiştir.

Prof.Dr. Mustafa Yaşar TINAR
Enstitü Müdürü

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “Veri Madenciliği Teknikleri İle Üniversite Öğrencilerinin Seçmeli Ders Eğilimlerine İlişkin Veri Desenlerinin Belirlenmesi: Ekonometri Bölümü Öğrencileri Üzerine Bir Uygulama ” adlı çalışmamın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

Tarih

.../.../.....

Dilayla BAYYURT

İmza

ÖZET

Yüksek Lisans Tezi

Veri Madenciliği Teknikleri İle Üniversite Öğrencilerinin Seçmeli Ders
Eğilimlerine İlişkin Veri Desenlerinin Belirlenmesi: Ekonometri Bölümü
Öğrencileri Üzerine Bir Uygulama
Dilayla BAYYURT

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Ekonometri Anabilim dalı

Ekonometri Programı

Günümüzde, gelişen teknoloji sayesinde kayıt altına alınan ve saklanan veri sayısı günden güne katlanarak artmaktadır. Bununla birlikte, ‘bilginin çıkarımı’ daha fazla önem kazanmaktadır. Veri madenciliği, büyük miktardaki veri içerisinden anlamlı bilginin elde edilmesini sağlayan bir süreçtir.

Yapılan tez çalışmasında, veri madenciliği ile ilgili kavramlar, birliktelik kuralları ve Apriori Algoritması detaylı bir şekilde incelenmiş ve kapsamlı bir literatür araştırması yapılmıştır. Özellikle Market Sepet Analizi uygulamalarında kullanılan Apriori Algoritması farklı bir veri kümesi üzerine uygulanmıştır. Üniversite öğrencilerinin seçmeli ders eğilimlerine ilişkin veri deseni elde etmeye yönelik bir anket oluşturulmuştur. Elde edilen verilere öncelikle gerekli ön işlemler uygulanarak analiz için uygun hale getirilmiştir. Düzenlenen veriler ile farklı model varyasyonları geliştirilmiş ve birçok değişken için bağlantı analizleri gerçekleştirilmiştir.

Anahtar Kelimeler: Veri Madenciliği, Birliktelik Kuralları, Apriori Algoritması, Veri Deseni, Seçmeli Ders.

ABSTRACT

Master's Thesis

**Determining of Data Patterns of Elective Course Tendencies of University
Students With Data Mining Techniques: An Application on Students of
Econometrics Department
Dilayla BAYYURT**

**Dokuz Eylül University
Graduate School of Social Sciences
Department of Econometrics
Econometrics Program**

Today, the number of data to be recorded and stored is increasing day by day thanks to improving technology. At the same time, "information inference" is coming into prominence. Data mining is a process by which to obtain meaningful information in large amounts of data.

In this thesis, concepts of data mining, association rules Apriori algorithm and examined in detail and a comprehensive literature search was conducted. In particular, when used in Apriori Algorithm Market Basket Analysis applications were applied on a different set of data. A survey aimed at obtaining data pattern for the elective courses trend of university students was formed. Primarily, the obtained data has become convenient for analysis applying the necessary preprocesses. Different model variations were developed with edited data and connection analysis was carried out for many variables.

Keywords: Data Mining, Association Rules, Apriori Algorithm, Data Pattern, Elective Courses.

**VERİ MADENCİLİĞİ TEKNİKLERİ İLE ÜNİVERSİTE
ÖĞRENCİLERİNİN SEÇMELİ DERS EĞİLİMLERİNE İLİŞKİN VERİ
DESENLERİNİN BELİRLENMESİ: EKONOMETRİ BÖLÜMÜ
ÖĞRENCİLERİ ÜZERİNE BİR UYGULAMA**

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
KISALTMALAR	ix
TABLolar LİSTESİ	x
ŞEKİLLER LİSTESİ	xii
EKLER LİSTESİ	xv
GİRİŞ	1

**BİRİNCİ BÖLÜM
KAPSAM VE LİTERATÜR**

1.1. AMAÇ	3
1.2. KAPSAM	3
1.3. LİTERATÜR ARAŞTIRMASI	3

**İKİNCİ BÖLÜM
VERİ MADENCİLİĞİ**

2.1. VERİ MADENCİLİĞİNİN TANIMI	9
2.2. VERİ MADENCİLİĞİNİN DİĞER DİSİPLİNLERLE İLİŞKİSİ	11
2.2.1. Veri Madenciliği ve İstatistik	11
2.2.2. Veri Madenciliği ve Makine Öğrenimi	12

2.2.3. Veri Madenciliği ve Veri Tabanı	12
2.3. VERİ MADENCİLİĞİNİN UYGULANABİLDİĞİ DEPOLAMA BİRİMLERİ	12
2.3.1. Veri Tabanları	12
2.3.1.1. İlişkisel Veri Tabanları	13
2.3.1.2. İşlemsel Veri Tabanları	13
2.3.1.3. Gelişmiş Veri Tabanları	14
2.3.2. Veri Ambarları	14
2.4. VERİ MADENCİLİĞİ SÜRECİ	16
2.4.1. Verilerin Hazırlanması (Veri Önleme)	18
2.4.1.1. Veri Temizleme	19
2.4.1.2. Veri Birleştirme	20
2.4.1.3. Veri Dönüştürme	21
2.4.1.4. Veri İndirgeme	21
2.5. VERİ MADENCİLİĞİ UYGULAMA ALANLARI	22
2.6. VERİ MADENCİLİĞİ TEKNİKLERİ	24
2.6.1. Kümeleme Teknikleri	25
2.6.1.1. Bölümleyici Teknikler	28
2.6.1.2. Hiyerarşik Teknikler	28
2.6.1.3. Yoğunluk Tabanlı Teknikler	28
2.6.1.4. Izgara Tabanlı Teknikler	28
2.6.1.5. Model Tabanlı Teknikler	29
2.6.2. Sınıflandırma Teknikleri	29
2.6.2.1. Diskriminant Analizi	31
2.6.2.2. NaiveBayes	32
2.6.2.3. Karar Ağaçları	32
2.6.2.4. Yapay Sinir Ağları	33
2.6.2.5. Kaba Kümeler	33
2.6.2.6. Genetik Algoritmalar	34
2.6.2.7. Regresyon Analizi	34
2.6.3. Birliktelik Kuralları Teknikleri	34

ÜÇÜNCÜ BÖLÜM BİRLİKTELİK KURALLARI

3.1. BİRLİKTELİK KURALLARININ TANIMI	35
3.2. BİRLİKTELİK KURALLARI İÇİN VERİ GÖSTERİMİ	35
3.3. BİRLİKTELİK KURALLARININ MATEMATİKSEL GÖSTERİMİ	37
3.3.1. K- Öğe Küme Ve Sık Geçen Öğe Küme	38
3.3.2. Destek (Support) ve Güven (Confidence) Değerleri	38
3.4. BİRLİKTELİK KURALLARININ AŞAMALARI	39
3.5. BİRLİKTELİK KURALLARI NOTASYONLARI	40
3.6. BİRLİKTELİK KURALLARI ALGORİTMALARI	40
3.6.1. Apriori Algoritması	41
3.6.2. AprioriTid Algoritması	49
3.6.3. AprioriHybrid Algoritması	53
3.6.4. AIS Algoritması	53
3.6.5. SETM Algoritması	54
3.6.6. FP- Growth Algoritması	54
3.6.7. Bölümleme Algoritması	54

DÖRDÜNCÜ BÖLÜM UYGULAMA

4.1. UYGULAMA GİRİŞ	55
4.2. VERİ SETLERİNİN OLUŞTURULMASI	55
4.3. VERİLERİN HAZIRLANMASI	56
4.4. VERİ SETİNE İLİŞKİN İSTATİSTİKLER	63
4.5. MODELLEME	73
4.6. BİRLİKTELİK DİYAGRAMLARI (AĞ GRAFİKLERİ)	84
SONUÇ	90
KAYNAKÇA	92
EKLER	

KISALTMALAR

SPSS	Statistical Packagesforthe Social Sciences
SQL	Structured Query Language
D	Veri tabanı
TID	İşlem Numarası
L_k	k- Sık Geçen Öğe Kümeleri
C_k	K- Öğeli Aday Kümeler
\hat{C}_k	K- Öğeli Ve TID' leri İçeren Aday Kümeler
FP-Growth	FrequentPatternGrowth
AIS	AgrawalImielinskiSwami
C&RT	ClassificationandRegressionTrees
CHAID	Chi-SquaredAutomaticInteractionDetector
ID3	Induction of DecisionTrees

TABLULAR LİSTESİ

Tablo 1: İşlemsel Veri Tabanı Örneği	s.14
Tablo 2: Kümeleme Tekniği Veri Seti	s.26
Tablo 3: Sınıflandırma Tekniği Eğitim Verileri	s.30
Tablo 4: Sınıflandırma Tekniği Eğitim Verilerinden Oluşturulan Kurallar	s.30
Tablo 5: Sebze Standında Yapılan İşlemler	s.36
Tablo 6: Sebze Standı Verileri İçin İşlemsel Veri Formatı	s.36
Tablo 7: Sebze Standı Verileri İçin Çizelge Tipi Veri Formatı	s.37
Tablo 8: Birliktelik Kuralları Notasyonları	s.40
Tablo 9: Birliktelik Kuralları Algoritmaları	s.40
Tablo 10: Apriori Algoritması İşlemsel Veriler	s.44
Tablo 11: Apriori Algoritması C_1 Kümesi	s.45
Tablo 12: Apriori Algoritması L_1 Kümesi	s.45
Tablo 13: Apriori Algoritması C_2 Kümesi	s.46
Tablo 14: Apriori Algoritması L_2 Kümesi	s.46
Tablo 15: Apriori Algoritması C_3 Kümesi	s.48
Tablo 16: Apriori Algoritması L_3 Kümesi	s.48
Tablo 17: AprioriTid Algoritması İşlemsel Verileri	s.50
Tablo 18: AprioriTid Algoritması \hat{C}_1 Kümesi	s.50
Tablo 19: AprioriTid Algoritması L_1 Kümesi	s.51
Tablo 20: AprioriTid Algoritması C_2 Kümesi	s.51
Tablo 21: AprioriTid Algoritması \hat{C}_2 Kümesi	s.51
Tablo 22: AprioriTid Algoritması L_2 Kümesi	s.52
Tablo 23: AprioriTid Algoritması C_3 Kümesi	s.52
Tablo 24: AprioriTid Algoritması \hat{C}_3 Kümesi	s.52
Tablo 25: AprioriTid Algoritması L_3 Kümesi	s.53
Tablo 27: Genel Özelliklere İlişkin Değişkenler ve Özellikleri	s.60
Tablo 28: Anabilim Dalı Bazında İlgi Alanlarına İlişkin Değişkenler ve Özellikleri	s.61
Tablo 29: Genel İlgi Alanlarına İlişkin Değişkenler ve Özellikleri	s.61
Tablo 30: Seçimlik Ders Seçim Sebeplerine İlişkin Değişkenler ve Özellikleri	s.61

Tablo 31: Likert Tipindeki Sorulara İlişkin Değişkenler ve Özellikleri	s.62
Tablo 32: Ders ve Seçim Sebeplerini İçeren Değişkenler ve Özellikleri	s.62
Tablo 33: Cinsiyet Dağılımı	s.63
Tablo 34: Öğretim Türü Dağılımı	s.63
Tablo 35: Öğrenim Dönemi Dağılımı	s.63
Tablo 36: Liseden Mezun Olunan Bölüm Dağılımı	s.64
Tablo 37: Mezun Olunan Lise Türü Dağılımı	s.64
Tablo 38: Çalışma Durumu Dağılımı	s.65
Tablo 39: Öğrencilerin Tekrar Şans Verilse Hangi Bölümde Okumak İsteyeceklerinin Dağılımı	s.65
Tablo 40: Öğrenim Dönemine Göre, Öğrencilerin Tekrar Şans Verilse Hangi Bölümde Okumak İsteyeceklerinin Dağılımı (Ekonometri & Diğer)	s.66
Tablo 41: Öğrencilerin Mezuniyet Sonrası Hedeflerinin Dağılımı	s.66
Tablo 42: ‘Seçimlik Dersleri Seçerken En Çok Neyi Referans Alıyorsunuz?’ Dağılımı	s.67
Tablo 43: Seçimlik Derslerin Dağılımı	s.67
Tablo 44: ‘En Kısa Zamanda Mezun Olma Baskısı Hissetmiyorum’ Dağılımı	s.68
Tablo 45: ‘En Kısa Sürede Mezun Olmaktansa, Donanımlı Mezun Olmayı Tercih Ederim’ Dağılımı	s.69
Tablo 46: ‘Ekonometri Bölümü’ ne İsteyerek, Bilerek, Araştırarak Geldim’ Dağılımı	s.69
Tablo 47: ‘Öğrenim Dönemine Göre, ‘Ekonometri Bölümü’ ne İsteyerek, Bilerek, Araştırarak Geldim’ Dağılımı	s.70
Tablo 48: ‘Okuduğum Bölümü Benimsedim ve Seviyorum’ Dağılımı	s.70
Tablo 49: Öğrenim Dönemine Göre, ‘Okuduğum Bölümü Benimsedim ve Seviyorum’ Dağılımı	s.71
Tablo 50: ‘Dokuz Eylül Üniversitesi’ ne İsteyerek Geldim’ Dağılımı	s.71
Tablo 51: ‘Araştırma Yapmayı Severim’ Dağılımı	s.72
Tablo 52: ‘Bilgisayar Kullanımına Erişimim Kolaydır’ Dağılımı	s.72
Tablo 53: ‘Yabancı Dil Düzeyimin Yeterli Olduğunu Düşünüyorum’ Dağılımı	s.72
Tablo 54: ‘Benim İçin Yeni Bir Şey Öğrenmekten Çok Dersi Geçmek Daha Önemlidir’ Dağılımı	s.73

Tablo 55: Apriori Algoritması İle Elde Edilen Modellere Ait Değişken Özellikleri	s.74
Tablo 56: 1. Model ile Elde Edilen Birliktelik Kuralları	s.75
Tablo 57: 2. Model ile Elde Edilen Birliktelik Kuralları	s.77
Tablo 58: 3. Model ile Elde Edilen Birliktelik Kuralları	s.79
Tablo 59: 4. Model ile Elde Edilen Birliktelik Kuralları	s.81
Tablo 60: 5. Model ile Elde Edilen Birliktelik Kuralları	s.83



ŞEKİLLER LİSTESİ

Şekil 1: Veri Madenciliği	s.10
Şekil 2: Veri Madenciliği Ve Diğer Disiplinler	s.11
Şekil 3: Tipik Bir Veri Ambarı Yapısı	s.15
Şekil 4: Veri Madenciliği Süreci	s.17
Şekil 5: Veri Önışleme Formları	s.19
Şekil 6: Veri Birleřtirme Örneęi	s.22
Şekil 7: Veri Madencilięi Tekniklerinin Gruplandırılması	s.25
Şekil 8: Kümeleme Teknięi Serpilme Diyagramı	s.27
Şekil 9: Sınıflandırma Teknięi Eğitim Verilerine Uygun Karar Ağacı	s.30
Şekil 10: Apriori Algoritmasını Gösteren Akış Diyagramı	s.41
Şekil 11: Apriori Algoritmasının Birleřtirme Özellięi	s.42
Şekil 12: Apriori Algoritmasının Budama Özellięi	s.43
Şekil 13: Apriori Algoritması Kodu	s.44
Şekil 14: AprioriTid Algoritması Kodu	s.49
Şekil 15: Clementine’ da SPSS Düęümü İle 1. Veri Setine Saęlanan Eriřim	s.56
Şekil 16: Uygulamada Kullanılan 1. Veri Setinin Clementine’ da Görüntülenmesi	s.57
Şekil 17: Filler Düęümü İle Boř Deęerlerin Doldurulması	s.58
Şekil 18: Boř deęerlerin doldurulması sonrası 1. veri setinin görüntüsü	s.58
Şekil 19: Veri Kalitesini Gösteren Quality Düęüm Çıktısı (1. Veri) seti	s.59
Şekil 20: Type Düęümü Görüntüsü	s.60
Şekil 21: 1. Model ile Elde Edilen Birliktelik Kuralları	s.75
Şekil 22: 2. Model İle Elde Edilen Birliktelik Kuralları	s.76
Şekil 23: 3. Model İle Elde Edilen Birliktelik Kuralları	s.78
Şekil 24: 4. Model İle Elde Edilen Birliktelik Kuralları	s.80
Şekil 25: 5. Model İle Elde Edilen Birliktelik Kuralları	s.82
Şekil 26: Ders Seęim Sebeplerine İliřkin Çizgi Seviyesi 0 İken Elde Edilen Ağ Grafięi	s.84
Şekil 27: Ders Seęim Sebeplerine İliřkin Çizgi Seviyesi 45 İken Elde Edilen Ağ Grafięi	s.85

Şekil 28: Ders Seçim Sebeplerine İlişkin En güçlü birliktelikler	s.85
Şekil 29: Derslere İlişkin Çizgi Seviyesi 0 İken Elde Edilen Ağ Grafiği	s.86
Şekil 30: Derslere İlişkin Çizgi Seviyesi 30 İken Elde Edilen Ağ Grafiği	s.87
Şekil 31: Derslere İlişkin En Güçlü Birliktelikler	s.87
Şekil 32: İlgi Alanlarına İlişkin Çizgi Seviyesi 0 İken Elde Edilen Ağ Grafiği	s.88
Şekil 33: İlgi Alanlarına İlişkin Çizgi Seviyesi 20 İken Elde Edilen Ağ Grafiği	s.88
Şekil 34: İlgi Alanlarına İlişkin En Güçlü Birliktelikler	s.89



EKLER LİSTESİ

EK 1: Arařtırmada Kullanılan Anket Formunun Ekran Görüntüsü ek s.1

EK 2: Seçimlik Dersler Listesi ek s.6



GİRİŞ

Gün geçtikçe teknolojideki gelişim ile birlikte elde edilen ve kayıt altına alınan veri miktarı artmakta ve bu artışla birlikte bilgiye ulaşmak daha önemli hale gelmektedir. Veri miktarındaki bu artış, ancak eldeki veri anlamlı ve işe yarar hale getirilebildiğinde önem arz etmekte ve avantaj sağlamaktadır. Veri madenciliği kavramı bu noktada önem kazanmaktadır.

Veri madenciliği çok büyük miktarlardaki veri içerisinde anlamlı ve önemli olan bilgiye ulaşmayı sağlayan bir süreçtir. Birçok veri madenciliği tekniği bulunmakla birlikte problemin yapısı ve elde edilmek istenilen sonuçlara göre uygun teknik seçilebilir. Veri madenciliğinin son dönemlerde önem kazanmasının en büyük nedenlerinden birisi giderek artan veri ve bu veri yığınının yararlı bilgiye dönüştürülebilmesi ihtiyacıdır.

Üniversite eğitiminde öğrenciler, öğrenim gördükleri programlarda müfredat gereği seçmeli dersler almaktadırlar. Seçmeli ders uygulamasında öğrenciler ders seçimlerinde kendilerine göre bazı kriterleri göz önünde bulundurmakta ve bu kriterler öğrenciden öğrenciye farklılık göstermektedir. Öğrenciler açısından ders seçim sebepleri ve birçok madde ilişkilendirildiğinde ve elde edilen ilişkiler incelendiğinde bu konuda gerekli düzenlemeler yapılarak eğitim kalitesinde önemli iyileştirmeler yapılabilir.

Yapılan çalışmanın amacı, üniversite öğrencilerinin seçmeli derslerle ilgili seçim kararı verirken hangi kriterleri göz önünde bulundukları ve bu anlamda genel eğilimlerini belirlemektir. Bu amaçla Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesinde bulunan Ekonometri Bölümü öğrencilerinden elde edilen verilere veri madenciliği teknikleri, birliktelik kuralları çıkarım algoritmalarından Apriori Algoritması kullanılarak uygulama gerçekleştirilmiştir.

Çalışmanın ilk bölümü tezin amacı, kapsamı ve literatür araştırmasından oluşmaktadır.

Çalışmanın ikinci bölümünde veri madenciliği tanımı, veri madenciliğinin diğer disiplinlerle olan ilişkisi, veri madenciliğinin uygulanabildiği depolama birimleri, veri hazırlama süreci, veri madenciliği uygulama alanları incelenmiştir.

Aynı zamanda veri madenciliği tekniklerine de ayrı ayrı değinilmiş ve bir örnek ile giriş niteliğinde gösterilmiştir.

Üçüncü bölümde birliktelik kuralları ayrıntılı bir biçimde ele alınmış, birliktelik kuralları çıkarım algoritmalarına değinilmiş ve bu algoritmalar içerisinde uygulamada kullanılan Apriori Algoritması ayrıntılı bir şekilde incelenmiştir.

Uygulamanın dördüncü bölümün ise uygulama bölümüdür. Bu bölümde, SPSS Clementine programı kullanılarak öncelikle anket verileri ön işleme sürecine tabii tutulmuştur. Daha sonra hazır hale getirilen veriler üzerinde birliktelik kurallarının en temel algoritmalarından olan Apriori Algoritması kullanılarak farklı değişkenlerle ve koşullarla beş ayrı model geliştirilerek öğrencilerinin seçmeli ders eğilimlerine ilişkin veri desenleri elde edilmiştir. En son ise, değişkenlerin ikili birliktelikleri arasındaki ilişkilerin gücünü ve grafiğini elde etmemizi sağlayan ağ grafiklerinden yararlanılmıştır. Ağ grafikleri kullanılarak ders seçim sebepleri, dersler ve ilgi alanlarının kendi aralarındaki ilişkileri elde edilmiştir.

BİRİNCİ BÖLÜM

KAPSAM VE LİTERATÜR

1.1. AMAÇ

Yapılan çalışmada amaç, üniversite öğrencilerinin seçmeli ders eğilimlerine ilişkin veri desenlerini veri madenciliği tekniklerinden birliktelik kuralları çıkarım algoritmalarından Apriori Algoritması kullanarak elde etmektir.

1.2. KAPSAM

Çalışmada örneklem, Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü'nde öğrenim gören 2. 3. ve 4. sınıf öğrencilerinden oluşmaktadır. Veri seti, öğrenciler için hazırlanan anket formu ile elde edilmiştir. Anket formunu toplamda 151 öğrenci cevaplamış ve toplamda 33 ders ile ilgili 862 değerlendirme elde edilmiştir.

1.3. LİTERATÜR ARAŞTIRMASI

Aşağıda veri madenciliği, birliktelik kuralları, öğrenci başarısı ve ders seçimi konularında incelenen literatür yıllara göre özetlenmiştir.

2013

Güngör, Yalçın ve Yurtay, üniversite öğrencilerinin teknik seçmeli ders seçimlerindeki nedenlerin belirlenmesi amacıyla 2013 yılında bir çalışma gerçekleştirmişlerdir. Bu amaçla bir anket hazırlamış ve hazırladıkları anketi Bilecik Şeyh Edebali Üniversitesi Bilgisayar Mühendisliği Bölümü öğrencileri üzerinde uygulamışlardır. Veri madenciliği tekniklerinden birliktelik kuralı çıkarım algoritması olan Apriori algoritmasıyla bazı kurallar elde etmişlerdir. Elde edilen kurallar, öğrencilerin ders seçimlerinde hangi hususları göz önünde bulundurdıkları, hangi derslerin seçilme eğiliminin yüksek olduğu, hangi derslerin birlikte seçildiği

gibi konularda bilgi içermektedir. Uygulamayı Matlab programlama dili ile gerçekleştirmişlerdir (Güngör ve diğerleri, 2013).

Budak tarafından 2013 yılında gerçekleştirilen çalışmada ise bilgisayar okuryazarlığını etkileyen faktörlerin belirlenmesi amacıyla veri madenciliği sürecinden de yararlanılarak birelektronik tablolama programı yardımıyla analiz edilmiştir. Çalışmada, Kocaeli Üniversitesi'nde 2012-2013 öğretim yılı güz döneminde sosyal bilimler ve sağlık bilimleri alanında lisans eğitim gören, Temel Bilgi Teknolojileri Kullanımı (TBTK) dersini alan öğrenciler arasından rastgele seçilen 144 örnekleme anket soruları uygulanmıştır. Elde edilen veriler, bir elektronik tablolama ve hesaplama programı olan Microsoft Office 2010 Excel programı kullanılarak analiz edilmiştir. Veri ön işleme teknikleri uygulandıktan sonra 132 öğrencinin cevapları değerlendirmeye alınmıştır. Anketten elde edilen verilere göre, ailenin ortalama aylık geliri, çocuk sayısı ankete cevap veren öğrencilerin bilgisayar sahipliği ve bilgisayar kullanım süreleri arasındaki korelasyonlara bakılarak düzeyleri analiz edilmiştir. Araştırma sonuçlarına göre, lisans öncesi eğitim süreçlerinin erken dönemlerinde de benzer içerikte eğitim almaları bilgisayar okuryazarlığı açısından olumlu etki yaratabileceği düşünülmektedir (Budak, 2013).

Taş, Adak ve Yurtay 2013 yılındaki çalışmalarında Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü'nde öğrenim görmüş/gören öğrencilerin 1999-2012 yılları arasındaki zorunlu donanım ve yazılım stajı kayıtları kullanılarak bölüm öğrencilerinin staj eğilimleri belirlenmeye çalışılmıştır. Çalışmada veri madenciliği algoritmalarından Apriori Algoritması kullanılarak birliktelik kuralları analizi gerçekleştirilmiştir. Elde edilen bulgulardan, staj etkinliğini arttırmaya ve bölüm staj politikasını iyileştirmeye yönelik kararlar almalarında destek sağlanmıştır (Taş ve diğerleri, 2013).

2011

Ünal, Ekim ve Köklü 2011 yılında yaptıkları çalışmada üniversite öğrencilerinin ortak zorunlu derslerdeki başarılarının analizini yapmışlardır. Çalışmada Selçuk Üniversitesi' nin 3 fakülte ve bir yüksekokulda okuyan öğrencilerinin verileri kullanılmıştır. Öğrenci notları, üniversitenin öğrenci işleri

biriminden ORACLE veri tabanı ortamında alınarak gerekli dönüşümler yapılmıştır. Elde edilen bu verilere SQL Server 2005 programı içindeki “Analysis Services” paketinde bulunan veri madenciliği modülleri, Visual Studio 2005 programı üzerinde çalıştırılarak analizleri yapılmıştır. Uygulamada hiyerarşik olmayan kümeleme tekniklerinden k-ortalama algoritması kullanılmıştır. İnceleme sonucunda sayısal bölümlerden oluşan Mühendislik Mimarlık Fakültesi öğrencilerinin sözel derslerde diğer fakülte ve yüksekokul öğrencilerine göre daha başarılı oldukları görülmüştür. Üniversiteye giriş puan türüne göre sözel olan Sosyal Bilimler Meslek Yüksek Okulu öğrencilerinin, sayısal ağırlıklı olan Mühendislik Mimarlık Fakültesi ve Fen Fakültesi öğrencilerine göre başarı oranlarının düşük olduğu görülmektedir. Sonuç olarak öğrencilerin üniversiteye giriş puanı ve durumlarının, ortak zorunlu derslerdeki başarıları ile doğru orantılı olduğu kanısınvarılmıştır (Ünal ve diğerleri, 2011).

Birtıl yüksek lisans tezinde kız meslek lisesi öğrencilerinin akademik başarısızlık nedenlerini veri madenciliği teknikleri ile incelemiştir. Uygulamada Kümeleme metodu ve algoritmaları kullanılmıştır. Araştırmada SPSS Clementine programı kullanılmıştır. Uygulama sonucunda öğrencilerin üç kümeye ayrıldığı gözlenmiştir. Sonuç olarak öğrencilerin başarısının rehberlik yardımıyla yükseltilebileceği sonucuna varılmıştır. Üç farklı küme için öğrencilerin başarılarının yükseltilmesi için ayrı ayrı stratejiler sunulmuştur (Birtıl, 2011).

Bir başka çalışmada veriler Boolean dönüşümü uygulanarak kullanılmıştır. Bu dönüşümle yapılan geliştirilmiş Apriori algoritmanın geleneksel Apriori algoritmasından farkı bütünleşik veri tabanını bir kez taramasıdır. Yapılan bu dönüşüm sayesinde veri tabanının bir kez taranması sonuçlara ulaşmak açısından faydalı olacaktır. Algoritma, Hadoop' un Platform üzerinde gerçekleştirilmiştir, böylece katlanarak algoritmanın etkinliğini artırabilir (Yu ve diğerleri, 2011).

2008

Dündar, öğrencilerin seçmeli derslerden hangisini alacaklarının belirlenmesi amacıyla Analitik Hiyerarşi Süreciteknğini kullanmıştır. Örneklem Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi İşletme bölümü 3. sınıfta okuyan 156

öğrenciden oluşturulmuştur. Bu amaçla öncelikle, öğrencilerin seçmeli ders seçiminde hangi kriterleri göz önüne aldıklarını belirlemiştir. Öğrenciler belirlenen kriterlerin, Analitik Hiyerarşi Süreci tekniğinin ölçeğine göre ikili karşılaştırmalarını yapmışlardır. Seçmeli derslerin her bir kritere ve bütün kriterlerin birlikte değerlendirilmelerine göre sıralaması belirlenmiştir (Dündar, 2008).

Seçmeli derslere ilişkin bir diğer çalışmada Tezcan ve Gümüş tarafından Gazi Üniversitesi Fen Edebiyat Fakültesi Kimya, Fizik ve Biyoloji bölümlerinde okuyan 3. ve 4. sınıf öğrencileri üzerinde gerçekleştirilmiştir. Öğrencilere likert tipi sorulardan oluşan 25 soruluk bir anket uygulanmıştır. Sonuçlar 3 bölüme göre ayrı ayrı değerlendirildiğinde en önemli faktörün, kimya bölümünde “öğrencinin öğretim üyeleri hakkındaki görüşleri”, fizik bölümünde, “daha önce aldığı ve başarılı olduğu derslere yakın olması”, biyoloji bölümünde “dersin işleniş yöntemi” olduğu, en az etkili faktörün ise her üç bölümde de “ailenin etkisi” olduğu görülmektedir (Tezcan ve Gümüş, 2008).

2007

Bir diğer çalışmada İbrahim ve Ruslı, öğrencilerin akademik başarılarının tahmini için yapay sinir ağlarını karşılaştırmışlardır. Çalışmada akademik performansın önemi, kümülatif not ortalaması ile ölçülmüştür. Çalışmada, öğrencilerin demografik profili ve lisans çalışmaları üzerinde ilk dönem kümülatif not ortalaması, yüksek lisans programındaki öğrencilerin akademik performansı için belirleyici değişken olarak kullanılmıştır. Uygulamada 206 öğrenciden elde edilen veri seti kullanılmıştır. Üç tahmini model; yapay sinir ağı, karar ağacı ve lineer regresyon SAS Enterprise Miner kullanılarak geliştirilmiştir. Çalışma üç modelinde %80 in üzerinde doğru sonuç ürettiğini göstermektedir. Üç model içerisinde en iyi performans gösteren teknik ise yapay sinir ağı tekniği olarak belirtilmiştir (İbrahim ve Ruslı, 2007).

Cavique, çalışmasında büyük veri setleri elde etmek için farklı bir yaklaşım önermiştir. Önerilen Similis algoritmasına göre özet veriler kullanılmakta ve böylece zaman karmaşıklığı giderilmektedir. Bu algoritma problemi grafik haline dönüştürmektedir. Çalışmada Similis algoritması sepet analizi konusuna

uygulanmıştır. Çalışmanın ilerleyen kısımlarında algoritma Apriori benzeri algoritmalarla perakendecilik sektörü için karşılaştırılmış, avantaj ve dezavantajlarına değinilmiştir (Cavique, 2007).

Durmuş ve İplikçi 2007 yılında yaptıkları çalışmada kümeleme algoritmalarını karşılaştırmış, kullanılan tüm veri setleri için en iyi kümelemeyi yapabilen bir algoritma bulunmadığı belirtilmiştir. Bunun sebebini, tüm kümeleme algoritmalarının performanslarının verilerin dağılımına bağlı olması olarak açıklamışlardır (Durmuş ve İplikçi, 2007).

2006

Durucasu, Aşan ve Er 2006 yılında üniversite öğrencilerinin yaz okuluna ilişkin genel bakışını ortaya koymak üzere çalışma yapmışlardır. Bu amaçla öğrencilerin genel profilinin belirlenmesi kümeleme analizi yardımı ile gerçekleştirilmiştir (Durucasu ve diğerleri, 2006).

2005

Erdoğan ve Timor çalışmalarında öğrencilerinin üniversite giriş sınavı sonuçlarıyla öğrencilerin başarılarının ilişkisini incelemişlerdir. Analiz için hiyerarşik kümeleme algoritmalarından k-ortalamlar algoritmasını kullanmışlardır. Matlab programının eklentisi MapToolBox kullanılarak kümeleme sonuçları gösterilmiştir (Erdoğan ve Timor, 2005).

2004

Karabatak ve İnce 2004 yılında yaptıkları çalışmada üniversite öğrencilerinin notlarını kullanarak başarı analizi yapmışlardır. Analiz için veri madenciliği algoritmalarından Apriori algoritmasını boolean verileri için kullanmışlar ve başarı kuralları ortaya çıkarmışlar. Uygulama Matlab ortamında bir program hazırlanarak yapılmıştır (Karabatak ve İnce, 2004).

2003

Tsay ve Chang-Chien yaptıkları çalışmada Apriori Algoritması ve CDAR tekniğini karşılaştırmışlardır. CDAR tekniğinin diğer bilinen birliktelik kuralı algoritmara göre daha büyük veri setlerine uygulanabilecek bir yöntem olduğunu belirtmişler. Bir uygulama ile çok büyük veri setlerinde Apriori Algoritması ve CDAR algoritmasının sonuçlarını göstermişlerdir ve çok büyük veri setleri için CDAR tekniğinin daha başarılı olduğunu göstermişlerdir (Tsay ve Chang-Chien, 2003).

1996

Demir, 1996 yılındaki çalışmasında seçmeli ders sisteminin öğrenciler ve öğretim üyeleri tarafından nasıl değerlendirildiğini araştırmıştır. Örneklem Orta Doğu Teknik Üniversitesi'nde yer alan 151 öğretim üyesi ve 1134 öğrenciden oluşturulmuş ve iki grup için ayrı ayrı formlar hazırlanarak uygulanmıştır. Öğretim üyelerine uygulanan form 20 sorudan, öğrencilere uygulanan form 25 sorudan oluşturulmuştur. Değerlendirme her soruya verilen yanıtların frekans ve yüzdesi çıkarılarak yapılmıştır. Sonuç olarak hem öğrencilerin hem de öğretim üyelerinin seçmeli ders sisteminden memnun oldukları belirlenmiştir (Demir, 1996).

Demir ve Ok 1996 yılındaki çalışmalarında, Orta Doğu Teknik Üniversitesi'nde uygulanan seçmeli ders sistemiyle ilgili öğretim üyeleri ve öğrencilerin görüşleri incelenmiştir. 91 öğretim üyesi 189 öğrenciye görüşleri sorulmuştur. İki grubun görüşleri ayrı ayrı incelenerek görüşlere ilişkin ortak noktalara dikkat çekilmiştir. Elde edilen bulgularla Orta Doğu Teknik Üniversitesi ve diğer üniversiteler için seçmeli ders uygulamasının iyileştirilmesine yönelik görüş ve öneriler sunulmuştur (Demir ve Ok, 1996).

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ

2.1. VERİ MADENCİLİĞİ TANIMI

Günümüz gelişen teknolojisi sayesinde kayıt altına alınan ve saklanan veri sayısı günden güne katlanarak artmaktadır. İnsanların her gün yaptıkları bir internet sayfasını ziyaret etmek gibi en basit hareketten, banka işlemlerine kadar her türlü işlem kayıt altına alınmakta ve saklanmaktadır. Ortaya çıkan miktarı oldukça hızlı bir şekilde artan bu çok büyük hacimdeki veriler veri tabanlarında, veri ambarlarında saklanılmaktadır (Freitas, 2002: 1).

Literatürde “Veri Madenciliği” ile ilgili yapılan birçok tanım bulunmaktadır. Tanımlardan bazıları şu şekildedir:

Veri madenciliği, büyük veri kümelerindeki yararlı desenleri ve eğilimleri keşfetme sürecidir (Larose D.T. ve Larose C.D., 2014: 2).

Giudici’ ye göre ise veri madenciliği, veri tabanı kullanıcılarına anlaşılır ve faydalı sonuçlar vermek amacıyla, büyük miktardaki verilerden daha önceden bilinmeyen ilişki ve kuralların keşfedilmesi sürecidir (Giudici, 2003: 2).

Veri madenciliği, kabul edilebilir sayısal verimlilik çerçevesinde, çok sayıdaki veri içerisinden anlamlı örüntüleri ortaya koyan veri analizi ve keşif algoritmalarının uygulanması ile oluşan bilgi keşfinin bir sürecidir (Fayyad ve diğerleri, 1996: 84).

Kantardzic ise veri madenciliğini yeni, değerli, belli bir önemi olan bilginin büyük veri içerisinden araştırılması olarak tanımlamıştır (Kantardzic, 2011: 2).

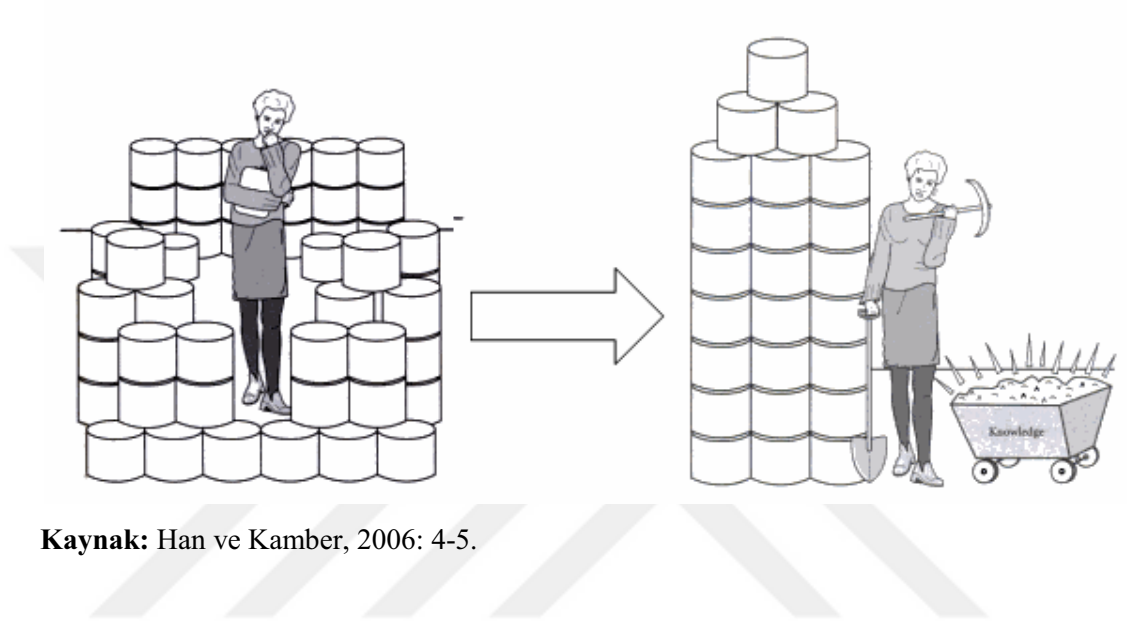
Veri madenciliği, büyük bir veri kümesinden gizli desenleri belirlemek için, farklı bakış açıları ile verileri analiz eden ve önemli bilgiler halinde özetleyen bir süreçtir (Borkar ve Rajeswari, 2013: 273).

Veri madenciliği, büyük veri setlerinden belli bir öneme sahip olan, önceden bilinmeyen, yararlı ve anlaşılır örüntülerin keşfidir (Büchhner ve diğerleri, 1997: 1).

Ho ve Park’ ın tanımına göre ise veri madenciliği, kullanıcının verdiği birçok kararı kapsayan ve çok sayıda adımı içinde bulunduran etkileşimli ve tekrarlamalı bir

süreçtir. Bu süreç problemin anlaşılması ve tanımı ile başlar ve rekabet avantajı kazanmak için sonuçları kullanır. Sonuçların analizi ve stratejilerin planlanması ile süreç sona erer (Ha ve Park, 1998: 4).

Şekil 1: Veri Madenciliği



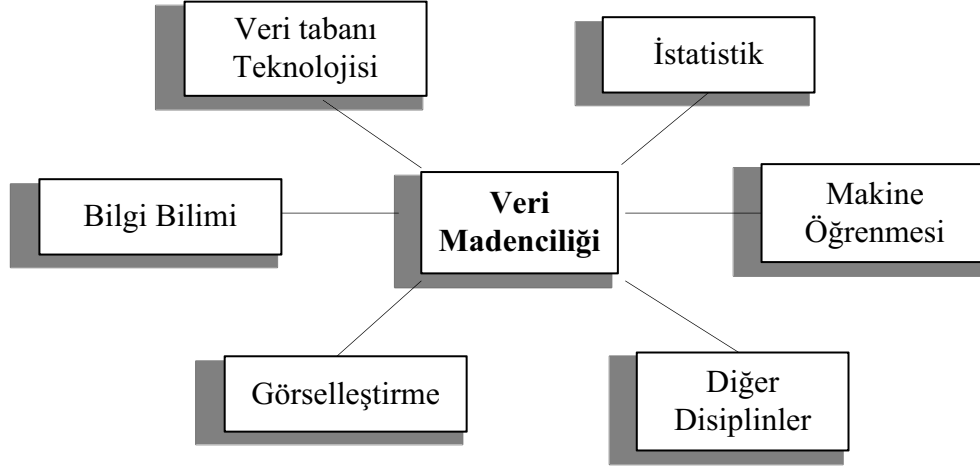
Kaynak: Han ve Kamber, 2006: 4-5.

Tanımlar incelediğinde hepsinin ortak noktasının çok büyük miktarda veriye sahip olunması ve bu büyük miktardaki veriler içerisinde anlamlı bilginin keşfedilmesi olduğu görülmektedir.

2.2. VERİ MADENCİLİĞİNİN DİĞER DİSİPLİNLERLE İLİŞKİSİ

Veri madenciliği, Şekil 2' de gösterildiği gibi veri tabanı teknolojisi, istatistik, makine öğrenmesi, görselleştirme, bilgi bilimi ve daha birçok konuyu kapsayan disiplinler arası bir alandır (Han ve Kamber, 2006: 29).

Şekil 2: Veri Madenciliği Ve Diğer Disiplinler



Kaynak: Han ve Kamber, 2006, 29.

2.2.1. Veri Madenciliği ve İstatistik

Veri madenciliği ve istatistiğin ortak amacı 'veriden öğrenme' dir. Her iki yaklaşımda da verilerden bir anlam çıkartmaya odaklanılır (Ganesh, 2002: 1).

Veri madenciliği ile istatistiksel veri analizini birbirinden farklı kılan en az üç özellik vardır (Giudici, 2003: 6). İstatistik ve veri madenciliği arasındaki en temel farklılık, veri madenciliğinin çok daha büyük veri kümeleri üzerinde çalışıyor olmasıdır. Bir istatistikçi için 'büyük' veri kümesi birkaç yüz veya bin veri içerirken, veri madenciliği ile çalışan kişiler için milyon veya milyarlık veri olası bir sayıdır. Büyük veri tabanları ile karşılaşmak gerçek hayatta sıklıkla karşımıza çıkabilecek bir durumdur.

İstatistik uygulamalarında veriler belli bir amaç doğrultusunda toplanmaktadır ve toplanan veriler belli sorulara yanıt bulmak amacıyla analiz edilmektedir. Veri madenciliğinde ise veriler belli bir amaç doğrultusunda toplanmamaktadır (Oğuzlar, 2003: 69).

Veri madenciliğinde hipoteze gerek duyulmazken, istatistikçiler genellikle çalışmalarında hipotezler kullanırlar.

İstatistiksel analizlerde sadece sayısal veriler kullanılır. Veri madenciliğinde ise metin ve ses gibi sayısal veriler dışında farklı tiplerde veriler de kullanılmaktadır (Demirel, 2010: 41).

2.2.2. Veri Madenciliği ve Makine Öğrenimi

Makine öğrenimi, yapay zekânın veri madenciliği ve istatistikle ilişkili bir dalıdır. Makine öğrenmesi teknikleri geçmişteki verileri kullanarak veriye en uygun modeli bulmaya çalışır, yeni gelen verileri de bu modele göre analiz ederler. Veri madenciliğinin de büyük miktarlardaki veriler üzerinde çalışarak, eldeki verilerden anlamlı ve işe yarayan bilgiler elde etmeyi amaçladığından daha öncede bahsetmiştik. Makine öğrenme teknikleri veri madenciliği algoritmalarında kullanılan tekniklerin temelini oluşturmaktadır (Yıldız, 2014: 24).

2.2.3. Veri Madenciliği ve Veri Tabanı

Veri madenciliği sorgularına girdi sağlamak amacıyla veri tabanı kullanılmaktadır. Veri tabanında var olan desenler için sorgular çalıştırılır. Veri madenciliğinde ise sorgular keşfe dayalıdır ve gizli desenleri keşfetmek için çalıştırılır (Tiryaki, 2015: 5).

2.3. VERİ MADENCİLİĞİNİN UYGULANABİLDİĞİ VERİ TÜRLERİ

Veri madenciliği her türlü depolama birimi üzerinde uygulanabilir. Veri ambarları (data warehouses), ilişkisel veri tabanları (relational databases), işlemsel veri tabanları (transactional databases), gelişmiş veri tabanları, dosyalar ve web üzerindeki hareketler bu depolama türlerinden bazılarıdır (Han ve Kamber, 2006: 9).

2.3.1. Veri Tabanları

Teknolojinin hızla gelişmesi son yıllarda elde edilen verinin hızla artmasına yol açmıştır. Günümüzde yapılan basit bir telefon konuşmasından internette ziyaret

edilen bir sayfaya, banka işlemlerinden markette yapılan alışverişlere kadar her türlü elektronik işlem kayıt altına alınmaktadır (Şentürk, 2006: 1).

İşte bu kadar detaylı şekilde verilerin kayıt altına alındığı düşünülürse devasa büyüklükteki veri yığınının saklanması, depolanması ve erişiminde klasik dosya sistemlerinin yetersiz kaldığı görülmektedir. Bu yetersizliği gidermek amacıyla veri tabanı kavramı ortaya çıkmıştır (Demirel, 2010: 4).

Birbiriyle ilişkisi olan verilerin tutulduğu, kullanım amacına uygun olarak düzenlenmiş veriler topluluğunun mantıksal ve fiziksel olarak tanımların olduğu bilgi depoları “veri tabanı” olarak tanımlanmaktadır. Veri tabanları da işlemsel veri tabanları ve ilişkisel veri tabanları olarak ikiye ayrılmaktadır (Şentürk, 2006: 5).

2.3.1.1. İlişkisel Veri Tabanları

İlişkisel veri tabanı, her birine benzersiz bir isim atanan tabloların bir bileşimidir. Her bir tablo niteliklerin bir kümesinden ve büyük bir kayıtlar setinden oluşur. İlişkisel bir tablodaki her bir kayıt benzersiz bir anahtar tarafından belirlenen ve bu nitelikler kümesi tarafından tanımlanan bir nesneyi temsil eder (Han ve Kamber, 2006: 10).

2.3.1.2. İşlemsel Veri Tabanları

İşlemsel veri tabanı, her bir kaydın bir işlemi temsil ettiği dosyalardan oluşur. Her bir işlem benzersiz bir kimlik numarası (islem_ID) ve işlemi oluşturan öğelerin listesini (örneğin, bir marketten satın alınan öğeler gibi) içermektedir. Tablo 1’ de örnek bir işlemsel veri tabanı tablosu gösterilmiştir (Han ve Kamber, 2006: 14).

Tablo 1: İşlemsel Veri Tabanı Örneği

İşlem ID	Öğelerin Listesi
I100	I1, I3, I8, I16
T200	I2, I8
...	...

Kaynak: Han ve Kamber, 2006: 15.

2.3.1.3. Gelişmiş Veri Tabanları

Veri tabanı teknolojisinin gelişmesi ile, gelişmiş veri ve bilgi sistemlerinin çeşitli türleri ortaya çıkmış ve yeni uygulamalar gereksinimi karşılayacak şekilde kullanılmaya başlanmıştır. Yeni veri tabanı uygulamaları, uzaysal veri (haritalar vb.), mühendislik tasarım verileri (bina tasarımları, sistem bileşenleri, devreler), çoklu ortam verileri, zaman eksenli veriler ve web verileri gibi veriler üzerinde işlem yapmaktadırlar. Bu tür uygulamalar, karmaşık nesne yapıları, değişken boyutlu kayıt yapıları, metin ve çoklu ortam verileri için verilerdeki dinamik değişimler açısından daha etkin veri yapıları gerektirmektedirler. Bu ihtiyaçlara cevap verebilmek için, gelişmiş veri tabanı sistemleri geliştirilmiştir. Nesne tabanlı veri tabanları, uzaysal veri tabanları, zaman serisi veri tabanları, metin veri tabanları gelişmiş veri tabanı sistemleridir (Han ve Kamber, 2006: 16).

2.3.2. Veri Ambarları

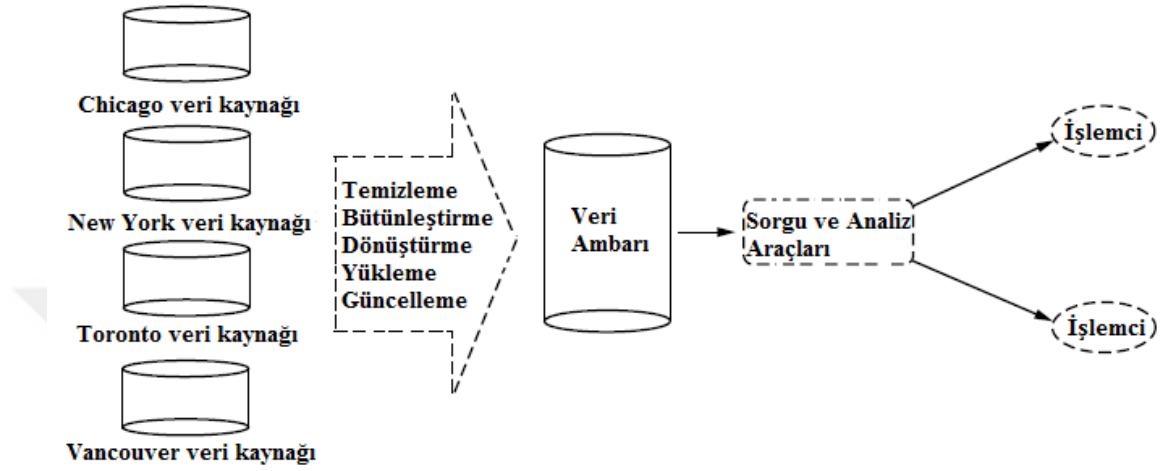
Birçok farklı kaynaktan elde edilen bilgilerin depolandığı arşivlere veri ambarı denilmektedir (Düzgünoğlu, 2006: 4).

Veri ambarları veri temizleme, veri bütünleştirme, veri dönüşümü, veri yükleme ve periyodik veri yenileme süreci yoluyla inşa edilir (Han ve Kamber, 2006: 10).

Bir veri ambarı bulunması veri madenciliği için bir ön koşul olmamasına rağmen, pratikte, veri madenciliği görevi, özellikle bazı büyük şirketler için, bir veri ambarına erişimi sağlayarak çok daha kolay yapılır. Bir veri ambarının temel amacı

bir karar sürecinin "akıllılığını" ve bu süreçte yer alan kişilerin bilgilerini artırmaktır (Kantardzic, 2011: 14).

Şekil 3: Tipik Bir Veri Ambarı Yapısı



Kaynak:Han ve kamber, 2006: 12.

Immon' un tanımına göre (1996) bir veri ambarı, temel olarak organizasyonel karar alımında kullanılan konu yönlendirmeli, bütünleştirilmiş, zaman değişimli, değişime uğramayan (kararlı) veri depolarıdır (Giudici, 2003: 20).

Tanımdan da anlaşılacağı gibi veri ambarları;

- Konu odaklıdır. Yani, aynı olayla ya da varlıkla ilgili veriler birbirine bağlanmıştır.
- Bütünleşiktir, birden fazla veri kaynağından gelen veriler aktarılarak veri tabanı ile bütünleştirilmiştir. Aynı zamanda, tekrar eden yer ve kişi adları tek bir alanda bir araya getirilmiş ve buna bağlı olarak dönüştürme ve normalizasyon işlemleri uygulanmıştır.
- Belirli bir döneme ve zaman dilimine aittir. Yani, eldeki veriler belli bir dönemin verilerini içerir (Örneğin; 5 yıllık veya son 10 yıllık veriler). Her veri bir şekilde zaman değişkeniyle ilişkili hale getirilmiştir.
- Geçici veya uçucu değildir. Yani, veri ambarlarındaki veriler silinmez, aynı zamanda yeni veri eklenmez. Yani, veri giriş çıkışı bulunmamaktadır (Silahtaroglu, 2013: 17-18).

Bir veri ambarı oluşturmak için iki yol vardır. İlk yaklaşım, bütün şirket bilgilerini toplayarak ve dışarıdan gelen bilgi ile bütünleştirerek, tek bir merkezi arşivin oluşturulmasına dayanmaktadır. İkinci yaklaşım başlangıçta kendi aralarında bağılı olmayan veri martı olarak adlandırılan farklı tematik veri tabanları ve veri ambarlarını bir araya getirmektir. İlk yaklaşım sistem yöneticilerine sürekli sunulan verilerin kalitesini kontrol etmeleri için olanak sağlar (Giudici, 2003: 21).

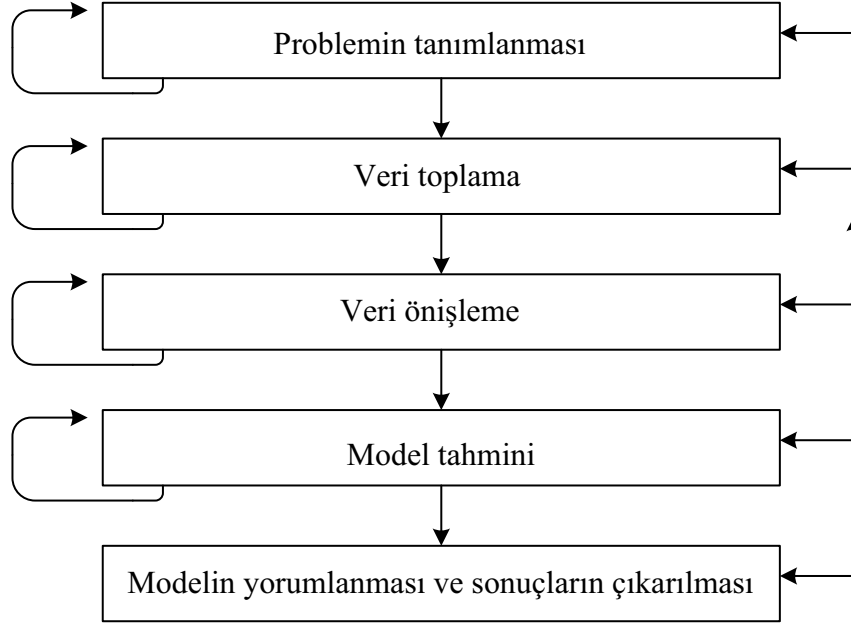
2.4. VERİ MADENCİLİĞİ SÜRECİ

Veri madenciliği süreci, verilerden doğru, yeni, faydalı ve anlaşılır modeller elde eden ve bu modelleri kullanarak anlamlı desenler elde etmekte kullanılan özel bir süreçtir.

Başarılı bir veri madenciliği projesinde izlenmesi gereken adımlar şu şekildedir (Kantardzic, 2011: 7-8):

1. Problemin tanımlanması,
2. Verilerin toplanması,
3. Veri ön işleme,
4. Modelin tahmini,
5. Modelin yorumlanması ve sunumu.

Şekil 4: Veri Madenciliği Süreci



Kaynak: Kantardzic, 2011: 9.

Veri madenciliği çalışmalarında başarılı olmak için uygulamanın amacı açık ve net bir şekilde tanımlanmalıdır. Daha sonra elde edilen verilere gerekli ön işlemler uygulanarak veriler uygun hale getirilmelidir. Bu aşama veri madenciliği sürecinin çok önemli bir adımını oluşturmaktadır. Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda model kurularak denenmesi sonucunda varılabilecek bir sonuçtur. Bu sebeple verilerin hazırlaması ve modelin kurulması aşamaları, en iyi olduğu düşünülen model elde edilinceye kadar tekrarlanan bir süreçtir. Modelin kurulması ve değerlendirilmesi sonucunda elde edilen sonuçlar hazırlanır ve süreç tamamlanır. Farklı veri kaynaklarından elde edilen veriler hazırlandıktan sonra hemen kullanılmayabilir. Bu yüzden hazırlanan veriler veri ambarında depolanabilir ve daha sonra başka problemin çözümü için kullanılabilir (Ayık ve diğerleri, 2007: 442; Gürgen, 2008: 7).

2.4.1. Verilerin Hazırlanması (Veri Önişleme)

Gerçek dünya verileri çok büyük miktarlarda ve dolayısıyla gürültülü, eksik ve tutarsız olma eğilimindedir. Verilerin kalitesinin düşük olması elde edilen sonuçlarında kalitesiz olmasına yol açmaktadır. Veri madenciliği uygulaması ile kaliteli sonuçlar elde etmek için analiz aşamasından önce verilere yönelik bazı işlemler gerçekleştirilmektedir (Han ve Kamber, 2006: 47).

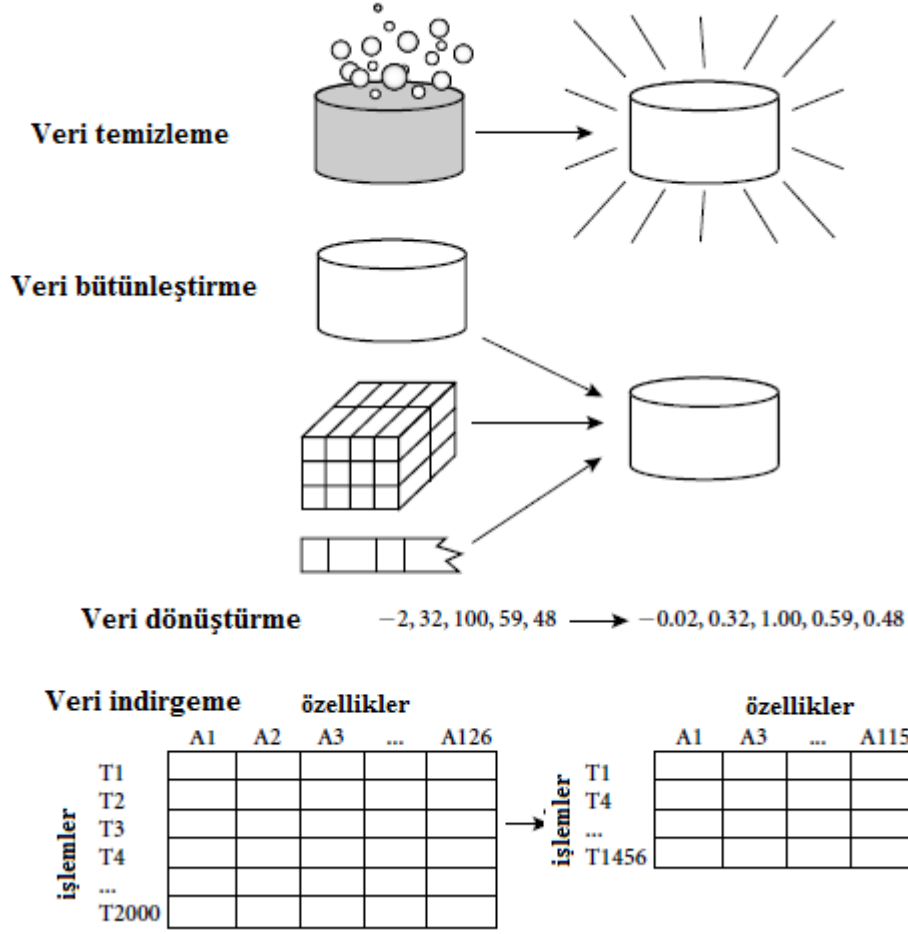
Veri ön işleme çok fazla verinin bulunduğu veri tabanı veya veri ambarlarındaki verileri analizden önce istatistiksel olarak sağlıklı hale getirmeyi amaçlamaktadır. Veriyi sağlıklı hale getirmek eksik, gürültülü, tutarsız, aykırı özellik taşıyanları belirlemek ve bu sorunları uygun tekniklerle ortadan kaldırmaktır (Şentürk, 2006: 11).

Veri madenciliği uygulamalarında kaynakların %80' i verilerin ön işleminden geçirilmesi ve temizlenmesi süreçleri için harcanmaktadır. Yani anlaşılacağı üzere veri önişleme veri madenciliğinin olmazsa olmaz bir adımı olmakla birlikte onu anlamlı kılan bir süreçtir (Piramuthu, 2004: 483).

Veri önişleme sürecinin basamakları şu şekilde sıralanabilir (Han ve Kamber, 2006: 47):

- Veri Temizleme
- Veri Birleştirme/ Bütünleştirme
- Veri Dönüştürme
- Veri İndirgeme

Şekil 5: Veri Önışleme Formları



Kaynak: Han ve Kamber, 2006: 50.

Şekil 5' te veri önışleme adımları gösterilmiştir.

2.4.1.1. Veri Temizleme

Veri temizlemede, eksik deęerler doldurularak, gürültülü veriler düzgünleřtirilerek, aykırı deęerler kaldırarak veya belirlenerek, tutarsızlıklar çözümlenerek veriler temiz hale getirilmeye çalıřılır (Han ve Kamber, 2006: 48).

Eksik deęerler; veri setindeki boş, girilmemiş deęerlerdir. Verilerin eksik olmasının birçok sebebi olabilmektedir. Veri tabanında yer alan verilerin anket verisi olması ve anket uygulanan kiřinin bilgi vermek istememesi, veri giriřini yapan

kişinin hatası, diğer veri özellikleriyle tutarsızlığı yüzünden verinin silinmesi gibi nedenler eksik veri oluşmasına neden olabilmektedir (Aydın, 2007: 28).

Herhangi bir değişkene ilişkin eksik değerlerin doldurulması için farklı yollar vardır. Eksik değerlerin bulunduğu kayıtlar için aşağıdaki adımlardan biri tercih edilebilir (Şentürk, 2006: 12):

1. Eksik değer içeren kayıt veya kayıtlar atılabilir.
2. İlgili değişkene ait ortalama eksik değerlerin yerine kullanılabilir.
3. Aynı sınıfa ait tüm örneklem için değişkenin ortalaması kullanılabilir.
4. Eksik değerlerin yerine en olası değer kullanılabilir.

Gürültü; büyük veri tabanlarında değerlerin birçoğu doğru olmayabilir. Veriler toplanırken ya da veri girişi yapılırken ortaya çıkan hatalara gürültü adı verilmektedir (Karagöz, 2007: 19). Gürültülü verilerin teşhis edilmesi ve düzleştirilmesi için uygulanabilecek tekniklerden bazıları bölümlenme, kümeleme ve regresyondur (Han ve Kamber, 2006: 63-64).

2.4.1.2. Veri Birleştirme

Veri birleştirme, farklı kaynaklı veri yapılarının uygun bir veri tabanında birleştirilmesidir. Veri birleştirilmesi esnasında dikkat edilmesi gereken bazı noktalar vardır. Şema birleştirme ve nesnelere eşleme hatalı olabilir. Örneğin; bir veri tabanındaki girişler 'tüketici-ID' şeklinde yapılmışken bir diğerinde 'tüketici-numarası' şeklinde olabilir. Bu tür problemler, varlık tanımlama problemi olarak adlandırılır ve üstesinden gelebilmek için meta veriler (veriye ilişkin veriler) kullanılır (Oğuzlar, 2003: 72).

Artıklık diğer bir önemli konudur. Bir değişken, başka bir değişkenden türetilmişse fazlalık olabilir. Değişkenlerdeki tutarsızlıklar da, sonuçta elde edilen veri kümesinde fazlalıklara neden olabilir. Bu fazlalıklar korelasyon analizi ile tespit edilebilir. Örneğin, daha önce bahsedilen "tüketici-ID" ile "tüketici-numarası" korelasyon katsayısına bakılabilir. Bu iki değişkene ait korelasyon katsayısı yüksek

ise, deęişkenlerden biri veri tabanından ıkarılarak indirgeme yapılır (Han ve Kamber, 2006: 67).

2.4.1.3. Veri Dönüştürme

Veri tabanlarında veya veri ambarlarındaki verilerin çoęunlukla orijinal biçimleri birbirinden farklı olabilmektedir. Veri dönüştürme ile veriler, veri madencilięi için uygun biçimlere getirilmektedirler. Veri dönüştürme; düzeltme, birleştirme, genelleştirme ve normalleştirme gibi deęişik işlemlerden biri veya bir kaçını içerebilir (Şentürk, 2006: 15).

- Düzeltme, verilerden gürültünün ayrıştırılmasını ve azaltılmasını sağlamaktadır. Bu tür tekniklere örnek olarak bölümlenme, kümeleme ve regresyon verilebilir.
- Birleştirme, verilerin özetlenmesi veya birleştirilmesi işlemlerini içerir. Örneęin, günlük satış deęerleri birleştirilerek aylık veya yıllık deęerler elde edilebilir.
- Genelleştirme, düşük düzeydeki deęişken veya ham verileri daha üst düzeyde deęişkenleri tanımlayacak şekilde dönüştürme işlemidir. Örneęin, yaş gibi sayısal nitelikli deęerler genç, orta yaşlı, yaşlı gibi üst düzey kavramlara eşlenmiş olabilir.
- Normalleştirme, en sık kullanılan veri dönüştürme işlemlerinden birisidir. Bu yöntemde amaç deęişken deęerlerini -1 ile 1 veya 0 ile 1 gibi belirli küçük aralıkta toplanmaktadır (Han ve Kamber, 2006: 70-71).

2.4.1.4. Veri İndirgeme

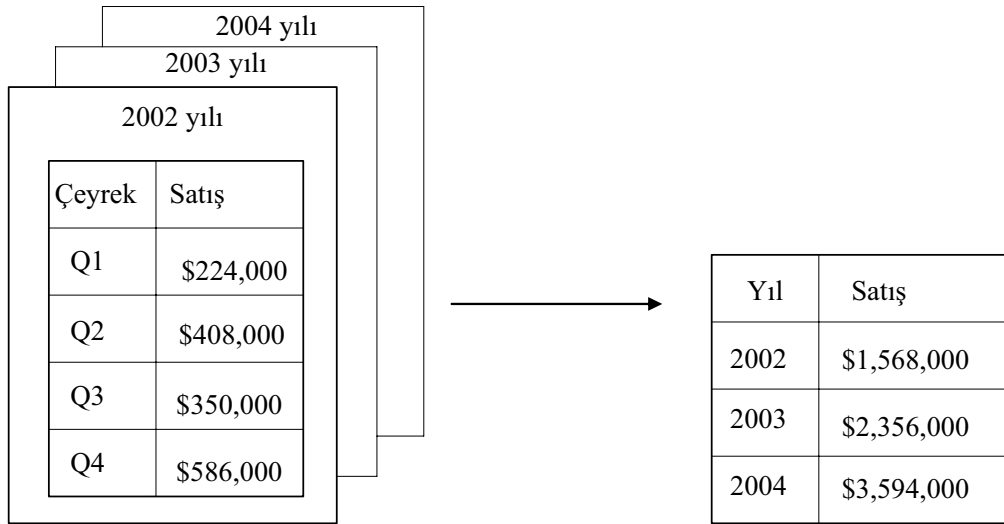
Veri indirgeme teknikleri ile veri setinin azaltılmış gösterimi elde edilir ve verilerin özgün nitelikleri korunur. Niteliklerdeki tutarsızlıklar elde edilen veri setinde fazlalıklara neden olabilir. Bu fazlalıkları tespit etmek için korelasyon analizi yapılmaktadır. Örneęin korelasyon analizine göre, müşteri kimlięi ile müşteri numarası isimli nitelikler arasında yüksek bir ilişki bulunursa deęişkenlerden birisi veri tabanından silinir ve indirgeme yapılır. Bu sayede elde edilen indirgenmiş veri

kümesine veri madenciliği teknikleri uygulanarak daha etkin sonuçlar elde edilebilir. (Oğuzlar, 2003: 73; Maimon ve Rokach, 2007: 95).

Veri indirgeme teknikleri aşağıdaki biçimde özetlenebilir (Oğuzlar, 2003: 73):

1. Veri birleştirme veya veri küpü
2. Boyut indirgeme
3. Veri sıkıştırma
4. Kesikli hale getirme

Şekil 6: Veri Birleştirme Örneği



Kaynak: Han ve Kamber, 2006: 74.

Han ve Kamber' in veri birleştirme için verdikleri örnekte sol tarafta 2002, 2003 ve 2004 yılına ait çeyreklik satış değerleri verilmiştir. Sağ tarafta ise veriler birleştirilerek yıllık satış değerleri elde edilmiştir (Han ve Kamber, 2006: 74).

2.5. VERİ MADENCİLİĞİ UYGULAMA ALANLARI

Veri madenciliği eldeki birçok veriden anlamlı bilgiyi elde etmeyi amaçlanmaktadır ve dolayısıyla birçok alanda veri madenciliğinden

yararlanılmaktadır. Veri madenciliğinin uygulama alanlarından bazıları aşağıda verilmiştir (Tuffery, 2011: 8-11; Demirel, 2010: 42-46; Han ve Kamber, 2006: 651-652)

Bankacılık sektöründe,

- Müşterilerin kredi riskleri hesaplanabilir ve böylelikle hangi müşterilerin riskli olduğu belirlenebilir.
- Kredi kartı harcamalarına göre, müşterilerden hangilerinin hangi kredi kartını kullandığı belirlenebilir ve geleceğe yönelik daha etkili kampanyalar düzenlenebilir.
- Kredi kartı dolandırıcılıkları önlenebilir.

Pazarlama sektöründe,

- Müşterilerin satın alma örüntüleri ile satın alma eğilimleri sistematik olarak analiz edilebilir. Aynı müşteriler tarafından farklı dönemlerde satın alınan mal dizileri gruplandırılabilir. Daha sonra veri madenciliği teknikleri ile eldeki müşterilerin korunması ve yeni müşterilerin kazanılması için stratejiler geliştirilebilir.
- İşletmeler en karlı ve en masraflı müşterilerini belirleyebilirler. En karlı müşterilerine özel kampanyalar geliştirebilir, en masraflı müşteriler masrafsız müşteri haline dönüştürülebilirler.
- Pazarlama sektöründe en çok kullanılan uygulamalardan biri de Pazar Sepet Analizidir. Bu analizde genellikle marketlere veya büyük alışveriş merkezlerine giden müşterilerin hangi ürünleri birlikte aldıkları veri madenciliği algoritmaları ile analiz edilir ve marketler bu şekilde en çok beraber satılan ürünleri aynı reyona koyarak satışlarını belli bir oranda arttırma imkânına sahip olurlar.
- Çapraz satış analizleri ve satış tahminleri yapılabilir.

Perakendecilik sektöründe,

- Perakende sektöründe de satış, müşterilere ait alışveriş detayları, nakliye, tüketim ve hizmet verilerinin toplandığı büyük veri kümelerinin olmasından dolayı veri madenciliğinin kullanımı yaygındır.
- Perakende sektöründe ürünlerin tanıtımı ve müşteri çekmek amacıyla reklam ve kupon gibi çeşitli satış kampanyaları düzenlenmektedir. Veri madenciliği teknikleriyle kampanya öncesi ve sonrası ele alınarak, kampanyaların etkinliği analiz edilebilir.

Diğer sektörlere ait uygulamalar,

- Mobil telefon sektöründe faturaların ödenmemesinin önlenmesi,
- Trafik tahminleme (gün ve saat bazında),
- Su veya elektrik tüketimi tahminleme,
- Bir telefon şebekesinin kalitesinin iyileştirilmesi (Başarısız aramaların nedenlerini keşfetmek);
- En uzun vadeli işsizlik riski altındaki işsiz kişileri tespit etmek ve kendi kişisel koşullarına uygun acil yardım sağlamak amacıyla iş arayanların profillemesi,
- Tıp alanında gen haritalarının incelenmesi ve kişilerin hastalık risklerinin belirlenmesi,

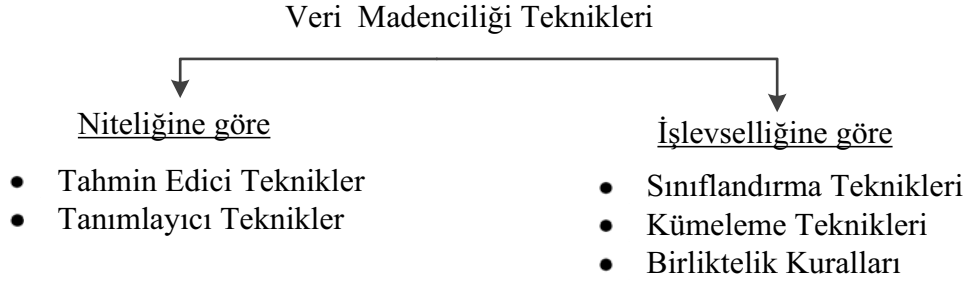
Ve daha birçok alanda veri madenciliği uygulamalarına rastlanılabilmektedir.

2.6. VERİ MADENCİLİĞİ TEKNİKLERİ

Veri madenciliği teknikleri 2 yaklaşıma göre gruplandırılırlar. Bu yaklaşımların ilki tekniklerin niteliklerine göre gruplandırılması, ikincisi ise işlevselliklerine göre gruplandırılmasıdır (Köktürk ve Dirsehan, 2012: 6).

Veri madenciliğinde kullanılan tekniklerin gruplandırılması Şekil 7’ de gösterilmiştir.

Şekil 7: Veri Madenciliği Tekniklerinin Gruplandırılması



Kaynak: Köktürk ve Dirsehan, 2012: 6.

Tanımlayıcı tekniklerde amaç eldeki verileri kullanarak gizli ilişkileri tanımlamak, kümeleri ve veriyi niteleyebilecek özellikleri keşfetmektir. Tahmin edici tekniklerde ise amaç, sonuçları bilinen eldeki veriler kullanılarak bir model geliştirilmesi ve elde edilen modelden yararlanılarak, sonuçları bilinmeyen veri kümeleri için yeni sonuçlar elde edilmesidir. Yapılan çalışmada ele alınacak tekniklerden kümeleme analizleri ve birliktelik kuralları tanımlayıcı teknikler sınıflandırma teknikleri ise tahmin edici teknikler kapsamına girmektedir (Özek ve diğerleri, 2010: 102; Gürsoy, 2010: 5).

Veri madenciliği konusunda çok sayıda teknik ve algoritma geliştirilmiştir. Bu tekniklerin birçoğu istatistiksel tabanlıdır. Yapılan çalışmada veri madenciliği modelleri temel olarak üç başlık altında toplanacaktır. Söz konusu veri madenciliği modelleri temel olarak şunlardır (Özkan, 2013: 44):

1. Sınıflandırma
2. Kümeleme
3. Birliktelik Kuralları.

2.6.1. Kümeleme Teknikleri

Kümeleme yaklaşımı, verileri mantıklı gruplar halinde düzenlemek için geliştirilmiş en temel yaklaşımlardan birisidir (Kantardzic, 2011: 250)

Kümeleme tekniğinde gruplar arası varyansın maksimum, gruplar içi varyansın minimum yapılması amaçlanır. Yani, kümeler kendi içinde benzer öğeleri bir araya getirirken, kümelerin diğer kümelerden öğeler açısından olabildiğince farklılaştırılması amaçlanmaktadır (Liu ve Shih, 2005: 387).

Kümeleme analizi pazar araştırması, örüntü tanıma, veri analizi ve görüntü işleme de dahil olmak üzere çok sayıda uygulamada yaygın olarak kullanılmaktadır (Han ve Kamber, 2006: 384).

9 araca ilişkin bilgilerin bulunduğu Tablo 2’ deki verileri ele alalım (Gürsoy, 2012: 15-16).

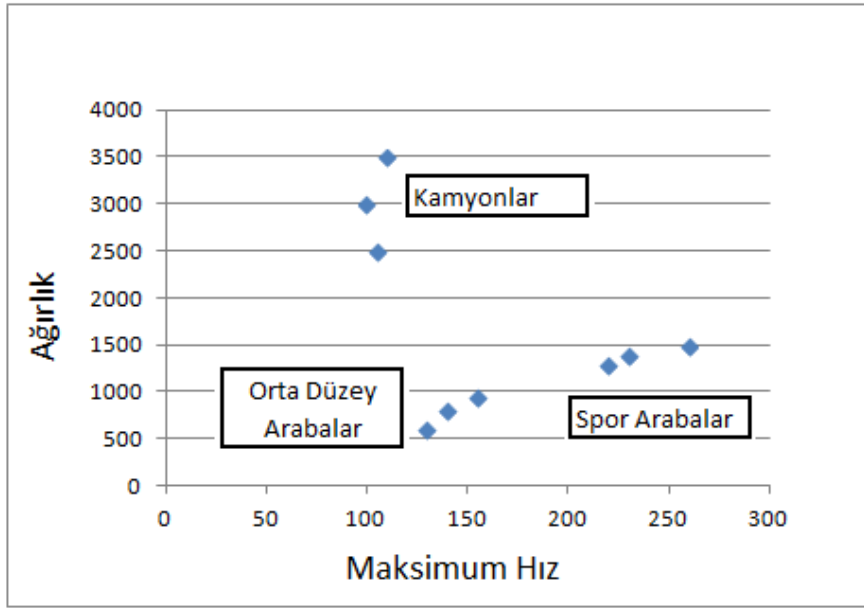
Tablo 2: Kümeleme Tekniği Veri Seti

Araç	Maksimum Hız (km/s)	Renk	Ağırlık (kg)
Araç 1	220	Kırmızı	1300
Araç 2	230	Siyah	1400
Araç 3	260	Kırmızı	1500
Araç 4	140	Gri	800
Araç 5	155	Mavi	950
Araç 6	130	Beyaz	600
Araç 7	100	Siyah	3000
Araç 8	105	Kırmızı	2500
Araç 9	110	Gri	3500

Kaynak: Gürsoy, 2012: 15-16.

Veri setindeki kayıtları kümelemek istediğimizi düşünelim. Bu amaçla, maksimum hız(km/s) ve ağırlık (kg) değişkenleri dikkate alınır ve bu değişkenlerle bir serpilme diyagramı çizilirse, Şekil 8’ deki grafik elde edilmektedir.

Şekil 8: Kümeleme Tekniği Serpilme Diyagramı



Kaynak: Gürsoy, 2012, 15-16.

Serpilme diyagramına bakıldığında veri setinin üç kümeye ayrılacağı açıkça anlaşılmaktadır. Elde edilen kümeler:

- Spor arabalar: Ağırlığı hafif, hızı yüksek olan araçlar,
- Kamyonlar: Ağırlığı fazla, hızı düşük olan araçlar,
- Orta düzey arabalar: ağırlığı hafif, hızı da ortada olan araçlar.

Veri kümesi çok daha fazla veriyi kapsadığında şekildeki gibi kümeleri bir bakışta anlamak mümkün olmamaktadır. Bunun için farklı teknikler geliştirilmiştir (Kantardzic, 2011: 251).

Literatürde pek çok kümeleme algoritması bulunmaktadır. Kullanılacak olan algoritmanın seçimi, veri tipine ve amaca göre değişmektedir. Başlıca kümeleme teknikleri şu şekildedir (Han ve Kamber, 2006: 408):

1. Bölümleyici Teknikler
2. Hiyerarşik Teknikler
3. Yoğunluk Tabanlı Teknikler
4. Izgara Tabanlı Teknikler
5. Model Tabanlı Teknikler

2.6.1.1. Bölümleyici Teknikler

Bölümleyici kümeleme tekniklerinde veri kümesinden n tane giriş parametresi kabul edilir ve bu parametreler k tane algoritmaya bölünür. Bölümleyici tekniklere ait algoritmalarda merkez noktaların kümeyi temsil etmesi esasına dayanılır. Bu tekniğe ait algoritmalar elde edilen sonuçların verimli olması ve kolay uygulanabilmesi açısından yaygın bir şekilde kullanım alanı bulmaktadır. Bölümleyici tekniğe ait algoritmalar: k -ortalamalar, k -medoids, c -ortalamalar, clara ve clarans algoritmalarıdır (Alzand ve Karacan, 2014: 2).

2.6.1.2. Hiyerarşik Teknikler

Hiyerarşik teknikler öğeleri ağaç yapısı şeklinde gruplandırma temeline dayanır. Hiyerarşik teknikler giriş parametresi olarak, elde edilecek küme sayısını belirten k değerine ihtiyaç duymazlar, ancak ağaç yapısı oluşturma işleminin ne zaman durdurulacağını belirten eşik değeri parametresine ihtiyaç duyarlar (Han ve Kamber, 2006: 408).

2.6.1.3. Yoğunluk Tabanlı Teknikler

Yoğunluk tabanlı teknikler, öğelerin doğal dağılımını bir yoğunluk fonksiyonu aracılığı ile tespit eder. Daha sonra belirlenen bir eşik yoğunluğunu aşan bölgeleri belirleyerek küme olarak adlandırır. Yoğunluk tabanlı algoritmalar düzgün şekilli olmayan kümeleri bulmadaki üstünlüğü, gürültü ve istisnalardan etkilenmeme ve tek tarama ile sonuca ulaşma gibi avantajlarından dolayı en başarılı kümeleme metotları arasında yer almaktadırlar (Gülce, 2010: 23).

2.6.1.4. Izgara Tabanlı Teknikler

Izgara tabanlı kümeleme teknikleri çok çözünürlüklü ızgara veri yapısını kullanır. Kullandıkları ızgara yapısından dolayı veri tabanındaki nesne sayısından bağımsızdırlar. Performanslarını etkileyen tek unsur kullanılan kare sayısıdır. Kare

sayısı arttıkça hesaplama zamanı artar ve dolayısıyla performans düşer. Izgara tabanlı tekniklerin en önemli avantajları işlem yükü az olduğu için hızlı ve çabuk sonuca ulaşabilmeleridir (Han ve Kamber, 2006: 424).

2.6.1.5. Model Tabanlı Teknikler

Model tabanlı teknikler verileri bir matematiksel model ile ifade etmeye çalışırlar. Bu tür teknikler kümelerin her biri için bir model varsayımında bulunur ve verilen model için verinin en uygununu bulurlar. En iyi bilinen ve en çok kullanılan bölme teknikleri k- ortalamalar tekniği, k-medoidstekniği ve bunların varyasyonlarıdır (Han ve Kamber, 2006: 429).

2.6.2. Sınıflandırma Teknikleri

Sınıflandırma veri madenciliğinde sık kullanılan tekniklerden biridir. Bu teknik verilerdeki gizli örüntülerin ortaya çıkarılmasını amaçlar.

Verilerin sınıflandırılması için 2 adımdan oluşan bir süreç izlenir. İlk olarak, eldeki verilerin bir kısmı eğitim amaçlı kullanılarak sınıflandırma kuralları oluşturulur. İkinci adımda ise elde edilen kurallar kullanılır ve bu kurallar yardımıyla yeni bir durum ortaya çıktığında nasıl karar verileceği belirlenir (Özkan, 2013: 45).

Aşağıda Özkan' ın örneğiyle sınıflandırma tekniklerinden basit bir karar ağacının nasıl oluşturulacağı gösterilmiştir. Örnekte bir bankanın kredi verdiği müşterilerinin risk durumunu belirlemek amacıyla karar ağaçları tekniği kullanılmıştır. Bu teknikle eldeki verilerle elde edilen kurallar ve karar ağacı yardımıyla yeni bir müşteri talebi geldiğinde kredi verilip verilmeyeceğinin kararı verilebilecektir.

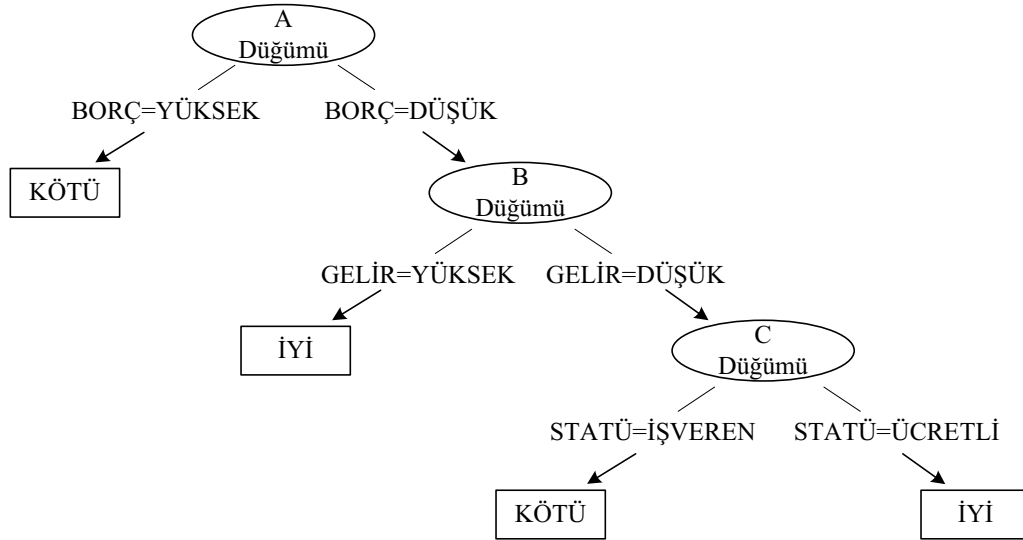
Tablo 3:Sınıflandırma Tekniği Eğitim Verileri

MÜŞTERİ	BORÇ	GELİR	STATÜ	RİSK
1	YÜKSEK	YÜKSEK	İŞVEREN	KÖTÜ
2	YÜKSEK	YÜKSEK	ÜCRETLİ	KÖTÜ
3	YÜKSEK	DÜŞÜK	ÜCRETLİ	KÖTÜ
4	DÜŞÜK	DÜŞÜK	ÜCRETLİ	İYİ
5	DÜŞÜK	DÜŞÜK	İŞVEREN	KÖTÜ
6	DÜŞÜK	YÜKSEK	İŞVEREN	İYİ
7	DÜŞÜK	YÜKSEK	ÜCRETLİ	İYİ
8	DÜŞÜK	DÜŞÜK	ÜCRETLİ	İYİ
9	DÜŞÜK	DÜŞÜK	İŞVEREN	KÖTÜ
10	DÜŞÜK	YÜKSEK	İŞVEREN	İYİ

Kaynak: Özkan, 2013: 45.

Karar ağacı aşağıda görüldüğü biçimde oluşturulabilir.

Şekil 9:Sınıflandırma Tekniği Eğitim Verilerine Uygun Karar Ağacı



Kaynak: Özkan, 2013: 46.

Elde edilen karar ağacı karar kuralları oluşturulmasında kullanılabilir.

Tablo 4:Sınıflandırma Tekniği Eğitim Verilerinden Oluşturulan Kurallar

KURAL 1: Eğer BORÇ = YÜKSEK ise RİSK = KÖTÜ	KURAL 3: Eğer BORÇ = DÜŞÜK ise ve Eğer GELİR = DÜŞÜK ise ve Eğer STATÜ = İŞVEREN ise RİSK = KÖTÜ
KURAL 2: Eğer BORÇ = DÜŞÜK ise ve Eğer GELİR = YÜKSEK ise RİSK = İYİ	KURAL 4: Eğer BORÇ = DÜŞÜK ise ve Eğer GELİR = DÜŞÜK ise ve Eğer STATÜ = ÜCRETLİ ise RİSK = İYİ

Kaynak: Özkan, 2013: 46.

Yeni bir müşteri geldiğinde eğitim kümesinden elde edilen kural tablosu kullanılarak müşteri hakkında karar verilebilir.

Temel sınıflama algoritmaları aşağıda verilmiştir (Gülce, 2010: 18):

1. Diskriminant analizi,
2. Naivebayes,
3. Karar ağaçları,
4. Sinir ağları,
5. Kaba kümeler,
6. Genetik algoritma,
7. Regresyon analizi.

2.6.2.1. Diskriminant Analizi

Diskriminant analizinde bir dizi gözlem önceden tanımlanmış sınıflara atanır. Model, ait oldukları sınıf bilinen ve öğrenme kümesi olarak adlandırılan gözlem kümesi üzerine kurulur. Öğrenme kümesi esas alınarak diskriminant fonksiyonu olarak bilinen doğrusal fonksiyonların bir kümesi oluşturulur. Diskriminant fonksiyonu kullanılarak yeni gözlemlerin ait olduğu sınıf belirlenir. Yeni bir gözlem olduğunda bütün diskriminant fonksiyonları hesaplanır ve yeni gözlem diskriminant fonksiyonunun değerinin en yüksek olduğu sınıfa atanır (Gülce, 2010: 18).

2.6.2.2. Naive Bayes

Naive Bayes, hedef deęişkenle baęımsız deęişkenler arasındaki ilişkiyi analiz eden bir sınıflandırma algoritmasıdır. Naive Bayes, sürekli veriler ile çalışmaz. Bundan dolayı Naive Bayes Tekniğinde, sürekli deęerleri içeren baęımlı ya da baęımsız deęişkenler kategorik hale getirilmelidir. Örneęin; baęımsız deęişkenlerden biri yaş ise, sürekli deęerler “<20”, “21–30”, “31–40” gibi yaş aralıklarına dönüştürülerek kategorik hale getirilmelidir.

Naive Bayes, modelin öğrenilmesi esnasında, her çıktının öğrenme kümesinde kaç kere meydana geldiğini hesaplar. Bulunan deęer, öncelikli olasılık olarak adlandırılır. Örneęin; bir banka, kredi kartı başvurularını “iyi” ve “kötü” risk sınıflarında gruplandırmayı amaçlamaktadır. İyi risk çıktısı toplam 5 vaka içinde 2 kere meydana geldiyse, iyi risk için öncelikli olasılık 0,4 olarak elde edilir. Elde edilen deęer şu şekilde yorumlanır: “Kredi kartı için başvuran biri hakkında hiçbir şey bilinmiyorsa, bu kişi 0,4 olasılıkla iyi risk grubundadır.” Naive Bayes aynı zamanda her baęımsız deęişken/ baęımlı deęişken birlikteliğinin meydana gelme sıklığını elde eder ve bu sıklıklar öncelikli olasılıklarla birleştirilerek tahminde kullanılır (Akbulut, 2006: 21-22).

2.6.2.3. Karar Ağaçları

Karar ağaçları ile sınıflandırma temel olarak bir karar ağacının oluşturulması, veri tabanındaki her kaydın bu ağaca uygulanması ve çıkan sonuca göre kaydın sınıflandırılmasına dayanır. Dięer tekniklerle karşılaştırıldığında karar ağaçlarının oluşturulması ve anlaşılması daha kolaydır. Karar ağacı oluşturulurken ağacın şekli kullanılan algoritmaya göre deęişmekte ve deęişik ağaç yapıları farklı sonuçlar verebilmektedir. Karar ağaçlarına baęlı olarak geliştirilen birçok algoritma bulunmaktadır. Bu algoritmalar kök, düğüm ve dallanma kriteri seçimlerinde izledikleri yol açısından farklılık göstermektedirler (Tapkan ve dięerleri, 2011: 249).

En sık kullanılan karar ağacı algoritmaları CHAID (Chi-Squared Automatic Interaction Detector, Kass 1980), C&RT (Classification and Regression Trees,

Breiman ve Friedman, 1984), ID3 (Induction of Decision Trees, Quinlan, 1986), C4.5 (Quinlan, 1993) algoritmalarıdır (Ay, 2009: 19).

2.6.2.4. Yapay Sinir Ağları

Yapay sinir ağları, biyolojik sinir ağlarından esinlenerek geliştirilmiş bir bilgi işleme sistemidir. Yapay sinir ağları, yapay sinir hücrelerinin birbirleriyle çeşitli şekilde bağlanmalarından oluşur ve genellikle katmanlar şeklinde düzenlenir (Silahtaroglu, 2013: 121-122).

Yapay sinir ağları doğrusal bir programın gerçekleştiremediği karmaşık görevleri de gerçekleştirebilir. Yapay sinir ağlarının çalışması için eğitime ihtiyacı vardır. Eğitim sürecinden sonra öğrenir ve yeniden programlanması gerekmez (Bajpai ve diğerleri, 2011: 27).

Yapay sinir ağları uygulama alanı oldukça geniş olan bir tekniktir ve öngörü, sınıflandırma ve modelleme alanlarında daha yoğun kullanılmaktadır. Son yıllarda kredi derecelendirme, savunma, sigortacılık, güvenlik, işlem modelleme ve yönetimi, kalite kontrol, ekonomik öngörü, konuşma ve yapı tanımlama gibi alanlarda da sıklıkla kullanılmaktadır (Şentürk, 2006: 33).

2.6.2.5. Kaba Kümeler

Kaba küme teorisi 1982 yılında Pawlak tarafından geliştirilmiştir (Pawlak, 1982: 341). Kaba küme teorisi matematiksel bir yaklaşımdır. Kaba kümeler amacı kesin olmayan ve belirsiz veri kaynaklarından kurallar biçiminde bilgi keşfetmektir. Klasik küme teorisinde küme tek olarak elemanları ile tanımlanmakta ve kümenin elemanları hakkında hiçbir ilave bilgi bulunmamaktadır. Kaba küme teorisi ise klasik küme teorisinin aksine, bir kümenin tanımlanabilmesi için başlangıçta evrenin elemanları hakkında bazı bilgilere gereksinim olduğu varsayımına dayanmaktadır. Nesnelere aynı bilgi ile nitelendiriliyorsa aynıdır veya ayırt edilemezdir. Ortaya konulan farksızlık ilişkisi, kaba küme teorisinin temelini oluşturmaktadır. Teori genellikle deneysel verilerle oluşturulmuş ikili ilişkileri kullanarak kümelerin veya

kavramların bir yaklaşımını sağlar (Aydoğan ve Gencer, 2007: 18; Olson ve Delen, 2008: 87)

2.6.2.6. Genetik Algoritmalar

Genetik algoritmalar doğadaki bilinen evrim fikrinden esinlenilerek geliştirilmiş algoritmalarlardır. Karmaşık çok boyutlu arama uzayında en güçlünün hayatta kalması ilkesine göre en iyi çözümü arar (Olson ve Delen, 2008: 125).

Probleme ilgili parametreler, genler halinde bir araya gelerek kromozomu oluşturmaktadır. Kromozomlardan bir başlangıç popülasyonu oluşturulur ve genetik algoritmanın üç işlemi olan çapraz değişim, mutasyon ve seçim kullanılarak yeni bir nesil oluşturulur. Genlerin kromozom içindeki dizilişi değiştirilerek yeni nesiller oluşturulmakta ve en iyi kromozoma ulaşılması amaçlanmaktadır (Hong ve diğerleri, 2014: 657).

2.6.2.7. Regresyon Analizi

Regresyon analizi bir veya daha fazla bağımsız değişken ile bağımsız değişken arasındaki ilişkiyi modellemek için kullanılan bir tekniktir. Bağımsız değişkenin bir tane olduğu regresyon tek değişkenli regresyon analizi, birden fazla olduğu regresyon çok değişkenli regresyon analizi olarak adlandırılmaktadır (Han ve Kamber, 2006: 354)

2.6.3. Birliktelik Kuralları

Birliktelik Kuralları en önemli veri madenciliği tekniklerinden biridir. Birliktelik kurallarının amacı veri tabanlarındaki önemli ilişkileri keşfetmektir (Mansoul ve Atmani, 2014: 105).

ÜÇÜNCÜ BÖLÜM

BİRLİKTELİK KURALLARI

3.1. BİRLİKTELİK KURALLARININ TANIMI

Birliktelik kuralları, büyük veri kümeleri arasındaki birliktelik ilişkilerini bulmaya yönelik çalışan bir veri madenciliği tekniğidir (Thabtah ve diğerleri, 2006: 414).

Veri miktarlarındaki artış sonucunda elde edilen veriler arasındaki ilişkilerin ortaya çıkarılarak anlamlı sonuçlar elde etme ihtiyacı artmıştır. Büyük kuruluşlar için karar verme açısından ilişkilerin ortaya koyulması çok büyük önem arz etmekte ve verimliliği arttırmaktadır (Silahtaroglu, 2013: 139).

Birliktelik kurallarının en önemli uygulama alanları arasında sepet analizi, çapraz-pazarlama, promosyon analizleri, katalog ve yerleşim düzeni tasarımları bulunmaktadır. Sepet analizi ile müşterilerin hangi ürünleri birlikte almaya eğilimli olduklarının belirlenmesi amaçlanmaktadır. Birlikte satın alınan ürünlerin belirlenmesiyle müşterilere yönelik stratejiler belirlenebilir ve dolayısıyla daha fazla ürün satılmasını sağlar. Böylece, bu şekilde firmanın karını arttırması sağlanmış olur (Erpolat, 2012: 138). Sepet analizi müşteri alışveriş verilerine uygulanabileceği gibi bir bankanın, bir seyahat acentesinin verilerine de uygulanabilir (Şimşek Gürsoy, 2012: 14).

3.2. BİRLİKTELİK KURALLARI İÇİN VERİ GÖSTERİMİ

Sepet analizi verilerinin işlemsel veri formatı veya çizelge tipi veri formatı kullanarak ele alınan iki temel gösterimi vardır. İşlemsel veri formatında sadece iki alan gerekmektedir. Bunlar her kayıt için sadece tek bir öğeyi temsil eden ID ve içerik alanıdır. Örneğin, Tablo 5' teki veriler Tablo 6' da işlemsel veri formatı şeklinde gösterilmiştir (Larose D.T. ve Larose C.D, 2014: 248).

Tablo 5: SebzeStandında Yapılan İşlemler

İşlem	Satın alınan öğeler
1	Brokoli, yeşil biber, mısır
2	Kuşkonmaz, kabak, mısır
3	Mısır, domates, fasulye, kabak
4	Yeşil biber, mısır, domates, fasulye
5	Fasulye, kuşkonmaz, brokoli
6	Kabak, kuşkonmaz, fasulye, domates
7	Domates, mısır
8	Brokoli, domates, yeşil biber
9	Kabak, kuşkonmaz, fasulye
10	Fasulye, mısır
11	Yeşil biber, brokoli, fasulye, kabak
12	Kuşkonmaz, fasulye, kabak
13	Kabak, mısır, kuşkonmaz, fasulye
14	Mısır, yeşil biber, domates, fasulye, brokoli

Kaynak:Larose D.T. ve Larose C.D, 2014: 249.

Çizelge tipi veri formatında, her bir kayıt ayrı bir işlemi temsil eder. Öğeler 0/1 şeklinde ikili değişken olarak kullanılmaktadır.

Tablo 5' in verileri Tablo7' de çizelge halinde veri formatına uygun olarak gösterilmiştir.

Tablo 6:Sebze Standı Verileri İçin İşlemsel Veri Formatı

İşlem	Satın alınan öğeler
1	Brokoli
1	Yeşil biber
1	Mısır
2	Kuşkonmaz
2	Kabak
...	...

Kaynak:Larose D.T. ve Larose C.D, 2014: 249.

Tablo 7: Sebze Standı Verileri İçin Çizelge Tipi Veri Formatı

İşlem	Kuşkonmaz	Fasulye	Brokoli	Mısır	Yeşil biber	Kabak	Domates
1	0	0	1	1	1	0	0
2	1	0	0	1	0	1	0
3	0	1	0	1	0	1	1
4	0	1	0	1	1	0	1
5	1	1	1	0	0	0	0
6	1	1	0	0	0	1	1
7	0	0	0	1	0	0	1
8	0	0	1	0	1	0	1
9	1	1	0	0	0	1	0
10	0	1	0	1	0	0	0
11	0	1	1	0	1	1	0
12	1	1	0	0	0	1	0
13	1	1	0	1	0	1	0
14	0	1	1	1	1	0	1

Kaynak: Larose D.T. ve Larose C.D, 2014: 250.

Birliktelik kuralları genel olarak sadece 0-1 ikili değişkenler için incelenmiştir. Sebze standı örneğinin çizelge tipi gösteriminde (Tablo 7) veriler basit bir Sepet Analizi örneğini yansıtacak şekilde 0-1 ikili şekilde idi. Ancak, birliktelik kuralları 0-1 ikili değişken türleri ile sınırlı değildir. Özellikle Apriori Algoritması kategorik veriler üzerinde uygulanabilmektedir (Larose D.T. ve Larose C.D, 2014: 255).

3.3. BİRLİKTELİK KURALLARININ MATEMATİKSEL GÖSTERİMİ

Birliktelik kurallarının matematiksel gösterimi 1993 yılında Agrawal, Imielinski ve Swami tarafından ortaya koyulmuştur (Srikant ve diğerleri, 1997: 67).

$I = \{I_1, I_2, \dots, I_m\}$ şeklinde ikili (binary) nitelikli öğeler kümesi olsun.

$D = \{t_1, t_2, \dots, t_m\}$ veri tabanındaki işlemlerin kümesi olsun. Her t_k 'nın alacağı değer 0 veya 1 dir. $t_k = 0$ olması, öğenin yokluğunu, $t_k = 1$ olması öğenin varlığını temsil etmektedir. Her işlem veri tabanında TID adı verilen bir kimlik tanımlayıcı ile ilişkilendirilir. X, I' da bir öğe kümesi olsun. T işlemi X öğe kümesini ancak ve ancak $X \subseteq T$ olduğunda içerir. X ve Y öğelerden oluşan ve 'öge kümeleri' olarak

adlandırılan kümeler olsun. X ve Y kümeleri, $X \subset I$, $B \subset I$ ve $X \cap Y \neq \emptyset$ olacak şekilde, $X \Rightarrow Y$ bir birliktelik kuralını gösterir (Agrawal ve diğerleri, 1996: 308).

3.3.1. K- Öğe Küme Ve Sık Geçen Öğe Küme

Eğer bir küme k tane öğe içeriyorsa küme 'k-öğe küme' olarak ifade edilir. K - öğe kümenin sık geçen kümeleri L_k şeklinde ifade edilir. Örnek olarak; $\{a,b,c,d\}$ 4 elemanlı bir öğe kümedir ve 4-öğe küme olarak ifade edilir (Han ve Kamber, 2006: 230).

3.3.2. Destek (Support) ve Güven (Confidence) Değeri

Birliktelik kuralları analizlerinde öğeler arasındaki ilişkileri ortaya koymak amacıyla 'destek' ve 'güven' ölçütlerinden yararlanılmaktadır. 'Kural destek ölçütü' bir ilişkinin tüm kayıtlar içinde hangi oranda tekrarlandığını, 'kural güven ölçütü' ise A öğesinin bulunduğu işlemlerde B öğesinin de bulunma durumunu gösterir. $A \Rightarrow B$ birliktelik kuralı için kural destek ölçütü 2. eşitlikte, kural güven ölçütü ise 3. eşitlikte aşağıda gösterilmiştir (Erpolat, 2012: 139):

$$\text{destek}(A) = P(A) = \frac{\text{sayı}(A)}{N} \quad (1)$$

$$\text{destek}(A \Rightarrow B) = P(A \cap B) = \frac{\text{sayı}(A,B)}{N} \quad (2)$$

$$\text{güven}(A \Rightarrow B) = P(B \setminus A) = \frac{\text{destek}(A,B)}{N} = \frac{\text{sayı}(A,B)}{\text{sayı}(A)} \quad (3)$$

Han ve Kamber' in örneğine göre, bilgisayar satın alan bir müşterinin aynı zamanda antivirüs programı satın alma eğilimi için oluşturulan bir birliktelik kuralı 4 numaralı eşitlikte şu şekilde verilmiştir:

$$\text{bilgisayar} \Rightarrow \text{antivirüs_programı} [\text{destek} = \%2, \text{güven} = \%60] \quad (4)$$

Dördüncü eşitlikteki kural, destek ve güven değerlerine göre şu şekilde yorumlanabilir:

- Destek değerinin %2 olması, bütün işlemlere bakıldığında bilgisayar ve antivirüs programının birlikte satın alındığı kayıtların oranını göstermektedir.
- Güven değerinin %60 olması, bilgisayar alan müşterilerin %60' ının aynı zamanda antivirüs programı da satın aldığı anlamına gelmektedir.

Birliktelik kurallarının destek değeri kuralın veriler için ne sıklıkta uygun olduğunu gösterir. Yüksek destek değeri ögeler arasındaki güçlü korelasyon anlamına gelmektedir. Güven değeri ise birliktelik kurallarının güvenilirliğinin ölçüsüdür (Liu ve Shih, 2005: 389).

Bir kuralın ilgi çekici olması için araştırmacı tarafından belirlenen minimum destek ve minimum güven ölçütlerini sağlaması gerekmektedir. Eşik değerler araştırmacı tarafından araştırma sırasında belirlenir. Yüksek güvenilirlik ve destek değerine sahip kurallar güçlü kurallar olarak adlandırılırlar (Han ve Kamber, 2006: 230).

3.4. BİRLİKTELİK KURALLARININ AŞAMALARI

Birliktelik kuralları genel olarak iki adımda incelenmektedir (Agrawal ve diğerleri, 1993: 208).

1. *Tüm sık geçen öge kümeleri bulunur.* Bir öge kümesinin sık geçenler kümesinde yer alabilmesi için, o ögenin destek değerinin önceden araştırmacı tarafından belirlenmiş olan minimum destek değerinden büyük olması gerekir.

2. *Sık geçen öge kümeleri kullanılarak güçlü ilişki kuralları oluşturulur.* Bir kuralın güçlü olması için yine araştırmacının belirlediği minimum güven değerini sağlaması gerekmektedir. Yani sonuç olarak kurallar minimum destek ve minimum güven durumunu sağlamalıdır.

3.5. BİRLİKTELİK KURALLARI NOTASYONLARI

Tablo 8: Birliktelik Kuralları Notasyonları

k- sık geçen öge kümesi	k- öge içeren bir öge kümesi
L_k	k- sık geçen öge kümeleri
C_k	köğeli aday kümeler
\hat{C}_k	k- öğeli ve TID' leri içeren aday kümeler

Kaynak: Agrawal ve Srikant, 1994: 489.

3.6. BİRLİKTELİK KURALLARI ALGORİTMALARI

Birliktelik kuralı algoritmaları temelde ardışık algoritmalar ve paralel algoritmalar olmak üzere iki başlık altında incelenmektedirler. Ardışık algoritmalar öge kümelerinin oluşturulduğu ve sayıldığı mantıksal ifadeleri içermekte, paralel algoritmalar ise büyük öge kümelerinin paralellik sağlayarak oluşturulmasını sağlamaktadır (Erpolat, 2012: 138).

Tablo 9: Birliktelik Kuralları Algoritmaları

Ardışık Algoritmalar	Paralel ve Dağıtılmış Algoritmalar
AIS	CD (Sayım Dağılımı)
SETM	PDM (Paralel VM)
Apriori	CCPD (Ortak Aday Bölünmüş Veri Tabanı)
Apriori-TID	DD (Veri Dağılımı)
Apriori-Hybrid	IDD (Akıllı Veri Dağılımı)
OCD (Sıradışı Aday Belirleme)	HPA (Bağlantı Kurallarının Çırpı Temelli Paralel Madenciliği)
Bölümleme Tekniği	PAR (Paralel Bağlantı Kuralları)
Örnekleme Tekniği	DMA (Dağıtılmış Madencilik Algoritması)
CARMA (Sürekli Bağlantı Kuralı Madenciliği)	Candidate Distribution (Aday Dağılımı)
DIC (Dinamik Nesne Sayımı)	SH (Çarpık Taşıma)
FP-Growth	HD (Hibrid Dağılımı)

Kaynak: Erpolat, 2012: 139.

3.6.1. Apriori Algoritması

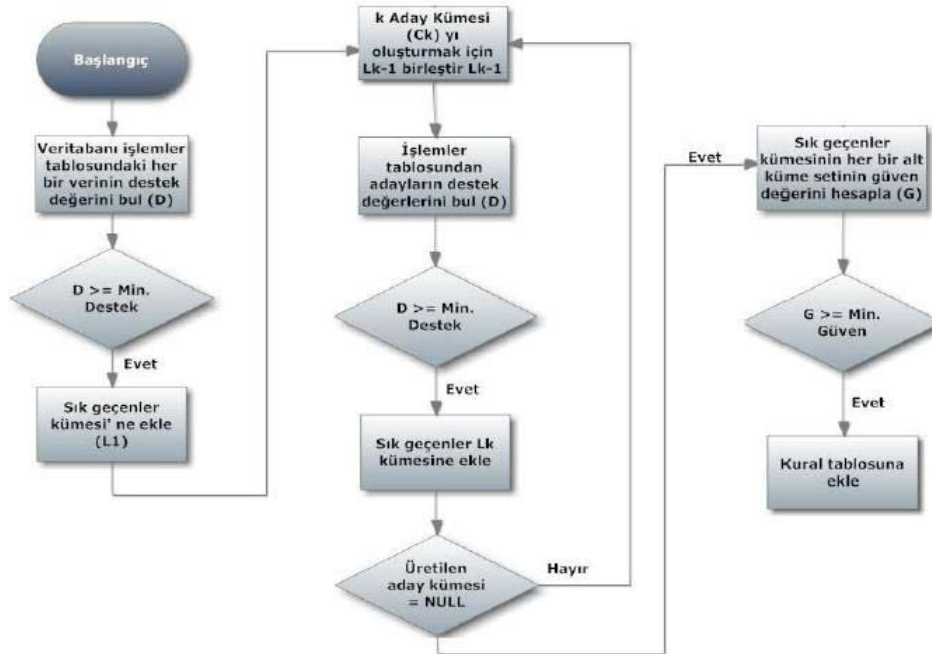
Apriori Algoritması Agrawal ve Srikant tarafından 1994 yılında önerilmiştir. Algoritmanın ismi, bilgileri bir önceki adımdan aldığı için 'prior' anlamına gelen Apriori'dir. Apriori (k +1) - öge kümelerini keşfetmek için k - öge kümelerini kullanılan iteratif bir niteliğe sahiptir (Han ve Kamber, 2006: 234-235).

Her sık geçen öge kümesi bulunmasında veri tabanının tam bir taranması gerekmektedir.

İlk taramada minimum destek ölçütünü sağlayan 1-sık geçen öge kümeleri (1 elemanlı sık geçen öge kümeleri) bulunmaktadır. Sonraki taramalarda bir önceki taramada bulunan sık geçen öge kümeleri aday kümeler adı verilen yeni potansiyel sık geçen öge kümelerini üretmek için kullanılmaktadır.

Aday kümelerin destek değerleri, tarama esnasında hesaplanmaktadır. Aday kümelerden minimum destek ölçütü sağlayan kümeler o geçişte üretilen sık geçen öge kümeleri olmaktadır. Bu süreç yeni bir sık geçen öge kümesi bulunamayana kadar devam etmektedir (Han ve Kamber, 2006: 235).

Şekil 10: Apriori Algoritmasını Gösteren Akış Diyagramı



Kaynak: Gülce, 2010:34.

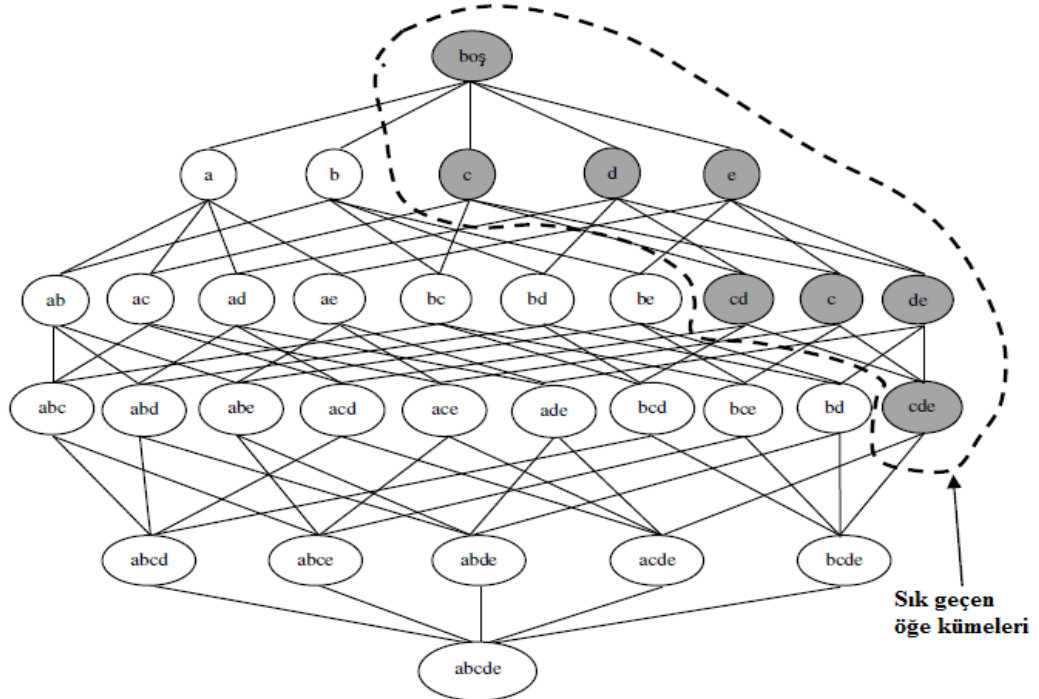
Apriori algoritmasında temel yaklaşım, eğer k-öge kümesi minimum destek ölçütünü sağlıyorsa, bu kümenin alt kümeleri de minimum destek ölçütünü sağlamaktadır. Yani, eğer bir öge kümesi sık geçen ise bu kümenin bütün alt kümeleri de sık geçen öge kümesi olmalıdır (Han ve Kamber, 2006: 235).

Apriori Algoritması Özelliği (Birleştirme ve Budama Özellikleri)

Apriori algoritmasının birleştirme ve budama özellikleri aşağıdaki şekillerle daha anlaşılır şekilde görülmektedir (Ay, 2009: 43-44).

Şekil 11’ de {c, d, e} kümesinin bir sık geçen küme olduğu varsayılır. Bu durumda {c, d, e} kümesini içeren her işlemin bu kümenin alt kümeleri olan {c, d}, {c, e}, {c}, {d} ve {e} kümelerini de içerdiği bilinmektedir. Sonuç olarak da, {c, d, e} sık geçen küme ise bütün alt kümeleri de sık geçen küme olmalıdır.

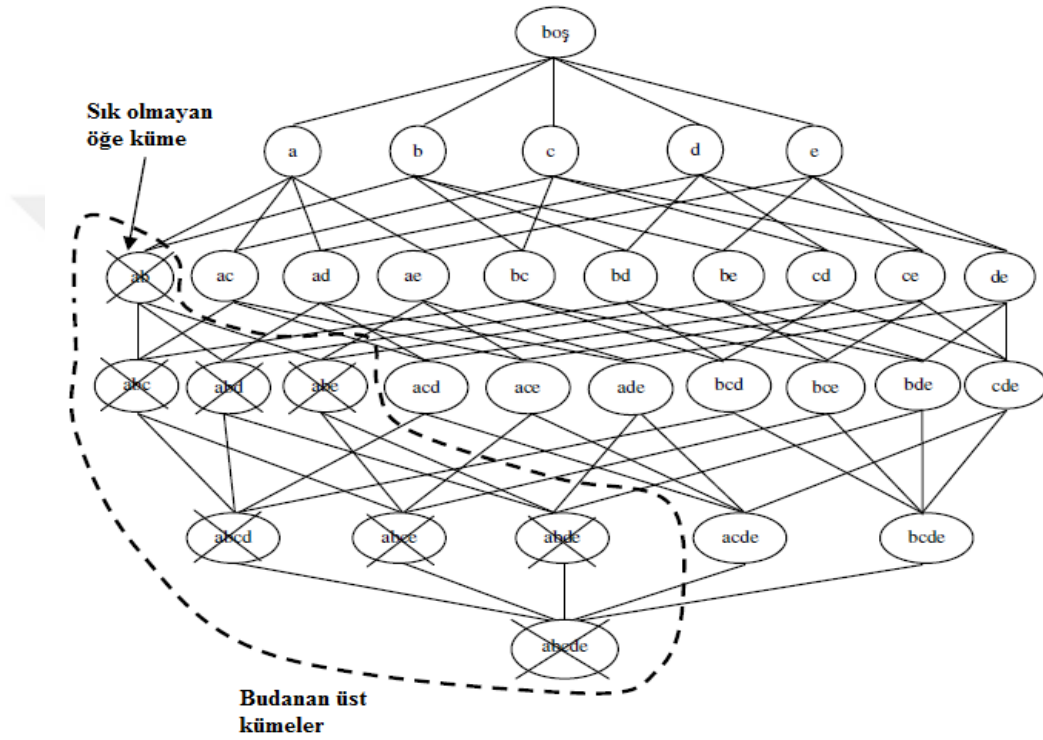
Şekil 11: Apriori Algoritmasının Birleştirme Özelliği



Kaynak: Ay, 2009: 43.

Bunun tersi olarak da, sık olmayan bir kümenin üst kümeleri de sık olmayan kümeler olarak alınır. Şekil 12’ de gösterildiği gibi {a, b} kümesinin sık olmayan küme olarak belirlendiği durumda {a, b} kümesini içeren bütün üst kümeler budanabilir.

Şekil 12: Apriori Algoritmasının Budama Özelliği



Kaynak: Ay, 2009: 44.

Şekil 13: Apriori Algoritması Kodu

```
(1) L1 = { Sık geçen 1- öge kümeleri };
(2) for (k= 2; Lk-1 ≠ ∅; k++) do begin
(3)   Ck = apriori-gen (Lk-1); // Yeni adaylar
(4)   forall işlemler t ∈ D do begin
(5)     Ct = altküme (Ck, t); // t' yi kapsayan adaylar
(6)     forall adaylar c ∈ Ct do
(7)       c.count++;
(8)     end
(9)   Lk = {c ∈ Ck | c.count ≥ min_destek}
(10) end
(11) Yanıt = Uk Lk;
```

Kaynak: Agrawal ve Srikant, 1994: 489.

Apriori algoritmasının işleyişini Han ve Kamber' in örneği ile aşağıdaki örnekle birlikte incelenmiştir. Aşağıdaki tabloda D veri tabanına ait işlemler verilmiştir. Veri tabanında 9 işlem bulunmaktadır (|D|=9) (Han ve Kamber, 2006: 236)

Tablo 10: Apriori Algoritması İşlemsel Veriler

TID	Öğeler
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Kaynak: Han ve Kamber, 2006: 236.

Adım 1: Minimum destek sayısı ve minimum güven değerinin belirlenmesi; *Minimum destek sayısı 2, minimum güven değeri ise %70* olarak seçilmiştir.

Adım 2: Öğe kümeleri içerisindeki her bir ögenin destek değerinin bulunması (C_1 kümesi)

Tablo 11: Apriori Algoritması C_1 Kümesi

Öğe kümesi	Destek değeri
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2

Kaynak: Han ve Kamber, 2006: 237.

Adım 3: 1-öğeli aday kümeler ile sık geçen L_1 oluşturulur. (C_1 ' de minimum destek değerini sağlayan öğeler L_1 i oluşturur). Bu örneğe göre C_1 kümesindeki tüm adaylar destek değerini karşılamaktadırlar. Böylece L_1 şu şekilde elde edilir.

Tablo 12: Apriori Algoritması L_1 Kümesi

Öğe kümesi	Destek değeri
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2

Kaynak: Han ve Kamber, 2006: 237.

Adım 4: 2-öğeli sık geçen öğe kümeleri (L_2) algoritma $L_1 \times L_1$ birleşimini kullanır ve 2-öğeli öğe kümeleri (C_2) oluşturur. C_2 , L_1 'in ikili kombinasyonlarından oluşur. Adayların her alt kümesinin de sık geçen olduğu için hiçbir adayın budama

adımı esnasında C_2 'den kaldırılmamaktadır. D veri tabanındaki işlemler taranarak C_2 kümesindeki öğelerin destekdeğerleri belirlenir.

Tablo 13: Apriori Algoritması C_2 Kümesi

Öğe kümesi	Destek değeri
{I1, I2}	4
{I1, I3}	4
{I1, I4}	1
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I2, I5}	2
{I3, I4}	0
{I3, I5}	1
{I4, I5}	0

Kaynak: Han ve Kamber, 2006: 237.

Adım 5: C_2 kümesindeki minimum destek değerine sahip olan 2 ögeli öğe kümeler L_2 'yi oluşturur.

Tablo 14: Apriori Algoritması L_2 Kümesi

Öğe kümesi	Destek değeri
{I1, I2}	4
{I1, I3}	4
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I2, I5}	2

Kaynak: Han ve Kamber, 2006: 237.

Adım 6: 3 ögeli adaylar kümesini (C_3) oluşturmak için $L_2 \times L_2$ birleşimi kullanılır. Birleştirme adımından, 3 ögeli adaylar kümesi $C_3 = L_2 \times L_2 = \{\{I1, I2,$

I_3 }, $\{I_1, I_2, I_5\}$, $\{I_1, I_3, I_5\}$, $\{I_2, I_3, I_4\}$, $\{I_2, I_3, I_5\}$, $\{I_2, I_4, I_5\}$ şeklinde elde edilir.

D' deki işlemler C_3 'ü oluşturmak için taranır.

a. *Birleştirme*: $C_3 = L_2 \circ L_2 = \{\{I_1, I_2\}, \{I_1, I_3\}, \{I_1, I_5\}, \{I_2, I_3\}, \{I_2, I_4\}, \{I_2, I_5\}\} \circ \{\{I_1, I_2\}, \{I_1, I_3\}, \{I_1, I_5\}, \{I_2, I_3\}, \{I_2, I_4\}, \{I_2, I_5\}\}$
 $= \{I_1, I_2, I_3\}, \{I_1, I_2, I_5\}, \{I_1, I_3, I_5\}, \{I_2, I_3, I_4\}, \{I_2, I_3, I_5\}, \{I_2, I_4, I_5\}$.

b. *Budama, Apriori özelliğini kullanarak*: Adayların herhangi sık olmayan bir alt kümesini var mı?

– $\{I_1, I_2, I_3\}$ kümesinin 2- öğeli alt kümeleri $\{\{I_1, I_2\}, \{I_1, I_3\}$ ve $\{I_2, I_3\}$ kümeleridir. $\{I_1, I_2, I_3\}$ kümesinin bütün 2- öğeli alt kümeleri L_2 ' nin üyesidir. Bu yüzden, $\{I_1, I_2, I_3\}$ kümesi C_3 ' te saklanır.

– $\{I_1, I_2, I_5\}$ kümesinin 2- öğeli alt kümeleri $\{\{I_1, I_2\}, \{I_1, I_5\}$ ve $\{I_2, I_5\}$ kümeleridir. $\{I_1, I_2, I_5\}$ kümesinin bütün 2- öğeli alt kümeleri L_2 ' nin üyesidir. Bu yüzden, $\{I_1, I_2, I_5\}$ kümesi C_3 ' te saklanır.

– $\{I_1, I_3, I_5\}$ kümesinin 2- öğeli alt kümeleri $\{\{I_1, I_3\}, \{I_1, I_5\}$ ve $\{I_3, I_5\}$ kümeleridir. $\{I_3, I_5\}$ kümesi L_2 ' nin üyesi değildir ve bu yüzden sık geçen öge kümesi değildir. Bu yüzden, $\{I_1, I_3, I_5\}$ kümesi C_3 ' ten çıkartılır.

– $\{I_2, I_3, I_4\}$ kümesinin 2- öğeli alt kümeleri $\{\{I_2, I_3\}, \{I_2, I_4\}$ ve $\{I_3, I_4\}$ kümeleridir. $\{I_3, I_4\}$ kümesi L_2 ' nin üyesi değildir ve bu yüzden sık geçen öge kümesi değildir. Bu yüzden, $\{I_2, I_3, I_4\}$ kümesi C_3 ' ten çıkartılır.

– $\{I_2, I_3, I_5\}$ kümesinin 2- öğeli alt kümeleri $\{\{I_2, I_3\}, \{I_2, I_5\}$ ve $\{I_3, I_5\}$ kümeleridir. $\{I_3, I_5\}$ kümesi L_2 ' nin üyesi değildir ve bu yüzden sık geçen öge kümesi değildir. Bu yüzden, $\{I_2, I_3, I_5\}$ kümesi C_3 ' ten çıkartılır.

– $\{I_2, I_4, I_5\}$ kümesinin 2- öğeli alt kümeleri $\{\{I_2, I_4\}, \{I_2, I_5\}$ ve $\{I_4, I_5\}$ kümeleridir. $\{I_4, I_5\}$ kümesi L_2 ' nin üyesi değildir ve bu yüzden sık geçen öge kümesi değildir. Bu yüzden, $\{I_2, I_4, I_5\}$ kümesi C_3 ' ten çıkartılır.

Bu durumda, budamadan sonra $C_3 = \{\{I_1, I_2, I_3\}, \{I_1, I_2, I_5\}\}$ olarak elde edilir.

Tablo 15: Apriori Algoritması C₃Kümesi

Öğe kümesi	Destek değeri
{I1, I2, I3}	4
{I1, I2, I5}	4

Kaynak: Han ve Kamber, 2006: 237.

Adım 7: C₃ kümesindeki minimum destek değerine sahip olan 3 ögeli öge kümeler L₃' ü oluşturur.

Tablo 16: Apriori Algoritması L₃Kümesi

Öğe kümesi	Destek değeri
{I1, I2, I3}	4
{I1, I2, I5}	4

Kaynak: Han ve Kamber, 2006: 237.

Adım 8: Algoritma L₃ ∞ L₃ birleşimini kullanarak 4- öge kümeli C₄ adaylar kümesini oluşturur. {I2, I3, I5} sık geçen olmadığı için, {I1, I2, I3, I5} birleştirme sonuçlarına rağmen budanır. Bunun için, C₄=∅ olmaktadır ve algoritma sonlandırır.

Sık geçen öge kümelerinden birliktelik kurallarının oluşturulması:

Verinin sık geçen öge kümesi I={I1, I2, I5} kümesini içerdiğini varsayalım. I' dan elde edilebilecek birliktelik kurallarını bulalım. I kümesinin boş olmayan alt kümeleri {I1, I2}, {I1, I5}, {I2, I5}, {I1}, { I2} ve {I5} kümeleridir. Birliktelik kuralları sonuçları ve her birine ait güven değeri aşağıda gösterilmektedir:

I1^ I2 ⇒ I5,	<i>güven</i> = 2/ 4= %50
I1^ I5 ⇒ I2,	<i>güven</i> = 2/ 2= %100
I2^ I5 ⇒ I1,	<i>güven</i> = 2/ 2= %100
I1⇒ I2^ I5,	<i>güven</i> = 2/ 6= %33
I2 ⇒ I1^ I5,	<i>güven</i> = 2/ 7= %29

$$I5 \Rightarrow I1 \wedge I2, \quad \text{güven} = 2/2 = \%100$$

Minimum güven eşik değeri %70 olarak belirlenirse sadece ikinci, üçüncü ve altıncı kurallar güçlü birliktelik kuralları olarak belirlenir.

Minimum güven değerini sağlayan kuralların yorumlanması:

Kural 2: I1 ve I5 in olduğu kümede I2' nin olma olasılığı %100 dür.

Kural 3: I2 ve I5 in olduğu kümede I1' nin olma olasılığı %100 dür.

Kural 6: I5' in olduğu kümede I1 ve I2' nin olma olasılığı %100 dür.

3.6.2. AprioriTid Algoritması

AprioriTid algoritması geçiş işlemleri başlamadan önce aday öge kümeleri tanımlamak için Şekil 14 'te verilen apriori-gen fonksiyonunu kullanır (Agrawal ve diğerleri, 1996: 315).

Şekil 14: AprioriTid Algoritması Kodu

```

(1)   L1 = {sık geçen 1-öge kümeler}
(2)   Ĉ1 = D veritabanı;
(3)   for (k=2; Lk-1 ≠ ∅; k++) do begin
(4)     Ck = apriori-gen (Lk-1); // Yeni adaylar
(5)     Ĉk = ∅;
(6)     forall kayıtlar t ∈ Ĉk-1 do begin
(7)       // TID numaralı t işlemindeki bütün aday kümeleri belirle
         Ct = {c ∈ Ck | (c[1] . c[2] . ... . c[k-1]) ∈ t.öge kümeleri ∧
              (c[1] . c[2] . ... . c[k-2] . c[k]) ∈ t.öge kümeleri };
(8)     forall adaylar c ∈ Ct do
(9)       c.count++;
(10)    if (Ct ≠ ∅) then Ĉk += <t.TID, Ct>;
(11)    end
(12)    Lk = {c ∈ Ck | c.count ≥ min_destek}
(13)    end
(14)    Yanıt = Uk Lk;

```

Kaynak: Agrawal ve diğerleri, 1996: 314.

AprioriTid algoritmasında, veri tabanı ilk geçişten sonra destek değeri hesaplamaları için kullanılmaz. Bunun yerine, ilk geçişten sonra veri tabanı yerine \hat{C}_1 kümesi kullanılarak geniş öge kümeler bulunur.

AprioriTid algoritması Agrawal ve diğerleri tarafından şekildeki veri tabanı dikkate alınarak aşağıdaki gibi açıklanmıştır (Agrawal ve diğerleri, 1996: 314).

Tablo 17: AprioriTid Algoritması İşlemsel Verileri

TID	Öğeler
T100	I1, I3, I4
T200	I2, I3, I5
T300	I1, I2, I3, I5
T400	I2, I5

Kaynak: Agrawal ve diğerleri, 1996: 316.

Örnekte minimum destek değeri 2 olarak belirlenmiştir. İlk olarak veri tabanı taranarak \hat{C}_1 öge kümeleri belirlenir.

Tablo 18: AprioriTid Algoritması \hat{C}_1 Kümesi

TID	Öge kümelerinin kümesi
T100	{ {I1}, {I3}, {I4} }
T200	{ {I2}, {I3}, {I5} }
T300	{ {I1}, {I2}, {I3}, {I5} }
T400	{ {I2}, {I5} }

Kaynak: Agrawal ve diğerleri, 1996: 316.

\hat{C}_1 ' deki kayıtlara göre L_1 hesaplanır.

Tablo 19: AprioriTid Algoritması L_1 Kümesi

TID	Destek
T100	2
T200	3
T300	3
T400	3

Kaynak: Agrawal ve diğerleri, 1996: 316.

Apriori-gen fonksiyonu kullanılarak L_1 ' den C_2 aday kümesi oluşturulur.

Tablo 20: AprioriTid Algoritması C_2 Kümesi

TID	Destek
{I1 I2}	1
{I1 I3}	2
{I1 I5}	1
{I2 I3}	2
{I2 I5}	3
{I3 I5}	2

Kaynak: Agrawal ve diğerleri, 1996: 316.

C_2 ' deki öge kümelerin destek değerleri hesaplamak amacıyla tüm veri tabanı yerine \hat{C}_1 taranır ve \hat{C}_2 oluşturulur.

Tablo 21: AprioriTid Algoritması \hat{C}_2 Kümesi

TID	Öge kümelerin kümesi
T100	{ {I1 I3} }
T200	{ {I2 I3}, {I2 I5}, {I3 I5} }
T300	{ {I1 I2}, {I1 I3}, {I1 I5}, {I2 I3}, {I2 I5}, {I3 I5} }
T400	{ {I2 I5} }

Kaynak: Agrawal ve diğerleri, 1996: 316.

Daha sonra \hat{C}_2 ' deki destek deęerleri sayılarak da L_2 elde edilir.

Tablo 22: AprioriTid Algoritması L_2 Kümesi

TID	Destek
{I1 I3}	2
{I2 I3}	2
{I2 I5}	3
{I3 I5}	2

Kaynak: Agrawal ve dięerleri, 1996: 316.

Yine apriori-gen fonksiyonu kullanılarak L_2 ' den C_3 oluřturulur.

Tablo 23: AprioriTid Algoritması C_3 Kümesi

TID	Destek
{ I2 I3 I5}	2

Kaynak: Agrawal ve dięerleri, 1996: 316.

Verinin üzerinden \hat{C}_2 ile geçilir ve C_3 ' ten \hat{C}_3 oluřturulur.

Tablo 24: AprioriTid Algoritması \hat{C}_3 Kümesi

TID	Öęe kümelerin kümesi
T200	{ {I2 I3 I5} }
T300	{ {I2 I3 I5} }

Kaynak: Agrawal ve dięerleri, 1996: 316.

\hat{C}_3 kümesinde 100 ve 400. işlemlerine ait giriş yapılmadığına dikkat ediniz. Bunun sebebi 100. ve 400. işlemlerdeki öęelerin C_3 aday kümesindeki öęe kümelerde yer almamasıdır. C_3 kümesindeki tek aday {2 3 5} öęe kümesi minimum destek deęerini saęlayarak L_3 kümesini oluřturur.

Tablo 25: AprioriTid Algoritması L₃Kümesi

TID	Destek
{I2 I3 I5}	2

Kaynak: Agrawal ve diğerleri, 1996: 316.

C₄ kümesi boş küme olacağından algoritma burada sonlandırılır (Agrawal ve diğerleri, 1996: 315).

3.6.3. AprioriHybrid Algoritması

Veri üzerindeki tüm geçişlerde aynı algoritmanın kullanılması zorunluluğu yoktur. Agrawal ve Srikant yaptıkları çalışmada kullandıkları veri tabanında Apriori ve AprioriTid algoritmalarının veri tabanında farklı geçişlerdeki işlem sürelerini elde etmişlerdir. İlk geçişlerde Apriori algoritmasının daha iyi olduğu gözlemlenmekle birlikte, son geçişlerde AprioriTid algoritmasının daha iyi performans gösterdiği görülmüştür.

Apriori ve AprioriTid algoritmalarının ikisi de aynı aday oluşturma prosedürünü kullanır ve aynı öge kümeleri sayar. Sonraki geçişlerde aday öge kümelerin sayısının azalmasına rağmen Apriori veri tabanındaki bütün işlemleri kontrol etmeye devam eder. Diğer taraftan, AprioriTid destek sayılarını elde etmek için veri tabanından daha küçük boyuttaki \hat{C}_k ' yi tarar. Bu incelemeler üzerine AprioriHybrid algoritması ortaya koyulmuştur.

AprioriHybrid algoritmasında ilk geçişlerde Apriori algoritmasını kullanılırken \hat{C}_k kümesinin belleğe uygun olabileceğinin anlaşılması ile AprioriTid' e geçilir (Agrawal ve Srikant, 1994: 496).

3.6.4. AIS Algoritması

AIS algoritması Agrawal, Imielinski ve Swami tarafından 1993 yılında geliştirilmiş bir algoritmadır. Algoritma geniş öge kümeleri üretmektedir. AIS algoritması aday öge kümeleri veri tabanı taranırken sayarak oluşturur. Her geçişte

bütün veri tabanı taranır. Bir işlem okunduktan sonra önceki geçişte sık bulunan öge kümeler genişletilir ve aday öge kümeler oluşturulur. Geniş bir öge kümesi l sadece geniş öge kümelerle genişletilir. AIS algoritması, l kümesindeki her bir ögenin alfabetik sıraya göre dizilmesi kısıtını taşımaktadır (Agrawal ve Srikant, 1994: 492).

3.6.5. SETM Algoritması

SETM algoritması geniş öge kümeleri bulmak için SQL kullanan bir algoritmadır. AIS algoritması gibi SETM algoritması da aday öge kümeleri veri tabanı taranırken sayarak oluşturmakta, ancak bu işlemi uygularken SQL kullanmaktadır (Agrawal ve Srikant, 1994: 492).

3.6.6. FP-Growth

Birliktelik kuralları analizini gerçekleştirmek üzere geliştirilmiş tekniklerden bir tanesi de FP-Growth (Frequent Pattern Growth) algoritmasıdır. Algoritma böl ve yönet stratejisini benimser. İlk olarak, sık geçen öge kümelerini bir FP- ağacına sıkıştırır. Daha sonra, şartlı veri tabanlarını oluşturarak her bir veri tabanını ayrı ayrı madenler (Han ve Kamber, 2006: 243).

3.6.7. Bölümleme algoritması

Bölümleme algoritması, veri tabanını küçük parçalara bölerek işgal edilen bellek alanını azaltır ve böylece daha hızlı sonuca varılmasını sağlar.

İlk iki geçişte veri tabanı üzerindeki bütün sık geçen kümeleri hesaplar. İlk veri tabanı geçişinde yerel sık geçen öge kümeleri belirlenir. İkinci geçişte ise yerel sık geçen öge kümelerinin birleşimi aday küme olarak kullanılır. İlk geçiş sık geçen öge kümelerinin bir üst kümesinin bulunmasını garanti etmektedir. İkinci geçiş sadece kümelerin frekanslarını hesaplamak için gereklidir (Toivonen, 1996: 136).

DÖRDÜNCÜ BÖLÜM

UYGULAMA

4.1. UYGULAMA GİRİŞ

Yapılan çalışmanın amacı, üniversite öğrencilerinin seçimlik derslerle ilgili karar verirken hangi kriterleri göz önünde bulundurdıkları, hangi derslerin birlikte seçilme eğiliminde olduğu, hangi özelliklerdeki öğrencilerin hangi dersleri seçme eğiliminde oldukları bilgisini elde etmek ve elde edilen bilgiler doğrultusunda öneriler sunmaktır. Bu amaçla oluşturulan seçimlik ders anketi Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü lisans öğrencilerine uygulanmıştır (EK 1' de verilmiştir). Anket, seçimlik derslerin başladığı dönem dikkate alınarak 2. 3. ve 4. sınıf öğrencilerine web üzerinden ulaştırılmış ve uygulanmıştır. Uygulamada 151 öğrencinin sonuçları değerlendirmeye alınmıştır.

Çalışmada IBM SPSS Statistics 20.0 ve SPSS Clementine 10.1 programları kullanılmıştır.

4.2. VERİ SETLERİNİN OLUŞTURULMASI

Elde edilen veriler düzenlenerek 2 ayrı veri seti oluşturulmuş ve üzerinde çalışılmıştır. İlk veri seti genel sorular ve derslerin seçim durumundan oluşmaktadır. Bu amaçla her ders ayrı bir boolean değişken olarak ele alınmıştır. Yani bir ders değerlendirmeye alınan öğrenci tarafından seçilmiş ise 1, seçilmemiş ise 0 değerini almaktadır. Birinci veri setinde değerlendirmeye katılan 151 öğrenci olduğu için 151 kayıt bulunmaktadır.

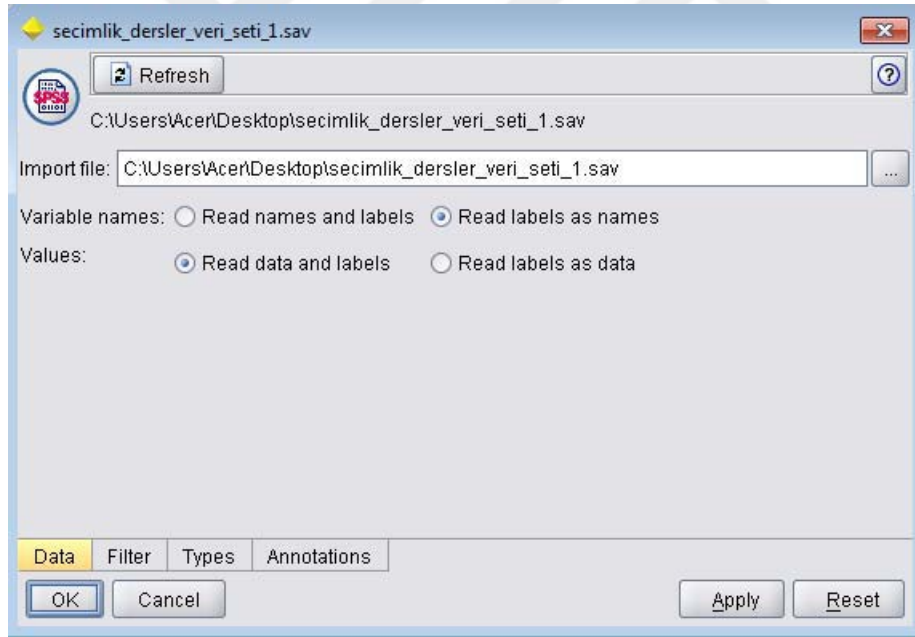
İkinci veri seti ise her dersin sebepleriyle birlikte ayrı kayıtlara bölünmesiyle elde edilmiştir. Yani, bir öğrenci toplamda altı adet seçmeli ders almışsa ve dolayısıyla altı ders için değerlendirme yapmışsa bir öğrencinin verilerinden 6 kayıt elde edilmiştir. Bu şekilde 2. veri seti için ise toplam 862 kayıt elde edilmiştir.

4.3. VERİLERİN HAZIRLANMASI

Öncelikle veri setleri SPSS Clementine programına aktarılmadan önce SPSS Statistics programında derslerle ilgili düzenlemeler yapılmıştır. Hiç seçilmeyen dersler listeden çıkartılmış, diğer seçeneği ile eklenen dersler listeye eklenmiştir. Özellikle listede yer almasına rağmen bazı derslerin farklı dönemlerde diğer seçeneği ile eklendiği gözlemlenmiştir. Bunun sebebi derslerin dönemlerinin revize edilmesidir. Bu şekilde ayrı olarak eklenmiş dersler aynı başlık altında toplanmıştır.

Veri setleri SPSS formatında oluşturulduğu için kaynak düğümlerinden ‘SPSS Düğümü’ ile veri setleri SPSS Clementine programına aktarılmıştır.

Şekil 15:Clementine’ da SPSS Düğümü İle 1. Veri Setine Sağlanan Erişim



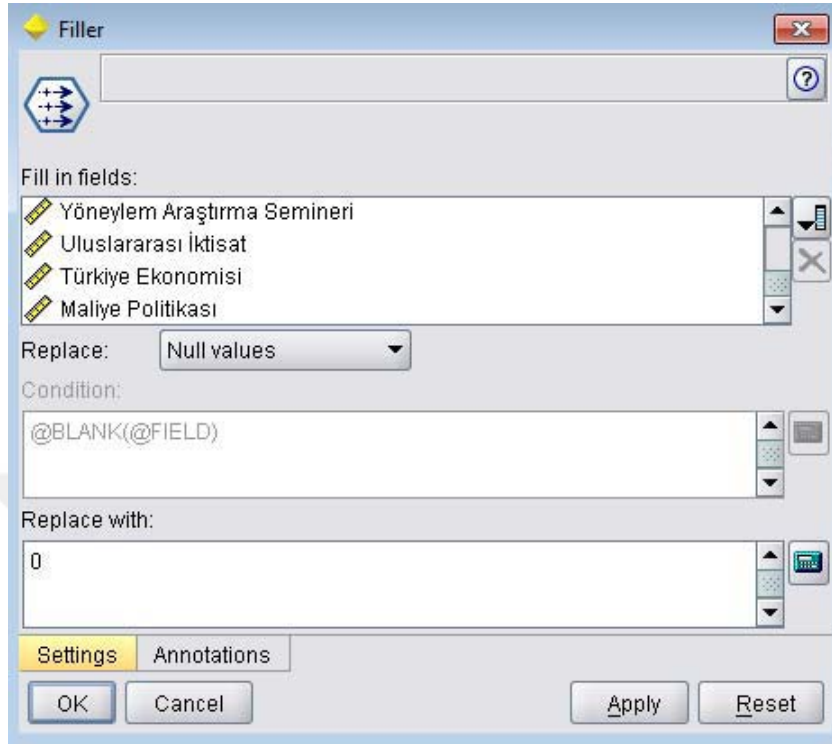
‘SPSS Düğümü’ne bir ‘Table Düğümü’ eklenerek çalıştırılır ve veri kontrolü yapılır. Şekil 16’ daki gibi veriler görüntüleniyorsa, veri kaynağına erişim sağlanmış olur.

Şekil 16: Uygulamada Kullanılan 1. Veri Setinin Clementine’da Görüntülenmesi

	Cin...	Öğr...	Dike...	Yata...	Farabi	Üniv...	Ça...	Mez...	Me...	Lise...	Tek...	Eko...	Ekon...	Ekon...	İgi al...
1	2	2	\$null\$	\$null\$	\$null\$	\$null\$	3	2	4	3	17	\$null\$	1	\$null\$	\$null\$
2	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	1	9	2	9	\$null\$	\$null\$	1	\$null\$
3	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	3	4	3	11	1	\$null\$	1	1
4	1	1	\$null\$	\$null\$	1	\$null\$	2	1	1	3	10	\$null\$	\$null\$	1	1
5	2	2	\$null\$	\$null\$	\$null\$	\$null\$	3	5	1	3	1	1	1	\$null\$	1
6	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	2	1	3	2	\$null\$	\$null\$	1	1
7	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	1	4	3	1	\$null\$	\$null\$	1	1
8	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	2	4	3	1	\$null\$	\$null\$	1	1
9	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	1	1	3	7	\$null\$	\$null\$	1	1
10	2	2	\$null\$	\$null\$	\$null\$	\$null\$	3	3	4	3	1	1	1	1	1
11	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	1	4	3	11	\$null\$	\$null\$	1	\$null\$
12	2	1	\$null\$	\$null\$	\$null\$	\$null\$	1	3	4	1	1	\$null\$	1	1	1
13	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	3	4	3	14	1	1	1	1
14	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	1	4	3	24	1	\$null\$	\$null\$	1
15	1	2	1	\$null\$	\$null\$	\$null\$	3	1	1	1	1	1	1	1	1
16	2	1	\$null\$	\$null\$	\$null\$	\$null\$	3	3	1	3	19	1	\$null\$	1	1
17	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	2	1	3	18	\$null\$	1	1	\$null\$
18	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	5	4	3	12	\$null\$	1	\$null\$	\$null\$
19	1	1	\$null\$	1	\$null\$	\$null\$	2	1	4	3	1	1	1	1	1
20	2	1	\$null\$	\$null\$	\$null\$	\$null\$	3	1	4	3	12	\$null\$	1	1	\$null\$
21	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	3	1	3	14	1	\$null\$	\$null\$	1
22	1	1	\$null\$	\$null\$	\$null\$	1	3	3	4	3	1	1	\$null\$	1	1
23	1	1	\$null\$	\$null\$	\$null\$	\$null\$	3	3	1	3	23	1	1	\$null\$	1
24	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	2	1	3	7	1	1	\$null\$	1
25	2	1	1	\$null\$	\$null\$	\$null\$	3	2	1	3	1	1	1	\$null\$	1
26	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	2	1	3	5	\$null\$	\$null\$	1	1
27	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	2	1	3	24	1	\$null\$	1	1
28	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	1	1	3	1	1	\$null\$	1	1
29	1	2	\$null\$	\$null\$	\$null\$	\$null\$	3	3	1	3	1	1	\$null\$	\$null\$	1
30	1	1	\$null\$	\$null\$	\$null\$	\$null\$	2	3	1	3	1	1	\$null\$	\$null\$	\$null\$

SPSS Clementine’ da verilerin gösteriminde Şekil 17’ de, boolean tipindeki değişkenlere ait null/boş değerler olduğu görülmektedir. Boş değerler sıfır olmalıdır. SPSS Clementine programında ‘Filler Düğümü’ ile boş değerler sıfır değeri ile doldurulur.

Şekil 17: Filler Düğümü İle Boş Değerlerin Doldurulması



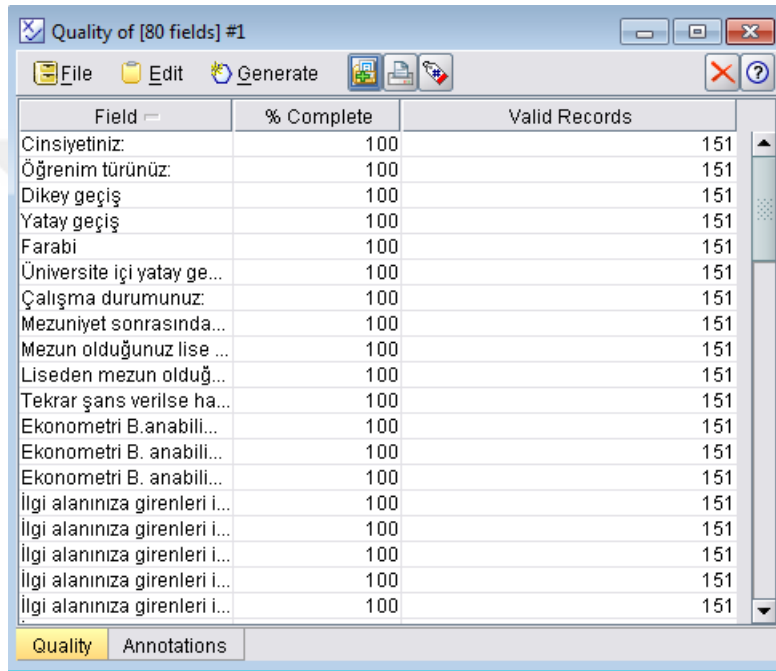
Şekil 18: Boş Değerlerin Doldurulması Sonrası 1. Veri Setinin Görüntüsü

	Cins...	Öğr...	Dikey...	Yata...	Farabi Üniv...	Çalı...	Mez...	Mezun...	Lise...	Tek...	Eko...	Eko...	Eko...	İlgil...	i...
1	2	2	0	0	0	0	3	2	4	3	17	0	1	0	0
2	1	1	0	0	0	0	3	1	9	2	9	0	0	1	0
3	1	1	0	0	0	0	3	3	4	3	11	1	0	1	1
4	1	1	0	0	1	0	2	1	1	3	10	0	0	1	1
5	2	2	0	0	0	0	3	5	1	3	1	1	1	0	1
6	1	1	0	0	0	0	3	2	1	3	2	0	0	1	1
7	1	2	0	0	0	0	3	1	4	3	1	0	0	1	1
8	1	1	0	0	0	0	3	2	4	3	1	0	0	1	1
9	1	2	0	0	0	0	3	1	1	3	7	0	0	1	1
10	2	2	0	0	0	0	3	3	4	3	1	1	1	1	1
11	1	1	0	0	0	0	3	1	4	3	11	0	0	1	0
12	2	1	0	0	0	0	1	3	4	1	1	0	1	1	1
13	1	1	0	0	0	0	3	3	4	3	14	1	1	1	1
14	1	2	0	0	0	0	3	1	4	3	24	1	0	0	1
15	1	2	1	0	0	0	3	1	1	1	1	1	1	1	1
16	2	1	0	0	0	0	3	3	1	3	19	1	0	1	1
17	1	1	0	0	0	0	3	2	1	3	18	0	1	1	0
18	1	1	0	0	0	0	3	5	4	3	12	0	1	0	0
19	1	1	0	1	0	0	2	1	4	3	1	1	1	1	1
20	2	1	0	0	0	0	3	1	4	3	12	0	1	1	0

Şekil 18’ de görüldüğü gibi boş değerler sıfır değeriyle doldurulmuştur. Aynı işlemler 2. veri seti için de gerçekleştirilir.

‘Quality Dügümü’ verinin kalitesi hakkında bilgi vermektedir. Veri setindeki boş değerlere ilişkin düzenleme yapıldıktan sonra bütün deęişkenlere ait tanımlı deęerleri görmek amacıyla ‘Quality Dügümü’ kullanılır.

Şekil 19: Veri Kalitesini Gösteren Quality Dügüm Çıktısı (1. VeriSeti)

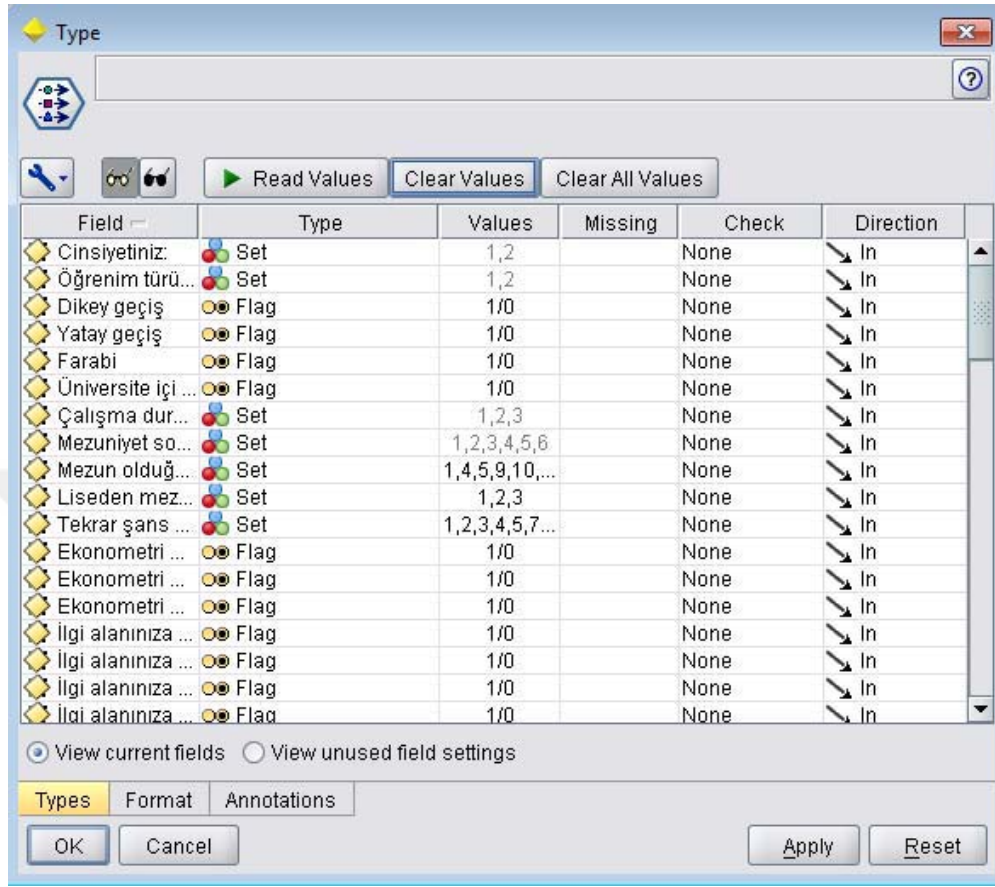


Field	% Complete	Valid Records
Cinsiyetiniz:	100	151
Öğrenim türünüz:	100	151
Dikey geçiş	100	151
Yatay geçiş	100	151
Farabi	100	151
Üniversite içi yatay ge...	100	151
Çalışma durumunuz:	100	151
Mezuniyet sonrasında...	100	151
Mezun olduğunuz lise ...	100	151
Liseden mezun olduğ...	100	151
Tekrar şans verilse ha...	100	151
Ekonometri B. anabili...	100	151
Ekonometri B. anabili...	100	151
Ekonometri B. anabili...	100	151
İlgi alanınıza girenleri i...	100	151
İlgi alanınıza girenleri i...	100	151
İlgi alanınıza girenleri i...	100	151
İlgi alanınıza girenleri i...	100	151
İlgi alanınıza girenleri i...	100	151

Şekil 19’ da verilerin bir kısmına ait bilgiyi görmekteyiz. Şekildeki deęişkenlerde hiç boş, eksik deęer yoktur yani tamamlanma oranı %100’ dür.

‘Type Dügümü’ ile verilerin tip ve yönleri belirlenir. SPSS Statistics programından farklı olarak SPSS Clementine programında ‘flag’ tanımlaması bulunmaktadır. Boolean tipindeki deęişkenler SPSS Clementine’ da ‘flag’ olarak tanımlanır.

Şekil 20: Type Düğümü Görüntüsü



Tablo 27: Genel Özelliklere İlişkin Değişkenler ve Özellikleri

Değişken	Tipi ve Özellikleri
Cinsiyet	Metin, küme (Kadın, Erkek)
Öğretim türü	Metin, küme (Örgün, İkinci Öğretim)
Dikey geçiş	Boolean, Evet Hayır (1/0)
Yatay geçiş	Boolean, Evet Hayır (1/0)
Farabi	Boolean, Evet Hayır (1/0)
Üniversite içi yatay geçiş	Boolean, Evet Hayır (1/0)
Çalışma durumunuz:	Metin, küme (...)
Mezuniyet sonrasındaki temel hedefiniz nedir?	Metin, küme (...)
Öğreniminizin kaçınıcı dönemindesiniz?	Metin, küme (3,4,5,6,7,8)
Mezun olduğunuz lise türü:	Metin, küme (...)
Liseden mezun olduğunuz bölüm:	Metin, küme (...)
Tekrar şans verilse hangi bölümde okumak istersiniz?	Metin, küme (...)

Tablo 28:Anabilim Dalı Bazında İlgi Alanlarına İlişkin Değişkenler ve Özellikleri

Değişken	Tipi ve Özellikleri
Ekonometri Anabilim Dalı	Boolean, Evet Hayır (1/0)
Yöneylem Araştırması Anabilim Dalı	Boolean, Evet Hayır (1/0)
İstatistik Anabilim Dalı	Boolean, Evet Hayır (1/0)

Tablo 29:Genel İlgi Alanlarına İlişkin Değişkenler ve Özellikleri

Değişken	Tipi ve Özellikleri
Ekonometri	Boolean, Evet Hayır (1/0)
Yöneylem Araştırması	Boolean, Evet Hayır (1/0)
İstatistik	Boolean, Evet Hayır (1/0)
Para- Banka	Boolean, Evet Hayır (1/0)
Finans	Boolean, Evet Hayır (1/0)
Pazarlama	Boolean, Evet Hayır (1/0)
Muhasebe	Boolean, Evet Hayır (1/0)
Maliye	Boolean, Evet Hayır (1/0)
Kamu Yönetimi	Boolean, Evet Hayır (1/0)
Kalite	Boolean, Evet Hayır (1/0)
Bilgisayar Bilimleri	Boolean, Evet Hayır (1/0)
İnsan Kaynakları Yönetimi	Boolean, Evet Hayır (1/0)
İktisat	Boolean, Evet Hayır (1/0)
Yönetim ve Organizasyon	Boolean, Evet Hayır (1/0)
Üretim Yönetim	Boolean, Evet Hayır (1/0)

Tablo 30:Seçimlik Ders Seçim Sebeplerine İlişkin Değişkenler ve Özellikleri

Değişken	Tipi ve Özellikleri
Seçimlik dersleri seçerken en çok neyi referans alıyorsunuz?	Metin, küme (...)
Seçimlik dersler hakkında ön bilgi verilmesini ister miydiniz?	Boolean, Evet Hayır (1/0)

Tablo 31: Likert Tipindeki Sorulara İlişkin Değişkenler ve Özellikleri

Değişken	Tipi ve Özellikleri
En kısa sürede mezun olma baskısı hissetmiyorum.	Metin, sıralı (5’li likert ölçeği)
En kısa sürede mezun olmaktansa, donanımlı mezun olmayı tercih ederim.	Metin, sıralı (5’li likert ölçeği)
Ekonometri bölümünü isteyerek, bilerek, araştırarak geldim.	Metin, sıralı (5’li likert ölçeği)
Okuduğum bölümü benimsedim ve seviyorum.	Metin, sıralı (5’li likert ölçeği)
Okuduğum bölüme aidiyet hissediyorum.	Metin, sıralı (5’li likert ölçeği)
Dokuz Eylül Üniversitesi' ne isteyerek geldim.	Metin, sıralı (5’li likert ölçeği)
Araştırma yapmayı severim.	Metin, sıralı (5’li likert ölçeği)
Bilgisayar kullanımına erişimim kolaydır.	Metin, sıralı (5’li likert ölçeği)
Yabancı dil düzeyimin yeterli olduğunu düşünüyorum.	Metin, sıralı (5’li likert ölçeği)
Benim için yeni bir şey öğrenmekten çok dersi geçmek daha önemlidir.	Metin, sıralı (5’li likert ölçeği)

Tablo 32: Ders ve Seçim Sebeplerini İçeren Değişkenler ve Özellikleri

Değişken	Tipi ve Özellikleri
Alınan dersin adı:	Metin, küme (...)
Kolay olması	Boolean, Evet Hayır (1/0)
Öğretim üyesi	Boolean, Evet Hayır (1/0)
Bilgisayar uygulamalı olması	Boolean, Evet Hayır (1/0)
İlgi alanıma yakın olması	Boolean, Evet Hayır (1/0)
Önemli kazanım sunuyor olması	Boolean, Evet Hayır (1/0)
Alan derslerimle bağlantılı olması	Boolean, Evet Hayır (1/0)
Alan derslerimden bağımsız olması	Boolean, Evet Hayır (1/0)
Yüksek geçme oranına sahip olması	Boolean, Evet Hayır (1/0)
Yüksek not alma ihtimalinin olması	Boolean, Evet Hayır (1/0)
Sözel ağırlıklı içeriğe sahip olması	Boolean, Evet Hayır (1/0)
Sayısal ağırlıklı içeriğe sahip olması	Boolean, Evet Hayır (1/0)
Güncel içeriğe sahip olması	Boolean, Evet Hayır (1/0)
İlgi çekici içeriğe sahip olması	Boolean, Evet Hayır (1/0)
Arkadaş tavsiyesi	Boolean, Evet Hayır (1/0)
Ödev- Proje içermesi	Boolean, Evet Hayır (1/0)
Ödev- Proje içermemesi	Boolean, Evet Hayır (1/0)
Haftalık Programa Uygun olması	Boolean, Evet Hayır (1/0)
Devam gerektirmemesi	Boolean, Evet Hayır (1/0)
Bu dersi tekrar seçerdim.	Boolean, Evet Hayır (1/0)
Bu dersi almaları için başkalarına da tavsiye ederim.	Boolean, Evet Hayır (1/0)

Tablo 32’ de alınan dersin adı 1. veri setinde dersler ayrı ayrı boolean hale getirilerek, 2. veri setinde ise küme (grup) değişken olarak kullanılmıştır. Derslerin seçim sebeplerine ilişkin değişkenler sadece 2. veri setinde yer almıştır. Dersler ve hangi dönemde yer aldıkları EK2’ de verilmiştir.

4.4. VERİ SETİNE İLİŞKİN İSTATİSTİKLER

'Distribution Düğümü' veri setindeki değişkenlere ait dağılımların elde edilmesini sağlar.

Tablo33:Cinsiyet Dağılımı

	Frekans	Yüzde (%)
Kadın	97	64.24
Erkek	54	35.76

– Tablo 33' e göre ankete katılan öğrencilerin 97' si kadın, 54' ü erkektir. Kadın öğrencilerin sayısı erkek öğrencilerin sayısının neredeyse iki katı kadardır.

Tablo34:Öğretim Türü Dağılımı

	Frekans	Yüzde (%)
Örgün	76	50.33
İkinci Öğretim	75	49.67

– Ankette yer alan öğrencilerin 76' sı örgün, 75' i ikinci öğretim öğrencisidir.

Tablo 35:Öğrenim Dönemi

	Frekans	Yüzde (%)
3. yarıyıl	3	2
4. yarıyıl	46	30.5
5. yarıyıl	3	2
6. yarıyıl	28	18.5
7. yarıyıl	0	0
8. yarıyıl	71	47

– Ankette yer alan öğrencilerin 71’ i öğreniminin 8. yarıyılında, 28’i 6. yarıyılında ve 46’ sı ise 4. yarıyılında öğrenimine devam etmektedir. Anketin uygulandığı dönem bahar yarıyılı olmasına rağmen güz yarıyılına ait dönemlerin olduğu görülmektedir. Öğrencilerin 3’ ü 3. yarıyılında, 3’ ü ise 5. yarıyılındadır. Bunun sebebi ise öğrencilerden en çok ilerleme gösterdikleri dönemin seçilmesinin istenmesidir.

Tablo 36:Liseden Mezun Olunan Bölüm Dağılımı

	Frekans	Yüzde (%)
Eşit Ağırlık	134	88.74
Sayısal	15	9.93
Sözel	2	1.32

– Tablo 36’ ya göre ankete katılan öğrencilerin %88.74’ ü lisede Eşit Ağırlık Bölümü’ nü, %9.93’ ü Sayısal Bölümü’ nü, %1.32’ si ise Sözel Bölümü’ nü tamamlamıştır.

Tablo 37:Mezun Olunan Lise Türü Dağılımı

	Frekans	Yüzde (%)
Genel Lise (Düz Lise)	87	57.62
Anadolu Lisesi	54	35.76
Anadolu Meslek Lisesi	4	2.65
Öğretmen Lisesi	2	1.32
Meslek Lisesi	1	0.66
Diğer	3	1.99

– Tablo 37’ ye göre ankete katılan öğrencilerin mezun olduğu lise türü için en yüksek oranın Genel Lise (Düz Lise)’ ye ait olduğu ve %57.62 değere sahip olduğu, ikinci sırada ise %35.76 değeriyle Anadolu Lisesi olduğu sonucuna ulaşılmaktadır.

Tablo 38:Çalışma Durumu Dağılımı

	Frekans	Yüzde (%)
Sadece öğrenciyim	132	87.42
Kısmi zamanlı bir işte çalışıyorum	16	10.6
Tam zamanlı bir işte çalışıyorum	3	1.99

– Tablo 38’ e göre öğrencilerin %87.42 gibi yüksek bir oranının sadece öğrenci olduğu ve herhangi bir işte çalışmadığı, %10.6’ sının kısmi zamanlı bir işte çalıştığı ve %1.99’ unun ise tam zamanlı bir işte çalıştığı sonucu elde edilmektedir.

Tablo 39:Öğrencilerin Tekrar Şans Verilse Hangi Bölümde Okumak İsteyeceklerinin Dağılımı

	Frekans	Yüzde (%)
Ekonometri	31	20.52
Mühendislik Fakültesi	18	11.92
Hukuk Fakültesi	12	7.95
Mimarlık Fakültesi	12	7.95
Eğitim Fakültesi	11	7.28
Güzel Sanatlar Fakültesi	11	7.28
Tıp Fakültesi	7	4.64
Yönetim Bilişim Sistemleri	7	4.64
Denizcilik Fakültesi	6	3.97
Hemşirelik Fakültesi	5	3.31
Edebiyat Fakültesi	4	2.65
Dil ve Tarih-Coğrafya Fakültesi	3	1.99
Maliye	3	1.99
İktisat	3	1.99
İşletme	3	1.99
İşletme Fakültesi	3	1.99
Fen Fakültesi	1	0.66
Turizm Fakültesi	1	0.66
Ziraat Fakültesi	1	0.66
Çalışma Ekonomisi ve Endüstri İlişkileri	1	0.66
Diğer	8	5.3

Tablo 39’ da görüldüğü gibi en yüksek oran Ekonometri Bölümü’ ne aittir ve 31 öğrenci kendilerine tekrar şans verilse, aynı bölümü tekrar tercih edeceklerini

belirtmişlerdir. Bununla birlikte, Ekonometri Bölümü'nden sonraki en yüksek oran Mühendislik Fakültesi' ne aittir ve 18 öğrenci tarafından tercih edilmiştir.

Tablo 40: Öğrenim Dönemine Göre, Öğrencilerin Tekrar Şans Verilse Hangi Bölümde Okumak İsteyeceklerinin Dağılımı (Ekonometri & Diğer)

	Ekonometri	Diğer
3. yarıyıl	0	3
4. yarıyıl	8	38
5. yarıyıl	1	2
6. yarıyıl	3	25
7. yarıyıl	0	0
8. yarıyıl	19	52

– Öğrenim dönemine göre, öğrencilerin tekrar şans verilmesi durumunda hangi bölümü tercih edeceklerine yönelik frekanslar Tablo 40' da verilmiştir. Tabloya göre tekrar Ekonometri Bölümü' nü tercih edecek öğrenciler en çok 8. yarıyıl öğrencilerinden oluşmaktadır ve 19 öğrencidir. Farklı bir bölümü tercih edeceğini belirten öğrenciler ise en çok 8. yarıyıldaki öğrencilerden oluşmakta ve 52 öğrencidir.

Tablo41: Öğrencilerin Mezuniyet Sonrası Hedeflerinin Dağılımı

	Frekans	Yüzde (%)
Özel sektörde çalışmak	63	41.72
Akademisyen olmak	37	24.5
KPSS ile atanabilmek	33	21.85
Kendi işini kurmak	7	4.64
Aile işinde çalışmak	2	1.32
Diğer	9	5.96

Öğrencilerin mezuniyet sonrası hedefleri dağılımı incelendiğinde en yüksek oranın 'özel sektörde çalışmak' olduğu sonucuna varılmaktadır. 63 öğrenci özel

sektörde çalışmayı, 37 öğrenci akademisyen olmayı, 33 öğrenci KPSS ile atanmayı, 7 öğrenci aile işinde çalışmayı ve 2 öğrenci kendi işini kurmayı hedeflemektedir. 9 öğrenci ise diğer seçeneğini seçmiş ve net bir karar bildirmemiştir.

Tablo 42: ‘Seçimlik Dersleri Seçerken En Çok Neyi Referans Alıyorsunuz?’ Dağılımı

	Frekans	Yüzde (%)
Kişisel değerlendirmelerim	65	43.05
Arkadaşlarımın tavsiyeleri	57	37.75
Hoca tavsiye ve yönlendirmeleri	29	19.21

– Tablo 42 incelendiğinde öğrencilerin seçimlik dersleri seçerken referans aldıkları konuların başında kişisel değerlendirmeleri geldiğini, çok yakın bir şekilde bunu arkadaşlarının tavsiyelerini izlediği sonucuna ulaşılmaktadır. Bununla birlikte öğrencilerin %19.21’ i ise ders seçiminde hoca tavsiye ve yönlendirmelerini dikkate almaktadırlar.

Tablo43: Seçimlik Derslerin Dağılımı

	Frekans	Yüzde (%)
Kamu Maliyesi	80	9.28
Kalite Geliştirme Araçları	72	8.35
Üretim Ve Tedarik Planlaması	61	7.08
Süreç Yönetimi	57	6.61
Çalışma Ekonomisi	54	6.26
Kalite Yönetim Sistemleri	53	6.15
Para Teorisi Ve Politikası	48	5.57
Sosyal Güvenlik	45	5.22
Mesleki Yabancı Dil 1	35	4.06
Para Ve Sermaye Piyasaları	34	3.94
Proje Yönetimi	32	3.71
Yabancı Dilde Okuma Ve Konuşma	32	3.71
Yönetim Ve Organizasyon	30	3.48
Ekonometri Semineri	29	3.36
Türkiye Ekonomisi	27	3.13
Maliye Politikası	26	3.02
Veri Madenciliği	22	2.55
Mesleki Yabancı Dil 2	21	2.44
İş Hayatı İçin Yabancı Dil	20	2.32
Yöneylem Araştırma Semineri	18	2.09

Tablo 43 (Devam):Seçimlik Derslerin Dağılımı

Uluslararası İktisat	17	1.97
Yatırım Projeleri Analizi	16	1.86
Borçlar Hukuku	13	1.51
Ticaret Hukuku	4	0.46
Kesit Verisi Ekonometrisi	3	0.35
Çok Kriterli Karar Verme	3	0.35
Coğrafi Bilgi Sistemleri	2	0.23
Finans Matematiği	2	0.23
Pazarlama Araştırması Teknikleri	2	0.23
Bilişim Teknolojisi Tabanlı Karar Verme (Erasmus)	1	0.12
Ekonometri (Erasmus)	1	0.12
Ekonometriye Giriş (Erasmus)	1	0.12
Çok Değişkenli İstatistik: İşletme Ve İktisat Uygulamalı (Erasmus)	1	0.12

Tablo43’ te seçimlik derslerin dağılımı gösterilmiştir. Dağılım dikkate alındığında en yüksek değerın Kamu Maliyesi dersi için olduđu görölmektedir.

Tablo 44: ‘En Kısa Zamanda Mezun Olma Baskısı Hissetmiyorum’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	18	11.92
Katılıyorum	35	23.18
Kararsızım	19	12.58
Katılmıyorum	41	27.15
Kesinlikle Katılmıyorum	38	25.17

– Tablo 44’ e göre öğrencilerin mezun olma baskısı hissedip hissetmemelerine yönelik herhangi bir eğilimin olmadığı ve farklı görüşlerin varlığının ortaya çıktığı sonucuna ulaşılabilir.

Tablo 45: ‘En Kısa Sürede Mezun Olmaktansa, Donanımlı Mezun Olmayı Tercih Ederim’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	50	33.11
Katılıyorum	57	37.75
Kararsızım	29	19.21
Katılmıyorum	11	7.28
Kesinlikle Katılmıyorum	4	2.65

– Tablo 45’ e göre öğrencilerin büyük bir bölümü, neredeyse %71’i donanımlı mezun olmayı, en kısa sürede mezun olmaya tercih etmekte, yaklaşık % 10’ u ise tam tersi yönde tercih belirtmektedirler.

Tablo 46: ‘Ekonometri Bölümü’ ne İsteyerek, Bilerek, Araştırarak Geldim’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	31	20.53
Katılıyorum	51	33.77
Kararsızım	19	12.58
Katılmıyorum	29	19.21
Kesinlikle Katılmıyorum	21	13.91

– Tablo 46’ ya göre öğrencilerin neredeyse %55’ i Ekonometri Bölümü’ ne isteyerek, bilerek ve araştırarak gelmiştir.

Tablo 47: Öğrenim Dönemine Göre, ‘Ekonometri Bölümü’ ne İsteyerek, Bilerek, Araştırarak Geldim’ Dağılımı

	Kesinlikle Katılmıyorum	Katılmıyorum	Kararsızım	Katılıyorum	Kesinlikle Katılıyorum
3. yarıyıl	1	1	1	0	0
4. yarıyıl	6	7	6	17	10
5. yarıyıl	0	1	1	1	0
6. yarıyıl	4	7	4	8	5
7. yarıyıl	0	0	0	0	0
8. yarıyıl	10	13	7	25	16

– Tablo 47’ de öğrencilerin öğrenim dönemlerine göre ‘Ekonometri Bölümü’ ne isteyerek, bilerek, araştırarak geldim’ ifadesine verdikleri yanıtlar verilmiştir.

Tablo48: ‘Okuduğum Bölümü Benimsedim ve Seviyorum’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	29	19.21
Katılıyorum	62	41.06
Kararsızım	36	23.84
Katılmıyorum	13	8.61
Kesinlikle Katılmıyorum	11	7.28

– Tablo 48’ e göre öğrencilerin büyük bir çoğunluğu Ekonometri Bölümü’ nün benimsemiş ve sevmişlerdir. Bununla birlikte yaklaşık % 15’ lik bir kısım ise aksi yönde görüş ifade etmişlerdir.

Tablo 49: Öğrenim Dönemine Göre, ‘Okuduğum Bölümü Benimsedim ve Seviyorum’ Dağılımı

	Kesinlikle Katılmıyorum	Katılmıyorum	Kararsızım	Katılıyorum	Kesinlikle Katılıyorum
3. yarıyıl	0	1	2	2	0
4. yarıyıl	3	3	15	15	10
5. yarıyıl	0	0	1	1	0
6. yarıyıl	2	3	5	5	4
7. yarıyıl	0	0	0	0	0
8. yarıyıl	6	6	13	31	15

– Tablo 49’ da öğrencilerin öğrenim dönemlerine göre ‘Okuduğum bölümü benimsedim ve seviyorum’ ifadesine verdikleri yanıtlar verilmiştir.

Tablo50: ‘Dokuz Eylül Üniversitesi’ ne İsteyerek Geldim’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	70	46.36
Katılıyorum	57	37.75
Kararsızım	12	7.95
Katılmıyorum	7	4.64
Kesinlikle Katılmıyorum	5	3.31

– Tablo 50’ ye göre öğrencilerin büyük bir çoğunluğu Dokuz Eylül Üniversitesi’ ne isteyerek geldiğini belirtmiştir. Tablo 46’ ya dönersek bunu şu şekilde yorumlayabiliriz: Öğrencilerin bir kısmı bölümden ziyade üniversite tercihi yapmışlardır.

Tablo 51: ‘Araştırma Yapmayı Severim’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	50	33.11
Katılıyorum	70	46.36
Kararsızım	19	12.58
Katılmıyorum	9	5.96
Kesinlikle Katılmıyorum	3	1.99

– Öğrencilerin yaklaşık %80’ inin araştırma yapmayı sevdikleri yönünde görüş bildirdikleri Tablo 51’ de görülmektedir.

Tablo 52: ‘Bilgisayar Kullanımına Erişimim Kolaydır’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	52	34.44
Katılıyorum	68	45.03
Kararsızım	21	13.91
Katılmıyorum	9	5.96
Kesinlikle Katılmıyorum	1	0.66

– Bilgisayar kullanımına erişimlerinin kolay olup olmadığıyla ilgili ifadeye ise öğrencilerin büyük bir çoğunluğu erişimlerinin kolay olduğunu belirtmişlerdir.

Tablo 53: ‘Yabancı Dil Düzeyimin Yeterli Olduğunu Düşünüyorum’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	9	5.96
Katılıyorum	18	11.92
Kararsızım	46	30.46
Katılmıyorum	55	36.42
Kesinlikle Katılmıyorum	23	15.23

– Yabancı dil düzeylerinin yeterliliği konusunda ise öğrencilerin büyük bir çoğunluğu negatif yönlü cevap vermişlerdir. Sadece yaklaşık % 18’ lik bir öğrenci dilimi yabancı dil konusunda yeterli olduklarını düşünmektedirler.

Tablo 54: ‘Benim İçin Yeni Bir Şey Öğrenmekten Çok Dersi Geçmek Daha Önemlidir’ Dağılımı

	Frekans	Yüzde (%)
Kesinlikle Katılıyorum	9	5.96
Katılıyorum	17	11.26
Kararsızım	39	25.83
Katılmıyorum	49	32.45
Kesinlikle Katılmıyorum	37	24.5

Tablo 54 incelendiğinde ise öğrencilerin büyük bir çoğunluğunun daha önceki elde edilen bilgileri destekler nitelikte olduğu ve öğrencilerin dersi geçmeye yönelik düşünceden ziyade yeni bir şey öğrenmeye yönelik isteklerinin ağır bastığı görülmektedir.

4.5. MODELLEME

Yapılan çalışmada birliktelik kuralları algoritmalarından Apriori Algoritması kullanılmıştır. Yapılan denemeler sonucunda Apriori Algoritması çeşitli destek ve güven değerleriyle ve çeşitli kombinasyonlarla her iki veri seti için de çalıştırılmıştır.

Elde edilen tablolarda ‘consequent’ kuralın sonucunu, ‘antecedent’ kuralın öncülünü, ‘confidence’ kuralın güven değerini ve ‘rulesupport’ ise kuralın destek değerini göstermektedir. Minimum değerler belirlenirken destek değeri için tanımlama yapılacaktır. Destek değeri ve kural destek değerinin birbirinden farklı olduğuna dikkat edilmelidir.

Öncelikle 5 modele ait değişkenlere yönelik belirlenen özellikler Tablo 55’ te verilmiştir. Değişkenlerin ‘input’ olarak belirlenmesi modelin öncül kısmında yer almasının istenilmesinden, ‘output’ olarak belirlenmesi sonuç kısmında yer almasının istenilmesinden, ‘both’ kısmında yer alması ise hem öncül hem sonuç kısmında yer

almasının istenilmesinden kaynaklanmaktadır. Yapılan bu düzenlemeler ‘Type Dügümü’ ile gerçekleştirilmektedir.

Tablo 55: Apriori Algoritması İle Elde Edilen Modellere Ait Değişken Özellikleri

	Input	Output	Both
1.Model	-	-	• Seçimlik dersler
2.Model	• Seçimlik dersler dışındaki değişkenler	• Seçimlik dersler	-
3.Model	• Ders seçimine etki eden faktörler dışındaki değişkenler	• Ders seçimine etki eden sebeplere ait değişkenler	-
4.Model	• İlgi alanları dışındaki değişkenler	• İlgi alanlarına ait değişkenler	-
5.Model	• Seçimlik dersler	• Ders seçimine etki eden sebeplere ait değişkenler • İlgi alanlarına ait değişkenler	-

1.model

1.veri seti kullanılarak, seçimlik dersler ‘Type Dügümü’ ile ‘both’ olarak belirlenmiştir. Minimum destek değeri %10, minimum güven değeri %75, öncül sayısı en fazla 5 olacak şekilde belirlenerek model çalıştırılmıştır.

Sonuç olarak Şekil 21’ de gösterilen şekilde 226adet kural elde edilmiştir. Kurallar incelenerek kurallardan bazıları Tablo 56’ da verilmiş ve yorumlanmıştır.

Şekil 21:1. Model ile Elde Edilen Birliktelik Kuralları

Consequent	Antecedent	Confidence %	Rule Support %
Kalite Yönetim Sistemleri	Türkiye Ekonomisi Sosyal Güvenlik	100.000	10.596
Kalite Yönetim Sistemleri	Türkiye Ekonomisi Kamu Maliyesi	100.000	12.583
Kalite Yönetim Sistemleri	Ekonometri Semineri Sosyal Güvenlik	100.000	10.596
Kamu Maliyesi	Para Ve Sermaye Piyasaları Para Teorisi Ve Politikası Kalite Yönetim Sistemleri	100.000	11.921
Kamu Maliyesi	Para Ve Sermaye Piyasaları Çalışma Ekonomisi Kalite Yönetim Sistemleri	100.000	12.583

Tablo 56:1. Model ile Elde Edilen Birliktelik Kuralları

Sonuç	Öncül	Güven%	Kural Destek%
–Kalite Yönetim Sistemleri	–Türkiye Ekonomisi –Sosyal Güvenlik	100.0	10.59
–Kalite Geliştirme Araçları	–Proje Yönetimi –Kalite Yönetim Sistemleri	88.23	9.93
–Kalite Geliştirme Araçları	–Kalite Yönetim Sistemleri	75.0	25.82

– Türkiye Ekonomisi ve Sosyal Güvenlik derslerini birlikte seçen öğrencilerin %100’ ü aynı zamanda Kalite Yönetim Sistemleri dersini de seçmişlerdir.Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 10.59’ dur.

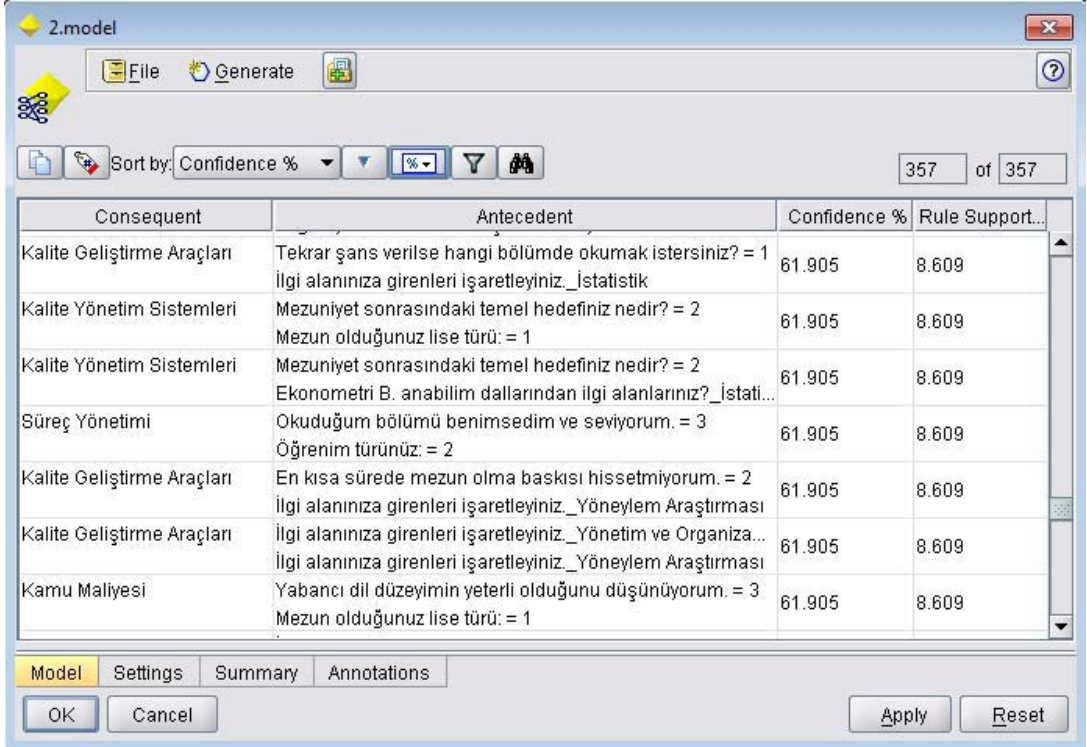
– Proje Yönetimi ve Kalite Yönetim Sistemleri derslerini birlikte seçen öğrencilerin % 88.23’ ü aynı zamanda Kalite Geliştirme Araçları dersini de seçmişlerdir.Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 9.93’ tür.

– Kalite Yönetim Sistemleri dersini seçen öğrencilerin % 75’ i aynı zamanda Kalite Geliştirme Araçları dersini de seçmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 25.82’ dir.

2.model

1.veri seti kullanılarak, seçimlik dersler ‘output’, veri setindeki diğer değişkenler ise ‘input olarak belirlenmiştir. Minimum destek değeri %10, minimum güven değeri %60, öncül sayısı en fazla 2 olacak şekilde belirlenmiştir. Model çalıştırılmıştır. Sonuç olarak Şekil 22’ de gösterilen şekilde 357 tane kural oluşturulmuştur. Kurallar incelenerek kurallardan bazıları Tablo 57’ de verilmiş ve devamında yorumlanmıştır.

Şekil 22:2. Model İle Elde Edilen Birliktelik Kuralları



Consequent	Antecedent	Confidence %	Rule Support...
Kalite Geliştirme Araçları	Tekrar şans verilse hangi bölümde okumak istersiniz? = 1 İlgi alanınıza girenleri işaretleyiniz_ İstatistik	61.905	8.609
Kalite Yönetim Sistemleri	Mezuniyet sonrasındaki temel hedefiniz nedir? = 2 Mezun olduğunuz lise türü: = 1	61.905	8.609
Kalite Yönetim Sistemleri	Mezuniyet sonrasındaki temel hedefiniz nedir? = 2 Ekonometri B. anabilim dallarından ilgi alanlarınız?_ İstati...	61.905	8.609
Süreç Yönetimi	Okuduğum bölümü benimsedim ve seviyorum. = 3 Öğrenim türünüz: = 2	61.905	8.609
Kalite Geliştirme Araçları	En kısa sürede mezun olma baskısı hissetmiyorum. = 2 İlgi alanınıza girenleri işaretleyiniz_ Yöneylem Araştırması	61.905	8.609
Kalite Geliştirme Araçları	İlgi alanınıza girenleri işaretleyiniz_ Yönetim ve Organiza... İlgi alanınıza girenleri işaretleyiniz_ Yöneylem Araştırması	61.905	8.609
Kamu Maliyesi	Yabancı dil düzeyimin yeterli olduğunu düşünüyorum. = 3 Mezun olduğunuz lise türü: = 1	61.905	8.609

Tablo 57:2. Model ile Elde Edilen Birliktelik Kuralları

Sonuç	Öncül	Güven%	Kural Destek%
-Kalite Geliştirme Araçları	-Tekrar şans verilse hangi bölümde okumak istersiniz? = Ekonometri -İlgi alanınıza girenleri işaretleyiniz. _İstatistik	61.90	8.60
-Kamu Maliyesi	-Yabancı dil düzeyimin yeterli olduğunu düşünüyorum. = Kararsızım -Mezun olduğunuz lise türü: = Düz lise	61.90	8.60
-Kalite Yönetim Sistemleri	-İlgi alanınıza girenleri işaretleyiniz. _Kalite -Öğretim türünüz: = Örgün	60.0	9.93

- Tekrar şans verilse okuyacağı bölüm olarak Ekonometri Bölümü' nü tercih edeceğini belirtmiş olan ve ilgi alanı İstatistik olan öğrencilerin % 61.90' ı Kalite Geliştirme Araçları dersini seçmişlerdir.Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 8.60' tır.

- Yabancı dil düzeyinin yeterli olduğu konusunda kararsız olduğunu belirtmiş olan ve mezun olduğu okul türü düz lise olan öğrencilerin% 61.90' ı Kamu Maliyesi dersini seçmişlerdir.Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 8.60' tır.

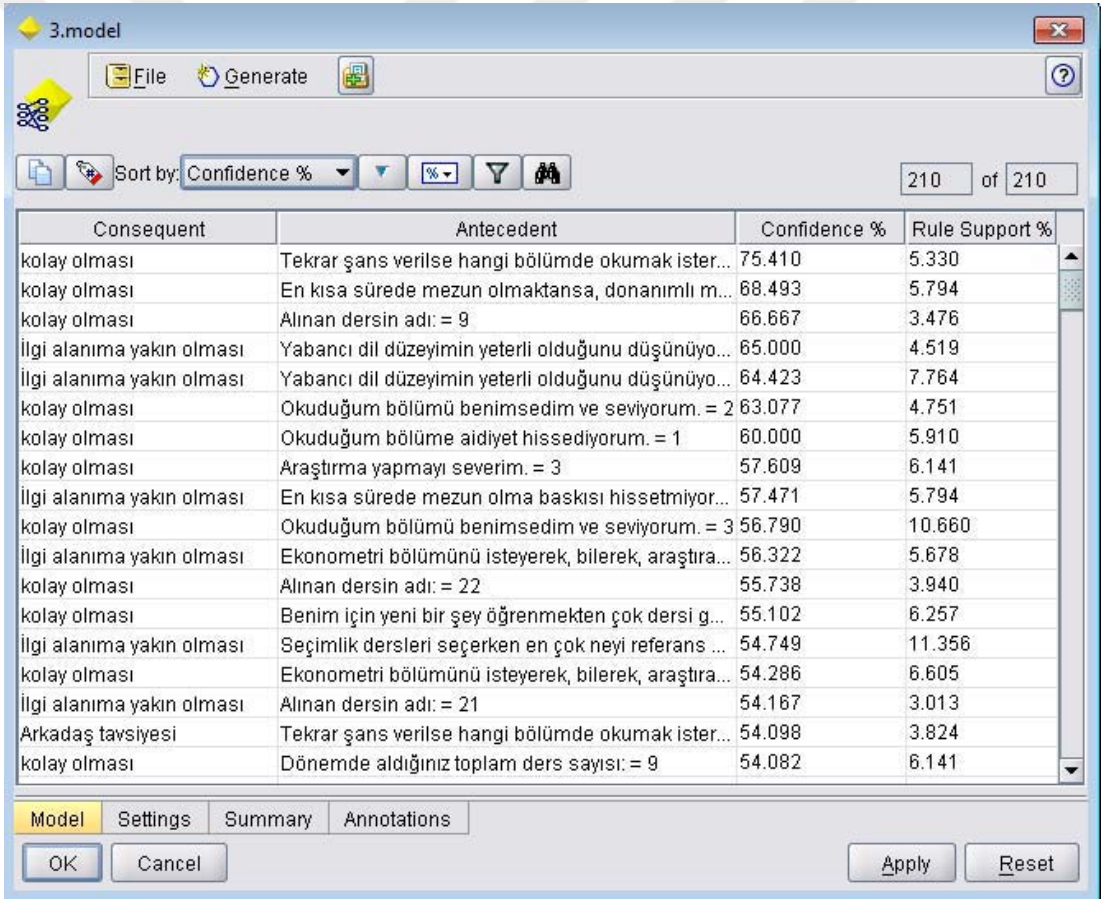
- İlgi alanı Kalite olan örgün öğretim öğrencilerininin % 60' ı Kalite Yönetim Sistemleri dersini seçmişlerdir.Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 9.93' tür.

Elde edilen kurallar incelendiğinde öğrencilerin ilgi alanları doğrultusunda ders seçimine yöneldikleri görülmektedir. Aynı zamanda yabancı dil düzeyinin yeterli olmadığını düşünen öğrencilerin aynı dönemde İngilizce dersini değil Kamu Maliyesi dersini seçmelerini kendilerini bu alanda yetersiz görmeleri olarak yorumlayabiliriz.

3.model

2. veri seti kullanılarak, ders seçimine etki eden nedenlere ait değişkenler 'output', veri setindeki diğer değişkenler 'input' olarak belirlenmiştir. Minimum destek değeri %5, minimum güven değeri %35, öncül sayısı en fazla 1 olacak şekilde belirlenerek model çalıştırılmıştır. Sonuç olarak Şekil 23' te gösterilen şekilde 210 tane kural oluşturulmuştur. Kurallar incelenerek kurallardan bazıları Tablo 58' de verilmiş ve devamında yorumlanmıştır.

Şekil 23:3. Model ile Elde Edilen Birliklik Kuralları



Consequent	Antecedent	Confidence %	Rule Support %
kolay olması	Tekrar şans verilse hangi bölümde okumak ister...	75.410	5.330
kolay olması	En kısa sürede mezun olmaksansa, donanımlı m...	68.493	5.794
kolay olması	Alınan dersin adı: = 9	66.667	3.476
İlgi alanıma yakın olması	Yabancı dil düzeyimin yeterli olduğunu düşünüyö...	65.000	4.519
İlgi alanıma yakın olması	Yabancı dil düzeyimin yeterli olduğunu düşünüyö...	64.423	7.764
kolay olması	Okuduğum bölümü benimsedim ve seviyorum. = 2	63.077	4.751
kolay olması	Okuduğum bölüme aidiyet hissediyorum. = 1	60.000	5.910
kolay olması	Araştırma yapmayı severim. = 3	57.609	6.141
İlgi alanıma yakın olması	En kısa sürede mezun olma baskısı hissetmiyor...	57.471	5.794
kolay olması	Okuduğum bölümü benimsedim ve seviyorum. = 3	56.790	10.660
İlgi alanıma yakın olması	Ekonometri bölümünü isteyerek, bilerek, araştıra...	56.322	5.678
kolay olması	Alınan dersin adı: = 22	55.738	3.940
kolay olması	Benim için yeni bir şey öğrenmekten çok dersi g...	55.102	6.257
İlgi alanıma yakın olması	Seçimlik dersleri seçerken en çok neyi referans ...	54.749	11.356
kolay olması	Ekonometri bölümünü isteyerek, bilerek, araştıra...	54.286	6.605
İlgi alanıma yakın olması	Alınan dersin adı: = 21	54.167	3.013
Arkadaş tavsiyesi	Tekrar şans verilse hangi bölümde okumak ister...	54.098	3.824
kolay olması	Dönemde aldığınız toplam ders sayısı: = 9	54.082	6.141

Tablo 58:3. Model ile Elde Edilen Birliktelik Kuralları

Sonuç	Öncül	Güven%	Kural Destek%
-Kolay olması	-Tekrar şans verilse hangi bölümde okumak istersiniz? = Eğitim Fakültesi	75.41	5.33
-Kolay olması	-Benim için yeni bir şey öğrenmekten çok dersi geçmek daha önemlidir. = Katılıyorum	55.10	6.25
-Kolay olması	-En kısa sürede mezun olma baskısı hissetmiyorum. = Kesinlikle Katılmıyorum	46.88	11.35
-Önemli kazanım sunuyor olması	-Seçimlik dersleri seçerken en çok neyi referans alıyorsunuz? = Hoca tavsiye ve yönlendirmeleri	43.01	8.92
-Önemli kazanım sunuyor olması	-Mezuniyet sonrasındaki temel hedefiniz nedir? = Akademisyen olmak	42.28	9.84

- Tekrar şans verilse Eğitim Fakültesi'nde okumayı tercih edeceğini belirten öğrencilerin % 75.41' i seçtiği dersi kolay olduğu için tercih ettiklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 5.33' tür.

- 'Benim için yeni bir şey öğrenmekten çok dersi geçmek daha önemlidir' ifadesine 'Katılıyorum' şeklinde yanıt veren öğrencilerin %55.10' u seçtiği dersi kolay olmasından dolayı seçtiklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 6.25' tir.

- 'En kısa sürede mezun olma baskısı hissetmiyorum' ifadesine 'Kesinlikle Katılmıyorum' şeklinde yanıt veren öğrencilerin % 46.88' i seçtiği dersi kolay olmasından dolayı seçtiklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 11.35' tir.

- 'Seçimlik dersleri seçerken en çok neyi referans alıyorsunuz' sorusuna 'hoca tavsiye ve yönlendirmeleri' cevabını veren öğrencilerin % 43.01' i seçtiği dersi

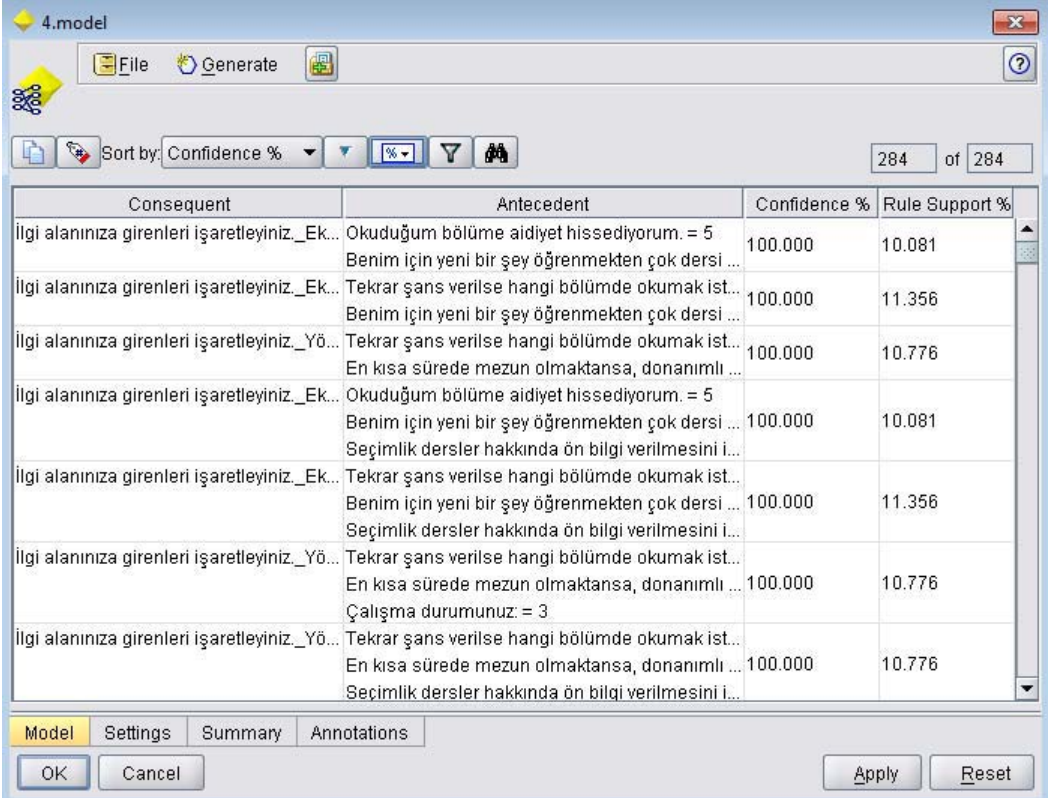
önemli kazanım sunuyor olmasından dolayı seçtiklerini belirtmişlerdir.Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 8.92' dir.

– Mezuniyet sonrasındaki temel hedefi akademisyen olmak olan öğrencilerin % 42.28'i seçtiği dersi önemli kazanım sunuyor olmasından dolayı seçtiklerini belirtmişlerdir.Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 9.84' tür.

4.model

2.veri seti kullanılarak, ilgi alanlarına ait değişkenler 'output', veri setindeki diğer değişkenler 'input' olarak belirlenmiştir. Minimum destek değeri %10, minimum güven değeri %90, öncül sayısı en fazla 5 belirlenerek model çalıştırılmıştır.Sonuç olarak Şekil 24' te gösterilen şekilde 284 tane kural oluşturulmuştur. Kurallar incelenerek bazıları Tablo 59' da verilmiş ve devamında yorumlanmıştır.

Şekil 24: 4. Model ile Elde Edilen Birliktelik Kuralları



Consequent	Antecedent	Confidence %	Rule Support %
İlgi alanınıza girenleri işaretleyiniz_Ek...	Okuduğum bölüme aidiyet hissediyorum. = 5 Benim için yeni bir şey öğrenmekten çok dersi ...	100.000	10.081
İlgi alanınıza girenleri işaretleyiniz_Ek...	Tekrar şans verilse hangi bölümde okumak ist... Benim için yeni bir şey öğrenmekten çok dersi ...	100.000	11.356
İlgi alanınıza girenleri işaretleyiniz_Yö...	Tekrar şans verilse hangi bölümde okumak ist... En kısa sürede mezun olmaktansa, donanımlı ...	100.000	10.776
İlgi alanınıza girenleri işaretleyiniz_Ek...	Okuduğum bölüme aidiyet hissediyorum. = 5 Benim için yeni bir şey öğrenmekten çok dersi ...	100.000	10.081
İlgi alanınıza girenleri işaretleyiniz_Ek...	Seçimlik dersler hakkında ön bilgi verilmesini i... Tekrar şans verilse hangi bölümde okumak ist... Benim için yeni bir şey öğrenmekten çok dersi ...	100.000	11.356
İlgi alanınıza girenleri işaretleyiniz_Yö...	Tekrar şans verilse hangi bölümde okumak ist... En kısa sürede mezun olmaktansa, donanımlı ...	100.000	10.776
İlgi alanınıza girenleri işaretleyiniz_Yö...	Çalışma durumunuz: = 3 Tekrar şans verilse hangi bölümde okumak ist... En kısa sürede mezun olmaktansa, donanımlı ...	100.000	10.776

Tablo 59:4. Model ile Elde Edilen Birliktelik Kuralları

Sonuç	Öncül	Güven%	Kural Destek%
–İlgi alanınıza girenleri işaretleyiniz._Yöneylem Araştırması	–Tekrar şans verilse hangi bölümde okumak istersiniz? = Ekonometri –En kısa sürede mezun olmaksansa, donanımlı mezun olmayı tercih ederim. = Katılıyorum	100.0	10.77
–İlgi alanınıza girenleri işaretleyiniz._Ekonometri	–Tekrar şans verilse hangi bölümde okumak istersiniz? = Ekonometri –Araştırma yapmayı severim= Kesinlikle Katılıyorum	97.98	11.24
–İlgi alanınıza girenleri işaretleyiniz._İstatistik	–Tekrar şans verilse hangi bölümde okumak istersiniz? = Ekonometri –Benim için yeni bir şey öğrenmekten çok dersi geçmek daha önemlidir. = Kesinlikle Katılmıyorum	91.83	10.42

– Tekrar şans verilse Ekonometri Bölümü’ nde okumayı tercih edeceğini ve ‘en kısa sürede mezun olmaksansa, donanımlı mezun olmayı tercih ederim’ ifadesine ‘katılıyorum’ şeklinde yanıt veren öğrencilerin % 100’ ü ilgi alanı olarak Yöneylem Araştırması’ nı seçmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 10.77’ dir.

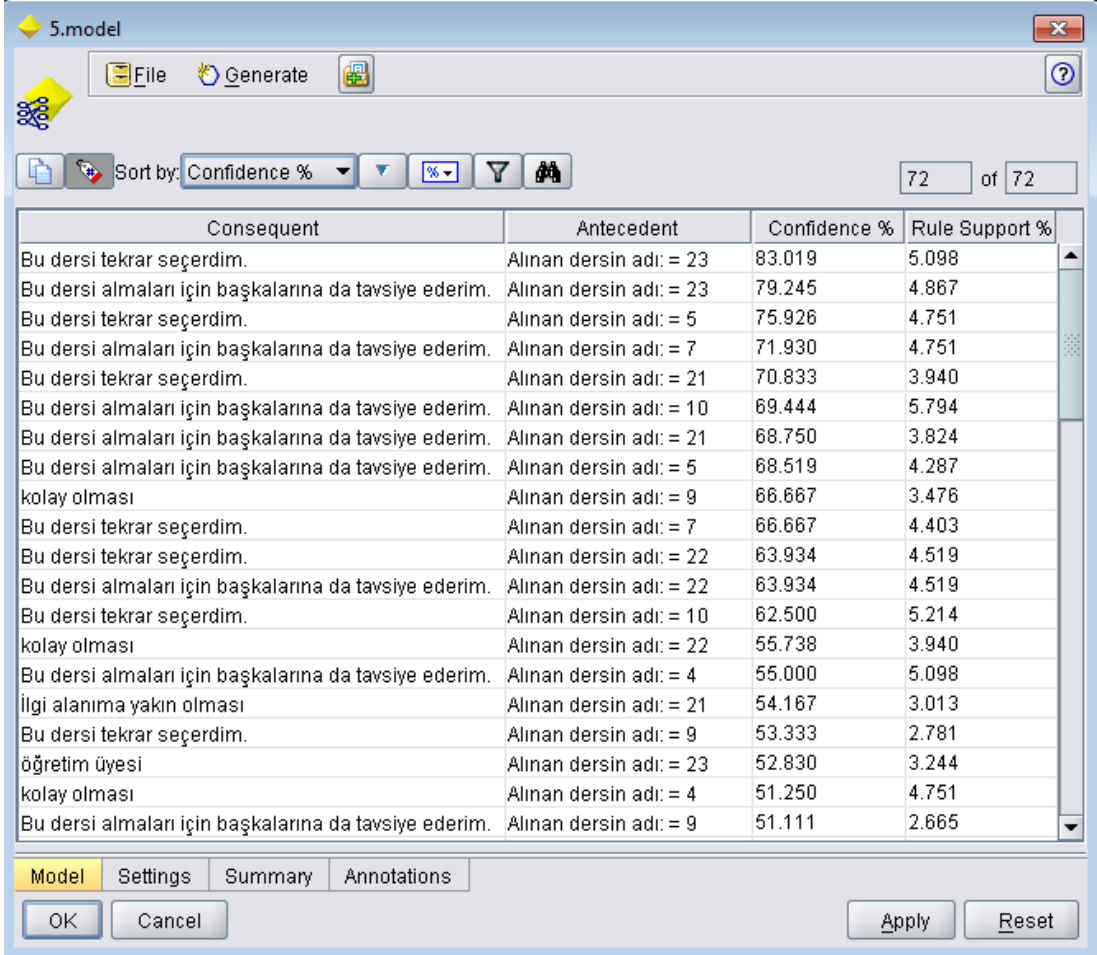
– Tekrar şans verilse Ekonometri Bölümü’ nde okumayı tercih edeceğini ve ‘araştırma yapmayı severim’ ifadesine ‘kesinlikle katılıyorum’ şeklinde yanıt veren öğrencilerin % 97.98’ i ilgi alanını İstatistik olarak seçmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 11.24’ tür.

– Tekrar şans verilse Ekonometri Bölümü’ nde okumayı tercih edeceğini ve ‘benim için yeni bir şey öğrenmekten çok dersi geçmek daha önemlidir’ ifadesine ‘Kesinlikle Katılmıyorum’ şeklinde yanıt veren öğrencilerin % 91.83’ ü ilgi alanını İstatistik olarak seçmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 10.42’ dir.

5.model

2. veri seti kullanılarak, ders seçimine etki eden sebeplere ait değişkenler ve ilgi alanlarına ait değişkenler 'output', seçimlik dersler 'input' olarak belirlenmiştir. Minimum destek değeri %5, minimum güven değeri %20, öncül sayısı en fazla 5 olarak belirlenerek model çalıştırılmıştır. Sonuç olarak Şekil 25' te gösterilen şekilde 72 tane kural oluşturulmuştur. Kurallar incelenerek bazıları Tablo 60' daverilmiş ve devamında yorumlanmıştır.

Şekil 25:5. Model ile Elde Edilen Birliktelik Kuralları



Consequent	Antecedent	Confidence %	Rule Support %
Bu dersi tekrar seçerdim.	Alınan dersin adı: = 23	83.019	5.098
Bu dersi almaları için başkalarına da tavsiye ederim.	Alınan dersin adı: = 23	79.245	4.867
Bu dersi tekrar seçerdim.	Alınan dersin adı: = 5	75.926	4.751
Bu dersi almaları için başkalarına da tavsiye ederim.	Alınan dersin adı: = 7	71.930	4.751
Bu dersi tekrar seçerdim.	Alınan dersin adı: = 21	70.833	3.940
Bu dersi almaları için başkalarına da tavsiye ederim.	Alınan dersin adı: = 10	69.444	5.794
Bu dersi almaları için başkalarına da tavsiye ederim.	Alınan dersin adı: = 21	68.750	3.824
Bu dersi almaları için başkalarına da tavsiye ederim.	Alınan dersin adı: = 5	68.519	4.287
kolay olması	Alınan dersin adı: = 9	66.667	3.476
Bu dersi tekrar seçerdim.	Alınan dersin adı: = 7	66.667	4.403
Bu dersi tekrar seçerdim.	Alınan dersin adı: = 22	63.934	4.519
Bu dersi almaları için başkalarına da tavsiye ederim.	Alınan dersin adı: = 22	63.934	4.519
Bu dersi tekrar seçerdim.	Alınan dersin adı: = 10	62.500	5.214
kolay olması	Alınan dersin adı: = 22	55.738	3.940
Bu dersi almaları için başkalarına da tavsiye ederim.	Alınan dersin adı: = 4	55.000	5.098
İlgi alanıma yakın olması	Alınan dersin adı: = 21	54.167	3.013
Bu dersi tekrar seçerdim.	Alınan dersin adı: = 9	53.333	2.781
öğretim üyesi	Alınan dersin adı: = 23	52.830	3.244
kolay olması	Alınan dersin adı: = 4	51.250	4.751
Bu dersi almaları için başkalarına da tavsiye ederim.	Alınan dersin adı: = 9	51.111	2.665

Tablo 60:5. Model ile Elde Edilen Birliktelik Kuralları

Sonuç	Öncül	Güven%	Kural Destek%
–Bu dersi tekrar seçerdim.	–Kalite Yönetim Sistemleri	83.01	5.09
–Bu dersi almaları için başkalarına da tavsiye ederim.	–Kalite Geliştirme Araçları	69.44	5.79
–İli alanıma yakın olması	–Para Teorisi Ve Politikası	54.16	3.01
–Öğretim üyesi	–Süreç Yönetimi	45.61	3.01
–Arkadaş tavsiyesi	–Üretim Ve Tedarik Planlaması	39.3	2.78
–Yüksek not alma ihtimalinin olması	–Kamu Maliyesi	30.00	2.78

– Kalite Yönetim Sistemleri dersini seçen öğrencilerin % 83.01’ i dersi tekrar olsa yine seçeceklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 5.09’ dur.

– Kalite Geliştirme Araçları dersini seçen öğrencilerin % 69.44’ ü bu dersi almaları için başkalarına da tavsiye edeceklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 5.79’ dur.

– Para Teorisi Ve Politikası dersini seçen öğrencilerin % 54.16’ sı bu dersi ilgi alanına yakın olmasından dolayı seçtiklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 3.01’ dir.

– Süreç Yönetimi dersini seçen öğrencilerin % 45.61’ i bu dersi öğretim üyesinden dolayı seçtiklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 3.01’ dir.

– Üretim Ve Tedarik Planlaması dersini seçen öğrencilerin % 39.34’ ü bu dersi arkadaş tavsiyesinden dolayı seçtiklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 2.78’ dir.

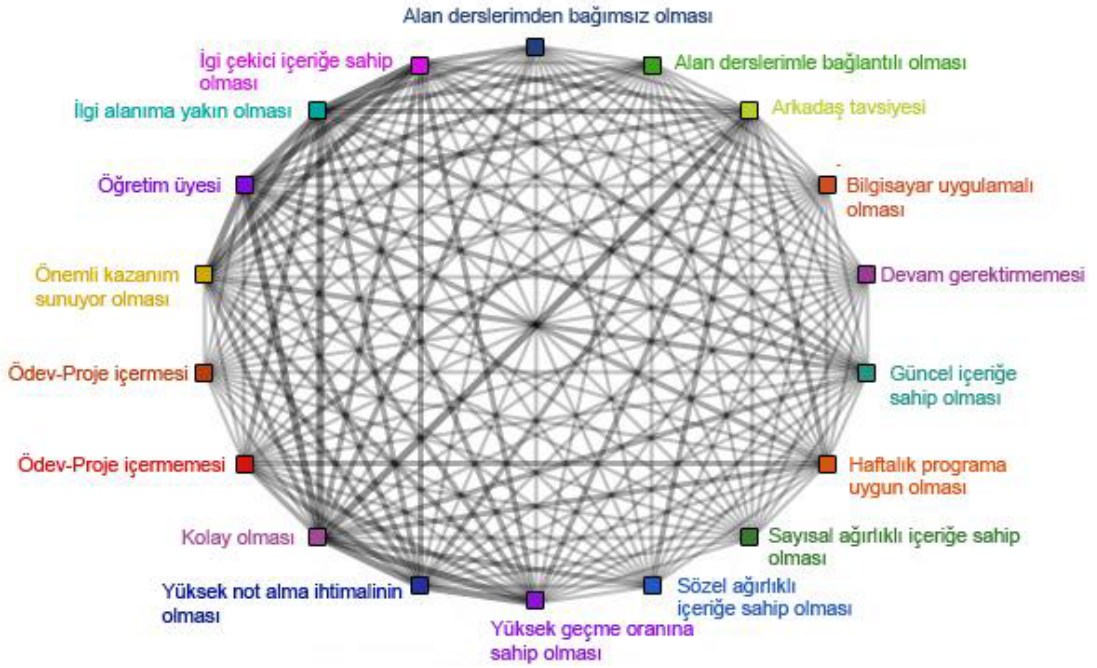
– Kamu Maliyesi dersini seçen öğrencilerin % 30’ u bu dersi yüksek not alma ihtimalinin olmasından dolayı seçtiklerini belirtmişlerdir. Bu kuralın veri setindeki işlemlerde birlikte bulunma olasılığı ise % 2.78’ dir.

4.6. BİRLİKTELİK DİYAGRAMLARI (AĞ GRAFİKLERİ)

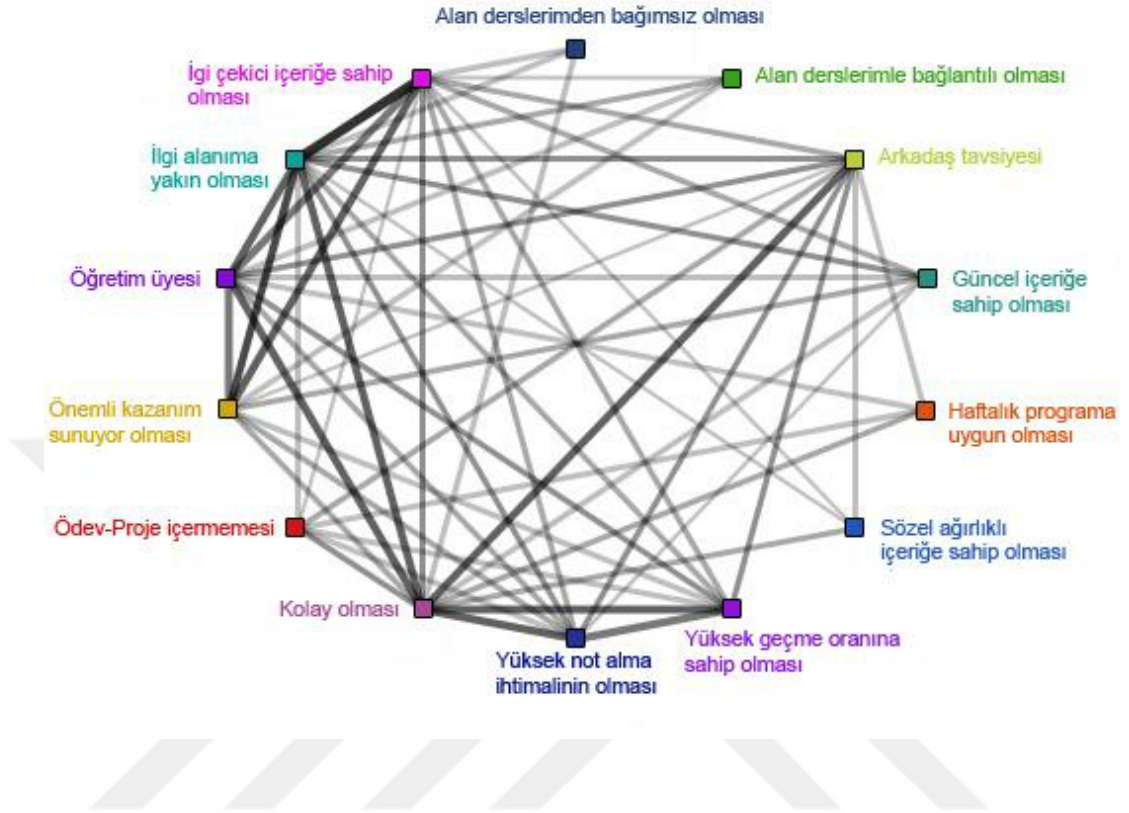
Ders seçim sebeplerinin arasındaki ilişkileri, seçilen dersler arasındaki ilişkileri ve ilgi alanları arasındaki ilişkileri görebilmek için Clementine’ da ağ grafiğinden yararlanılmıştır. Ağ grafiğindeki çizgi kalınlığı ilişkinin kuvvetini göstermektedir. Çizgi kalınlığı arttıkça ilişkinin kuvveti de artmaktadır.

Öncelikle ders seçim sebeplerinin aralarındaki ilişkileri görmek için 862 kaydın yer aldığı 2. veri seti için ağ grafiği kullanılmıştır. Şekil 26’ da çizgi değeri 0 iken ilişkinin en karmaşık hali gösterilmiş, Şekil 27’ de ise çizgi seviyesi arttırılarak ilişkiler daha görünür hale getirilmiştir.

Şekil 26: Ders Seçim Sebeplerine İlişkin Çizgi Seviyesi 0 İken Elde Edilen Ağ Grafiği



Şekil 27:Ders Seçim Sebeplerine İlişkin Çizgi Seviyesi 45 İken Elde Edilen Ağ Grafığı



Şekil 28:Ders Seçim Sebeplerine İlişkin En Güçlü Birliktelikler

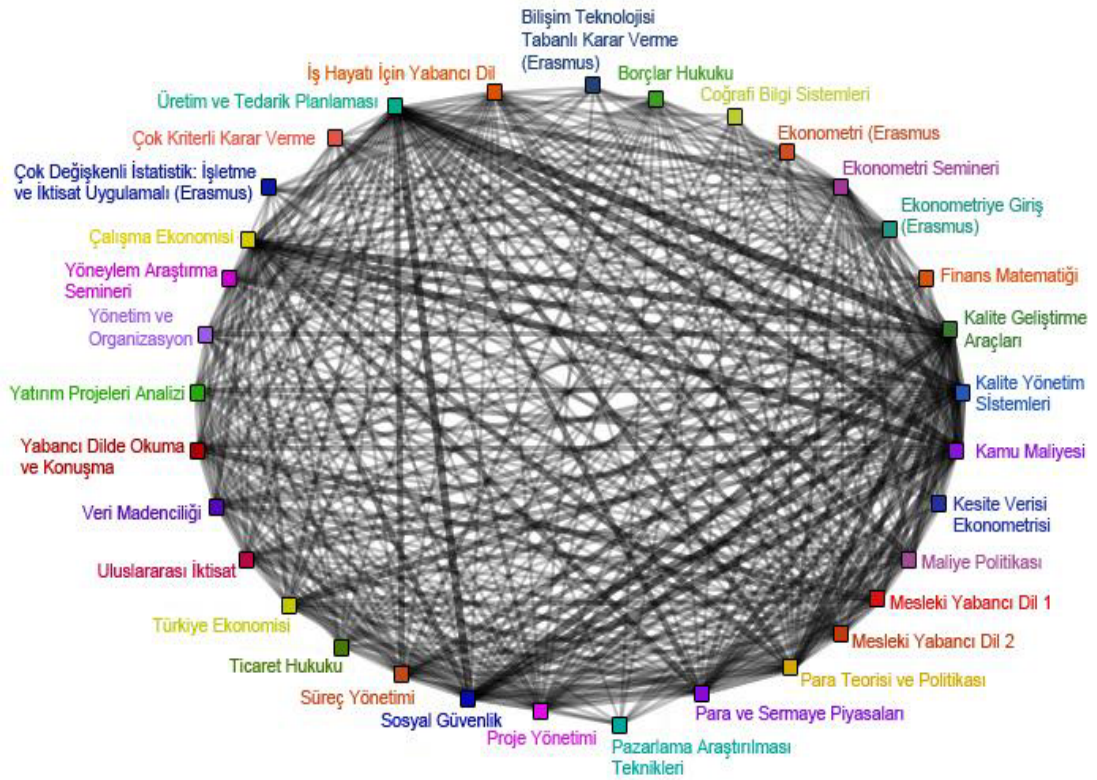
Strong Links

Links	Field 1	Field 2
172	İlgili alanıma yakın olması = "1"	İlgili çekici içeriğe sahip olması = "1"
148	İlgili alanıma yakın olması = "1"	Önemli kazanım sunuyor olması = "1"
135	Önemli kazanım sunuyor olması = "1"	İlgili çekici içeriğe sahip olması = "1"
134	kolay olması = "1"	öğretim üyesi = "1"
131	kolay olması = "1"	Yüksek geçme oranına sahip olması = "1"
129	kolay olması = "1"	İlgili alanıma yakın olması = "1"
128	kolay olması = "1"	Arkadaş tavsiyesi = "1"
126	öğretim üyesi = "1"	İlgili alanıma yakın olması = "1"
122	kolay olması = "1"	Yüksek not alma ihtimalinin olması = "1"
112	öğretim üyesi = "1"	Önemli kazanım sunuyor olması = "1"
107	öğretim üyesi = "1"	İlgili çekici içeriğe sahip olması = "1"
104	Yüksek geçme oranına sahip olması = "1"	Yüksek not alma ihtimalinin olması = "1"
93	kolay olması = "1"	Ödev- Proje içermemesi = "1"
92	İlgili alanıma yakın olması = "1"	Yüksek not alma ihtimalinin olması = "1"
89	öğretim üyesi = "1"	Arkadaş tavsiyesi = "1"
88	İlgili alanıma yakın olması = "1"	Güncel içeriğe sahip olması = "1"
86	Yüksek not alma ihtimalinin olması = "1"	Arkadaş tavsiyesi = "1"
84	kolay olması = "1"	İlgili çekici içeriğe sahip olması = "1"
84	İlgili alanıma yakın olması = "1"	Arkadaş tavsiyesi = "1"
84	öğretim üyesi = "1"	Yüksek geçme oranına sahip olması = "1"
84	Yüksek geçme oranına sahip olması = "1"	Arkadaş tavsiyesi = "1"
79	öğretim üyesi = "1"	Yüksek not alma ihtimalinin olması = "1"
77	Güncel içeriğe sahip olması = "1"	İlgili çekici içeriğe sahip olması = "1"
75	Yüksek not alma ihtimalinin olması = "1"	İlgili çekici içeriğe sahip olması = "1"

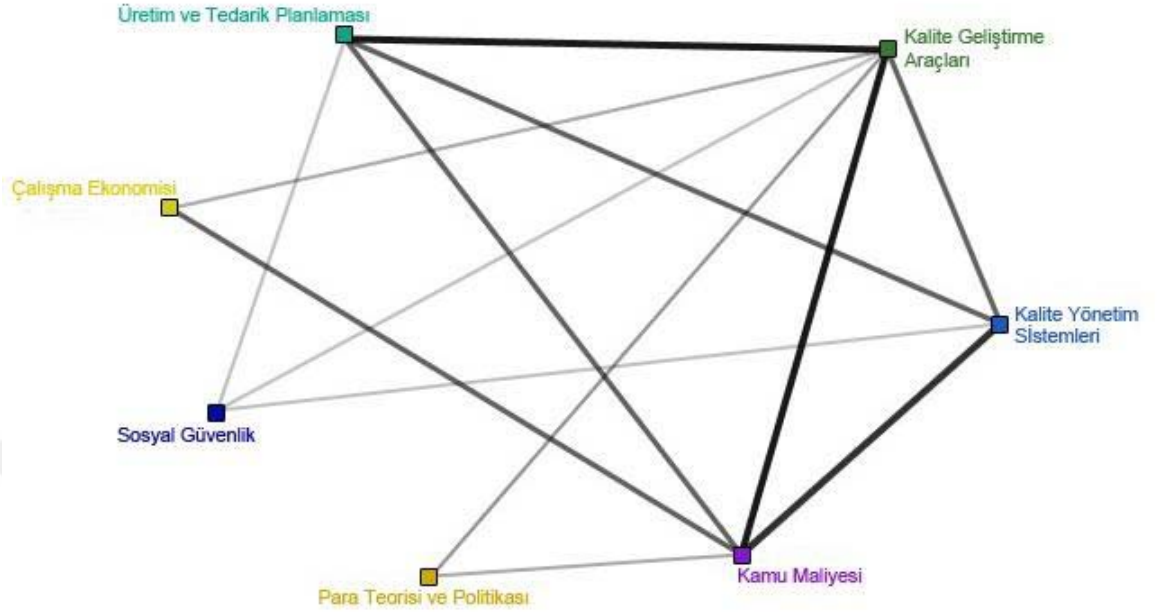
Şekil 28’ de ders seçim sebepleri ile ilgili güçlü birliktelikler görülmektedir. En güçlü birliktelik 173 bağlantı ile dersin ilgi alanına yakın olması ve ilgi çekici içeriğe sahip olmasıdır.

İkinci olarak derslerin aralarındaki ilişkileri görmek için 151 kaydın yer aldığı 1. veri seti için ağ grafiği kullanılmıştır. Şekil 29’ da çizgi değeri 0 iken ilişkinin en karmaşık hali gösterilmiş, Şekil 30’ da ise aynı şekilde çizgi seviyesi artırılarak ilişkiler daha görünür hale getirilmiştir.

Şekil 29: Derslere İlişkin Çizgi Seviyesi 0 İken Elde Edilen Ağ Grafiği



Şekil 30: Derslere İlişkin Çizgi Seviyesi 30 İken Elde Edilen Ağ Grafığı



Şekil 31: Derslere İlişkin En Güçlü Birliktelikler

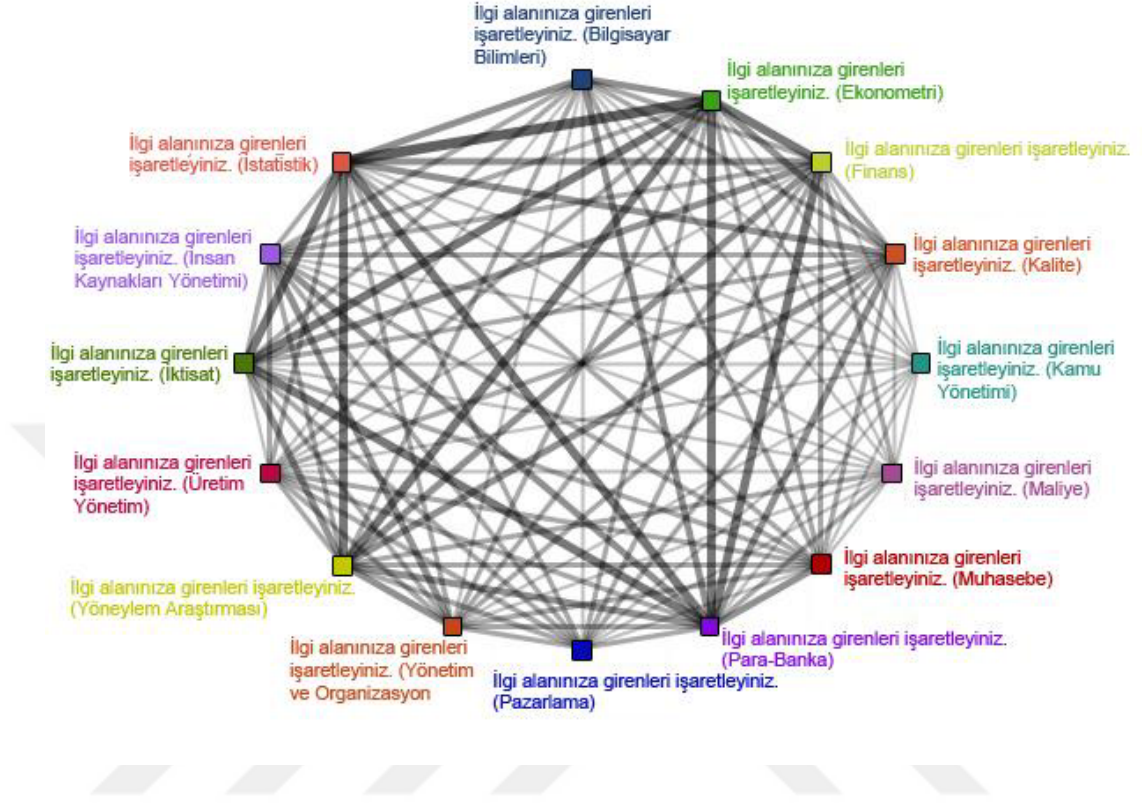
Strong Links

Links	Field 1	Field 2
45	Kalite Geliştirme Araçları = "1"	Üretim Ve Tedarik Planlaması = "1"
44	Kamu Maliyesi = "1"	Kalite Geliştirme Araçları = "1"
42	Kamu Maliyesi = "1"	Kalite Yönetim Sistemleri = "1"
39	Kamu Maliyesi = "1"	Çalışma Ekonomisi = "1"
39	Kamu Maliyesi = "1"	Üretim Ve Tedarik Planlaması = "1"
39	Kalite Geliştirme Araçları = "1"	Kalite Yönetim Sistemleri = "1"
38	Üretim Ve Tedarik Planlaması = "1"	Kalite Yönetim Sistemleri = "1"

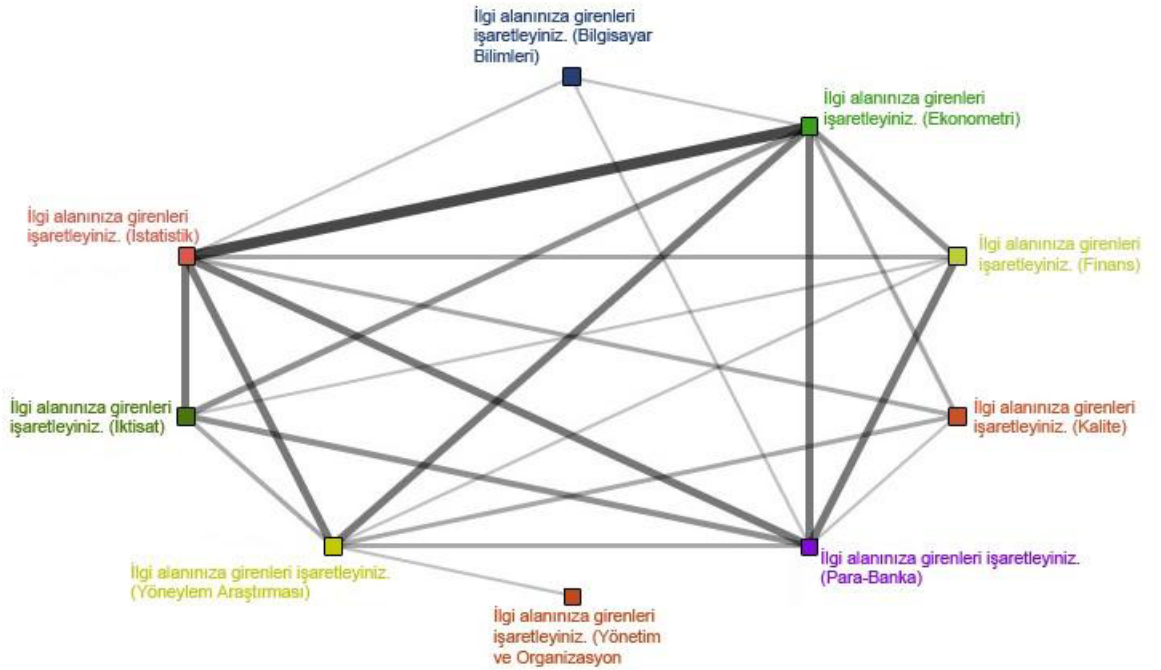
Şekil 31’ de seçimlik derslerle ilgili güçlü birliktelikler verilmiştir. En güçlü birliktelik 45 bağlantı ile Kalite Geliştirme Araçları ve Üretim Ve Tedarik Planlaması derslerine aittir ve Şekil 30’ da çizgi kalınlığından güçlü birliktelikler açıkça görülmektedir.

Üçüncü ve son olarak ise ilgi alanlarının aralarındaki ilişkileri görmek 151 kaydın yer aldığı 1. veri seti için ağ grafığı kullanılmıştır. Şekil 32’ de çizgi değeri 0 iken ilişkinin en karmaşık hali gösterilmiş, Şekil 33’ te ise aynı şekilde çizgi seviyesi artırılarak ilişkiler daha görünür hale getirilmiştir.

Şekil 32: İlgili Alanlarına İlişkin Çizgi Seviyesi 0 İken Elde Edilen Ağ Grafiği



Şekil 33: İlgili Alanlarına İlişkin Çizgi Seviyesi 20 İken Elde Edilen Ağ Grafiği



Şekil 34:İlgi Alanlarına İlişkin En Güçlü Birliktelikler

Links	Field 1	Field 2
50	İlgi alanınıza girenleri işaretleyiniz_Ekonometri = "1"	İlgi alanınıza girenleri işaretleyiniz_İstatistik = "1"
38	İlgi alanınıza girenleri işaretleyiniz_İstatistik = "1"	İlgi alanınıza girenleri işaretleyiniz_Para- Banka = "1"
37	İlgi alanınıza girenleri işaretleyiniz_Yöneylem Araştırması = "1"	İlgi alanınıza girenleri işaretleyiniz_İstatistik = "1"
37	İlgi alanınıza girenleri işaretleyiniz_Para- Banka = "1"	İlgi alanınıza girenleri işaretleyiniz_Finans = "1"
36	İlgi alanınıza girenleri işaretleyiniz_İstatistik = "1"	İlgi alanınıza girenleri işaretleyiniz_İktisat = "1"

Şekil 34' te ilgi alanları ile ilgili güçlü birliktelikler verilmiştir. En güçlü birliktelik 50 bağlantı ile Ekonometri ve İstatistik alanlarına aittir ve Şekil 33' te bu açıkça görülmektedir.



SONUÇ

Veri madenciliğinin amacı, çok büyük ve karmaşık veri kümelerindeki gizli veri desenlerini ortaya çıkarmak ve anlamlı sonuçlar elde etmektir. Yapılan çalışmada birliktelik kuralı çıkarım algoritmalarından Apriori Algoritması kullanılarak üniversite öğrencilerinin ders seçimlerindeki genel eğilimlerinin belirlenmesi amaçlanmıştır. Oluşturulan anket içerik ve kapsam açısından dikkate alındığında elde edilen veri sayısı ve çeşitliliği veri madenciliği amacıyla örtüşmekte ve bu alanda yararlı sonuçların elde edilmesine yatkın nitelikte olduğu görülmektedir.

İncelenen literatür çalışması da dikkate alındığında, yapılan çalışmanın anket süreci ve verilerin ele alış biçimi oldukça kapsamlı ve detaylıdır. Aynı zamanda, değişken sayısının ve veri sayısının fazla olması bu durumu açıkça göstermektedir. Yapılan çalışmada farklı model varyasyonları incelenmiş ve birçok değişkene ait bağlantı analizleri gerçekleştirilmiştir. Seçimlik derslere ilişkin tercih nedenlerinin ve diğer derslerle ilişkilerinin belirlenmesi de aynı zamanda kapsamlı bir şekilde ele alınmıştır.

Üniversite eğitiminde öğrenciler, öğrenim gördükleri programlarda müfredat gereği seçmeli dersler almaktadırlar. Uygulamada üniversite öğrencilerinin ders seçimlerine etki eden faktörlerin elde edilmesi amacıyla ders bazında değerlendirme yapılmasını sağlayacak kapsamlı bir anket formu hazırlanmıştır. Hazırlanan anket formu Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü'nde öğrenimine devam etmekte olan 151 öğrenciye uygulanmış ve 33 dersle ilgili 862 değerlendirme elde edilmiştir. Elde edilen veriler üzerinde SPSS Clementine programında gerekli ön işleme teknikleri uygulanmış, veriler iki ayrı veri seti şeklinde kullanılarak farklı destek, güven değerleri ve farklı değişkenler için çalıştırılmış ve birçok kural elde edilmiştir. Kurallar incelenmiş ve bir kısmı tablolar aracılığıyla uygulama kısmında sunulmuştur. Daha sonra ders seçim sebepleri, derslerle ilgili ilişkiler ve ilgi alanlarının, görselleştirilmesi ve ikili bağlantıların elde edilmesi amacıyla SPSS Clementine programındaki ağ grafiklerinden yararlanılmış ve birliktelik diyagramları farklı derecelere sunulmuştur.

Yapılan analizler kapsamında, genel olarak bakıldığında öğrenciler ders seçimlerinde ilgi alanları ve yeterlilikleri doğrultusunda karar vermektedirler. Diğer bir taraftan ders yükü ağır olan ve bölümü benimsememiş öğrencilerin seçimlerinde dersin kolay olması, geçme oranının yüksek olması, ödev- proje içermemesi gibi kriterlerin önemli olduğu sonucu elde edilmiştir.

Derslerin birlikteliklerine bakıldığında ise birbirine yakın alanlardaki derslerin birlikte seçilme eğiliminde oldukları gözlemlenmiştir. Bunun yanında çoğu öğrencinin ders seçiminde ilk etapta yüzeysel kararlar verdiği gözlemlenmiş, derslerin içerik ve işleyişleriyle ilgili bir ön bilgilendirmenin yararlı olabileceği tespit edilmiştir. Öğrencilere yöneltilen dersi tekrar seçip seçmeyeceklerine yönelik sorudan ise ilginç sonuçlar elde edilmiştir. Öğrencilerden bazıları bazı dersler için, dersten beklediklerinden düşük not almalarına rağmen dersi tekrar seçeceklerini belirtmişlerdir. Bu da öğrenciler için dersi geçmekten çok, edindikleri kazanımların onlar için önemli olduğu gerçeğini ortaya koymuştur. Bir diğer çıkarımımız ise, ders seçimlerinde öğretim üyelerinin de önemli derecede etkili olduğudur. Ayrıca, bölümü benimsememiş, aidiyet hissetmeyen öğrencilerin öğretim üyeleriyle ilişkilerinin zayıf olduğu sonucu elde edilmiştir.

Ankette yer alan ve elde edilen bilgiler sonucunda Apriori Algoritması kullanılarak 5 adet model geliştirilmiş ve birliktelik diyagramları ile birliktelik analizleri gerçekleştirilmiştir. Eldeki verilerin çok yönlülüğü ve zenginliği sayesinde diğer Veri Madenciliği Teknikleri de kullanılarak farklı modeller de geliştirilebilir. Aynı zamanda analiz gelecek dönemler için de tekrarlanırsa elde edilen sonuçların karşılaştırılması ile çok daha kapsamlı yorumlaryapılabilir ve önemli çıkarsamalar sağlanabilir.

KAYNAKÇA

Agrawal, R., Imieliński, T., ve Swami, A. (1993). Mining Association Rules Between Sets Of Items In Large Databases. *ACM SIGMOD Record*. 22(2): 207-216.

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. ve Verkamo, A. I. (1996). Fast Discovery of Association Rules. *Advances In Knowledge Discovery And Data Mining*. 12(1): 307-328.

Agrawal, R., ve Srikant, R. (1994). Fast Algorithms For Mining Association Rules. *Proc. 20th Int. Conf. Very large data bases, VLDB*. 1215: 487-499.

Akbulut, S. (2006). *Veri Madenciliği Teknikleri İle Bir Kozmetik Markanın Ayrılan Müşteri Analizi Ve Müşteri Segmentasyonu*. (Yüksek Lisans Tezi). Ankara: GÜ Fen Bilimleri Enstitüsü.

Alzand, H.R.A. ve Karacan, H. (2014). Bölümleyici Kümeleme Algoritmalarının Farklı Veri Yoğunluklarında Karşılaştırılması. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi*. 30(1): 1-7.

Ay, D. (2009). *Veri Madenciliği ve Apriori Algoritması İle Süpermarket Analizi*. (Yüksek Lisans Tezi). Sakarya: Sakarya Üniversitesi Fen Bilimleri Enstitüsü.

Aydın S. (2007). *Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama*. (Doktora Tezi). Eskişehir: Anadolu Üniversitesi.

Ayık, Y. Z., Özdemir, A., ve Yavuz, U. (2007). Lise Türü Ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2).

Bajpai, S., Jain, K. ve Jain, N. (2011). Artificial Neural Networks. *International Journal of Soft Computing and Engineering (IJSCE)*. 1: 28- 31.

Bırtıl, F. S. (2011). *Kız Meslek Lisesi Öğrencilerinin Akademik Başarısızlık Nedenlerinin Veri Madenciliği Teknikleri İle Analizi*. (Yüksek Lisans Tezi). Afyonkarahisar: Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü.

Borkar, S. ve Rajeswari, K. (2013). Predicting Students Academic Performance Using Education Data Mining. *International Journal of Computer Science and Mobile Computing*. 2(7): 273-279.

Buchner, A. G., Anand, S. S., ve Hughes J. G. (1997). Data Mining in Manufacturing Environments: Goals, Techniques and Applications. *Studies in Informatics and Control*. 6(10): 319-328.

Budak E. Ç. (2013). Üniversite Öğrencileri İçin Bilgisayar Okuryazarlığını Etkileyen Faktörlerin Etkisinin Veri Madenciliği ile Analizi. *AJIT-e: Online Academic Journal of Information Technology*. 4(11): 57-70.

Cavique, L. (2007). A Scalable Algorithm For The Market Basket Analysis. *Journal of Retailing and Consumer Services*. 14(6): 400-407.

Demir, A. (1996). Üniversitedeki Seçmeli Ders Uygulamasının Öğrenciler ve Öğretim Üyelerinde Değerlendirilmesi. *Türk Psikolojik Danışma ve Rehberlik Dergisi*. 2(7): 24-31.

Demir, A., ve Ok, A. (1996). Orta Doğu Teknik Üniversitesindeki Öğretim Üye Ve Öğrencilerinin Seçmeli Dersler Hakkındaki Görüşleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*. 12: 121-125.

Demirel, B. (2010). *Veri Madenciliğinde Chaid Algoritmasının Sosyal Güvenlik Kurumu Veri Tabanına Uygulanması*. (Yüksek Lisans Tezi). Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

Demirel, B. (2010). *Veri Madenciliğinde Chaid Algoritmasının Sosyal Güvenlik Kurumu Veri Tabanına Uygulanması*. (Yüksek Lisans Tezi). Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

Durmuş, M. S. ve İplikçi, S. (2004). Veri Kümeleme Algoritmalarının Performansları Üzerine Karşılaştırmalı Bir Çalışma, *Akademik Bilişim'07, IX. Akademik Bilişim Konferansı Bildirileri* (ss. 393- 400), 31 Ocak - 2 Şubat 2007.

Durucasu, H., Aşan, Z., ve Er, F. (2006). Öğrencilerin Yaz Okulu Hakkındaki Görüşleri İçin Kümeleme Analizi. *Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi*. 7(1): 97-101.

Dündar, S. (2008). Ders Seçiminde Analitik Hiyerarşi Proses Uygulaması. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*.13(2): 217-226.

Düzgünoğlu, S. (2006). *Veri Ambarı ve OLAP Teknolojilerinden Yararlanılarak Karar Destek Amaçlı Raporlama Aracı Gerçekleştirimi*. (Yüksek Lisans Tezi). Ankara: Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.

Erdoğan, Ş. Z., ve Timor, M. (2005). A Data Mining Application in a Student Database. *Journal of aeronautics and space Technologies*. 2(2): 53-57.

Erpolat, S. (2012). Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması. *AÜ Sosyal Bilimler Dergisi*. 12(1): 151-166.

Fayyad, U. M., Piatetsky-Shapiro, G., ve Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *KDD*. 96: 82-88.

Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Verlag: Springer Yayınevi.

Ganesh, S. (2002). Data Mining: Should It Be Included In The ‘Statistics’ Curriculum?. *The Sixt International Conference on Teaching Statistics*. Cape Town. South Africa, 7–12 July.

Giudici, P. (2003). *Applied Data Mining: Statistical Methods For Business And Industry*. Chichester: John Wiley & Sons Yayınevi.

Gülce, A.C. (2010). *Veri Madenciliğinde Apriori Algoritması ve Apriori Algoritmasının Farklı Veri Kümelerinde Uygulanması*. (Yüksek Lisans Tezi). Edirne: Trakya Üniversitesi Fen Bilimleri Enstitüsü.

Güngör, E., Yalçın, N. ve Yurtay, N. (2013). Apriori Algoritması ile Teknik Seçmeli Ders Seçim Analizi, *UZEM 2013 Ulusal Uzaktan Eğitim ve Teknolojileri Sempozyum Kitabı*. 01-03.

Gürgen, G. (2008). Birliktelik Kuralları ile Sepet analizi ve Uygulaması. (Yüksek Lisans Tezi) İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü.

Gürsoy, U. T. Ş. (2011). *Uygulamalı Veri Madenciliği Sektörel Analizler*. Ankara: Pegem Akademi Yayıncılık.

Ha, S. H. ve Park, S. C. (1998). Application Of Data Mining Tools To Hotel Data Mart On The Intranet For Database Marketing. *Expert Systems With Applications*. 15(1): 1-31.

Han, J. ve Kamber, M. (2006). *Data Mining Concepts and Techniques*. San Francisco: Elsevier.

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, And Algorithms*. Hoboken: John Wiley & Sons Yayınevi.

Karabatak, M., ve İnce, M. C. (2004). Apriori Algoritması ile Öğrenci Başarısı Analizi. *Eleco " Elektrik-Elektronik ve Bilgisayar Mühendisleri Sempozyumu*, Bursa.

Köktürk, M. K. ve Dirsehan, T. (2012). *Veri Madenciliği İle Pazarlama Etkileşimi*. Ankara: Nobel Akademik Yayıncılık.

Larose, D. T. ve Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken: John Wiley & Sons Yayınevi.

Liu, D.R., Shih, Y.Y. (2005). Integrating AHP And Data Mining For Product Recommendation Based On Customer Lifetime Value. *Information & Management*. 42(3): 387-400.

Maimon, O., ve Rokach, L. (2007). *Data Mining And Knowledge Discovery Handbook*. New York: Springer Science+Business Media Inc.

Mansoul, A. ve Atmani, B. (2014). A Boolean Modeling For Improving The Algorithm Apriori. *Computer Science & Information Technology (CS & IT)*. 105-114.

Oğuzlar, A. (2003). Veri Ön İşleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 2(21): 67-76.

Özek, M. B., Daş, B. ve Akpolat, Z. H. (2010). Zaman Serisi Tahmininde Tip-2 Bulanık Mantık Tabanlı Veri Madenciliği Uygulaması. *E-Journal of New World Sciences Academy*. 5(2): 100-109.

Özkan, Y. (2008). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Yayıncılık Eğitim.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences*. 11(5): 341-356.

Piramuthu, S. (2004). Evaluating Feature Selection Methods For Learning In Data Mining Applications. *European Journal of Operational Research*. 156(2): 483-494.

Silahtaroglu, G. (2008). *Veri Madenciliği (Kavram ve Algoritmalar)*. İstanbul: Papatya Yayınları.

Srikant, R., Vu, Q. ve Agrawal R. (1997). Mining Association Rules With Item Constraints. *KDD.97*: 67-73.

Şentürk, A. (2006). *Veri Madenciliği Kavram ve Teknikler*. Bursa: Ekin Kitabevi.

Tapkan, P., Özbakır, L. ve Baykasoğlu, A. (2011). *Weka İle Veri Madenciliği Süreci Ve Örnek Uygulama* (ss. 247- 262), Endüstri Mühendisliği Yazılımları ve Uygulamaları Kongresi İzmir. 30 Eylül-01/02 Ekim 2011.

Taş, M., Adak, M., F., Yurtay, N. (2013). Öğrencilerin Staj Verileri Üzerine Uygulanan Apriori Algoritması ile Birliktelik Kurallarının Çıkarılması ve Staj Eğiliminin Belirlenmesi'. *Ist International Symposium on Innovative Technologies in Engineering and Science*, Akademik Platform. 3(10): 1086- 1095.

Tezcan, H. ve Gümüş, Y. (2008). Üniversite Öğrencilerinin Seçmeli Ders Tercihlerine Etki Eden Faktörlerin Araştırılması. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*. 28(1): 1-17.

Thabtah, F., Cowling, P. ve Hammoud, S. (2006). Improving Rule Sorting, Predictive Accuracy and Training Time in Associative Classification. *Expert Systems with Applications*. 31: 414-426.

Tiryaki, S. (2015). *Lojistik Alanında Bir Veri Madenciliği Uygulaması*. (Yüksek Lisans Tezi). İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

Toivonen, H. (1996). Sampling Large Databases For Association Rules. *VLDB.* 96: 134-145.

Tsay, Y. J., ve Chang-Chien, Y. W. (2004). An Efficient Cluster And Decomposition Algorithm For Mining Association Rules. *Information Sciences.*160(1): 161-171.

Tufféry, S. (2011). *Data Mining And Statistics For Decision Making.* John Wiley & Sons Yayınevi.

Ünal, Y., Ekim, U. ve Köklü, M. (2011). ‘Üniversite Öğrencilerinin Ortak Zorunlu Derslerdeki Başarılarının K-Means Algoritması ile İncelenmesi’, *e- Journal of New World Sciences Academy*, 6(1): 342-347.

Yıldız, O. (2014). *Makine Öğrenmesi İle Uzaktan Eğitim Öğrencilerinin Performanslarının Değerlendirilmesi.* (Doktora Tezi). İstanbul: İstanbul Üniversitesi Fen Bilimleri Enstitüsü.

Yu, H., Wen, J., Wang, H. ve Jun, L. (2011). An Improved Apriori Algorithm Based On the Boolean Matrix and Hadoop. *Procedia Engineering.* 15: 1827 – 1831.

Ziadeh, I., ve Rusli, D. (2007). Predicting Student Academic Performance: Comparing Artificial Neural Network, Decision Tree And Linear Regression. *Journal of Legal Studies Education.* (2): 225-239.



EKLER

EKLER

EK 1: Arařtırmada Kullanılan Anket Formunun Ekran Görüntüsü

Seçmeli Ders Eğilimlerine İlişkin Desenlerin Veri Madenciliği ile Belirlenmesi



Sevgili arkadaşlar,

Ekonometri Bölümü öğrencilerinin seçmlik ders tercihleri üzerine arařtırma yapmaktayız. Arařtırmada eğitim kalitesini arttırmak için öğrencilerin beklentileri ve tercihlerini tespit etmeyi amaçlamaktayız. Elde edilecek bilgiler doğrultusunda bilimsel gereklilikler ile birlikte öğrenci tercihlerinin dikkate alındığı; ders havuzu, ders içerik ve yöntem güncellemesi yapılması mümkün olacaktır. Arařtırmada sizlerin bilgilerinize ihtiyaç duymaktayız. Sizin için hazırlanmış sorulara özenle cevap vermenizi umuyoruz. Teşekkürler...

Doç. Dr. Mehmet AKSARAYLI

[Devam](#)

EK 1 (Devam): Arařtırmada Kullanılan Anket Formunun Ekran Görüntüsü

0 25 50 75 100

Cinsiyetiniz:

Kadın Erkek

Öğrenim türünüz:

Örgün İkinci öğretim

Bölüme aşağıdaki deęişimlerden faydalanarak geldiyseiniz seçiniz.

Dikey geçiř

Yatay geçiř

Farabi

Üniversite içi yatay geçiř

Çalıřma durumunuz:

Lütfen Seçiniz ▼

Mezuniyet sonrasındaki temel hedefiniz nedir?

Lütfen Seçiniz ▼

Öğreniminizin kaçınıcı dönemindedesiniz?

Kayıt yaptırdığınız en üst dönemi seçiniz.

Lütfen Seçiniz ▼

Mezun olduğunuz lise türü:

Lütfen Seçiniz ▼

Liseden mezun olduğunuz bölüm:

Lütfen Seçiniz ▼

[Surveey.com](https://www.surveey.com) [Devam](#)

EK 1 (Devam): Araştırmada Kullanılan Anket Formunun Ekran Görüntüsü

Tekrar şans verilse hangi bölümde okumak istersiniz?

Ekonometri Maliye Çalışma Ekonomisi ve Endüstri İlişkileri

İşletme İktisat Kamu Yönetimi

Yönetim Bilişim Sistemleri Fen Fakültesi Eğitim Fakültesi

Hukuk Fakültesi Tıp Fakültesi Güzel Sanatlar Fakültesi

Mühendislik Fakültesi Mimarlık Fakültesi Edebiyat Fakültesi

Orman Fakültesi Ziraat Fakültesi Dil Tarih Coğrafya Fakültesi

Denizcilik Fakültesi Turizm Fakültesi İlahiyat Fakültesi

İşletme Fakültesi Hemşirelik Fakültesi Diğer

Ekonometri Bölümü içerisindeki anabilim dallarından ilgi alanlarınız nelerdir? (Birden çok seçenek işaretleyebilirsiniz)

Ekonometri Anabilim Dalı

Yöneylem Araştırması Anabilim Dalı

İstatistik Anabilim Dalı

Aşağıdaki konulardan ilgi alanınıza girenleri işaretleyiniz. (Birden çok seçenek işaretleyebilirsiniz)

Ekonometri Yöneylem Araştırması İstatistik

Para- Banka Finans Pazarlama

Muhasebe Maliye Kamu Yönetimi

Kalite Bilgisayar Bilimleri İnsan Kaynakları Yönetimi

İktisat Yönetim ve Organizasyon Üretim Yönetim

Seçimlik dersleri seçerken en çok neyi referans alıyorsunuz?

Arkadaşlarının tavsiyeleri

Hoca tavsiye ve yönlendirmeleri

Kişisel değerlendirmelerim

Seçimlik dersler hakkında ön bilgi verilmesini ister miydiniz?

Evet Hayır

Aşağıdaki soruları yandaki ölçek doğrultusunda cevaplandırınız.

	Kesinlikle Katılmıyorum	Katılmıyorum	Kararsızım	Katılıyorum	Tamamen Katılıyorum
En kısa sürede mezun olma baskısı hissetmiyorum.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En kısa sürede mezun olmaktansa, donanımlı mezun olmayı tercih ederim.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ekonometri bölümünü isteyerek, bilerek, araştırarak geldim.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Okuduğum bölümü benimsedim ve seviyorum.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Okuduğum bölüme aidiyet hissediyorum.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dokuz Eylül Üniversitesi' ne isteyerek geldim.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Araştırma yapmayı severim.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bilgisayar kullanımına erişimim kolaydır.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Yabancı dil düzeyimin yeterli olduğunu düşünüyorum.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Benim için yeni bir şey öğrenmekten çok dersi geçmek daha önemlidir.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[Surveey.com](#)

EK 1 (Devam): Araştırmada Kullanılan Anket Formunun Ekran Görüntüsü

0 25 50 75 100

.....3. YARIYIL.....

Dönem başındaki not ortalamanız:
(virgül kullanınız)

Dönemde aldığınız toplam ders sayısı:

Dönemde başarılı olduğunuz toplam ders sayısı:

Aldığınız dersi/ dersleri işaretleyiniz.

Yabancı Dilde Okuma Ve Konuşma
 Yönetim Ve Organizasyon
 Borçlar Hukuku
 Kamu Maliyesi
 Ders listede yoksa yazınız.

1.dersi (seçiminize göre listede önce yer alan ders) seçmenizde etkili olan sebepleri işaretleyiniz.

<input type="checkbox"/> Kolay olması	<input type="checkbox"/> Öğretim üyesi	<input type="checkbox"/> Bilgisayar uygulamalı olması
<input type="checkbox"/> İlgili alanıma yakın olması	<input type="checkbox"/> Önemli kazanım sunuyor olması	<input type="checkbox"/> Alan derslerimle bağlantılı olması
<input type="checkbox"/> Alan derslerimden bağımsız olması	<input type="checkbox"/> Yüksek geçme oranına sahip olması	<input type="checkbox"/> Yüksek not alma ihtimalinin olması
<input type="checkbox"/> Sözel ağırlıklı içeriğe sahip olması	<input type="checkbox"/> Sayısal ağırlıklı içeriğe sahip olması	<input type="checkbox"/> Güncel içeriğe sahip olması
<input type="checkbox"/> İlgili çekici içeriğe sahip olması	<input type="checkbox"/> Arkadaş tavsiyesi	<input type="checkbox"/> Ödev- Proje içermesi
<input type="checkbox"/> Ödev- Proje içermemesi	<input type="checkbox"/> Haftalık Programa Uygun olması	<input type="checkbox"/> Devam gerektirmemesi

1. ders için harf notunuz:

1. ders için harf notu beklentiniz:

1.ders ile ilgili başarı notunuz beklentinize göre nasıl oldu?

Beklentimin altında idi.
 Beklediğim düzeyde oldu.
 Beklentimin üzerinde oldu.

Bu dersi tekrar seçerdim.

Evet Hayır

Bu dersi almaları için başkalarına da tavsiye ederim.

Evet Hayır

EK 1 (Devam): Araştırmada Kullanılan Anket Formunun Ekran Görüntüsü

2.dersi (seçiminize göre listede ikinci sırada yer alan ders) seçmenizde etkili olan sebepleri işaretleyiniz.

<input type="checkbox"/> Kolay olması	<input type="checkbox"/> Öğretim üyesi	<input type="checkbox"/> Bilgisayar uygulamalı olması
<input type="checkbox"/> İlgi alanıma yakın olması	<input type="checkbox"/> Önemli kazanım sunuyor olması	<input type="checkbox"/> Alan derslerimle bağlantılı olması
<input type="checkbox"/> Alan derslerimden bağımsız olması	<input type="checkbox"/> Yüksek geçme oranına sahip olması	<input type="checkbox"/> Yüksek not alma ihtimalinin olması
<input type="checkbox"/> Sözel ağırlıklı içeriğe sahip olması	<input type="checkbox"/> Sayısal ağırlıklı içeriğe sahip olması	<input type="checkbox"/> Güncel içeriğe sahip olması
<input type="checkbox"/> İlgi çekici içeriğe sahip olması	<input type="checkbox"/> Arkadaş tavsiyesi	<input type="checkbox"/> Ödev- Proje içermesi
<input type="checkbox"/> Ödev- Proje içermemesi	<input type="checkbox"/> Haftalık Programa Uygun olması	<input type="checkbox"/> Devam gerektirmemesi

2. ders için harf notunuz:

Lütfen Seçiniz ▼

2. ders için harf notu beklentiniz:

Lütfen Seçiniz ▼

2.ders ile ilgili başarı notunuz beklentinize göre nasıl oldu?

Beklentimin altında idi.

Beklediğim düzeyde oldu.

Beklentimin üzerinde oldu.

Bu dersi tekrar seçerdim.

Evet Hayır

Bu dersi almaları için başkalarına da tavsiye ederim.

Evet Hayır

[Survey.com](#)

EK 2: Seçimlik Dersler Listesi

Seçimlik Dersin Adı	Seçimlik Dersin Yarıyılı
Yabancı Dilde Okuma Ve Konuşma	3. yarıyıl
Yönetim Ve Organizasyon	3. yarıyıl
Borçlar Hukuku	3. yarıyıl
Kamu Maliyesi	3. yarıyıl
Çalışma Ekonomisi	4. yarıyıl
Mesleki Yabancı Dil 1	4. yarıyıl
Süreç Yönetimi	4. yarıyıl
Ticaret Hukuku	4. yarıyıl
Sosyal Güvenlik	5. yarıyıl
Kalite Geliştirme Araçları	5. yarıyıl
Mesleki Yabancı Dil 2	5. yarıyıl
Ekonometriye Giriş (Erasmus)	5. yarıyıl
Finans Matematiği	5. yarıyıl
Yatırım Projeleri Analizi	5. yarıyıl
Ekonometri (Erasmus)	6. yarıyıl
İş Hayatı İçin Yabancı Dil	6. yarıyıl
Proje Yönetimi	6. yarıyıl
Para Teorisi Ve Politikası	6. yarıyıl
Üretim Ve Tedarik Planlaması	6. yarıyıl
Kalite Yönetim Sistemleri	7. yarıyıl
Veri Madenciliği	7. yarıyıl
Çok Kriterli Karar Verme	7. yarıyıl
Bilişim Teknolojisi Tabanlı Karar Verme (Erasmus)	7. yarıyıl
Çok Değişkenli İstatistik: İşletme Ve İktisat Uygulamalı (Erasmus)	7. yarıyıl
Para Ve Sermaye Piyasaları	7. yarıyıl
Pazarlama Araştırması Teknikleri	7. yarıyıl
Kesit Verisi Ekonometrisi	7. yarıyıl
Ekonometri Semineri	8. yarıyıl
Coğrafi Bilgi Sistemleri	8. yarıyıl
Yöneylem Araştırma Semineri	8. yarıyıl
Uluslararası İktisat	8. yarıyıl
Türkiye Ekonomisi	8. yarıyıl
Maliye Politikası	8. yarıyıl