

**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE
MÜHENDİSLİK FAKÜLTESİ UZAKTAN EĞİTİM
BÖLÜMLERİNİN ANALİZİ:
KARABÜK ÜNİVERSİTESİ ÖRNEĞİ**

**2014
YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ**

Abdulkadir TAŞDELEN

**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE MÜHENDİSLİK FAKÜLTESİ
UZAKTAN EĞİTİM BÖLÜMLERİNİN ANALİZİ:
KARABÜK ÜNİVERSİTESİ ÖRNEĞİ**

Abdulkadir TAŞDELEN

**Karabük Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalında
Yüksek Lisans Tezi
Olarak Hazırlanmıştır**

**KARABÜK
Eylül 2014**

Abdulkadir TAŞDELEN tarafından hazırlanan "VERİ MADENCİLİĞİ YÖNTEMLERİ İLE MÜHENDİSLİK FAKÜLTESİ UZAKTAN EĞİTİM BÖLÜMLERİNİN ANALİZ EDİLMESİ: KARABÜK ÜNİVERSİTESİ ÖRNEĞİ" başlıklı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

Yrd. Doç. Dr. Bekir S. GÜR

Tez Danışmanı, Bilgisayar Mühendisliği



Bu çalışma, jürimiz tarafından oy birliği / oy çokluğu ile Bilgisayar Mühendisliği Anabilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir. 09.09/2014

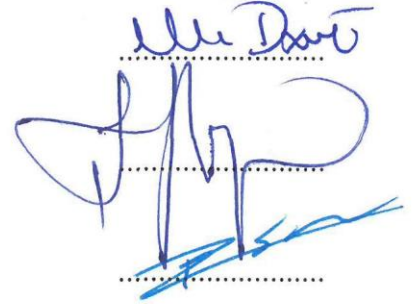
Unvanı, Adı SOYADI (Kurumu)

Başkan: Yrd. Doç. Dr. Murat ÖZOĞLU (YBÜ)

Üye : Yrd. Doç. Dr. Şafak BAYIR (KBÜ)

Üye : Yrd. Doç. Dr. Bekir S. GÜR (YBÜ)

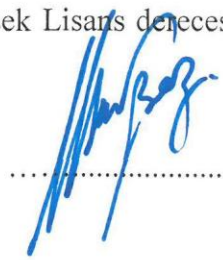
İmza



Tarih/...../2014

KBÜ Fen Bilimleri Enstitüsü Yönetim Kurulu, bu tez ile, Yüksek Lisans derecesini onamıştır.

Prof. Dr. Mustafa BOZ
Fen Bilimleri Enstitüsü Müdürü



“Bu tezdeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.”

Abdulkadir TAŞDELEN

ÖZET

Yüksek Lisans Tezi

**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE MÜHENDİSLİK FAKÜLTESİ
UZAKTAN EĞİTİM BÖLÜMLERİNİN ANALİZİ:
KARABÜK ÜNİVERSİTESİ ÖRNEĞİ**

Abdulkadir TAŞDELEN

**Karabük Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı**

Tez Danışmanı:

Yrd. Doç. Dr. Bekir S. GÜR

Eylül 2014, 83 Sayfa

Teknolojinin hızla gelişmesi ve daha yaygın kullanılması ile başta kişisel veriler olmak üzere birçok veri dijital ortamlarda saklanabilmektedir. Zamanla saklanan bu verilerden anlamlı ilişkiler ve örüntüler çıkartılmaya başlanmıştır. Veriden bilgi keşfini sağlayan birçok veri madenciliği yöntem ve tekniği geliştirilmiş ve farklı alanlarda kullanılmıştır. Veri madenciliği yöntemlerinin uygulanabileceği alanlardan bir tanesi yükseköğretim alanıdır. Yükseköğretim kurumlarında yer alan otomasyon sistemleri öğrencilere ait demografik bilgiler başta olmak üzere akademik başarıları, devam durumları, ders seçimleri vb. gibi bilgileri bir otomasyon sisteminde tutmaktadır. Ayrıca uzaktan eğitim veren kurumlarda bu bilgilere ilave olarak öğrencilerin öğrenim yönetim sistemindeki ders içi ve ders dışı etkinlikleri de kayıt altında tutulmaktadır. Tüm bu kayıtlardan öğrencilere ait çeşitli tanımlamalar ve tahminler yapılabilmektedir.

Ülkemizde son yıllarda üniversite sayısı günden güne artmaktadır. Bu üniversitelerde yüzyüze eğitim veren programların yanı sıra uzaktan eğitim veren farklı önlisans ve lisans programları da açılmaktadır. Uzaktan eğitim programlarına sahip üniversitelerden bir tanesi de Karabük Üniversitesidir. Aynı zamanda bu üniversite uzaktan eğitim ile mühendislik eğitimi vermeye başlayan iki üniversiteden biri olma özelliğini taşımaktadır.

Bu çalışmada Karabük Üniversitesi uzaktan eğitim mühendislik bölümleri veri madenciliği yöntemleri ile analiz edilmiştir. Bu yöntemlerden demografik özelliklerin analizi tanımlayıcı istatistiksel yöntemlerden yararlanılarak, üniversite giriş puanları ile not ortalamalarının analizleri parametrik ve parametrik olmayan testlerden yararlanılarak yapılmıştır. Araştırma sonucunda bölümlerin kendi içerisinde yapılan analizler birinci öğretim, ikinci öğretim ve uzaktan eğitim programlarına ait üniversite giriş puanları arasında anlamlı fark olduğu sonucuna varılmıştır. Ayrıca farklı iki bölüme ait uzaktan eğitim programlarına üniversite giriş puanları arasında da istatistiksel olarak anlamlı bir fark olduğu sonucuna varılmıştır. Üniversite giriş puanının ağırlıklı genel not ortalamasını pozitif yönde fakat düşük oranda yordadığı sonucuna varılmıştır. Bu da üniversite giriş puanının tek başına akademik başarıyı yordama da yeterli olmadığını göstermektedir. Ayrıca not ortalamaları analiz edildiğinde her bölümün birbirinden farklılaştığı sonucuna varılmıştır. En son olarak lojistik regresyon analizi ile geliştirilen model farklı değerler verilerek örneklendirilmiştir.

Anahtar Sözcükler : Veri madenciliği, mühendislik eğitimi, uzaktan eğitim.

Bilim Kodu : 902.1.063

ABSTRACT

M. Sc. Thesis

AN ANALYSIS OF DISTANCE EDUCATION PROGRAMS OF ENGINEERING FACULTY BY DATA MINING METHODS: KARABUK UNIVERSITY SAMPLE

Abdulkadir TAŞDELEN

**Karabük University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering**

Thesis Advisor:

Assist. Prof. Dr. Bekir S. GÜR

September 2014, 83 pages

As a result of new developments and common use of technology, lots of data, specifically personal data, can be stored on digital platforms. By the time of progress, meaningful relations and patterns has been deduced from these stored data. Lots of data mining methods and techniques discovering knowledge from data have been developed and used in different fields. One of the fields of the data mining methods used is higher education. All information about students such as academic success, course selection, drop out, and particularly demographic information are stored in the student information system in higher education institutions. Additionally, students' course activities both internal and external are stored in the learning management systems in distance education institutions. Some predictions and descriptions can be made from these records.

The number of universities in our country is increasing in recent years. These universities may have both formal education and distance education courses. Karabuk University is one of the universities which has distance education courses. Also, this university is one of the two universities which has started our country's first distance engineering courses.

In this study, Karabuk University distance education engineering courses are analyzed by the data mining methods. Demographical features have been analyzed by descriptive statistic methods, university entrance examination score and cumulative grade point average have been analyzed by parametric and nonparametric test methods. In the end of the study, significant difference has been found among all programs both daytime education, evening education and distance education in terms of university entrance examination score. Also, it has found that there is a significant difference between two distance education programs regarding university entrance examination score. It has found that there is a significant positive but poor correlation between university entrance examination score and cumulative grade point average. This shows that university entrance examination score is not enough to predict academic success. Additionally, when grade points have analyzed, it has found that each programs have been differentiated. Finally, the model being developed by logistic regression analyze, has been illustrated by giving various values.

Keywords : Data mining, engineering education, distance education.

Science Code : 902.1.063

TEŞEKKÜRLER

Bu tez çalışmasının her aşamasında ilgi ve desteğini hiçbir zaman esirgemeyen, bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temellerde oluşturmamı sağlayan, hocam Sayın Yrd. Doç. Dr. Bekir S. GÜR'e teşekkürlerimi sunarım.

Tez jürimde yer alan değerli hocalarım Sayın Yrd. Doç. Dr. Şafak BAYIR ve Yrd. Doç. Dr. Murat ÖZOĞLU'na teşekkürlerimi sunarım.

Çalışmamda kullandığım verilerin temini ve kullanımı ile ilgili göstermiş olduğu kolaylıklar ve vermiş olduğu izinden dolayı Mühendislik Fakültesi Dekanı Sayın Prof. Dr. Erol ARCAKLIOĞLU'na teşekkür ederim.

Çalışmam sırasında yardımlarını esirgemeyen, Karabük Üniversitesi Bilgi İşlem Daire Başkanı Recep KILINÇ, Arş. Gör. Oğuzhan MENEMENCİOĞLU ve Arş. Gör. Ferhat ATASOY'a teşekkür ederim.

Veri işleme sürecinde her türlü bilgisini paylaşıp hiçbir zaman yardımını esirgemeyen değerli arkadaşım Türker TOKU'ya teşekkür ederim.

Veri analizi aşamasında kıymetli vakitlerini ayırıp çalışmamın tamamlanmasında yardımcı olan Murat ÖZTÜRK ve Muharrem KAPLAN hocalarıma teşekkür ederim.

Son olarak, sevgili eşim Emine'ye ve değerli aileme her türlü konuda bana yardımcı oldukları ve tezimi bitirme hususundaki destekleri için tüm kalbimle teşekkür ederim.

İÇİNDEKİLER

Sayfa

KABUL.....	ii
ÖZET	iv
ABSTRACT	vi
TEŞEKKÜR	viii
İÇİNDEKİLER.....	ix
ŞEKİLLER DİZİNİ	xii
ÇİZELGELER DİZİNİ	xiii
SİMGELER VE KISALTMALAR DİZİNİ.....	xv
BÖLÜM 1.	1
GİRİŞ	1
1.1. AMAÇ	1
1.2. ÖNEM.....	2
1.3. YÖNTEM.....	3
1.4. SINIRLILIKLAR	8
1.5. TANIMLAR.....	8
1.6. ETİK	9
BÖLÜM 2.	10
VERİ MADENCİLİĞİ.....	10
2.1. GİRİŞ.....	10
2.1.1. Veri.....	10
2.1.2. Verilerin Saklanması	11
2.1.2.1. Geleneksel Dosyalama Sistemleri	11
2.1.2.2. Veri Tabanı Yönetim Sistemleri.....	12
2.1.3. Veri Ambarı	13
2.2. VERİ MADENCİLİĞİ NEDİR.....	15
2.3. VERİ MADENCİLİĞİ UYGULAMA ALANLARI	16
2.4. VERİ MADENCİLİĞİ SÜRECİ.....	17

	<u>Sayfa</u>
2.5. VERİ MADENCİLİĞİ MODELLERİ	19
2.5.1. Tanımlayıcı Modeller	19
2.5.1.1. İlişki Analizi	19
2.5.1.2. Kümeleme Analizi	25
2.5.2. Tahmin Edici Modeller	30
2.5.2.1. Sınıflandırma	30
2.5.2.2. İstatistiksel Tahmin Modelleri.....	35
BÖLÜM 3.	40
EĞİTİMDE VERİ MADENCİLİĞİ UYGULAMLARI.....	40
3.1. EĞİTİM VE VERİ MADENCİLİĞİ	40
3.1.1. Müfredat Düzenlemeleri.....	41
3.1.2. Derse Kayıt ve Devam	42
3.1.3. Akademik Başarı.....	43
3.1.4. Öğrenci Davranışları ve Performansları.....	45
3.2. EĞİTSEL VERİ MADENCİLİĞİ YAZILIMLARI.....	47
3.3. UZAKTAN EĞİTİM İLE MÜHENDİSLİK EĞİTİM.....	48
3.4. KARABÜK ÜNİVERSİTESİ UZAKTAN EĞİTİM MODELİ	49
BÖLÜM 4.	54
BULGULAR	54
4.1. VERİ MADENCİLİĞİ SÜRECİ.....	54
4.1.1. Verilerin Toplanması ve Hazırlanması	54
4.1.1.1. KBÜ Öğrenci Otomasyonu Veri Tabanı.....	54
4.1.1.2. KBUZEM Öğrenim Yönetim Sistemi Veri Tabanı	59
4.1.2. Modelleme	61
BÖLÜM 5.	72
SONUÇLAR.....	72
KAYNAKLAR.....	74
EK AÇIKLAMALAR A. ARAŞTIRMA İZİN BELGESİ	80

	<u>Sayfa</u>
ÖZGEÇMİŞ.....	83

ŞEKİLLER DİZİNİ

Sayfa

Şekil 2.1. Veritabanından bilgi keşfi aşamaları	15
Şekil 2.2. Veri madenciliği süreci	16
Şekil 2.3. CRIPS-DM veri madenciliği süreci.....	18
Şekil 2.4. AprioriTid algoritması geniş küme ve aday küme aşamaları.....	24
Şekil 2.5. Örnek karar ağacı yapısı.....	31
Şekil 2.6. Yapay sinir ağı yapısı.....	34
Şekil 3.1. Eğitim sistemlerinde veri madenciliği uygulama döngüsü	40
Şekil 3.2. KBUZEM web sayfası	49
Şekil 4.1. Örnek bir derse ait etkinlik kayıtları	60
Şekil 4.2. Ders kayıtları tablosu	60
Şekil 4.3. Ders katılım oranı tablosu.....	61

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 1.1. Sınıflara göre öğrenci dağılımları	3
Çizelge 1.2. Üniversite giriş puanlarına ait betimsel istatistik	5
Çizelge 2.1. OLTP ve OLAP karşılaştırması	13
Çizelge 2.2. Apriori algoritması notasyonu	21
Çizelge 3.1. Eğitsel veri madenciliği, istatistik ve görselleştirme araçları	47
Çizelge 3.2. KBÜ Uzaktan eğitim önlisans programları	49
Çizelge 3.3. KBÜ uzaktan eğitim lisans programları	50
Çizelge 3.4. UE mühendislik programları ve öğrenci alım yılları	51
Çizelge 3.5. KBÜ mühendislik bölümleri ve öğrenci alım yılları.....	52
Çizelge 4.1. Öğrenci bilgilerine ait değişkenler	55
Çizelge 4.2. “bölüm” değişkeni veri dönüşümü	55
Çizelge 4.3. “giriş_yılı” değişkeni veri dönüşümü.....	56
Çizelge 4.4. “durum” değişkeni veri dönüşümü.....	56
Çizelge 4.5. “cinsiyet” değişkeni veri dönüşümü.....	56
Çizelge 4.6. “ayrılış_nedeni” değişkeni veri dönüşümü	57
Çizelge 4.7. Ders bilgilerine ait değişkenler	57
Çizelge 4.8. Derslere ait “bölüm” değişkeni veri dönüşümü	58
Çizelge 4.9. Derslere ait “ders_yılı” değişkeni veri dönüşümü	58
Çizelge 4.10. Derslere ait “ders_dönemi” değişkeni veri dönüşümü	58
Çizelge 4.11. Derslere ait “ders_notu_harf” değişkeni veri dönüşümü.....	59
Çizelge 4.12. Derslere ait “başarı_durumu” değişkeni veri dönüşümü.....	59
Çizelge 4.13. Giriş yıllarına göre öğrenci sayıları ve bölümler	61
Çizelge 4.14. Bilgisayar mühendisliğine ait giriş puanları analizi	62
Çizelge 4.15. Endüstri mühendisliğine ait giriş puanları analizi.....	63
Çizelge 4.16. Bilgisayar ve endüstri mühendisliği UE giriş puanları analizi	63
Çizelge 4.17. Bilgisayar mühendisliği not ortalamaları analiz sonuçları	64
Çizelge 4.18. Endüstri mühendisliğine ait giriş puanları analizi.....	65
Çizelge 4.19. Bilgisayar ve endüstri mühendisliği UE not ortalamaları analizi	65
Çizelge 4.20. Analiz edilen olaylar	66

	<u>Sayfa</u>
Çizelge 4.21. Hosmer ve Lemeshow testi.....	67
Çizelge 4.22. Başlangıç sınıflandırma tablosu	67
Çizelge 4.23. Sınıflandırma tablosu.....	68
Çizelge 4.24. Omnibus testi	68
Çizelge 4.25. Model özeti	69
Çizelge 4.26. Eşitlikteki değişkenlerin analizi	69

SİMGELER VE KISALTMALAR DİZİNİ

SİMGELER

α : Alfa

β : Beta

d : Uzaklık

D : Çap

R : Yarıçap

r : Korelasyon katsayısı

r^2 : Belirleme katsayısı

ε : Epsilon

λ : Lamda

η^2 : eta kare

χ^2 : ki kare

n : Örneklem sayısı

p : probability

KISALTMALAR

ABNO : Akademik Başarı Not Ortalaması

ACT : American College Testing (Amerikan Kolej Testi)

AID : Automatic Interaction Detection (Otomatik Etkileşim Belirleme)

AGNO : Ağırlıklı Genel Not Ortalaması

C&RT : Classification and Regression Trees (Sınıflandırma ve Regresyon Araçları)

CHAID : Chi-Squared Automatic Interaction Detector (Ki-kare Otomatik Etkileşim Belirleme)

ColAT : Collaborative Analysis Tool (İşbirlikçi Analiz Aracı)

CRISP-DM : Cross-Industry Standart Process For Data Mining (Veri Madenciliği Çapraz Endüstri Standart Süreci)

ID3	: Induction of Decision Trees (Karar Ağaçları İndüksiyonu)
İTE	: İnternet Tabanlı Eğitim
İVYTS	: İlişkisel Veri Tabanı Yönetim Sistemleri
KBUZEM	: Karabük Üniversitesi Uzaktan Eğitim Uygulama ve Araştırma Merkezi
KBÜ	: Karabük Üniversitesi
LRA	: Lojistik Regresyon Analizi
MYO	: Meslek Yüksekokulu
OBP	: Ortaöğretim Başarı Puanı
OLAP	: On-Line Analytical Processing (Çevrim-içi Analitik İşleme)
OLTP	: On-Line Transaction Processing (Çevrim-içi Etkileşimsel İşleme)
OR	: Odds Ratio (Risk Oranı)
ÖSS	: Öğrenci Seçme Sınavı
ÖSYM	: Ölçme Seçme ve Yerleştirme Merkezi
ÖYS	: Öğrenim Yönetim Sistemi
QUEST	: Quick, Unbiased, Efficient, Statistical Trees (Hızlı, Tarafsız, Etkili, İstatistiksel Ağaçlar)
sd.	: serbestlik derecesi
SPRINT	: Scalable Parallelizable Induction of Decision Trees (Ölçeklenebilir, Parallelenebilir Karar Ağaçları İndüksiyonu)
SPSS	: Statistical Package for the Social Sciences (Sosyal Bilimler İçin İstatistik Paketi)
SQL	: Structured Query Language (Yapılandırılmış Sorgu Dili)
TID	: Transaction Identification (Etkileşim Kimliği)
UE	: Uzaktan Eğitim
WTE	: Web Tabanlı Eğitim
vb.	: ve benzeri
vd.	: ve diğerleri
VM	: Veri Madenciliği
VT	: Veri Tabanı
VTBK	: Veri Tabanlarında Bilgi Keşfi

BÖLÜM 1

GİRİŞ

1.1. AMAÇ

Eğitim sistemlerinde karşılaşılan sorunların çözümünde daha başarılı olma amacı politikacıları ve diğer karar vericileri yeni arayışlara ve yeniliklere yönlendirmektedir. Bu amaca ulaşmak için sarf edilen çabalarının temelinde iki temel inanç yatmaktadır. Bunlardan ilki eğitimde fırsat eşitliği inancı diğeri ise eğitimin bireysel veya toplumsal amaçların gerçekleşmesinde katkıda bulunabilecek bir etmen olmasıdır. Bu iki inancın yerine getirilmesi amacıyla ortaya atılan alanlardan biri uzaktan eğitimidir. Uzaktan eğitim ile bireysel ve toplumsal amaçların gerçekleşmesinin daha kolay olacağı ve bireylerin eğitime erişmesinde fırsat eşitliği sağlanacağını ifade edilmektedir (Kaya, 2006). Benzer şekilde İşman (2008) da uzaktan eğitimin eğitimde karşılaşılan bazı sorunlara karşı bir alternatif olabileceği ve yaygınlaşması ile temel eğitim ve yüksek öğrenimden yararlanamayan bireyler için fırsat eşitliği sağlanacağını belirtmektedir.

Yükseköğretim kurumlarında açılan uzaktan eğitim programlarının özellikle zaman ve mekân konusunda öğrenci için daha esnek yapıda olması bu programları daha cazip hale getirme potansiyeli taşımaktadır. İş hayatına atılan veya herhangi başka bir sebepten dolayı yüksek öğretime erişim imkânı bulamayan bireyler uzaktan eğitim ile eğitim hayatlarına devam etme fırsatına sahip olabilmektedirler.

Son yıllarda ülkemizdeki mühendislik fakültelerinde uzaktan eğitim yöntemleri ile mühendislik eğitimi veren bölümler açılmaktadır. Bu bölümler beraberinde bazı eleştirileri getirmektedir. Yapılan eleştiriler uzaktan eğitimin mühendis yetiştirmede başarısız olacağı, mühendislik eğitimin bazı uygulama alanlarına ve laboratuvarlara

ihtiyaç duyduğu, dolayısıyla uzaktan eğitim ile mühendis yetiştirilemeyeceği yönünde dile getirilmektedir (Makina Mühendisleri Odası, 2009).

Yapılan bu eleştirilerin ne derece doğru olduğunu ortaya koymak için, uzaktan eğitim yöntemi ile eğitim veren bu bölümlerin analizinin önemlidir. Öz ve Yüksel (2007) makine mühendisliği örneğinde yaptıkları çalışmada farklı yöntemlerle verilen aynı mühendislik dersi üzerinde uzaktan eğitim yönteminin geleneksel öğretim yönteminden daha etkili olduğunu kanıtlamışlardır. Bu durumu destekleyen başka çalışmalar da bulunmaktadır (Göksu, 2012). Bu nedenlerden uygulanan uzaktan eğitim modelinin, öğrencilerin akademik başarılarının ve demografik dağılımlarının analizinin, kısmen de olsa yapılan eleştirilere ışık tutacağı düşünülmektedir.

Bu çalışmada Karabük Üniversitesi uzaktan eğitim bilgisayar ve endüstri mühendisliği bölümlerinin analiz edilmesi amaçlanmaktadır. Araştırmada şu sorulara cevap aranmaktadır;

- Bölümlere göre öğrencilerin üniversiteye giriş puanları arasında anlamlı bir fark var mı?
- Uzaktan eğitim öğrencileri ile yüzyüze eğitim öğrencilerinin not ortalamaları arasında anlamlı bir fark var mı?
- Uzaktan eğitim bilgisayar ve endüstri mühendisliği bölümleri öğrencilerinin akademik başarıları arasında anlamlı bir fark var mı?
- Üniversite giriş puanı ile not ortalaması arasında nasıl bir ilişki bulunmaktadır?
- Uzaktan eğitim derslerine katılım oranı, üniversite giriş puanı, yaş ve cinsiyet öğrencinin dersi geçme durumunu nasıl yordamaktadır?

1.2. ÖNEM

Ülkemizde mühendislik eğitiminin uzaktan eğitim yöntemiyle verilmesine son yıllarda başlanmıştır. Bu eğitim yöntemi üniversitelerdeki birinci ve ikinci öğretimden farklılık göstermektedir. Uzaktan eğitim ile lisans eğitimi alan öğrencilerin akademik başarılarının mühendislik eğitimi konusunda uzaktan eğitimin başarılı olup olmadığının ortaya konulmasında önemli olduğu düşünülmektedir.

Ayrıca öğrencilerin ders geçme durumlarını tahmin etmek için geliştirilen modelin öğrencilere ve öğretim elemanlarına geri bildirimlerde bulunabileceği önceden bazı önlemlerin alınmasına yardımcı olabileceği düşünülmektedir.

1.3. YÖNTEM

Araştırmada problemlerin çözümü amacıyla veri madenciliği yöntemleri kullanılmaktadır. Bu yöntemlerden demografik özelliklerin analizi tanımlayıcı istatistiksel yöntemlerden yararlanılarak, üniversite giriş puanları ile not ortalamalarının analizleri parametrik ve parametrik olmayan testlerden yararlanılarak yapılmış, öğrencilerin üniversiteye giriş puanları ile akademik başarıları arasındaki ilişki incelenmiş ve son olarak da derslere ait farklı değişkenlerin başarı üzerindeki etkisi lojistik regresyon analizi ile modellenmiştir.

Verilerin analiz edilmesi ve oluşturulan modelin uygulanmasında SPSS 21.0 istatistik yazılımı kullanılmıştır. Analizlerde istatistikî anlamlılık düzeyi $p < 0.05$ olarak kabul edilmiştir.

Kullanılan veriler Karabük Üniversitesi Bilgisayar ve Endüstri Mühendisliği bölümlerine ait 2009 ile 2013 yılları arasındaki hem yüzyüze eğitim hem de uzaktan eğitim öğrencilerine ait verilerden oluşmaktadır.

Öğrencilerin bölümlerine ve sınıflara göre dağılım istatistikleri Çizelge 1.1’de yer almaktadır.

Çizelge 1.1. Sınıflara göre öğrenci dağılımları.

Bölüm Adı	Sınıf	Sıklık	Yüzde	Yığılmalı Yüzde
Bilgisayar Mühendisliği (%30 İngilizce)	Hazırlık	6	2,3	2,3
	1. Sınıf	88	34,0	36,3
	2. Sınıf	69	26,6	62,9
	3. Sınıf	65	25,1	88,0
	4. Sınıf	31	12,0	100,0
	Toplam	259	100,0	

Çizelge 1.1. (devam ediyor).

Bilgisayar Mühendisliği (%30 İngilizce) (İÖ)	Hazırlık	5	1,9	1,9
	1. Sınıf	88	32,7	34,6
	2. Sınıf	87	32,3	66,9
	3. Sınıf	67	24,9	91,8
	4. Sınıf	22	8,2	100,0
Toplam	269	100,0		
Bilgisayar Mühendisliği (UE)	1. Sınıf	24	7,2	7,2
	2. Sınıf	91	27,2	34,3
	3. Sınıf	110	32,8	67,2
	4. Sınıf	110	32,8	100,0
	Toplam	335	100,0	
Endüstri Mühendisliği	1. Sınıf	79	29,4	29,4
	2. Sınıf	86	32,0	61,3
	3. Sınıf	73	27,1	88,5
	4. Sınıf	31	11,5	100,0
	Toplam	269	100,0	
Endüstri Mühendisliği (İÖ)	Hazırlık	3	1,1	1,1
	1. Sınıf	83	30,4	31,5
	2. Sınıf	74	27,1	58,6
	3. Sınıf	85	31,1	89,7
	4. Sınıf	28	10,3	100,0
Toplam	273	100,0		
Endüstri Mühendisliği (UE)	1. Sınıf	8	3,1	3,1
	2. Sınıf	96	37,5	40,6
	3. Sınıf	102	39,8	80,5
	4. Sınıf	50	19,5	100,0
	Toplam	256	100,0	
Bilgisayar Mühendisliği (%100 İngilizce)	Hazırlık	1	,5	,5
	1. Sınıf	84	42,0	42,5
	2. Sınıf	74	37,0	79,5
	3. Sınıf	41	20,5	100,0
	Toplam	200	100,0	
Bilgisayar Mühendisliği (% 100 İngilizce) (İÖ)	Hazırlık	6	3,0	3,0
	1. Sınıf	87	42,9	45,8
	2. Sınıf	74	36,5	82,3
	3. Sınıf	36	17,7	100,0
	Toplam	203	100,0	

Çizelge 1.1’den de anlaşılacağı üzere analiz için kullanılan en az öğrenci sayısına 200 öğrenci ile Bilgisayar Mühendisliği (%100 İngilizce) programı sahipken en fazla öğrenci sayısına ise 335 öğrenci ile Bilgisayar Mühendisliği (UE) programı sahiptir.

Çizelge 1.2’de bilgisayar ve endüstri mühendisliği öğrencilerine ait üniversite giriş puanlarının yıllara göre betimsel istatistikleri yer almaktadır. Burada yer alan üniversite giriş puanları üniversiteyi kazanıp kayıt yaptıran öğrencilere ait puanlardan oluşmaktadır.

Çizelge 1.2. Üniversite giriş puanlarına ait betimsel istatistik.

Bölüm Adı	Giriş Yılı	Ortalama	Ortanca	Tepe Değer	St. Sapma	Genişlik	Min. Değer	Maks. Değer
Bilgisayar Müh. (%30 İngilizce)	2009-2010	314,58	313,76	311,89*	2,91	13,82	311,89	325,71
	2010-2011	402,45	402,15	365,84*	7,25	66,49	365,84	432,34
	2011-2012	369,46	370,21	256,26*	20,29	151,43	256,26	407,69
	2012-2013	316,72	316,09	242,95*	10,05	98,48	242,95	341,44
Bilgisayar Müh. (%30 İngilizce) (İÖ)	2009-2010	306,06	306,22	302,85*	2,57	10,20	302,85	313,05
	2010-2011	381,25	379,43	372,03*	7,41	29,49	372,03	401,52
	2011-2012	345,42	344,18	338,28*	6,29	36,01	338,28	374,30
	2012-2013	298,07	296,30	296,54	4,29	21,44	293,53	314,98
Bilgisayar Müh. (UE)	2009-2010	272,56	269,16	264,14*	8,77	32,01	264,14	296,15
	2010-2011	306,14	301,37	253,91*	20,14	129,43	253,91	383,35
	2011-2012	269,51	260,98	244,11*	28,40	150,38	244,11	394,50
	2012-2013	254,05	248,79	230,80*	22,10	146,92	230,80	377,73
Endüstri Müh.	2009-2010	308,26	307,88	307,86	2,44	12,61	305,29	317,91
	2010-2011	396,18	393,74	387,71*	8,11	43,18	387,71	430,90
	2011-2012	372,38	370,22	363,12*	7,88	35,36	363,12	398,48
	2012-2013	327,62	324,63	319,66*	8,04	33,89	319,66	353,56
Endüstri Müh.(İÖ)	2009-2010	300,89	300,16	297,60*	2,50	8,47	297,60	306,08
	2010-2011	373,56	372,30	364,21*	7,04	29,46	364,21	393,68
	2011-2012	349,00	346,26	338,95*	7,55	33,33	338,95	372,29
	2012-2013	303,35	302,22	297,82*	4,41	20,68	297,82	318,51
Endüstri Müh. (UE)	2010-2011	267,24	266,71	248,27*	13,72	53,15	248,27	301,43
	2011-2012	247,37	242,41	223,56*	20,51	89,86	223,56	313,42
	2012-2013	233,81	228,58	211,03*	19,91	144,03	211,03	355,06
Bilgisayar Müh. (%100 İngilizce)	2010-2011	408,52	406,26	404,47	9,82	42,96	395,73	438,69
	2011-2012	374,80	372,09	248,33*	19,81	163,63	248,33	411,96
	2012-2013	329,93	324,77	266,79*	17,23	137,51	266,79	404,30

Çizelge 1.2. (devam ediyor).

Bilgisayar Müh. (%100 İngilizce) (İÖ)	2010-2011	378,40	375,88	366,94*	9,76	43,44	366,94	410,38
	2011-2012	350,79	348,20	338,80*	11,52	66,36	338,80	405,16
	2012-2013	302,64	300,93	296,39*	6,57	33,43	296,39	329,83

*Birden fazla tepe noktası olduğundan minimum değer alınmıştır.

Araştırmada kullanılan verilerden öğrencilerin demografik bilgileri, üniversite giriş puanları, not ortalamaları ile ders bilgileri ve sınav sonuçları öğrenci otomasyonu veritabanından; derslere katılım ve diğer ders etkinlikleri ise öğrenim yönetim sistemine ait veritabanından yararlanılarak elde edilmiştir.

Verilerin analizinde parametrik ve parametrik olmayan testlerden yararlanılmıştır. Bu testlerin uygulanmasında bazı varsayımların sağlanması gerekmektedir. Örneğin parametrik bir test olan tek yönlü ANOVA testinin yapılabilmesi için, gruplara ait değişkenlerin normal dağılım göstermesi gerekirken, ortalamaları kıyaslanacak grupların varyanslarının da eşit olması gerekmektedir. ANOVA testi yapılabilmesi için gereken şartlar yerine gelmediği zaman veriler üzerinde dönüşüm işlemleri yapılabileceği gibi parametrik olmayan testlerden de yararlanılabilmektedir. Parametrik olmayan testlerden Kruskal Wallis testiyle de analiz edilen gruplar arasında istatistiksel olarak anlamlı bir fark olup olmadığı analiz edilebilmektedir (Can, 2013).

İki veri grubu arasındaki ilişkinin yönünün ve miktarının istatistiksel olarak gösterilmesine korelasyon denir ve elde edilen katsayı korelasyon katsayısı olarak adlandırılır. Bu katsayı -1 ile +1 arasında bir değer alır ve r sembolü ile gösterilir. Korelasyon katsayısı 1'e ne kadar yaklaşırsa ilişki o kadar güçlü olur. İlişkinin yönü ise bu katsayının aldığı pozitif veya negatif değerlerden anlaşılmaktadır. Değerin negatif yönlü olması ters yönlü ilişkiyi, pozitif yönlü olması ise doğru yönlü ilişkiyi ifade etmektedir (Can, 2013).

Veri grupları arasındaki ilişkiyi tespit etmek için çeşitli yöntemler kullanılmaktadır. Bunlardan bir tanesi Pearson Moment Çarpım Korelasyonu'dur. Bir diğeri ise Spearman Sıra Farkları Korelasyonu'dur. En az aralık ölçeğinde olan verilerde bazı

varsayımların karşılanması durumunda örneğin verilerin normal dağılım göstermesinde Pearson korelasyonu kullanılabilir. Bu varsayımların karşılanmadığı durumlarda ise Spearman korelasyonu kullanılabilir (Can, 2013). Üniversite giriş puanları ile not ortalamaları arasındaki ilişkiyi tespit etmek için yapılan analizde Pearson korelasyonunun hesaplanması için gereken varsayımlar karşılanmadığı için Spearman korelasyonu kullanılmıştır.

Model oluşturulurken lojistik regresyon analizinin seçilmiştir. Bu analizin kullanılması mevcut veri dağılımlarının ve özelliklerinin yapılabilecek alternatif analizlerin varsayımlarını büyük ölçüde sağlamamasından kaynaklanmaktadır. Örneğin verilerin normal dağılım göstermesi şartı ile sürekli veya kesirli olma şartı bu analizde aranmamaktadır. Bu da farklı tür veriler üzerinde analiz yapmayı kolaylaştırmaktadır. Bu nedenlerden uzaktan eğitim öğrencilerin ders geçme durumları ile üniversite giriş puanı, yaş, cinsiyet, uzaktan eğitime katılım oranı arasındaki ilişki lojistik regresyon analizi ile analiz edilerek bir model geliştirilmiştir. Bu modelde yer alan verilerde diğer faktörlerin etkisini azaltmak amacıyla hem yüzyüze eğitim hem de uzaktan eğitimde aynı öğretim elemanlarının verdiği aynı derslerden oluşmaktadır. Oluşturulan veri grubunda bu özelliklere uyan en uygun veriler seçilmeye çalışılmış ve yapılan incelemede en uygun veriler 2011-2012 eğitim öğretim yılının güz ve bahar dönemine ait bilgisayar mühendisliği ikinci sınıfa ait zorunlu derslerinden elde edilmiştir.

Lojistik regresyon analizinde (LRA) modelleme yapıldığında ilk önce teorik modelin veriyi temsil etmesi bakımından uygun olup olmadığının incelenmesi gerekmektedir. Bunun için istatistik programı ile Hosmer ve Lemeshow testi yapılmakta ve bu testle “teorik model veriyi temsil etmektedir” hipotezi test edilmektedir. Bu amaçla model oluşturulurken Hosmer ve Lemeshow testi yapılmış ve sonucunda modelin veriyi temsil ettiği anlaşılmıştır ($p>0,05$).

Oluşturulan modelin veriyi iyi temsil ettiğinin test edilmesinden sonra uygulanacak diğer test, modelin parametrelerinin anlamlılığı testi olan Omnibus testidir. Yapılan analiz sonucunda model parametrelerinin anlamlı olduğu anlaşılmıştır ($p<0,05$).

Model parametrelerinin anlamlılığının testinden sonra modelin denklemini oluşturmak için analiz sonuçlarında yer alan Eşitlikteki Değişkenler tablosundan yararlanılmıştır. Bu tabloya göre cinsiyet değişkeni hariç diğer tüm bağımsız değişkenlere ait katsayıların modeli açıklamada anlamlı olduğu anlaşılmıştır ($p<0,05$).

1.4. SINIRLILIKLAR

Bu araştırma Karabük Üniversitesi örgün ve uzaktan eğitim bilgisayar ve endüstri mühendisliği bölümlerinde 2009-2013 yılları arasında kayıt yaptırmış öğrencilere ait demografik bilgiler ile üniversiteye giriş puanı, sınav sonucu, ağırlıklı genel not ortalaması ve uzaktan eğitim öğrencileri için derslere katılım oranı değişkenleri ile sınırlıdır.

Araştırmaya yatay geçiş, dikey geçiş, mühendislik tamamlama ve çift anadal programları ile gelen öğrenciler dâhil edilmemiştir.

1.5. TANIMLAR

Yüzyüze Eğitim (YE) : Ders ve etkinliklere devam etme zorunluluğu olan bir eğitim-öğretim türüdür.

Birinci Öğretim (BÖ): Yükseköğretim kurumlarında mesai saatleri içerisinde yapılan (yüzyüze) öğretimdir.

İkinci Öğretim (İÖ): Yükseköğretim kurumlarında birinci öğretimi takiben yapılan (yüzyüze) öğretimdir.

Uzaktan Eğitim (UE): Öğrenci ve öğretim elemanının aynı ortamda bulunmasının gerekmediği ve eğitimin eş-zamanlı veya eş-zamansız olarak verilebildiği eğitimdir.

Eş-zamanlı (Senkron) ders: Aynı zamanda yapılan ders veya etkinliklerdir.

Eş-zamansız (Asenkron) ders: Farklı zamanda yapılan ders veya etkinliklerdir.

Ağırlıklı Genel Not Ortalaması (AGNO): Bir öğrencinin üniversiteye giriş yaptırmışından itibaren bulunduğu döneme kadar sorumlu olduğu derslerin ağırlıklı notlarının ortalamasıdır. Bu ifade tezin hipotezlerinde not ortalaması olarak kullanılmıştır.

Katılım Oranı (KO): Uzaktan eğitim öğrencilerinin uzaktan eğitim yöntemi ile aldıkları derslere ve bu derslere ait diğer etkinliklere katılımlarının oranıdır.

Öğrenim Yönetim Sistemi (ÖYS): Uzaktan eğitim yöntemi ile verilen eğitimin organizasyonunun web teknolojileri kullanılarak yürütülmesi ve yönetilmesini sağlayan sistemdir.

1.6. ETİK

Araştırmada kullanılan öğrencilere ait tüm veriler için Karabük Üniversitesi Mühendislik Fakültesi dekanlığından gerekli izinler yazılı olarak alınmıştır (Ek Açıklamalar A). Kullanımına izin verilen bütün veriler büyük bir titizlikle korunmuş ve üçüncü şahıslar ile kesinlikle paylaşılmamıştır. Analizlerin kontrolü için gerekli durumlarda ise bu veriler, hangi verinin hangi öğrenciye ait olduğu araştırmacı haricindeki kişilerce anlaşılacak şekilde, belirli bir algoritma ile dönüştürülerek kullanılmıştır.

BÖLÜM 2

VERİ MADENCİLİĞİ

2.1. GİRİŞ

Günümüzde bilişim ve iletişim teknolojilerindeki yeni gelişmeler son kullanıcının bu teknolojilere erişiminin daha kolay ve ekonomik olmasını sağlamakla birlikte daha çok bilginin daha hızlı saklanmasına, iletilmesine ve kullanılmasına imkân sağlamaktadır. Günlük hayatımızın birçok alanına giren bu teknolojilerle başta kişisel bilgilerimiz olmak üzere birçok veri dijital ortamlarda saklanmakta ve çeşitli amaçlarla kullanılabilir (Şentürk, 2006). Örneğin süpermarketlerde; raf düzenlemeleri, indirim ve kampanya yapılması gibi konularda müşterilerin alışveriş davranışlarından yararlanabilmektedir. Bunun için de kayıt altında tutulan müşteri verileri analiz edilmektedir. Pazarlama alanında özellikle daha fazla kâr amacıyla yapılan bu analizler, tıp alanında hastalıkların erken dönemde tespiti için yapılabilmekteyken bankacılık alanında dolandırıcılık tespiti gibi amaçlarda kullanılabilir (Özkan, 2013).

Müşteri davranışları arasındaki ilişki ve örüntüleri, erken dönemde hastalık teşhisi, dolandırıcılık tespiti vb. gibi konularda bilgi keşfini sağlamada karşılaştığımız şu soruların cevapları veri madenciliğinin temellerini oluşturmaktadır: Veri nedir? Nasıl, ne şekilde ve hangi ortamlarda kayıt altında tutulmaktadır? Verilerden farklı bilgiler nasıl keşfedilmektedir?

2.1.1. Veri

Veri, “muhakeme, tartışma, ya da hesaplama için temel olarak kullanılan olgusal bilgi” anlamında kullanılmaktadır. Latince *datum* kelimesinin çoğulu olan *data* kelimesi ile ifade edilmektedir. Veri, kullanışlı olabileceği gibi ilgisiz veya gereksiz

de olabilir. Bu yüzden anlamlı hale gelebilmesi için işlenmesi gerekmektedir (Merriam-Webster Online Dictionary, 2013).

Şentürk (2006)'e göre veri, bilgiyi elde etmeye yarayan işlenmemiş ham malzemedir. Yukarıdaki iki tanımdan da anlaşıldığı üzere veri, kullanışlı veya kullanışsız olabilmekle birlikte anlamlı bilginin temelini oluşturmaktadır.

Tüzüntürk (2010)'e göre verilerin düzenlenmiş haline enformasyon (*information*) denilmektedir.

Düzenlenen ve enformasyon haline dönüşen verinin, anlamlı hale gelmesi için analiz edilmesi ve sentezlenmesi sonucunda, karar vermeye yönelik daha üst seviyeli gerçekleri içermesi haline ise bilgi (*knowledge*) denilmektedir (Çağiltay, 2010).

2.1.2. Verilerin Saklanması

Verilerin bilgisayar ortamında saklanması ve istenildiği zaman erişilebilmesi için iki farklı sistem bulunmaktadır. Bu sistemlerde ilki Geleneksel Dosyalama Sistemleridir. İkincisi ise Veri Tabanı Yönetim Sistemleridir.

2.1.2.1. Geleneksel Dosyalama Sistemleri

Bilgisayar ortamında verilerin saklanması ilk zamanlar geleneksel dosyalama sistemleri ile yapılmaktaydı. Bu dosyalama sistemleri sıralı erişimli, dizinli ve hesaba dayalı dosya sistemi olmak üzere üç şekilde sınıflandırılmaktadır. Sıralı erişimli dosya sisteminde her bir kayıt sıralı halde dosyada yer alır ve istenildiği zaman veriye ulaşmak için tüm dosyanın taranması gerekmektedir. Dizinli dosya sisteminde her bir kayda ait bir dizin numarası yer alır ve erişilmek istenilen veriye bu dizin numarası ile doğrudan erişilir. Hesaba dayalı dosya sistemindeki dosyalarda ise bir dizin numarası verisinin tutulmasına gerek yoktur. Erişilmek istenen veriye doğrudan erişmek için bir hesap algoritması kullanılmaktadır (Aydın Ö. , 2014).

Geleneksel dosyalama sistemleri veri boyutlarının artması ile veri bütünlüğü ve güvenliğinin sağlanması konusundaki ihtiyacı karşılayamaz hale gelmiştir. Bunun sonucunda da yeni dosyalama sistemleri geliştirilmiştir (Aydın Ö. , 2014).

2.1.2.2. Veri Tabanı Yönetim Sistemleri

Birbiriyle ilişkili verilerin tutulduğu, kullanım amacına uygun olarak düzenlenmiş verilerin mantıksal ve fiziksel olarak tanımlarının olduğu bilgi depolarına *veritabanı* denilmektedir (Şentürk, 2006).

Veritabanları yapılarına göre İşlemsel, İlişkisel, Ağ Modeli ve Hiyerarşik Model gibi farklı çeşitlere ayrılmaktadır. Günümüzde en fazla kullanılan model ilişkisel veritabanı modelidir (Gözüdeli, 2005).

Veritabanlarını tanımlamak, oluşturmak, kullanmak, değiştirmek ve veri tabanı sistemleri ile ilgili her türlü işletimsel ihtiyaçları karşılamak için kullanılan yazılım sistemine Veri Tabanı Yönetim Sistemi (VTYS) denilmektedir. Bu sistemlerin geleneksel sistemlere karşı üstün özelliklerinden bazıları şunlardır (Yarımağan, 2010):

- Veri tekrarını azaltma
- Veri bütünlüğünü sağlama
- Veri paylaşımını sağlama
- Kullanımı yaygınlaştırma
- Verilerin gizliliğini ve güvenliğini sağlama
- Standartları uygulayabilme

İlişkisel veritabanlarının yönetimi için geliştirilen sistemlere İlişkisel Veri Tabanı Yönetim Sistemleri (İVTYS) denilmektedir. Bu sistemlerin temeli 1970’li yıllarda E. F. Codd tarafından yapılan çalışmalarla atılmıştır. Bu sistemlerde yer alan veritabanlarının en önemli yanı veritabanının tablolardan oluşmasıdır. İlişkisel kavramı da bu tablolar arasında oluşturulan ilişkilerden gelmektedir. Birden fazla tabloya ayrılan fazla veri, tablolar arasındaki ilişkilendirmeler sayesinde daha az ve

kolay yönetilebilir şekilde saklanabilmektedir. Günümüzde en sık kullanılan İVTYS'den bazıları şunlardır (Gözüdeli, 2005):

- MySQL
- MS SQL
- Oracle
- Sybase
- Informix

İlişkisel veriler genellikle On-Line Analytical Processing (OLTP) sistemler tarafından kullanılmaktadır. OLTP sistemler günlük olarak verilerin işlendiği sistemlerdir. Bu sistemlerde kullanılan veritabanına ilgili veriler eklenebilir, gerekli görüldüğünde düzeltilbilir veya silinebilmektedir. Bu sistemlere örnek olarak stok takip sistemini verebiliriz. Bir firmanın tutmuş olduğu ürün stoku günlük olarak değişikliğe uğrar; yeni gelen ürünler sisteme kaydedilirken, stok durum bilgileri de güncellenebilir. Aynı zamanda mevcut stok durumu ile ilgili raporlamalar ve sorgulamalar da yapılabilir (Özkan, 2013).

2.1.3. Veri Ambarı

Belirli bir döneme ait, konu odaklı olarak düzenlenmiş, birleştirilmiş ve sabitleştirilmiş işlemlere ait veritabanlarına *veri ambarı* denilmektedir. Veri ambarı üzerinde çeşitli analizler ve sorgulamalar yapmayı kolaylaştıracak ve karar almada yardımcı olacak işlemlere Çevrimiçi Analitik İşlemler (On-Line Analytical Processing - OLAP) denilmektedir (Silahtaroglu, 2008).

Çizelge 2.1. OLTP ve OLAP karşılaştırması.

	OLTP	OLAP
Kullanıcı	Bilişim Personeli	Bilgi Analisti
Fonksiyon	Günlük İşler	Karar verme
Veritabanı Tasarımı	Uygulamaya yönelik	Karar vermeye yönelik
Veri	Güncel, detaylı, ilişkisel	Zamana dayalı, özet, çok boyutlu
Kullanım	Tekrarlı sorgular	Anlık Sorgular

Çizelge 2.1. (devam ediyor).

Erişim	Okuma, yazma, değişiklik	Okuma
Birim İş	Kısa ve basit hareketler	Karmaşık sorgular
Erişilen Kayıt Sayısı	Onlarca	Milyonlarca
Kullanıcı Sayısı	Binlerce	Yüzlerce
Veritabanı Büyüklüğü	100-MB-GB	100GB-TB
Ölçüm	Hareketlilik Performansı	Sorgu Performansı

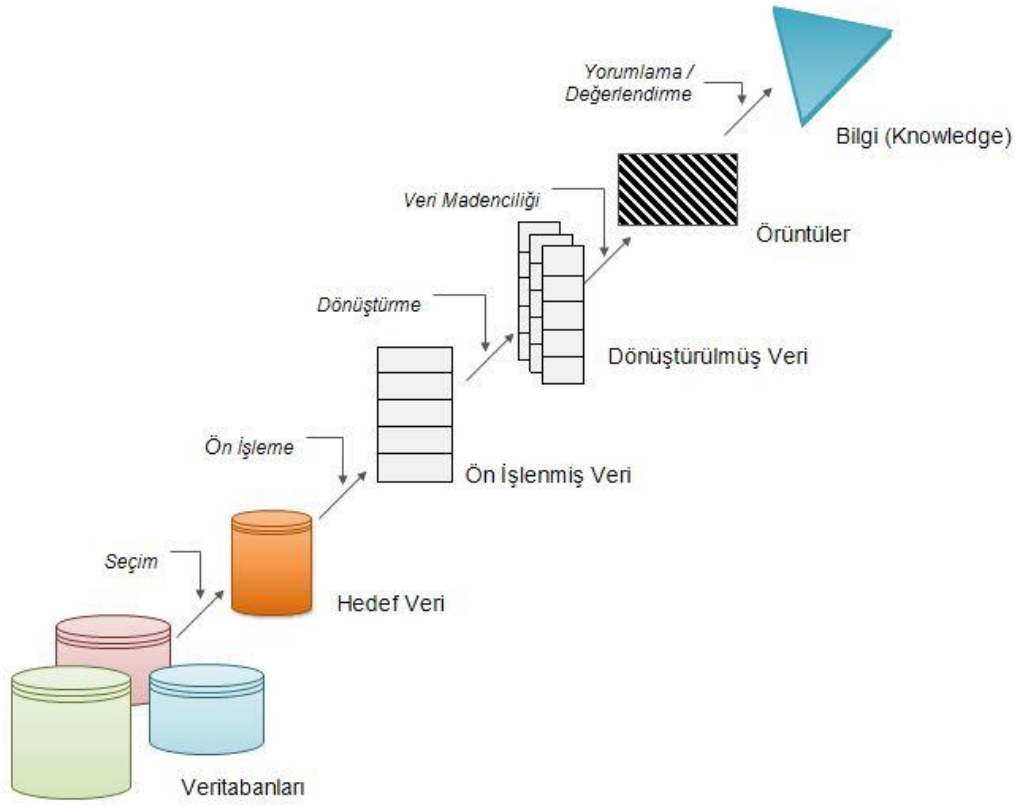
Kaynak: Çağıltay (2010) s. 42'den alınmıştır.

Çizelge 2.1'de de görüldüğü gibi OLTP işlemleri gündelik işler için uygulamaya yönelik kullanılmaktayken OLAP işlemleri karar vermeye yönelik kullanılmaktadır. Aynı zamanda OLTP'de veriler kaydedilip değişiklik yapılabilmekteyken, OLAP'da veriler süreli olarak ve değiştirilemez halde saklanmaktadır.

Teknolojik gelişmelerle paralel olarak verilerin elde edilmesi, düzenlenmesi ve muhafaza edilmesi daha kolay hale gelmiştir. VTYS'nde daha çok verinin daha güvenilir ortamlarda uzun yıllar saklanması ile birlikte de verilerden bilgi keşfi süreci hızlanmıştır. Elde edilen veri arttıkça ve verilerdeki karmaşıklık arttıkça da daha iyi çözümleme tekniklerine ihtiyaç duyulmaya başlanmıştır.

Veritabanındaki verilerin kullanılarak bilgi elde edilmesi için gerekli olan bütün süreçleri ifade etmek amacıyla Veritabanlarında Bilgi Keşfi (VTBK) kavramı kullanılmıştır. Veri madenciliği (VM) ise veriden bilgi keşfetme sürecindeki önemli adımlardan birisini oluşturmaktadır (Gürsoy, 2009).

Veritabanlarından bilginin keşfedilmesi için gerekli aşamalar Şekil 2.1'de görülmektedir. Bu süreçte hedef verinin seçimi, önışleme yapılması, dönüştürülmesi, veri madenciliği yöntemleri ile veriler arasındaki örüntülerin bulunması ve bulunan bu örüntülerin yorumlanarak bilginin keşfedilmesi yer almaktadır.



Kaynak: Gürsoy (2009) s. 41'den uyarlanmıştır.

Şekil 2.1. Veritabanından bilgi keşfi aşamaları.

Veri madenciliği Şekil 2.1'de görüldüğü gibi VTBK sürecinin bir basamağıdır. Gizli kalmış örüntüleri ortaya çıkarması ile zamanla bu kavram bilgi keşfinden daha çok ön plana çıkar hale gelmiştir (Şentürk, 2006).

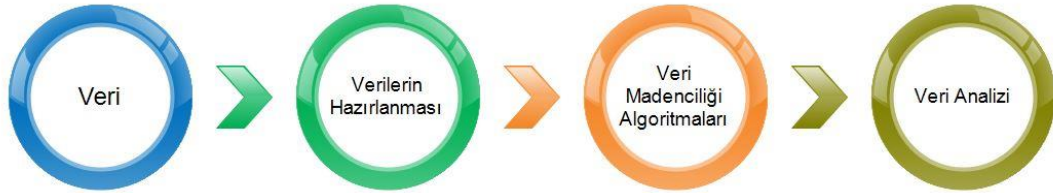
2.2. VERİ MADENCİLİĞİ NEDİR?

Özkan'a (2013) göre veri madenciliği; büyük boyutlu veriler arasından değeri olan bir bilginin elde edilmesi işidir. Silahtaroglu'na (2008) göre daha önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin veritabanlarından elde edilmesi ve bu bilgilerin işletme kararları verirken kullanılmasıdır. Gürsoy'a (2009) göre ise anlamlı örüntü ve kurallar çıkarmak amacıyla büyük miktardaki verinin analiz edilmesidir.

Sumathi ve Sivanandam (2006) VM sürecinin verilerin hazırlanması, VM algoritmalarının uygulanması ve verilerin analizi olmak üzere üç önemli aşamadan oluştuğunu belirtmektedir:

- Verilerin hazırlanması aşaması: Verilerin alan uzmanları tarafından seçildiği, temizlendiği ve ön işleme tabi tutulduğu aşamadır. Veri madenciliğinde süreç olarak en uzun süreyi bu aşama oluşturmaktadır.
- Veri madenciliği algoritmaları aşaması: Hazırlanan verilerden yararlanılarak belirli bir veri madenciliği algoritması / algoritmalarının kullanıldığı ve bütün değerli bilgi parçacıklarının tanımlandığı aşamadır.
- Veri analizi aşaması: Veri madenciliği sonuçlarının ve eğer varsa veri madenciliği algoritmaları ile bulunan gerçeklerin bağlı önemini tanımlamak ve keşfedilen bilgileri anlamak için değerlendirildiği aşamadır.

Şekil 2.2’de yukarıda bahsedilen VM ile ilgili süreç yer almaktadır.



Kaynak: Sumathi and Sivanandam (2006) s. 40’dan uyarlanmıştır.

Şekil 2.2. Veri madenciliği süreci.

2.3. VERİ MADENCİLİĞİ UYGULAMA ALANLARI

Silahtaroglu’na (2008) göre veri madenciliğinin uygulandığı alanlar şunlardır:

- Pazarlama Yönetimi: Pazarlama alanında yapılan bazı veri madenciliği uygulamalarına ait bazı örnekler şu şekildedir: müşterilerin satın alma örüntülerinin belirlenmesi, müşterilere ait demografik özellikler arasındaki bağıntıların bulunması, posta kampanyalarına cevap verme oranının artırılması, mevcut müşterilerin korunması, yeni müşterilerin kazanılması ve pazar sepeti analizidir.
- Risk Yönetimi ve Dolandırıcılık Saptama: Kredi kartı dolandırıcılığı, sigorta dolandırıcılığı, telefon dolandırıcılığı ve bilgisayar sistemleri ile bilgisayar ağlarına girmeyeyle yapılan dolandırıcılıkları saptama bu alandaki bazı örneklerdir.

- İşaret İşleme: Telefon hatlarındaki parazitlerden oluşan kayıpları azaltma ve oluşan gürültüyü yok etmeyi sağlayan uygulamalar bu alandaki örneklerdendir.
- Biyoloji: Özellikle genler üzerine yapılan çalışmalarda kullanılan VM yöntemleri ile yapılan analizler bu alandaki örneklerdendir.
- Tıp: Bazı hastalıkların erken dönemde tanısının konulabilmesinin güçlüğünden kaynaklanan durumlarda var olan bazı verilerden yararlanarak VM yöntemleri ile çeşitli risklerin tahmin edilmesi mümkün olabilmektedir.

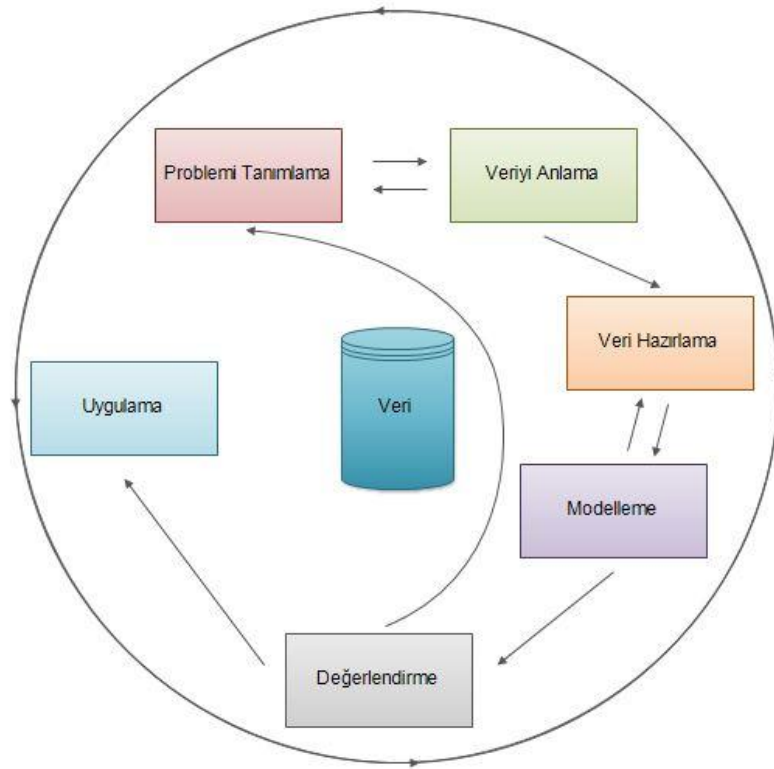
Veri madenciliği yöntemleri yukarıda da görüldüğü gibi birçok alanda farklı amaçlar için kullanılmaktadır.

2.4. VERİ MADENCİLİĞİ SÜRECİ

1996 yılının sonlarına doğru The Cross Industry Standart Process for Data Mining (CRISP-DM) konsorsiyumu kurulmuştur (Gürsoy, 2009). Bu konsorsiyum tarafından belirlenen VM süreci altı aşamadan oluşmaktadır (Chapman et al., 2000).

- Problemin tanımlanması
- Verinin anlaşılması
- Veri hazırlama
- Modelleme
- Değerlendirme
- Uygulama

Şekil 2.3’de CRIPS-DM tarafından belirlenen VM süreci yer almaktadır.



Kaynak: Chapman et al. (2000)'den uyarlanmıştır.

Şekil 2.3. CRIPS-DM veri madenciliği süreci.

Şekil 2.3'de yer alan VM sürecine ait aşamalar ayrıntılı olarak şu şekildedir:

- Problemin tanımlanması: VM sürecinin en önemli aşamasıdır. Bu aşamada problemin iyi bir şekilde tanımlanması gerekmektedir. Eğer problem iyi olarak tanımlanmazsa diğer bütün süreçler olumsuz etkilenebilmektedir.
- Verinin anlaşılması: Kullanılacak verilerin toplandığı, anlaşıldığı ve verilerde yer alan varsa problemlerinin belirlenmesi gibi işlemlerinin yapıldığı aşamadır.
- Veri hazırlama: Modellemede kullanılacak veri setinin oluşturulması aşamasıdır. Bu aşamada mevcut veriler arasından gerekli olan veriler seçilir ve varsa kayıp veya hatalı veriler düzeltilir.
- Modelleme: VM yöntemlerinin seçilip uygulandığı aşamadır. Gürsoy (2009)'a göre VM sürecinin önemli bir aşaması olan modellemede kullanılan modeller ikiye ayrılmaktadır. Bu modeller tanımlayıcı ve tahmin edici modellerdir. Tanımlayıcı modeller; ilişki analizi ve kümeleme analizi olmak üzere iki başlıkta incelenmektedir. Tahmin edici modeller ise sınıflandırma ve

istatistiksel tahmin modelleri olmak üzere iki ana başlık altında incelenmektedir.

- Değerlendirme: Uygulanan modelin değerlendirildiği ve varsa eksikliklerinin tamamlandığı veya hatalarının giderildiği aşamadır. Bu aşamada gerekirse problemin tanınması aşamasına geri dönülebilmektedir.
- Uygulama: VM sürecinin son aşaması olan bu aşamada oluşturulan model kullanılmaktadır.

2.5. VERİ MADENCİLİĞİ MODELLERİ

VM modelleri iki başlıkta incelenmektedir. Bunlardan ilki tanımlayıcı modeller ikincisi ise tahmin edici modellerdir. Tanımlayıcı modeller; ilişki analizi ve kümeleme analizi olmak üzere iki başlıkta incelenirken tahmin edici modeller ise sınıflandırma ve istatistiksel tahmin modelleri olmak üzere iki başlık altında incelenmektedir (Gürsoy, 2009).

2.5.1. Tanımlayıcı Modeller

Tanımlayıcı/Betimleyici modeller (*Descriptive Models*), karar vermeye rehberlik edecek mevcut verilerdeki örüntülerin tanımlanmasını sağlar. İlişki analizi ve kümeleme analizi olmak üzere iki başlık altında incelenebilmektedir (Gürsoy, 2009).

2.5.1.1. İlişki Analizi

İlişki analizi veritabanlarındaki bir dizi verinin diğer verilerle olan ilişkisini açıklayan işlemler dizisidir (Silahtaroglu, 2008). Örneğin alışverişte satın alınan bir ürünün yanında başka bir ürünün de satın alınması arasındaki ilişki bu analiz sonucunda ortaya çıkarılabilir.

Birliktelik Kuralları

Birliktelik kuralları verideki potansiyel ilişkilerin tanımlanmasında kullanılır. İlişkiyi tanımlamak için kurallar oluşturulur ve bu kurallar destek ve güvenilirlik seviyelerine

göre deęerlendirmeye alınır. Oluřturulan kuralın gl olması byk destek ve yksek gven seviyesine sahip olmasına baęlıdır (Grsoy, 2009). Burada sz konusu olan *destek* veride baęıntının ne kadar sık olduęunu, *gven* de bir rn/malı alan bir kiřinin hangi olasılıkla dięer bir malı alacaęını ifade etmektedir. rneęin bir rnn alınması ile ilgili olarak hesaplanan *gven* ve *destek* ařaęıdaki gibi hesaplanmaktadır (Alpaydın, 2000). Hesaplanan bu deęerlerin anlamlı olabilmesi iin eřik deęerlere ihtiya vardır. Hem destek hem de gven iin hesaplanan deęerlerden eřik deęerlerin byk olması beklenmektedir (zkan, 2013).

Destek:

$P(X \text{ ve } Y) = X \text{ ve } Y \text{ rnn satın alan mřteri sayısı} / \text{Toplam mřteri sayısı}$

Gven:

$P(X|Y) = X \text{ ve } Y \text{ rnn satın alan mřteri sayısı} / Y \text{ malını satın alan mřteri sayısı}$

Birliktelik kuralları sol ve saę olmak zere iki kısımdan oluřmaktadır. Sol kısım ncl veya kořul olarak adlandırılırken saę kısım da baęlı kořul olarak adlandırılmaktadır. ncl doęru olduęunda baęlı kořulda doęru olmalıdır (řentrk, 2006).

Birliktelik kurallarının ıkarılması iin eřitli algoritmalar geliřtirilmiřtir. Bu algoritmalar; AIS, SETM, Apriori ve AprioriTid algoritmalarıdır (Silahtaroglu, 2008).

AIS Algoritması

AIS algoritması 1993 yılında geniř nesne kmeleri retmek iin geliřtirilmiř bir algoritmadır. Nesne isimleri algoritma veritabanındaki nesnelere teker teker tarayarak geniř nesne kmesini oluřturur. Burada nesnelere A'dan Z'ye doęru sıralanmaktadır. Ardından minimum destek seviyesine eřit veya daha byk olanlar aday kmesini oluřturur. Sonraki ařamada, bir nceki taramada geniř oldukları belirlenen nesne kmeleri ile o iřlemin nesnelere arasında ortak kmeler belirlenir. Ardından ortak kme nesnelere ile iřleminde mevcut olan dięer nesne kmeleri birleřtirilerek yeni aday

kümeler oluşturulur. Aday kümeler içindeki gereksiz kümelerin silinmesi için budama tekniği kullanılır. Budama da kümenin sahip olması gereken minimum destek seviyesi göz önüne alınır ve bu destek seviyesinden küçük olanlar geniş nesne kümesinde çıkartılır (Silahtaroglu, 2008).

SETM Algoritması

SETM algoritması AIS algoritması gibi veritabanını tarar ve geniş nesne kümeleri oluşturur. Bu sırada AIS algoritmasından farklı olarak nesne isimleri ile birlikte nesneye atanan bir kimlik numarası (TID) tutar. Nesnelerin sıralanmasında tutulan bu kimlik numaraları ile gerçekleştirilir (Silahtaroglu, 2008).

Apriori Algoritması

Apriori Algoritması, birliktelik kuralları oluşturulurken sıkça tekrarlanan öğeleri bulmak için kullanılan en temel yöntemdir (Gursoy, 2009).

Geniş nesne kümeleri belirlenirken, her bir nesnenin destek seviyesi hesaplanarak kullanıcı tarafından belirlenen minimum destek seviyesi ile karşılaştırılır. Bu karşılaştırma sonucunda geniş olmayan kümeler budanarak aday nesne kümeleri oluşturulur ve bu işlem başka geniş nesne kümeleri bulunamayınca kadar devam eder. Aday kümeler oluşturulurken tüm veri tabanı tekrar taranmaz sadece kendinden önce oluşturulan geniş küme dikkate alınarak işlem yapılır (Silahtaroglu, 2008). Çizelge 2.2’de Apriori algoritmasına ait notasyon ve kaba kod aşağıda yer almaktadır (Agrawal and Srikant, 1994):

Çizelge 2.2. Apriori algoritması notasyonu.

k -itemset	k elemanlı küme
L_k	k elemanlı Geniş Küme (minimum desteği sağlayanlar) Her bir öğe iki alana sahiptir. i) eleman, ii) destek sayısı
C_k	k elemanlı Aday Küme (potansiyel geniş küme) Her bir öğe iki alana sahiptir. i) eleman, ii) destek sayısı
\bar{C}_k	İşlemleri oluştururken verilen TID ile birlikte adayların yer aldığı k elemanlı Aday Küme

Apriori Algoritması kaba kodu şu şekildedir (Agrawal and Srikant, 1994):

```
 $L_1 = \{\text{Geniş 1-elemanlı kümeler}\}$ 
for ( $k = 2; L_{k-1} \neq 0; k++$ ) do begin
   $C_k = \text{apriori-gen}(L_{k-1})$  // yeni adaylar
  forall işlemler  $t \in D$  do begin
     $C_t = \text{altküme}(C_k, t);$  //  $t$  içindeki adaylar
    forall adaylar  $c \in C_t$  do
       $c.\text{count}++;$ 
  end
   $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
end
Answer =  $\bigcup_k L_k;$ 
```

Apriori-gen fonksiyonuna ait birleştirme ve budama kaba kodları şu şekildedir (Agrawal and Srikant, 1994):

Birleştirme Adımı

Bu aşamada birleştirme işlemi yapılır.

insert into C_k

select $p.\text{item}_1, p.\text{item}_2, \dots, p.\text{item}_{k-1}, q.\text{item}_{k-1}$

from $L_{k-1} p, L_{k-1} q$

where $p.\text{item}_1 = q.\text{item}_1, \dots, p.\text{item}_{k-2} = q.\text{item}_{k-2}, p.\text{item}_{k-1} < q.\text{item}_{k-1};$

Budama Adımı

Bu aşamada C aday kümesine ait $(k-1)$ -elemanlı alt kümelerinde geniş $L(k-1)$ kümesinde olmayan elemanlar çıkartılır.

forall $c \in C_k$ elemanlar **do**

forall c 'nin $(k-1)$ -altkümesi s **do**

if ($s \notin C_k$) **then**

delete c from C_k

AprioriTid Algoritması

AprioriTid algoritması SETM algoritmasına benzer bir şekilde her bir nesneyi işlem kimlik numarası ile tutar. Burada nesnenin bir kimlik numarası ile tutulması algoritmanın her seferinde veritabanını nesnelerin destek seviyesini belirlemek için taramıyor olmasıdır (Silahtaroglu, 2008).

AprioriTid Algoritması kaba kodu şu şekildedir (Agrawal and Srikant, 1994):

```
 $L_1 = \{\text{Geniş 1-elemanlı kümeler}\}$ 
 $\bar{C}_k = \text{database } D;$ 
for ( $k = 2; L_{k-1} \neq 0; k++$ ) do begin
   $C_k = \text{apriori-gen}(L_{k-1})$  // yeni adaylar
   $\bar{C}_k = 0;$ 
  forall kayıtlar  $t \in \bar{C}_{k-1}$  do begin
     $C_t = \{c \in C_{t-1} \mid (c - c[k]) \in t.\text{elemanlı küme} \wedge (c - c[k-1]) \in t.\text{elemanlı küme}\}$ 
    forall adaylar  $c \in C_t$  do
       $c.\text{count}++;$ 
      if( $C_t \neq 0$ ) then  $\bar{C}_k += \langle t.\text{TID}, C_t \rangle;$ 
  end
   $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
end
Answer =  $\bigcup_k L_k;$ 
```


AprioriTid Algoritması, geniş küme ve aday küme aşamalarına ait örnek bir gösterim Şekil 2.4’de yer almaktadır (Agrawal and Srikant, 1994):

Veritabanı		\bar{C}_1	L_1		
TID	Elemanlar	TID	Alt Küme	Elemanlar	Destek
100	1, 3, 4	100	{ {1}, {3}, {5} }	{1}	2
200	2, 3, 5	200	{ {2}, {3}, {5} }	{2}	3
300	1, 2, 3, 5	300	{ {1}, {2}, {3}, {5} }	{3}	3
400	2,5	400	{ {2},{5} }	{5}	3

C_2	\bar{C}_2	L_2		
Elemanlar	TID	Alt Küme	Elemanlar	Destek
{1,2}	100	{ {1,3} }	{1,3}	2
{1,3}	200	{ {2,3}, {2,5}, {3,5} }	{2,3}	2
{1,5}	300	{ {1,2}, {1,3}, {1,5}, {2,3}, {2,5}, {3,5} }	{2,5}	3
{2,3}	400	{ {2,5} }	{3,5}	2
{2,5}				
{3,5}				

C_3	\bar{C}_3	L_3		
Elemanlar	TID	Alt Küme	Elemanlar	Destek
{2, 3, 5}	200	{ {2, 3, 5} }	{2, 3, 5}	2
	300	{ {2, 3, 5} }		

Kaynak: Agrawal and Srikant (1994) s. 9’den uyarlanmıştır.

Şekil 2.4. AprioriTid algoritması geniş küme ve aday küme aşamaları.

Ardışık Zamanlı Örüntüler

Örüntü, sürekli devam eden ve tekrarlanan yapılara denilmektedir. Birliktelik kurallarından farklı olarak ardışık zamanlı örüntülerde zaman faktörü dikkate alınmaktadır. Örneğin bir müşterinin satın aldığı bir üründen belli bir süre sonra başka bir ürün satın alması ardışık zamanlı bir örüntü oluşturmaktadır (Gürsoy, 2009).

2.5.1.2. Kümeleme Analizi

Kümeleme analizi, araştırma konusu olan birey veya nesneleri, aralarındaki benzerlikleri ile bir araya getirip kümeler oluşturmayı amaçlar. Oluşturulan kümelerin özellikleri, küme içinde maksimum derecede homojen olurken, diğer kümeler ile maksimum derecede heterojen olur (Gürsoy, 2009). Özkan (2013) da benzer şekilde kümelemeyi birbirine benzeyen verileri ayırma işlemi olarak tanımlamaktadır.

Sınıflandırma ile kümeleme analizi arasındaki en belirgin fark, sınıflandırmada sınıfların önceden belli olmasıyken kümeleme analizinde sınıfların eldeki veri setine göre oluşmasıdır (Gürsoy, 2009).

Veriler birbirinden ayrılırken genellikle verilerin birbiri arasındaki uzaklıktan yararlanılmaktadır (Özkan, 2013).

Uzaklık ve Benzerlik

Veritabanındaki iki kaydın birbirine benzer olup olmadığı diğer kayıtlara olan uzaklığı ile belirlenmektedir (Gürsoy, 2009).

Kümeleme analizinde kullanılan uzaklık ölçülerinden en önemlileri şunlardır (Şentürk, 2006):

- Öklid Uzaklığı: Uygulamalarda en çok bu uzaklık ölçüsü kullanılmaktadır. Şekil 2.1’de görüldüğü i ve j noktaları için uzaklık (d_{ij}) Pisagor teoreminin bir uygulaması olarak hesaplanmaktadır (Özkan, 2013).

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

- Minkowski Uzaklığı: İki den fazla veri arasındaki uzaklığın hesaplanmasında kullanılır (Özkan, 2013).

$$d_{ij} = \sum_{k=1}^p \left[|x_{ik} - x_{jk}|^\lambda \right]^{1/\lambda} \quad \lambda \geq 1 \quad (2.2)$$

- Manhattan (City-Blok) Uzaklığı: Bu uzaklık ölçüsünde veriler arasındaki mutlak uzaklıkların toplamından yararlanılmaktadır (Özkan, 2013).

$$d_{ij} = \sum_{k=1}^p (|x_{ik} - x_{jk}|) \quad \lambda = 1 \quad (2.3)$$

- Mahalanobis D^2 Uzaklığı: Veriler arasındaki korelasyonların da dikkate alınması ile hesaplanan uzaklık ölçüsüdür (Şentürk, 2006).

$$d_{ij} = D^2 = (x_i - x_j)' \Sigma^{-1} (x_i - x_j) \quad (2.4)$$

Verilerin farklı uzaklıklara göre kümelere ayrılması her zaman için kümelerin gerçekten farklı özelliklere sahip oldukları anlamına gelmeyebilir. Bu durumda oluşturulan kümeler arasındaki benzerlik de ölçülmelidir. Benzerlik, iki veri arasındaki yakınlığı göstermektedir. Kümeleme analizinde kullanılan benzerlik ölçülerinden bazıları şunlardır (Silahtaroglu, 2008):

- Dice Benzerliği:

$$benzerlik(X_m, X_j)_{DICE} = \frac{2 \sum_{i=1}^n x_{mi} x_{ji}}{\sum_{i=1}^n x_{mi}^2 + \sum_{i=1}^n x_{ji}^2} \quad (2.5)$$

- Jaccard Benzerliği:

$$benzerlik(X_m, X_j)_{JACCARD} = \frac{\sum_{i=1}^n x_{mi} x_{ji}}{\sum_{i=1}^n x_{mi}^2 + \sum_{i=1}^n x_{ji}^2 - \sum_{i=1}^n x_{mi} x_{ji}} \quad (2.6)$$

- Cosine Benzerliği:

$$benzerlik(X_m, X_j)_{COSINE} = \frac{2 \sum_{i=1}^n x_{mi} x_{ji}}{\sqrt{\sum_{i=1}^n x_{mi}^2 \sum_{i=1}^n x_{ji}^2}} \quad (2.7)$$

- Overlap Benzerliği:

$$benzerlik(X_m, X_j)_{OVERLAP} = \frac{2 \sum_{i=1}^n x_{mi} x_{ji}}{\min(\sum_{i=1}^n x_{mi}^2, \sum_{i=1}^n x_{ji}^2)} \quad (2.8)$$

Kümeleme analizinde kümeler arası uzaklık ve benzerlik konularında aşağıdaki formüllerden yararlanılmaktadır (Silahtaroglu, 2008):

- Merkez:

$$Merkez = X_0 = \frac{\sum_{i=1}^N x_{mi}}{N} \quad (2.9)$$

- Yarıçap:

$$Yarıçap = R = \sqrt{\frac{\sum_{i=1}^N (x_{mi} - X_0)^2}{N}} \quad (2.10)$$

- Çap:

$$Çap = D = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^N (x_{mi} - x_{mj})^2}{N(N-1)}} \quad (2.11)$$

Kümeleme analizinde uygulanacak adımlar şunlardır (Gürsoy, 2009):

- Problemin formüle edilmesi
- Bir benzerlik ölçüsünün belirlenmesi
- Kümeleme tekniğinin belirlenmesi
- Seçilen teknik için hangi kümeleme yönteminin kullanılacağına belirlenmesi
- Küme sayısına karar verilmesi
- Kümelerin yorumlanması ve tanımlanması
- Kümelerin geçerliliğinin test edilmesi

Kümeleme teknikleri kümelerin oluşturulma şekline göre ayrılabilirler gibi veri tipine veya yapılacak çalışmanın amacına göre de farklılık göstermektedir. Genel olarak kümeleme teknikler iki gruba ayrılmaktadır:

A. Hiyerarşik Yöntemler

- Toplaşım Kümeleme Algoritmaları
 - Merkezi Kümeleme (Centroid) Yöntemi
 - Tek Bağlantı (En Yakın Komşu – Single Linkage) Yöntemi
 - Tam Bağlantı (En Uzak Komşu – Complete Linkage) Yöntemi
 - Ortalama Bağlantı (Average Linkage) Yöntemi

- Ward Yöntemi
 - Bölünür (Divisive) Kümeleme Algoritmaları
 - Bölünmüş (Splinter-Average Distance) Ortalamalar Yöntemi
 - Otomatik Etkileşim Dedektörü (Automatic Interaction Detection - AID) Yöntemi
- B. Hiyerarşik Olmayan Yöntemler
- K-Ortalamlar (K-Means) Yöntemi
 - Metoid Parçalama Yöntemi
 - Yığıma Kümeleme Yöntemi
 - Bulanık Kümeleme Yöntemi

Hiyerarşik Yöntemler

Hiyerarşik yöntemlerde veritabanındaki elemanlarından bir küme ağacı oluşturulur. Ağaçlar düğüm noktalarına sahiptir ve dallara ayrılarak yapraklarla bu dallanma son bulur. Hiyerarşik yöntemler ağaç yapısının dallanmasına göre farklılık göstermektedir. Başlangıçta her bir elemanın küme olarak ele alınıp, ortak özellikleri olan diğer elemanlar ile birleştirilerek yeni kümeler oluşturulması ve sonucunda tek bir kümeye ulaşılmasını sağlayan algoritmalara Toplaşım Kümeleme Algoritmaları denilmektedir. Başlangıçta veritabanının tek bir küme olarak ele alınarak, her taramada birbirine benzeyen elemanların küme dışına çıkarılarak, başlangıçta belirlenen küme sayısı kadar küme oluşturulmasını sağlayan algoritmalara Bölünür Kümeleme *Algoritmaları* denilmektedir (Silahtaroglu, 2008).

Hiyerarşik yöntemlerde kümeler, her bir eleman başlangıçta bir küme olacak şekilde belirlenip, bütün elemanlar tek bir küme olacak şekilde birleşinceye kadar kendisine en yakın küme ile birleşmesi ile oluşmaktadır (Gürsoy, 2009). Küme sayısının belirlenmesinde kullanılan en pratik yollardan birisi, özellikle küçük örneklemeler için, şu şekildedir (Gürsoy, 2009):

$$K = \left(\frac{n}{2}\right)^{\frac{1}{2}} \quad (2.12)$$

$K = \text{küme sayısı}, \quad n = \text{gözlem sayısı}$

Küme sayısının belirlenmesinde kullanılan en pratik yollardan bir diğeri ise, özellikle büyük örneklem için, şu şekildedir (Gürsoy, 2009):

$$M = k^2 |W| \quad (2.13)$$

$M =$ küme sayısı, $W =$ grup içi kareler toplamı matrisi

Toplaşım Kümeleme Algoritmaları

- Merkezi Kümeleme (*Centroid*) Yöntemi
- Tek Bağlantı (En Yakın Komşu – *Single Linkage*) Yöntemi
- Tam Bağlantı (En Uzak Komşu – *Complete Linkage*) Yöntemi
- Ortalama Bağlantı (*Average Linkage*) Yöntemi
- Ward Yöntemi

Bölünür (Divisive) Kümeleme Algoritmaları

- Bölünmüş (*Splinter-Average Distance*) Ortalamalar Yöntemi
- Otomatik Etkileşim Dedektörü (*Automatic Interaction Detection - AID*) Yöntemi

Hiyerarşik Olmayan Yöntemler

Hiyerarşik olmayan yöntemlerde kümeler, sayısı daha önceden bilindiği varsayılan sayıda ayrılırken, birimlerin uygun oldukları kümede toplanmaları hedeflenmektedir (Gürsoy, 2009). Hiyerarşik olmayan kümeleme yöntemlerinden en fazla kullanılanı K-Ortalamlar yöntemidir. Bunun yanında Metoid Parçalama yöntemi, Yığma Kümeleme yöntemi, Bulanık Kümeleme yöntemi de bulunmaktadır (Gürsoy, 2009).

K-Ortalamlar (K-Means) Yöntemi

Bu yöntemde her bir veri önceden belirlenen küme sayısına bağlı olarak kendisine en yakın kümeye atanır. Buradaki kümeye yakınlık veya uzaklık yukarıda bahsedilen formüllerden yararlanılarak hesaplanmaktadır (Şentürk, 2006).

2.5.2. Tahmin Edici Modeller

Tahmin edici modeller, verilerden bazı tahminlerde bulunmaya yarar. Sınıflandırma ve istatistiksel tahmin modelleri olmak üzere iki başlık altında incelenebilmektedir (Gürsoy, 2009).

2.5.2.1. Sınıflandırma

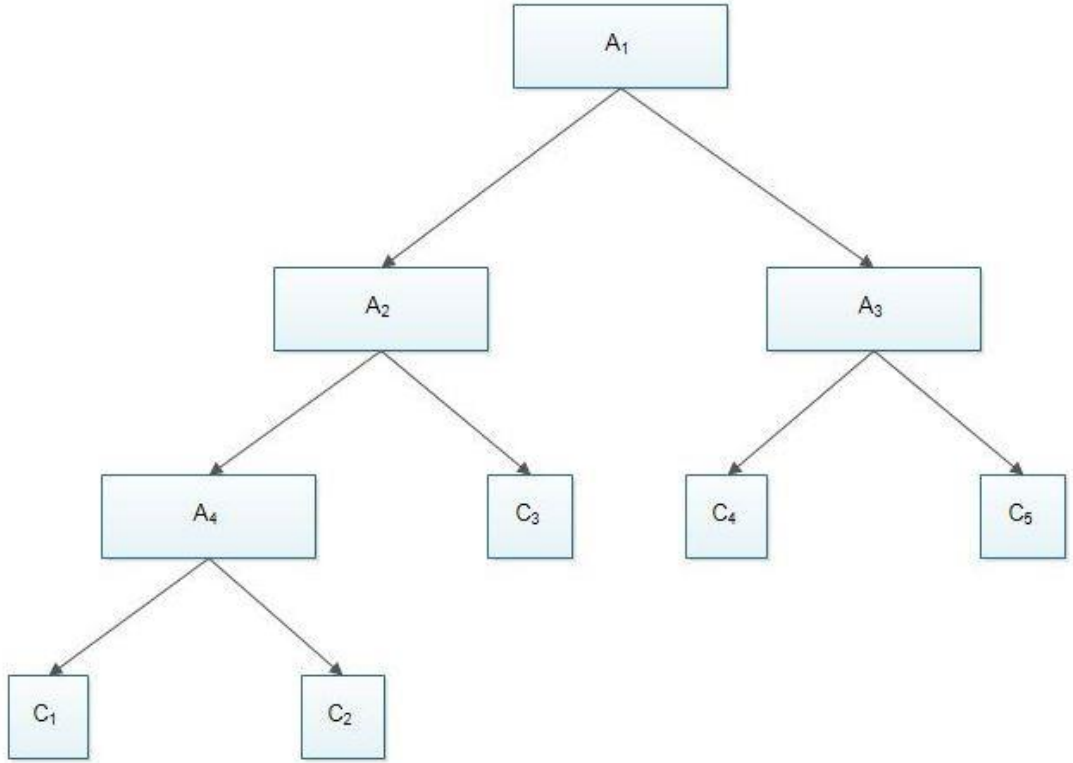
Sınıflandırmada amaç yeni bir nesnenin belirli sınıflar içinde hangi sınıfa ait olup olmadığını belirleyecek bir sınıflayıcı oluşturmak ve belirlenen sınıfların alınan verilerin sınıflandırılması için model geliştirilmesinde kullanılmasıdır (Gürsoy, 2009).

Karar Ağaçları

Karar ağaçları ilk olarak 1973 yılında Bierman ve Friedman tarafından önerilmiştir. Bu teknikte en kuvvetli ilişkiye sahip bağımsız değişken bulunduğu veri kümesi ikiye ayrılarak bölünmelerin tamamlanmasına kadar devam ettirilmesi ile bir ağaç yapısı oluşturulmaktadır (Şentürk, 2006).

Karar ağaçları kök düğüm denilen başlangıç düğümü ile başlar ve dallara ayrılarak ya farklı düğümlere ayrılır ya da yapraklara ulaşıp son bulur. Burada amaç en kısa yoldan ağacın istenilen yanıtı veya sınıfa ulaşmasıdır (Silahtaroglu, 2008).

Şekil 2.5'te A_1 kök düğüm, A_2 , A_3 ve A_4 ise diğer düğümleri göstermektedir. C_1 , C_2 , C_3 , C_4 ve C_5 ise dalları göstermektedir.



Kaynak: Silahtaroglu (2008) s. 47'den alınmıştır.

Şekil 2.5. Örnek karar ağacı yapısı.

Karar ağaçlarında kullanılan temel algoritma yapısı şu şekildedir (Şentürk, 2006):

- Başlangıçta tüm noktalar ağacın kökünde toplanır.
- Tüm örnekler aynı sınıfa ait olması durumunda, düğüm yaprak haline dönüşür ve aynı ismi alır.
- Aksi durumda test yapılır ve bir bölümlenme meydana gelir.
- Kategorik veriler kullanıldığı için sürekli değişkenlerin kesikli değişken haline dönüştürülür.
- Bir dal, test değişkeninin tüm değerleri için oluşturulmakta ve örneklem buna göre yapılır.
- Örneklemin her bölümlenmesinde aynı algoritma yinelenir.
- Bir düğümde bulunan bütün örneklem aynı sınıfa ait olunca, bölümlenmenin yapılacağı başka değişken kalmayınca veya başka örneklem kalmayınca bölümlenme sona erer.

Karar ağaçlarında kullanılan önemli algoritmalar şunlardır (Gürsoy, 2009) (Şentürk, 2006):

- CART (C&RT): C&RT (Sınıflandırma ve Regresyon Ağaçları) tekniği karar ağaçlarını en iyi dallara ayırmak için ID3 algoritması gibi entropiden yararlanır (Silahtaroglu, 2008).
- CHAID: Bu metod 1980 yılında Kaas tarafından geliştirilmiştir. En uygun bölümlendirmeyi oluşturmak için ki kare testini kullanmaktadır (Albayrak ve Yılmaz, 2009).
- C4.5: İlk olarak 1993 yılında Quinlan tarafından ortaya atılmıştır. Bu algoritmada, ID3 algoritmasının değişkenleri birçok alt bölüme ayırması ile ortaya çıkabilecek aşırı öğrenmeyi önlemek için *kazanım* yerine *kazanım oranı* kullanılmaktadır. *Kazanım Oranı* ve *Ayırma Bilgisi* şu şekilde hesaplanmaktadır (Silahtaroglu, 2008):

$$Kazanım Oranı (D, S) = \frac{Kazanım(D, S)}{Ayırma Bilgisi(D, S)} \quad (2.14)$$

$$Ayırma Bilgisi (D, S) = H\left(\frac{|D_1|}{|D|}, \frac{|D_2|}{|D|}, \dots, \frac{|D_s|}{|D|}\right) \quad (2.15)$$

- QUEST: (*Quick, Unbiased, Efficient, Statistical Trees*) Wei-Yin Loh ve Yu-Shan Shih tarafından veri madenciliği ve sınıflandırma için geliştirilmiş iki bölmeli karar ağacı algoritmasıdır (Shih, 2014).
- ID3: İlk olarak J. Ross tarafından Sydney Üniversitesinde geliştirilmiştir. ID3 algoritması sınıflandırma yaparken en ayırıcı özelliğe sahip değişkeni bulmak için *entropi* kavramından yararlanır. Entropi 0 ile 1 arasında değerler almaktadır. Bütün olasılıkların eşit olduğu durumlarda entropi maksimum olmaktadır (Silahtaroglu, 2008).

$$H(p_1, p_2, \dots, p_n) = \sum \left(p_i \log \left(\frac{1}{p_i} \right) \right) \quad (2.16)$$

$\langle p_1, p_2, \dots, p_n \rangle$ her bir değişkenin olasılığı

ID3 algoritması ile doğru sınıflandırma yapabilmek için entropi ile bölümlendirme yapıldıktan sonraki entropi arasındaki fark kullanılarak düğüm noktalarına ve dallanmalara karar verilir. Bu aradaki farka *kazanım* denilir. Kazanım şu şekilde hesaplanılır (Silahtaroglu, 2008):

$$Kazanım(D, S) = H(D) - \sum_{i=1}^n P(D_i) H(D_i) \quad (2.17)$$

Her bir sınıf için hesaplanan kazanım miktarlarından en fazla olanı düğüm noktası olarak ele alınıp, her bir dallanma için aynı işlem tekrarlanmaktadır.

- **SLIQ:** SLIQ (Araştırmada Denetimli Öğrenme) algoritması 1996 yılında IBM Almaden Araştırma Merkezinde Mehta M., Agrawal R. ve Rissanen J. tarafından önerilmiştir.
- **SPRINT:** SPRINT (*Scalable Paralellizable Induction of Decision Trees*) algoritmasında bölme işlemi tek bir niteliğe göre yapılmaktadır (Emel ve Taşkın, 2005).

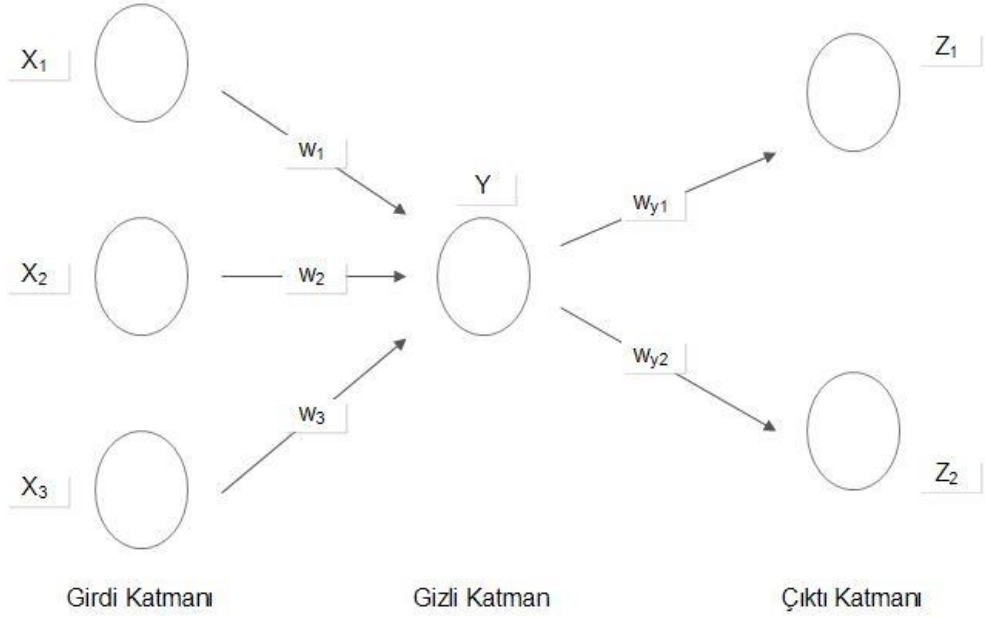
Yapay Sinir Ağları

Yapay sinir ağları, öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri, otomatik olarak gerçekleştirebilmek amacıyla geliştirilen bilgisayar sistemleridir (Öztemel, 2012).

Yapay sinir ağları biyolojik sinir ağlarından esinlenerek geliştirilmiştir. Yapay sinir hücrelerinin birbirleriyle çeşitli şekillerde bağlanmasından oluşurlar. Yapay sinir ağlarının en belirgin özellikleri birbirine bağlı nöronlar, bağlantılar arasındaki ağırlıkların belirlenmesi ve ateşleme fonksiyonudur (Silahtaroglu, 2008).

Yapay sinir ağları, ağıın yapısına göre İleri Beslemeli (*feed forward*) ve Geri Beslemeli (*feed back*), öğrenme türüne göre ise Denetimli Öğrenme (*supervised learning*) ve Denetimsiz Öğrenme (*unsupervised learning*) olmak üzere ikiye ayrılmaktadır (Gürsoy, 2009).

Şekil 2.6’da örnek bir yapay sinir ağı yapısı yer almaktadır.



Kaynak: Silahtaroglu (2008) s. 69’ dan alınmıştır.

Şekil 2.6. Yapay sinir ağı yapısı.

Genetik Algoritmalar

Genetik algoritmalar ile ilgili kavramlar şunlardır (Gürsoy, 2009):

- **Kromozom ve Gen:** Genetik algoritma ile çözülmesi istenilen problemin her bir çözümünü göstermektedir. Bir probleme ait birçok çözüm arasından en iyisi genetik algoritma ile bulunmaya çalışılır. Çözüme ait her bir özellik gen olarak adlandırılmaktadır.
- **Çözüm Havuzu:** Problemin çözümü için kullanılan ve rastgele olarak belirlenen çözüm kümesinden oluşmaktadır.
- **Çaprazlama (Cross-Over):** Çözüm havuzunda bulunan çözümlerin birleştirilerek oluşturulan yeni çözümlerdir.
- **Mutasyon (Mutation):** Çaprazlama ile ulaşılamayan çözümlere ait oluşturulan kromozomun bir özelliğinin/genin değiştirilmesi işlemidir.
- **Uygunluk Fonksiyonu:** Çözümlerin uygunluğunun ölçülmesini sağlayan fonksiyondur.

- Yeniden Üretim: Üretilen yeni çözümlerin çözüm havuzundaki gereksiz görülen çözümlerin atılması sonucu tekrar çaprazlanarak yeni çözümlerin üretilmesi işlemidir.

Genetik algoritmalar genellikle üç alanda uygulanmaktadır. Deneysel çalışmalarda optimizasyon amacıyla, endüstriyel uygulamalarda ve sınıflandırmada kullanılmaktadır (Gürsoy, 2009).

2.5.2.2. İstatistiksel Tahmin Modelleri

İstatistiksel tahmin yöntemleri iki alt başlıkta incelenebilmektedir. Bunlardan ilki Regresyon Analizi bir diğeri de Diskriminant Analizidir (Gürsoy, 2009).

Regresyon Analizi

Herhangi bir bağımlı değişkenin bir veya daha fazla bağımsız değişken ile arasındaki ilişkinin matematiksel olarak gösterilmesine regresyon analizi denilmektedir. Bağımlı değişken ile bağımsız değişkenler arasındaki ilişki gücü yüzde olarak gösterilen korelasyon katsayısı bulunmaktadır. Regresyon analizinde değişkenler arasındaki ilişki doğrusal olduğunda Doğrusal Regresyon Modeli, doğrusal olmadığına ise Doğrusal Olmayan Regresyon Modeli olarak ifade edilmektedir (Gürsoy, 2009).

Basit Doğrusal Regresyon Analizi

Bağımlı bir değişkenin bağımsız tek bir değişken ile olan ilişkisinin doğrusal fonksiyon ile ifade edilmesidir. Basit doğrusal regresyon denkleminin genel ifadesi şu şekildedir (Gürsoy, 2009):

$$y = a + bx + \epsilon \quad (2.17)$$

y = bağımlı değişken

x = bağımsız değişken

a

= Bağımsız değişkenin sıfır değerini alması durumunda bağımlı değişkenin alacağı değer

b = Bağımsız değişkendeki değişimin, bağımlı değişkende meydana getirdiği değişim

ε = hata

Çoklu Doğrusal Regresyon Analizi

Bağımlı bir değişkenin bağımsız birden fazla değişken ile olan ilişkisinin doğrusal fonksiyon ile ifade edilmesidir. Basit doğrusal regresyon denkleminin genel ifadesi şu şekildedir (Gürsoy, 2009):

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n + \varepsilon \quad (2.18)$$

y = bağımlı değişken

x = bağımsız değişken

n = bağımsız değişken sayısı

a = Bağımsız değişkenin sıfır değerini alması durumunda bağımlı değişkenin alacağı değer

b = Bağımsız değişkendeki değişimin, bağımlı değişkende meydana getirdiği değişim

ε = hata

Çoklu regresyon analizi genel olarak şu amaçlar için yapılır (Can, 2013):

- Bağımlı bir değişken üzerinde etkisi olan birden çok bağımsız değişken ile bu bağımlı değişken arasındaki ilişkileri açıklayan bir model oluşturmak
- Bir değişken üzerinde etkisi olan birden çok değişkenin etkilerinin derecelerini ve bu değişkendeki değişimi açıklamada öncelikleri belirlemek
- Bir değişken üzerinde etkisi olan bir dizi değişkenin etkilerinin en uygun değerlerini içeren ilişkiler denklemini ortaya koyarak, istenmeyen etkileri kontrol etmek, istenen etkileri belirginleştirmektir.

Çoklu doğrusal regresyon analizinin doğru sonuçlar verebilmesi için (Can, 2013):

- En az aralık ölçeğinde olan değişkenler normal dağılım sergilemelidir.
- Bağımsız değişkenlerin her birinin bağımlı değişkenle arasındaki ilişki doğrusal olmalıdır.
- Bağımlı değişkenler birbirinden bağımsız olmalıdır.

- Yordama işleminde, tahmin edilen değerlerle gözlenen değerler arasındaki farklar normal dağılım sergilemelidir.

Lojistik Regresyon Analizi

Çoklu regresyon analizinin temel varsayımları genellikle araştırmalarda sağlanamamaktadır. Özellikle verilerin dağılımının normal olmaması ve analiz edilecek değişkenlerin sürekli veya kesirli değişken olmaması sonucu analizde zorluklar ortaya çıkmaktadır. Çoklu regresyon analizinin bu varsayımlarını içermeyen ve kullanımı daha basit olan lojistik regresyon analizi kullanılmaktadır (Gürsoy, 2009).

LRA'da kullanılan bazı tanımlar şunlardır (Sümbüloğlu, 2014):

- Odds: Bir olayın veya başarının görülme olasılığının (p), görülme olasılığına ($1-p$) olan oranıdır.
- Odds Ratio (OR): İki odds'un birbirine olan oranına denir.
- Lojit: OR'nin doğal logaritmasıdır. Asimetrik olan OR'yi simetrik hale getirmek için kullanılır.

Lojistik regresyon analizinde bağımlı değişken kategorik olmakla birlikte bağımlı değişkenler hem kategorik hem de sürekli veya kesirli veri olabilmektedir. Lojistik regresyon modeli şu şekildedir (Gürsoy, 2009):

$$L = \ln \frac{P_i}{1-P_i} = b_0 + b_1 X_i + e_i \quad (2.19)$$

Tek Bağımsız değişken için;

$$L = \ln \frac{P_1}{1-P_1} = b_0 + b_1 X_1 + e_1 \quad \text{veya} \quad P = \frac{1}{1+e^{-(b_0+b_1 X_1)}} \quad (2.20)$$

Modeli doğrudan olasılık (P) şeklinde aşağıdaki gibi gösterebiliriz:

$$P = \frac{(e^{\alpha + \beta x})}{1 + (e^{\alpha + \beta x})} \quad (2.21)$$

P = olasılık

x = bağımsız değişken

a

= Bağımsız değişkenin sıfır değerini alması durumunda bağımlı değişkenin alacağı değer

β = Bağımsız değişkendirdeki değişimin, bağımlı değişkende meydana getirdiği değişim

Lojistik regresyonda değişken seçimi için iki temel yöntem kullanılmaktadır. Bu yöntemlerden ilki standart, diğer ise adımsal yöntemdir. Standart yöntem *enter* olarak, adımsal yöntem ise *stepwise* olarak isimlendirilmektedir. Adımsal yöntemler de kendi arasında ikiye ayrılmaktadır. Bunlardan birincisi geriye doğru (*backward*) diğer ise ileriye doğru (*forward*) yöntemlerdir. Standart yöntemlerde tüm ortak değişkenler yer alır ve bir bütün halinde parametre kestirimleri hesaplanır. Adımsal yöntemlerde ise değişkenler teker teker modele dâhil edilerek veya çıkartılarak model oluşturulur. İleriye doğru yöntemde ilk önce sadece sabit değerden başlayarak her seferinde bir değişken daha eklenerek model oluşturulurken, geriye doğru yöntemde ise başlangıçta tüm değişkenlerden yola çıkılır ve her seferinde bir değişken çıkartılarak model oluşturulur (Çokluk, 2010).

Diskriminant Analizi

Diskriminant analizi, önceden sınıflandırılmış iki veya daha fazla grubu birbirinden ayırmaya yarayan özellikleri belirleyip dışarıdan alınan bir gözlemin hangi gruba ait olabileceğini gösteren çok değişkenli bir analizdir (Gürsoy, 2009).

Diskriminant analizinin kullanım amaçları şunlardır (Gürsoy, 2009):

- Önceden belirlenen grupları birbirinden ayırmaya yarayan en iyi bağımsız değişkenlerin ayırma fonksiyonlarını belirlemek.
- Bağımsız değişkenler için, gruplar arası bir farkın olup olmadığını test etmek
- Gruplar arası farkı belirleyen en iyi değişkeni saptamak.
- Fonksiyonun tahmin etme yeteneğinin eldeki gözlemleri doğrulama derecesini saptamak.

- Yeni gözlemlerin hangi gruba üye olacağı konusunda öngöründe bulunmak.

Diskriminant fonksiyonu şu şekildedir (Gürsoy, 2009):

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \dots W_nX_{nk} \quad (2.21)$$

Z_{jk} = Diskriminant değeri

a = Kesim noktası

W_i = Bağımsız değişkenlerin ağırlıkları

X_{ik} = Bağımsız değişkenler

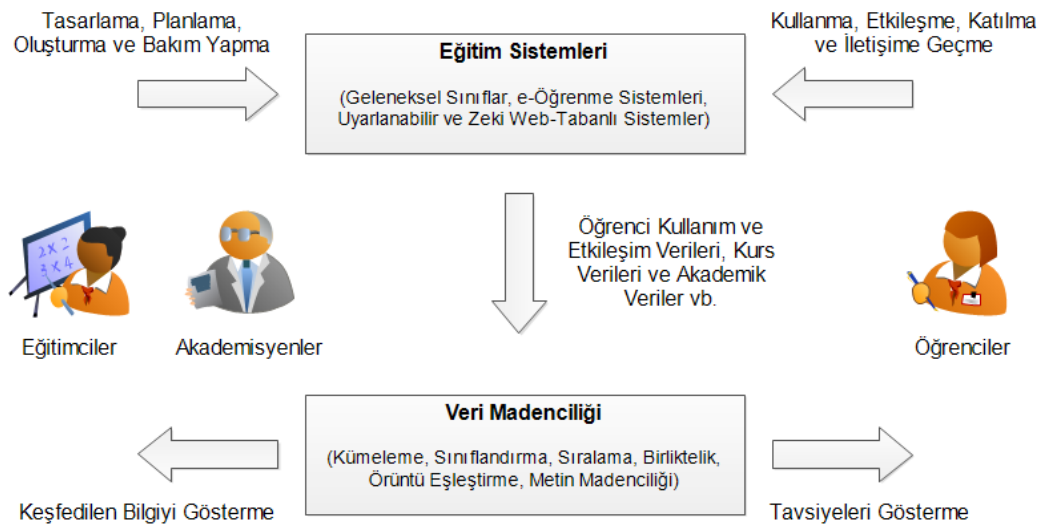
BÖLÜM 3

EĞİTİMDE VERİ MADENCİLİĞİ UYGULAMLARI

3.1. EĞİTİM VE VERİ MADENCİLİĞİ

Temel eğitimden yüksek öğretime kadar birçok alanda ders içi ve ders dışı etkinliklere ait veriler kayıt altında tutulmaktadır. Bu veriler bilgi ve iletişim teknolojilerindeki gelişmelere bağlı olarak daha çok çeşitlenmekte ve daha büyük hacimlerde saklanabilmektedir.

Eğitimde VM yöntemlerinin uygulanması iki farklı eğitim sisteminden gelen veriler göz önünde bulundurularak değerlendirilebilir. Bu eğitim sistemlerinden birincisini geleneksel sınıflardan gelen veriler, diğerini ise uzaktan eğitim sisteminden gelen veriler oluşturmaktadır. Uygulanan VM yöntemlerinin belirlenmesinde bu sistemlerin farklı hedefleri ve veri kaynakları olduğunun da değerlendirilmesi gerekmektedir (Romero ve Ventura, 2007).



Not: Romero and Ventura (2007) s.136'dan uyarlanmıştır.

Şekil 3.1. Eğitim sistemlerinde veri madenciliği uygulama döngüsü.

Şekil 3.1’de görüldüğü üzere dersteki öğrenci etkileşimlerine ait veriler üzerinde VM yöntemleri uygulanmakta ve sonuçta öğrenciye gerekli tavsiyeler gösterilirken, akademisyen ve öğretmenlere de keşfedilen bilgiler sunulmaktadır.

Chen, Liu, Ou, and Liu, (2000) geleneksel sınıflarda öğretmenler dersleri öğrencilerin öğrenme süreçlerini gözlemleyerek geliştirmeye çalıştıklarını ve öğrenci performanslarının gözlemlere ve öğrencilere ait kâğıtlara göre analiz edildiğini belirtmektedirler. Ayrıca uzaktan eğitim sistemlerindeki web kayıtlarından ve diğer verilerden öğrenci performanslarının belirlenebileceği fakat büyük verilerin analiz edilmesinin öğretmenler açısından zor olacağı bunun için de VM yöntemlerinden yararlanılması gerektiğini ifade etmektedirler. Zaiane and Luo (2001) eğitimde yapılan bu VM çalışmalarının amacı öğrencilerin daha aktif öğrenen öğrenciler haline getirilmesi çabalarıdır.

Literatürde yer alan ve eğitimde kullanılan VM yöntem ve tekniklerin uygulandığı çalışmalar; müfredat, derse kayıt ve devam, akademik başarı, öğrenci davranışları ve performansları başlıkları altında incelenmiştir.

3.1.1. Müfredat Düzenlemeleri

Becker, Ghedini, and Terra (2000), Brezilya üniversitelerindeki müfredat değişikliklerinin öğrencileri nasıl etkilediğini anlamak ve açıklamak amacıyla VTBK uygulaması gerçekleştirmişlerdir. Araştırmada 5 farklı programa (Mimarlık, Makine Mühendisliği, İnşaat Mühendisliği, Muhasebe ve İktisat) ait mezun öğrenci verileri kullanılmış ve müfredat değişikliklerinin etkisi birliktelik kuralları ve sınıflandırma gibi VM yöntemleri kullanılarak incelenmiştir. Çalışmanın sonunda bireysel müfredat değişikliklerinin kısmen faydalı olsa da çoğu zaman öğrencileri etkilemediği sonucuna varılmıştır. Ancak, müfredat değişikliklerinin, öğrencilerin mezun olmaları için gerekli olan şartları sağlayabilmesi için daha fazla ders almalarına neden olduğu fakat mezun olma süresini etkilemediği sonucuna varılmıştır.

3.1.2. Derse Kayıt ve Devam

Sanjeev and Zytchow (1995) yapmış oldukları çalışmada öğrencilerin üniversiteye devam durumlarını anlama ve devamlılığı arttırmada yardımcı olacak bilgi keşfini sağlamayı amaçlamışlardır. Çalışmada bağımsız değişkenler üç kategoride incelenmiştir. Birincisi; öğrencilerin demografik bilgileri, ikincisi öğrencilerin performansları ve sonuncusu ise öğrencilerin üniversite performanslarıdır. VM yöntemleri uygulanarak değişkenler arasında çeşitli örüntüler bulunmuştur. Bu örüntülerden bir tanesi, lise mezuniyet notu ile üniversitede alınan ders kredileri arasındaki örüntüdür. Bir diğeri öğrencilere verilen burs miktarı ve süresinin derse katılım oranına etkisidir. Sonuncusu ise öğrencilerin aldıkları ek derslerin derse katılımlarını nasıl etkilediğidir. Çalışmanın sonucunda iyi lise öğrencilerinin çok kredili dersler için iyi birer kaynak olduğu fakat bir kısmının dersleri bıraktığı, telafi derslerinin de öğrencilere bu konuda yardımcı olamadığı ve finansal yardımın da öğrencilerin derslere devamında yardımcı olmadığı bulguları bulunmuştur. Benzer bir çalışmayı Hamalainen, Suhonen, Sutinen, and Toivonen (2004) yapmış, VM sistemlerinin uzaktan eğitim sistemlerindeki derslere ait alıştırmaların geliştirilmesinde, dersin düzenlenmesinde ve muhtemel üniversite terkleri erken dönemde tanımlayabileceğini göstermişlerdir.

Luan (2002) yapmış olduğu çalışmada yüksek öğretim kurumlarının yapay sinir ağları, C&RT ve C5.0 gibi VM algoritmalarını kullanarak, öğrenci ve personel gibi kaynakları daha verimli kullanabileceklerini önermektedir. Sınıflandırma ve tahmin etme modelleri ile elde edilen bu bilgiler ile yüksek öğretim kurumları, öğrencilerin kayıt sildirme, ders saydırma ve ders tercihleri gibi bilgileri tahminde bulunabilecekleri politikalarında bu bilgilerin kendilerine yardımcı olabileceğini ifade etmişlerdir.

Ma, Liu, Wong, Yu, and Lee (2000) Singapur Eğitim Bakanlığının Üstün Yetenekliler Eğitim Programında yer alan ve ek ders alan öğrencilere ait verileri kullanarak derslerde başarısız olan ve ek derse ihtiyacı olan öğrencileri seçmek için VM yöntemleri ile bir yöntem geliştirmişlerdir. Araştırma ek derse ihtiyacı olan öğrencilerin genellikle zayıf not alan öğrencilerden seçildiği, bu seçimin olması

gerekenden daha fazla öğrenci seçilerek hem öğretmenlere hem de öğrencilere ilave yük getirdiğinden yola çıkılarak yapılmıştır. Çalışmanın sonunda geliştirilen model ile ek derse ihtiyacı olan öğrencilerin seçimleri daha başarılı bir şekilde gerçekleştirilmiştir.

3.1.3. Akademik Başarı

Erdoğan ve Timor (2005) Maltepe Üniversitesi öğrencilerinin üniversiteye giriş sınavından aldığı puanlar ile başarıları arasındaki ilişki analiz etmiştir. Kümeleme Analizi ve K-Means algoritmasının kullanıldığı çalışmada veri seti olarak da 2003 yılına ait 722 öğrenci verisi kullanılmıştır. Analiz sonuçlarında üniversite girişte öğrencinin sahip olduğu puanın yüzdelik dilimi ile başarı notlarının kümelenmesinde bir ilişki olduğu göstermektedir. Buna göre giriş yüzdesi düşük olan öğrencilerin yani üniversite giriş puanı yüksek olan öğrencilerin başarı puanları da yüksek olmaktadır.

Taşdemir (2012) üniversite öğrencilerinin başarısına etki eden faktörleri regresyon analizi ile tespit etmek için yapmış olduğu çalışmada, üniversiteye giriş yaşı ile not ortalaması ve öğrencinin kaldığı ders sayısı arasında negatif yönlü kuvvetli bir ilişki olduğunu ortaya koymuştur. Bu çalışmaya göre öğrencinin üniversiteye giriş yaşı arttıkça not ortalaması azalmakta fakat kaldığı ders sayısı da azalmaktadır. Yine aynı çalışmaya göre öğrencinin üniversite giriş puanı ile not ortalaması arasında pozitif yönlü kuvvetli bir ilişki olduğu ortaya konulmuştur. Höçük (2011) de uzaktan eğitim programlarına katılan önlisans öğrencilerin akademik başarısını, kişisel özelliklere ait faktörlerden, en fazla yordayan değişkenin yaş olduğunu ortaya koymuştur.

Göksu (2012) web tabanlı eğitim ile yüzyüze eğitim alan öğrencilerin başarıları arasında anlamlı bir fark olup olmadığını analiz etmek için deney ve kontrol grubu oluşturmuş ve eğitim öncesinde öntest, eğitim sonunda da sontest uygulayarak oluşturulan grupların başarıları arasındaki ilişkiyi incelemiştir. Yapılan çalışma sonucunda verilen eğitimin her iki grup içinde anlamlı olduğu anlaşılmıştır. WTE alan öğrencilerin başarıları ile yüzyüze eğitim alan öğrencilerin başarıları arasında anlamlı bir fark olup olmadığı test edilmiş ve WTE alan öğrencilerin daha başarılı

olduđu sonucuna varılmıřtır. řen ve Uçar (2012) ise uzaktan eđitim đrencileri ile yzyzye eđitim alan đrencilerin derslerdeki bařarılarını karřılařtırılmıř ve yzyzye eđitim alan đrencilerin daha bařarılı olduđu sonucuna varmıřtır.

Noble and Sawyer (2014) üniversite birinci sınıf đrencilerinin akademik bařarılarını aıklamada ACT notlarının, lise mezuniyet notlarından daha iyi yordadıđını ifade etmiřtir. Ayrıca hem ACT hem de lise mezuniyet notunun beraber kullanımının ise akademik bařarıyı yordamada, notların tek bařına kullanımlarından daha fazla aıklayıcı olduđu sonucuna varmıřtır. Myers and Pyles (1992) yapmıř oldukları alıřmada ACT notunun akademik bařarıyı aıklamada tek bařına yeterli olmadıđını, ACT notunun lise mezuniyet notu ile birlikte iyi bir yordayıcı olduđunu ifade etmiřlerdir. Ullah and Wilson (2009) ACT notu ve yařın akademik bařarıyı yordamada nemli bir yordayıcı olduđunu ortaya koymuřlardır. Buna gre ACT notundaki %1'lik bir artıř akademik bařarıyı %5 arttırmakta, đrenci yařındaki bir yıllık fark ise đrencinin akademik bařarısını %2 arttırmaktadır.

Akhun (1980) alıřmasında yksek okula eleme ve giriř sınavlarından alınan puanlaıyla giren đrencilerin akademik bařarıları ile birinci sınıf not ortalamaları arasındaki korelasyon katsayısı dřk ıkmıř, sonuta bu sınav sonularının tek bařına akademik bařarının kestirilmesinde iyi bir deđiřken olmadıđını ortaya ıkarmıřtır.

Karakaya (2011) đrencilerin đrenci Seme Sınavı (SS) puanları ve Ortađretim Bařarı Puanı (OBP) ile akademik bařarı not ortalamaları (ABNO) arasındaki iliřkiyi analiz etmiř ve SS'ye ait puanların ve OBP'nin đrencilerin ABNO'yu dřk dzeyde aıkladıđı ve dřk dzeyde iliřkili olduđu sonucuna varmıřtır.

Yapılan alıřmalar gstermektedir ki üniversiteye giriř puanı, lise mezuniyet notu ve yař deđiřkenlerinin akademik bařarıyı yordamada nemli deđiřkenlerdir.

3.1.4. Öğrenci Davranışları ve Performansları

Zaiane, Xin, and Han (1998) bir Sanal Üniversite projesi olan TeleLearning-NCE kapsamında son kullanıcılara ait web sayfalarına erişim bilgilerini kullanarak *WebLogMiner* adında bir VM aracı geliştirmiştir. Geliştirilen bu araç ile web sayfalarının erişim geçmişlerine ait veriler ile kullanıcı davranışları analiz edilmiş ve daha kaliteli bir hizmet sunmaya yardımcı olacak örüntüler keşfedilmeye çalışılmıştır. Benzer şekilde Ha, Bae, and Park (2000), Chen, Liu, Ou, and Liu (2000), Mostow, et al. (2005) öğrenci davranışlarındaki örüntüleri web kayıtlarından VM yöntemleri ile keşfetmeye çalışmış ve keşfedilen bilgiler ile ÖYS’de yeni stratejiler belirlenebileceğini ve sayfa tasarımlarının yeniden organize edilebileceğini ifade etmiştir. Romero, Ventura, and Bra (2004) ve Muehlenbrock (2005) ise bu konuda öğrenci davranışlarını keşfetmeye yarayan bir web aracı geliştirmiştir.

Beck and Woolf (2000) öğrencilerin matematik eğitiminde bir problemi ne kadar doğru cevaplayacaklarını ve bu cevaplama süresinin ne kadar süreceğini belirlemek amacıyla bir öğrenme ajanı (*learning agent*) geliştirmiştir. Model oluşturulurken makine öğrenmesi için önceki öğrencilerin bilgilerinden yararlanılmıştır. Oluşturulan model cevapların ne kadar doğru olacağı ve cevaplama süresinin ne kadar olacağı konusunda tahminde bulunmaktadır.

Tang, et al. (2000) uzaktan eğitimin kişiselleştirilmesi amacıyla yaptıkları çalışmada web madenciliği yöntemleri ile 5 farklı algoritma geliştirmiştir. Algoritmalar basit bir web eğitim ağacı oluşturmak, bu web ağacına ait konu başlıklarını ve kişiselleştirilmiş web eğitim ağacını oluşturmakta kullanılmıştır. Zaiane and Luo (2001), Pahl and Donnellan (2003) hem uzaktan eğitim sisteminin kullanılabilirliğini ve etkinliğini hem de öğrencilerin öğrenme süreçlerini keşfetmek amacıyla yapmış oldukları çalışmalarda farklı VM yöntemlerinin hangi tür veriler ile nasıl kullanılabileceği incelemişlerdir.

Merceron and Yacef (2003) Logic-ITA adını verdikleri bir araç geliştirmişler, bu araç ile çevrimiçi eğitim araçlarının öğrenme ve öğretmenin geliştirilmesine nasıl katkısı olabileceğini göstermeye çalışmışlardır. Birliktelik kurallarının kullanıldığı

bu arata hangi ğrencilerin kolayca eđitim aldıkları, hangilerinin zorlandıkları ve hataların nerelerde yapıldığı gibi bilgileri veri tabanından keşfedilmeye alışılmıştır. Baker, Corbett, and Koedinger (2004) ğrencilerin zeki ğretim sistemlerini hatalı kullanımlarından kaynaklanan yetersiz ğrenmeleri engellemeyi amaçlayan bir model tasarlamışlardır. Lu (2004) ise ğrenci ihtiyaçlarını belirleyecek ve bu ihtiyaçlar dođrultusunda en iyi ğrenme materyalini sunacak bir kişisel ğrenme tavsiyecisi geliştirmiştir.

Avouris, Komis, Fiotakis, Margaritis, and Voyiatzaki (2005) teknoloji temelli ğrene aktivitelere ait kompleks verilerin analizi amacıyla geliştirilen Synergo Analysis Tool ve ColAT (*Collaborative Analysis Tool*) analiz araçlarını incelemişlerdir. Bu araçlardan ilki Synergo, işbirliğine dayalı ğrenme ortamına ait kayıt verilerini analiz etmekte kullanılmakta bir diđeri ColAT ise oklu medya ortamlarına ait veriler üzerinde analiz yapmakta kullanılmıştır.

Dringus and Ellis (2005) asenkron tartışma forumarında yer alan verilerin eđitmen tarafından deđerlendirilmesinde karşılaşılan zorluklardan yola ıkarak yapmış olduđu alışmada, veri ve metin madenciliđi yöntemleri kullanarak eđiticilerin tartışmaları daha etkin bir şekilde nasıl deđerlendirebileceklerini tartışmışlardır.

Aydın (2007) Anadolu Üniversitesi Uzaktan Eđitim Sisteminde eđitim gören ğrencilere yönelik yapmış olduđu alışmada; ğrencilerin performanslarını ğrenci özellikleri ve e-ğrenme faaliyetlerine bakarak tahmin etmeyi ve ğrencilerin mezuniyet süreleri ve özellikleri arasındaki varsa örüntüleri belirlemeyi amaçlamıştır. alışma sonucunda ğrenci performansını en iyi tahmin eden modelin C5.0 algoritması olduđu ortaya konulmuştur. Mezuniyet süreleri ve özellikleri arasındaki örüntüyü bulmak için K-means kümeleme algoritması kullanılmış, sonuçta bilgisayar ve internet kullanan bekar erkek ğrencilerin mezuniyet sürelerinin diđer ğrencilere göre daha kısa olduđu anlaşılmıştır. Mezuniyeti en geç olan ğrencilerin ise yaşı büyük, bilgisayarı sınırlı kullanan ve internet kullanmayan ğrenciler olduđu anlaşılmıştır.

Altıntop (2006) Apriori algoritması kullanarak, öğrencilerin internet tabanlı eğitim sistemi üzerindeki sisteme giriş hareketleri ile değerlendirme sorularına verilen cevapları kendi içerilerinde analiz ederek, en fazla giriş yapılan sayfalar arasındaki birliktelikler ile en fazla yanlış cevaplanan sorular arasındaki birliktelikleri analiz etmiş, sonuçta internet tabanlı eğitim sisteminin öğrenci ve öğretim elemanlarının performanslarını arttıracak şekilde yeniden düzenlenmesinin yapılabileceğini ifade etmiştir.

Şengür ve Tekin (2013) yapay sinir ağları ve karar ağaçları yöntemlerini kullanarak öğrencilerin mezuniyet notlarının erken tahmin edilmesini sağlayan bir model geliştirmişlerdir. Öğrencilerin mezuniyet durumlarının tahmin edilmesi ile belirli bir notun altında kalan öğrencilerin yönlendirilmesi ve farklı ders seçimlerinin yaptırılarak öğrenci başarısının artırılmasının sağlanabileceği belirtilmiştir.

3.2. EĞİTSEL VERİ MADENCİLİĞİ YAZILIMLARI

VM yöntemlerinin uygulanabileceği çeşitli yazılımlar geliştirilmiştir. Bu yazılımlardan en önemlileri DBMiner, Clementine, Intelligent Miner, Weka yazılımlarıdır. Bu yazılımlar her ne kadar VM yöntemlerinin uygulanmasını sağlasa da özellikle pedagojik amaçlar için tasarlanmamıştır ve VM yöntemleri konusunda kapsamlı bilgiye sahip olmayan eğitimciler için elverişsiz olduğu düşünülmektedir. Bu yüzden eğitim alanında kullanılacak bazı araçlar geliştirilmiştir (Romero and Ventura, 2007).

Eğitsel VM için geliştirilen bu araçların bazıları Çizelge 3.1’de yer almaktadır.

Çizelge 3.1. Eğitsel veri madenciliği, istatistik ve görselleştirme araçları.

Araç Adı	Yazar(lar)	VM Yöntemi
Mining tool	Zaiane ve Luo (2001)	Birliktelik ve Örüntü
MultiStar	Silva ve Vieira (2002)	Birliktelik ve Sınıflandırma
Data Analysis Center	Shen v.d. (2002)	Birliktelik ve Sınıflandırma
EPRules	Romero v.d. (2003)	Birliktelik
KAON	Tane v.d. (2004)	Metin Madenciliği ve Kümeleme

Çizelge 3.1. (devam ediyor).

TADA-ED	Merceron ve Yacef (2005)	Sınıflandırma ve Birliktelik
O3R	Becker v.d. (2005)	Ardışık Örüntüler
Synergo/ColAT	Avouris v.d. (2005)	İstatistik ve Görselleştirme
GISMO/CourseVis	Mazza ve Milani (2005)	Görselleştirme
Listen tool	Mostow v.d. (2005)	Görselleştirme
TAFPA	Damez v.d. (2005)	Sınıflandırma
iPDF_Analyzer	Bari ve Benzater (2005)	Metin Madenciliği

Kaynak: Romero and Ventura (2007) s. 137'den uyarlanmıştır.

3.3. UZAKTAN EĞİTİM İLE MÜHENDİSLİK EĞİTİMİ

Mühendislik mesleği; eğitim, deneyim ve uygulama ile edinilen matematik ve doğa bilimleri bilgisinin, doğal güç ve kaynakların insanlık yararına ve sürdürülebilirlik ilkeleri ile mühendislik etiği dikkate alınarak kullanılması için yöntemler geliştirilme çabası olarak tanımlanmaktadır (TMMOB ve Mühendislik Eğitimi, 2006).

Özkul (2003) e-öğrenme teknolojilerinin uzaktan mühendislik eğitimi için gerekli ortam ve araçları sağlayacak düzeye gelmeye başladığını ve ülkemizde de eğitsel etkinliği arttırmak, yükseköğretimde ek kapasite yaratmak ve mezuniyet sonrası eğitim fırsatları oluşturmak için diğer alanlarla birlikte mühendislik ve mesleki teknik eğitimde de e-öğrenme yöntemlerinden yararlanmanın yaygınlaşması gerektiğini vurgulamaktadır.

3.4. KBÜ UZAKTAN EĞİTİM MODELİ

Karabük Üniversitesi (KBÜ) 29 Mayıs 2007 yılında kurulmuştur. Bünyesinde 12 fakülte, 4 enstitü, 3 yüksekokul, 7 meslek yüksekokulu ve 7 araştırma merkezi bulunmaktadır. Bu araştırma merkezlerinden uzaktan eğitimin sürdürülmesinden ve teknik altyapısından sorumlu olan Karabük Üniversitesi Uzaktan Eğitim Uygulama ve Araştırma Merkezi (KBUZEM)'dir (KBÜ e-Katolog, 2014).



Şekil 3.2. KBUZEM web sayfası.

KBÜ’de açılan uzaktan eğitim önlisans programları ve bu programların öğrenci almaya başladıkları akademik yıllar Çizelge 3.2’de yer almaktadır (KBUZEM, 2014).

Çizelge 3.2. KBÜ uzaktan eğitim önlisans programları.

Bölüm/Program Adı	Yüksekokul	Akademik Yıl
Bilgisayar Programcılığı	Meslek Yüksekokulu	2010-2011
Elektronik Teknolojisi	Meslek Yüksekokulu	2010-2011
İşletme Yönetimi	Safranbolu MYO	2010-2011
Muhasebe ve Vergi Uygulamaları	Safranbolu MYO	2010-2011
İş Sağlığı ve Güvenliği	Meslek Yüksekokulu	2011-2012
Çocuk Gelişimi	Safranbolu MYO	2011-2012
Uygulamalı İngilizce ve Çevirmenlik	Safranbolu MYO	2012-2013
Sağlık Kurumları İşletmeciliği	Sağlık Hizmetleri MYO	2012-2013
İş Sağlığı ve Güvenliği	Eskipazar MYO	2012-2013

KBÜ’de açılan uzaktan eğitim lisans programları ve bu programların öğrenci almaya başladıkları tarihler Çizelge 3.3’de yer almaktadır (KBUZEM, 2014).

Çizelge 3.3. KBU uzaktan eğitim lisans programları.

Bölüm/Program Adı	Fakülte	Akademik Yıl
Bilgisayar Mühendisliği	Mühendislik Fakültesi	2009-2010
Endüstri Mühendisliği	Mühendislik Fakültesi	2010-2011
Tarih	Edebiyat Fakültesi	2010-2011
İngiliz Dili ve Edebiyatı	Edebiyat Fakültesi	2010-2011
Coğrafya	Edebiyat Fakültesi	2011-2012
Sosyoloji	Edebiyat Fakültesi	2011-2012
Türk Dili ve Edebiyatı	Edebiyat Fakültesi	2011-2012
Matematik	Fen Fakültesi	2011-2012
İşletme	İktisadi ve İdari Bilimler Fakültesi	2011-2012
Uluslararası İlişkiler	İktisadi ve İdari Bilimler Fakültesi	2013-2014

Kaynak: ÖSYM (2009, 2010, 2011, 2012, 2013) verilerinden elde edilmiştir.

KBÜ bünyesinde uzaktan eğitim veren önlisans ve lisans programlarının yanında Bilgisayar Mühendisliği Tezli ve Tezsiz Yüksek Lisans ve İşletme Tezsiz Yüksek Lisans programları da mevcuttur. Ayrıca İşyeri Hekimliği ve İş Güvenliği Uzmanlığı Eğitimi Sertifika Programının bir kısmı uzaktan eğitim yöntemi ile verilmektedir (KBUZEM, 2014). Tüm bu programlara baktığımızda KBÜ’nün uzaktan eğitim konusunda günümüze kadar birçok alanda farklı tecrübeler edindiğini ve önemli bir birikime sahip olduğunu söyleyebiliriz.

Karabük Üniversitesi’nde mühendislik eğitimi veren iki uzaktan eğitim programı mevcuttur. Bu programlardan ilki Çizelge 3.3’te görüldüğü üzere 2009 yılında ilk öğrencileri ile eğitime başlayan Bilgisayar Mühendisliği bölümü, ikincisi ise 2010 yılında eğitime başlayan Endüstri Mühendisliği bölümüdür.

Çizelge 3.4. UE mühendislik programları ve öğrenci alım yılları.

Üniversite Adı	Bölüm	Akademik Yıl
Karabük Üniversitesi	Bilgisayar Mühendisliği	2009-2010
Sakarya Üniversitesi	Bilgisayar Mühendisliği	2009-2010
Sakarya Üniversitesi	Endüstri Mühendisliği	2009-2010
Hoca Ahmet Yesevi Üniversitesi	Bilgisayar Mühendisliği	2009-2010
Hoca Ahmet Yesevi Üniversitesi	Endüstri Mühendisliği	2009-2010
Karabük Üniversitesi	Bilgisayar Mühendisliği	2010-2011
Beykent Üniversitesi	Endüstri Mühendisliği	2011-2012
Süleyman Demirel Üniversitesi	Bilgisayar Mühendisliği	2011-2012
Zirve Üniversitesi	Bilgisayar Mühendisliği	2012-2013
Zirve Üniversitesi	Endüstri Mühendisliği	2012-2013

Kaynak: ÖSYM (2009, 2010, 2011, 2012) verilerinden elde edilmiştir.

Çizelge 3.4 incelendiğinde Türkiye’deki uzaktan eğitim yöntemiyle eğitim veren ilk mühendislik programları Karabük Üniversitesi, Sakarya Üniversitesi ve Hoca Ahmet Yesevi Üniversitesi olduğu anlaşılmaktadır. Bu üniversitelerden Hoca Ahmet Yesevi Üniversitesi, Türkiye ve Kazakistan Cumhuriyetleri arasında ikili anlaşmayla kurulan uluslar arası, özerk statüye sahip, ortak devlet üniversitesidir (Ahmet Yesevi Üniversitesi, 2014). Bu üniversiteyi de ele alacak olursak 2009-2010 yıllarında 3 farklı üniversitede toplam 5 uzaktan eğitim veren mühendislik lisans programı açılmıştır. Günümüzde 4 farklı üniversitedeki 5 program ile bu sayı toplam 7 farklı üniversitede 10 programa ulaşmıştır. Ayrıca açılan programların %60’ını bilgisayar mühendisliği oluştururken, %40’ı da endüstri mühendisliğinden oluşmaktadır.

Uzaktan eğitim mühendislik bölümlerinin analiz edilmesinde kullanılacak ve karşılaştırma yapılacak bölümler aynı üniversitede eğitime devam eden yüzyüze eğitim verilen bilgisayar ve endüstri mühendislikleri bölümleridir. Bu bölümler ile bölümlerin öğrenci almaya başlangıç yılları Çizelge 3.5’de yer almaktadır.

Çizelge 3.5. KBÜ mühendislik bölümleri ve öğrenci alım yılları.

Bölüm Adı	Akademik Yıl
Bilgisayar Mühendisliği (%30 İngilizce) (BÖ)	2009-2010
Bilgisayar Mühendisliği (%30 İngilizce) (İÖ)	2009-2010
Bilgisayar Mühendisliği (UE)	2009-2010
Endüstri Mühendisliği (BÖ)	2009-2010
Endüstri Mühendisliği (İÖ)	2009-2010
Endüstri Mühendisliği (UE)	2010-2011
Bilgisayar Mühendisliği (%100 İngilizce) (BÖ)	2010-2011
Bilgisayar Mühendisliği (%100 İngilizce) (İÖ)	2010-2011

Uzaktan eğitim programlarının web ortamında yürütülmesi KBUZEM'in desteği ve organizesi ile gerçekleştirilmektedir. Ders içeriklerinin hazırlanması, web ortamının eğitime hazır hale getirilmesi ve eğitim yönetim sisteminin sürdürülebilir hale getirilmesinden de bu merkez sorumludur (KBUZEM, 2014).

KBÜ uzaktan eğitim modelinde lisans eğitimi şu şekilde uygulanmaktadır;

- Eş zamanlı eğitim
- Eş zamansız eğitim
- Yüzyüze eğitim

Eş zamanlı eğitim: Tanımlar kısmında işaret edildiği üzere, kayıtlı olan öğrencilerin web üzerinden eş zamanlı eğitim alabilmesini ifade eden eğitimidir. Bu derslerde öğrenciler aktif olarak derse katılabilirler. Bu derslerin yüzyüze eğitimden farkı sanal ortamda gerçekleştiriliyor olmasıdır.

Eş zamansız eğitim: Derslere ait eğitim materyallerinin kaydedilerek/yüklenerek öğrencilerin istediği zaman erişmesinin ve dersi takip etmesinin sağlandığı eğitimidir. Örneğin dersini kayda alıp video formatında öğretim elemanının dersin sayfasına yüklemesi gibi. Burada ders videosuna öğrenci istediği zaman erişebilmektedir.

Yüzyüze Eğitim: Üniversitede yüz yüze gerçekleştirilen eğitimidir. Bu eğitimler uzaktan eğitim programları için oranları ve tarihleri önceden belirlenecek şekilde belirlenirler. Örneğin bilgisayar mühendisliğinde bu oran 1/3 şeklinde belirlenmiştir.

Bu da 14 haftalık bir eğitimin yaklaşık 4 haftasının yüzyüze eğitim ile gerçekleşeceği anlamına gelmektedir.

KBÜ uzaktan eğitim modeli değerlendirildiğinde, yapılan eğitimin aslında sadece uzaktan eğitim olmadığı hem uzaktan hem de yüzyüze eğitim yapıldığı anlaşılmaktadır.

Eğitim modelinin bu şekilde olmasının mühendislik eğitimi konusundaki tartışmalar için önemli olduğu düşünülmektedir. Nitekim yüzyüze eğitim, bu modelde görüldüğü gibi derslerin üçte birinden oluşmaktadır. Bu oran da uygulamalı derslerin, özellikle laboratuvar ortamında yapılması gereken derslerin, yüzyüze yapılabileceği anlamına gelmektedir.

BÖLÜM 4

BULGULAR

4.1. VERİ MADENCİLİĞİ SÜRECİ

Bölüm 2.4.'de belirtilen VM sürecine ilişkin aşamalar araştırmanın bu aşamasında ilgili veri grubu üzerinde uygulanmıştır.

4.1.1. Verilerin Toplanması ve Hazırlanması

Uzaktan eğitim ve yüzyüze eğitim öğrencilere ait veriler iki farklı veritabanından elde edilmiştir. Bu veritabanları, MS SQL Server VTYS kullanan Öğrenci Otomasyonu ile Mysql VTYS kullanan Öğrenim Yönetim Sistemidir.

Farklı VTYS'nde yer alan veritabanlarındaki tablolardan SQL sorgulamaları ile analiz için gerekli olan veriler toplanmış, düzenlemeler ve kontroller yapılmak amacıyla Mysql veritabanı kullanılmıştır.

MS Excel programına aktarılan veriler, analizinin yapılacağı SPSS programına göre kontrol edilip gerekli durumlarda yeniden düzenlenmiştir. Bu düzenlemeler yapılırken veritabanlarından elde edilen değişkenlere (özellikle kategorik değişkenlere) ait veriler sayısal değerlerle ifade edilmiştir.

4.1.1.1. KBÜ Öğrenci Otomasyonu Veri Tabanı

Öğrenci otomasyon sistemi, öğrencilere ait tüm iş ve işlemlerin yürütüldüğü web tabanlı bir sistemdir. İçerisinde öğrencilere ait kimlik bilgileri, ders bilgileri, sınav bilgileri başta olmak üzere ilgili tüm veriler yer almaktadır. Öğrenci otomasyon

sisteminden seçilen öğrenci bilgilerine ait değişkenler ve bu değişkenlere ait açıklamalar Çizelge 4.1’de yer almaktadır.

Çizelge 4.1. Öğrenci bilgilerine ait değişkenler.

Değişken Adı	Açıklama
<i>öğrenci_no</i>	Öğrenci numarası
<i>adi</i>	Öğrenci adı
<i>soyadı</i>	Öğrenci soyadı
<i>fakülte</i>	Öğrencinin okuduğu fakülte
<i>bölüm</i>	Öğrencinin okuduğu bölüm
<i>sınıf</i>	Öğrencinin okuduğu sınıf
<i>dönem</i>	Öğrencinin okuduğu dönem
<i>giriş_yılı</i>	Öğrencinin kayıt yaptırdığı yıl
<i>durum</i>	Öğrencinin üniversiteye devam durumu
<i>giriş_puanı</i>	Öğrencinin üniversite giriş puanı
<i>cinsiyet</i>	Öğrencinin cinsiyeti
<i>doğum_tarihi</i>	Öğrencinin doğum tarihi
<i>not_ortalaması</i>	Öğrencinin ağırlıklı genel not ortalaması
<i>ayrılış_tarihi</i>	Üniversiteden ayrılış tarihi
<i>ayrılış_nedeni</i>	Üniversiteden ayrılış nedeni

Çizelge 4.1’deki dönüşüm yapılan değişkenlerin aldıkları değerler ve bu değişkenlerin, analiz için atanan yeni değerleri şu şekildedir:

- “bölüm” değişkeni: Bu değişkene atanan yeni değerler aynı zamanda bölüm kodu olan değerler kullanılarak yapılmıştır. Çizelge 4.2’de bölüm değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.2. “bölüm” değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
Bilgisayar Mühendisliği (%30 İngilizce)	205
Bilgisayar Mühendisliği (%30 İngilizce) (İÖ)	206
Bilgisayar Mühendisliği (%100 İngilizce)	213
Bilgisayar Mühendisliği (%100 İngilizce) (İÖ)	217
Endüstri Mühendisliği	208
Endüstri Mühendisliği (İÖ)	209

Çizelge 4.2. (devam ediyor).

UE Endüstri Mühendisliği	211
UE Bilgisayar Mühendisliği	207

- “giriş_yılı” değişkeni: Öğrencinin üniversiteye giriş yılını ifade eder. Çizelge 4.3'te giriş_yılı değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.3. "giriş_yılı" değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
2009-2010	1
2010-2011	2
2011-2012	3
2012-2013	4

- “durum” değişkeni: Öğrencinin üniversiteye devam durumunu ifade eder. Çizelge 4.4'de durum değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.4. "durum" değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
Ayrıldı (Kendi İsteği İle)	1
Beklemeli (Katkı Payı ve Öğrenim Ücreti Ödemedi)	2
Devamlı Öğrenci	3
Disiplin Cezası (Kınama)	4
Kaydı Silindi	5
Kaydı Silindi (Yatay Geçiş)	6
Kayıt Dondurdu	7
Mezun (Lisans)	8

- “cinsiyet” değişkeni: Öğrenci cinsiyet bilgisini içerir. Çizelge 4.5'te cinsiyet değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.5. “cinsiyet” değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
Erkek	1
Kız	2

- “ayrılış_nedeni” değişkeni: Öğrenci üniversiteden ayrıldı ise ayrılma nedenini ifade eder. Çizelge 4.6’da ayrılış_nedeni değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.6. “ayrılış_nedeni” değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
Kendi İsteği İle	1
Mezuniyet	2

Öğrenci otomasyon sisteminden seçilen ders bilgilerine ait değişkenler ve bu değişkenlere ait açıklamalar Çizelge 4.7’de yer almaktadır.

Çizelge 4.7. Ders bilgilerine ait değişkenler.

Değişken Adı	Açıklama
<i>fakülte</i>	Öğrencinin okuduğu fakülte
<i>bölüm</i>	Öğrencinin okuduğu bölüm
<i>öğrenci_no</i>	Öğrenci numarası
<i>adı</i>	Öğrenci adı
<i>soyadı</i>	Öğrenci adı
<i>ders_kodu</i>	Dersin kodu
<i>ders_adi</i>	Dersin tam adı
<i>kredi</i>	Dersin kredisi
<i>ders_yılı</i>	Dersin açıldığı yıl
<i>ders_dönemi</i>	Dersin açıldığı dönem
<i>dersi_açan</i>	Dersi açan öğretim elemanı
<i>ders_notu</i>	Öğrencinin derste aldığı yılsonu notu (yüzlük not)
<i>ders_notu_harf</i>	Öğrencinin derste aldığı yılsonu notu (harf notu)
<i>basari_durumu</i>	Öğrenci ders başarısı

Çizelge 4.7’deki dönüşüm yapılan değişkenlerin aldıkları değerler ve bu değişkenlerin, analiz için atanan yeni değerleri şu şekildedir:

- “bölüm” değişkeni: Analiz için gerekli olan bu değişkene ait yeni değerler aynı zamanda bölüm kodu olan değerler kullanılarak düzenleme yapılmıştır. Çizelge 4.8’de bölüm değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.8. Derslere ait “bölüm” değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
Bilgisayar Mühendisliği (%30 İngilizce)	205
Bilgisayar Mühendisliği (%30 İngilizce) (İÖ)	206
Bilgisayar Mühendisliği (%100 İngilizce)	213
Bilgisayar Mühendisliği (%100 İngilizce) (İÖ)	217
Endüstri Mühendisliği	208
Endüstri Mühendisliği (İÖ)	209
UE Endüstri Mühendisliği	211
UE Bilgisayar Mühendisliği	207

- “ders_yılı” değişkeni: Dersin ait olduğu akademik yılı ifade eder. Çizelge 4.9’da ders_yılı değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.9. Derslere ait "ders_yılı" değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
2009-2010	1
2010-2011	2
2011-2012	3
2012-2013	4

- “ders_dönemi” değişkeni: Dersin ait olduğu dönemi ifade eder. Çizelge 4.10’da ders_dönemi değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.10. Derslere ait "ders_dönemi" değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
Güz Dönemi	1
Bahar Dönemi	2

- “ders_notu_harf” değişkeni: Ders geçme notlarını ifade eder. Çizelge 4.11’de ders_notu_harf değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.11. Derslere ait "ders_notu_harf" değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
A1	4,00
A2	3,50
B1	3,00
B2	2,75
C	2,50
F1	99
F2	99
F3	0

- “başarı_durumu” değişkeni: Ders geçme durumunu ifade eder. Çizelge 4.12’de başarı_durumu değişkenine ait veri dönüşümleri yer almaktadır.

Çizelge 4.12. Derslere ait “başarı_durumu” değişkeni veri dönüşümü.

Mevcut Değer	Yeni Değer
Başarılı	1
Başarısız	0

4.1.1.2. KBUZEM Öğrenim Yönetim Sistemi Veri Tabanı

ÖYS’de yer alan öğrenci kayıtları veritabanında yer almaktadır. Buradaki etkinlik kayıtları aşağıdaki Şekil 4.1’de görüldüğü gibi ders ekranından erişilebilmektedir. Burada öğrenciye ait erişim zamanı, kullanıcı adı, eylem adı ve eyleme ait ayrıntılı bilgi yer almaktadır. Öğrencilere ait katılım oranı bilgileri ÖYS’de yer alan farklı tür ders materyallerine erişim oranları verilerinden elde edilmiştir. Genel olarak katılım oranı farklı tür ders materyallerine ait erişim sayısının, o ders materyaline ait toplam ders materyal sayısına bölümü ile elde edilen ortalamalardan oluşmaktadır. Katılım oranının temel varsayımı, ders materyaline erişen öğrencinin o materyali çalışmış olduğudur. Örneğin birinci haftanın ders notuna bakma eylemini gerçekleştirmiş öğrencinin o ders notunu okumuş/çalışmış olduğu varsayılmaktadır.

Şekil 4.1’de ÖYS’deki bir derse ait kullanıcı eylemlerinin örnek raporlama sayfası yer almaktadır.

Diferansiyel Denklemler 2011-2012: Bütün katılımcılar, Bütün günler (Sunucunun yerel zamanı)

Diferansiyel Denklemler 2011-2012 [dahası] Bütün katılımcılar Bütün günler Bütün etkinlikler
 Bütün değişiklikler Sayfada göster Bu kayıtları getir

150 kayıt gösteriliyor
Sayfa: (Önceki) 1 2

Zaman	IP Adresi	Tam adı	Eylem	Bilgi
Çrş 12 Ekim 2011, 15:51	10.3.0.101	Caner ÖZCAN	resource add	Senkron Ders Videosu
Çrş 12 Ekim 2011, 10:51	10.3.1.37	Semra BAYRAKCI	course update mod	flv 1144
Çrş 12 Ekim 2011, 10:51	10.3.1.37	Semra BAYRAKCI	flv update	Ders Videosu
Çrş 12 Ekim 2011, 10:40	10.3.1.37	Semra BAYRAKCI	course update mod	flv 1144
Çrş 12 Ekim 2011, 10:40	10.3.1.37	Semra BAYRAKCI	flv update	Ders Videosu
Çrş 12 Ekim 2011, 10:39	10.3.1.37	Semra BAYRAKCI	course add mod	flv 1144
Çrş 12 Ekim 2011, 10:39	10.3.1.37	Semra BAYRAKCI	flv add	Ders Videosu
Çrş 12 Ekim 2011, 10:37	10.3.1.37	Semra BAYRAKCI	course add mod	flv 1143
Çrş 12 Ekim 2011, 10:37	10.3.1.37	Semra BAYRAKCI	flv add	Ders Videosu
Çrş 12 Ekim 2011, 10:32	10.3.1.37	Semra BAYRAKCI	course add mod	flv 1140
Çrş 12 Ekim 2011, 10:32	10.3.1.37	Semra BAYRAKCI	flv add	Ders Videosu
Sal 11 Ekim 2011, 17:06	10.3.1.37	Semra BAYRAKCI	course add mod	resource 12480
Sal 11 Ekim 2011, 17:06	10.3.1.37	Semra BAYRAKCI	resource add	ÇALIŞMA SORULARI
Sal 11 Ekim 2011, 17:05	10.3.1.37	Semra BAYRAKCI	resource add	resource 12479
Sal 11 Ekim 2011, 17:05	10.3.1.37	Semra BAYRAKCI	resource add	ÇÖZÜMLÜ ÇALIŞMA SORULARI - 3
Sal 11 Ekim 2011, 17:05	10.3.1.37	Semra BAYRAKCI	upload upload	/usr/moodledata/290/calisma_sorulari/Coezueumlu_Calisma_Sorulari-3.pdf
Sal 11 Ekim 2011, 17:03	10.3.1.37	Semra BAYRAKCI	course add mod	resource 12478

Şekil 4.1. Örnek bir derse ait etkinlik kayıtları.

SQL sorgulama dili ile öğrenim yönetim sistemi veri tabanındaki farklı tablolardan elde edilen etkinlik kayıtları sorgusu ile elde edilen verilerden iki farklı tablo oluşturularak yeni bir veri tabanına kaydedilmiştir.

Bu tablolardan birincisi, ders kayıtlarına ait bilgilerin yer aldığı “ders kayıtları” tablosudur. Diğer bir tablo ise derse katılım oranlarının tutulduğu “katılım oranı” tablosudur. Veri analizinde kullanılacak bu tablolar öğrenci otomasyon sisteminde yer alan öğrencilere ait demografik bilgiler, sınav sonuçları, giriş puanları ve not ortalamalarını ile birleştirilerek tek bir tablo oluşturulmuştur. Oluşturulan bu son tablo SPSS programı ile analiz yapılmaya hazır hale gelen verilerden oluşmaktadır. Ders kayıtlarının kaydedildiği örnek tablo Şekil 4.2’de yer almaktadır:

+ Seçenekler

	id	bolumadi	bolum	bolumid	ipadres	adsoyad	eylem	eylemadi
<input type="checkbox"/> Düzenle <input type="checkbox"/> Kopyala <input type="checkbox"/> Sil	1	BM	292	11	193.140.85.150	Abdulkadir TAŞDELEN	resource view	1.Hafta Ders Notu PDF belgesi
<input type="checkbox"/> Düzenle <input type="checkbox"/> Kopyala <input type="checkbox"/> Sil	2	BM	292	11	193.140.85.150	Abdulkadir TAŞDELEN	resource view	Senkron Ders
<input type="checkbox"/> Düzenle <input type="checkbox"/> Kopyala <input type="checkbox"/> Sil	3	BM	292	11	193.140.85.150	Abdulkadir TAŞDELEN	course view	Mantık Devreleri I 2011-2012
<input type="checkbox"/> Düzenle <input type="checkbox"/> Kopyala <input type="checkbox"/> Sil	4	BM	292	6	10.12.3.213	Abdulkadir TAŞDELEN	course view	Mantık Devreleri I 2011-2012
<input type="checkbox"/> Düzenle <input type="checkbox"/> Kopyala <input type="checkbox"/> Sil	5	BM	292	21	10.12.3.33	Abdulkadir TAŞDELEN	course view	Mantık Devreleri I 2011-2012
<input type="checkbox"/> Düzenle <input type="checkbox"/> Kopyala <input type="checkbox"/> Sil	6	BM	292	12	10.12.3.60	Abdulkadir TAŞDELEN	course view	Mantık Devreleri I 2011-2012

Şekil 4.2. Ders kayıtları tablosu.

Her bir ders için kullanıcının derse katılım oranı verilerini içeren örnek tablo Şekil 4.3'te yer almaktadır:

bolun	bolunAdi	adsoyad	oran
686	Demir Çelik Endüstrisinde Endüstr	ABDULKADİR DÖNMEZ	76.7
321	Uluslararası Ağ Yapılar ve Kümele	ABDULKADİR DÖNMEZ	60
292	Mantık Devreleri I	Abdulkadir TAŞDELEN	5.9
697	Mühendislik İstatistiği II	Abdulkadir TAŞDELEN	2.7
522	Elektronik I	Abdulkadir TAŞDELEN	0
315	Ekonomi I	Abdulkadir TAŞDELEN	0

Şekil 4.3. Derse katılım oranı tablosu.

4.1.2. Modelleme

Veri seti içerisinde yer alan bölümler, öğrenci alım yılları ve öğrenci sayılarına ait veriler Çizelge 4.13'te yer almaktadır:

Çizelge 4.13. Giriş yıllarına göre öğrenci sayıları ve bölümler.

Bölüm Adı	Giriş Yılı				Toplam
	2009-2010	2010-2011	2011-2012	2012-2013	
Bilgisayar Müh. (%30 İng.) (BÖ)	40	75	71	73	259
Bilgisayar Müh. (%30 İng.) (İÖ)	39	76	74	80	269
Bilgisayar Müh. (UE)	45	95	92	103	335
Endüstri Müh. (BÖ)	40	74	73	82	269
Endüstri Müh. (İÖ)	39	77	75	82	273
Endüstri Müh. (UE)	0	64	88	104	256
Bilgisayar Müh. (%100 İng.) (BÖ)	0	57	66	77	200
Bilgisayar Müh. (%100 İng.) (İÖ)	0	57	65	81	203
Toplam	203	575	604	682	2064

Çizelge 4.13'de yer alan verilere göre veri setinde 2009-2010 ile 2012-2013 akademik yılları arasında eğitim alan ve/veya devam eden toplam 2064 öğrenciye ait veriden yararlanılmıştır. Bu öğrencilerden 335'i bilgisayar mühendisliği ve 256'sı endüstri mühendisliği olmak üzere toplam da 591 uzaktan eğitim öğrencisi

bulunmaktadır. Diğer 1473 öğrenci ise aynı bölümlere ait yüzyüze eğitim alan öğrencilerdir.

Öğrencilerin üniversiteye giriş puanları ortalamaları tek yönlü ANOVA testi ile karşılaştırılarak bölümler arasında bir fark olup olmadığı analiz edilmek istenmiştir. Yapılan analiz sonucunda ilişkisiz örnekler için tek yönlü varyans analizi ANOVA testinin yapılabilmesi için gereken şartlardan biri olan dağılımların normalliği şartının sağlanmadığı anlaşılmıştır ($p < 0.05$).

Parametrik bir test olan tek yönlü ANOVA testi yapılabilmesi için gereken şartlar yerine gelmediği zaman veriler üzerinde dönüşüm işlemleri yapılabileceği gibi parametrik olmayan testlerden de yararlanılabileceği Bölüm 1.3'te belirtilmiştir. Aynı bölümde belirtilen parametrik olmayan Kruskal Wallis testiyle gruplar arasında istatistiksel olarak anlamlı bir fark olup olmadığı analiz edilebildiğinden bu test kullanılmıştır.

2011-2012 yılında üniversiteyi kazanan bilgisayar mühendisliği öğrencilerinin giriş puanları arasında anlamlı bir fark olup olmadığını test etmek için yapılan Kruskal Wallis testinde, grupların üniversiteye giriş puanları ortalamaları arasında anlamlı bir fark olduğu gözlenmiştir ($\chi^2_{(4)} = 257,829$, $p < 0.05$). Mann Whitney U testi ile yapılan çoklu karşılaştırmalar sonucu bu farkın, bütün gruplar arasında olduğu tespit edilmiştir.

Bilgisayar Mühendisliği bölümüne ait analiz sonuçları Çizelge 4.14'de gösterilmiştir.

Çizelge 4.14. Bilgisayar mühendisliğine ait giriş puanları analizi.

Grup No	Grup Adı	n	Sıra Ortalaması	sd	χ^2	p	Anlamlı Fark
1	Bilgisayar Müh. (%30 İng.)	67	260,36	4	257,829	0,000	1-2, 1-3, 1-4, 1-5
2	Bilgisayar Müh. (%30 İng.) (İÖ)	73	135,52				2-3, 2-4, 2-5
3	Bilgisayar Müh. (UE)	75	46,29				3-4, 3-5
4	Bilgisayar Müh. (%100 İng.)	66	277,76				4-5
5	Bilgisayar Müh. (%100 İng.) (İÖ)	63	161,89				

2011-2012 yılında üniversiteyi kazanan endüstri mühendisliği öğrencilerin giriş puanları arasında anlamlı bir fark olup olmadığını test etmek için yapılan Kruskal Wallis testinde, grupların üniversiteye giriş puanları ortalamaları arasında anlamlı bir fark olduğu gözlenmiştir ($\chi^2_{(2)}=191,156$, $p<0.05$). Mann Whitney U testi ile yapılan çoklu karşılaştırmalar sonucu bu farkın, bütün gruplar arasında olduğu tespit edilmiştir.

Endüstri mühendisliği bölümüne ait analiz sonuçları Çizelge 4.15’de yer almaktadır.

Çizelge 4.15. Endüstri Mühendisliğine ait giriş puanları analizi.

Grup No	Grup Adı	n	Sıra Ortalaması	sd	χ^2	p	Anlamlı Fark
1	Endüstri Müh. (BÖ)	72	260,36	2	191,156	0,000	1-2, 1-3
2	Endüstri Müh. (İÖ)	73	135,52				2-3
3	Endüstri Müh. (UE)	73	46,29				

2011-2012 yılında üniversiteyi kazanan endüstri mühendisliği öğrencileri ile bilgisayar mühendisliği öğrencilerinin giriş puanları arasında anlamlı bir fark olup olmadığını test etmek için yapılan Mann Whitney U testinin sonuçlarına göre, bilgisayar mühendisliği UE öğrencileri giriş puanları ortalamaları (ortanca=260,988) ile endüstri mühendisliği UE öğrencileri giriş puanları ortalamaları (ortanca=242,418) arasında istatistiksel olarak anlamlı bir fark olduğu gözlenmiştir ($U=1152$, $p<0,05$).

Çizelge 4.16. Bilgisayar ve endüstri mühendisliği UE giriş puanları analizi.

Grup No	Grup Adı	n	Sıra Ortalaması	Sıra Toplamı	U	p
1	Bilgisayar Müh. (UE)	75	95,64	7173	1152,00	0,000
2	Endüstri Müh. (UE)	73	52,78	3853		

Bölümlerin kendi içerisinde yapılan analizler birinci öğretim, ikinci öğretim ve uzaktan eğitim programlarına ait giriş puanları arasında anlamlı fark olduğu göstermektedir. Ayrıca farklı iki bölüme ait uzaktan eğitim programları arasında da istatistiksel olarak anlamlı bir fark olduğu anlaşılmaktadır.

Bölgümlere ait öđrenci üniversite giriş puanları ortalamalarının analizinden sonra analiz edilen ikinci bir deđişken de öđrencilerin ađırlıklı genel not ortalamalarıdır. 2011 – 2012 yılında üniversiteye giren öđrencilere ait ađırlıklı genel not ortalamaları arasında anlamlı bir fark olup olmadığını test etmek için yapılan normallik testine göre bilgisayar mühendisliğine ait not ortalamalarının normal dađılım gösterdiğini ($p>0,05$) endüstri mühendisliği bölümlerinin ise normal dađılım göstermediğini anlaşılmıştır ($p<0,05$). Bu sonuçlara göre bilgisayar mühendisliği için ilişkisiz örnekler için tek yönlü varyans analizi ANOVA kullanılmış, endüstri mühendisliği için ise Kruskal Wallis analizi kullanılmıştır. İki farklı uzaktan eğitim programının ortalamalarının karşılaştırılmasında da yine Mann Whitney U testi kullanılmıştır.

Bölüm 1.3'te belirtildiđi üzere ANOVA testinin yapılabilmesi için, hem gruplara ait deđişkenlerin normal dađılım göstermesi hem de ortalamaları kıyaslanacak grupların varyanslarının eşit olması gerekmektedir. Normallik testi sonucunda gruplara ait dađılımların normalliğinin sağlandığını tespitinden sonra ($p>0,05$) gruplara ait varyansların eşit olup olmadığını test etmek için Levene Testi uygulanmış ve grup varyansları arasında anlamlı bir fark olmadığı anlaşılmıştır ($p>0,05$). ANOVA testi sonucunda da gruplar arasında anlamlı bir fark yoktur hipotezi reddedilmiş ve en az iki grup arasında anlamlı bir fark olduğu anlaşılmıştır ($F_{(4-331)}=12,95$, $p<0,05$). Test sonucunda hesaplanan etki büyüklüğünün ($\eta^2=0,135$) bu farkın geniş düzeye yakın olduğunu göstermektedir. Yapılan Tukey Çoklu Karşılaştırma Testi sonucunda Bilgisayar Mühendisliği (%30 İngilizce) Birinci, İkinci Öğretim ve (%100 İngilizce) Birinci Öğretim programlarının homojen bir alt grup oluşturduğu ve aralarında anlamlı bir fark olmadığı anlaşılmıştır ($p>0,05$). Aynı şekilde Bilgisayar Mühendisliği (UE) ile %100 İngilizce İÖ programının bir diğer alt grubu oluşturduğu ve aralarında anlamlı bir fark olmadığı anlaşılmıştır ($p>0,05$).

Çizelge 4.17. Bilgisayar mühendisliği not ortalamaları analiz sonuçları.

Varyansın Kaynađı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	Anlamlı Fark*
Gruplar arası	25,069	4	6,267	12,950	0,000	1-3,1-5
Gruplar içi	160,187	331	0,484			2-3,2-5
Toplam	185,256	335				3-4, 4-5

*1= Bilgisayar Müh. (%30 İng.) (BÖ), 2= Bilgisayar Müh. (%30 İng.) (İÖ)

3= *Bilgisayar Müh. (%30 İng.) (UE)*, 4= *Bilgisayar Müh. (%100 İng.) (BÖ)*
5= *Bilgisayar Müh. (%100 İng.) (İÖ)*

2011-2012 yılında üniversiteyi kazanan endüstri mühendisliği öğrencilerin ağırlıklı genel not ortalamaları arasında anlamlı bir fark olup olmadığını test etmek için yapılan Kruskal Wallis testinde, grupların ortalamaları arasında anlamlı bir fark olduğu gözlenmiştir ($\chi^2_{(2)}=106,242$, $p<0.05$). Mann Whitney U testi ile yapılan çoklu karşılaştırmalar sonucu bu farkın, bütün gruplar arasında olduğu tespit edilmiştir.

Çizelge 4.18. Endüstri mühendisliğine ait giriş puanları analizi.

Grup No	Grup Adı	n	Sıra Ortalaması	sd	χ^2	p	Anlamlı Fark
1	Endüstri Müh. (BÖ)	71	162,61	2	106,242	0,000	1-2, 1-3
2	Endüstri Müh. (İÖ)	72	130,00				2-3
3	Endüstri Müh. (UE)	83	57,17				

2011-2012 yılında üniversiteyi kazanan endüstri mühendisliği öğrencileri ile bilgisayar mühendisliği öğrencilerinin ağırlıklı genel not ortalamaları arasında anlamlı bir fark olup olmadığını test etmek için yapılan Mann Whitney U testinin sonuçlarına göre, bilgisayar mühendisliği (UE) öğrencileri not ortalamaları (ortanca=2,020) ile endüstri mühendisliği (UE) öğrencileri not ortalamaları (ortanca=2,025) arasında istatistiksel olarak anlamlı bir fark olmadığı gözlenmiştir ($U=3299,5$, $p>0,05$).

Çizelge 4.19. Bilgisayar ve endüstri mühendisliği UE not ortalamaları analizi.

Grup No	Grup Adı	n	Sıra Ortalaması	Sıra Toplamı	U	p
1	Bilgisayar Müh. (UE)	86	88,13	7579,50	3299.5	0,397
2	Endüstri Müh. (UE)	83	81,75	6785,50		

Çizelge 4.19’da bilgisayar ve endüstri mühendisliği UE bölümlerinin not ortalamaları analiz sonuçları yer almaktadır.

Üniversite giriş puanları ile öğrencileri ağırlıklı not ortalamaları arasında ilişki olup olmadığını test etmek için yapılan Spearman Sıra Farkları Korelasyon işlemine göre üniversite giriş puanı ile ağırlıklı not ortalaması arasında pozitif yönde ve anlamlı düzeyde düşük bir ilişki olduğu tespit edilmiştir ($r=0,37$, $p<0.01$). Bu durumda üniversiteye giriş puanı not ortalamasının %13'ünü açıklayabilmektedir ($r^2 =0,13$). Bu ilişki Karakaya (2011)'nin belirttiği sonuçlar ile benzerlik göstermektedir.

Bölüm 1.3'te belirtildiği gibi uzaktan eğitim öğrencilerin ders geçme durumları ile üniversite giriş puanı, yaş, cinsiyet, uzaktan eğitime katılım oranı arasındaki ilişki lojistik regresyon analizi ile analiz edilerek bir model geliştirilmiştir.

Öğrencilerin ders geçme durumları *girisPuani*, *katilimOrani*, *yas* ve *cinsiyet* değişkenlerine göre lojistik regresyon analizi ile modellemede kullanılmıştır. Bu değişkenler ve özellikleri şu şekildedir;

- dersGecme: Öğrencinin dersi geçmesi (1) veya dersten kalmasını (0) ifade eden kategorik değişkendir.
- girisPuani: Üniversite giriş puanını ifade eden sürekli değişkendir.
- yas: Öğrencinin yaşını ifade eden sürekli değişkendir.
- cinsiyet: Öğrencinin cinsiyetini ifade eden kategorik değişkendir. Erkek 1 sayısı ile kadın 2 sayısı ile temsil edilmektedir.

Analiz edilen toplam 914 olaydan (*case*) hiçbiri kayıp veri olarak değerlendirilmemiş dolayısıyla hepsi analize dâhil edilmiştir. Bu olaylara ait bilgiler Çizelge 4.20'de yer almaktadır.

Çizelge 4.20. Analiz edilen olaylar.

		Sayı	Yüzde
Seçilen Olaylar	Analize eklenen	914	100,0
	Kayıp Olay	0	,0
	Toplam	914	100,0
Seçilmeyen Kayıtlar		0	,0
Toplam		914	100,0

Yapılan analizde ilk önce teorik modelin veriyi temsil etmesi bakımından uygun olup olmadığı incelenmiştir. Bu amaçla yapılan Hosmer ve Lemeshow testi sonuçlarına göre, model veriyi iyi temsil etmektedir ($p>0,05$).

Çizelge 4.21’de Hosmer ve Lemeshow testine ait adımlar ve bu adımlardaki sonuçlar yer almaktadır.

Çizelge 4.21 Hosmer ve Lemeshow testi.

Adım	Chi-square	sd	p
1	22,267	8	,004
2	12,346	8	,136
3	2,939	8	,938

Başlangıç adımında yer alan sınıflandırmaya göre modelin doğru sınıflandırma yüzdesi %66,4’dür (Çizelge 4.22).

Çizelge 4.22. Başlangıç sınıflandırma tablosu.

Gözlenen	Tahmin Edilen		
	dersGecme		Doğru Yüzdesi
	0	1	
0	607	0	100,0
Adım 0 dersGecme 1	307	0	,0 66,4

Sınıflandırma tablosu incelendiğinde (Çizelge 4.23), modelin doğru sınıflandırmasının üçüncü adımda tamamlandığı ve sınıflandırma başarısının da %69,4 olduğu anlaşılmaktadır. Bu değer modelin genel tahmin etme oranının %69,4 olduğunu göstermektedir.

Çizelge 4.23. Sınıflandırma tablosu.

Gözlenen			Tahmin Edilen		
			dersGecme		Doğru Yüzdesi
			0	1	
Adım 1	dersGecme	0	564	43	92,9
		1	245	62	20,2
	Genel Yüzde				68,5
Adım 2	dersGecme	0	561	46	92,4
		1	242	65	21,2
	Genel Yüzde				68,5
Adım 3	dersGecme	0	560	47	92,3
		1	233	74	24,1
	Genel Yüzde				69,4

Çizelge 4.24’de görüldüğü üzere modelin parametrelerinin anlamlılığı testi olan Omnibus testi sonucunda model parametrelerinin anlamlı olduğu anlaşılmıştır ($p < 0,05$).

Çizelge 4.24. Omnibus testi.

		ki-kare	sd	<i>p</i>
	Adım	57,392	1	,000
Adım 1	Block	57,392	1	,000
	Model	57,392	1	,000
	Adım	11,024	1	,001
Adım 2	Block	68,417	2	,000
	Model	68,417	2	,000
	Adım	9,490	1	,002
Adım 3	Block	77,907	3	,000
	Model	77,907	3	,000

Bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi Negelkerke R^2 göstermektedir. Buna göre oluşturulan modelin bu ilişkiyi açıklama gücü %11,3’dür. Bu değer modelde yer alan değişkenlerin ders geçme durumunun %11,3’ü açıkladığını göstermektedir (Çizelge 4.25).

Çizelge 4.25. Model özeti.

Adım	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1109,363	,061	,084
2	1098,339	,072	,100
3	1088,849	,082	,113

Model oluşturulurken kullanılan yöntem ileriye dönük adımsal yöntemdir. Bölüm 2.5.2.2’de belirtildiği üzere bu yöntemde sabit değerden başlamak üzere her seferinde bir değişken daha ekleyerek model oluşturulmaktadır. Çizelge 4.25’de de görüldüğü gibi eşitliğe dâhil edilen değişkenler *girisPuani*, *katilimOrani* ve *yas* değişkenleridir. Cinsiyet değişkeni ise eşitlikte anlamlı bulunmadığı için ($p>0,05$) model içerisinde yer almamaktadır.

Çizelge 4.26. Eşitlikteki değişkenlerin analizi.

	B	S. hata	Wald	sd	p	Exp(B)
girisPuani	,004	,001	9,426	1	,002	1,004
katilimOrani	,018	,003	45,371	1	,000	1,018
yas	,053	,017	9,420	1	,002	1,054
Sabit	-3,631	,528	47,297	1	,000	,026

Oluşturulan modelin kestirim eşitliği şu şekildedir:

$$\log\left(\frac{P}{1-P}\right) = \alpha + \beta (\text{girisPuani}) + \beta (\text{katilimOrani}) + \beta (\text{yasYeni}) \quad (4.1)$$

Formül 4.1’de yer alan α ve β değerleri yerine kestirilmiş değerleri yazdığımızda formülümüz şu şekilde olmaktadır:

$$\log\left(\frac{\hat{P}}{1-\hat{P}}\right) = a + b1 (\text{girisPuani}) + b2 (\text{katilimOrani}) + b3 (\text{yasYeni}) \quad (4.2)$$

$$\log\left(\frac{\hat{P}}{1-\hat{P}}\right) = -3.631 + 0.004 (\text{girisPuani}) + 0.018 (\text{katilimOrani}) + 0.053 (\text{yasYeni}) \quad (4.3)$$

Formülde yer alan \hat{P} işareti kestirilen olasılık olarak anlamında kullanılmaktadır.

Modelde yer alan β değerlerinin kestirimi olan b değerlerinden hiç biri negatif değer almamaktadır. Bu değer negatif olmaması, ders geçme en çok olabilirliğinin bütün bağımsız değişkenler için arttığını göstermektedir.

Modelde yer alan değişkenleri Formül 2.20'deki yerlerine yazarsak aşağıdaki formülü elde etmiş oluruz.

$$\hat{P} = \frac{(e^{-3.631 + 0.004 \text{ girişPuani} + 0.018 \text{ katilimOrani} + 0.053 \text{ yasYeni}})}{1 + (e^{-3.631 + 0.004 \text{ girişPuani} + 0.018 \text{ katilimOrani} + 0.053 \text{ yasYeni}})} \quad (4.5)$$

Formül 4.5'den yararlanarak örnek değerler ile öğrencilerin sınıf geçme olasılığını hesaplandığı çeşitli örnekler aşağıda yer almaktadır.

Örnek 4.1: $\text{girişPuani} = 300$, $\text{katilimOrani} = 0$, $\text{yasYeni} = 24$ olan bir öğrencinin ders geçme olasılığı şu şekilde hesaplanır:

$$\hat{p} = \frac{(e^{-3,631 + 0,004 \cdot 300 + 0,018 \cdot 0 + 0,053 \cdot 24})}{1 + (e^{-3,631 + 0,004 \cdot 300 + 0,018 \cdot 0 + 0,053 \cdot 24})}$$
$$\hat{p} = 0,24$$

Yukarıdaki örnekte yer alan öğrencinin ders geçme olasılığı %24'tür.

Örnek 4.2: $\text{girişPuani} = 300$, $\text{katilimOrani} = 50$, $\text{yasYeni} = 24$ olan bir öğrencinin ders geçme olasılığı şu şekilde hesaplanır:

$$\hat{p} = \frac{(e^{-3,631 + 0,004 \cdot 300 + 0,018 \cdot 50 + 0,053 \cdot 24})}{1 + (e^{-3,631 + 0,004 \cdot 300 + 0,018 \cdot 50 + 0,053 \cdot 24})}$$
$$\hat{p} = 0,44$$

Yukarıdaki örnekte yer alan öğrencinin ders geçme olasılığı %44'tür.

Örnek 4.3: $\text{girişPuani} = 300$, $\text{katilimOrani} = 100$, $\text{yasYeni} = 24$ olan bir öğrencinin ders geçme olasılığı şu şekilde hesaplanır:

$$\hat{\pi} = \frac{(e^{-3,631 + 0,004 \cdot 300 + 0,018 \cdot 100 + 0,053 \cdot 24})}{1 + (e^{-3,631 + 0,004 \cdot 300 + 0,018 \cdot 100 + 0,053 \cdot 24})}$$

$$\hat{\pi} = 0,65$$

Örnek 4.3’de yer alan öğrencinin ders geçme olasılığı %65’tür.

Örneklerden de görüldüğü gibi değişkenlere ait farklı değerler modelde test edildikçe öğrencinin dersi geçme olasılığı değişmektedir. Bu model için bir yazılım geliştirilip ilgili değişken bilgileri belirli aralıklarla dikkate alınarak öğrenci ve öğretim elemanlarını uyaran bir sistem geliştirilebilir. Bu şekilde öğrencilerin ders geçmesinde faydalı olacak farklı önlemlerin alınmasının sağlanabilir.

BÖLÜM 5

SONUÇLAR

Bölümlerin kendi içerisinde yapılan analizler birinci öğretim, ikinci öğretim ve uzaktan eğitim programlarına ait üniversite giriş puanları arasında anlamlı fark olduğu sonucuna varılmıştır. Ayrıca farklı iki bölüme ait uzaktan eğitim programlarına üniversite giriş puanları arasında da istatistiksel olarak anlamlı bir fark olduğu sonucuna varılmıştır.

Öğrencileri ağırlıklı genel not ortalamaları dikkate alındığında Bilgisayar Mühendisliği (%30 İngilizce) Birinci Öğretim, İkinci Öğretim ve (%100 İngilizce) Birinci Öğretim programlarının homojen bir alt grup oluşturduğu ve aralarında anlamlı bir fark olmadığı anlaşılmıştır. Aynı şekilde Bilgisayar Mühendisliği Uzaktan Eğitim ile (%100 İngilizce) İkinci Öğretim programının bir diğer alt grubu oluşturduğu ve aralarında anlamlı bir fark olmadığı anlaşılmıştır. Bu farkın Endüstri Mühendisliğinde tüm programlar arasında olduğu sonucuna varılmıştır. Bunun yanında Uzaktan Eğitim Bilgisayar ve Endüstri Mühendisliği arasında ağırlıklı genel not ortalamaları bakımından anlamlı bir fark olmadığı anlaşılmıştır. Bu sonuçlar göstermektedir ki sadece uzaktan eğitim programları değil diğer birinci ve ikinci öğretim programları da akademik başarı bakımından birbirinden farklılık gösterebilmektedir.

Üniversite giriş puanları ile öğrencileri ağırlıklı not ortalamaları arasında pozitif yönde ve anlamlı düzeyde düşük bir ilişki olduğu tespit edilmiştir. Bu durumda üniversiteye giriş puanı ağırlıklı genel not ortalamasının sadece %13'ünü açıklayabilmektedir. Bu oran da oldukça düşük bulunmaktadır.

Uzaktan eğitim öğrencilerin ders geçme durumları ile üniversite giriş puanı, yaş, cinsiyet, uzaktan eğitime katılım oranı arasındaki ilişki lojistik regresyon analizi ile

analiz edilerek bir model geliştirilmiştir. Geliştirilen modelin doğru tahmin etme ihtimalinin %67,7 olduğu ve bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi açıklama gücü %11,9 olarak belirlenmiştir.

Sonuçlar göstermektedir ki, mühendislik bölümlerinin birinci öğretim, ikinci öğretim veya uzaktan eğitim olması öğrencilerin bu bölümleri tercih etme durumları farklılık göstermektedir. Dolayısıyla üniversiteye giriş puanları arasında farklılıklar ortaya çıkmaktadır. Fakat bu giriş puanlarındaki farklılıkların bölümlere devam eden öğrencilerin başarılarını açıklamada ilişkisi incelendiğinde, bu ilişkinin yani üniversiteye giriş puanı ile ağırlıklı genel not ortalamasının düşük bir ilişkiye sahip olduğu ve sadece üniversiteye giriş puanının bu ilişkiyi açıklamada tek başına yeterli olmadığı anlaşılmıştır.

Uzaktan eğitim programlarının geleceği düşünüldüğünde akademik başarıyı açıklayan diğer değişkenlerin varlığının göz ardı edilmemesinin ve çalışmaların bu değişkenlerin de belirlendiği ve değerlendirildiği bir şekilde sürdürülmesinin önemli olduğu düşünülmektedir.

Uzaktan eğitim mühendislik bölümlerinin henüz yeni mezun vermeye başlamışlardır. İleriki çalışmalarda bu bölümlerin mezun öğrencilerine ait verilerle değerlendirilmesinin uygun olacağı düşünülmektedir. Ayrıca üniversitelerde uygulanan farklı uzaktan eğitim modellerinin karşılaştırılması için çalışmalar yapılabilir ve başarılı bir mühendislik eğitimi için nasıl bir uzaktan eğitim modeli benimsenmesi gerektiği konusunda önerilerde bulunabilir. Bu şekilde, eğitimde fırsat eşitliğini ve erişilebilirliği sağlamada önemli bir rol üstlenebilecek olan uzaktan eğitim bölümlerinin başarısız olmasını beklemek yerine tüm değişkenlerin göz önüne alındığı bir model oluşturmaya çalışmanın daha önemli olacağı düşünülmektedir.

İleriki çalışmalarda, oluşturulan model için bir yazılım geliştirilip ilgili değişken bilgileri belirli aralıklarla güncellenerek öğrenci ve öğretim elemanlarını uyarıcı bir sistem geliştirilmesi düşünülmektedir. Bu şekilde öğrencilerin ders geçmesinde faydalı olacak farklı önlemlerin alınmasının sağlamak amaçlanmaktadır.

KAYNAKLAR

Agrawal, R. and Srikant, R., “Fast algorithms for mining association rules”, *20. Very Large Data Bases (VLDB) Konferansı*, Şili, 1215: 487-499 (1994).

Akhun, İ., “Akademik başarının kestirilmesi”, *Ankara Üniversitesi Eğitim Fakültesi Dergisi*, 88: 325-334 (1980).

Albayrak, A. S. ve Yılmaz, Ş. K., “Veri madenciliği: karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14 (1): 31-52 (2009).

Altıntop, Ü., “İnternet tabanlı öğretimde veri madenciliği tekniklerinin uygulanması”, Yüksek Lisans Tezi, *Kocaeli Üniversitesi Fen Bilimleri Enstitüsü*, Kocaeli (2006).

Avouris, N., Komis, V., Fiotakis, G., Margaritis, M. and Voyiatzaki, E. “Why logging of fingertip actions is not enough for analysis of learning activities”, *Workshop on Usage Analysis in Learning Systems AIED Conference*, Amsterdam, 1-8 (2005).

Aydın, S., “Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama”, Doktora Tezi, *Anadolu Üniversitesi Sosyal Bilimler Enstitüsü*, Eskişehir (2007).

Baker, R., Corbett, A. and Koedinger, K., “Detecting student misuse of intelligent tutoring systems”, *7th International Conference On Intelligent Tutoring Systems*, Brazil, 3220: 531-540 (2004).

Beck, J. and Woolf, B., “High-level student modeling with machine learning”, *5th International Conference On Intelligent Tutoring Systems*, Canada, 1839: 584-593 (2000).

Becker, K., Ghedini, C. G. and Terra, E. L., “Using KDD to analyze the impact of curriculum revisions in a Brazilian university” *SPIE 14th Annual International Conference*, Orlando, 412-419 (2000).

Can, A., “SPSS ile Bilimsel Araştırma Sürecinde Nicel Veri Analizi”, *Pegem Akademi Yayınları*, Ankara (2013).

Chen, G., Liu, C., Ou, K. and Liu, B., “Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology”, *Journal Of Educational Computing Research*, 23 (3): 305-332 (2000).

Çağiltay, N. E. “İş Zekası ve Veri Ambarı Sistemleri”, *ODTÜ Yayıncılık*, Ankara (2010).

Çokluk, Ö., “Lojistik regresyon analizi: kavram ve uygulama”, *Kuram ve Uygulamada Eğitim Bilimleri*, 10 (3): 1357-1407 (2010).

Dringus, L. and Ellis, T., “Using data mining as a strategy for assessing asynchronous discussion forums”, *Computer & Education Journal*, 45: 141-160 (2005).

Emel, G. G. ve Taşkın, Ç., “Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması”, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6 (2): 221-239 (2005).

Erdoğan, Ş. ve Timor, M., “A data mining application in a student database”, *Havacılık ve Uzay Dergisi*, 2 (2): 57-64 (2005).

Göksu, İ., “Web tabanlı öğrenme ortamında veri madenciliğine dayalı öğrenci değerlendirmesi”, Yüksek Lisans Tezi, *Fırat Üniversitesi Eğitim Bilimleri Enstitüsü*, Elazığ (2012).

Gözüdeli, Y., “MS SQL ile Temel Veritabanı Programlama, Tasarım ve Gerçekleme”, *Seçkin Yayıncılık*, Ankara (2005).

Gürsoy, U. T., “Veri Madenciliği ve Bilgi Keşfi”, *Pegem Akademi Yayınları*, Ankara (2009).

Gürsoy, U. T., “Uygulamalı Veri Madenciliği Sektörel Analizler”, *Pegem Akademi Yayınları*, Ankara (2012).

Ha, S. H., Bae, S. M., and Park, S.-C., “Web mining for distance education”, *Proceedings of the 2000 IEEE International Conference*, 2: 715-719 (2000).

Hamalainen, W., Suhonen, J., Sutinen, E., and Toivonen, H., “Data mining in personalizing distance education courses” *World Conference On Open Learning and Distance Education*, Hong Kong (2004).

Höçük, S., “Ankara üniversitesi uzaktan eğitim programına katılan öğrencilerin akademik başarılarını etkileyen faktörler”, Yüksek Lisans Tezi, *Ankara Üniversitesi Eğitim Bilimleri Enstitüsü*, Ankara (2011).

İnternet: Ahmet Yesevi Üniversitesi, “Ahmet Yesevi Üniversitesi Tarihçesi”, http://www.yesevi.edu.tr/index.php?option=com_content&view=article&id=1&Itemid=118 (2014).

İnternet: Alpaydın, E. “Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri”, http://www.cmpe.boun.edu.tr/~ethem/files/papers/veri-maden_2k-notlar.doc (2000).

İnternet: Aydın, Ö., “Dosya Organizasyonu”, http://ozlemaydin.trakya.edu.tr/DO_1 (2014).

İnternet: Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R., “CRISP-DM 1.0 Step-By-Step Data Mining Guides”, <ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf> (2000).

İnternet: KBUZEM, “Resmi Web Adresi”, <http://kbuzem.karabuk.edu.tr> (2014).

İnternet: KBÜ e-Katalog, “Karabük Üniversitesi e-Katalog”, <http://www.karabuk.edu.tr/katalog/index.html> (2014).

İnternet: Makina Mühendisleri Odası, “Uzaktan Eğitimle Mühendis Yetiştirilmez!”, http://www.mmo.org.tr/resimler/bizden/orj/4993_14_57_44.jpg (2009).

İnternet: Merriam-Webster Online Dictionary, “Data”, <http://www.merriam-webster.com/dictionary/data> (2013).

İnternet: ÖSYM, “Tablo 4 Merkezi Yerleştirme ile Öğrenci Alan Yükseköğretim Lisans Programları”, http://dokuman.osym.gov.tr/pdfdokuman/arsiv/2009/2009_OSYS_TERCIH_KILAVUZU/tablo4.pdf (2009).

İnternet: ÖSYM, “Tablo 4 Merkezi Yerleştirme ile Öğrenci Alan Yükseköğretim Lisans Programları”, http://dokuman.osym.gov.tr/pdfdokuman/2010/OSYS/Kilavuzlar/TercihKilavuzu/2010_OSYS_Tablo4.pdf (2010).

İnternet: ÖSYM, “Tablo 4 Merkezi Yerleştirme ile Öğrenci Alan Yükseköğretim Lisans Programları”, <http://osym.gov.tr/dosya/1-57952/h/2011tablo4-2172011.pdf> (2011).

İnternet: ÖSYM, “2012 Yükseköğretim Programları ve Kontenjanları Kılavuzu”, <http://dokuman.osym.gov.tr/pdfdokuman/2012/OSYS/2012OSYSKONTKILAVUZ.pdf> (2012).

İnternet: Shih, Y., “QUEST Classification Tree”, <http://www.math.ccu.edu.tw/~yshih/quest.html> (2014).

İnternet: Sumbüloğlu, K., “Lojistik Regresyon Analizi”, http://78.189.53.61/-/bs/ess/k_sumbuloglu.pdf (2014).

İnternet: Türk Mühendis ve Mimar Odaları Birliği, “TMMOB ve Mühendislik Eğitimi”, http://tmmob.org.tr/resimler/ekler/7bc2f75bf1bcfe8_ek.pdf (2006).

İşman, A., “Uzaktan Eğitim”, *Pegem Akademi*, Ankara (2008).

Karakaya, İ., “Öğretmenlik programlarındaki öğrencilerin ÖSS puanları ile akademik başarıları arasındaki ilişkinin incelenmesi”, *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2 (1): 155-163 (2011).

Kaya, Z., “Öğretim Teknolojileri ve Materyal Geliştirme”, *Pegem A Yayıncılık*, Ankara (2006).

Lu, J., “Personalized e-learning material recommender system”, *International Conference On Information Technology for Application*, Harbin, 374-379 (2004).

Luan, J., “Data mining, knowledge management in higher education, potential applications.” *Workshop Associate of Institutional Research International Conference*, Toronto, 1-18 (2002).

Mazza, R. and Milani, C., “Exploring usage analysis in learning systems: Gaining insights from visualisations”, *Workshop On Usage Analysis In Learning Systems At 12th International Conference On Artificial Intelligence In Education*, Amsterdam, The Netherlands, 65-72 (2005).

Merceron, A. and Yacef, K., “A web-based tutoring tool with mining facilities to improve learning and teaching.”, *Proceedings Of 11th International Conference On Artificial Intelligence In Education*, 201-208 (2003).

Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E. and Heiner, C., “An educational data mining tool to browse tutor–student interactions: Time will tell!” *Proceedings Of The Workshop On Educational Data Mining*, 15-22 (2005).

Muehlenbrock, M., “Automatic action analysis in an interactive learning environment.”, *Conference On Artificial Intelligence In Education*, Amsterdam, The Netherlands (2005).

Myers, R. S. and Pyles, M. R., “Relationships among high school grades, ACT test scores, and collage grades.”, *Mid-South Educational Research Association*, Knoxville, TN (1992).

Noble, J. and Sawyer, R., “Predicting different levels of academic success in college using high school GPA and ACT composite score”, *ACT Research Report Series*, (2014).

Oğuzlar, A., “Temel Metin Madenciliği”, *Dora Yayınları*, Bursa (2011).

Öz, R. ve Yüksel, B., “Uzaktan eğitim: Makine mühendisliği örneği.”, *Erzincan Eğitim Fakültesi Dergisi*, 9 (2): 167-182 (2007).

Özkan, Y., “Veri Madenciliği Yöntemleri”, *Papatya Yayıncılık*, İstanbul (2013).

Özkul, A. E., “E-Öğrenme ve Mühendislik Eğitimi”, TMMOB *Elektrik Mühendisleri Odası*, 419 (2003).

Öztemel, E. , “Yapay Sinir Ağları.”, *Papatya Yayıncılık*, İstanbul (2012).

Pahl, C. and Donnellan, C., "Data mining technology for the evaluation of web-based teaching and learning systems.", *Proceedings Of The Congress E-Learning*, Montreal, Canada (2003).

Romero, C. and Ventura, S. "Educational data mining: A survey from 1995 to 2005.", *Expert Systems With Applications*, 33 (1): 135-146 (2007).

Romero, C., Ventura, S., Pechenizkiy, M., and Baker, R., "Handbook of Educational Data Mining", *Chapman and Hall/CRC Press*, US (2010).

Romero, C. and Ventura, S., "Data mining in course management systems: Moodle case study and tutorial", *Computers & Education*, 55: 368–384 (2008).

Romero, C., Ventura, S. and Bra, P. D., "Knowledge discovery with genetic programming for providing feedback to courseware author", *The Journal Of Personalization*, 14 (5): 425–464 (2004).

Sanjeev, A. P. and Zytow, J. M., "Discovering enrollment knowledge in university databases", *Proceedings Of The First International Conference On Knowledge Discovery And Data Mining AAAI Press*, California, 246-251 (1995).

Shen, R., Yang, F. and Han, P., "Data analysis center based on e-learning platform", *Proceedings Of The 5th International Workshop*, 19-28 (2002).

Silahtaroglu, G., "Kavram ve Algoritmalarıyla Temel Veri Madenciliği.", *Papatya Yayıncılık*, İstanbul (2008).

Sumathi, S. and Sivanandam, S., "Introduction to data mining and its applications", *Springer*, Berlin (2006).

Şen, B. ve Uçar, E., "Evaluating the achievements of computer engineering department of distance education students with data mining methods", *Procedia Technology*, 262-267 (2012).

Şengür, D. ve Tekin, A., "Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini", *Bilişim Teknolojileri Dergisi*, 6 (3): 7-16 (2013).

Şentürk, A., "Veri Madenciliği Kavram ve Teknikler.", *Ekin Yayınevi*, Bursa (2006).

Tang, C., Lau, R. W., Li, Q., Yin, H., Li, T., and Kilis, D., "Personalized courseware construction based on web data mining", *Proceedings of the First International Conference on Web Information Systems Engineering*, 204-211 (2000).

Taşdemir, M. "Veri madenciliği (öğrenci başarısına etki eden faktörlerin regresyon analizi ile tespiti)", Yüksek Lisans Tezi, *Dicle Üniversitesi Sosyal Bilimler Enstitüsü*, Diyarbakır (2012).

Tüzüntürk, S., "Veri madenciliği ve istatistik", *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 29 (1): 65-90 (2010).

Ullah, H., and Wilson, M. A., “Öğrencilerin akademik başarısı ve bunun öğrencinin öğrenmeye katılımı, fakülte ve arkadaş ilişkileriyle bağlantısı”, Çeviri Editörü: Dam, H, *Dinbilimleri Akademik Araştırma Dergisi*, 9 (4): 179-195 (2009).

Yarımağan, Ü., “Veri Tabanı Sistemleri”, *Akademi Yayıncılık*, Ankara (2010).

Zaiane, O., and Luo, J., “Web usage mining for a better web-based learning environment”, *In Proceedings Of Conference On Advanced Technology For Education*, Banff, Alberta, 60-64 (2001).

Zaiane, O., Xin, M., and Han, J., “Discovering web access patterns and trends by applying OLAP and data mining technology on web logs”, *Research And Technology Advances In Digital Libraries*, 19-29 (1998).

EK AÇIKLAMALAR A.

ARAŞTIRMA İZİN BELGESİ



T.C.
KARABÜK ÜNİVERSİTESİ
Mühendislik Fakültesi Dekanlığı

12 Şubat 2015

Sayı : 23534334-900/ 52-511
Konu : Bilgilendirme

REKTÖRLÜK MAKAMINA
(Öğrenci İşleri Daire Başkanlığı)

Üniversitemiz fen Bilimleri Enstitüsü Bilgisayar Mühendisliği A.B.D. 2009628126001 numaralı öğrencisi Abdulkadir TAŞDELEN' in Dekanlığımıza vermiş olduğu dilekçesine istinaden Yüksek Lisans Tez konusu "Veri Madenciliği Yöntemleri ile Mühendislik Fakültesi Uzaktan Eğitim Bölümlerinin Analiz Edilmesi" başlıklı çalışması için Fakültemiz Bilgisayar ve Endüstri Mühendisliği öğrencilerine ait MOODLE-Öğrenim Yönetim Sistemi kayıt bilgilerinin kullanılması uygun görülmüştür.

Bilgilerinizi ve gereğini arz ederim.


Prof. Dr. Erol ARCAKLIOĞLU
Dekan

EK: Dilekçe (1 sayfa)



T.C.
KARABÜK ÜNİVERSİTESİ REKTÖRLÜĞÜ
Öğrenci İşleri Dairesi Başkanlığı

Sayı :73548361 – 900/1137

Konu :Bilgilendirme


12 NİSAN 2013

Sayın: Öğr. Gör. Abdulkadir TAŞDELEN

İlgi: Mühendislik Fakültesi Dekanlığının 12/02/2013 tarih ve 52-511 sayılı yazısı.

11/02/2013 tarihli dilekçenize istinaden Başkanlığımıza gönderilen ilgi yazı gereğince; istemiş olduğunuz bilgiler ekte sunulmaktadır.

Bilgilerinizi rica ederim.


Mustafa CÜNÜK
Öğrenci İşleri Dairesi Başkanı

EKLER:
EK -1 Yazı örneği (2 sayfa)
EK -2 CD (1 adet)

12/04/2013 Bilg. İşl. : S.İŞİK Sİ
12/04/2013 Şub.Müd. : H.DURAK

Kalp ve damar hastalıklarından korunmak için sağlıklı beslenmeye özen gösteriniz.

Karabük Üniversitesi Rektörlüğü Öğrenci İşleri Daire Başkanlığı 78050 KARABÜK Tel: (370) 433 83 19 Faks: (370)433 83 70
ogrenci.isleri@karabuk.edu.tr

ÖZGEÇMİŞ

Abdulkadir TAŞDELEN 1985 yılında Hatay ili Antakya ilçesinde doğdu; ilkokulu Yahya Çavuş İlkokulu'nda tamamladı. Ortaokula Ezine İmam Hatip Lisesi'nde başlayıp Konya İmam Hatip Lisesi'nde devam etti. Lise eğitimini Karatay SDMP Anadolu Lisesi'nde tamamladı. Lisans eğitimine 2003 yılında Selçuk Üniversitesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü'nde başlayıp 2007 yılında mezun oldu. Aynı zamanda ikinci üniversite olarak Anadolu Üniversitesi İşletme Bölümü'ne 2004 yılında kaydolup 2008 yılında mezun oldu.

2007-2010 yılları arasında bilişim teknolojileri öğretmeni olarak Milli Eğitim Bakanlığı'na bağlı Orhanlı İlköğretim Okulu'nda çalıştı. 2010 yılında Karabük Üniversitesi'nde öğretim görevlisi olarak görev aldı ve 2010-2013 yılları arasında uzaktan eğitim merkezinde değişik kademelerde görev yaptı. 2013 yılında Yükseköğretim Kurulu Başkanlığı'nda görevlendirildi ve görevine aynı kurumda devam etmektedir. Evli ve iki çocuk babasıdır.

Adres: Karabük Üniversitesi

Balıklarkayası Mevkii / KARABÜK

Tel : (505) 360 5044

E-posta: abdulkadirtasdelen@gmail.com