

Ders Yönetim Sistemi Tabanlı Veri Madenciliđi ve Öğrenme Analitiđi

Seda Tuzcu

**YÜKSEK LİSANS TEZİ**

Endüstri Mühendisliđi Anabilim Dalı

Nisan 2018



Course Management System based Data Mining and Learning Analytics

Seda Tuzcu

**MASTER OF SCIENCE THESIS**

Department of Industrial Engineering

April 2018

Ders Yönetim Sistemi Tabanlı Veri Madenciliği ve Öğrenme Analitiği

Seda Tuzcu

Eskişehir Osmangazi Üniversitesi

Fen Bilimleri Enstitüsü

Lisansüstü Yönetmeliği Uyarınca

Endüstri Mühendisliği Anabilim Dalı

Yöneylem Araştırması Bilim Dalında

YÜKSEK LİSANS TEZİ

Olarak Hazırlanmıştır

Danışman: Prof. Dr. Muzaffer Kapanoğlu

Nisan 2018

## ONAY

Endüstri Mühendisliği Anabilim Dalı Yüksek Lisans öğrencisi Seda Tuzcu'nun YÜKSEK LİSANS tezi olarak hazırladığı "Ders Yönetim Sistemi Tabanlı Veri Madenciliği ve Öğrenme Analitiği" başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek oybirliği ile kabul edilmiştir.

**Danışman** : Prof. Dr. Muzaffer Kapanoğlu

**İkinci Danışman** : —

### **Yüksek Lisans Tez Savunma Jürisi:**

**Üye** : Prof. Dr. Muzaffer Kapanoğlu

**Üye** : Prof. Dr. Hasan Durucasu

**Üye** : Dr. Öğr. Üyesi Yeliz Buruk Şahin

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ..... tarih ve  
..... sayılı kararıyla onaylanmıştır.

Prof. Dr. Hürriyet ERŞAHAN  
Enstitü Müdürü

## ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. Muzaffer Kapanođlu danışmanlığında hazırlamış olduđum “Ders Yönetim Sistemi Tabanlı Veri Madenciliđi ve Öğrenme Analitiđi” başlıklı YÜKSEK LİSANS tezimin özgün bir çalışma olduđunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandıđımı; tezimde verdiđim bilgileri, verileri akademik ve bilimsel etik ilke ve kurallara uygun olarak elde ettiđimi; tez çalışmamda yararlandıđım eserlerin tümüne atıf yaptıđımı ve kaynak gösterdiđimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduđumu beyan ederim. 27/04/2018

Seda Tuzcu

## ÖZET

Ülkemizde yükseköğretim kurumlarında, ders yönetim sistemleri henüz yakın bir geçmişte kullanılmaya başlanmıştır. Bu konuda çok aktif üniversiteler olduğu gibi pek çok üniversitede de böyle bir sistem bulunmamaktadır. Çalışmanın gerçekleştirildiği Eskişehir Osmangazi Üniversitesinde ise Enformatik Bölümü sorumluluğunda 2005 yılından beri Moodle Ders Yönetim Sistemi kullanılmaktadır. Günümüzde giderek artan büyük veri kavramı, Ders Yönetim Sistemi veri tabanlarında da karşımıza çıkmaktadır. Büyük verinin yönetime, akademik personele, öğrencilere ve eğitim sürecine hiçbir katkı sağlamadan depolanması bu alanda veri madenciliği uygulamalarını kullanmanın gerekliliğine işaret etmektedir. Veri madenciliği konusu ele alındığında ise ders yönetim sistemleri ile ilgili çalışmaların diğerlerinden sayıca geride kaldığı gözükmemektedir. Bu nedenle, özellikle yükseköğretim kurumlarında ders yönetim sistemlerinin kurulumu ve kullanımının önemine, yapılan akademik çalışmalarla dikkat çekilmesi şüphesiz eğitim sürecinin kalitesini artıracaktır. Bu çalışmada ESOGÜ Enformatik Bölümü Ders Yönetim Sisteminden elde edilen gerçek kullanıcı verileri RapidMiner veri bilimi yazılımı ile analiz edilmiştir. Öğrencilerin dersler bazında başarı tahmini yapılmış ve algoritma performansları karşılaştırılmıştır. Kullanılan her bir sınıflandırma algoritmasının farklı veri setlerinde farklı performanslar gösterdiği gözlenmiştir. Ders Yönetim Sisteminden elde edilen değişkenlerin farklı veri setlerinde kullanıldığında sınıflandırma üzerindeki etkileri tartışılmıştır.

**Anahtar Kelimeler:** Veri Madenciliği, Öğrenme Analitiği, Ders Yönetim Sistemi, Moodle, RapidMiner

## SUMMARY

Course management systems have begun to be used in not too distant past in higher education of Turkey. Beside the fact that there are many universities using these systems actively, majority of them don't have a system at all. Moodle Course Management System has been used in Eskişehir Osmangazi Üniversitesi in which this study takes place since 2005 by Department of Informatics. We run across the concept of big data which is growing rapidly nowadays in course management systems' databases too. Storing the big data without contributing to management, academic staff, students, or education process shows the necessity of using data mining techniques in this field. Considering data mining, there are the studies about course management systems but few in number compared to others. For this reason, the studies pointing out the significance of using course management systems especially in higher education will surely raise the quality of education. In this study, real user data gathered from ESOGÜ Informatics Course Management System has been analyzed with RapidMiner data science software. Students' performances are predicted in terms of courses and performances of algorithms are compared. It is observed that the classification algorithms used in the study shows different performances in different datasets. The importance of attributes obtained from the course management system is discussed in terms of prediction process.

**Keywords:** Data Mining, Learning Analytics, Course Management System, Moodle, RapidMiner

## TEŞEKKÜR

Veri madenciliđi konusuyla tanışmamı sađlayan, tez alıřması boyunca deđerli fikirleriyle yol gsterip desteđini esirgemeyen tez danıřmanım Sayın Prof. Dr. Muzaffer Kapanođlu'na en iten teřekkrlerimi sunarım. alıřmanın uygulama kısmında ihtiya duyulan ham verinin, ESOG Enformatik Ders Ynetim Sisteminden elde edilmesine yardımcı olan Sayın đr. Gr. zgr Gltekin'e katkılarından tr teřekkr ederim. Son olarak, her kořulda ve her zaman yanımda olan ailem ve eřime gsterdikleri anlayıř ve destekleri iin ok teřekkr ederim.

Seda Tuzcu

## İÇİNDEKİLER

	<b><u>Sayfa</u></b>
<b>ÖZET</b> .....	vi
<b>SUMMARY</b> .....	vii
<b>TEŞEKKÜR</b> .....	viii
<b>İÇİNDEKİLER</b> .....	ix
<b>ŞEKİLLER DİZİNİ</b> .....	xii
<b>ÇİZELGELER DİZİNİ</b> .....	xiii
<b>SİMGELER VE KISALTMALAR DİZİNİ</b> .....	xiv
<b>1. GİRİŞ VE AMAÇ</b> .....	1
<b>2. VERİ MADENCİLİĞİ</b> .....	3
2.1. Genel Bakış .....	3
2.2. Veri Madenciliği Tanımı .....	4
2.3. Bilgi Keşfi .....	5
2.4. Değişken Kavramı .....	7
2.5. Veri Madenciliğinde Karşılaşılan Problemler .....	8
2.6. Veri Madenciliği Yöntemleri .....	8
2.6.1. Gözetimli öğrenme .....	9
2.6.1.1. <u>Sınıflandırma</u> .....	9
2.6.1.2. <u>Regresyon</u> .....	10
2.6.2. Gözetimsiz öğrenme .....	10
2.6.2.1. <u>Kümeleme</u> .....	10
2.6.2.2. <u>Birliktelik kuralları</u> .....	11
2.7. Veri Madenciliği Yazılımları .....	11
2.8. Veri Madenciliğinin Uygulama Alanları .....	11
2.9. Eğitim Alanında Veri Madenciliği .....	12
<b>3. ÖĞRENME ANALİTİĞİ</b> .....	14
<b>4. DERS YÖNETİM SİSTEMLERİ</b> .....	15
4.1. Bilgisayar Destekli Eğitim .....	15
4.2. Ders Yönetim Sistemi .....	15

## İÇİNDEKİLER (devam)

	<b><u>Sayfa</u></b>
4.3. Ders Yönetim Sistemi Yazılımları .....	16
4.3.1. Moodle .....	17
4.4. ESOĞÜ Enformatik Ders Yönetim Sistemi .....	17
<b>5. LİTERATÜR ARAŞTIRMASI .....</b>	<b>19</b>
<b>6. MATERYAL VE YÖNTEM .....</b>	<b>23</b>
6.1. Uygulamada Kullanılan Yazılım .....	23
6.1.1. Avantajlar .....	23
6.1.2. Zayıf yanlar .....	24
6.1.3. Genel bakış .....	24
6.1.3.1. <u>Ekran ve çalışma ortamı</u> .....	24
6.1.3.2. <u>Operatörler ve süreçler</u> .....	24
6.1.3.3. <u>Modelleme</u> .....	25
6.1.3.4. <u>Veri aktarımı</u> .....	25
6.1.3.5. <u>Filtreleme ve sıralama</u> .....	26
6.1.3.6. <u>Grup birleştirme/yaratma ve silme</u> .....	27
6.1.3.7. <u>Tip/rol değiştirme</u> .....	28
6.2. Uygulamada Kullanılan Algoritmalar .....	29
6.3. Model Performansı .....	30
6.3.1. Doğruluk ve hata oranı .....	31
6.3.2. ROC analizi .....	32
6.4. Verilerin Elde Edilmesi ve Hazırlık Süreci .....	32
6.4.1. Veri setleri .....	32
6.4.1.1. <u>Assign modülü</u> .....	35
6.4.1.2. <u>Forum modülü</u> .....	35
6.4.1.3. <u>Message modülü</u> .....	35
6.4.1.4. <u>Quiz modülü</u> .....	35
6.4.1.5. <u>Resource modülü</u> .....	36
6.4.1.6. <u>Folder modülü</u> .....	36

**İÇİNDEKİLER (devam)**

	<b><u>Sayfa</u></b>
6.4.1.7. <u>Files modülü</u> .....	36
6.4.1.8. <u>Stats modülü</u> .....	36
6.4.2. Hazırlık süreci .....	36
6.4.3. Model .....	37
<b>7. BULGULAR VE TARTIŞMA</b> .....	<b>39</b>
7.1. İncelenen Dersler .....	39
7.1.1. Doğrusal sistemler .....	39
7.1.2. Diferansiyel denklemler .....	41
7.1.3. İşbilim .....	44
<b>8. SONUÇ VE ÖNERİLER</b> .....	<b>46</b>
<b>KAYNAKLAR DİZİNİ</b> .....	<b>48</b>

## ŞEKİLLER DİZİNİ

<b><u>Sekil</u></b>	<b><u>Sayfa</u></b>
2.1. Veri madenciliğinin diğer disiplinler ile ilişkisi .....	5
2.2. Veri tabanlarında bilgi keşfi süreci .....	6
6.1. RapidMiner Studio çalışma ortamı .....	25
6.2. Veri aktarım ekranından bir kesit .....	26
6.3. Filtreleme ve sıralama .....	27
6.4. Grup birleştirme/yaratma ve silme .....	28
6.5. Titanic veri seti .....	29
6.6. Sistemde en fazla kullanıcı hareketine sahip ilk 25 bölüm .....	34
7.1. Sınıflandırmada kullanılan algoritma performansları (doğruluk) .....	40
7.2. İlk durumda algoritmaların performansı (doğruluk) .....	42
7.3. İkinci durumda algoritmaların performansı (doğruluk) .....	42
7.4. K-means algoritması uygulanmamış halde performanslar .....	43
7.5. K-means algoritması uygulandıktan sonra performanslar .....	44
7.6. Algoritma performansları .....	45

## ÇİZELGELER DİZİNİ

<b><u>Cizelge</u></b>	<b><u>Sayfa</u></b>
2.1. Veri madenciliğinde kullanılan yöntemler .....	9
6.1. İki sınıflı bir veri kümesine ait karışıklık matrisi .....	31



**SİMGELER VE KISALTMALAR DİZİNİ**

<b><u>Kısaltmalar</u></b>	<b><u>Açıklama</u></b>
DYS	Ders yönetim sistemi
ESOGÜ	Eskişehir Osmangazi Üniversitesi
KDD	Knowledge discovery in databases
LMS	Learning management system
VLE	Virtual learning environments
SCORM	Sharable content object reference model
ROC	Receiver operating characteristic
TP	True positive
FN	False negative
FP	False positive
TN	True negative
GLM	Genelleştirilmiş lineer model
LR	Lojistik regresyon
VMT	Virtual math teams

## 1. GİRİŞ VE AMAÇ

Genellikle yükseköğretim kurumlarında olmak üzere, eğitimcilerin eğitimle ilgili her türlü içeriği web ortamı üzerinden öğrencilerle ve diğer kullanıcılarla paylaşmasını sağlayan sistemler Ders Yönetim Sistemleri olarak adlandırılmaktadır. Geleneksel bir eğitim-öğretim sürecindeki zaman ve mekân sınırlılıkları ders yönetim sistemleri ile aşılmaktadır. Teknolojinin de giderek gelişmesiyle bu gibi sistemlere olan talep artmaktadır.

Günümüzde internet ve bilgisayar kullanımının giderek yaygınlaşması sonucu sahip olunan veri miktarı devasa boyutlara ulaşmıştır. Bu boyutlardaki veriden, anlamlı bilgi elde edilmediği sürece veriye sahip olmanın ya da miktarının bir anlamı olmayacaktır. Karar vericinin bu denli devasa veriyi kendince inceleyip anlamlı bilgiye ulaşması imkânsız olduğundan tam da bu noktada veri madenciliği teknikleri devreye girmektedir.

Ders yönetim sistemi kullanılan kurumlarda veri tabanlarındaki kullanıcılara ilişkin verilerin yönetime, eğitimcilere, öğrencilere ve eğitime hiçbir katkı sağlamadan depolanması bu alanda veri madenciliği uygulamalarını kullanmanın gerekliliğine işaret etmektedir. Söz konusu uygulamalar sonunda elde edilebilecek anlamlı bilgiler ışığında sistemleri geliştirecek müdahalelerde bulunmak mümkündür.

Bu çalışmada ESOGÜ Enformatik Ders Yönetim Sistemi (DYS) üzerinde bir veri madenciliği uygulaması yapılmıştır. Çalışmanın motivasyonu, ders yönetim sistemi kullanılan bir yükseköğretim kurumunda, herhangi bir derse kayıt yaptıran herhangi bir öğrencinin performansı hakkında henüz akademik dönem bitmeden anlamlı bilgiye ulaşmanın mümkün olup olmadığının araştırılmasıdır. Çalışmanın genel amaçları ise aşağıdaki gibidir:

- Elle toplanamayacak karmaşıklıkta ve devamlı artış gösteren veride gizli olan anlamlı ve değerli bilginin keşfi,
- Eğitim alanında veri madenciliğinin kullanım alanlarını araştırmak,
- Konuyla ilgili geçmiş çalışmaları araştırıp analiz etmek,
- Örnek bir ders yönetim sistemi üzerinde veri madenciliği gerçekleştirmek,

- Gerçekleştirilen veri madenciliği algoritmaları sonucu elde edilen sonuçları değerlendirmektir.

Uygulama kısmında ESOGÜ Enformatik Bölümü Ders Yönetim Sisteminden elde edilen gerçek kullanıcı verileri RapidMiner veri bilimi yazılımı ile analiz edilmiştir. Öğrencilerin dersler bazında başarı tahmini yapılmış ve algoritma performansları karşılaştırılmıştır. Ders Yönetim Sisteminden elde edilen değişkenlerin farklı veri setlerinde kullanıldığında sınıflandırma üzerindeki etkileri tartışılmıştır.

Çalışma sekiz bölümden oluşmaktadır. İlk bölümde ele alınan konu ve ele alınma gerekçelerinden bahsedilmektedir. İkinci bölümde konuya genel bakış, veri madenciliğinin tanımı, bilgi keşfi, değişken kavramı, veri madenciliğinde karşılaşılan problemler, veri madenciliği yöntemleri, veri madenciliği yazılımları, veri madenciliğinin uygulama alanları ve eğitim alanında veri madenciliği ele alınmaktadır. Üçüncü bölümde öğrenme analitiği konusu işlenmektedir. Dördüncü bölümde bilgisayar destekli eğitim, ders yönetim sistemi, ders yönetim sistemi yazılımları ve ESOGÜ Enformatik Ders Yönetim Sistemi hakkında bilgi verilmektedir. Beşinci bölümde konuyla ilgili geçmiş çalışmalardan bahsedilmektedir. Altıncı bölümde uygulamada kullanılan yazılım, bu yazılımın tercih edilme nedenleri, uygulamada kullanılan algoritmalar, verilerin elde edilmesi ve hazırlık süreci hakkında bilgi verilmektedir. Yedinci bölümde elde edilen bulgular açıklanmaktadır. Son bölümde ise çalışma değerlendirilerek sonuç ve öneriler verilmektedir.

## 2. VERİ MADENCİLİĞİ

### 2.1. Genel Bakış

Veri, tek başına bir anlam ifade etmeyen ya da kullanılmayan fakat bilgiye temel oluşturan, ilişkilendirilmeye, gruplandırılmaya, yorumlanmaya, anlamlandırılmaya ve analiz edilmeye gereksinim duyulan ham bilgidir.

Bilgisayarlı veri toplama teknolojilerinin gelişmesi, dağıtık ve mobil teknolojiler, bilgi sistemlerinin kurumların temel bileşenlerinden biri haline gelmesi, yanı sıra yazılım ve donanım yönlü maliyetlerin gitgide azalması ile veri okyanusları ortaya çıkmıştır. Yönetimlerin ve yönetsel araçların bilgisayarların varlığına rağmen veri açlığı çektiği aşağı yukarı elli yıllık bir dönem bitmiş, 2000’li yıllarla birlikte her türlü olgunun kayıt altına alındığı yeni bir dönem başlamıştır. İnsan ve ait olduğu evrene ilişkin hemen her şeyin elektronik veri tabanlarında tutulmaya başlandığı bu dönem, büyük veri (big data) dönemi olarak adlandırılmaktadır (Kapanoğlu vd., 2016). Devasa boyutlardaki bu verilerin ele alınmasındaki en temel zorluk bu büyük hacimdeki veriyi çalıştırmak, uygun analizi bulmak, yapılan analizle yararlı bilgiyi çıkarıp gelecekte oluşabilecek olaylar için çıkarım yapmaktır. Günümüzde farklı sektörlerden karar vericiler büyük oranda ve hızla artan verilerin daha doğru ve etkin analizi için yarışmaktadır. Çağımızda en değerli kaynağın zaman olduğunu göz önünde bulundurursak büyük veriyi kısa sürede analiz edecek yöntem ve teknolojiye duyulan ihtiyaç hayli fazladır.

Büyük veri, geleneksel veri işleme yöntemleriyle ele alınamayacak büyüklükte ve karmaşık veri yığınlarından oluşmaktadır. Büyük verinin karakteristik özellikleri aşağıdaki gibidir (Kudyba, 2014):

- Hacim (volume): Veri miktarıdır.
- Hız (velocity): Verinin üretim ve yayılmasındaki hızı temsil eder. Günümüzde gerçek zamanla eş sayılacak raddededir.
- Çeşitlilik (variety): Verinin yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış farklı sınıflardan oluşan verileri temsil eder.

- Değer (value): Veri içindeki yapı ve ilişkinin diğer bir deyişle veriden bilginin keşfidir. Her karar verici geleceğe dair sağlam adımlar atmak isteyeceğinden bu bilgiye erişmek son derece önemlidir.

## 2.2. Veri Madenciliği Tanımı

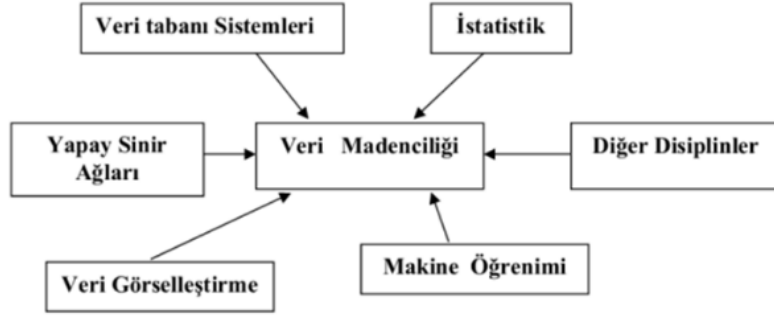
Veri madenciliği terimi zaman içinde bilgi çıkarımı, bilgi keşfi, bilgi toplama, veri arkeolojisi ve veri model işleme gibi birçok isimle anılsa da 90'lı yıllara gelindiğinde veri tabanı komitelerince veri madenciliği terimi kullanılmaya başlanmıştır (Fayyad vd., 1996).

Madencilik kavramı, veri madenciliği sürecinin yer altındaki değerli taşların aranması sürecine benzetilmesinden kaynaklanmaktadır (Berry ve Linoff, 2004). Günümüzde veriye verilen önemin giderek artması ve içinde gizlenen cevher niteliğindeki bilgiye erişme isteği veri madenciliğinin popülaritesini artırmıştır. Veri madenciliği mevcut durumun analizini sağlayarak geleceğe yönelik çıkarımlarla karar verme mercilerinin işini kolaylaştırmaktadır.

Veri madenciliği ile ilgili tanımlardan bazıları aşağıdaki gibidir:

- Veriden önceden bilinmeyen, muhtemelen yararlı sayılacak bilginin çıkartılması işlemidir (Piatetsky-Shapiro, 1989).
- Veriden örüntü ya da desen elde etmek için özel algoritmaların kullanıldığı veri tabanlarından anlamlı bilginin keşfi sürecinin bir aşamasıdır (Fayyad vd., 1996).
- İstatistik, veri tabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veri tabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizidir (Hand, 1998).
- Veri tabanlarında, veri ambarlarında veya diğer veri depolarındaki büyük veri içindeki ilginç bilgileri keşfetme sürecidir (Han ve Kamber, 2006).
- Veri tabanlarında, veri ambarlarında ve diğer veri depolarında saklanan büyük miktardaki veriden ilginç bilgiler keşfetme sürecidir (Wahbeh vd., 2011).

- Veri tabanlarından bilgi derlemek ve derlenen bu bilgiyi anlamlı kurallara dönüştürerek anlaşılmayı kolaylaştırmak için kullanılan teknikler bütünüdür (Sewaiwar ve Verma, 2015).



Şekil 2.1. Veri madenciliğinin diğer disiplinler ile ilişkisi (Han ve Kamber, 2006)

Veri madenciliğinin diğer disiplinlerle ilişkisi Şekil 2.1.'de verilmiştir.

Veri madenciliğinin karar vericilere sağladığı olanaklar (Kapanoğlu vd., 2016):

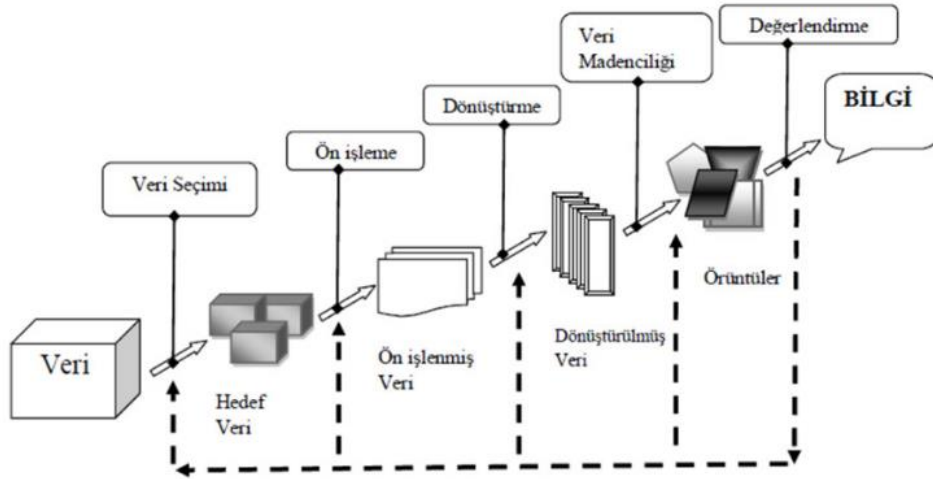
- Data surfing: Karar vericilerin verilere göz gezdirmesi
- Drill-down: Sadece gerektiğinde ayrıntıya inebilmesi
- Alerts: Anormal sapmalar hakkında uyarılması
- Forecasting: Geleceğe ilişkin güvenilir tahminlerin yapılabilmesi
- Olsa-ne olur analizleri: Senaryo analizleri uygulanabilmesi

### 2.3. Bilgi Keşfi

Veri madenciliği ile büyük veri yığınlarında gizli, maden niteliğindeki anlamlı ve değerli bilginin elde edilmesi hedeflenmektedir. 80'li yıllarla birlikte büyük veriden bilgi elde edebilme düşüncesi sorgulanmaya başlanmıştır. Bu bağlamda Shapiro tarafından veri tabanlarında bilgi keşfi (KDD) terimi ortaya atılmıştır (Li, 2016). Veri madenciliği bu sürecin bir adımıdır.

Veri tabanlarında bilgi keşfi sürecini meydana getiren aşamalar aşağıdaki gibidir (Han ve Kamber, 2006):

- Veri temizleme: Tutarsız, hatalı ya da eksik verilerin elenmesi işlemidir.
- Veri bütünleştirme: Farklı veri tabanlarından çekilen verilerin analiz için birleştirilmesi işlemidir.
- Veri seçme: Veri madenciliğinde kullanılacak verilerin seçilmesi işlemidir.
- Veri dönüştürme: Seçilen verilerin analiz için hazır olmadığı durumlarda, sağlıklı sonuçlardan kaçınmak için bu verileri analize hazırlama işlemidir.
- Veri madenciliği: Anlamlı bilgiye ulaşmak için kullanılan metotların oluşturduğu analiz işlemidir.
- Örüntü değerlendirme: Veri madenciliği sonucunda gizli kalmış bir kural, ilişki ya da örüntü olup olmadığının tespit edilmesi işlemidir.
- Bilgi sunumu: Bulunan örüntülerin kişilere sunumu için uygun tekniklerin belirlenmesi işlemidir.



Şekil 2.2. Veri tabanlarında bilgi keşfi süreci (Anonim, 2017)

Veri tabanlarında bilgi keşfi süreci Şekil 2.2.'de verilmiştir. Veri, durum bilgisi ve yöntem bilgisi ilişkisi ise aşağıdaki gibidir (Kapanoğlu vd., 2016):

- Veri: Faaliyetlerin içeriğine uygun veri performansı (bellek ve hız) ilkeleri gözetilerek tasarlanmış ve genellikle ilişkisel veri tabanlarında tutulan kayıtlardır. Bir kaydı veri yapan şey, bir işlem, bir formül, bir hesaplama içermemesidir.
- Durum bilgisi: Anlamlandırılabilen, verilerden bir sorgu, bir formül, bir hesaplama veya bir işlem sonucu elde edilen bilgiler için kullanılır. Karar

vericinin kararını etkileme, belirleme ve deęiřtirme yeteneęi olan bilgi türüdür.

- Yöntem bilgisi: Problem çözme ve karar vermede durum bilgilerinin nasıl kullanılacağı konusundaki bilgileri de içeren türde bilgilerdir. Hangi durum bilgilerinin hangi eylemlerle ve nasıl ilişkilendirilmesi gerektięi üzerinedir. Deneyim ve eğitimle geliştirilebilir. Bu yüzden kişiye göre deęiřir.

## 2.4. Deęiřken Kavramı

Üzerinde ölçüm yapılacak varlıęın, sayılarla ya da ait olduęu kategori veya grup ile ifade edilecek şekilde farklı deęerler alabilen özellięine deęiřken denir (Anonim, 2016). Deęiřken yapıları birbirinden farklıdır, her biri farklı şekillerde, farklı birimlerle ve kavramlarla ifade edilir. Bu nedenle yapısal olarak nitel (kategorik) ve nicel (sayısal) olmak üzere iki alt gruba ayrılmıřlardır.

Nicel deęiřken; aldıęı deęerin miktarı derecelendirilebilen, sıralanabilen, işlem yapılabilen bir sayı ile ifade edilebilen deęiřkendir. Kendi içerisinde iki alt sınıfa ayrılır:

- Sürekli deęiřken: Sayı doğrusunda her noktada bulunma olasılıęı olan deęiřken (ölçülebilir)
- Kesikli deęiřken: Sayı doğrusunda, tamsayılarla örtüşen noktaların dışında, her noktada yer alma olasılıęı olmayan deęiřken (sayılabilir)

Nitel deęiřken; bir miktar ile ifade edilemeyen, bir kategori veya sınıfı belirten deęiřkendir.

Neden-sonuç ilişkisi gözetilen deneysel arařtırmalarda söz konusu bir dięer deęiřken ayrımı da ařaęıdaki gibidir:

- Baęımlı deęiřken: Manipüle edilemeyen, arařtırmanın asıl konusu olup, ölçülecek olan deęiřken
- Baęımsız deęiřken: Karar vericinin amaçları doğrultusunda manipüle edilebilen bir dięer deyiřle istenen grup ve düzeylerde tanımlanabilen ve baęımlı deęiřken üzerindeki etkisi arařtırılmak istenen deęiřken

Değişken kavramından hareketle veri; değişkenin aldığı değerlere denir (Anonim, 2016).

## 2.5. Veri Madenciliğinde Karşılaşılan Problemler

Veri madenciliğinde karşılaşılan problemler aşağıda sıralanmıştır:

- Veri tabanı boyutu; veri madenciliği sistemleri için önemli problemlerden biridir. Bu sorunu çözmek için uygulanabilecek yöntemlerden bazıları: veri madenciliği sisteminin sezgisel bir yaklaşımla arama uzayını taraması, örneklemin yatay ve dikey boyutta indirgenmesi olarak sayılabilir (Oğuz, 2000).
- Gürültülü veri; veri toplanması esnasında oluşan sistem dışı hatalar sonucu oluşur.
- Eksik veri; veri kümesindeki kayıtların eksikliğinden ya da bazı kayıtlar için bazı değişkenlere ait değerlerin olmamasından kaynaklanır.
- Artık veri; veri kümesindeki, eldeki probleme uygun olmayan gereksiz değişkenlere ait verilerdir. Bu bağlamda değişken seçimi yapılması uygulanan veri madenciliği yönteminin kalitesini artırmaktadır.
- Null veri; değişkenin bilinmeyen bir değere sahip olmasıdır. Uygulama esnasında bu değerler tümüyle göz ardı edilebilmekte ya da olası bir değer atanabilmektedir. Null değerler, söz konusu değişkendeki en yüksek frekansa sahip bir değer ya da ortalama bir değer alınabilir (Quinlan, 1986).
- Dinamik veri; içeriği sürekli olarak değişen veri tabanlarındaki değişen verileri belirtmektedir.

## 2.6. Veri Madenciliği Yöntemleri

Veri madenciliğinde kullanılan yöntemler tahminleyici (predictive) ve tanımlayıcı (descriptive) modeller olmak üzere iki sınıfa ayrılmaktadır. Bu sınıfların detayları Çizelge 2.1.'de gösterilmektedir.

Çizelge 2.1. Veri madenciliğinde kullanılan yöntemler

<b><u>Tahminleyici/Gözetimli öğrenme</u></b>	<b><u>Tanımlayıcı/Gözetimsiz öğrenme</u></b>
Sınıflandırma (classification)	Kümeleme (clustering)
Regresyon (regression)	Özetleme (summarization)
Zaman serisi analizi (time series analysis)	Birliktelik kuralları (association rules)
Kestirim (prediction)	Sıralılık keşfi (sequence discovery)

### 2.6.1. Gözetimli öğrenme

Tahminleyici yöntemler önceden sonuçları bilinen verilerden bir model geliştirip geliştirilen modeli sonucu bilinmeyen bir veri seti üzerinde uygulayarak bir sonuç tahmin etmektedir. Bir diğer deyişle gözetimli öğrenme (supervised learning), girdi ve çıktılardan hareketle bir model oluşturup, yeni bir girdi verisi verildiğinde oluşacak çıktıyı tahmin etme işlemidir. Veriler ortak özelliklerine göre etiketlenerek belirli sınıflara atanmaktadır. Gözetim kelimesinden kasıt ise süreci denetleyen bir eğitim mekanizmasının olmasındandır. Girdi değişkenleri bağımsız, çıktı değişkenleri ise bağımlı değişkenlerdir.

#### 2.6.1.1. Sınıflandırma

Sınıflandırma, bir nesnenin özelliklerinin incelenerek belirli sınıflar içerisinde hangisine ait olduğunun tespit edilmesi anlamına gelmektedir (Alpaydın, 2000). Sınıflandırmak istenen değişken bağımlı ve onun dışında kalanlar bağımsız değişken olarak adlandırılmaktadır. Buradaki amaç; tahmin edici olarak kullanılan girdi değişkenlerinin bulunduğu bir modelde çıktı olarak bağımlı bir değişkenin olduğu anlamlı bir model kurulmasıdır. Bağımlı değişken sayısal değil kategorik bir değer ise problem bir sınıflandırma problemi, bağımlı değişken sayısal bir değer ise regresyon problemi olarak adlandırılmaktadır (Gülçe, 2010).

### **2.6.1.2. Regresyon**

Sınıflandırma yönteminde olduğu gibi regresyonda da amaç bilinmeyen bir çıktı değişkeni için bir değer belirlemektir. Sınıflandırmada çıktı değişkeni kategorik iken regresyonda nümeriktir. Örneğin, bir banka müşterilerinin kredi geri ödemelerini güvenli ya da riskli olarak kategorize etmek istediğinde sınıflandırma modeli; bir satış yöneticisi demografik bilgileri bilinen müşterilerin harcamalarını tahmin etmek istediğinde regresyon modeli kullanabilmektedir.

### **2.6.2. Gözetimsiz öğrenme**

Tanımlayıcı yöntemler, karar vermeye yönelik olarak veriler arasındaki örüntülerin tanımlanması sağlamaktadır. Gözetimsiz öğrenme (unsupervised learning) söz konusu olduğunda çıktı değişkeni mevcut değildir fakat model oluşturmak için kullanılan tüm değişkenler bağımsız değişkenlerdir yani girdilerdir. Gözetimsiz öğrenmede amaç girdi değişkenleri ile çıktı değişkenlerini tahmin etmektir (Gorunescu, 2011). Gözetimli öğrenmeden farkı bir denetim mekanizması olmamasıdır.

#### **2.6.2.1. Kümeleme**

Sınıflandırma tekniğinde sınıflar daha önceden belirli iken, kümeleme tekniğinde sınıflar daha önceden belli değildir. Yani sınıflandırma tekniğinden farklı olarak kümeleme analizinde ne kadar grup oluşacağı da belirli değildir (Şekeroğlu, 2010). Kümelemede tahmin edilecek bir hedef değer yani etiket yoktur. Amacı aynı grup içindeki nesnelerin birbirleriyle benzerliklerini (ilişkilerini) ve gruplar arasındaki farkı tespit etmektir. Bir grup içindeki benzerlik (homojenlik) ve gruplar arasındaki fark ne kadar büyük olursa o küme o kadar iyi veya belirgin olur. İki nesnenin ne kadar benzer olduğunu bulmak için mesafe ölçüleri kullanır. Aynı kümede bulunan nesnelerin birbirine yakın olması beklenir. Dolayısıyla nesnelere benzerse mesafeleri de kısadır. En çok kullanılan mesafe ölçüleri Öklid ve Manhattan mesafeleridir (Bhan vd., 2013). Kümeleme yöntemindeki amaç; üyelerinin birbirine benzer olduğu ancak özelliklerinin birbirinden çok farklı olduğu kümelerin tespit edilmesi ve veri tabanındaki bilgilerin elde edilen bu farklı kümelere bölünmesidir.

Kümeleme analizi yapılırken, veri tabanındaki kayıtların hangi kümelere ayrılacağı veya kümeleme işleminin hangi değişken özellikleri dikkate alınarak yapılacağı bir uzman tarafından belirlenebileceği gibi kümeleme yazılımları tarafından da yapılabilmektedir (Han ve Kamber, 2006).

Gözetimsiz öğrenmenin gözetimli öğrenme için bir değerlendirme aracı olarak kullanılması alışlagelmiştir. Birçok veri başlangıçta sınıf değişkenine sahip olmayabilir. Bu gibi bir durumda kümeleme yöntemi kullanılarak sınıf tayin edilmesi ve ardından uygulanacak sınıflandırma veya regresyon analizleri ile veri çözümlemesi yapılabilir.

### **2.6.2.2. Birliktelik kuralları**

Olayların birlikte gerçekleşmiş olduğu durumlarını çözümleyen veri madenciliği yöntemleri de birliktelik kuralları (association rules) olarak adlandırılır (Şekeroğlu, 2010). Veri kümesindeki kayıtlar arası bağlantıları aramayı sağlar. Sıklıkla kullanıldığı perakende sektöründe pazar sepet analizi olarak adlandırılır.

## **2.7. Veri Madenciliği Yazılımları**

Veri madenciliği yazılımları gruplanacak olursa ticari ve ücretsiz (açık kaynak kodlu) yazılımlar olarak ikiye ayırmak mümkündür. Açık kaynak kodlu yazılımların kullanıcılara verdiği geliştirme yetkisi sayesinde bu yazılımlar daha hızlı gelişmekte ve eksiklikler daha kolay güncellenebilmektedir.

Ticari yazılımlardan en bilinenleri; SPSS Clementine, MS Excel, SAS, SQL Server, Oracle, MATLAB olarak sıralanabilir. Açık kaynak kodlu popüler veri madenciliği yazılımları ise; WEKA, RapidMiner, Orange, KNIME ve R gibi yazılımlardır.

## **2.8. Veri Madenciliğinin Uygulama Alanları**

Veri madenciliğinin en önemli faydalarından biri de hemen hemen her alanda kullanılabilir olmasıdır (Gorunescu, 2011). Veri madenciliğinin en bilindik kullanım

alanları; bankacılık, finans, borsa, sağlık, mühendislik, kamu, internet, eğitim, perakende ve pazarlama süreçleridir.

Veri madenciliğinin kullanıldığı alanlarda başarılı çözüme ulaştığı örnek problemler aşağıdaki gibidir:

- Benzer satın alma örüntüleri gösteren müşterilerin doğru segmentlerde ifade edilmesi,
- Bireysel ilişki yönetimi için etkin bir şekilde müşteri profillerinin çıkarılması,
- Maillere gelen yanıt oranının artırılması,
- Hangi müşterilerin daha sadık ve hangilerinin belirli promosyonlara cevap verme olasılığının daha yüksek olduğunu tanımlayacak müşteri profillerinin oluşturulması,
- Bir müşteriyi rakip şirketi tercih etmeye iten faktörlerin neler olduğunu anlamak,
- Satın alma örüntülerini, ödemeleri ve tepki oranlarını etkileyen faktörleri keşfetmek,
- Bir kredi kartı işleminin ya da sigorta talebinin sahtekârlıkla sonuçlanıp sonuçlanmayacağını tahmin etmek,
- Kredi kartı müşterilerinin verilen bir zaman çerçevesinde hesaplarını başka bir bankaya nakledip nakletmeyeceklerini tahmin etmek,
- Geçmişlerine ve karakteristiklerine bakarak müşterilerin gelecekteki davranışlarını tahmin etmek,
- Tıp merkezlerine ve sigorta şirketlerine hangi prosedürlerin birleşiminin en çok istenen çıktıları üreteceğini belirleyerek maliyet yönetiminde yardım etmek (Han ve Kamber, 2006).

## 2.9. Eğitim Alanında Veri Madenciliği

Günümüzde her eğitim kademesinde eğitim uygulamalarının önemli bir parçası olan öğrencilerin; kişisel bilgileri, notları, başarılı ve başarısız olduğu dersler gibi birçok bilgi geniş veri tabanlarında tutulmaktadır. Anlamlı ilişkilerin araştırılabileceği ve önemli bilgilerin elde edilebileceği bu veri yığınları eğitimde aksaklıklara neden olan problemlerin

tespitinde ve eğitimin kalitesinin arttırılmasında kullanılabilir. Eğitimdeki bu verilerin analiz edilmesi ve veriler arasındaki örüntülerin ortaya çıkarılması veri madenciliği yoluyla gerçekleştirilebilir (Özby, 2015).

Eğitimde veri madenciliği uygulamaları geleneksel sınıf ortamında ve uzaktan eğitim ortamında olmak üzere iki şekilde gerçekleştirilmektedir. Eğitimde veri madenciliği uygulamaları öğrencilerin tam olarak izlenmesinin güç olduğu geleneksel eğitimde daha az kullanılırken, öğrencilerin izlenmesinin daha kolay olduğu ve öğrenme ortamlarında gerçekleştirilen pek çok öğrenci davranışının kaydedildiği uzaktan eğitimde daha geniş uygulama potansiyeli bulmaktadır (Zaiane, 2001).

Geleneksel sınıf ortamında veri madenciliğinin kullanımı ilk olarak Sanjeev ve Zytchow tarafından 1995 yılında gerçekleştirilmiştir. Araştırmacılar üniversite veri tabanından elde ettikleri öğrenci kayıtlarını anlamlandırmaya çalışmışlardır. Elde ettikleri sonuçları üniversite yönetimine sunarak kurumun stratejik kararlar vermelerinde bu bilgilerden yararlanmalarını sağlamışlardır (Özby, 2015).

### 3. ÖĞRENME ANALİTİĞİ

Teknoloji destekli öğrenmenin hayatımıza girmesiyle birlikte “öğrenme analitikleri” adlı yeni bir araştırma alanı ortaya çıkmıştır (Elias, 2011). Eğitsel veri madenciliği, akademik analitikler, sosyal ağ analizleri gibi farklı dallardan tekniklerin devşirilmesiyle ortaya çıkan bu kavram yararlı bilgi elde etmek, kişinin önceki öğrenmelerini yansıtmak ve eğitim-öğretimi geliştirmek için kullanılır (Dyckhoff vd., 2012).

Dijital bir platformda gerçekleştirilen her türlü aktivite dijital bir ayak izi bırakır. Bu noktadan hareketle ders yönetim sistemlerinde de kullanıcıların gerçekleştirdiği her bir eylemin dijital kaydının olduğu söylenebilir. Öğrenme analitikleri tam bu noktada devreye girer ve riskte olan öğrencilerin durumu yahut geleceği hakkında eğitimciye geribildirim verir. (Bahçeci, 2015).

Öğrenme analitiği kavramı ilk kez 2010’da Siemens tarafından; öğrenme üzerinde tahmin ve tavsiye yapabilmek için akıllı veri, öğrenenin ürettiği veri, bilgi ve sosyal bağlantıları keşfetmek için analizlerin kullanılması olarak tanımlanmıştır. Daha sonra düzenlenen 1. Uluslararası Öğrenme Analitikleri ve Bilgi Konferansı (2011)’nda yeniden tanımlanmıştır. Buna göre öğrenme analitiği; öğrenme ve öğrenmenin gerçekleştiği ortamları anlamak ve optimize etmek amacıyla; öğrenenler hakkında verilerin toplanması, ölçümü, analizi ve raporlanmasıdır (Bahçeci, 2015).

Yeni bir alan olması nedeniyle bu konuda ülkemizde yayınlanmış tez sayısı sadece iki tanedir: Adanır (2016), doktora tezinde bilgisayar destekli ortaklaşa öğrenme ortamı olan Virtual Math Teams (VMT) çevrimiçi ortamındaki öğrenme sürecinin öğrenme analitiklerinin kullanılarak değerlendirilmesi üzerine bir çalışma gerçekleştirmiştir. Çalışmada, VMT ortamında gerçekleşen ortaklaşa öğrenme sürecinin analizi için sosyo-tekniik yaklaşım kullanılmıştır. Kayabaş (2017), doktora tez çalışmasında Öğrenme Bulutu olarak adlandırılan yeni bir öğrenme analitikleri sistemi geliştirilerek kullanılabilirliğine ilişkin araştırmalar yapmıştır.

## 4. DERS YÖNETİM SİSTEMLERİ

### 4.1. Bilgisayar Destekli Eğitim

Bilgisayar destekli eğitim, bilgisayarların sistem içine programlanan dersler yoluyla öğrencilere bir konu ya da kavramı öğretmek ya da önceden kazandırılan davranışları pekiştirmek amacıyla kullanılmasıdır (Yalın, 2006).

Bilgisayar destekli eğitimin bazı avantajları aşağıdaki gibidir:

- Büyük kitlelere ulaşmak,
- Eğitimde fırsat ve imkân eşitliği sağlamak,
- Bölgesel uzmanlarla sınırlı kalmamak,
- Eğitim sürecindeki mesafe boyutunu ortadan kaldırmak,
- Eğitim maliyetlerini azaltmak.

### 4.2. Ders Yönetim Sistemi

Ders yönetim sistemi; öğretmen, öğrenci, yönetici üçlüsünün birbiri ile etkileşimini sağlayan, her zaman verilen yetkiler doğrultusunda çift yönlü (öğrenci-öğrenci, öğrenci-öğretmen, öğretmen-yönetici) iletişim içinde olmasını sağlayan ağ teknolojilerinin ve standartlarının, programlama dillerinin ve veri tabanlarının kullanıldığı yazılımlardır. Bu yazılımlar eş zamanlı olmayan öğrenme materyali sunma, sunulan materyali değişik biçimlerde paylaşma ve tartışma, derslere kayıt olma, ödevler alma, sınavlara girme, bu ödev ve sınavlara ilişkin geri bildirim sağlama, öğrenme materyallerini düzenleme, öğrenci, öğretmen ve sistem kayıtlarını tutma, raporlar alma gibi olanakların ağ üzerinden otomatik olarak gerçekleşmesini sağlayan yazılımlardır (Çekiç, 2010).

Bir DYS'nin genel yapısı aşağıdaki özelliklere sahip olmalıdır (Tekin, 2007):

- Kullanıcıların tanımlanması ve yönetilmesi,
- Ders içeriklerinin hazırlanması,
- Derslerin yönetilmesi,

- Öğrenciye özel programların açılması,
- Ödev ve proje verilmesi/teslimi
- Sınav/testlerin hazırlanması ve uygulanması,
- Öğrenci davranışlarının izlenmesi ve incelenmesi,
- Öğrencilerin başarı durumlarının değerlendirilmesi,
- Etkileşimli iletişim ortamlarının oluşturulması ve yönetilmesi.

DYS'lerde aranılan içeriğe kolayca ulaşılması, farklı üreticilerden alınan içerik ve araçların birlikte çalıştırılması, bir ÖYS'de hazırlanan içeriğin başka bir ÖYS'ye taşınabilmesi için standartlar ve spesifikasyonlar oluşturulmaktadır. Spesifikasyonlar, standartlardan daha az gelişmiş olup kişilere yaptıkları iş hakkında bilgi vermektedir. Standartlar ise tanınan veya bilinen organizasyonlar tarafından onaylanmış tanım veya biçimdir. Bu anlamda SCORM (Paylaşılabilir İçerik Nesne Referans Modeli) standartları, eğitsel içeriğin en küçük birimlerinden sistemin çalışma ortamına kadar çok kapsamlı bir standartlar çatısı ortaya koymaktadır. Bu nedenle bir DYS tasarlanırken mutlaka göz önünde bulundurulması gerekir. E-öğrenme alanında geliştirilmiş birçok standart ve tanımlamalar tek bir başvuru modeli çatısı altında toplanmaya başlamış ve oluşturulan bu modele SCORM denilmiştir (Çoban, 2016).

#### 4.3. Ders Yönetim Sistemi Yazılımları

DYS yazılımları, öğrenme yönetim sistemi (learning management system) ya da sanal öğrenme ortamları (virtual learning environments) olarak da anılmaktadır. Web üzerinde öğrenme aktivitelerinin yönetimini sağlayan bu yazılımlar ile öğrenci, öğretmen ve yöneticinin çevrimiçi öğrenim hizmetlerine erişimi organize edilir. Söz konusu yazılımlar ticari ya da açık kaynak kodlu olmak üzere ikiye ayrılmakta olup en popüler olanları aşağıdaki gibidir:

- Ticari: Blackboard LMS, WebCT, Angel, GlobalLearning, Ilinc LMS, Oracle LMS, Plateau, SAP, SkillSoft.
- Açık kaynak kodlu: Moodle, Sakai LMS, Spaghettilearning, OpenUSS, ILIAS, ATutor, Dokeos, dotLRN.

### 4.3.1. Moodle

Moodle, Modular Object Oriented Dynamic Learning Environment (Modüler Nesne Yönelimli Dinamik Öğrenme Ortamı) yazılımını dünya üzerinde 232 ülkede, 99.517 kurumda, 128.622.274 kullanıcı tarafından kullanılmaktadır. 1999 yılında Martin Dougiamas tarafından Avustralya'daki Perth Üniversitesinde geliştirilen Moodle, 77 farklı dil desteği sunmaktadır (Moodle, 2018).

Moodle, 2005 yılından beri çekirdek bir takım ve Moodle Partners organizasyon yapısı altında gönüllü kullanıcılar ile geliştirilmeye devam edilmektedir. PHP desteği olan sunucu ortamlarında çalışabilmekte ve veri tabanı olarak MySQL ya da PostgreSQL kullanılmaktadır. Platform bağımsız olarak çalışan uygulama kolay yönetilebilen ve görsel tema zenginliği ile rahat özelleştirilebilen esnek bir yapıya sahiptir. En önemli özelliği ve tercih sebebi herkes tarafından çok kolay şekilde kullanılmasıdır (Özarslan, 2008).

Moodle, modüllerden oluşan bir yazılımdır ve içerdiği standart modüllerin yanı sıra yazılımın resmi internet sitesinden ihtiyaca göre çok sayıda farklı modül indirmeye ve sisteme uyarlamaya elverişlidir. Türkiye'de 372'si özel olmak üzere toplamda 571 kurumda Moodle kullanılmaktadır (Moodle, 2018).

### 4.4. ESOĞÜ Enformatik Ders Yönetim Sistemi

Enformatik Bölümünün YÖK tarafından tanımlı görevi; üniversite ve ileri teknoloji enstitülerinin tüm öğrencilerine temel bilgi teknolojileri konularının öğretilmesi için gerekli örgütlenmeler ile düzenlemeler yapmak ve fiilen dersleri koordine etmek ve/veya gerçekleştirmektir. Bu bağlamda Eskişehir Osmangazi Üniversitesi Enformatik Bölümü 2005 yılından beri üniversitedeki tüm derslerin ders yönetim sistemleri ile desteklenmesi hizmetini sunmaktadır. 2016-2017 Bahar Dönemi itibariyle son beş yılda yüz yüze eğitime destek olarak ders yönetim sistemi üzerinde açılan sanal sınıf sayısı 70.000'e yaklaşmıştır (Eskişehir Osmangazi Üniversitesi Enformatik Bölümü, 2017).

Eskişehir Osmangazi Üniversitesinde Moodle yazılımı kullanılmakta olup, Enformatik Bölümü tarafından düzenli aralıklarla Ders Yönetim Sistemi Eğitimleri verilmektedir. Sisteme <http://enf.ogu.edu.tr/golddys/> adresinden ulaşılmaktadır.



## 5. LİTERATÜR ARAŞTIRMASI

Zaiane (2001), eğitimciler ve öğrenme süreçlerinin daha iyi değerlendirilmesi amacıyla web-tabanlı öğrenme ortamının tasarımında veri madenciliği ve makine öğrenmesi tekniklerinin kullanılabilirliğini tartışmıştır.

Minaei-Bidgoli vd. (2003), genetik algoritmalar kullanarak, web tabanlı eğitsel bir veri tabanında bulunan verilere bağlı olarak öğrencilerin final sınavında alacakları notu tahmin etmiştir.

Karabatak ve İnce (2004), öğrencilerin tüm derslerini dikkate alarak tüm ders notları arasındaki ilişkileri birliktelik kuralı ile ortaya çıkarmayı hedefleyen bir çalışma yapmıştır.

Erdoğan ve Timor (2005), Maltepe Üniversitesi öğrenci verilerini kullanarak yaptıkları çalışmada öğrencileri üniversitedeki başarı durumlarına göre kümelere ayrılmıştır. Ortaya çıkan kümelere öğrenciler ÖSS sınavında elde ettikleri yüzdelik dilimlere göre benzer gruplar içerisinde yer almıştır. En başarılı küme olarak Fen Edebiyat Fakültesi'ndeki öğrenciler yer almaktayken (bu öğrenciler burslu olarak öğrenim görmektedir), en başarısız grup İBBF ve İletişim Fakültesi'nde öğrenim gören öğrenciler oluşturmaktadır (ÖSS'de düşük puanla yerleşmiş öğrenciler).

Ben-Zadok vd. (2007), öğrencilerin öğrenme davranışlarını analiz ederek final sınavları öncesi risk altındaki öğrencilerin uyarılmasını sağlayan bir çalışma önermişlerdir.

Romero vd. (2008), çalışmalarında ders yönetim sistemi olan Moodle verilerine birliktelik kuralı, kümeleme, sınıflandırma gibi teknikleri uygulamışlardır.

Ayesha vd. (2010), tarafından geliştirilen yazılım ile okula kayıt yaptıracak öğrencilerin sayıları karar ağacı yöntemi ile tahmin edilerek, idarecilerin gerekli kaynakları hazırlayıp yönetmesi kolaylaşmıştır.

Zhang vd. (2010), veri madenciliğinin risk altındaki öğrencilere nasıl yardım edebileceğini, ders uygunluğunun nasıl değerlendirileceğini ve son olarak elde edilen sonuçların öğrencilere nasıl uyarlanacağını inceleyen bir çalışma gerçekleştirmiştir.

Bozkır vd. (2010), kullanıcıların Facebook kullanma süresi ve sıklığını tahmin eden bir çalışma önererek, Facebook'un eğitime katkısını öğrenci görüşlerine bağlı olarak irdelemiştir.

Bharadwaj ve Pal (2011), veri madenciliği yöntemleri ile bilgisayar uygulama dersini alan 5 farklı üniversiteden toplam 300 lisans öğrencisinin başarımlarını değerlendirmesini gerçekleştirmiştir. Çalışmada 17 değişken kullanılmıştır.

Mamcenko vd. (2011), veri madenciliğindeki birliktelik kuralları ve kümeleme yöntemlerini kullanarak elektronik sınav verilerini analiz etmeyi önermişlerdir. Önerilen çalışmanın amacı elde edilen sonuçları değerlendirmek, yorumlamak ve tanımlayıcı bir model kullanarak elektronik sınav sistemini geliştirmektir.

Ekim (2011), Selçuk Üniversitesinden elde edilen veriler üzerinden, öğrenciler hakkında gelecekle ilgili tahmin yapılabilmesi için gerekli birliktelik kurallarını çıkarmıştır. Selçuk Üniversitesini yeni kazanan bir öğrencinin, üniversitedeki başarısına etki eden faktörler araştırılmış, çalışma sonucunda ailenin eğitim seviyesinin ve gelir düzeyinin öğrencinin başarısında en etkili faktörler olduğu görülmüştür.

Alan (2012), Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü öğrencilerine ait veriler üzerinde bir veri madenciliği uygulaması yapmıştır. Hem verileri en başarılı sınıflandıran algoritma, hem de öğrencilerin programı, cinsiyeti, Sivas ilinden ya da başka bir ilden olması, kadrosunun araştırma görevlisi olup olmaması ve ders döneminin farklı olmasının notlarını etkileyip etkilemediği tespit edilmeye çalışılmıştır.

Romero vd. (2013), öğrencilerin final performanslarının tahmin tutarlılığının iyileştirilmesi için bir çalışma gerçekleştirmiştir. Örnek ve değişkenlerin seçiminin, verinin alınma zamanının sonucu nasıl etkilediği üzerinde durulmuştur.

Alan (2014), başka bir çalışmada karar ağaçları yöntemiyle Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi öğrencilerine ait verilerden yararlanarak, hem bu verileri en başarılı sınıflandıran karar ağaçlarına ait algoritma, hem de bu algoritmanın üreteceği sınıfları tespit etmeye çalışmıştır.

Akçapınar (2014), doktora tezi çalışmasında öğrencilerin derse ilişkin akademik performanslarını geçti/kaldı olarak sınıflandıran, dersten başarısız olacakları tahmin eden, benzer öğrenci profiline sahip öğrencileri gruplandıran bir model geliştirmeye çalışmıştır.

Özarıslan (2014), Kırıkkale Üniversitesinde okuyan birinci sınıf öğrencilerinin Temel Bilgi Teknolojileri Kullanımı dersi için akademik performansları incelemiştir. İnceleme, dersi geleneksel bir yöntem olan yüz yüze eğitim ile alan öğrenciler ile yeni bir yöntem olan uzaktan eğitim ile alan 672 öğrenciye ait veriler veri madenciliği sınıflandırma algoritmaları ile incelenmiştir. Sonuçlara göre karar ağacı oluşturularak öğrenci başarısına etki eden faktörler belirlenmiştir.

Natek ve Zwilling (2014), veri madenciliği araçları ile öğrenci başarı oranı tahminine yönelik bir çalışma gerçekleştirmiştir.

Belsis vd. (2014), kümeleme ve birliktelik kurallarından yararlanarak yükseköğretimde geçirilen süre analizi üzerine bir çalışma gerçekleştirmiştir.

Campagni vd. (2015), üniversiteden mezun olan öğrencilerin kariyer analizleri üzerine bir veri madenciliği metodolojisi sunmuştur.

Öztürk (2015), çalışmasında 2014-2015 eğitim-öğretim yılında Anadolu Üniversitesi Açıköğretim Sistemine kayıtlı öğrencileri kümelere ayırarak, bu kümelere uygun stratejilerin belirlenmesine yön göstermeyi amaçlamıştır.

Kılınç (2015), mevzuattaki başarısızlık nedeniyle öğrencilikten çıkarılma politikasına, öğrencilerin parasal durumlarının ve demografik özelliklerinin etkilerini incelemiştir. Eskişehir Osmangazi Üniversitesi Bilgisayar Mühendisliği Bölümünde 2008-

2011 arasında birinci sınıf derslerine başlayan öğrencilerin verileri kullanılmıştır. Öğrencilerin eğitim sürelerinin, burs veya kredi alınmasıyla değişiklik gösterdiği ve öğrencilerin parasal durumlarıyla annelerinin meslekleri arasında bağlantı olduğu görülmüştür.

Hassana ve Al-Razgan (2016), öğrencilerin üniversite öncesi girdikleri sınav sonuçlarının üniversite not ortalamalarına etkisi olduğunu ortaya koyan bir vaka çalışması gerçekleştirmiştir.

Hamsa vd. (2016), karar ağacı ve bulanık genetik algoritma teknikleri kullanarak; bilgisayar, elektronik ve iletişim bölümleri lisans ve yüksek lisans öğrencilerinin akademik performans tahminine ilişkin bir çalışma gerçekleştirmiştir.

Kınay (2016), Yaşar Üniversitesi e-öğrenme sistemi üzerindeki öğrenci hareketliliğini veri madenciliği teknikleriyle inceleyerek öğrencinin dersten kalıp kalmayacağını öngören bir sistem tasarlanmıştır.

## 6. MATERYAL VE YÖNTEM

### 6.1. Uygulamada Kullanılan Yazılım

Çalışmada kullanılan RapidMiner yazılımı, YALE Üniversitesi tarafından Java dili kullanılarak geliştirilen, verinin ön işlemlerden geçirilip en son raporlama ve görselleştirmeye kadar götürülebildiği bir veri madenciliği yazılımıdır. Açık kaynak kodlu bir yazılım olmasının yanı sıra yakın zamanda profesyonel sürümü de piyasaya sunulmuştur. <http://community.rapidminer.com/> adresli forumu kullanıcı etkileşimi ve geri bildirimler açısından oldukça etkin kullanılmakta ve yeni başlayanlar için öğretici dokümantasyonu mevcuttur (RapidMiner, 2014).

#### 6.1.1. Avantajlar

RapidMiner yazılımının avantajları aşağıda sıralanmıştır:

- Öğrenci ve araştırmacılar için özel bir sürüme (Educational) sahip olup, ücretsiz olması,
- Görselleştirme araçları ve ara yüz bakımından kullanıcı dostu olması,
- Anlaşılır ve modüler (ihtiyaca göre şekillendirilebilen) yapısı,
- Veri ön işleme sürecine yardımcı çok sayıda metoda sahip olması,
- Süreç tasarımı konusunda, gerçek hayat problemlerinin gerekliliklerini karşılamada muazzam esnekliğe sahip olması (Benzer uygulamalara kıyasla daha fazla operatör, analiz basamağı vb.),
- Hız, operatörler, algoritmalar, uyumluluk vb. açılardan sık sık geliştirilen bir yazılım olması,
- Büyük veriyle çalışıldığında da yüksek performans göstermesi,
- RapidMiner geliştiricilerinin kullanıcı soru/isteklerine geri dönüş konusunda oldukça hızlı olmaları
- Oracle, Microsoft SQL Server, PostgreSQL, MySQL gibi veri tabanlarından veri aktarımı yapılabilmesi.

### 6.1.2. Zayıf yanlar

RapidMiner ile ilgili farklı ortamlarda pek çok olumlu yorum bulunmasına karşın olumsuz yorumlar da mevcut olup bunlar öğretici dokümantasyonunun geliştirilmesi gerektiğinde birleşmiştir.

### 6.1.3. Genel bakış

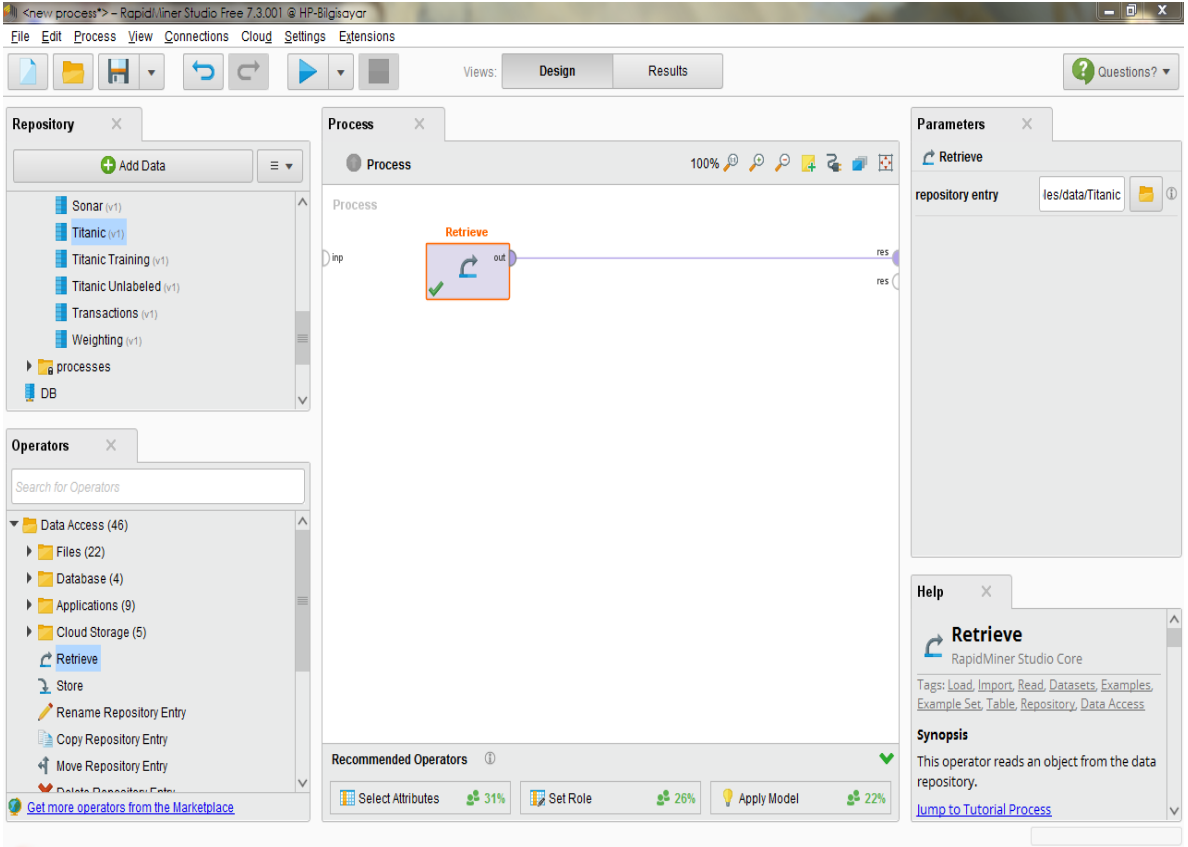
#### 6.1.3.1. Ekran ve çalışma ortamı

Yazılımda veri seti, data table; değişkenler attribute, veriler ise row olarak gösterilmektedir. Çalışma ortamında kullanıcıya yardımcı paneller aşağıdaki gibidir:

- Toolbar: Araç çubuğu
- Repository: Örnek/kullanıcı veri setlerinin depolandığı panel
- Operators: Veri madenciliği operatörlerini içeren panel
- Parameters: Operatör parametre ayarlarının yapıldığı panel
- Help: Yardım
- Process: Süreç akışının tasarlandığı panel

#### 6.1.3.2. Operatörler ve süreçler

- Yazılımda süreçler, operatörlerin eklenmesi ve birbirine bağlanması ile yaratılır.
- Operatörler birbirine port'lar aracılığıyla bağlanır.
- Run butonu ile operatörler aksiyonlarını gerçekleştirir.
- Bir operatörün çıktısını görmek için, operatörün out (output) portu, Process panelindeki res (result) portuna bağlanarak Run butonu tıklanmalıdır.
- Retrieve operatörü, Repository panelinden veri yükleme işlemini gerçekleştirir. Retrieve operatörü ile hangi veri seti yüklendiyse, Run butonu sonrası verilerin ham hali Results sekmesinde incelenebilir hale gelir. Statistics sekmesinde ise veri seti ile ilgili çeşitli istatistiksel özetler bulunmaktadır (Şekil 6.1.).



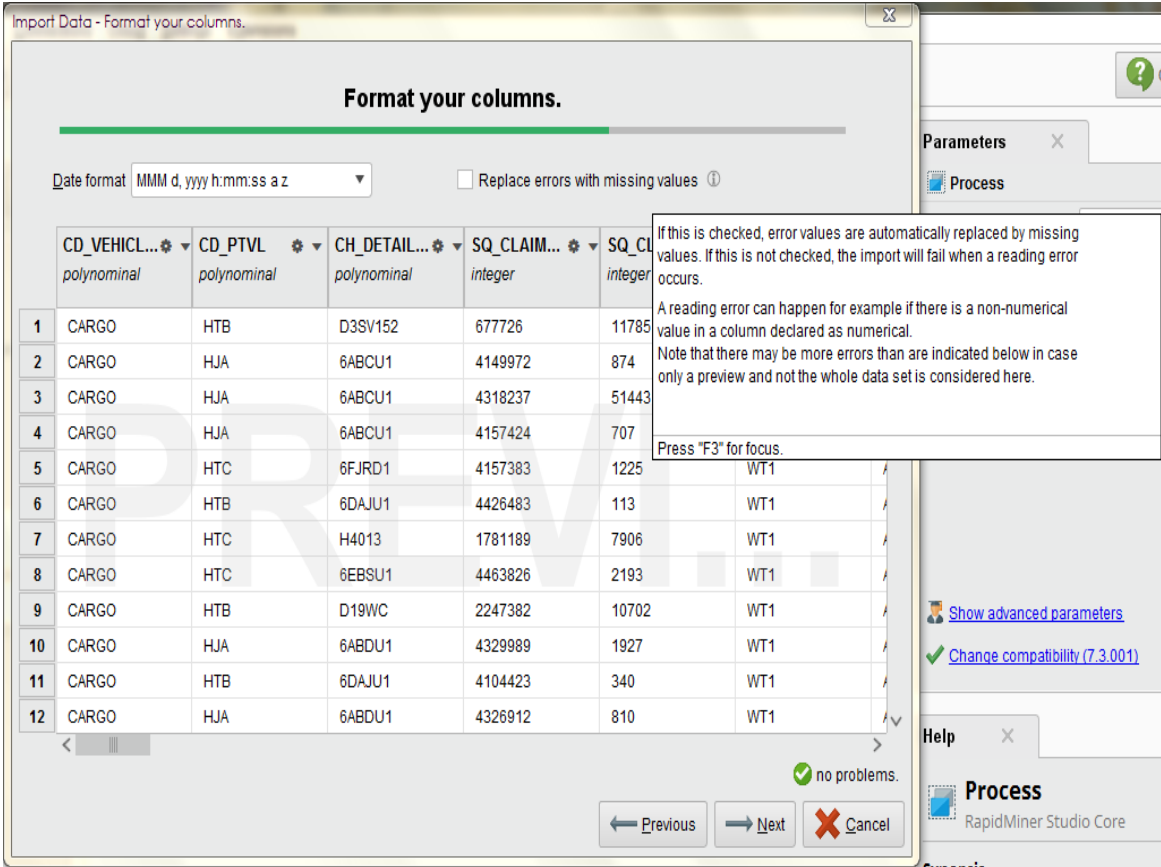
Şekil 6.1. RapidMiner Studio çalışma ortamı

### 6.1.3.3. Modelleme

Modelleme (modeling), çeşitli operatörleri birleştirerek istatistiksel bir model yaratma sürecidir. Örneğin karar ağacı algoritması, bir veri setindeki saklı örüntüleri bulan popüler bir modelleme tekniğidir. Yazılım içinde bunun gibi pek çok algoritma bulunmaktadır.

### 6.1.3.4. Veri aktarımı

Repository panelinde bulunan, Add Data sekmesi tıklanarak istenen veri içe aktarılabilir (Şekil 6.2.). Aktarılan veri Local Repository alanında saklanmaktadır. Veri setinde boş hücreler varsa onlar da '?' ile belirtilerek, missing value (kayıp veri) olarak nitelendirilir.

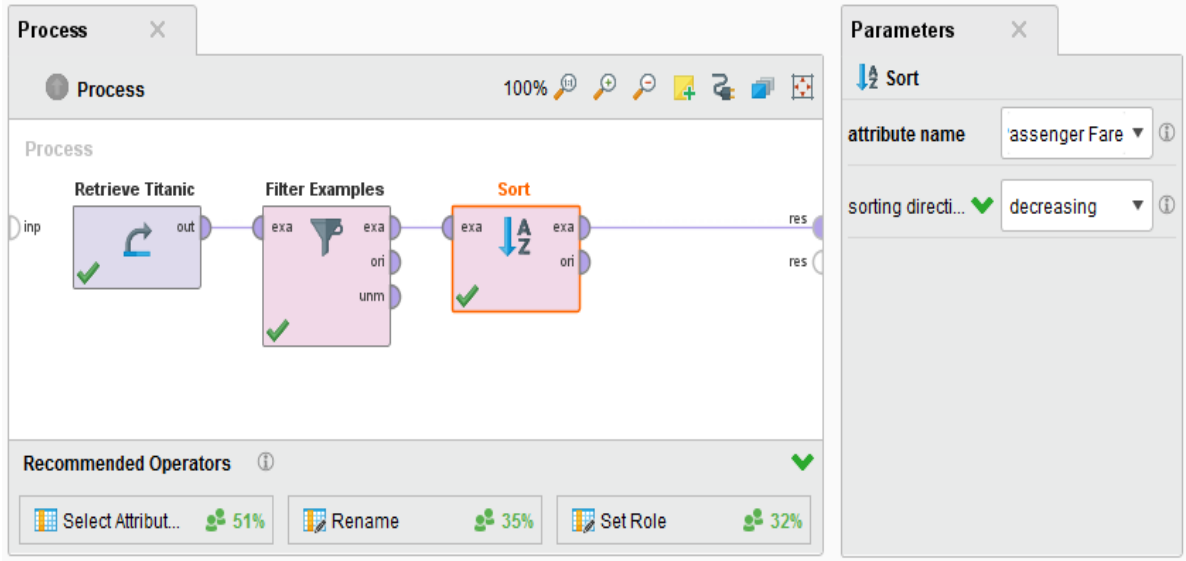


Şekil 6.2. Veri aktarım ekranından bir kesit

### 6.1.3.5. Filtreleme ve sıralama

Operators panelinden Filter Examples operatörü seçilerek ilgili veri kümesine bağlanır. Parameters panelindeki; Add Filters seçeneği kullanılarak veri kümesi üzerinde istenen filtre tanımlanır.

Benzer şekilde Operators panelinden Sort operatörü eklenerek de istenen değişkene göre sıralama gerçekleştirilebilir. İlgili ekran Şekil 6.3.'de verilmiştir.

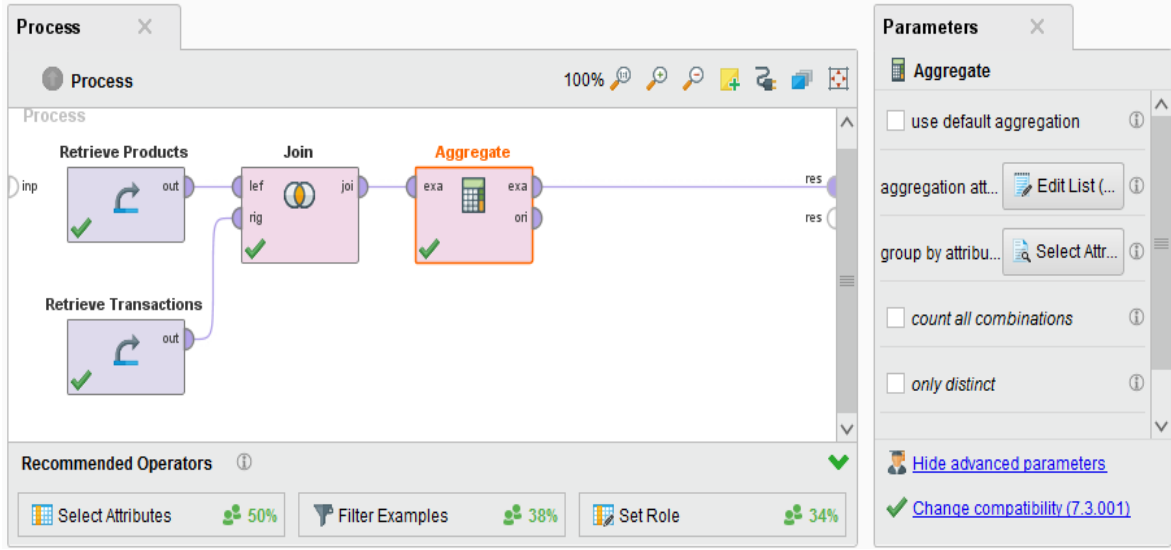


Şekil 6.3. Filtreleme ve sıralama

#### 6.1.3.6. Grup birleştirme/yaratma ve silme

Kimi durumlarda kullanıcılar bazı veri setlerini birleştirerek gereksiz sütunları ayıklamak, gereklileri bir araya getirerek sadece ihtiyaç duyulan bilgileri sağlamak isteyebilirler. Bu gibi durumlar için Join ve Aggregate operatörleri kullanılmaktadır. Join operatörü ile iki veri seti birleştirilebilir. Aggregate operatörü ise bilindik veri tabanı yönetim sistemlerindeki group-by-function fonksiyonuna denktir.

İlgili RapidMiner ekran görüntüsü Şekil 6.4.'te verilmiştir.



Şekil 6.4. Grup birleştirme/yaratma ve silme

### 6.1.3.7. Tip/rol değiştirme

Her bir değişken (attribute) o sütun için mümkün değerleri tanımlayan bir tipe (type) sahiptir. Bu tipler; polynominal, binominal, real, integer, date\_time, date ya da time şeklinde değişken özelliğine göre tanımlanabilmektedir (Şekil 6.5.).

Değişken rolü (role), o değişkenin makine öğrenmesi algoritmaları tarafından nasıl kullanılacağını tanımlar. Rolü olmayan değişkenler (regular attribute) makine öğrenmesi algoritmaları tarafından eğitim sırasında girdi olarak kullanılırken, id attribute (her veri için ona özel bir değere sahip)'ler genellikle görmezden gelir. Label (target/class) attribute ise model sonucunda tahmin edilmek istenen değişkeni belirtmektedir.

Örneğin, Discretize by Binning operatörü; nümerik tipe sahip değişkeni polynominal tipe çevirmek için kullanılır. Bölünmek istenen değişim aralığı (range) sayısı da Parameters sekmesindeki number of bins alanında belirtilebilir. Benzer şekilde Nominal to Numerical operatörü vd. ile değişken tipleri istenen şekilde kolaylıkla revize edilebilmektedir.

Stored at:  
Process: result 1 (result 1)  
Meta data: Data Table  
● Source: //Tutorials99 Data/Titanic

Number of examples = 1309  
12 attributes:  
Note: Some of the nominal values in this set were discarded due to performance reasons. You can change this behaviour in the preferences (zap:dataset.general\_md\_nominal\_values\_limit)  
Generated by: [Retrieve Titanic output](#)

Role	Rol	Name	İsim	Type	Tip	Range	Aralık	Missings	Kayıp	Comment
		Passenger Class		polynomial		= [First, Second, Third]		= 0		<b>Yorum</b>
		Name		polynomial		= [Abbing, Mr. Anthony, Abbott, Master, Eugene J...		= 0		
		Sex		binominal		= [Female, Male]		= 0		
		Age		real		= [0.167 - 80]		= 263		
		No of Siblings or Spouses on Board		integer		= [0 - 8]		= 0		
		No of Parents or Children on Board		integer		= [0 - 9]		= 0		
		Ticket Number		polynomial		= [110152, 110413, 110465, 110469, 110489, 1...		= 0		
		Passenger Fare		numeric		= [0 - 512.329]		= 1		
		Cabin		polynomial		= [A10, A11, A14, A16, A18, A19, A20, A21, A23, ...		= 1014		
		Port of Embarkation		polynomial		= [Cherbourg, Queenstown, Southampton]		= 2		
		Life Boat		polynomial		= [1, 10, 11, 12, 13, 13 15, 13 15 B, 14, 15, 15 16...		= 823		
		Survived		binominal		= [No, Yes]		= 0		

Şekil 6.5. Titanic veri seti

## 6.2. Uygulamada Kullanılan Algoritmalar

RapidMiner, kullanıcıya elindeki veri setini analiz edip, belirlenen amaç doğrultusunda mevcut veri setinde en efektif çalışacak algoritmalarla ilgili bilgi vermektedir. Bu doğrultuda çalışmada kullanılan algoritmalar aşağıda verilmiştir:

- Karar ağaçları (decision trees): Karar ağaçları, ağaç yapısına benzer şekilde olan bir akış şeması biçimindedir. Karar ağacı üzerinde bulunan her bir düğüm; bir nitelik üzerindeki bir test işlemi temsil ederken, her bir dallanma test işleminin sonucunu temsil etmektedir. Ağacın sonlanması ise sınıflar ile olmaktadır (Telcioğlu, 2007). Karar ağacı algoritmalarından bazıları; random forest, gradient boosted trees, rotation forest olarak sayılabilir. Ayrıca ID3, C4.5, CHAID gibi algoritmalar da öne çıkan karar ağacı öğrenme algoritmalarındandır.
- Naive Bayes: Naive Bayes algoritması ile sınıfları daha önceden belirlenmiş mevcut veriler kullanılarak, eldeki verinin mevcut sınıflardan herhangi birine girip girmeme olasılığı hesaplanır.
- Genelleştirilmiş lineer model (Generalized linear model): Genelleştirilmiş lineer model kavramı ilk olarak Nelder ve Wedderburn (1972) tarafından

geliştirilmiştir (Koç ve Cengiz, 2012). Regresyon analizinin en temel hali lineer regresyondur. Lineer regresyon modelinde normal dağılım varsayımı önemli bir rol oynamaktadır. Bağımlı değişkenin sayı gibi kesikli değişken olduğu durumlarda normallik varsayımı sağlanmayabilir. Bir başka durum olarak da bağımlı değişkenin ikili olduğu durumlarda bağımlı değişken sürekli değildir. Ayrıca bağımlı değişkenin sürekli olduğu fakat normal dağılım göstermediği durumlar da olabilir. Bu tür verilerin analizine imkân sağlayacak modeller genelleştirilmiş lineer modellerdir (Koç ve Cengiz, 2012).

- Lojistik regresyon: Doğrusal regresyon değişkenlerinin 0-1 değerlerini almış hali lojistik regresyon olarak isimlendirilmektedir. Genel olarak ikili değişkenler ve nadiren de çok sınıflı değişkenlerin tahmin edilmesi için kullanılmaktadır (Büyükakın, 2005).
- Derin öğrenme (Deep learning): Derin öğrenme, yapay sinir ağlarının geliştirilmesiyle ortaya çıkmıştır. Bir ya da daha fazla gizli katman içeren yapay sinir ağları ve benzeri makine öğrenme algoritmalarını kapsamaktadır.
- K-means: K-means bir kümeleme algoritması olup, Algoritmanın adı, her biri  $c_j$  olmak üzere  $k$  adet kümenin  $c_j$  ile ifade edilen ortalamalarının alınmasından gelir. Aranan küme sayısını ifade eden  $k$  önceden bilinen ve kümeleme işlemi bitene kadar değeri değişmeyen bir sabit olmalıdır (Bilgin ve Çamurcu, 2005).

### 6.3. Model Performansı

Model başarısı değerlendirilirken hata oranı, kesinlik, duyarlılık, F-ölçütü ya da ROC analizi kullanılabilir. Modelin başarı oranı, doğru sınıfa atanan örnek sayısı ve yanlış sınıfa atanan örnek sayısı ile ifade edilir.

Modelin performans testinin, test kümesi adı verilen, öğrenme veri kümesinin dışında bir veri kümesi ile yapılması gerekir. Bu nedenle kullanılan veri kümesi, öğrenme ve test veri kümesi olarak ayrılır. Öğrenme veri kümesi üzerinde model oluşturulduktan sonra oluşan model test veri kümesinde sınanır.

Algoritmalar test edildikten sonra elde edilen sonuçlara göre algoritma performansları karışıklık matrisi (confusion matrix) ile ifade edilebilir. Karışıklık matrisinde satırlar test kümesindeki örneklere ait gerçek sayıları, kolonlar ise tahmin sonuçlarını ifade eder (Çizelge 6.1.).

Çizelge 6.1. İki sınıflı bir veri kümesine ait karışıklık matrisi

	ÖNGÖRÜLEN SINIF		
	Sınıf = 1	Sınıf = 0	
DOĞRU SINIF	Sınıf = 1	a (TP örnek sayısı)	b (FN örnek sayısı)
	Sınıf = 0	c (FP örnek sayısı)	d (TN örnek sayısı)

Sınıf sayısı ikiden fazla olduğunda 2x2 boyutundaki bu matris, n sınıf sayısı olmak üzere, n x n boyutlarında genişletilmiş bir matris şeklini alacaktır. TP ve TN değerleri doğru sınıflandırılmış örnek sayısıdır. FP, aslında 0 (negatif) sınıfındayken 1 (pozitif) olarak tahmin edilmiş örneklerin sayısıdır. FN ise 1 (pozitif) sınıfındayken 0 (negatif) olarak tahmin edilmiş örneklerin sayısını ifade eder. Genel olarak n x n boyutlarındaki bir karışıklık matrisinde ana köşegen doğru tahmin edilen örnek sayılarını; ana köşegen dışında kalan matris elemanları ise hatalı sonuçları ifade etmektedir (Coşkun ve Baykal, 2011).

### 6.3.1. Doğruluk ve hata oranı

Model performansının ölçülmesinde en çok kullanılan, basit ve belirleyici ölçüt, modelin doğruluk değeridir. Doğruluk değeri, doğru sınıflandırılmış örnek sayısının (TP +TN), toplam örnek sayısına (TP+TN+FP+FN) oranı şeklinde ifade edilir (Coşkun ve Baykal, 2011).

Hata oranı ise bu değerlerin birimsel tamlayanıdır. Başka bir deyişle yanlış sınıflandırılmış örnek sayısının (FP+FN), toplam örnek sayısına (TP+TN+FP+FN) oranıdır (Coşkun ve Baykal, 2011).

### 6.3.2. ROC analizi

ROC eğrisi, ikili sınıflandırma sistemlerinde ayırım eşik değerinin farklılık gösterdiği durumlarda, hassasiyetin kesinliğe olan oranıyla ortaya çıkmaktadır. Daha basit anlamda ise doğru pozitiflerin yanlış pozitiflere olan kesri olarak da ifade edilebilir.

Bir ROC eğrisi, farklı eşik değerleri için dikey eksen üzerinde doğru pozitifler (duyarlılık) ve yatay eksen üzerinde yanlış pozitiflerin (özgüllük) oranlarının yer aldığı bir eğridir. ROC eğrisi üzerindeki her nokta, farklı eşik değerlerine karşılık gelen duyarlılık ve özgüllük değerlerini ortaya koyar. Genelde düşük yanlış pozitiflik oranlarını veren eşik değerleri, düşük doğru pozitiflik oranına da sahiptir. Doğru pozitif oranının yüksek, yanlış pozitif oranının düşük olduğu sonuçlar başarılı sonuçlar olarak adlandırılmaktadır.

ROC eğrisi, FP rate (yatay) eksenine yaklaştıkça başarımın seviyesi düşer. Sistemin başarımının tek bir değer ile ifade edilmesi ROC eğrisinin altında kalan alan ile ifade edilmektedir. Bu alan değeri ne kadar büyük ise sistemin güvenilirlik değeri de o kadar yüksek olur (Anonim, 2015).

## 6.4. Verilerin Elde Edilmesi ve Hazırlık Süreci

### 6.4.1. Veri setleri

ESOGÜ Enformatik Ders Yönetim Sisteminden alınan toplam veri seti sayısı 200'ün üzerinde olup bu veri setlerinden bir kısmı veri içermemektedir. Bu nedenle gereksiz veri setleri ayrılarak analiz için içlerinden aşağıda verilen 44 tanesi kullanılmıştır:

- mdl\_assign: 2.897 kayıttan oluşmaktadır.
- mdl\_assign\_grades: 34.562 kayıttan oluşmaktadır.
- mdl\_assign\_submission: 97.315 kayıttan oluşmaktadır.
- mdl\_assign\_user\_flags: 34.514 kayıttan oluşmaktadır.
- mdl\_assign\_user\_mapping: 128.020 kayıttan oluşmaktadır.
- mdl\_assignsubmission\_file: 60.147 kayıttan oluşmaktadır.
- mdl\_comments: 559 kayıttan oluşmaktadır.

- mdl\_course: 18.756 kayıttan oluşmaktadır.
- mdl\_course\_categories: 1180 kayıttan oluşmaktadır.
- mdl\_course\_modules: 28.721 kayıttan oluşmaktadır.
- mdl\_event: 3148 kayıttan oluşmaktadır.
- mdl\_files: 314.280 kayıttan oluşmaktadır.
- mdl\_folder: 863 kayıttan oluşmaktadır.
- mdl\_forum: 9.432 kayıttan oluşmaktadır.
- mdl\_forum\_discussions: 4.687 kayıttan oluşmaktadır.
- mdl\_forum\_posts: 4.936 kayıttan oluşmaktadır.
- mdl\_forum\_subscriptions: 263 kayıttan oluşmaktadır.
- mdl\_grade\_categories\_history: 13.390 kayıttan oluşmaktadır.
- mdl\_grade\_grades: 250.453 kayıttan oluşmaktadır.
- mdl\_grade\_grades\_history: 586.032 kayıttan oluşmaktadır.
- mdl\_grade\_items: 9.653 kayıttan oluşmaktadır.
- mdl\_grade\_items\_history: 1.310.996 kayıttan oluşmaktadır.
- mdl\_groups: 57 kayıttan oluşmaktadır.
- mdl\_groups\_members: 292 kayıttan oluşmaktadır.
- mdl\_label: 742 kayıttan oluşmaktadır.
- mdl\_message: 12.743 kayıttan oluşmaktadır.
- mdl\_message\_read: 67.586 kayıttan oluşmaktadır.
- mdl\_quiz: 492 kayıttan oluşmaktadır.
- mdl\_quiz\_attempts: 44.760 kayıttan oluşmaktadır.
- mdl\_quiz\_feedback: 501 kayıttan oluşmaktadır.
- mdl\_quiz\_grades: 41.666 kayıttan oluşmaktadır.
- mdl\_resource: 13.658 kayıttan oluşmaktadır.
- mdl\_role: 9 kayıttan oluşmaktadır.
- mdl\_role\_assignments: 513.030 kayıttan oluşmaktadır.
- mdl\_stats\_daily: 4.866.396 kayıttan oluşmaktadır.
- mdl\_stats\_monthly: 634.788 kayıttan oluşmaktadır.
- mdl\_stats\_user\_daily: 595.949 kayıttan oluşmaktadır.
- mdl\_stats\_user\_monthly: 570.921 kayıttan oluşmaktadır.
- mdl\_stats\_user\_weekly: 624.360 kayıttan oluşmaktadır.

- mdl\_stats\_weekly: 1.810.748 kayıttan oluşmaktadır.
- mdl\_user: 65.596 kayıttan oluşmaktadır.
- mdl\_user\_enrolments: 576.453 kayıttan oluşmaktadır.
- mdl\_user\_info\_data: 65.547 kayıttan oluşmaktadır.
- mdl\_user\_lastaccess: 161.818 kayıttan oluşmaktadır.

Veri setleri; 2015-2016 Güz, 2015-2016 Bahar ve 2016-2017 Güz yarıyıllarını kapsamaktadır. Ders Yönetim Sistemi kayıtlarında olan Eskişehir Osmangazi Üniversitesinin mevcut öğrenci kümesinin toplam kayıt sayısı 61.232 olup, bunlardan 18.596'sı sistemi aktif olarak kullanmaktadır.

Şekil 6.6.'da ESOĞÜ Enformatik Ders Yönetim Sisteminde en fazla kullanıcı hareketliliğine sahip ilk 25 bölüm gösterilmektedir.

Index	İndeks	Nominal value	Sembolik değer	Absolute count	Mutlak değer	Fraction	Fraksiyon
1		İnşaat Mühendisli	değer	1478	değer	0.085	
2		Elektrik Elektronik Mühendisliği		1103		0.063	
3		Makine Mühendisliği		1091		0.062	
4		İktisat		1069		0.061	
5		Endüstri Mühendisliği		1042		0.060	
6		İşletme		971		0.056	
7		Maliye		908		0.052	
8		Rehberlik ve Psikolojik Danışm...		675		0.039	
9		Matematik-Bilgisayar		601		0.034	
10		Maden Mühendisliği		533		0.031	
11		Zihin Engelliler Öğretmenliği		459		0.026	
12		Kimya Mühendisliği		427		0.024	
13		Bilgisayar Mühendisliği		408		0.023	
14		İstatistik		407		0.023	
15		Sınıf Öğretmenliği		373		0.021	
16		Türk Dili ve Edebiyatı		371		0.021	
17		Metalurji ve Malzeme Mühendis...		361		0.021	
18		Karşılaştırmalı Edebiyat		353		0.020	
19		İlköğretim Fen Bilgisi Öğretme...		334		0.019	
20		Uluslararası İlişkiler		333		0.019	
21		İlköğretim Matematik Öğretmen...		329		0.019	
22		Siyaset Bilimi ve Kamu Yönetimi		291		0.017	
23		Jeoloji Mühendisliği		272		0.016	
24		Bilgisayar ve Öğretim Teknoloji...		240		0.014	
25		İngilizce Öğretmenliği		224		0.013	

Şekil 6.6. Sistemde en fazla kullanıcı hareketine sahip ilk 25 bölüm

#### **6.4.1.1. Assign modülü**

Assign, ders içeriğine karşılığında notlandırma yapılan bir assignment (ödev, uygulama vb.) eklemeye yarayan modüldür. Yaratılan her bir assignment için tutulan veriler, mdl\_assign veri setinde yer almaktadır. Öğrenci, assignment yüklemesini gerçekleştirdiği an ise mdl\_assign\_submission veri setinde yeni bir kayıt açılmaktadır. İlgili assignment üzerinden kullanıcıya verilen not (grade) kayıtlarının tutulduğu veri seti ise mdl\_assign\_grades'dir.

#### **6.4.1.2. Forum modülü**

Forum, ders içeriğine bir forum girdisi eklemeye yarayan modüldür. Genellikle duyuru, bilgilendirme vb. ihtiyaçlar için kullanılmaktadır. Yaratılan her bir forum için tutulan veriler, mdl\_forum veri setinde yer almaktadır. Kullanıcıların forum girdileri üzerinden yaptıkları yorum, tartışma ya da geribildirimlerin kayıtlarının tutulduğu veri seti ise mdl\_forum\_discussions'dır. Yaratılan forumdaki mesaj, duyuru vb. içeriklerin kaydı ise mdl\_forum\_posts veri setinde tutulmaktadır.

#### **6.4.1.3. Message modülü**

Message, bir kullanıcının diğer kullanıcılara mesaj göndermek istediğinde kullandığı modüldür. Herhangi bir kullanıcı tarafından gönderilen ve henüz gönderilen kişi tarafından okunmamış mesajların kaydının tutulduğu veri seti mdl\_message'dır. Gönderilen kişi tarafından okunmuş mesajların kaydının tutulduğu veri seti ise mdl\_message\_read'dir.

#### **6.4.1.4. Quiz modülü**

Quiz, ders içeriğine sistem üzerinden gerçekleştirilecek olan bir mini sınav eklemeye yarayan modüldür. Yaratılan her bir quiz kaydı mdl\_quiz veri setinde, her bir kullanıcının quiz katılım kaydı mdl\_quiz\_attempts veri setinde, her bir kullanıcı için quiz notu sisteme girildiğinde oluşan kayıtlar mdl\_quiz\_grades veri setinde yer almaktadır.

#### **6.4.1.5. Resource modülü**

Resource, öğrenmeyi desteklemek amacıyla ders yürütücüsü tarafından, ders içeriğine bir kaynak (kitap, dosya, klasör, url vb.) eklenmesini sağlayan modüldür. Eklenen her bir resource kaydı mdl\_resource veri setinde saklanmaktadır.

#### **6.4.1.6. Folder modülü**

Folder, ders içeriğine eklenecek dosyaları organize etmek ya da farklı uzantılara sahip birden çok kaynağı bir arada yönetebilmek amacıyla kullanılan modüldür. Eklenen her bir folder kaydı mdl\_folder veri setinde saklanmaktadır.

#### **6.4.1.7. Files modülü**

Files, ders içeriğine tekil bir dosya eklemek amacıyla kullanılan modüldür. Eklenen her bir files kaydı mdl\_files veri setinde saklanmaktadır.

#### **6.4.1.8. Stats modülü**

Stats, belirli bir zaman zarfında; roller, kullanıcılar ve dersler bazında gerçekleştirilen eylem istatistiklerinin kayıt altına alınmasını yöneten modüldür. İlgili veri setlerinden derse kayıtlı öğrenci sayısı, ders içeriğinde yapılan yazma (statswrites) ya da okuma (statsreads) aktivitelerinin istatistikleri gibi veriler elde edilmiştir. Buradaki okumadan kasıt, kullanıcının içeriklere erişim yapması, okuması ya da download etmesi yani ders içeriğinde herhangi bir değişiklik yapmamasıdır. Yazmadan kasıt ise ders içeriğine bir ekleme yapmak, revize etmek ya da öğrencilerin quiz ya da assignment teslim durumlarıdır.

#### **6.4.2. Hazırlık süreci**

Veri madenciliği teknikleri uygulanmadan önce, verileri analize hazır hale getirmek için bir hazırlık süreci söz konusudur. Büyük veriyi veri tabanından çekildiği haliyle kullanmak imkânsızdır.

Bu noktada en önemli adım veri temizleme aşamasıdır. Bu aşamada, gereksiz, tutarsız vb. veriler silinmiştir. Bu verilere; sistemi test amaçlı açılan ders kayıtları, sisteme hiçbir zaman giriş yapmamış (firstaccess = 0) kullanıcı kayıtları vb. örnek verilebilir.

Veri dönüştürme aşamasında kullanılan metot Z-transformation olup, değişken değerleri belirli aralıklara dönüştürülmüştür.

Son olarak outlier (aykırı) veriler, RapidMiner'da Detect Outlier operatörü kullanılarak belirlenmiş ve filtrelenerek analize katılmamaları garanti altına alınmıştır.

### 6.4.3. Model

Öğrencinin dersteki başarısının tahminine yönelik modeller RapidMiner veri bilimi yazılımı üzerinde oluşturulmuştur. Veri setinden, ders yönetim sistemi aktif olarak kullanılan farklı dersler alınarak modeller ders bazında ele alınmıştır.

Dersler için ele alınan değişkenler farklılık gösterebilmektedir. Bunun nedeni ise bir dersteki içeriklerin diğer dersten farklı olmasıdır. A dersinde hiç assignment kaydı yokken, B dersinde olması ya da A dersinde birden fazla quiz yapılırken, B dersinde sadece tek bir quiz yapılmış olması bu duruma örnek olarak verilebilir. Nitekim ne tür bir model kurulursa kurulsun, kullanılan değişkenlerin amaca yönelik kaliteli veriler içermesi oldukça önemlidir. Modelde yer alan değişkenlerin seçimi oldukça kritiktir. RapidMiner bu konuda, değişken değerlerinin dikkat edilmesi gereken özelliklerini kullanıcıya ön bilgilendirme olarak sağlayan özelliğiyle karar verme sürecine yardımcı olmaktadır. İstenildiğinde bu özellik kullanılarak; kırmızı, sarı ve yeşil baloncuklarla değişken değerlerinin kalitesi görselleştirilmektedir. Burada; kırmızı renk, zayıf kalitede ve çoğu durumda kullanılmaması gereken değişkeni, sarı renk, label değişken üzerinde çok düşük (%0,01'den düşük, muhtemelen tahmine fayda sağlamayacak) ya da tam tersi çok yüksek (%50 üzeri, label değişken üzerinde direk bir neden-sonuç ilişkisine sahip) korelasyona sahip değişkeni, yeşil ise bir sorun gözükmeyen değişkeni ifade etmektedir. Yine de tüm bu değişkenlerin kullanılıp kullanılmama durumu karar vericinin inisiyatifindedir.

Modelde kullanılan doğrulama metodu 10-kat apraz doğrulama (10-fold cross validation) yntemidir. Model performanslarının deęerlendirilmesinde doęruluk (accuracy) deęeri kullanılmıřtır.



## 7. BULGULAR VE TARTIŞMA

### 7.1. İncelenen Dersler

#### 7.1.1. Doğrusal sistemler

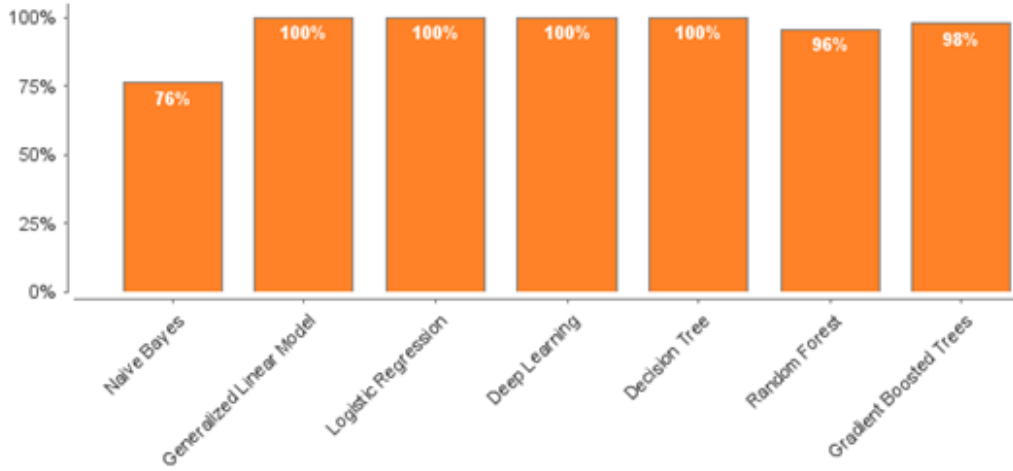
Bu kısımda, 2015-2016 Bahar Döneminde açılan 9450 courseid'li, Endüstri Mühendisliği Bölümü ikinci sınıfa ait olan Doğrusal Sistemler dersinin analizi gerçekleştirilmiştir. Bu ders, ESOGÜ Endüstri Mühendisliği Bölümünün, ders yönetim sistemi kullanımında en aktif derslerinden biridir. Ders içeriğinde yapılan quiz ortalamalarından hareketle öğrencilerin başarı durumlarının farklı algoritmalarla tahmin performansları incelenmiştir.

Derse ait veri setinde yer alan değişkenler aşağıdaki gibidir:

- userid (id değişken)
- statsreads (öğrencinin sistemde gerçekleştirdiği okuma sayısı)
- statswrites (öğrencinin sistemde gerçekleştirdiği yazma sayısı)
- derece (öğrencilerin, öğrenci kimlik numaralarındaki hane ayrımından hareketle üniversiteye yerleştirme sınavından aldıkları sonuçlara göre bölüme yerleştirilme dereceleri)
- files (öğrencinin sisteme yüklediği dosya sayısı)
- gelenmesajokumaorani (öğrencinin kendisine gönderilen mesajları okuma oranı)
- eposta (öğrencinin sisteme e-posta adresini girmesi durumunda var, diğer durumda yok)
- girisyili (yine öğrenci kimlik numarası verilerinden hareketle elde edilen, öğrencinin üniversiteye giriş yılı)
- grade (first, second, vb. şeklinde sayısı derse göre değişebilen, öğrencilerin quiz notları)
- ortalama (başarı ortalamasını gösteren label değişken)

Label deęişken olan ortalama deęişkeni, öğrencinin notlarından hareketle ortalama üstü ve ortalama altı olmak üzere iki sınıftan oluşmaktadır. Analizdeki amaç bu sınıfların önceden tahmin edilmesinin mümkün olup olmamasıdır. Algoritmaların tahmin performansları Şekil 7.1.'de verilmiştir.

### Accuracy Doğruluk



Model	Accuracy
Naive Bayes	76.1%
Generalized Linear Model <b>Genelleştirilmiş Lineer Model</b>	100.0%
Logistic Regression <b>Lojistik Regresyon</b>	100.0%
Deep Learning <b>Derin Öğrenme</b>	100.0%
Decision Tree <b>Karar Ağacı</b>	100.0%
Random Forest	95.7%
Gradient Boosted Trees	97.8%

Şekil 7.1. Sınıflandırmada kullanılan algoritma performansları (doęruluk)

Öğrencinin ilk quiz notu başarı üzerinde büyük oranda belirleyici olmuştur. Ondan sonra sırasıyla derece ve sistemde gerçekleştirdiği yazma işlemi sayısı (statswrites) gelmektedir. Sisteme e-posta bildirimini neredeyse derse kayıtlı tüm öğrenciler tarafından yapıldığı için bu deęişkenin sonuca bir etkisi olmamıştır. Aynı şekilde, öğrencilerin gelen mesajı okuma oranları da oldukça benzer deęerlere sahip olduğundan deęişken etkisi önemsiz kalmıştır.

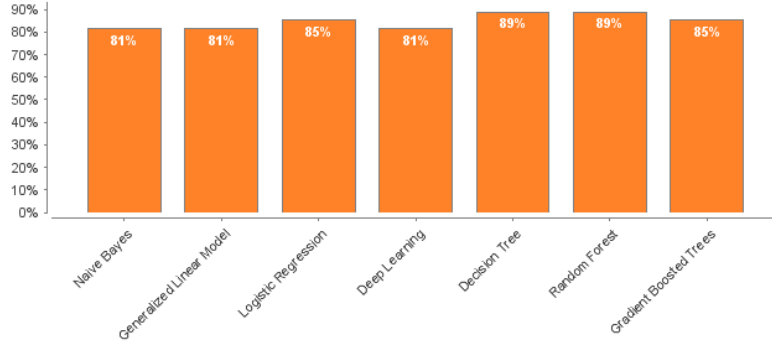
Dersi alan öğrencilerin çoğunluğu 2014 girişlidir. Giriş yılı geriledikçe, öğrenci başarı ortalamasının düştüğü gözlenmiştir. Dersi üstten alan sadece 2 öğrenci olduğundan onlarla ilgili anlamlı bir çıkarım yapılamamaktadır.

### 7.1.2. Diferansiyel denklemler

Sıradaki uygulamada ise İnşaat Mühendisliği Bölümünde ikinci sınıfa ait olan Diferansiyel Denklemler dersinin önce 3513 courseid'li 2015-2016 Güz Döneminde açılan şubesi, daha sonra da 16104 courseid'li 2016-2017 Güz Döneminde açılan şubesi analiz edilmiştir. Bu ders, ders yönetim sistemindeki tüm dersler arasında sistem üzerinde en fazla quiz gerçekleştirilen derstir.

3513 courseid'li ders için iki farklı analiz yapılmıştır. İlk durumda not ortalamalarından hareketle ortalama altı ve ortalama üstü olarak sınıflandırılan başarı durumu bu şekilde belirli sınıflandırma algoritmalarına tabi tutulup algoritmaların tahmin performansları karşılaştırılmıştır. İkinci durumda ise, iki aşamalı bir yöntem izlenmiş olup üstteki gibi sınıflandırılmış veri setine önce k-means kümeleme algoritması uygulanıp, yanlış sınıfa atanan kayıtlar veri setinden temizlenmiştir. Diğer aşamadaysa bu yeni veri setine yine belirli sınıflandırma algoritmaları uygulanarak öğrenci başarısını tahmin performansları değerlendirilmiştir. İlgili RapidMiner çıktıları Şekil 7.2. ve Şekil 7.3. aşağıda verilmiştir.

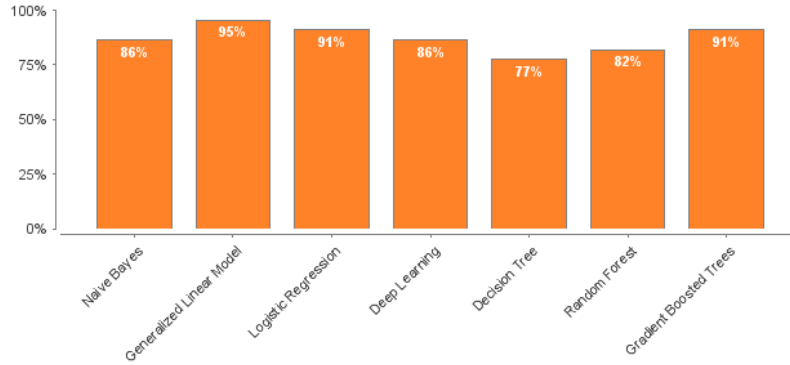
### Accuracy



Model	Accuracy
Naive Bayes	81.5%
Generalized Linear Model	81.5%
Logistic Regression	85.2%
Deep Learning	81.5%
Decision Tree	88.9%
Random Forest	88.9%
Gradient Boosted Trees	85.2%

Şekil 7.2. İlk durumda algoritmaların performansı (doğruluk)

### Accuracy



Model	Accuracy
Naive Bayes	86.4%
Generalized Linear Model	95.5%
Logistic Regression	90.9%
Deep Learning	86.4%
Decision Tree	77.3%
Random Forest	81.8%
Gradient Boosted Trees	90.9%

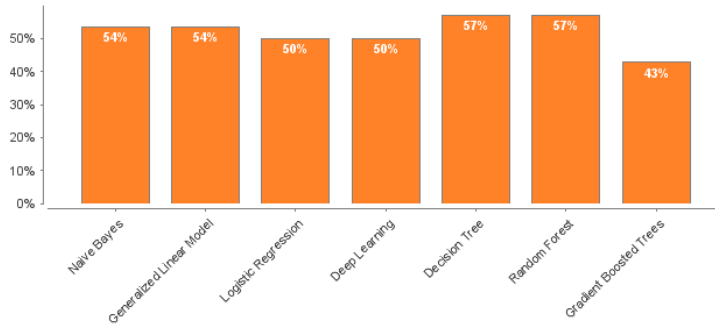
Şekil 7.3. İkinci durumda algoritmaların performansı (doğruluk)

İkinci durumda, veri setine K-means kümeleme algoritması uygulandıktan sonra yanlış sınıflandırılan 25 kayıt silinmiştir. Şekil 6.3. ve Şekil 6.4.'ten görüleceği üzere bu adım bazı algoritmaların tahmin performansını yükseltirken bazılarınınkini de düşürmüştür. Yine de GLM algoritması baz alınacak olursa bu adımın tahmin kalitesini artırdığı söylenebilir.

Analiz sonucunda bu derse kayıtlı öğrenci başarılarının; yapılan quizlere katılım oranları, sistemde gerçekleştirdikleri okuma sayıları (bir diğer deyişle ders yönetim sisteminde aktif olma sayısı), kendilerine gelen mesajı okuma oranları ve sisteme e-postalarını eklemeleri durumlarıyla pozitif yönde ilişkili olduğu görülmüştür.

Sıradaki uygulamada ise 16104 courseid'li ders analiz edilerek, aynı dersi farklı yıllarda alan öğrenci başarılarını etkileyen değişkenlerin aynı olup olmadığı görülmek istenmiştir. Dersin bir önceki yılı için uygulanan analiz adımları aynı şekilde uygulanmıştır. Bu defa K-means algoritması sonrasında yanlış sınıflara atanan 22 kayıt silindikten sonra tekrar sınıflandırma modelleri çalıştırılmıştır. İlgili RapidMiner çıktıları Şekil 7.4. ve Şekil 7.5. aşağıda verilmiştir.

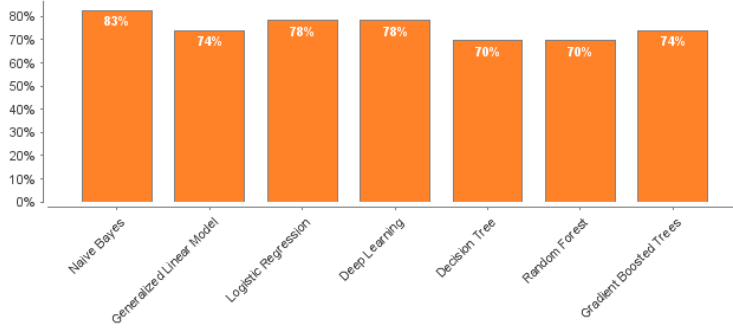
#### Accuracy



Model	Accuracy
Naive Bayes	53.6%
Generalized Linear Model	53.6%
Logistic Regression	50.0%
Deep Learning	50.0%
Decision Tree	57.1%
Random Forest	57.1%
Gradient Boosted Trees	42.9%

Şekil 7.4. K-means algoritması uygulanmamış halde performanslar

### Accuracy



Model	Accuracy
Naive Bayes	82.6%
Generalized Linear Model	73.9%
Logistic Regression	78.3%
Deep Learning	78.3%
Decision Tree	69.6%
Random Forest	69.6%
Gradient Boosted Trees	73.9%

Şekil 7.5. K-means algoritması uygulandıktan sonra performanslar

Şekillerden görüleceği üzere aynı dersin bir sonraki yılına ait tahmin sonuçları önceki yıla nispeten daha düşüktür. Değişken önemleri analiz edildiğinde ise; yapılan quizlere katılım oranları, sistemde gerçekleştirdikleri okuma sayıları, sisteme yükledikleri dosya sayıları ve derece değişkenlerinin sonuç üzerinde en önemli değişkenler olduğu görülmüştür.

Dersin önceki yılından farklı olarak, öğrencinin sisteme e-postasını kaydetmiş olup olmaması bir sonraki yılda öğrenci başarısı üzerinde kayda değer önem teşkil etmemektedir. Yine gelen mesajları okuma oranları da bu defa önemsiz bir değişken olarak kalmıştır. Bunun nedeni ise bu değişkenler için öğrencilerin tamamının neredeyse özdeş değerlere sahip olmasıdır.

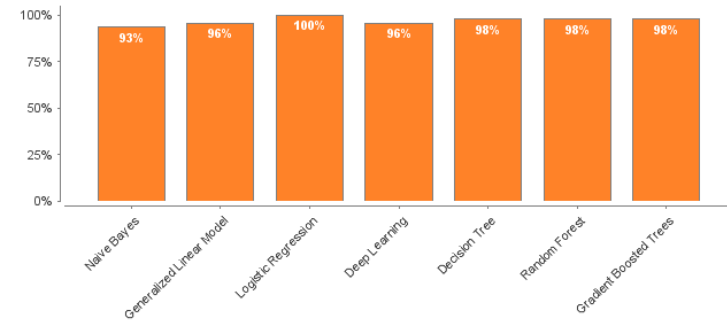
### 7.1.3. İşbilim

Bu bölümde iki farklı öğretim görevlisi tarafından ortak yürütülen, Endüstri Mühendisliği Bölümü 2015-2016 Bahar Dönemine ait İşbilim dersi analiz edilmiştir. İkinci ve normal öğretime ait ikişer şube bulunmaktadır, toplamda dört şube birleştirilerek, toplu bir veri setiyle modellenmiştir.

Önceki analizlerden farklı olarak bu defa tip (ikinci ya da normal öğretim öğrencisi) ve ders yürütücüsü değişkenleri analize dâhil edilmiştir. Bu dersi alan öğrenci başarısında ders yürütücüsünün ya da öğrencinin öğretim tipinin öneminin olup olmadığı gözlenmek istenmiştir.

Algoritmaların sınıflandırma başarıları oldukça yüksektir (Şekil 7.6.).

#### Accuracy



Model	Accuracy
Naive Bayes	93.3%
Generalized Linear Model	95.6%
Logistic Regression	100.0%
Deep Learning	95.6%
Decision Tree	97.8%
Random Forest	97.8%
Gradient Boosted Trees	97.8%

Şekil 7.6. Algoritma performansları

Analiz sonucunda bu derse kayıtlı öğrenci başarıları ile ders yürütücüsü ya da öğretim tipinin arasında kayda değer bir ilişki saptanmamıştır. Bir öncekine kıyasla, son quiz notunun önemi daha yüksek çıkmıştır. Bunun dışında öğrencinin sistemde gerçekleştirdiği okuma ve yazma sayıları da başarıya pozitif yönde etki etmektedir. Son olarak öğrencinin bölüme yerleştirilme derecesi arttıkça başarı durumunun da yükseldiği gözlenmiştir.

## 8. SONUÇ VE ÖNERİLER

21. yüzyılda eğitim; herkes için karşılanması gereken bir ihtiyaç halini almıştır. Eğitim ihtiyacı doğan grup giderek artarken, var olan kaynaklar (öğretim görevlileri, derslik, teçhizat vb.) kısıtlıdır. Kaliteli bir eğitim sürecinin temini için var olan kaynakların etkin kullanımı sağlanmalıdır. Hal böyle iken 21. yüzyıl üniversitelerinde DYS kullanımı kaçınılmazdır.

Eğitim-öğretimde ders yönetim sistemi yazılımlarından faydalanan yükseköğretim kurumlarında kullanıcılara ait günden güne artan veriler bulunmaktadır. Bu veriler analiz edilerek fayda sağlanmaya çalışılmadığında, büyük verinin veri tabanlarında saklanması hiçbir anlamı yoktur. Büyük veride anlamlı örüntüler yakalamaya çalışan veri madenciliği yöntemleri ile öngörülen/öngörülemeyen durumlar ortaya çıkarılmaktadır.

Bu çalışmada ESOGÜ Enformatik Bölümü Ders Yönetim Sisteminden elde edilen gerçek kullanıcı verileri RapidMiner veri madenciliği yazılımı ile analiz edilmiştir. Öğrencilerin ders bazında başarı tahmini yapılmış ve algoritma performansları karşılaştırılmıştır. Ders Yönetim Sisteminden elde edilen değişkenlerin farklı veri setlerinde kullanıldığında sınıflandırma üzerindeki etkileri tartışılmıştır.

Çalışma sonunda; öğrenci başarısına etki eden faktörlerin farklı kümelerdeki (farklı dersler) öğrenciler için değişiklik gösterdiği anlaşılrsa da genele bakıldığında; sistemde daha aktif olan öğrencilerin daha başarılı oldukları söylenebilmektedir. Kullanılan algoritmaların performansları da her bir veri seti için farklılık göstermiştir, fakat karar ağacı algoritmalarının tüm veri setlerinde yüksek performans gösterdiği saptanmıştır. Sınıflandırma öncesinde kümeleme (k-means) algoritmasının kullanılması ile tahmin performansı %86 oranında artmıştır. Diferansiyel Denklemler dersini 2015-2016 yılında alan öğrencilerin gelen mesajları okuma oranlarının ortalaması %87 iken, 2016-2017 yılında alan öğrencilerin %98'dir. Bu kümeden hareketle öğrencilerin DYS'ye ilgilerinin arttığı söylenebilmektedir. Bu örnek özelinden yola çıkarak; ders yürütücülerine öğrenciler ve

derslerle ilgili sađlanan bu gibi geribildirimler DYS kullanıcıların farkındalıđının artmasını sađlayarak DYS'ye ilgilerini sürekli kılacaktır.

DYS kullanıcılarının sınıf dıřı aktivitelerinin tamamı DYS veri tabanında toplanmaktadır. Veri madenciliđi ile DYS kullanıcılarının sistemi kullanma alışkanlıklarından hareketle, DYS kullanıcılarına bilgi geribildirimi sađlanabilmektedir. Bu şekilde DYS kullanıcılarının olađan dıřı sapmalara karřı da önceden uyarılabilmeleri mümkün gözükmemektedir. Böylece farkındalıđı artan DYS kullanıcılarının sürekli iyileřtirme bilinci ile eđitim kalitesinin artırılması sađlanabilecektir.

Gelecekte, ders yönetim sistemlerinden akan gerçek zamanlı verilerin anlık analizlerini gerçekteřtirip; öğrenci, öğretim görevlisi ve yöneticilere düzenli geribildirimlerde bulunan uyarı/bilgilendirme sistemleri yaratmak mümkündür.

## KAYNAKLAR DİZİNİ

- Adanır, G. A., 2016, Virtual math teams (VMT) çevrimiçi ortamındaki öğrenme sürecinin öğrenme analitiklerinin kullanılarak değerlendirilmesi, Doktora tezi, Orta Doğu Teknik Üniversitesi Enformatik Enstitüsü, 236 s.
- Akçapınar, G., 2014, Çevrimiçi öğrenme ortamındaki etkileşim verilerine göre öğrencilerin akademik performanslarının veri madenciliği yaklaşımı ile modellenmesi, Doktora tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, 96 s.
- Alan, M. A., 2012, Veri madenciliği ve lisansüstü öğrenci verileri üzerine bir uygulama, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 33, 165-174.
- Alan, M. A., 2014, Karar ağaçlarıyla öğrenci verilerinin sınıflandırılması, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 28, 4, 101-112.
- Alpaydın, E., 2000, Zeki veri madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri. Bilişim 2000 Eğitim Semineri, [https://www.cmpe.boun.edu.tr/~ethem/files/papers/veri-maden\\_2k-notlar.doc](https://www.cmpe.boun.edu.tr/~ethem/files/papers/veri-maden_2k-notlar.doc), erişim tarihi: 10.04.2018.
- Anonim, 2015, ROC (Receiver Operating Characteristic) Analizi, <http://portal.netcad.com.tr/display/HELP/ROC+Analizi>, erişim tarihi: 10.04.2018.
- Anonim, 2016, Genel İstatistik, <https://www.onlineistatistik.com/single-post/2016/12/01/istatistik-nedir-temel-kavramlari-nelerdir>, erişim tarihi: 10.04.2018.
- Anonim, 2017, <http://yazilimagiris.com/2017/11/page/2/?print=print-search>, erişim tarihi: 10.04.2018.
- Ayesha S., Tasleem M., Sattar A.R., Khan M.I., 2010, Data mining model for higher education system, European Journal of Scientific Research, 43, 1, 24-29.
- Bahçeci, F., 2015, Öğrenme Yönetim Sistemlerinde Kullanılan Öğrenme Analitikleri Araçlarının İncelenmesi, Turkish Journal of Educational Studies, 2, 1, 43-44, <http://dergipark.gov.tr/download/article-file/403131>, erişim tarihi: 10.04.2018.
- Belsis, P., Chalaris, I., Chalaris, M., Skourlas, C., Tsolakidis, A., 2014, The Analysis of the Length of Studies in Higher Education based on Clustering and the Extraction of Association Rules, Procedia - Social and Behavioral Sciences, 147, 567-575.
- Ben-Zadok G., Hershkovitz, A., Mintz, R., Nachmias, R., 2007, Examining online learning processes based on log files analysis: a case study, [https://www.researchgate.net/publication/242472530\\_Examining\\_online\\_learning\\_processes\\_based\\_on\\_log\\_files\\_analysis\\_A\\_ca\\_se\\_study](https://www.researchgate.net/publication/242472530_Examining_online_learning_processes_based_on_log_files_analysis_A_ca_se_study), erişim tarihi: 10.04.2018.

### KAYNAKLAR DİZİNİ (devam)

- Berry, M. J. A., Linoff, G. S., 2004, Data Mining Techniques for Marketing, Sales, and Customer Relationship Management, Wiley Publishing Inc., 5-16.
- Bhan, N., Tech, M., Mehrotra, D., 2013, Comparative Study Of Em And k-means Clustering Techniques In WEKA Interface, International Journal of Advanced Technology & Engineering Research, 3, 4, 40-44.
- Bharadwaj, B.K., Pal, S., 2011, Mining Educational Data to Analyze Students' Performance, International Journal of Advance Computer Science and Applications (IJACSA), 2, 6, 63-69.
- Bozkır, A.S., Mazman, S.G., Sezer, E.A., 2010, Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case, IMCW2010, CCIS (Communications in computer and information science), 96, 145-153.
- Bilgin, T. T., Çamurcu, Y., 2005, DBSCAN, OPTICS ve K-Means Kümeleme Algoritmalarının Uygulamalı Karşılaştırılması, Politeknik Dergisi, 8, 2, 139-145, <http://dergipark.gov.tr/download/article-file/384439>, erişim tarihi: 10.04.2018.
- Büyükakın, A., 2005, Bulanık Mantık ile Veri Madenciliği, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, 86 s.
- Campagni, R., Merlini, D., Sprugnoli, R., Verri, M. C., 2015, Data mining models for student careers, Expert Systems with Applications, 42, 5508–5521.
- Coşkun, C., Baykal, A., 2011, Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması, Akademik Bilişim'11 - XIII. Akademik Bilişim Konferansı Bildirileri, Akademik Bilişim, 51-58.
- Çekiç, U., 2010, Uzaktan Eğitim Sistemi Tasarımı, Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimler Enstitüsü, 113 s.
- Çoban, S., 2016, Üniversitelerde Öğretim Yönetim Sistemleri Yazılımları Kullanımına Yönelik Bir İnceleme, Batman Üniversitesi Yaşam Bilimleri Dergisi, 6, 1, 2-3, <http://dergipark.gov.tr/download/article-file/313306>, erişim tarihi: 10.04.2018.
- Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., Schroeder, U., 2012, Design and Implementation of a Learning Analytics Toolkit for Teachers, Educational Technology & Society, 15, 3, 58–76, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.298.9599&rep=rep1&type=pdf>, erişim tarihi: 10.04.2018.

### KAYNAKLAR DİZİNİ (devam)

- Ekim, U., 2011, Veri madenciliği algoritmalarını kullanarak öğrenci verilerinden birliktelik kurallarının çıkarılması, Yüksek lisans tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 61 s.
- Elias, T., 2011, Learning Analytics: Definitions, Processes and Potential, <https://www.slideshare.net/bordi/learning-analytics-definitions-processes-potential>, erişim tarihi: 10.04.2018.
- Erdoğan Ş., Timor M., 2005, A Data Mining Application in a Student Database, Havacılık ve Uzay Dergisi, 2, 2, 57-64, <https://pdfs.semanticscholar.org/04ae/b32415878a4e9292682447dd3e239874f943.pdf>, erişim tarihi: 10.04.2018.
- Eskişehir Osmangazi Üniversitesi Enformatik Bölümü, 2017, <https://enformatik.ogu.edu.tr/Sayfa/Index/10/bolum-hakkinda>, erişim tarihi: 10.04.2018.
- Fayyad, U.M, Piatesky-Shapiro, G., Smyth P., 1996, From Data Mining to Knowledge Discovery: An Overview Advances in Knowledge Discovery and Data Mining, AAAI Press/The MIT Press, 1-34.
- Gorunescu, F., 2011, Data Mining Concepts, Models, Methods and Algorithms, Springer, 185-317.
- Gülçe, G., 2010, Veri Ambarı ve Veri Madenciliği Teknikleri Kullanılarak Öğrenci Karar Destek Sistemi Oluşturma, Yüksek lisans tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, 109 s.
- Hamsa, H., Indiradevi, S., Kizhakkethottam, J. J., 2016, Student academic performance prediction model using decision tree and fuzzy genetic algorithm, Procedia Technology, 25, 326-332.
- Hassana, S. M., Al-Razgan, M. S., 2016, Pre-University Exams Effect on Students GPA: A case Study in IT Department, Procedia Computer Science, 82, 127-131.
- Han, J., Kamber, M., 2006, Data Mining: Concepts and Techniques, Morgan Kaufmann, 1-39.
- Hand, J. D., 1998, Data Mining: Statistics and More ?, The American Statistician, 18-50.
- Kapanoğlu, M., Abdalla, S., Gültekin, Ö., Er, F., Sönmez, H., 2016, İşletme Analitiği, Anadolu Üniversitesi Yayınları, 3-20.
- Karabatak, M., İnce, M. C., 2004, Apriori Algoritması ile Öğrenci Başarısı Analizi, ELECO, 348-352.

**KAYNAKLAR DİZİNİ (devam)**

- Kayabaş, İ., 2017, Açık ve uzaktan öğrenmede öğrenme analitikleri kontrol panelinin geliştirilmesi: Öğrenme Bulutu, Doktora tezi, Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, 172 s.
- Kılınç, Ç., 2015, Üniversite öğrenci başarısı üzerine etki eden faktörlerin veri madenciliği yöntemleri ile incelenmesi, Yüksek lisans tezi, Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, 79 s.
- Kınay, E., 2016, Yaşar Üniversitesi uzaktan eğitim sistemi analizi, Yüksek lisans tezi, Yaşar Üniversitesi Fen Bilimleri Enstitüsü, 55 s.
- Koç, T., Cengiz, M. A., 2012, Genelleştirilmiş Lineer Karma Modellerde Tahmin Yöntemlerinin Uygulamalı Karşılaştırılması, Karaelmas Fen ve Mühendislik Dergisi, 2, 2, 47-52, <http://dergipark.gov.tr/download/article-file/304587>, erişim tarihi: 10.04.2018.
- Kudyba, S., 2014, Big Data, Mining, and, Analytics, CRC Press, 1-15.
- Mamcenko, J., Sileikiene, I., Lieponiene, J., Kulvietiene, R., 2011, Analysis of E-Exam Data Using Data Mining Techniques, Proc of 17th International Conference on Information and Software Technologies (IT 2011), 215–219.
- Minaei-Bidgoli, B., Kashy, D. A., Kortmeyer, G., Punch, W. F., 2003, Predicting student performance: An application of data mining methods with an educational web-based system, In the proceedings of the 33rd ASEE/IEEE frontiers in education conference, <http://lectureonline.cl.msu.edu/papers/v5-FIE-paper.pdf>, erişim tarihi: 10.04.2018.
- Moodle, 2018, Moodle statistics, <https://moodle.net/stats/?lang=tr>, erişim tarihi: 10.04.2018.
- Moodle, 2018, Registered Moodle sites, <https://moodle.net/sites/index.php?country=TR>, erişim tarihi: 10.04.2018.
- Natek, S., Zwilling, M., 2014, Student data mining solution–knowledge management system related to higher education institutions, Expert Systems with Applications, 41, 6400–6407.
- Oğuz, B., 2000, Eslestirme haznelemesinin biçimsel kavram analizi ile modellenmesi, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, 95 s.
- Özarıslan, S., 2014, Öğrenci performansının veri madenciliği ile belirlenmesi, Yüksek lisans tezi, Kırıkkale Üniversitesi, (yayımlanmamış).

**KAYNAKLAR DİZİNİ (devam)**

- Özarslan, Y., 2008, Uzaktan Eğitim Uygulamaları için Açık Kaynak Kodlu Öğrenme Yönetim Sistemleri, inet-tr'08 - XIII. Türkiye'de İnternet Konferansı Bildirileri, [http://www.academia.edu/1060120/Uzaktan\\_E%C4%9Fitim\\_Uygulamalar%C4%B1\\_i%C3%A7in\\_A%C3%A7%C4%B1k\\_Kaynak\\_Kodlu\\_%C3%96%C4%9Frenme\\_%C3%B6netim\\_Sistemleri](http://www.academia.edu/1060120/Uzaktan_E%C4%9Fitim_Uygulamalar%C4%B1_i%C3%A7in_A%C3%A7%C4%B1k_Kaynak_Kodlu_%C3%96%C4%9Frenme_%C3%B6netim_Sistemleri), erişim tarihi: 10.04.2018.
- Özbay, Ö., 2015, Veri Madenciliği Kavramı ve Eğitimde Veri Madenciliği Uygulamaları, Uluslararası Eğitim Bilimleri Dergisi, 2, 5, 262-272, [http://www.inesjournal.com/Makaleler/92710351\\_19-id-162.pdf](http://www.inesjournal.com/Makaleler/92710351_19-id-162.pdf), erişim tarihi: 10.04.2018.
- Öztürk, A., 2015, Açık ve uzaktan öğrenme sistemlerinde kümeleme analizi yöntemiyle öğrenen gruplarının belirlenmesi, Yüksek lisans tezi, Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, 131 s.
- Piatetsky-Shapiro. G., 1989, Knowledge Discovery in Real Databases, A Report on the IJCAI- 89 Workshop, AI Magazine, 11, 5, 68-70, <https://www.aaai.org/ojs/index.php/aimagazine/article/view/873>, erişim tarihi: 10.04.2018.
- Quinlan, J. R., 1986, Induction of Decision Trees, Machine Learning, 1, 81-106, <http://hunch.net/~coms-4771/quinlan.pdf>, erişim tarihi: 10.04.2018.
- RapidMiner, 2014, User Manual, <https://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf>, erişim tarihi: 10.04.2018.
- Romero, C., Lopez, M., Luna, J., Ventura, S., 2013, Predicting students' final performance from participation in on-line discussion forums, Computers & Education, 68, 458-472.
- Romero, C., Ventura, S., Salcines, E., 2008, Data mining in course management systems: Moodle case study and tutorial, Computer & Education, 51, 1, 368-384.
- Sewaiwar, P., Verma, K. K., 2015, Comparative Study of Various Decision Tree Classification Algorithm Using WEKA, International Journal of Emerging Research in Management & Technology, 4, 10, 87-91, [https://www.ermt.net/docs/papers/Volume\\_4/10\\_October2015/V4N10-113.pdf](https://www.ermt.net/docs/papers/Volume_4/10_October2015/V4N10-113.pdf), erişim tarihi: 10.04.2018.
- Şekeroğlu, S., 2010, Hizmet Sektöründe Bir Veri Madenciliği Uygulaması, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, 135 s. (yayımlanmamış).

**KAYNAKLAR DİZİNİ (devam)**

- Tekin, Y., 2007, Ders Yönetimi Bilgi Sistemi Destekli Web Tabanlı Öğretimin Kursiyerlerin Akademik Başarılarına ve Web Tabanlı Öğretime Yönelik Tutumlarına Etkisi (K.K.K. Lığı Mebs Okulu Örneği), Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, 168 s.
- Telcioğlu, M. B., 2007, Veri Madenciliğinde Genetik Programlama Temelli Yeni Bir Sınıflandırma Yaklaşımı Ve Uygulaması, Yüksek Lisans Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, 145 s.
- Wahbeh, A. H., Al-Radaideh, Q. A., Al-Kabi, M. N., Al-Shawakfa, E. M., 2011, A comparison Study between Data Mining Tools over some Classification Methods, International Journal of Advanced Computer Science and Applications, 18-25, <http://thesai.org/Downloads/SpecialIssueNo3/Paper%204-A%20Comparison%20Study%20between%20Data%20Mining%20Tools%20over%20some%20Classification%20Methods.pdf>, erişim tarihi: 10.04.2018.
- Yalın, H.İ., 2006, Öğretim Teknolojileri Ve Materyal Geliştirme, Nobel Yayın Dağıtım, 1-54.
- Zaiane, O. R., 2001, Web Usage Mining for a Better Web-based Learning Environment, Conference on Advanced Technology for Education, 60-64, <https://webdocs.cs.ualberta.ca/~zaiane/postscript/CATE2001.pdf>, erişim tarihi: 10.04.2018.
- Zhang Y., Oussena S., Clark T., Kim H., 2010, Use data mining to improve student retention in higher education - a case study, ICEIS 2010: Proceedings of the 12th International Conference on Enterprise Information Systems, 1, 190-197.