



**ÜNİVERSİTE KÜTÜPHANESİ VERİLERİ
ÜZERİNDE VERİ MADENCİLİĞİ
YÖNTEMLERİNİN UYGULANMASI**

YÜKSEK LİSANS TEZİ

Ahmet Gökhan GÜREL

Danışman

Dr. Öğr. Üyesi Ertuğrul ERGÜN

İNTERNET VE BİLİŞİM TEKNOLOJİLERİ YÖNETİMİ

ANABİLİM DALI

Haziran 2019

AFYON KOCATEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ÜNİVERSİTE KÜTÜPHANESİ VERİLERİ ÜZERİNDE VERİ
MADENCİLİĞİ YÖNTEMLERİNİN UYGULANMASI

Ahmet Gökhan GÜREL

Danışman
Dr. Öğr. Üyesi Ertuğrul ERGÜN

İNTERNET VE BİLİŞİM TEKNOLOJİLERİ YÖNETİMİ
ANABİLİM DALI

Haziran 2019

TEZ ONAY SAYFASI

Ahmet Gökhan GÜREL tarafından hazırlanan “Üniversite Kütüphanesi Verileri Üzerinde Veri Madenciliği Yöntemlerinin Uygulanması” adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca 24/06/2019 tarihinde aşağıdaki jüri tarafından **oy birliği** ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **İnternet ve Bilişim Teknolojileri Yönetimi Anabilim Dalı’nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Dr. Öğr. Üyesi Ertuğrul ERGÜN

Başkan : Prof. Dr. Çavuş ŞAHİN
Çanakkale Onsekiz Mart Üniv., Eğitim Fak.

Üye : Prof. Dr. Gürbüz OCAK
Afyon Kocatepe Üniv., Eğitim Fak.

Üye : Dr. Öğr. Üyesi Ertuğrul ERGÜN
Afyon Kocatepe Üniv., Uzaktan Eğitim M.Y.O.

İmza



Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetim Kurulu’nun
...../...../.... tarih ve
..... sayılı kararıyla onaylanmıştır.

.....
Prof. Dr. İbrahim EROL
Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİM SAYFASI

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

24 / 06 / 2019



Ahmet Gökhan GÜREL

ÖZET

Yüksek Lisans Tezi

ÜNİVERSİTE KÜTÜPHANESİ VERİLERİ ÜZERİNDE VERİ MADENCİLİĞİ YÖNTEMLERİNİN UYGULANMASI

Ahmet Gökhan GÜREL

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü

İnternet ve Bilişim Teknolojileri Yönetimi Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Ertuğrul ERGÜN

Günümüzde her alanda üretilen verilerin çoğu yapılandırılmamış biçimde veri yığını olarak veri tabanlarında veya veri ambarlarında tutulmaktadır. Bu durum büyük miktardaki veriden bilginin çıkarılmasını zorlaştırmaktadır. Akademik kütüphanelerde kullanıcı ile ilgili yapılandırılmamış büyük miktarda verinin yönetimi de zor ve önemli hale gelmiştir. Bu bağlamda veri madenciliği yöntemleri kullanılarak üniversite kütüphanesinde öğrencilerin kitap alma davranışları üzerinde çalışmalar yapılması önemli hale gelmiştir.

Bu çalışmada, Afyon Kocatepe Üniversite kütüphanesinden alınan veri setinde öğrencilerin cinsiyet, yaş, bölüm, sınıf gibi bilgileri ve hangi kitapları aldıkları, alınan kitapların ne kadar süre onlarda kaldığı gibi bilgiler veri olarak kullanılarak, veri madenciliği yöntemleri çalıştırılmıştır. Öncelikle ham verilerin temizlenme aşamaları ele alınarak, gereksiz verilerin çıkartılması, işe yarar verilerin analizlere aktarılması sağlanmıştır. Düzensiz ve karmaşık haldeki ham veriler düzenlenip, homojen kümeler oluşturularak verilerin anlamlı ve ilişkilendirilebilir hale getirilmesi hedeflenmiştir.

Model seçimi aşamasında veri madenciliğinin sınıflandırma ve kümeleme algoritmaları uygulanmıştır. Daha sonra Weka programı yardımıyla öğrencilerin bölüm, fakülte/enstitü, iade gün sayısı ve alınan kitabın türü kriter olarak kullanılarak sonuçlar görselleştirilmiştir. Aynı zamanda veriler, sınıflandırma ve kümeleme algoritmalarından çalışmanın amacına en uygun olan algoritmalar kullanılarak incelenmiştir.

Bu araştırma sonunda, öğrencilerin kitap alma davranışları incelenmiş, bölümlerine göre hangi kitapları tercih ettikleri, hangi derslerde daha çok araştırma yapıldığı, alınan kitapların birbirleri ile karşılaştırılması ve üniversite genelinde bölüm ve fakültelere göre kitapların dağılımı gibi istatistiksel sonuçların grafiklerle gösterilmesi sağlanmıştır. Bu çalışma sonunda elde edilen bilgi ve örüntülerle geliştirilebilecek kitap önerisi sistemleri, yoğun kullanıcı ve büyük hacimli kitaba sahip üniversite kütüphaneleri için çözümlerden biri haline gelebilir. Aynı zamanda geliştirilebilecek bir kullanıcı profili ödünç alma sistemine dayalı bir kütüphanede kitap öneri sistemi kullanmak kitap erişimini kolaylaştırabilecektir.

2019, x + 67 sayfa

Anahtar Kelimeler: Veri Madenciliği, Kümeleme, Sınıflandırma, Diferansiyel Hiyerarşik Kümeleme, Kütüphane Verisi.

ABSTRACT

M.Sc. Thesis

APPLICATION OF DATA MINING METHODS ON UNIVERSITY LIBRARY DATA

Ahmet Gökhan GÜREL

Afyon Kocatepe University

Graduate School of Natural and Applied Sciences

Department of Internet and Management of Information Technologies

Supervisor: Asst. Prof. Ertuğrul ERGÜN

Today, most of the data generated in each field is stored in databases or data warehouses as a data stack in unstructured form. This makes it difficult to extract information from large amounts of data. Management of large unstructured amounts of user-related data in academic libraries has also become difficult and important. In this context, it has become important to conduct studies on students ' book-borrow behavior in the university library using data mining methods.

In this study, data mining methods were used in the data set taken from the University Library by using information such as gender, age, department, class and which books were taken and how long the books were kept in them. First of all, the steps of cleaning up the raw data, removing unnecessary data and transferring the useful data to the analysis were provided. It is aimed to make the data meaningful and relatable by organizing raw data in irregular and complex states and creating homogeneous sets.

Classification and clustering algorithms of data mining were applied during the model selection phase. Later, with the help of the weka program, the results were visualized using the students ' department, faculty/Institute, number of return days and the type of book received as criteria. At the same time, the data was studied using algorithms that best fit the purpose of the study from classification and clustering algorithms.

At the end of this research, students ' book borrow behavior was examined, which books they preferred according to their departments, which courses were more researched, comparison of the books received with each other and the distribution of the books according to departments and faculties throughout the university, such as statistical results were shown with graphics. Book proposal systems that can be developed with the information and patterns obtained at the end of this study may become one of the solutions for university libraries with large volumes of books with intensive users. Using a book recommendation system in a library based on a user profile borrowing system that can also be developed will facilitate book access.

2019, x + 67 pages

Keywords: Data Mining, Clustering, Differential Hierarchical Clustering, Classification, Library Data.

TEŞEKKÜR

Bu araştırmanın konusu, deneysel çalışmaların yönlendirilmesi, veri madenciliği çalışmaları ve yazım süresince yardımlarını esirgememesi, elde edilen sonuçların değerlendirilmesi ve yazımı aşamasında göstermiş olduğu büyük katkılarından dolayı tez danışmanım Sayın Dr. Öğr. Üyesi ERTUĞRUL ERGÜN'e, araştırmamda kullandığım veri setinin temininde yardımcı olan Afyon Kocatepe Üniversitesi Merkez Kütüphanesi Daire Başkanı Sayın Mustafa AKYÜZ'e ve Şube Müdürü Fatma Çakır'a sonsuz teşekkürlerimi sunarım.

Sevgili aileme manevi hiçbir yardımı esirgemediğim yanımda oldukları için tüm kalbimle teşekkür ederim.

Bu çalışmada kullandığım WEKA programını ücretsiz paylaştığı için Waikato Üniversitesine teşekkür ederim.

Ahmet Gökhan GÜREL
AFYONKARAHİSAR, 2019

İÇİNDEKİLER DİZİNİ

Sayfa

ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER DİZİNİ	vi
KISALTMALAR DİZİNİ	viii
ŞEKİLLER DİZİNİ	ix
ÇİZELGELER DİZİNİ	x
1. GİRİŞ	1
2. LİTERATÜR BİLGİLERİ	4
2.1 Temel Kavramlar	4
2.1.1 Veri, Enformasyon ve Bilgi	4
2.1.2 Veri Tabanı Yönetim Sistemi.....	5
2.1.3 Veri Tabanlarında Bilgi Keşfi.....	6
2.2. Veri Ambarı	6
2.2.1 Veri Ambarının Tanımı	6
2.2.2 Veri Ambarının Tarihçesi ve Kullanımı	7
2.3 Veri Madenciliği (Data Mining)	8
2.3.1 Veri Madenciliği Tanımı.....	8
2.3.2 Veri Madenciliğinin Tarihsel Gelişimi	8
2.3.3 Veri Madenciliğinin Kullanıldığı Sektörler	9

2.3.4 Veri Madenciliğini Etkileyen Faktörler	14
2.3.5 Veri Madenciliğinde Karşılaşılan Problemler	15
2.3.6 Veri Madenciliği Süreci	15
2.4 Veri Madenciliğinde Kullanılan Yöntemler ve Teknikler	17
2.4.1 Veri Madenciliği Yöntemleri	17
2.4.2 Veri Madenciliği Teknikleri	20
2.5 Weka	25
2.5.1 Genel Bilgi ve Kullanımı	25
2.5.2 Weka Programı Modüllerinin Tanımları	26
2.6. İlgili Araştırmalar	29
3. MATERYAL ve METOT	33
3.1 Araştırmanın Modeli	33
3.2 Araştırmanın Evreni ve Örneklemi	33
3.3 Verilerin Elde Edilmesi	33
3.4 Veri Ön işleme	34
3.5 Model Seçimi	36
3.6 Modelin Değerlendirilmesi ve Yorumlanması	37
4. BULGULAR	39
5. TARTIŞMA ve SONUÇ	55
6. KAYNAKLAR	59
ÖZGEÇMİŞ	66
EKLER	67

KISALTMALAR DİZİNİ

Kısaltmalar

ARFF	Attribute-Relation File Format
CRISP-DM	Cross-Industry Standard Process for Data Mining
CSV	Comma Separated Values (Text file)
SQL	Structured Query Language
VTBK	Veri Tabanlarında Bilgi Keşfi
VTYS	Veri Tabanı Yönetim Sistemi
WEKA	Waikato Enviroment for Knowledge Analysis
YSA	Yapay Sinir Ağları



ŞEKİLLER DİZİNİ

Sayfa

Şekil 2.1	Veri Ambarı Temel Yapısı (İnt.Kyn.4).....	7
Şekil 2.2	Veri Madenciliğinin Tarihsel Gelişimi (İnt.Kyn.6).	9
Şekil 2.3	Veri Madenciliği Aşamaları (İnt.Kyn.9).	16
Şekil 2.4	Karar Ağacının Yapısına Örnek (Daş ve Türkoğlu 2014).	22
Şekil 2.5	K-means Adımları (İnt.Kyn.11).....	23
Şekil 2.6	Birimlerin Arasındaki Bağların Gösterildiği Saklı Katmanlı Sinir Ağı (Karataş 2004).	24
Şekil 2.7	Hava Durumu İçin Örnek Arff Dosyası İçeriği	26
Şekil 3.1	Ham Veriler	34
Şekil 3.2	Verilerin Filtreleme İşleminden Geçmiş Hali	35
Şekil 3.3	J48 Algoritmasının Fakülte/Enstitü Özniteliğine Göre Sonuçları	36
Şekil 4.1	Weka Sınıflandırma J48 Algoritmasının Fakülte/Enstitü Özniteliği Seçilerek Sonuçlarının Ağaç Görseli	53

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 2.1 Veri Madenciliğinin Kullanıldığı Alanlar (İnt.Kyn.7).	10
Çizelge 4.1 Ödünç Alınan Kitapların Türlerine Göre Dağılımları.....	39
Çizelge 4.2 Ödünç Alınan Kitapların Bölümlere Göre Dağılımları.....	40
Çizelge 4.3 Ödünç Alınan Kitapların Kurumlara Göre Dağılımları	40
Çizelge 4.4 Ödünç Alınan Kitapların Cinsiyete Göre Dağılımları.....	41
Çizelge 4.5 Ödünç Alınan Kitapların Yıllara Göre Dağılımları	41
Çizelge 4.6 NaiveBayes Algoritması ile Ödünç Alınan Kitapların Bölümlere Göre Dağılımları	42
Çizelge 4.7 J48 Algoritması Sonucunun Bölüm ve Kurumlara Göre Dağılımları.....	45
Çizelge 4.8 Beden Eğitimi ve Spor (YL) Bölümünün İade Gün Sayısına Göre Dağılımları	47
Çizelge 4.9 Bilgisayar Teknolojileri ve Programcılığı Bölümünün Cinsiyete Göre Dağılımları	48
Çizelge 4.10 Dış Ticaret Bölümünün İade Günü ve Cinsiyete Göre Dağılımları	48
Çizelge 4.11 İşletme Bölümünün Kurumlara, İade Güne ve Cinsiyete Göre Dağılımları	49
Çizelge 4.12 FilteredClusterer, MakeDensitBasedClusterer ve Simple K-Means Kümeleme Algoritmasının Ortak Sonucu	54

1. GİRİŞ

Günümüzde dünya çapındaki ekonomik kalkınmalara paralel olarak, bilişim ve teknoloji alanında önemli değişimler ve gelişmeler tüm hızıyla devam etmektedir. Bu gelişmeler birçok sektörün ihtiyaç duyduğu verilerin toplanmasına ve depolanmasına olanak sağlamaktadır. Dolayısıyla sürekli biriken veriler yığın haline gelmekte ve bunlar da işleminden geçmeden bir anlam ifade etmemektedir. Toplanan ve depolanan verilerden anlamlı bilgiler ve örüntüler çıkartabilmek için veri madenciliğinden yararlanılmaktadır. Veri madenciliğinin etkili yöntem ve teknikleri yardımıyla veriler arasında ilişki kurma ve anlamlandırma işlemleri yapılabilmektedir.

Dünyadaki kurumlar, işletmeler ve bireyler arasındaki tüm etkileşimler sırasında her gün büyük miktarda veri oluşmaktadır. Bu etkileşim sadece iş hayatında değil, aynı zamanda eğitim ve öğretim sektöründe de gözlenmektedir. Eğitim alanındaki veriler genellikle analiz edilmese de, konuyla ilgili bilgi içerir ve işleme ihtiyaç duyar. Eğitim veri setleri arasında, üniversitelerin ve diğer akademik kurumların kütüphanelerinde ve arşiv veritabanlarında oluşturulan veri setleri önemli bir yer tutmaktadır. Ağ teknolojilerinin gelişmesi ile birlikte üniversite kütüphanelerinin yapısının dijital kütüphanelere dönüşme süreci başlamıştır. Kütüphanelerde otomatik bilgi yönetim sistemi ile, okuyucuların kolayca kitap bilgilerini sorgulayabilmesini ve kitap ödünç takibinin yanı sıra bibliyografik bilgiye de erişimi de sağlanabilmektedir.

Akademik kurumların kütüphane veri yönetim sistemi, kullanıcılar tarafından günlük olarak yapılan ödünç, iade ve sorgular hakkında bilgi içeren verileri üretir ve depolar. Kütüphane yönetimi verilerden belirli aralıklarla çeşitli raporlar oluştursa da, halihazırda verilerden gizli bilgileri elde etmek ve bu verileri dönüştürmek için yararlı bir yöntem kullanmamaktadır. Bilginin hazinelerinden yararlanmaya ve yararlı bir bilgi haline gelmek için rafine edilmeye yardımcı olacak yeni nesil teknolojilere ve yöntemlere acil ihtiyaç duyulmaktadır. Veri madenciliği, bilgi keşfinin temel parçasıdır. Veri madenciliğinin teknikleri ve yöntemleri, dijital kütüphanede geniş uygulama alanı ve değere sahiptir. Veri madenciliği teknolojisi, insanların çok miktarda bilgiyi derinlemesine geliştirmelerine yardımcı olabilir (Zhang 2011).

Üretilen verileri işlemek ve analiz etmek için Bilgi Veri Keşfi (BVK) gibi metodolojilerin uygulanması gerekmektedir. Bu metodolojiler, değerli öngörülerini ortaya çıkarmak için veri madenciliği ve makine öğrenmesi algoritmalarının kullanılmasını sağlamaktadır. Veri madenciliği, üç model halinde uygulanmaktadır. Bu modeller sınıflandırma, kümeleme ve birliktelik kurallarıdır. Teknik ve algoritma olarak incelendiğinde veri madenciliği tanımlayıcı ve tahmin edici olarak da incelenebilmektedir (Cano *et al.* 2019). Kütüphane kullanıcılarından elde edilen verilerle okurların ihtiyaçlarının analiz edilip tahmin edilmesi, veri kümesine veri madenciliği teknolojisi uygulanarak yararlı bilgiye ulaşılması ve trendlerin bulunması sağlanabilir (Chen 2014).

Kütüphanelerde veri yönetim sisteminde veri madenciliği yöntemleri uygulanarak, okuyucuların kolayca kitap bilgilerini sorgulayabilmesi ve kitap ödünç takibinin yanı sıra bibliyografik bilgiye de erişimi sağlanabilecektir. Okurların ihtiyaçlarının analiz edilip tahmin edilmesi, kitaplar arasındaki ilişkiler ve karşılaştırmalar, yararlı bilgiye ulaşılması ve trendlerin bulunması sağlanabilecektir.

Problem Cümlesi

“Afyon Kocatepe Üniversitesi Merkez Kütüphanesinden alınan kitap ödünç alma davranış veri setinde kitap kodu, iade günü, cinsiyet, bölüm, fakülte/enstitü öznelikleri seçilip veri madenciliği yöntemleri uygulanarak sonuçlar incelendiğinde veriler arasında bir ilişki veya örüntü var mıdır?”

Alt Problemler

1. Kitap ödünç alma davranış veri setine veri madenciliği tekniklerinden görselleştirmenin uygulanmasıyla kitap kodlarına, bölümlere, fakültelere/enstitülere ve cinsiyete göre ayrı ayrı elde edilen istatistiksel sonuçlar arasında nasıl bir ilişki vardır?
2. Kitap ödünç alma davranış veri setine veri madenciliğinin 3 ana yöntemi olan birliktelik kuralları, sınıflandırma ve kümelemeden hangisini uygulamak daha iyi sonuç verecektir?

3. Birlikte kuralları, sınıflandırma ve kümeleme yöntemlerinin WEKA programı içerisindeki hangi algoritmaları daha başarılı sonuç verecektir?
4. Sınıflandırma algoritmalarından; bayes sınıflandırıcı algoritması NaiveBayes ve trees sınıflandırıcı algoritması J48 veriler arasındaki ilişki ve örüntüyü hangi doğrultuda ön plana çıkartmaktadır?
5. Kümeleme algoritmalarından; Farthest First, Filtered Clusterer, Make Density Based Clusterer ve Simple KMeans veriler arasındaki ilişki ve örüntüyü hangi doğrultuda ön plana çıkartmaktadır?

Sınırlılıklar

Araştırmadaki örneklem, Afyon Kocatepe Üniversitesi Merkez Kütüphanesi'nin 10 yıllık kitap ödünç alma veri seti 402918 kayıt ve içeriğinde Demirbaş, Sınıflama, Kitap Ad, Ödünç Tarihi, İade Tarihi, Sıra No, Üye, Cinsiyet, Bölüm, Enstitü, Fakülte öznitelikler ile sınırlıdır. Ayrıca araştırmada uygulanan modeller ile elde edilen bulgularla sınırlıdır.

Araştırmanın Amacı

Araştırmadaki temel amaç, kütüphaneden alınan veri seti üzerinde öğrencilerin kitap alma davranışları incelenerek, bölümlerine göre hangi kitapları tercih ettikleri, hangi derslerde daha çok araştırma yapıldığı, alınan kitapların birbirleri ile karşılaştırılması ve üniversite genelinde bölüm ve fakültelere göre kitapların dağılımı gibi verilere WEKA programı içerisindeki veri madenciliği yöntemlerinden sınıflandırma ve kümeleme algoritmaları uygulanarak istatistiksel sonuçların grafiklerle gösterilmesidir.

Bu tez çalışmasında kütüphane kaynaklarından elde edilen veriler üzerinde bir veri madenciliği araştırması gerçekleştirilmiş ve genel olarak kitap ödünç alma davranışları üzerinden, öğrencilerin bölümlerine göre hangi kitapları tercih ettikleri, hangi derslerde daha çok araştırma yaptığı, alınan kitapların birbirleri ile karşılaştırılması ve üniversite genelinde bölüm ve fakültelere göre kitapların dağılımı gibi sonuçlara ulaşılmıştır. WEKA programı ile veri madenciliği yöntemlerinden kümeleme ve sınıflandırma algoritmaları uygulanarak istatistiksel sonuçların grafiklerle gösterilmesi sağlanmıştır.

2. LİTERATÜR BİLGİLERİ

Bu bölümde araştırmaya temel teşkil eden veri madenciliği bilim dalı ile ilgili bazı teorik bilgiler, uygulama sürecinde kullanılan algortima, yöntem ve metodlar ve alanla ilgili yapılmış bazı araştırmalar yer verilmiştir.

2.1 Temel Kavramlar

2.1.1 Veri, Enformasyon ve Bilgi

Veri, ham haldeki, işlenmemiş kayıtlar olarak adlandırılmaktadır. Verilerin temel amacı, faaliyetleri veya durumları kaydetmek, gerçek resmi veya gerçek olayı yakalamaya çalışmaktır.

Enformasyon, karar veya eylem için ilgili anlam, ima veya girdi içeren bir mesajdır. Enformasyon hem mevcut (iletişim) hem de tarihsel (işlenmiş veriler veya ‘yeniden oluşturulmuş resim’) kaynaklarından gelmektedir. Özü, bilgi amaçlı bir fırsat sunmak veya problem çözmeye yardım etmektir.

Bilgi, zihin içinde veya beyinde bulunan biliş veya tanıma, hareket etme kapasitesi ve anlayıştır. Bilginin amacı hayatımızı iyileştirmektir. İşletme bağlamında, bilginin amacı, işletme ve tüm paydaşları için değer yaratmak veya arttırmaktır. Kısacası, bilginin nihai amacı değer yaratmak içindir.

Veri, enformasyon ve bilgi arasındaki karmaşık ilişkiyi anlamamanın anahtarı, veri ve enformasyon kaynağında yatmaktadır. Her ikisinin de kaynağı iki yönlüdür: faaliyetler ve durumlar. Enformasyon ve bilgi arasındaki ilişkiyi anlamamanın anahtarı ise, enformasyonun nerede olduğunu bilmektir. Bilginin özünde faaliyetlerden ve durumlardan üretilen bir mesaj bulunmaktaydı. Ancak, enformasyon depolama ortamında (veri tabanı, yazıcı, video kasetleri vb.) veri biçiminde veya insan zihninde bilgi olarak bulunmaktadır. Eğer durum buysa, o zaman veri ve enformasyon, enformasyon ve bilgi arasındaki örtüşme belirginleşmektedir, yani aynı anda farklı yerlerde yer kaplamaktadırlar (Liew 2007).

2.1.2 Veri Tabanı Yönetim Sistemi

Bir veri tabanı yönetim sistemi (VTYS) veri tabanlarını oluşturmak ve yönetmek için kullanılan sistem yazılımıdır. VTYS, kullanıcılara ve programcılara verileri oluşturmak, almak, güncellemek ve yönetmek için sistematik bir yol sağlar. Bir VTYS, son kullanıcıların bir veri tabanındaki verileri oluşturmasını, okumasını, güncellemesini ve silmesini mümkün kılar. VTYS, verinin sürekli organize ve kolayca erişilebilir kalmasını sağlayarak, veri tabanı ve son kullanıcılar veya uygulama programları arasında bir ara yüz olarak hizmet vermektedir.

VTYS, üç önemli şeyi yönetir: bunlar veriler, verilere erişilmesini, kilitlenmesini ve değiştirilmesini sağlayan veri tabanı altyapısı ve veri tabanının mantıksal yapısını tanımlayan veri tabanı şemasıdır. Bu üç temel öge eşzamanlılık, güvenlik, veri bütünlüğü ve tekdüzen yönetim yordamları sağlamaya yardımcı olur. VTYS tarafından desteklenen tipik veri tabanı yönetim görevleri değişiklik yönetimi, performans izleme ya da ayarlama, yedekleme ve kurtarmayı içerir. Birçok veri tabanı yönetim sistemleri de otomatik geri alma, yeniden başlatma ve kurtarmanın yanı sıra günlük ve denetim faaliyetinden sorumludur.

VTYS, birden çok kullanıcı tarafından birden çok konumdan kontrollü bir şekilde erişilebilen merkezi bir veri görünümü sağlamak için belki de en kullanışlıdır. VTYS, son kullanıcının gördüğü verilerin yanı sıra son kullanıcının verileri nasıl görüntüleyebileceğini sınırlayabilir ve tek bir veri tabanı şemasının birçok görünümünü sağlar. Son kullanıcılar ve yazılım programları, verilerin fiziksel olarak nerede bulunduğunu veya VTYS'nin tüm istekleri işlediği için hangi tür depolama ortamı bulunduğunu anlamak zorunda değildir.

VTYS, mantıksal ve fiziksel veri bağımsızlığını sunabilir. Bu, kullanıcıların ve uygulamaların verilerin nerede saklandığını bilmesine veya verilerin fiziksel yapısındaki değişikliklerden (depolama ve donanım) endişe duymasına neden olabileceği anlamına gelir. Programlar VTYS tarafından sağlanan veri tabanı için uygulama programlama arabirimini (API) kullandığı sürece, geliştiriciler yalnızca veri tabanında değişiklikler yapıldığı için programları değiştirmek zorunda kalmazlar (İnt.Kyn.1).

2.1.3 Veri Tabanlarında Bilgi Keşfi

Geleneksel olarak, veri madenciliği ve bilgi bulma el ile gerçekleştirilirdi. Zaman geçtikçe, birçok sistemdeki veri miktarı terabayt boyutundan daha büyük hale geldi ve artık manuel olarak muhafaza edilemedi. Dahası, herhangi bir işletmenin başarılı bir şekilde varlığı için, verilerdeki temel kalıpları keşfetmesi önemlidir. Sonuç olarak, gizli verileri keşfetmek ve yapay zekanın bir parçasını oluşturan varsayımları yapmak için çeşitli yazılım araçları geliştirilmiştir.

VTBK süreci son 10 yılda zirveye ulaşmıştır. Şimdi endüktif öğrenme, Bayes istatistikleri, semantik sorgu optimizasyonu, uzman sistemler için bilgi edinme ve bilgi teorisi içeren keşif için birçok farklı yaklaşım barındırmaktadır. Nihai hedefi, düşük seviyeli verilerden üst düzey bilgi elde etmektir.

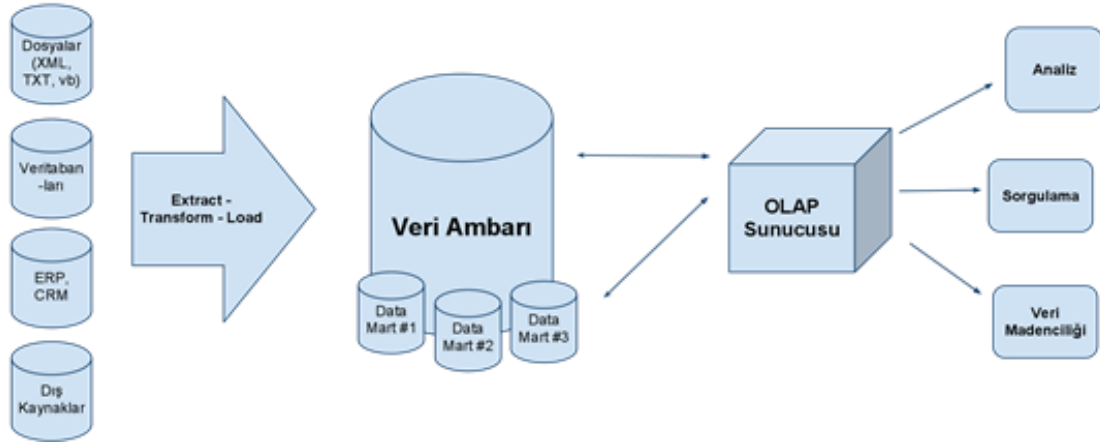
VTBK çok disiplinli “multidisipliner” faaliyetleri içerir. Bu, veri depolama ve erişimi, algoritmaları büyük veri kümelerine ölçeklendirmeyi ve sonuçları yorumlamayı kapsar. Veri depolama alanına dahil olan veri temizleme ve veri erişim süreci VTBK sürecini kolaylaştırır. Yapay zeka, deney ve gözlemlerden ampirik yasaları keşfederek VTBK’yi de desteklemektedir. Verilerde tanınan desenler yeni veriler üzerinde geçerli olmalı ve bir dereceye kadar kesinliğe sahip olmalıdır. Bu desenler yeni bilgi olarak kabul edilir (İnt.Kyn.2).

2.2. Veri Ambarı

2.2.1 Veri Ambarının Tanımı

Bir veri ambarı, operasyonel sistemler ve dış veri kaynaklarından elde edilen kurumsal bilgi ve verilerin bir koleksiyonudur. Veri ambarı, farklı toplama düzeylerinde veri birleştirme, analiz ve raporlamaya izin vererek iş kararlarını desteklemek üzere tasarlanmıştır. Veri çıkarma, dönüşüm ve yükleme işlemleri yoluyla veri ambarı içine doldurulma işlemi sağlanır. (İnt.Kyn.3)

Veri ambarının şematik gösterimi Şekil 2.1’de görülmektedir.



Şekil 2.1 Veri Ambarı Temel Yapısı (İnt.Kyn.4).

2.2.2 Veri Ambarının Tarihçesi ve Kullanımı

Veri ambarı mimarisi 1980’lerde operasyonel sistemlerden karar destek sistemlerine veri akışını desteklemek için tasarlanmış bir mimari model olarak doğmuştur. Bu sistemler, şirketler tarafından zamanla biriken büyük miktarlarda heterojen verilerin analizini gerektirir.

Bir veri ambarında, birçok heterojen kaynaktan elde edilen veriler tek bir alana çıkarılır, karar destek sistemi ihtiyaçlarına göre dönüştürülür ve depoya saklanır. Örneğin, bir şirketteki çalışanlar, maaşlar, gelişmiş ürünler, müşteri bilgileri, satış ve faturalarla ilgili bilgileri depolar. CEO, en son maliyet azaltma önlemleriyle ilgili bir soru sormak isteyebilir; cevaplar tüm bu verilerin analizini içerecektir. Bu, veri ambarının ana hizmetidir, yani yöneticilerin tüm bu farklı ham veri öğelerine dayanarak iş kararlarına ulaşmasına izin verir. Böylece, bir veri ambarı gelecekteki karar verme sürecine katkıda bulunur. Yukarıdaki örnekte olduğu gibi, bir firma yöneticisi belirli bir ürünün piyasa talebini, coğrafi bölgeye göre satış verilerini veya diğer soruları yanıtlamak için depo verilerini sorgulayabilir. Bu, belirli bir ürünü daha etkili bir şekilde pazarlamak için gerekli adımlar hakkında bilgi sağlar. Operasyonel veri deposundan farklı olarak, bir veri ambarı, kritik iş kararlarına ulaşmak için analiz edilebilen toplam geçmiş verileri içerir. İlişkili maliyetlere ve çabalara rağmen, günümüzde çoğu büyük şirket veri depolarını kullanmaktadır (İnt.Kyn.3).

2.3 Veri Madenciliği (Data Mining)

2.3.1 Veri Madenciliği Tanımı

Veri madenciliği; veri tabanlarında bilgi keşfi, bilgisayar bilimlerinde ise büyük hacimli verilerde yararlı kalıpları ve ilişkileri keşfetme süreci olarak da adlandırılır. Veri setleri olarak bilinen büyük dijital koleksiyonları analiz etmek için veri tabanı yönetimi ile istatistik ve yapay zeka (örneğin sinir ağları ve makine öğrenimi gibi) araçlarını birleştirir. Veri madenciliği yaygın olarak sektör bazında sigorta, bankacılık, perakende, bilim ve araştırma bazında astronomi, tıp ve hükümet güvenliği bazında suçluların ve teröristlerin tespitinde kullanılmaktadır. Ayrıca, çok sayıda hükümet ve özel büyük veri tabanlarının çoğalmasi ile bireysel kayıtların yetkisiz görüntülenmesi, tahrifattan çıkan verilerin doğru ve güvenli olmasını sağlamak için teknikler sağlar. Çoğu veri madenciliği türü, belirli bireyler hakkında bilgi yerine bir grup hakkında genel bilgi bulmaya yöneliktir. Örnek olarak bir süpermarket, birçok kişiye birçok öğe satmaktan çok bir kişiye bir öğe satmaktan daha az endişe duymaktadır. Desen analizi, dolandırıcılık veya diğer suç faaliyetleri gibi anormal bireysel davranışları ayırt etmek için de veri madenciliği kullanılabilir (İnt.Kyn.5).

2.3.2 Veri Madenciliğinin Tarihsel Gelişimi

“Veri madenciliği” terimi 1990’larda tanıtılmış, ancak veri madenciliği uzun bir geçmişli olan bir alanın evrimidir.

Verideki kalıpları tanımlamanın erken yöntemleri arasında Bayes teoremi (1700) ve regresyon analizi (1800) lerde bulunmuştur. Bilgisayar teknolojisinin çoğalmasi, her yerde bulunması ve artan gücü, veri toplama, depolama ve manipölasyonları artırmıştır. Veri setleri boyut ve karmaşıklık içinde büyüdükçe, doğrudan uygulamalı veri analizi, dolaylı, otomatik veri işleme ile giderek artmaktadır. Sinir ağları, kümeleme, genetik algoritmalar (1950), karar ağaçları (1960) ve destek vektör makineleri (1990) gibi bilgisayar bilimindeki diğer keşiflerle desteklenmiştir. Veri madenciliği köklerinin üç ana hattı: klasik istatistik, yapay zekâ ve makine öğrenimi boyunca uzanmaktadır (Kumar *et al.* 2011). Tarihsel gelişim Şekil 2.2’de gösterilmiştir.

1950'ler	• İlk bilgisayarlar (sayım için)
1960'lar	• Veritabanı ve verilerin depolanması • perseptonlar
1970'ler	• İlişkisel Veri Tabanı Yönetim Sistemleri • Basit kurallara dayanan uzman sistemler ve Makine öğrenimi
1980'ler	• Büyük miktarda veri içeren veri tabanları • SQL sorgu dili
1990'lar	• Veri Tabanlarında Bilgi Keşfi Çalışma Grubu ve Sonuç bildirgesi • Veri madenciliği için ilk yazılım
2000'ler	• Tüm alanlar için veri madenciliği uygulamaları

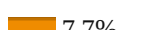
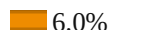
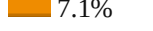
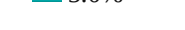
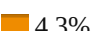
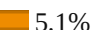
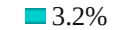
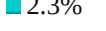
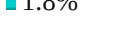
Şekil 2.2 Veri Madenciliğinin Tarihsel Gelişimi (İnt.Kyn.6).

2.3.3 Veri Madenciliğinin Kullanıldığı Sektörler

Veri madenciliği, bankacılık, borsa, pazarlama yönetimi, perakende satış, sinyal işleme, sigorta, telekomünikasyon, elektronik ticaret, sağlık, tıp, biyoloji, genetik, sanayi, inşaat, eğitim, zekâ, bilim ve mühendislik gibi birçok dalda uygulama alanlarına sahiptir. Çeşitli işletmeler rekabet avantajı kazanmak ve işletmelerin büyümesine yardımcı olmak için kritik iş süreçlerine veri madenciliği uygular (Özkan 2008, Silahtaroğlu 2008, Kaya ve Özel 2014).

Söz konusu sektörleri kapsayan ana sektörlerdeki veri madenciliği uygulamaları, veri madenciliğinin nasıl ve ne amaçla kullanıldığı sektör bazında 2014, 2015 ve 2016 yıllarında yapılan çalışma verileri Şekil 2.3'de verilmiştir. Yapılan çalışmalardan Eğitim alanındaki verilerin 2016 da düşmüş, reklamcılık alanının ise yükselmiş olduğu gözlenmiştir (İnt.Kyn.7).

Çizelge 2.1 Veri Madenciliğinin Kullanıldığı Alanlar (İnt.Kyn.7).

<i>VM Uygulandığı Alanlar</i>	 % 2016	 % 2015	 % 2014
CRM / Tüketim hesaplamaları (90)	 16.3%	 18.6%	 22.2%
Finans (83)	 15.0%	 15.4%	 10.9%
Bankacılık (74)	 13.4%	 14.3%	 16.7%
Reklamcılık (66)	 12.0%	 8.9%	 10.4%
Bilim (66)	 12.0%	 11.7%	 13.6%
Sağlık Hizmetleri (66)	 12.0%	 13.4%	 16.3%
Dolandırıcılık Algılama (61)	 11.1%	 10.0%	 13.6%
Perakende (57)	 10.3%	 9.1%	 13.6%
Sigorta (51)	 9.2%	 7.4%	 8.6%
E-ticaret (49)	 8.9%	 10.3%	 9.5%
Telekomünikasyon / Kablo (46)	 8.3%	 7.7%	 9.0%
Sosyal Medya / Sosyal Ağlar (46)	 8.3%	 10.3%	 8.6%
Program sağlaması (40)	 7.2%	 6.0%	 7.2%
BT / Ağ Altyapısı (40)	 7.2%	 6.6%	na
Petrol / Gaz / Enerji (39)	 7.1%	 8.9%	 9.5%
Eğitim (39)	 7.1%	 10.0%	 7.7%
Kredi Puanlaması (38)	 6.9%	 7.1%	 8.1%
Tedarik Zinciri (35)	 6.5	na	na
Tıbbi / İlaç (36)	 6.5%	 6.0%	 7.2%
Diğer (35)	 6.3%	 8.9%	 13.6%
Yatırım / Borsa (34)	 6.2%	 4.3%	 5.0%
Biyoteknoloji / Genomik (32)	 5.8%	 4.9%	 6.8%
İmalat (31)	 5.6%	 6.9%	 9.0%
Devlet / Askeri (31)	 5.6%	 7.1%	 6.3%
Arama / Web İçerik Madenciliği (30)	 5.4%	 6.0%	 6.3%
Otomobil/Kendinden Sürüş Otomobil (25)	 4.5%	 4.3%	 5.9%
Doğrudan Pazarlama (24)	 4.3%	 5.1%	 7.2%
Madencilik (23)	 4.2%	 3.7%	na
Seyahat / Konaklama (22)	 4.0%	 2.6%	 3.2%
Eğlence / Müzik / TV / Filmler (22)	 4.0%	 3.1%	 1.8%
İK / iş gücü analiz (20)	 3.6%	 6.3%	 5.9%
Mobil uygulamalar (18)	 3.3%	 1.4%	 2.3%
Oyunlar (16)	 2.9%	 4.0%	 1.8%
Güvenlik / Terörle Mücadele (15)	 2.7%	 2.3%	 2.3%
Sosyal Politika / Anket analizi (10)	 1.8%	 1.7%	 1.8%

2.3.3.1 Bankacılık ve Finans Sektörü

Veri madenciliği çoğunlukla bankacılık ve finans sektörlerinde, müşteri profilinin ne, ne zaman ve neden tercih ettiği sorularını belirlemek için kullanılır. Aynı zamanda, bu alanlarda, doğru talep yaratmak, doğru zaman taleplerini sunmak ve uygun çözümler bulmak için de kullanılır. Ayrıca finansal tahmin, hisse senedi fiyatlarının tahmini, yeni yatırımların yönetimi, yatırım portföyünün belirlenmesi, pazarlama stratejilerinin oluşturulması, risk analizleri yapma, insan kaynakları açısından doğru seçimi yapma, kredi ve kredi kartı dolandırıcılık tahminleri, kredi limiti belirleme ve ücret yönetimi uygulamalarında kullanılır. Veri madenciliği uygulamaları, bankacılık sektöründeki bazı temel iş süreçlerinin performansını artırmak için yaygın olarak kullanılmaktadır. Bazı bankalar, örneğin Türkiye'de Garanti Bankası, müşterilerin davranış modelleri hakkında bilgi sahibi olmak için veri madenciliği yöntemlerini kullanır ve müşteri “kredi kartları” seçiciliği ile karakterleri arasındaki ilişkiyi inceleyerek uygun ve başarılı promosyonlar sunar (Bhambri 2011, Pulakkazhy and Balan 2013).

2.3.3.2 Eğitim Sektörü

Eğitim sektöründeki birçok çalışmada veri madenciliğinin kullanıldığı gözlenmiştir. Öğrencilerin geçiş ve başarısızlık durumunu belirleme, üniversiteye kayıt yapan öğrencilerin başarısını etkileyen faktörler, yeni kayıtlı öğrencilerin tercih sırasını etkileyen aktörleri belirleme, üniversite bölümünün tercihini oluşturma, demografik ve kişisel özelliklerine göre bir meslek seçimi, öğrencilerin başarısız olmasını önleme ve başarıyı etkileyen faktörleri belirleme, öğrencilerin mezun olduğu okul türü ile üniversite bölümleri arasındaki ilişkileri belirleme, uzaktan eğitim öğrencilerinin çalışma faaliyetlerini değerlendirme, üniversite giriş sınavına giren öğrencilerin profillerini ve tercihlerini belirleme, akademik başarı ile üniversite öğrencilerinin ders dışı etkinliklerine katılım arasındaki ilişkinin belirlenmesi, öğrencilerin sosyoekonomik düzeyi ile akademik öğrenme düzeyi arasındaki ilişkilerin belirlenmesi, öğrenci giriş puanları ve okul başarısı arasında bir ilişki olup olmadığını belirleme konularında kullanılır. Eğitim sektöründeki bu kullanım alanları, öğretmenlerin sınıflarını yönetmelerine, öğrencilerinin öğrenmelerini kolaylaştırma ve öğrencilere proaktif geribildirim sağlamalarına yardımcı olur (Çiftçi 2006, Ayık vd. 2007, Dolgun vd. 2007, Savaş vd. 2012, Kurt ve Erdem 2012).

2.3.3.3 Telekomünikasyon Sektörü

Veri madenciliği telekomünikasyon sektöründe kullanılabilir. İletişim sektöründe mobil kullanıcı hareketlerini tahmin etmek, mobil kullanıcıların gelecekteki hareketlerini belirlemek, sahtekarlıkları tespit etmek, insan temelli analizin çoğunu azaltmak, belirli zamanlarda daha fazla arama yapan müşterileri etkileyen faktörleri belirlemek, sosyal ağ kullanımında kullanıcı şablonlarını belirlemek, demografik verileri kullanarak yeni olasılıkları tanımlamak, süspansiyon veya deaktivasyon gibi özel eyleme ihtiyaç duyan müşterilerin özelliklerini tanımlamak, müşteri kaybını önlemek konularında kullanılır. Örneğin müşteri kaybını önlemek için telekomünikasyon kuruluşları, strateji geliştirmeye, düşük maliyete ve etkili kampanyalara başvurarak bu sorunu ortadan kaldıracaktır. Bu sektörde veri madenciliğinin müşteri kaybı ile ilgili olarak kullanılmasına iyi bir örnek, Amerika'nın en büyük kablosuz iletişim sağlayıcısı olan Verizon'dur. Çalışmada Verizon, kaybedecekleri müşterileri ve müşteri kaybına neden olan faktörleri tanımlamak için veri madenciliği yöntemlerine başvurmuştur (Shahbaz and Rahman 2008, Bozkır vd. 2010).

2.3.3.4 Sağlık Sektörü

Veri madenciliği teknikleri ve uygulama araçları, sağlık sektörü için daha değerlidir. Veri madenciliği uygulamaları, sağlık verileri üzerinde yapılan veri işleme çalışmasının karmaşıklığını azaltmak için yaygın olarak kullanılmaktadır. Hastalığı teşhis etme, hastalığa uygulanacak tedavi yöntemini belirleme, hastanelerde kaynak kullanımını ve hasta numaralarını tahmin etme, hastanede uygulanan tedavi yöntemlerinin başarısını belirleme, hasta verilerini yaş, cinsiyet, ırk ve tedavi gibi faktörlere göre sınıflandırma, ameliyatlarda yüksek risk faktörlerini belirleme, hastane harcamalarında yolsuzluğu önleme konularında kullanılır. Sağlık alanında veri madenciliği çalışmalarının en iyi örneklerinden biri San Francisco Kalp Enstitüsü'nde yapılan bir çalışmadır. Bu çalışmada, hastanın durumu, laboratuvar verileri ve hastalardan elde edilen diğer tıbbi veriler gibi bazı veriler, hasta sonuçlarını iyileştirmek ve hastanın hastanede kalış süresini azaltmak için veri madenciliği yöntemleri ile bilgiye dönüştürülmüştür (Durairaj and Ranjani 2013, Biçer vd. 2014, Aydın ve Keleş 2017, Aydın 2017).

2.3.3.5 Kamu Sektörü

Veri madenciliği genellikle kamu sektöründe kamu güvenliği ve güvenlik sorunlarını tahmin etmek için kullanılır. Veri madenciliği teknikleri, kararları optimize etmek için kamu sektörü için açık fırsatlar sunar. Bu kararlar genel geçmiş deneyim ve geçmiş verilerden çıkarılan eğilimlerdir. Bunun dışında, vergi ile ilgili yolsuzluk belirleme, bütçe vergi sistemindeki değişikliklerin etkisini tahmin etme, atıkları belirleme ve atıkların neden olduğu hasarı önlemek için nüfusu tahmin etme, hava durumu tahmin etme, yeni iş fırsatları belirleme, çalışanların performansını ölçme, iş süreçlerini yönetme, kamu harcamalarını sınıflandırma, kaynakların doğru kullanımını planlama, kamu yatırımının geleceğini tahmin etme, savunma endüstrisindeki verileri analiz etme, güvenlik açısından hangi suçluların suç işleyeceğini belirleme konularında kullanılır. Bu alandaki en önemli örneklerden biri Türkiye'de e-devlet uygulamasıdır. Veri madenciliği yöntemleri, e-Devlet üzerindeki web sayfalarının ziyaret edildiği bilgi ve siparişin eşzamanlı erişimini belirleyerek, kullanıcının geçmişteki davranışlarına göre web sayfasının yeniden düzenlenmesi ile birlikte daha hızlı geri bildirim sağlamaktadır (Wang *et al.* 2008).

2.3.3.6 İnşaat Sektörü

Veri madenciliği inşaat sektöründe yapı, proje yönetimi, hidrolik, iş sağlığı ve güvenliği uygulamaları, deprem verilerinin analizi, zemin çalışmaları ve diğer birçok alanda kullanılmaktadır. Bu bağlamda yapılan çalışmalar ışığında, proje belgelerinde bilgi sınıflandırma şemasını oluşturma, atıkları belirleme ve atıkların neden olduğu hasarı önleme, karayolu inşaat maliyetini tahmin etme, nüfus tahmini yapma, yeni iş fırsatları belirleme, çimento ürününün basınç dayanımını tahmin etme, çalışanların performansını ölçme, iş süreçlerini yönetme, inşaat sektöründe iş kazalarının özelliklerini tanımlama, kaynakların doğru kullanımını planlama, işçi verimliliğini ölçme, beton basınç dayanımını belirleme, kamu yatırımının geleceğini tahmin etme, savunma endüstrisindeki verileri analiz etme, liderlik ilişkisini belirleme, şef ve işçi arasındaki motivasyon, inşaat yönetiminde veri madenciliği yönteminin yerini belirleme konularında kullanılır (Caldas *et al.* 2002, Wilmot and Cheng 2003, Baykasoğlu 2005, Liaoa and Perng 2008, Kaya vd. 2013, Keleş ve Kaya 2014, Özel ve Topsakal 2014, Keleş 2016, Keleş 2017).

2.3.3.7 Mühendislik ve Bilim Sektörü

Astronomi, biyoinformatik, bilgisayar, ceza bilimi, mühendislik, yerbilimleri, matematik, yazılım vb. bilimsel alanlardan büyük miktarlarda veri toplanmıştır. Veri madenciliği, mühendislik ve bilim sektöründe de dahil olmak üzere birçok avantaj sağlar. Firmalarda kullanılan yazılım sürecini yönetme, görev sayısını azaltma, yazılım geliştirme yaşam döngüsünün hızını arttırma, zaman ve emek tasarrufu, organizasyon tahmin analizi ile kuruluşlara rekabet avantajı sağlanması, üretim sürecinin iyileştirilmesi, biyolojik literatür analizi, uzaktan algılama, toprak kalitesi analizi, suç tespit deseni, türbülanslı akış analizi için gerçek zamanlı özellik ekstraksiyonu, kaliteli tohum elde etme, gelişen yeni ürün ırkları, astronomik nesneleri sınıflandırma, ekosistem modelleme, soğuk depoların en iyi kullanımı ve kanal suyunun kullanımı için ilişkileri keşfetme, biyoinformatik dizilerin sınıflandırılması konularında kullanılır. Mühendislik ve bilim sektöründeki bu kullanım alanları, kullanıcıların kullanılan yazılımın sistem performansını iyileştirmelerine, kullanılan mühendislik yazılım geliştirme süreçlerinin birçok bölümüne bilgi vermelerine ve gelecekteki karar verme sürecini planlamalarına yardımcı olur (Shahbaz and Rahman 2008).

2.3.4 Veri Madenciliğini Etkileyen Faktörler

Veri Madenciliğini etkileyen 5 ana faktörün tanımları aşağıda verilmiştir.

Veri, söz konusu veri madenciliği olduğunda bulunması gereken en önemli faktördür. Bu da veri madenciliğinin iyileşmesinde ve gelişmesinde etkisinin fazla olduğu anlamına gelmektedir.

Donanım, belirli bir süre işlenmemiş verilerin faaliyete tekrar geçirilmesini mümkün kılan bir faktördür.

Bilgisayar ağları, internet ile yapılandırılan bilgisayar ağları alanı dağılmış biçimdeki verileri incelemek ve analiz etmeyi sağlayan faktördür.

Bilimsel hesaplamalar, bilimde üçüncü yol olarak bilim insanları tarafından bulunmuştur. Veri madenciliği, bilgi keşfi, teori, deney gibi kavramları birbirine bağlamada kullanılan en önemli faktördür.

Ticari eğilimler, şirketler varoluşlarını devam ettirmeye başlamaları ile devamlı birbirleri ile rekabetteler. Rekabet içerisinde kendileri ve işletmelerinde varlıklarını devam ettirmek amacıyla daha yüksek kalitede hizmet sunmak ve daha hızlı hareket etmek için görev üstlenmişlerdir. Bunların hepsini yaparken de işletmelerini kayıp ve zarara uğratmayacak biçimde maliyetlerini ve iş gücünü ona göre ayarlamaya dikkat etmelidir (İnt.Kyn.8).

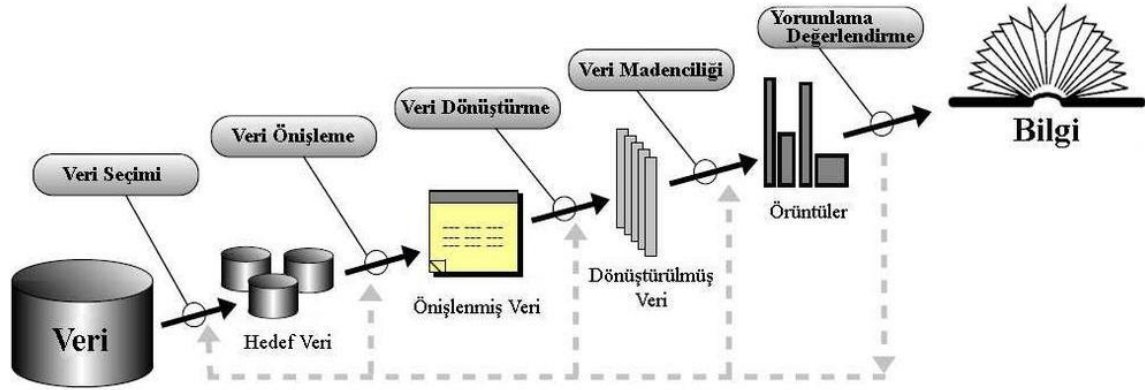
2.3.5 Veri Madenciliğinde Karşılaşılan Problemler

Veri ortamında büyük boyutlu verilerin olması problemlerin oluşmasına neden olabilmektedir. Örneğin, boyutu küçük olan veri kümeleri üzerinde oluşturulan veri madenciliği sistemleri, büyük boyutlu, gürültülü, belirsiz gibi verilerin yer aldığı ortamlarda yanlış şekilde çalışma yapabilir. Bu nedenle, veri madenciliği sistemlerinin hazırlanması işleminde ortaya çıkabilecek olan problemlerin incelenmesi ve bu problemlerin çözülmesi gerekmektedir. Bu veri madenciliği uygulamalarında karşılaşılabilecek problemler: “artık veri, belirsizlik, dinamik veri, eksik veri, farklı tipteki verileri ele alma, gürültülü ve kayıp değerler, sınırlı bilgi ve veri tabanı boyutu” olarak tanımlanabilir (Savaş vd. 2012).

2.3.6 Veri Madenciliği Süreci

Veri madenciliğinin diğer bir adı da süreçtir. Veri yığınları içerisinde, soyut kazı çalışmaları ile anlamlı verinin ortaya çıkartılmasının yanında, bilgi keşfi süresince örüntüleri ayrıştırarak süzer ve bir sonraki adıma hazır hale getirir. Bu da sürecin bir aşamasıdır. Veri ambarlarında ya da veri tabanlarında bulunan veriler üzerinde veri madenciliği tekniklerinin uygulanabilmesi bu verilerin belirli aşamalardan geçmesi gerekir.

Veri madenciliği çalışmasının başarılı bir biçimde uygulanması gereken aşamalar Şekil 2.3'te gösterilmiştir.



Şekil 2.3 Veri Madenciliği Aşamaları (İnt.Kyn.9).

Veri seçimi, veri madenciliğinin aşamalarından en çok zaman harcanan aşamadır. Bilgi sistemlerinde ortaya çıkan bilgi dikkatli incelenmelidir ve sorunla doğrudan ilişkilendirilmelidir. Bu aşama, incelemeyi yapan insanın veri kalitesini ölçmesi bakımından önemlidir. Veri madenciliği uygulaması için büyük hacimli verilerin tek bir veri ambarında ya da veri tabanında bir araya getirilmesi gereklidir. Veri seçimi, veri filtreleme olarak da adlandırılır.

Veri Ön İşleme, veri madenciliğinin başarısı için önemli bir aşamadır. Verinin sonraki aşamalarda kullanılması için elverişli ve düzgün biçimde işlenir. Bu aşamanın başarısı sonuçtaki başarıya doğrudan etki eder. Ön işleme aşamasının başarılı geçmesiyle net ve kesin sonuçlara ulaşılabilir.

Veri Dönüştürme, veri kümesinden yararlı ve doğru sonuca ulaşmak verinin dönüştürülme aşaması uygulanmalıdır. Verinin büyük bir parçası, ön işleme aşamasından geçse de ileriki aşamalarda kullanılacak biçimde değildir. Bu nedenle kullanılabilecek biçime dönüştürülmelidir.

Veri Madenciliği, veri madenciliği sürecinin bütün olarak çalıştığı aşamadır. Bu aşamada veri doğru ve kullanılabilir durumdadır. Çalışmadaki amaca göre bu aşamaya

gelindiğinde veri madenciliği metotlarından biri ya da birkaçı veri kümesine uygulanır. Gereken vaziyetlerde farklı metotlar bir araya getirilerek uygulanabilir.

Yorumlama ve Değerlendirme, veri kümesi üzerinde veri madenciliği aşaması tamamlandıktan sonra elde edilen sonuçlar yorumlanır ve çalışmanın doğru sonuca ulaşp ulaşmadığı kontrol edilir. Genel olarak bu aşamada farklı teknikler uygulanmışsa onlar arasında karşılaştırma yapılır. Bundan sonra sonuçlar daha önce yapılmış olan diğer çalışmalardan elde edilen sonuçlarla karşılaştırılır ve değerlendirilir (İnt.Kyn.10).

2.4 Veri Madenciliğinde Kullanılan Yöntemler ve Teknikler

2.4.1 Veri Madenciliği Yöntemleri

Veri madenciliği, yazılım tekniklerini temel alarak veriler üzerinde analiz yapılmasıyla ilgilenir. Analizin yapılması için de keşfetmeye dayalı yöntem ya da doğrulamaya dayalı yöntem gereklidir. Yöntemler ise, veriler üzerinden bilgiye ulaşılmasında ne tür yol takip edileceğini bildirmektedir.

Keşfetmeye dayalı yöntem, veri tabanları içerisindeki verilerin dönüştürülerek yeni bilgilerin üretilmesini mümkün kılar. Dolayısıyla yeni bilgilerin tanımlanması ya da tahmin edilmesini sağlar. Bu bağlamda tahmin edici “predictiv” ve tanımlayıcı “descriptive” olarak incelenir. Doğrulamaya dayalı yöntem ise, yeni bilgiyi üretmeye uğraşmaz ve geleneksel veri tabanlarıyla ilgilenir. Genellikle basit istatistiksel analizler kullanılarak, sorgulara ve raporlamaya dayalı bir yöntem uygular. Bu yöntem, bir varsayımın veri tabanları içerisindeki ilişkiler yardımıyla desteklenmesidir.

Tahmin edici yöntemler elde olan verileri kullanıp, bilinmeyen bir değer tahmin edilmesinde kullanılırlar. Bu yöntemlerin amacı, bazı değişken davranışlarının gelecekte durumunu bulabilmek için veriler üzerindeki desenlerin belirlenmesine odaklanır. Sınıflama “classification”, regresyon “regression” ve zaman serileri “time series” gibi yöntemler ile gelecekteki durumu tahmin edilebilir. Bir diğer yöntem, tanımlayıcı yöntemler ise veriler içerisindeki saklı ortak özellikleri ve bunlar arasındaki ilişkileri

araştırır. Kümeleme “clustering”, özetleme “summarization” ve birliktelik kuralı “association rule mining” gibi yöntemler ile de saklı ortak özellikler ve ilişkiler tanımlanabilir. Her iki yöntem veri madenciliği teknikleriyle desteklenir (Akpınar 2000, Anane 2001, Bounsaythip *et al.* 2001).

2.4.1.1 Tahmin Edici Yöntemler

Veri yığınlarının kullanılıp bir model geliştirilmesinde ve seçilen modelinde kullanılıp sonuçlarının daha önce bilinmeyen veri yığınları hakkında sonuçların tahmin edilmesini amaçlar. Tahmin edici modellerin en önemli özelliği karar alma aşamasında rol almasıdır. Örneğin bir bankada, müşterilerinin geçmiş dönemlerinde aldıkları kredilerinin verileri tutulmaktadır. Verilerle uyumlu şekilde seçilen model ile müşterilerin gelecekteki kredi talepleri üzerinden verilebilecek olan kredinin geri ödenip ödenemeyeceğinde tahmin yürütülebilir. Tahmin edici modeller eğri uydurma ve sınıflandırma olarak ikiye ayrılır. Tahmin edici modeller olarak en yaygın kullanılan model ise sınıflandırma modelidir (Han ve Kamber 2006, Arslan 2008, Onat 2008, Seven 2009).

Sınıflandırma Yöntemi

Sınıflandırma, veri madenciliğinde yaygın kullanılan bir yöntemdir. Verinin niteliğini inceler ve veriyi belirlenmiş bir sınıfa atar. Önemli bir nokta, her bir sınıfın özellikleri daha önce belirlenmiştir. Belirlenmiş sınıflar, veri tabanından yada veri ambarından çekilen verilerin sınıflandırılması işlemi yardımıyla yöntem geliştirme işleminde kullanılır. Bir değişken istenildiğinde bağımlı değişken şeklinde atanır, diğerleri ise bağımsızdır. Bağımsız değişkenler tahmin edici değişkenlerdir. Bu yöntem, tahmin edici değişkenin bulunduğu model üzerinde, bağımlı değişken değerinin yer aldığı anlamlı bir yöntem kurmayı amaçlar. Sınıflandırmadan sonra kayıtlar içerisindeki kişilerin, nesnelerin, kurumların, alanların niteliklerini ortaya çıkartması bu yöntemin en önemli işlevidir. Bu sınıflar belirli kullanıcı veya müşteri grupları için oluşturulmuşsa, burada öncelikli amaç her grubun profilini çıkarmak olmalıdır (Goebel and Gruenwald 1999, Ganti *et al.* 1999, Bergeron 2002).

Regresyon Analizi

Bir bağımlı değişkenin, birden fazla bağımsız değişken ile ilişki kurarak bir fonksiyon halinde ifade edilmesi ve bu fonksiyon sayesinde bağımlı değişkenin varabileceği değerlerin tahmin edilme çalışmasıdır. Tahmin edilen bağımlı değişkenin süreklilik gösteren ya da kategorik olan bir değerinin olması sınıflandırma ve regresyon modelleri arasındaki temel farktır. Eğer tahmin edilecek alan sürekli değişkense bu regresyon, kategorikse de sınıflandırma problemidir. Her ikisinin de kullanıldığı teknikler benzer olduğundan birbirlerine yakındırlar (Orhunbilge 1996, Fayyad and Stolorz 1997, Thuarisingham 2003).

2.4.1.2 Tanımlayıcı Yöntemler

Karar verme sürecinde tanımlayıcı yöntemler rehberlik edilmede kullanılabilecek veri örüntülerini tanımlamada kullanılır. Genel olarak sepet analizlerinde, alışveriş sepeti içerisindeki ürünler arasındaki ilişkinin bulunması ve elektronik alışveriş sitesi üzerinde ürününün seçildiği zaman müşteriye o ürüne yakın başka ürünlerin de önerilmesi tanımlayıcı yöntemlerinin veri madenciliğinde kullanıldığının birer örneğidir. Kümeleme, özetleme, birliktelik kuralları ve sıra örüntüleri en yaygın kullanılan tanımlayıcı yöntemleridir (Dunham 2003, Aynekin 2006, Bozkır 2009, Erten 2015).

Kümeleme Yöntemi

Veri tabanlarında büyük hacimli verilerin var olmasının yanında karmaşık yapıda verilerde bulunabilir. Karmaşık verilerde veri madenciliği tekniklerinin anlamlı sonuçlar elde etmesi zor olacaktır. Bu zorluk için izlenilebilecek yöntem verileri parçalamak, alt kümelere bölmek ve ilk önce alt bölümleri kullanarak çözüm üretmektir. Verilerin bölümlenebileceği ve gruplara ayrılabilmesinin nasıl yapılacağını saptamak amacı ile birtakım yöntemler geliştirilmiştir. Kümeleme yöntemi, bölümleme ve demetleme olarak da bilinmektedir. Bir diğer tanımıyla, birbirine benzeyen verilerin bir araya getirilerek homojen yapıda alt kümeler oluşturulması ve öngörülen bölümde heterojen halde olan verilerin belirlenmesi işlemi kümeleme olarak adlandırılır. Bu yöntemin amacı, karmaşık

haldeki veride değişkenler arasındaki benzerliğe ya da uzaklıklarına bağlı şekilde gruplanarak işlenebilecek hale getirilmesidir. Kümelemenin sınıflandırmadaki gibi daha önce tanımlanan sınıflara göre değil de, veride gizli bulunan fakat daha önceden tanımlanan sınıflara göre ayırım yapması sınıflandırma ve kümeleme yöntemleri arasındaki farktır. Dolayısıyla sınıflandırma tahmin edici, kümeleme tanımlayıcı bir yöntemdir. Kümeleme analizi, bunlardan ayrı olarak tıp, biyoloji, istatistik, psikoloji, sosyoloji, arkeoloji gibi birçok alanda kullanılmaktadır (Han and Kamber 2001, Getoor 2003, Vatansever ve Büyüklü 2009).

Birliktelik Kuralları

En yaygın olarak kullanılan veri madenciliği yöntemi birliktelik kurallarıdır. Veri yapıları arasındaki ilişkileri biçim olarak tanımlamaya yarayan bir yöntemdir. Değişkenler arasındaki ilişkinin büyüklüğü ve yakınlığı saptanmaya çalışılır. Başka bir deyişle, verinin potansiyel ilişkilerini tanımlar. Veri tabanlarının hacmi ile birliktelik kuralı yönteminden başarılı bir sonuç elde etmek birbiri ile doğru orantılıdır. Bu da birliktelik kurallarını büyük veri tabanlarında bulmanın zor olmadığını gösterir. Buna rağmen uzak ya da yakın ilişkili birden fazla birliktelik kuralının ortaya çıkması bir zorluk çıkartacaktır. Ortaya çıkartılacak birliktelik kurallarının sayısı sonsuz olacaktır. Verimli bir kural olabilmesi için en önce bir eşik değeri bulunmalıdır. Eksik, kayıp, gürültülü verilerden değerli, gereken veriyi ayırmak ve bu eşik değerinin bulabilmesi zor bir iştir. Dolayısıyla ilişkileri tanımlayabilmek kullanılan bazı ölçütlerin belirlenmesine bağlıdır. İstatistiğin sayısal kavramı destek ve güven değeri bu ölçütlerdir. Güçlü bir kuralın yüksek seviyeli bir güvenliği ve çok büyük desteği sahiptir. Büyük miktarda yatırımların yapıldığı özel bir veri madenciliği yöntemidir. Mühendislik, fen ve sağlık, ticaret sektörlerinde ve daha birçok alanda kullanılmaktadır (Agrawal *et al.* 1993, Adriaans and Zantinge 1996, Dunham 2003).

2.4.2 Veri Madenciliği Teknikleri

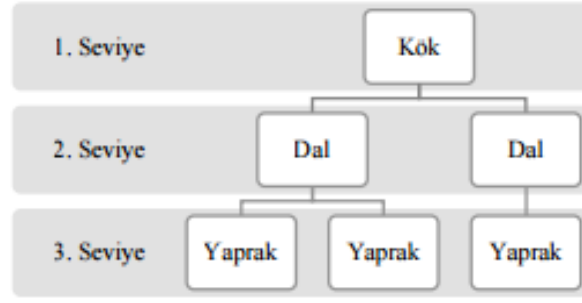
Veri madenciliği teknikleri denetimli modelleme ve denetimsiz modelleme şeklinde iki ana kategoriden oluşur. Bu modeller üzerinde iyi tanımlanmış ya da kesin bir hedef olduğunda denetimli modelleme olarak ifade edilir. Sonuca göre özel bir tanımlama

yapılmamışsa veya belirsizlik söz konusu ise denetimsiz modelleme ile ifade edilir. Denetimli veri madenciliği teknikleri: En yakın k komşuluk, K-ortalamalar kümeleme, Regresyon analizi, Karar ağaçları ve Sinir ağları tekniklerinden oluşur. Denetimsiz veri madenciliği teknikleri ise: Aşamalı kümeleme, Kendi kendini düzenleyen haritalar tekniklerinden oluşur. Veri madenciliğinin klasik teknikleri ağırlıklı olarak istatistiksel yöntemlerdir. Bu yöntemler de istatistiği temel alan ama daha çok makine öğrenimi ve yapay zekâ destekli yeni nesil yöntemlerdir. Bunlar: Regresyon, K - En Yakın Komşuluk ve Kümeleme yöntemleridir. Diğer Veri Madenciliği yöntemleri şu şekildedir: Temel Bileşenler Analizi, Diskriminant Analizi, Faktör Analizi, Kohonen Ağları, Bulanık Mantığa Dayalı Yöntemler, Genetik Algoritmalar, Bayesci Ağlar ve Pürüzlü Küme Teorisine Dayalı yöntemlerdir (Tsipris and Chorianopoulos 2009)

Kısaca, bilgi keşfinde kullanılan tüm teknikler veri madenciliğinin bir yöntemidir. Yaygın olarak kullanılan denetimli ve denetimsiz modelleme kullanılan veri madenciliği yöntemleri açıklamalarıyla aşağıda verilmiştir.

2.4.2.1 Karar Ağaçları

Veri madenciliği tekniklerinden karar ağaçları, kolay anlaşılabilen kural üretebilmesi ve kuralları görselleştirilebilen yapısı nedeniyle en sık kullanılan ve etkili bir sınıflandırma yöntemidir. İlk nüfusu yinelemeli olarak bölerek çalışır. Her bölme için, otomatik olarak en önemli belirleyiciyi, hedef alana göre en iyi ayrımı sağlayan belirleyiciyi seçerler. Amaçları ardışık bölümler aracılığıyla çıktı açısından homojen davranışa sahip saf alt segmentler üretmektir. Belki de en popüler sınıflandırma tekniğidir. Popüleritelerinin bir kısmı, incelenen etkinliğe bir bakış açısı sunarak kolayca yorumlanabilir şeffaf sonuçlar üretmeleridir. Üretilen sonuçlar iki eşdeğer biçime sahip olabilir. Bir ağaç biçiminde, kurallar grafiksel olarak ilk popülasyon (kök düğümü) ardışık olarak terminal düğümlerine veya hedef alana ilişkin benzer davranışa sahip alt segmentlerin yapraklarına bölünen bir ağaç olarak temsil edilir. Karar ağacı algoritmaları hız ve ölçeklenebilirlik sağlar. C5.0, CHAID, Classification and Regression Trees, CHAID ve QUEST kullanılan algoritmalar (Tsipris and Chorianopoulos 2009, Altaş ve Gülpınar 2012). Şekil 2.4’de bir karar ağacının yapısı gösterilmektedir.



Şekil 2.4 Karar Ağacının Yapısına Örnek (Daş ve Türkoğlu 2014).

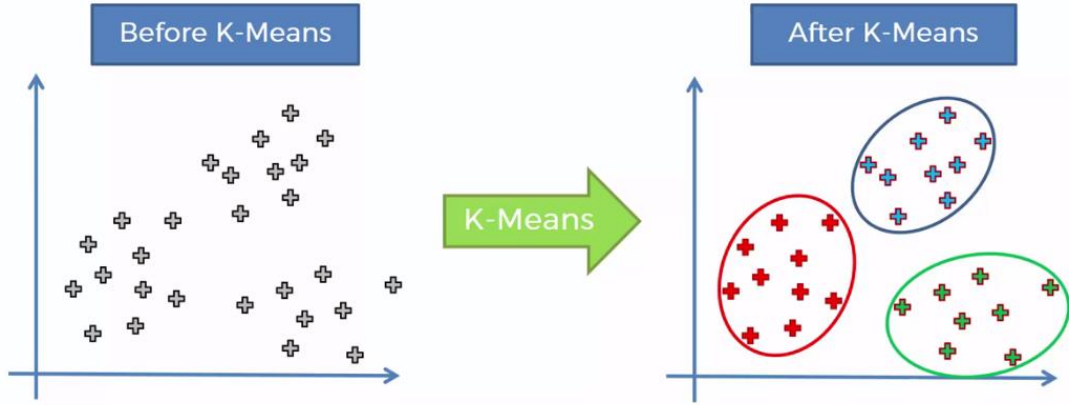
2.4.2.2 K-means

Hem büyük hacimli (birçok kayıt) hem de geniş veri kümelerini (birçok veri boyutu ve giriş alanı) işleyebilen verimli ve belki de en hızlı kümeleme algoritmasıdır. Mesafe tabanlı bir kümeleme tekniğidir ve hiyerarşik algoritmanın aksine, tüm kayıt çiftleri arasındaki mesafeleri hesaplamaya gerek duymaz. Oluşturulacak kümelerin sayısı önceden belirlenir ve kullanıcı tarafından önceden belirtilir. Genellikle en uygun onaylamadan önce bir dizi farklı çözüm denenmeli ve değerlendirilmelidir. Sürekli kümeleme alanlarını işlemek için en iyisidir (Tsitsis and Chorianopoulos 2009).

K-means amacı, küme içindeki toplam varyans yada aşağıda verilen denklem ile hesaplanan kare hata fonksiyonunun sonucunu minimuma indirmektir.

$$J = \sum_{j=1}^k \sum_{i=1}^n \|X_i^{(j)} - C_j\|^2 \quad (2.1)$$

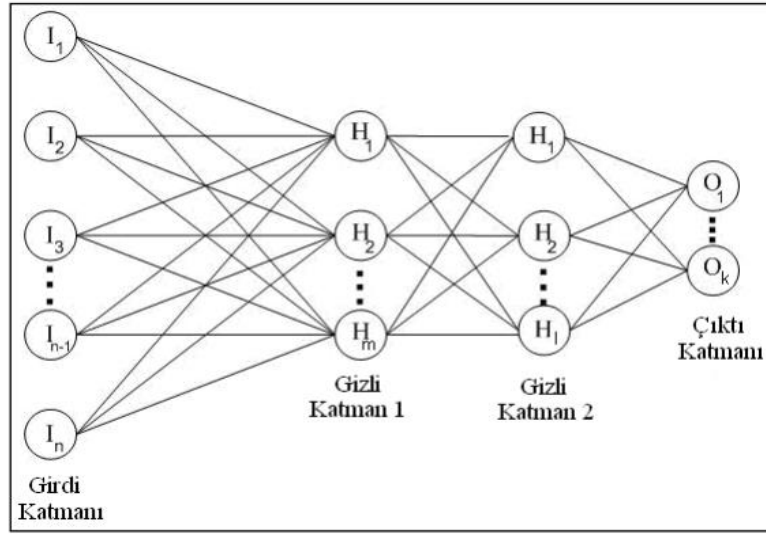
K-means algoritmasının adımları Şekil 2.5’de gösterilmektedir.



Şekil 2.5 K-means Adımları (İnt.Kyn.11).

2.4.2.3 Yapay Sinir Ağları

Teknoloji, insanlığın doğayı araştırma çabasının en son ürünüdür. İnsanlar tarafından gerçekleştirilen, çevreden gelen olaylara karşı nasıl tepkiler üretileceğini belirleyebilen ve olayları öğrenebilen bilgisayar sistemleridir. Sinir Ağları, tahmin ve sınıflandırma için karmaşık, doğrusal olmayan haritalama işlevlerini kullanan güçlü makine öğrenme algoritmalarıdır. Katmanlar halinde düzenlenmiş nöronlardan oluşurlar. Giriş katmanı belirleyicileri veya giriş nöronlarını içerirler. Çıktı katmanı hedef alanı içerirler. Bu modeller, tahmin edicileri (giriş katmanı) çıktıya bağlayan ağırlıkları tahmin ederler. Daha karmaşık topolojilere sahip modeller, ara, gizli katmanlar ve nöronları da içerebilirler. Eğitim prosedürü yinelemeli bir süreçtir. Giriş kayıtları, bilinen sonuçlarla birlikte, ağı sunulur ve gözlenen sonuçlara göre model tahmini değerlendirilir. Gözlemlenen hatalar, ilk ağırlık tahminlerini ayarlamak ve optimize etmek için kullanılır. Tahminlerinin bir açıklamasını sağlamadıkları için opak veya kara kutu çözümleri olarak kabul edilirler. Sadece giriş alanlarının tahmini önemini özetleyen bir duyarlılık analizi sağlarlar. Asgari istatistiksel bilgi gerektirir, ancak soruna bağlı olarak eğitim için uzun bir işlem süresi gerektirebilir (Tsiptsis and Chorianopoulos 2009, Altaş ve Gülpınar 2012). Sinir ağları katmanlardan oluşurlar, her katmanındaki nöronlarda bir diğer katmanın nöronlarına bağlı şeklindedir (Şekil 2.6).



Şekil 2.6 Birimlerin Arasındaki Bağların Gösterildiği Saklı Katmanlı Sinir Ağı (Karataş 2004).

2.4.2.4 Genetik Algoritmalar

Canlıların genetik yapısı üzerinden tasarlanarak geliştirilen bilgisayar teknolojilerinin ilerlemesiyle etkili eniyileme “optimizasyon” yöntemleri ortaya çıkmıştır. Bu yöntemler mühendislik, teknik, tıp ve siyaset bilimi vb. pek çok alanda başka arama tekniklerinin yerine genel bir çözüm olarak tanınmıştır. Genetik algoritmalar, etkili ve kesin sonuçlara varılmaya çalışılan rastgele arama metotlarıdır. Karmaşık olan problemlerin çözümü için geliştirilmiş, kromozomların üretilmesine dayanan sezgisel araştırma yöntemidir. Temel aldığı durum biyolojik evrimdir. En iyi kromozoma işlem sonucunda ulaşmaya çalışır. En iyi kromozoma vardığıdaysa genleri yer değiştirerek yeni sonuçlar üretir (Özçakar 1998, Saruhan 2004).

2.4.2.5 Veri Görselleştirme

Verilerin birbirleri arasındaki ilişkisinin belirlenmesi veri madenciliği uygulamalarında önemlidir. Geleneksel yöntemlerin sonucu ile elde edilen karmaşık ilişkilerin kolay algılanabilmesi ve rahatça anlaşılabilmesi gerekir. Verinin algılanabilirliğinin artırılması için bilgisayar desteğini de alarak görsel olarak temsil edilmesi veri görselleştirme olarak ifade edilir. Veri görselleştirme tekniği, veri içerisinden yeni yorumlanabilir örüntüler elde etmek için veriyi alışlagelmiş istatistiksel ve sayısal formdan soyutlayıp grafik arayüzleri ile oluşturur (Vatansever ve Büyüklü 2009).

2.5 Weka

Makine öğrenimi için WEKA, Waikato Üniversitesi tarafından geliştirilmiştir. "Waikato Environment for Knowledge Analysis" sözcüklerinin ilk harfleri kullanılarak oluşturulmuş programın adıdır. Yaygın olarak kullanılan birçok makine öğrenimi yöntemlerini ve algoritmalarını içerir.

Java platformunda geliştirilmesi, kütüphanelerin “.jar” uzantılı dosya biçiminde gelmesi ve Java dili üzerinde hazırlanan projelere kolay bir şekilde entegrasyonu sağlandığından kullanımı çoğalmıştır. Java diliyle yazılmış olan Weka, Unix, Linux, Macintosh ve Windows vb. işletim sistemlerinde kullanılabilir. Aynı zamanda dağıtık veya büyük hacimli veri tabanları üzerinde kullanılır. Ayrıca veri hazırlama süreci, sınıflandırma ve kümeleme algoritmaları, birliktelik analizleri, öznitelik değer seçimi işlemlerinin yapılabilmesini sağlamaktadır.

En önemli ve güçlü özelliği olarak çok sayıda sınıflandırma metotlarını içerisinde barındırmasıdır. Dolayısıyla sınıflandırma yöntemleri arasında karşılaştırma yapabilme imkânı sunmaktadır. Aynı zamanda uygulamaların komut girilerek gerçekleştirilmesine de imkân tanır. WEKA programı arff, csv, c45 dosya biçimlerini desteklemektedir. (İnt.Kyn.12).

2.5.1 Genel Bilgi ve Kullanımı

Veri madenciliği metotlarının tümü Weka içerisinde bulunmaktadır. Bunlar ana başlıklar olarak ön işleme, sınıflandırma, kümeleme, birliktelik kuralları, nitelik seçme ve görselleştirme metotlarıdır. ARFF dosya formatını giriş olarak kabul etmektedir. Ayrıca istatistiksel ve makine öğrenme bazındaki kütüphanelerde içerisinde mevcuttur.

Arff'yi tanımlarsak; üzerinde değişken tanımlaması sunan ve makine öğrenmesi kullanılabilecek ASCII formatlı metin dosyasıdır. “@relation” veri setindeki genel amacı ya da ismi belirtir, “@attribute” verilerin özelliklerini belirtir ve “@data” ham verileri bulundurur. Ayrıca harf duyarlılığı da bulunmaktadır. Şekil 2.7’de “@relation” hava durumu, “@attribute” sıcaklık vb., “@data” güneşli gibi verileri barından arff dosya içeriği örnek olarak verilmiştir.

```
@relation weather.symbolic

@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
```

Şekil 2.7 Hava Durumu İçin Örnek Arff Dosyası İçeriği

2.5.2 Weka Programı Modüllerinin Tanımları

Weka’da veri kümesi öğrenme metotları uygulanabilir ve verinin içeriğinden bilgi edinmek amacıyla çıkış incelenebilir. Tahminler yapmak için öğrenilmiş modeller yeni örnekler üzerinde kullanılabilir. Tahminler arasında karşılaştırma yapmak ve değişik algoritmaları kullanarak öğrenmeyi kolaylaştırmak amacıyla kullanılmaktadır.

İşler ve Narin (2012)’a göre WEKA programında sırasıyla explorer, experimenter, knowledge flow ve simple cli olarak 4 uygulama vardır. Bunlar aşağıda sırasıyla verilmiştir.

2.5.2.1 Explorer

Weka’nın kullanıcılar için tasarladığı ara yüz kısmıdır. İçerisinde sırasıyla ön işlem “*Preprocess*”, sınıflandırma “*Classify*”, kümeleme “*Cluster*”, ilişkilendirme “*Associate*”, öznitelik seçimi “*Select Attributes*” ve görselleştirme “*Visualize*” kısımlarından oluşmaktadır.

“Preprocess” sekmesi, veri setinin eklenmesi, veri filtrelemesi işlemi, özniteliklerin seçilmesi ve istatistikler vb. seçenek bulunmaktadır. Weka’nın desteklediği dosya biçimleri CSV, ARFF, C4.5 ve LibSVM’dir. Veri setine uygulanacak yöntemlere verilerin nominal değere dönüştürülmesi, yeniden örnekleme ve dalgacık analizi, ayrık değerlerin ayrıştırılması gibi birden fazla örnek verilebilir.

“Classify” sekmesi, birden çok farklı sınıflandırma algoritma seçeneklerin sunulduğu kullanıcılar için tasarlanan ara yüz ekranıdır. *“Classifier”* başlığında *“BayesNet”*, *“SMO”*, *“LibSVM”*, *“IBk”*, *“AdaBoostM1”*, *“OneR”*, *“ZeroR”* ve *“J48”* gibi birçok sınıflandırma algoritmasından birisi seçilebilmektedir. Test seçenekleri başlığı altında eğitim seti oluşturulması, başka bir veri seti seçilmesi, parça ayrıştırması ve belirli bir yüzdenin test edilmesi gibi uygulamalara fırsat veren seçenekleri bulundurmaktadır. Sınıflandırmanın verilerdeki performans raporlamasını sonuç listesi ekranı üzerinde işleme giren veri setinin ismi, örneklem sayısı, öznitelik sayısı, doğru ve yanlış biçimde sınıflandırılma işlemi gören örnek sayısı, doğruluk tablosu, ortalama mutlak hata gibi sonuçlar gösterilir. Sınıflandırma sekmesinin çıkış bölümü ise çalışma bilgilerini, detaylı doğruluk tablosunu, hata matrisini ve sınıflandırmanın özet bilgilerini gösterir.

“Cluster” sekmesi, kümelendirme algoritmalarının seçenekler halinde verildiği ara yüzü ekranıdır. Burada *“EM”*, *“SimpleKMeans”*, *“OPTICS”*, *“HierarchicalClusterer”* algoritmaların birisi seçilebilmektedir. Kümelendirme algoritmaları seçerken sınıflandırma sekmesinde de önceden verilen bilgi gibi bazı ayarlar parametreler üzerinden değiştirilebilir. Ayrıca algoritmalar konusunda detay içeren bilgiler alınır. *“Clusterer mode”* başlığında sınıflandırma sekmesindeki seçeneklere ek olarak sınıf bilgisi alınarak kümeler üzerindeki değerlendirilme işleminin *“Classes to clusters evaluation”* test için kullanılmasını sağlayacak seçenekler vardır. Bunlardan farklı olarak öz niteliklerin işleme alınmaması için seçilmesini *“Ignore attributes”* ve bununla beraber geriye kalan öz niteliklerin başarımının ne oranda etkilendiğinin gözlemlenmesini sağlamaktadır.

“Associate” sekmesi, birliktelik kuralları uygulanması ile alakalı komutların uygulanabildiği kullanım ara yüz ekranıdır. “Associator” ile birliktelik kuralı seçimi ve seçilen kurala göre ayarlar yapılabilir. “Associator output”, “Result list” kısımlarında çalıştırılan algoritmanın sonuçlarını göstermektedir.

“Select Attributes” sekmesi, öznitelik seçimi yapılabilir. “Attribute Evaluator” ve “Search Method” metotları seçilir. İlk olarak öz niteliklerin alt kümelerindeki uygunluk ölçütünü belirlemek için uygulanacak değerlendirme metodunu ve diğer öz niteliklerin arasında ne şekilde arama sağlanacağını ayarlar. “Attribute selection mode” başlığında öz niteliklerin değerlendirilmesi eğitim veri kümesi veya çapraz doğrulama metodu yardımıyla seçilenler ayarlanır. Sonuçlar seçilmiş öz niteliklerin listesini “Attribute selection output” başlığında görüntüler.

“Visualize” seçeneği, veri setinde örneklerin öznitelikler ışığında ne şekilde dağıtıldığını iki boyutlu olarak görüntüler. Burada öz nitelik seçimi yapıp görsel inceleme yapılması da sağlanabilir.

2.5.2.2 Experimenter

Çok sık geliştirilen bir algoritmanın birçok veri setindeki başarımlarının gösterilme işlemi ya da birçok algoritmanın aynı veri setindeki başarımlarının karşılaştırılması biçimindeki uygulamalar Weka programının “Experimenter” seçeneğinde gerçekleştirilebilmektedir. Burada deneyler için uygulanması beklenen veri tabanı, uygulanması beklenen algoritmaların seçildiği “Setup”, deneylerin sürdürülebildiği “Run” ve sürdürülen uygulamanın sonuçların incelenebileceği “Analyse” sekmesi vardır.

2.5.2.3 Knowledge Flow

“Explorer” seçeneğinde yapılan işlemlerin görsel olarak sürükle-bırak mantığı ile ayarlanabilmesi amacıyla geliştirilmiş çalışma ortamına “Knowledge Flow” denir. “Explorer” da yapılamayan daha fazla seçeneği burada bulmak mümkündür. Örnek olarak “Explorer” bütün bir şekilde veri setini işleyebilirken, “Knowledge Flow” ayrı

olarak verileri kademeli işleme özelliği vardır. “Explorer” çalışmalarda büyük veriler için sıkıntılar yaşatırken bu durumu kademeli olarak yaparak “Knowledge Flow” kolaylaştırmaktadır. Burada “DataSources”, “DataSinks”, “Filters”, “Classifiers”, “Clusterers”, “Evaluation” ve “Visualization” bölümleri vardır. Çalışma ile uyumlu ara işlemlerin sürükle-bırak metodu ile “Knowledge Flow Layout” uygulamasına bırakıp işlemlerin tamamlanmasından sonra “Arff Loader” a sağ tıklanarak çalışma sürdürülür.

2.5.2.4 Simple CLI

Komut vererek çalışmaların yapıldığı bölümdür.

2.6. İlgili Araştırmalar

Cano vd. (2019) tarafından yapılan araştırmada, akademik kurumlar tarafından veri madenciliği araçları sadece öğrenci kayıt veya kaynak tahsisi gibi sadece eğitim süreç yönetiminde kullanılmamış, aynı zamanda literatür tercihi, rotasyonu ve kullanımı gibi bilgi sağlayan yüksek kaliteli veri toplama bilgi teknoloji uygulamalarını da kullanmıştır. Bu teknoloji, Sekiz İngiltere üniversitesinde ve Honk Kong Baptist Üniversitesinde yapılan önceki bir çalışmaya ek olarak akademik performansa pozitif katkı sağlayan kütüphane kullanımını artırmak için kullanılabilir. Bu çalışmanın amacı, üniversite kütüphanelerinin veri tabanlarından bilgi edinmek ve kullanıcılar adına daha iyi kararlar almak için bilgi veri keşfine dayalı yararlı bir metodoloji sağlamaktır. Bu şekilde kütüphaneciler, literatürün tahsisine, konu başına teklif miktarına ve talebi karşılamak için gereken insan kaynaklarına karar verebileceklerdir. Öncelikle ana tekniklerden bahsedilmiş ve önceden tanımlanmış tekniklerle yeni yöntem üretilmiştir. Son olarak, bilgi veri keşfi metodolojisi Perulu bir üniversite kütüphanesinin örnek çalışmasına uygulanmıştır. Araştırma yapmak için kullanılan veri kümesi, bir akademik dönemden yakalanan yarı yapılandırılmış bir kümedir. Veriler dört kaynaktan oluşmaktadır: literatür ödünç kayıtları, literatür iade kayıtları, kullanıcı sorguları ve literatür kullanılabilirliğidir. Edebiyat ders kitapları, dergi makaleleri, gazeteler, dergiler, raporlar ve görsel-işitsel verilerden oluşmaktadır. Araştırma sonunda kümeleme algoritmaları yardımıyla belirli profiller tanımlanarak, kullanıcıların karakterizasyonunun elde edilebileceği önerisi sunulmuştur. Aynı şekilde birliktelik kuralları yardımıyla kullanıcı davranışlarının ilişkisini gösterebileceği belirtilmiştir.

Luo (2017) çalışmasında, öğretmen ve öğrencilerin kitaplarda ne aradıklarını bulmalarını kolaylaştırmak için kütüphaneden DBSCAN algoritmasıyla veri toplamıştır. DBSCAN algoritmasının uygulandığı model ilk olarak kütüphane veri madenciliğini önermiş ve daha sonra bu konu üzerine geliştirmeler yapmıştır. Bu algoritma kullanılarak; hedeflenen kitapların önerilebilmesi için, kullanıcıların ilgi alanlarını ve hobilerini, profesyonel geçmişlerini ve diğer profillerini çıkarmak için veri örnekleri kümesini analiz etmiştir. Aynı zamanda kitap fiyatı, stok seviyesi, kitapların sınıflandırılması, kitap alım borçlarının sıklığı gibi bilgiler kriter olarak kullanmıştır. Bu çalışmada, veri madenciliği için birliktelik kuralları analizi ve sınıflandırma analizi gibi diğer analiz yöntemlerinin yanı sıra küme analizini de kullanmıştır.

Demiral vd. (2017) tarafından yapılan çalışmada, veri madenciliği yardımı ile verileri toplama ve anlamlandırma aşamalarına bir örneklem sunulmuştur. Kümeleme yöntemlerinden birisi olan hiyerarşik kümeleme yöntemi kullanılmıştır. 2007-2017 tarihine ait Uşak Üniversitesi Merkez Kütüphanesi verilerinde uygulama yapılmış ve öğrencilerin kitapları ödünç alma davranışları üzerine çeşitli istatistiki bilgiler elde edilmiştir. Veriler üzerinde ayırıcı hiyerarşik kümeleme tekniği ile 11 adet küme oluşturulmuştur. Kümeler kendi içinde homojen, kümeler arasında ise heterojen şeklinde ayarlanmıştır. Elde edilen istatistiksel veriler ile veri madenciliği yöntemlerinden görselleştirme uygulanarak birbirinden farklı grafikler ve tablolar oluşturulmuş; bunlardan elde edilen çıkarımlar ile kitap ödünç alma davranışları incelenmiştir.

Durmuşoğlu (2017) çalışmasında son 10 yıldaki dergilerde “Science Citation Index-SCI, Science Citation Index-expanded, Social Sciences Citation Index-SSCI ve Arts and Humanities Citation Index-AHCI” ve konferanslarda “Conference Proceedings Citation Index” yayınlanan veri madenciliği çalışmalarını ele almış ve ülke, yazar, yöntem vb. dağılımını ortaya koymuştur. Buna ek olarak, Türkiye adresli veri madenciliği çalışmaları ayrı bir başlık altında ele alınmış ve çalışmaların nitelik ve nicelikleri incelenmiştir.

Doğan (2017) çalışmasında, Türkiye’deki üniversitelerde erişimi bulunan ve “veri madenciliği” anahtar kelimesi ile ulaşılabilen yüksek lisans ve doktora tezleri farklı açılardan incelenmiştir. Çalışmada, sosyal bilimler alanında veri madenciliği sorgusu ile 116 lisansüstü tez belirlenmiştir. Bu tezlerin 22’si araştırmaya dahil edilmemiştir.

Erişilmiş 94 adet tezdten “tez tipi yüksek lisans/doktora, yazım dillerinin ne olduđu, hazırlandıkları üniversiteler ve anabilim dalları, uygulandıkları alanlar” incelenmiştir. Ayrıca tezlerde bir uygulama varsa bu uygulamanın hangi veri madenciliđi algoritmaları ve yazılımları ile yapıldığı gibi kriterler incelenmiştir. Buna ek olarak, tezler yıllara göre alan ve uygulama alanlarındaki deđişiklikler açısından incelenmiştir.

Siguenza-Guzman vd. (2015) tarafından yapılan çalışmada, akademik kütüphanelere uygulanan veri madenciliđi teknikleri için kapsamlı bir literatür taraması ve sınıflandırma yöntemi kullanılmıştır. Bunun için, 1998-2014 dönemi boyunca 41 pratik katkı tespit edilmiş ve bunlarla doğrudan ilgisi olan nedenler gözden geçirilmiştir. Her makale veri madenciliđi yöntemlerinden kümeleme, birliktelik kuralları, sınıflandırma ve regresyona göre kategorize edilmiş ve bunlar hizmetler, kalite, koleksiyon ve kullanım davranışı şeklinde dört ana kütüphane alanına uygulanmıştır. Bulgular, hem toplama hem de kullanım davranışı analizlerinin, özellikle web sitelerinin ve çevrimiçi hizmetlerin koleksiyon geliştirme ve kullanılabilirliği ile ilgili araştırmaların çođunu aldığını göstermektedir. Ayrıca, sınıflandırma ve regresyon modelleri kütüphane ayarlarında uygulanan en sık kullanılan iki veri madenciliđi yöntemidir. Buna ek olarak; sonuçlar, akademik kütüphanelerde veri madenciliđi tekniklerinin uygulanması üzerine yayınlanan makalelerin ilk 6 dergisinin: Üniversite ve araştırma kütüphaneleri, akademik kütüphanecilik dergisi, bilgi işleme ve yönetimi, Hi Tech kütüphanesi, uluslararası bilgi, kültür ve deđişim yönetimi dergisi ve elektronik kütüphane olduğunu göstermektedir. Scopus, tanımlanan dergi makalelerinin en iyi kapsamını sağlayan çok disiplinli veri tabanıdır. Bu çalışma, akademik kütüphanelere uygulanan veri madenciliđi tekniklerinin ilk sistematik, tanımlanabilir ve kapsamlı akademik literatür incelemesini temsil etmektedir.

Chen (2014) çalışmasında üniversite kütüphanelerinin veri madenciliđi yeteneğinden yoksun olduğunu tespit etmiştir. Bu duruma dayanarak veri madenciliđi teknolojisine dayalı kişiselleştirilmiş bilgi sistemi elde etmek için kütüphane sistemi ile kombine model yapısı tasarlamıştır. Her bir modülün kompozisyonunu teorik bir bakış açısıyla, fonksiyonel açıklama ve iş akışının açıklamasından analiz etmiştir. Sonuç olarak; model seçimi, üniversite kütüphanelerinin geliştirilmesi için daha iyi bir teorik temel sağlamış, böylece üniversite kütüphanelerinin uygulama becerisini geliştirmiştir.

Zhang (2011) alışmasında kütüphane verileri üzerinde yaptığı veri madencilięi uygulamasını üç açıdan incelemiştir; uygulamanın anlamının ortaya çıkartılması, uygulamadaki temel problemlerin analiz edilmesi ve sonunda karar ağacı algoritmasının uygulanmasını araştırmıştır. Linyi Üniversitesi kütüphanesindeki düşük kullanım okuyucularının analizi ile birleştirmesi sonucu bir alışma örneęi sunmuştur.



3. MATERYAL ve METOT

3.1 Araştırmanın Modeli

Bu çalışma iki aşamadan oluşmaktadır. İlk aşamada gerekli resmi yazışmalar yapılarak, kişisel veriler içermeyecek şekilde kütüphane veri sistemi üzerinden veriler elde edilmiştir. İkinci aşamada ise bu veriler veri madenciliği uygulama adımlarına uygun bir şekilde veri madenciliği yöntemleri ile incelenmiştir.

3.2 Araştırmanın Evreni ve Örneklemi

Bu çalışmanın araştırma evreni 2009-2019 eğitim öğretim yılında Afyon Kocatepe Üniversitesi Merkez Kütüphanesinde araştırma yapan öğrencilerin kitap ödünç alma davranışlarını içeren veri seti oluşturmaktadır. Araştırma bir veri madenciliği çalışması olduğu için veriden örneklem alınmamış, tüm veriler veri madenciliği algoritmaları tarafından değerlendirilmiştir.

3.3 Verilerin Elde Edilmesi

Afyon Kocatepe Üniversitesi Merkez Kütüphanesi'nin 10 yıllık verileri, kütüphane yazılımından Microsoft Office Excel ortamına aktarılması ile elde edilmiştir. Kütüphane veri tabanından alınan bu veriler; Excel çalışma kitabında, 402918 kayıt ve Demirbaş, Sınıflama, Kitap Ad, Ödünç Tarihi, İade Tarihi, Üye, Cinsiyet, Bölüm, Enstitü, Fakülte sütunlarından oluşmaktadır (Şekil 3.1).

Demirbaş	Sınıflama	Kitap Ad	Ödünç Tarihi	İade Tarihi	Üye	Cinsiyet	Bölüm	Enstitü	Fakülte
0033308	29952	Düşünce ve dil / L.S. Vygotsky;	18.01.2010	19.01.2010	061016xxx	01	OKUL ÖNCESİ ÖĞRETMENLİĞİ		Eğitim Fakültesi
0011520	29952	Düşünce ve dil / L.S. Vygotsky;	04.05.2009	22.05.2009	071010xxx	02	Sosyal Bilgiler Öğretmenliği		Eğitim Fakültesi
0081776	29952	Düşünce ve dil / L.S. Vygotsky;	07.03.2011	17.03.2011	071010xxx	01	Sosyal Bilgiler Öğretmenliği		Eğitim Fakültesi
0022304	29952	Düşünce ve dil / L.S. Vygotsky;	21.04.2010	04.05.2010	090645xxx	02	Sınıf Öğretmenliği (İlköğretim)	Sosyal Bilimler Enstitüsü	
0064361	29952	Düşünce ve dil / L.S. Vygotsky;	04.05.2010	24.05.2010	090645xxx	02	Sınıf Öğretmenliği (İlköğretim)	Sosyal Bilimler Enstitüsü	
0059050	29952	Düşünce ve dil / L.S. Vygotsky;	13.03.2010	05.04.2010	093101xxx	02	Beden Eğitimi ve Spor Öğretmenliği		Beden Eğitimi ve Spor Yüksekokulu
0071894	29952	Düşünce ve dil / L.S. Vygotsky;	30.10.2013	14.11.2013	111013xxx	02	Sınıf Öğretmenliği		Eğitim Fakültesi
0083139	HT 166/	Arazi Topulaştırması ve Taşını	18.02.2009	06.04.2009	040302xxx	01	Metal Eğitimi		Teknik Eğitim Fakültesi
0053268	HT 166/	Arazi Topulaştırması ve Taşını	08.10.2009	23.10.2009	062902xxx	01	Jeodezi ve Fotogrametri Mühendisliği		Mühendislik Fakültesi
0059478	HT 166/	Arazi Topulaştırması ve Taşını	07.02.2013	09.05.2013	092913xxx	01	Harita Mühendisliği (İÖ)		Mühendislik Fakültesi
0020056	RD 93/	A Acil Cerrahi / ed. Cemalettin E	18.05.2017	24.05.2017	163366xxx	02	Cerrahi Hastalıkları Hemşireliği (İÖ) (YL Sağlık Bilimleri Enstitüsü)		
0082234	AC 140/	Havra Sokağı / Orhan Kaya Çet	19.02.2013	02.04.2013	091018xxx	01	Bilgisayar ve Öğretim Teknolojileri Öğretmenliği		Eğitim Fakültesi
0045977	AC 140/	Havra Sokağı / Orhan Kaya Çet	06.10.2010	11.10.2010	100112xxx	02	Kimya		Fen Edebiyat Fakültesi
0020038	AC 140/	14 Eylül yayınları 4 Eylül ortak k	04.03.2013	25.03.2013	091018xxx	01	Bilgisayar ve Öğretim Teknolojileri Öğretmenliği		Eğitim Fakültesi
0050797	AC 140/	14 Eylül yayınları 4 Eylül ortak k	13.02.2018	27.02.2018	150125xxx	02	Felsefe (İÖ)		Fen Edebiyat Fakültesi
0065572	AC 140/	Kaid-i Azam : M. Ali Cinnah'ın	27.02.2009	04.03.2009	081014xxx	02	Sınıf Öğretmenliği (İÖ)		Eğitim Fakültesi
0051173	AC 140/	Konferanslar / Türk Kültürünü	27.02.2009	04.03.2009	081014xxx	02	Sınıf Öğretmenliği (İÖ)		Eğitim Fakültesi
0063789	AC 140/	Konferanslar / Türk Kültürünü	22.02.2010	22.02.2010	090664xxx	02	Orta Öğretim Sosyal Alanlar Eğitimi	Sosyal Bilimler Enstitüsü	
0039513	AC 140/	Prof.Dr. Sabri F. Ülgener'e arm	07.12.2011	07.12.2011	080110xxx	02	Sosyoloji		Fen Edebiyat Fakültesi
0053790	AC 140/	Prof.Dr. Sabri F. Ülgener'e arm	11.01.2012	16.01.2012	080113xxx	01	Sosyoloji (İÖ)		Fen Edebiyat Fakültesi
C0000588	AC 141	*Yapı Taşları ve Kaderimizin Çağ	28.04.2017	03.05.2017	160201xxx	02	İktisat		İktisadi ve İdari Bilimler Fakültesi
C0000850	AC 141	*Yapı Taşları ve Kaderimizin Çağ	06.12.2018		181290xxx	02	Yerel Yönetimler		Afyon Meslek Yüksekokulu
0063187	AC 141	*Yapı Taşları ve Kaderimizin Çağ	06.12.2018		181290xxx	02	Yerel Yönetimler (İÖ)		Afyon Meslek Yüksekokulu
0020831	AC 141	*Kırk Ambar: Lehçe-t-ül Hakayil	24.03.2016	28.04.2016	130115xxx	02	Çağdaş Türk Lehçeleri ve Edebiyatları		Fen Edebiyat Fakültesi
0007865	AC 141	*Kırk Ambar: Lehçe-t-ül Hakayil	01.12.2017	19.12.2017	140110xxx	02	Sosyoloji		Fen Edebiyat Fakültesi
0005596	AC 141	*Kırk Ambar: Lehçe-t-ül Hakayil	18.03.2017	27.03.2017	144201xxx	01	Makine Mühendisliği		Teknoloji Fakültesi

Şekil 3.1 Ham Veriler

3.4 Veri Önışleme

Veri setinin veri madenciliği programlarında kullanılmak üzere uygun olmayışı bu çalışmada elde edilen kütüphane verileri ile alakalı bir problemdir.

Bu problemin çözülmesi amacıyla öncelikle ihtiyaç duyulan verilerin belirlenmesi gerekmektedir. Bunun için sınıflama sütunu içerisinde kitap kodları ayrıştırılmış ve yeni sütun adı “kod” olarak değiştirilerek bu sütuna aktarılmıştır. İade Tarihinden Ödünç Tarihi matematiksel fark işlemi yapılarak sonuçlar iadegun sütunu oluşturulup aktarılmıştır. Fakülte ve enstitü isimleri numaralandırılmış ve son olarak sütun isimleri “kod, iadegun, cinsiyet, bolumno, okulno” öznitelik olarak oluşturulmuştur. Bu veriler ile yeni bir excel dosyası oluşturulmuştur (Şekil 3.2). Oluşturulan excel dosyası daha sonra .csv formatında kaydedilmiştir.

kod	iadegun	cinsiyet	bolumno	okulno
G	13 2		206	1
HD	9 2		206	1
G	13 2		206	1
HJ	100 1		206	1
HF	14 2		206	1
HD	5 2		206	1
HJ	100 1		206	1
G	13 2		206	1
HF	14 2		206	1
HJ	100 1		206	1
HF	14 2		206	1
HJ	100 1		206	1
HJ	100 1		206	1
HD	100 2		1	1
HD	100 2		1	1
G	14 2		1	1
NA	0 2		1	1
HD	100 2		1	1
G	14 2		1	1
PL	14 1		1	1
G	6 2		1	1
G	4 2		1	1
TX	6 2		1	1
TX	6 2		1	1
NK	0 2		1	1
QH	2 2		2	38
QM	100 2		2	38
QM	100 2		2	38
QM	0 1		2	38
QM	8 2		2	38
GV	100 1		2	38
QM	100 2		2	38
QM	100 2		2	38
QP	100 2		2	38
QM	100 2		2	38
QL	100 2		2	38
QM	14 2		2	38
RC	8 2		2	38
PS	15 2		2	38

Şekil 3.2 Verilerin Filtreleme İşleminde Geçmiş Hali

Şekil 3.2’deki verileri anlamlandırmak amacıyla; kitap kodlarının hangi alanlara karşılık geldiğinin belirlenmesi için Afyon Kocatepe Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı internet sayfasındaki LC Sınıflama şemasından yararlanılmıştır. İade gün sayısında fark işleminin sonucu yazıldığından ödünç tarihi değeri boş olunca sonuç negatif, iade tarihi boş olduğunda ise sonuç fazla çıktığından değişiklikler yapılmıştır. Burada gün sayısı 15 günü geçenler için 100 ve negatif değerler için ise 101 değeri verilmiştir. İade sütunundaki 0 değeri aynı gün içerisinde kitabın alınıp verildiğini göstermektedir. Cinsiyet sütununda değerlerin başındaki 0 rakamları silinmiştir. Bölümler sütunu A’dan Z’ye sıralanarak örgün ve ikinci öğretim olmasına göre birleştirilerek numaralandırılmıştır. Fakülte ve Enstitü sütunları birleştirilerek numaralandırılmıştır.

Veri setindeki veriler Excel ile .csv formatında kaydedilip, veri madenciliği yöntemlerini uygulamak üzere dosya biçimi .arff ye dönüştürülüp WEKA ya aktarılacaktır.

3.5 Model Seçimi

Verilerin elde edilmesi ve filtreleme yöntemlerinden sonra model seçimi işlemine geçilir. Boyutu büyük olan verilerin işlenmesi ve aralarındaki saklı örüntü ve eğilimlerin keşfedilmesini sağlayan veri madenciliği süreci farklı görevleri yerine getirmek için farklı algoritmalar kullanmaktadır. Algoritmalar verileri inceleyerek en uygun modeli seçerler. Dolayısıyla farklı modeller veri kümeleri üzerinde denenerek başarısı en yüksek olan model seçilmektedir.

Model seçilirken; verilerin belirli bölümü eğitilir, eğitilmiş veriler üzerinden oluşan örüntüler kullanılır ve geri kalan veriler kontrol edilir. Çalışmada örüntüler bulmak ve bunları kontrol etmek için WEKA sınıflandırma algoritmaları kullanılmıştır. Algoritmaların tüm analizleri öncelikle “use training set” yöntemi ile yapılmıştır. Örneğin Weka’da algoritma olarak J48 sınıflandırıcısı seçilmiş ve öznitelikler değiştirilerek test edilmiştir. En iyi sonuç Fakülte/Enstitü özniteliği seçildiğinde alındığından bu yöntem seçilmiştir. Sonuçlar Şekil 3.3’te verilmiştir.

```
=== Evaluation on training set ===

Time taken to test model on training data: 2.02 seconds

=== Summary ===

Correctly Classified Instances      371366           92.1943 %
Kappa statistic                     0.9025
Mean absolute error                  0.0069
Root mean squared error              0.0588
Relative absolute error              16.5538 %
Root relative squared error          40.6867 %
Total Number of Instances           402808
```

Şekil 3.3 J48 Algoritmasının Fakülte/Enstitü Özniteliğine Göre Sonuçları

Elde edilen bilgiler bulgular bölümünde verilmiştir.

Bütün veriler modeli eğitmek ve test etmekte kullanıldığı için sonuçlar yanıltıcı olabileceğinden literatürde sıklıkla kullanılan “Cross Validation” yöntemi ile sınıflandırma analizleri tercih edilmiştir. Ayrıca parçalara ayırma yöntemi olan cross validation yönteminde; veri seti 10 bölüme ayrılmış ve bu bölümler test verisi niteliğinde, geriye kalan 9 bölüm öğrenme verisi niteliğinde kullanılmıştır. Model seçildikten sonra

karşılaştırma yapılırken doğruluk değerine bakılmıştır. Verileri işlerken tahmin edilebilir alan için doğruluk performansına bakılarak yüksek olan algoritmalar kategorik ve orijinal veriler için ayrı ayrı elde edilmiştir. Sınıflandırma ve kümeleme algoritmalarının çalışma performansının test edilmesi amacıyla bütün algoritmalar denenmiştir. Bu algoritmalar aşağıda verilmiştir:

Model oluşturulurken sınıflandırma algoritmaları; lazy sınıflandırma algoritmaları “IBk, KStar vb.”, bayes sınıflandırma algoritmaları “BayesNet, NaiveBayes, NaiveBayesSimple, NaiveByes.Updateable”, functions sınıflandırma algoritmaları “SMO vb.”, misc sınıflandırma algoritmaları “HyperPipes, VFI”, meta sınıflandırma algoritmaları “FilteredClassifier, MultiClassClassifier vb.”, rules sınıflandırma algoritmaları “DecisionTable, ZeroR vb.”, trees sınıflandırma algoritmaları (SimpleCart, J48, RandomForest, RandomTree vb.) veri setine uygulanarak sonuçları incelenmiştir.

Model oluşturulurken kümeleme işlemleri için; “EM, Farthest First, Make Densit Based Clusterer, Filtered Clusterer, DBSCAN ve Simple K-Means” algoritmalarından yararlanılmıştır.

3.6 Modelin Değerlendirilmesi ve Yorumlanması

Model seçildikten sonra doğruluğunun testinde çalışılan en kolay yöntem basit geçerlilik testi “simple validation test” dir. “Correctly Classified Instances”, sınıflandırma algoritması eğitim verilerinin doğru ve yanlış biçimde sınıflandırma oranını göstermektedir. “Kappa değeri”, nicel olarak gözlemler arasındaki eşleşmenin büyüklüğünü gösterdiğinden, 1 değerine yaklaşıldıkça uyum oranı artmaktadır.

Algoritma karşılaştırmaları “RMSE (Root Mean-Squared Error, Ortalama Karesel Hatanın Karekökü)” ve “MAE (Mean Absolute Error, Ortalama Mutlak Hata)” değerlerine göre yapıp en küçük değer baz alınmaktadır. “RMSE” tek tek veri örneğindeki gerçek değerleriyle tahmin edilen değeri arasındaki değerin farkının alınması ile elde edilip değer ne kadar az ise hata o kadar az olacaktır. Doğruluk değeri tahmin modeli için karşılaştırma amacıyla kullanılmaktadır. “MAE (Mean Absolute Error, Ortalama Mutlak Hata)” hatası standart seçilerek değeri minimum olan tekniğin en iyi hata tahmini tekniğidir.

F1 skoru / F skoru, ölçüt olarak F, verilen sınıfın sınıflandırma algoritması ile ne kadar doğru sınıflandırma tahminin yapılabileceğini göstermektedir. Dolayısıyla değer 1'e yakınsadıkça verilen sınıfın algoritma ile öğrenilmesinin o ölçütte başarılı olduğu nu ifade etmektedir. F ölçütü 0,5 ve üzeri olan sınıflar modellenebilmektedir. En yüksek doğrulukla tahmin yapabilen WEKA sınıflandırma algoritmalarının öncelikle doğruluk yüzdelere bakılmış, eşit olan algoritmalarda ise RMSE değeri düşük olan algoritma seçilmiştir.

Algoritmaların analizi aşamasında bazı algoritmalar sürekli, bazıları ise kategorik verilerle çalıştığından; bir algoritma; veri setinin orjinal hali, kategorik hale getirilmiş ve normalize edilmiş halleri ile veriler için ayrı ayrı WEKA programında analiz edilerek, karşılaştırmada her veri seti kategorik ve orijinal veri değerleri kendi içinde değerlendirmeye alınmıştır. WEKA programı üzerinde tüm veri grupları bulunan tüm algoritmaların testinden geçirilmiştir.

4. BULGULAR

Üniversite Merkez Kütüphanesine ait veri seti üzerinde kitap ödünç alma ilişkileri veri madenciliği programlarından WEKA ile analiz edilip, değerlendirilmiştir.

Kod, bolumno, okulno, cinsiyet ve iadegun öznitelik olarak kabul edilmiş bunlara göre sınıflandırma ve kümeleme algoritmaları uygulanmış ve sonuçlar eklenmiştir.

WEKA üzerinde önışlem kısmında dağılım görselleri üzerinden özniteliklere göre ilk 10 sıralama analiz edilmiştir. Alınan kitapların kodları, konuları, sıklıkları ve yüzdelere göre ilk 10 sıralaması Çizelge 4.1’de verilmiştir:

Çizelge 4.1 Ödünç Alınan Kitapların Türlerine Göre Dağılımları

<i>LC Kodu</i>	<i>Konu</i>	<i>Frekans</i>	<i>Yüzde</i>
PL	Türk Dili ve Edebiyatı	85415	21,20
D	Tarih	54243	13,46
B	Felsefe, Psikoloji, Din	29683	7,37
HM	Sosyoloji	21192	5,26
L	Eğitim	15951	3,96
QA	Matematik. İstatistik. Bilgisayar	15866	3,94
HF	Ticaret	13552	3,36
PS	Amerikan edebiyatı	12265	3,04
HD	Toprak kullanımı. Tarım. Sanayi	11936	2,96
J	Siyaset bilimi	10654	2,64

Çizelge 4.1 incelendiğinde en çok ödünç alınan ilk 5 kitap sınıfının sosyal bilimler alanında olduğu görülmektedir. Listedeki teknik tek alan altıncı sırada (%3,94) yer alan “Matematik, İstatistik, Bilgisayar” alanıdır. Teknik alanlarda kitap ödünç alma işleminin daha az olduğu söylenebilir. İlk on alan toplam ödünç almaların %67’sini içermektedir. “Türk Dili ve Edebiyatı” ve “Tarih” alanları dışındaki alanların yüzde olarak payı düşük miktardadır.

Ödünç alınan kitapların bölümlere göre dağılımları sıralamasındaki ilk on bölüm Çizelge 4.2’de görülmektedir.

Çizelge 4.2 Ödünç Alınan Kitapların Bölümlere Göre Dağılımları

<i>Bölüm</i>	<i>Frekans</i>	<i>Yüzde</i>
Tarih	43809	10,87
Sosyoloji	37339	9,27
Türk Dili ve Edebiyatı	35872	8,90
İşletme	20534	5,10
Kamu Yönetimi	19751	4,90
İktisat	14219	3,53
Maliye	13479	3,35
Sosyal Bilgiler Öğretmenliği	11082	2,75
Sınıf Öğretmenliği	10414	2,58
Çağdaş Türk Lehçeleri ve Edebiyatları	10367	2,57

Çizelge 4.2 incelendiğinde ilk on bölüm içerisinde herhangi bir teknik bölüm olmadığı görülmektedir. Bölümlerin Fen Edebiyat Fakültesi, İktisadi ve İdari Bilimler Fakültesi ve Eğitim Fakültesindeki bölümlerden oluştuğu söylenebilir. İlk on bölüm toplam ödünç alma sayısının %54'lük bir kısmını göstermektedir.

Ödünç alınan kitapların kurumlara göre dağılımları sıralamasındaki ilk on kurum Çizelge 4.3'de görülmektedir.

Çizelge 4.3 Ödünç Alınan Kitapların Kurumlara Göre Dağılımları

<i>Bölüm</i>	<i>Frekans</i>	<i>Yüzde</i>
Fen Edebiyat Fakültesi	142402	35,34
İktisadi ve İdari Bilimler Fakültesi	67596	16,78
Eğitim Fakültesi	52386	13,00
Sosyal Bilimler Enstitüsü	35629	8,84
Mühendislik Fakültesi	28731	7,13
Afyon Meslek Yüksekokulu	11872	2,95
Teknoloji Fakültesi	11298	2,80
Tıp Fakültesi	8712	2,16
Güzel Sanatlar Fakültesi	7369	1,83
Afyon Sağlık Yüksekokulu	6538	1,62

Çizelge 4.3 incelendiğinde ilk on kurum içerisinde ağırlığın sosyal alanlarda eğitim veren kurumlar olduğu görülmektedir. Ödünç alınan kitapların %35'lik bir bölümü "Fen

Edebiyat Fakültesi” öğrencileri tarafından alınmaktadır. Herhangi bir teknik bölüm olmadığı görülmektedir. İlk on kurum toplam ödünç alma sayısının %92’lik bir kısmını göstermektedir.

Ödünç alınan kitapların cinsiyete göre dağılımları Çizelge 4.4’de görülmektedir.

Çizelge 4.4 Ödünç Alınan Kitapların Cinsiyete Göre Dağılımları

<i>Cinsiyet</i>	<i>Frekans</i>	<i>Yüzde</i>
Kadın	256147	63,57
Erkek	146661	36,43

Çizelge 4.4 incelendiğinde ödünç alınan kitapların çoğunluğunun kadın (yaklaşık %64) öğrenciler tarafından alındığı görülmektedir.

Ödünç alınan kitapların yıllara göre dağılımları Çizelge 4.5’de görülmektedir.

Çizelge 4.5 Ödünç Alınan Kitapların Yıllara Göre Dağılımları

<i>Yıl</i>	<i>Ödünç İşlem Sayısı</i>	<i>Yüzde</i>
2009	41311	10,25%
2010	37165	9,22%
2011	33530	8,32%
2012	32083	7,96%
2013	30496	7,57%
2014	34424	8,54%
2015	46277	11,49%
2016	45263	11,23%
2017	51321	12,74%
2018	51047	12,67%
Toplam	402917	100

Çizelge 4.5 incelendiğinde ödünç alınan kitapların yıllara göre dağılımının yaklaşık olarak eşit olduğu görülmektedir. En yüksek kitap ödünç işlemi 2017 yılında (%12,74), en az kitap ödünç alma işlemi 2013 yılında (%7,57) görülmektedir.

Bölümlere göre hangi kategorideki kitapların en çok alındığının tespit edilmesi amacıyla sınıflandırma algoritmalarından; bayes sınıflandırıcı algoritması NaiveBayes seçilmiştir. Beş öznitelik değiştirilerek sınıf olarak Weka'ya verilerek çıkan sonuçlar analiz edilmiştir. Örnek olarak bolumno ve kod öznitelikleri karşılaştırılmıştır. Bu karşılaştırma sonucunda hangi bölümlerde hangi kitapların daha fazla alındığı elde edilmeye çalışılmıştır (Çizelge 4.6).

Çizelge 4.6 NaiveBayes Algoritması ile Ödünç Alınan Kitapların Bölümlere Göre Dağılımları

<i>LC Kodu ve Konusu</i>	<i>Bölüm</i>	<i>Sıklık</i>
G: Coğrafya. Antropoloji. Turizm	Coğrafya	1196
QK: Botanik		127
QE: Jeoloji		212
HD: Toprak Kullanımı. Tarım. Sanayi	İşletme	2329
HF: Ticaret		3203
PS: Amerikan Edebiyatı		972
PR: İngiliz Edebiyatı		575
PE: İngiliz Dili		633
TS: İmalat		114
HG: Finans		1100
PF: Alman Dili		35
T: Teknoloji		98
HJ: Kamu Maliyesi	Maliye	2502
NA: Mimarlık	İç Mimarlık ve Çevre Tasarımı	379
PL: Türk Dili ve Edebiyatı	Türk Dili ve Edebiyatı	27143
P: Dil ve Edebiyat		1160
PQ: Fransız Edebiyatı		599
GR: Folklor		239
PK: Uzakdoğu Dilleri		214
TX: Ev Ekonomisi	Gastronomi ve Mutfak Sanatları	876
NK: Dekoratif Sanatlar. Uygulamalı Sanatlar.	Seramik	176
Dekorasyon ve Süsleme		
TP: Kimya Teknolojisi	Biyoloji	196
QH: Biyoloji		312
QL: Zooloji		155
QZ: Pataloji		7
QM: İnsan Anotomisi	Tıp	1371
QP: Biyokimya		503

Çizelge 4.6 (Devam) NaiveBayes Algoritması ile Ödünç Alınan Kitapların Bölümlere Göre Dağılımları

R: Tıp		4349
QT: Fizyoloji		17
QR: Mikrobiyoloji	Tıp	752
QV: Farmakoloji		4
QW: Mikrobiyoloji ve İmmünoloji		13
GV: Rekreasyon. Spor	Beden Eğitimi ve Spor Öğretmenliği	352
QA: Matematik. İstatistik. Bilgisayar	Matematik	5923
PA: Yunan Edebiyatı		72
D: Tarih		28880
CT: Biyografi		53
U: Askerlik		127
Z: Kaynakçalar. Kütüphanecilik	Tarih	68
CC: Arkeoloji		37
HE: Taşımacılık. Telekomünikasyon		91
V: Denizcilik		14
C: Tarihe Yardımcı Bilimler		117
L: Eğitim	Sınıf Öğretmenliği	2948
PT: Alman Edebiyatı		167
B: Felsefe. Psikoloji. Din		6059
Q: Bilim		347
HM: Sosyoloji		11801
HC: Ekonomik Tarih	Sosyoloji	902
CB: Uygarlık Tarihi		318
H: Sosyal Bilimler		230
A: Genel Yapıtlar		334
GN: Antropoloji		202
PG: Rus Dili ve Edebiyatı	Çağdaş Türk Lehçeleri ve Edebiyatları	787
E: Amerika Tarihi		45
J: Siyaset Bilimi	Kamu Yönetimi	3976
TD: Çevre Teknolojisi		72
QU: Biyokimya ve Beslenme	Beslenme ve Diyetetik	52
TR: Fotoğrafçılık	Sinema ve Televizyon	146
S: Tarım	Veteriner	403
TN: Maden Mühendisliği	Maden Mühendisliği	430
HB: Ekonomik Teori	İktisat	2391

Çizelge 4.6 (Devam) NaiveBayes Algoritması ile Ödünç Alınan Kitapların Bölümlere Göre Dağılımları

K: Hukuk	Hukuk	2355
TA: Mühendislik. İnşaat Mühendisliği	İnşaat Mühendisliği	1997
TH: Yapı ve İnşaat		92
TC: Hidrolik Mühendisliği	İnşaat Mühendisliği	141
TF: Demiryolu Mühendisliği		9
TE: Karayolu Mühendisliği		52
QD: Kimya	Kimya	1278
QC: Fizik	Fizik	1622
M: Müzik	Piyano	122
N: Güzel Sanatlar	Resim	1248
QB: Astronomi	Harita Mühendisliği	118
HA: İstatistik	İstatistik	72
RS: Eczacılık. Tıbbi Maddeler	Moleküler Biyoloji ve Genetik	11
TK: Elektrik Mühendisliği. Elektronik. Nükleer Mühendislik	Elektrik Mühendisliği	1305
TJ: Mekanik Mühendisliği	Makine Mühendisliği	360
TL: Havacılık. Uzay Havacılığı	Otomotiv Mühendisliği	33
PC: Fransız Dili	Uluslararası Ticaret ve Finansman	9
TT: El Sanatları	Okul Öncesi Öğretmenliği	14
TG: Köprü Mühendisliği	Yapı Öğretmenliği	5

Çizelge 4.6 incelendiğinde NaiveBayes algoritması sonucunda üç ve üçten fazla alanda araştırma yapılan ödünç alınan kitaplar Coğrafya, İşletme, Türk Dili ve Edebiyatı, Biyoloji, Tıp, Tarih, Sosyoloji, Kamu Yönetimi, İnşaat Mühendisliği bölümü öğrencileri tarafından alındığı görülmektedir. Ödünç alınan kitapların sıklık sayısına bakıldığında da en çok D: Tarih alanındaki kitapların 28880 sıklıkla Tarih bölümünde alındığı görülmektedir.

Kurumlara göre hangi kategorideki kitapların en çok alındığının tespit edilmesi amacıyla sınıflandırma algoritmalarından Trees “ADTree, Id3, BFTree, DesicionStump, FT, J48, j48graft, LADTree, LMT, NBTree, RandomForest, RandomTree, REPTree, SimpleCart” içerisinde **J48** algoritması ile testler edilmiştir.

J48, “böl ve fethet algoritması” iki durumun farklı sınıflara ait olmalarına rağmen her bir öznitelik için aynı değerlere sahip olduğu daha fazla bölünmenin imkânsız olacağı duruma kadar ya da her yaprak tek sınıf durumu içerinceye kadar veriyi parçalar. Sonuçta, çelişkili durumlar yoksa karar ağacı tüm eğitim durumlarını doğru sınıflandırır. Aşırı uyum, üretildikten sonra karar ağacının bir kısım yapısını kaldırarak ya da bazı eğitim durumları kümelerini alt bölümlere ayrılmaktan koruyan durdurma kriterleri ile önlenabilir (Elomaa and Kaariainen 2001).

J48 algoritmasının Weka’da Fakülte/Enstitü özniteliği seçilerek sonuç alındığında 10 yıl içerisinde bölüm, fakülte/enstitü öznitelikleri ve frekansları yüksek olan sonuçlar üzerinde kütüphane kullanımının hangi bölümler ve enstitüler üzerindeki dağılımları Çizelge 4.7’de verilmiştir.

Çizelge 4.7 J48 Algoritması Sonucunun Bölüm ve Kurumlara Göre Dağılımları

<i>Bölüm</i>	<i>Fakülte/Enstitü</i>	<i>Frekans</i>
Çocuk Gelişimi	Afyon Meslek Yüksekokulu	1022.0/119.0
Muhasebe	Afyon Meslek Yüksekokulu	1026.0/232.0
Yerel Yönetimler	Afyon Meslek Yüksekokulu	1322.0/8.0
Mimari Restorasyon	Afyon Meslek Yüksekokulu	1343.0/27.0
Muhasebe ve Vergi Uygulamaları	Afyon Meslek Yüksekokulu	1615.0/99.0
Hemşirelik	Afyon Sağlık Yüksekokulu	1083.0/11.0
Beslenme ve Diyetetik	Afyon Sağlık Yüksekokulu	1518.0
Fizyoterapi ve Rehabilitasyon	Afyon Sağlık Yüksekokulu	1548.0/25.0
Sağlık Kurumları Yöneticiliği	Afyon Sağlık Yüksekokulu	2144.0/20.0
Beden Eğitimi ve Spor Öğretmenliği	Beden Eğitimi ve Spor Yüksekokulu	1280.0
Sınıf Öğretmenliği	Eğitim Fakültesi	10414.0/1326.0
Sosyal Bilgiler Öğretmenliği	Eğitim Fakültesi	11082.0/810.0
Makine Resim ve Konstrüksiyon Öğretmenliği	Eğitim Fakültesi	1118.0/12.0
Elektrik Öğretmenliği	Eğitim Fakültesi	1192.0/4.0
Yapı Öğretmenliği	Eğitim Fakültesi	1480.0/9.0
Fen Bilgisi Öğretmenliği	Eğitim Fakültesi	2458.0
Bilgisayar ve Öğretim Teknolojileri Öğretmenliği	Eğitim Fakültesi	3127.0
Okul Öncesi Öğretmenliği	Eğitim Fakültesi	8974.0
Türkçe Öğretmenliği	Eğitim Fakültesi	9426.0

Çizelge 4.7 (Devam) J48 Algoritması Sonucunun Bölüm ve Kurumlara Göre Dağılımları

Metal Öğretmenliği	Eğitim Fakültesi	949.0/9.0
İlköğretim Matematik Öğretmenliği	Eğitim Fakültesi	3653.0
Moleküler Biyoloji ve Genetik	Fen Edebiyat Fakültesi	1009.0/47.0
Çağdaş Türk Lehçeleri ve Edebiyatları	Fen Edebiyat Fakültesi	10367.0/68.0
Biyoloji	Fen Edebiyat Fakültesi	2066.0/232.0
Felsefe	Fen Edebiyat Fakültesi	2567.0/9.0
Kimya	Fen Edebiyat Fakültesi	2779.0/155.0
Fizik	Fen Edebiyat Fakültesi	2843.0/350.0
İstatistik	Fen Edebiyat Fakültesi	3454.0/320.0
Türk Dili ve Edebiyatı	Fen Edebiyat Fakültesi	35872.0/2111.0
Sosyoloji	Fen Edebiyat Fakültesi	37339.0/4712.0
Tarih	Fen Edebiyat Fakültesi	43809.0/6557.0
Coğrafya	Fen Edebiyat Fakültesi	6589.0/352.0
Matematik	Fen Edebiyat Fakültesi	8997.0/418.0
İç Mimarlık ve Çevre Tasarımı	Güzel Sanatlar Fakültesi	1483.0
Resim	Güzel Sanatlar Fakültesi	1999.0
Sinema ve Televizyon	Güzel Sanatlar Fakültesi	3013.0
Seramik	Güzel Sanatlar Fakültesi	970.0/164.0
Hukuk	Hukuk Fakültesi	3831.0
Maliye	İktisadi ve İdari Bilimler Fakültesi	13479.0/2105.0
İktisat	İktisadi ve İdari Bilimler Fakültesi	14219.0/1070.0
Kamu Yönetimi	İktisadi ve İdari Bilimler Fakültesi	19751.0/2090.0
Uluslararası Ticaret ve Finansman	İktisadi ve İdari Bilimler Fakültesi	8773.0/21.0
İslami İlimler	İslami İlimler Fakültesi	3387.0
Malzeme Bilimi ve Mühendisliği	Mühendislik Fakültesi	2562.0/21.0
Maden Mühendisliği	Mühendislik Fakültesi	2586.0/95.0
Elektrik Mühendisliği	Mühendislik Fakültesi	3055.0
Kimya Mühendisliği	Mühendislik Fakültesi	3414.0/100.0
Gıda Mühendisliği	Mühendislik Fakültesi	3834.0/181.0
Biyomedikal Mühendisliği	Mühendislik Fakültesi	4108.0/64.0
Harita Mühendisliği	Mühendislik Fakültesi	4157.0/7.0
İnşaat Mühendisliği	Mühendislik Fakültesi	4428.0/147.0
Yönetim ve Organizasyon	Sosyal Bilimler Enstitüsü	1162.0/5.0
Muhasebe - Finansman	Sosyal Bilimler Enstitüsü	1197.0/5.0
Eğitim Bilimleri	Sosyal Bilimler Enstitüsü	1227.0/11.0
Üretim Yönetimi ve Pazarlama	Sosyal Bilimler Enstitüsü	1571.0/3.0
Otomotiv Öğretmenliği	Teknik Eğitim Fakültesi	1295.0/390.0

Çizelge 4.7 (Devam) J48 Algoritması Sonucunun Bölüm ve Kurumlara Göre Dağılımları

Metalurji ve Malzeme Mühendisliği	Teknoloji Fakültesi	1266.0/137.0
Mekatronik Mühendisliği	Teknoloji Fakültesi	1630.0/9.0
Otomotiv Mühendisliği	Teknoloji Fakültesi	1929.0/15.0
Makine Mühendisliği	Teknoloji Fakültesi	2999.0/339.0
Elektrik-Elektronik Mühendisliği	Teknoloji Fakültesi	3414.0/183.0
Tıp	Tıp Fakültesi	8725.0/22.0
Turizm Rehberliği	Turizm Fakültesi	1194.0/63.0
Gastronomi ve Mutfak Sanatları	Turizm Fakültesi	3034.0
Seyahat İşletmeciliği	Turizm İşletmeciliği ve Otelcilik Yüksekokulu	1098.0/274.0
Konaklama İşletmeciliği	Turizm İşletmeciliği ve Otelcilik Yüksekokulu	2182.0/116.0
Veteriner	Veteriner Fakültesi	1873.0/32.0

Çizelge 4.7’de bölüm ve fakülte/enstitü’nün haricinde frekans sütununda algoritmanın hangi oranda olumlu/olumsuz çalıştığı gösterilmektedir. Son 10 yılda Sözel alanda Türk Dili ve Edebiyatı, Sosyoloji ve Tarih bölümü öğrencileri kütüphaneyi en çok kullanan ilk 3’tür. Sayısal alanda ise Elektrik, Kimya, Gıda, Biyomedikal, Harita, İnşaat, Makine, Elektrik-Elektronik Mühendislik bölümleri ve Tıp öğrencileri kütüphaneyi en çok kullananlardır. J48’in diğer sonuçları diğer kriterleri de içermektedir.

Çizelge 4.8 Beden Eğitimi ve Spor (YL) Bölümünün İade Gün Sayısına Göre Dağılımları

LC Kodu ve Konusu	İade	Bölüm	Fakülte/Enstitü	Frekans
GV : Rekreasyon. Spor	15	Beden Eğitimi ve Spor (YL)	Beden Eğitimi ve Spor Yüksekokulu	3.0
GV : Rekreasyon. Spor	15+	Beden Eğitimi ve Spor (YL)	Sağlık Bilimleri Enstitüsü	18.0/3.0
L : Eğitim	4	Beden Eğitimi ve Spor (YL)	Beden Eğitimi ve Spor Yüksekokulu	3.0
L : Eğitim	15+	Beden Eğitimi ve Spor (YL)	Sağlık Bilimleri Enstitüsü	2.0

Çizelge 4.8’de **GV : Rekreasyon. Spor** ve **L : Eğitim** alanındaki kitaplar Sağlık Bilimleri Enstitüsü Beden Eğitimi ve Spor (YL) bölümündeki öğrenciler tarafından 15 günden fazla alınıp kütüphane sistemine göre cezaya kaldıkları tespit edilmiştir. Ayrıca Sağlık Bilimleri Enstitüsü Beden Eğitimi ve Spor (YL) öğrencileri **GV : Rekreasyon. Spor** alanındaki kitaplar daha popülerdir.

Çizelge 4.9 Bilgisayar Teknolojileri ve Programcılığı Bölümünün Cinsiyete Göre Dağılımları

<i>LC Kodu ve Konusu</i>	<i>Cinsiyet</i>	<i>Bölüm</i>	<i>Fakülte/Enstitü</i>	<i>Frekans</i>
HD : Toprak kullanımı. Tarım. Sanayi	*	Bilgisayar Teknolojileri ve Programcılığı	Afyon Meslek Yüksekokulu	5.0
HF : Ticaret	Erkek	Bilgisayar Teknolojileri ve Programcılığı	Afyon Meslek Yüksekokulu	11.0
HF : Ticaret	Kadın	Bilgisayar Teknolojileri ve Programcılığı	Bolvadin Meslek Yüksekokulu	2.0
PL : Türk Dili ve Edebiyatı	*	Bilgisayar Teknolojileri ve Programcılığı	Afyon Meslek Yüksekokulu	18.0/1.0
QA : Matematik. İstatistik. Bilgisayar	Erkek	Bilgisayar Teknolojileri ve Programcılığı	Emirdağ Meslek Yüksekokulu	35.0/7.0
QA : Matematik. İstatistik. Bilgisayar	Kadın	Bilgisayar Teknolojileri ve Programcılığı	Afyon Meslek Yüksekokulu	33.0/7.0

Çizelge 4.9’da **HF** : Ticaret alanındaki kitaplar Afyon Meslek Yüksekokulu Bilgisayar Teknolojileri ve Programcılığı Erkek öğrencileri tarafından Bolvadin Meslek Yüksekokulu aynı bölümdeki Kadın öğrencilerden daha çok ödünç alınmıştır. **QA** : Matematik. İstatistik. Bilgisayar alanındaki kitaplar Emirdağ Meslek Yüksekokulu Erkek öğrencileri tarafından Afyon Meslek Yüksekokulu aynı bölümdeki Kadın öğrencilerden daha çok ödünç alınmıştır. *: Boş veri anlamına gelmektedir.

Devlet Konservatuvarı Güzel Sanatlar Fakültesi öğrencileri **M**: Müzik (18.0) olarak en çok bu alandaki kitapları tercih etmiştir.

Çizelge 4.10 Dış Ticaret Bölümünün İade Günü ve Cinsiyete Göre Dağılımları

<i>LC Kodu ve Konusu</i>	<i>İade</i>	<i>Cinsiyet</i>	<i>Bölüm</i>	<i>Fakülte/Enstitü</i>	<i>Frekans</i>
*	*	Erkek	Dış Ticaret	Bayat Meslek Yüksekokulu	5.0
*	*	Kadın	Dış Ticaret	Bolvadin Meslek Yüksekokulu	11.0
*	*	Erkek	Dış Ticaret	Bolvadin Meslek Yüksekokulu	2.0
*	*	Kadın	Dış Ticaret	Bayat Meslek Yüksekokulu	18.0/1.0
HC : Ekonomik Tarih	11	*	Dış Ticaret	Bolvadin Meslek Yüksekokulu	2.0
HB : Ekonomik Teori	11	*	Dış Ticaret	Bayat Meslek Yüksekokulu	3.0
HG : Finans	11	*	Dış Ticaret	Bolvadin Meslek Yüksekokulu	4.0
*	15+	*	Dış Ticaret	Bayat Meslek Yüksekokulu	62.0/5.0

Dış Ticaret bölümünde Erkek ve Kadın öğrencileri kitap ödünç almalarına bakıldığında Bayat Meslek Yüksekokulunda ve Bolvadin Meslek Yüksekokulunda kadın öğrencilerin daha çok kitap ödünç aldıkları tespit edilmiştir. **HG** : Finans alanındaki kitaplar iade günü 11 gün olarak Bolvadin Meslek Yüksekokulu öğrencileri tarafından alınmıştır. Kitaplar Bayat Meslek Yüksekokulu Dış Ticaret bölümündeki öğrenciler tarafından 15 günden fazla alınıp kütüphane sistemine göre cezaya kaldıkları tespit edilmiştir.

Sosyal Bilimler Enstitüsü Felsefe Öğretmenliği öğrencileri **L** : Eğitim (8.0) alanındaki kitapları Eğitim Fakültesi (2.0) ve Fen Edebiyat Fakültesi (3.0) öğrencilerinden daha çok ödünç almıştır. Tablo 5.10'da İşletme bölümünün kurumlara göre dağılımları verilmiştir.

Çizelge 4.11 İşletme Bölümünün Kurumlara, İade Güne ve Cinsiyete Göre Dağılımları

<i>LC Kod ve Konu</i>	<i>İade</i>	<i>Cinsiyet</i>	<i>Bölüm</i>	<i>Fakülte/Enstitü</i>	<i>Frekans</i>
HD : Tarım. Sanayi	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	2328.0/1178.0
HF : Ticaret	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	3202.0/941.0
PL : Türk Dili ve Edebiyatı	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	3186.0/297.0
TX: Ev Ekonomisi	1	*	İşletme	Turizm İşletmeciliği ve Otelcilik Y.	5.0/2.0
TX : Ev Ekonomisi	3	*	İşletme	Sosyal Bilimler Enstitüsü	2.0
TX : Ev Ekonomisi	6	*	İşletme	Afyon Meslek Yüksekokulu	14.0/4.0
TX : Ev Ekonomisi	7	Erkek	İşletme	Afyon Meslek Yüksekokulu	2.0
TX : Ev Ekonomisi	7	Kadın	İşletme	Sosyal Bilimler Enstitüsü	10.0
TX : Ev Ekonomisi	9	Erkek	İşletme	Turizm İşletmeciliği ve Otelcilik Y.	2.0
TX : Ev Ekonomisi	9	Kadın	İşletme	Sosyal Bilimler Enstitüsü	4.0/1.0
TX : Ev Ekonomisi	11	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	4.0/1.0
TX : Ev Ekonomisi	12	*	İşletme	Afyon Meslek Yüksekokulu	2.0
TX : Ev Ekonomisi	13	*	İşletme	Turizm İşletmeciliği ve Otelcilik Yüksekokulu	5.0
TX : Ev Ekonomisi	14	Erkek	İşletme	İktisadi ve İdari Bilimler Fakültesi	6.0
TX : Ev Ekonomisi	14	Kadın	İşletme	Turizm İşletmeciliği ve Otelcilik Y.	5.0/1.0
TX : Ev Ekonomisi	15	*	İşletme	Turizm İşletmeciliği ve Otelcilik Y.	3.0
TX : Ev Ekonomisi	15+	Erkek	İşletme	Sosyal Bilimler Enstitüsü	15.0/9.0
TX : Ev Ekonomisi	15+	Kadın	İşletme	Turizm İşletmeciliği ve Otelcilik Y.	33.0/15.0
NK : Dekoratif ve Uygulamalı Sanatlar.	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	7.0/1.0
QH : Biyoloji	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	30.0/3.0
GV : Rekreasyon. Spor	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	27.0/3.0
R : Tıp	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	53.0/12.0
PS : Amerikan Edebiyatı	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	971.0/54.0
QA : Matematik. İstatistik.	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	220.0/36.0
PR : İngiliz Edebiyatı	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	574.0/45.0
PA : Yunan Edebiyatı	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	17.0/2.0

Çizelge 4.11 (Devam) İşletme Bölümünün Kurumlara, İade Güne ve Cinsiyete Göre Dağılımları

L : Eğitim	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	279.0/143.0
PE : İngiliz Dili	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	632.0/41.0
PT : Alman Edebiyatı	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	157.0/11.0
PG : Rus Dili ve Edebiyatı	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	425.0/29.0
B : Felsefe, Psikoloji, Din	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	1304.0/143.0
E : Amerika Tarihi	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	20.0/2.0
D : Tarih	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	1027.0/148.0
P : Dil ve Edebiyat	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	306.0/36.0
PQ : Fransız Edebiyatı	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	576.0/38.0
Q : Bilim	*	Erkek	İşletme	İktisadi ve İdari Bilimler Fakültesi	22.0/8.0
Q : Bilim	*	Kadın	İşletme	Sosyal Bilimler Enstitüsü	21.0/10.0
HM : Sosyoloji	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	467.0/95.0
HC : Ekonomik Tarih	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	408.0/91.0
CB : Uygurlık Tarihi	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	23.0/4.0
J : Siyaset Bilimi	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	204.0/53.0
H : Sosyal Bilimler	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	161.0/43.0
TR : Fotoğrafçılık	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	13.0/3.0
CT : Biyografi	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	25.0/1.0
S : Tarım	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	19.0/1.0
TN : Maden Mühendisliği	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	14.0/8.0
HB : Ekonomik Teori	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	872.0/54.0
GR : Folklor	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	10.0/2.0
TS : İmalat	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	113.0/17.0
K : Hukuk	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	633.0/55.0
HG : Finans	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	1099.0/387.0
TA : İnşaat Mühendisliği	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	33.0/9.0
QD : Kimya	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	14.0
U : Askerlik	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	18.0/3.0
A : Genel Yapıtlar	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	41.0/12.0
QC : Fizik	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	29.0/1.0
M : Müzik	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	19.0/5.0
PK : Uzakdoğu Dilleri	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	34.0/5.0
QR : Mikrobiyoloji	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	10.0/1.0
N : Güzel Sanatlar	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	44.0/5.0
QB : Astronomi	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	8.0/1.0
HA : İstatistikler	*	Erkek	İşletme	Sosyal Bilimler Enstitüsü	8.0/1.0
HA : İstatistikler	*	Kadın	İşletme	İktisadi ve İdari Bilimler Fakültesi	24.0/4.0
RS : Eczacılık, Tıbbi Maddeler	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	5.0/1.0
QV : Farmakoloji	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	2.0/1.0
TD : Çevre Teknolojisi	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	6.0/2.0

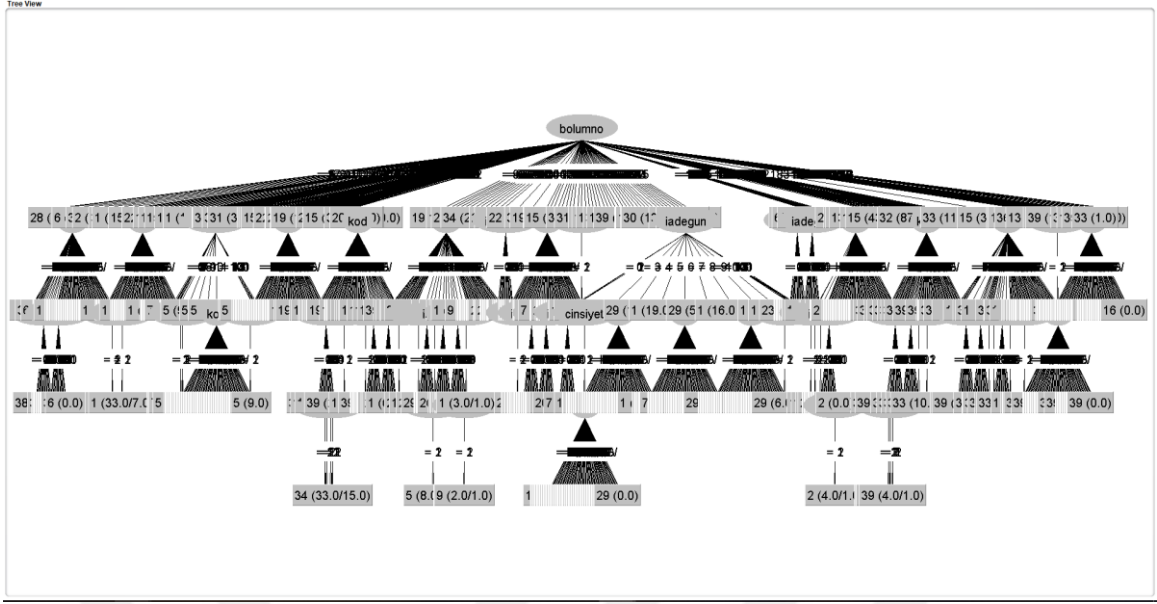
Çizelge 4.11 (Devam) İşletme Bölümünün Kurumlara, İade Güne ve Cinsiyete Göre Dağılımları.

TK : Elektrik Elektronik. Nükleer Mühendislik	0	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	4.0
TK : Elektrik Elektronik. Nükleer Mühendislik	7	*	İşletme	Sosyal Bilimler Enstitüsü	5.0/1.0
TK : Elektrik Elektronik. Nükleer Mühendislik	8	*	İşletme	Sosyal Bilimler Enstitüsü	3.0
TK : Elektrik Elektronik. Nükleer Mühendislik	10	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	4.0
TK : Elektrik Elektronik. Nükleer Mühendislik	12	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	2.0
TK : Elektrik Elektronik. Nükleer Mühendislik	15+	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	3.0
GN : Antropoloji	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	5.0/1.0
Z : Kaynakçılar. Kütüphanecilik	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	9.0/4.0
PF : Alman Dili	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	34.0/5.0
TJ : Mekanik Mühendisliği	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	15.0/5.0
T : Teknoloji	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	97.0/21.0
TL : Havacılık. Uzun Havacılığı	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	6.0/1.0
PC: Fransız Dili	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	6.0
QK : Botanik	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	4.0
TH : Yapı, İnşaat	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	5.0
CC : Arkeoloji	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	2.0/1.0
TC : Hidrolik Mühendisliği	*	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	2.0
HE : Taşımacılık. Telekomünikasyon	5	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	2.0
HE : Taşımacılık. Telekomünikasyon	6	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	2.0
HE : Taşımacılık. Telekomünikasyon	8	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	7.0/1.0
HE : Taşımacılık. Telekomünikasyon	15	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	3.0
HE : Taşımacılık. Telekomünikasyon	15+	*	İşletme	İktisadi ve İdari Bilimler Fakültesi	6.0/2.0

Çizelge 4.11 incelendiğinde İktisadi ve İdari Bilimler Fakültesi İşletme bölümü öğrencileri **HF**: Ticaret alanındaki kitapları bu algoritmanın olumlu/olumsuz (3202.0/941.0) durumlarına göre daha çok tercih edildiği görülmektedir. Bu bölümde üzerinde en araştırma yapılan ilk üç kitap **HF**: Ticaret, **PL**: Türk Dili ve Edebiyatı, **HD**: Toprak kullanımı. Tarım. Sanayi alanında olduğu tespit edilmiştir. Genel olarak İşletme bölümünün bulunduğu kurumlar Afyon, Çay ve Turizm İşletmeciliği ve Otelcilik Meslek Yüksekokulları, İktisadi ve İdari Bilimler Fakültesi ve Sosyal Bilimler Enstitüsü

kurumları arasında karşılaştırma yapılabilir. **TX:** Ev Ekonomisi, **TK:** Elektrik Mühendisliği. Elektronik. Nükleer Mühendislik ve **HE:** Taşımacılık. Telekomünikasyon kitapları arasında iade gün ve cinsiyet kriterlerine göre karşılaştırma yapılabilir. Buna göre **TX:** Ev Ekonomisi kitabını Afyon Meslek Yüksekokulu İşletme öğrencileri 6 gün içerisinde bu alandaki kitapları teslim etmiştir. İade günü 7 ve 9 gün olanlarda Sosyal Bilimler Enstitüsü İşletme bölümü Kadın öğrencileri Afyon Meslek Yüksekokulu ve Turizm İşletmeciliği ve Otelcilik Yüksekokulu İşletme bölümündeki Erkek öğrencilerden daha çok araştırma yaptığı tespit edilmiştir. Sosyal Bilimler Enstitüsü İşletme bölümündeki Erkek ve Turizm İşletmeciliği ve Otelcilik Yüksekokulu İşletme bölümündeki Kadın öğrenciler TX kodlu kitapları 15 günden fazla iade etmeyerek cezaya kalmışlardır ancak kadın öğrenciler bu kitabı daha çok ödünç almışlardır. **Q:** Bilim alanındaki kitaplar İktisadi ve İdari Bilimler Fakültesi İşletme bölümündeki Erkek ve Sosyal Bilimler Enstitüsü İşletme bölümündeki Kadın öğrenciler tarafından birbirine yakın miktarda kitap ödünç alımında bulunmuşlardır. **HA:** İstatistikler araştırma alanı cinsiyet olarak karşılaştırıldığında ise; İktisadi ve İdari Bilimler Fakültesi İşletme bölümündeki Kadın öğrenciler Sosyal Bilimler Enstitüsü İşletme bölümündeki Erkek öğrencilerden daha çok araştırma yapmıştır. **TK:** Elektrik Mühendisliği. Elektronik. Nükleer Mühendislik alanındaki kitap en çok Sosyal Bilimler Enstitüsü İşletme bölümü öğrencileri tarafından kullanılmıştır. **TK:** Elektrik Mühendisliği. Elektronik. Nükleer Mühendislik ve aynı zamanda **HE:** Taşımacılık. Telekomünikasyon alanındaki kitaplar İktisadi ve İdari Bilimler Fakültesi öğrenciler tarafından iade günü limitini aştığı için cezaya kalmıştır.

Aşağıda tüm veri setine uygulanan J48 algoritması sonucunun ağaç görseli verilmiştir. (Şekil 4.1)



Şekil 4.1 Weka Sınıflandırma J48 Algoritmasının Fakülte/Enstitü Özniteliği Seçilerek Sonuçlarının Ağaç Görseli

Ağaç görseli çok detaylı olduğundan ve okunması amacıyla çok fazla yakınlaştırma gerektirdiğinden burada yer verilememiştir. Yakından incelendiğinde ağaçta bolumno, iade gün ve cinsiyete göre dallanmalar gözlenmektedir.

Kümeleme işlemleri için; Farthest First, Filtered Clusterer, Make Densit Based Clusterer ve Simple KMeans kümeleme analizlerinden faydalanılmıştır.

FarthestFirst kümeleme algoritmasının sonucu olarak iki küme oluşturulmuştur:

İlk kümede Sosyal Bilimler Enstitüsünde Özel Hukuk (YL) (TEZLİ) (ANADOLU UNV. ORTAK) bölümünde **K: Hukuk** alanındaki kitapları alan Erkek öğrenciler iade günü 15 günü geçen kitap ödünç almışlardır. Diğer kümede ise Afyon Meslek Yüksekokulu bölümlerine genel olarak bakıldığında **G: Coğrafya. Antropoloji. Turizm** alanındaki kitapları alan Kadın öğrenciler kitapları 13 günlüğüne ödünç almışlardır.

Filtered Clusterer, Make Densit Based Clusterer ve Simple K-Means kümeleme algoritmalarında rastgele(random) sonucu olarak iki küme oluşturulmuştur:

İlk kümede Afyon Sağlık Yüksekokulunda Fizyoterapi ve Rehabilitasyon bölümünde **R: Tıp** alanındaki kitapları alan Kadın öğrenciler iade günü 10 gün olarak kitap ödünç almışlardır. Diğer kümede ise Sosyal Bilimler Enstitüsünde Tarih bölümünde **B: Felsefe. Psikoloji. Din** alanındaki kitapları alan Kadın öğrenciler iade günü 3 gün olarak kitap ödünç almışlardır.

Filtered Clusterer, Make Densit Based Clusterer ve Simple K-Means kümeleme algoritmasının genel olarak sonucu Çizelge 4.12’de verilmiştir:

Çizelge 4.12 FilteredClusterer, MakeDensitBasedClusterer ve Simple K-Means Kümeleme Algoritmasının Ortak Sonucu

Öznitelik	Tüm Veri (402808.0)	0 (338352.0)	1 (64456.0)
kod	PL: Türk dili ve edebiyatı	PL: Türk dili ve edebiyatı	D: Tarih
iadegun	15 günden fazla	15 günden fazla	15 günden fazla
cinsiyet	Kadın	Kadın	Kadın
bolumno	Tarih	Türk Dili ve Edebiyatı	Tarih
okulno	Fen Edebiyat Fakültesi	Fen Edebiyat Fakültesi	Fen Edebiyat Fakültesi

Çizelge 4.12 incelendiğinde ödünç alınan kitaplar bütün hali ve iki kümeye ayrılmış hali görülmektedir. Tüm veriye bakıldığında en çok Fen Edebiyat Fakültesi Tarih bölümündeki kadın öğrenciler iade günü 15 günden fazla olacak şekilde PL: Türk dili ve edebiyatı alanında kitap ödünç aldığı görülmektedir. Ayrılan iki kümede Fen Edebiyat Fakültesi Türk Dili ve Edebiyatı ve Tarih bölümü karşılaştırıldığında; birinci kümeye bakıldığında Türk Dili ve Edebiyatı bölümünde Kadın öğrenciler iade günü 15 günü geçen şekilde PL: Türk dili ve edebiyatı alanında kitap ödünç almıştır. İkinci kümeye bakıldığında Tarih bölümünde Kadın öğrenciler iade günü 15 günden fazla olacak şekilde D: Tarih alanında kitap ödünç almıştır.

5. TARTIŞMA ve SONUÇ

Afyon Kocatepe Üniversitesi Merkez Kütüphanesinden öğrencilerin kitap ödünç alma davranışları veri seti olarak alınmıştır. Bu veri setinde kitap sınıflama kodu, iade günü, cinsiyet, bölüm ve fakülte/enstitü öznitelik olarak WEKA programına verilecek hale getirilmiştir. Daha sonra sınıflandırma ve kümeleme algoritmaları uygulanarak alınan sonuçların görselleştirilmesi sağlanmıştır. Bu çalışmada veri madenciliği yöntemleri kullanılmıştır. Kullanılan bu yöntemlerin içinde Sınıflandırma algoritmalarından Naive Bayes, J48 ve Kümeleme algoritmalarından Farthest First, Filtered Clusterer, Make Densit Based Clusterer ve Simple KMeans uygulanarak elde edilen sonuçlardan yüksek doğruluk oranına sahip ve uygulanabilir olması nedeniyle bazı modeller seçilmiştir.

Veri yığınlarının kullanılıp bir model geliştirilmesinde ve seçilen modelinde kullanılıp sonuçlarının daha önce bilinmeyen veri yığınları hakkında sonuçların tahmin edilmesi ile ilgili yöntemlere başvurulmuştur. Karar verme sürecinde tanımlayıcı yöntemler rehberlik edilmede kullanılabilecek veri örüntülerini tanımlamada kullanılmıştır. Dolayısıyla tahmin edici ve tanımlayıcı yöntemler kullanılarak kitap ödünç alma davranışlarından veriler arasında bir ilişki ve örüntü çıkarılmaya çalışılmıştır.

Weka programının görselleştirme uygulaması ile özniteliklerin veri seti içerisindeki dağılımları incelenmiştir. İlk 10 kitap sıralamasına bakılmıştır. En çok ödünç alınan ilk 5 kitap sınıfının sosyal bilimler alanında olduğu elde edilmiştir. Ödünç alınan kitaplar arasında teknik tek alanın (%3,94) yüzdesi ile “Matematik, İstatistik, Bilgisayar” alanı olduğuna varılmıştır. “Türk dili ve edebiyatı” ve “Tarih” alanları dışındaki alanların yüzde olarak payının düşük olduğu tespit edilmiştir. Daha sonra ilk 10 bölüm içerisinde herhangi bir teknik bölüm olmadığı gözlenmiştir. Bölümlerin Fen Edebiyat Fakültesi, İktisadi ve İdari Bilimler Fakültesi ve Eğitim Fakültesindeki bölümlerden oluştuğu tespit edilmiştir. İlk 10 kurum arasında ağırlığın sosyal alanlarda eğitim veren kurumlar olduğu görülmektedir. Ödünç alınan kitapların %35’lik bir bölümü “Fen Edebiyat Fakültesi” öğrencileri tarafından alınmaktadır. Buradan herhangi bir teknik bölüm olmadığı sonucuna varılmıştır. Genel olarak ödünç alınan kitapların çoğunluğunun kadın (yaklaşık %64) öğrenciler tarafından alındığı sonucu elde edilmiştir.

Veri madenciliği yöntemleri kullanılarak elde edilen sonuçlara bakıldığında, NaiveBayes algoritması ile üç ve üçten fazla alanda araştırma yapılan ödünç alınan kitaplar Coğrafya, İşletme, Türk Dili ve Edebiyatı, Biyoloji, Tıp, Tarih, Sosyoloji, Kamu Yönetimi, İnşaat Mühendisliği bölümü öğrencileri tarafından alındığı sonucu elde edilmiştir. Ödünç alınan kitapların sıklık sayısında ise en çok **D: Tarih** alanındaki kitapların 28880 sıklıkla Tarih bölümünde alındığı incelenmiştir. J48 algoritması kullanılarak öznitelik olarak kurumlar seçilmiştir. Bu algoritmanın sonucunda %92 doğruluk oranında sınıflandıma işlemini gerçekleştirdiği tespit edildiğinden bu öznitelik ile yapılan model seçilmiştir. Modelin sonuçları değerlendirildiğinde kitap kodu, iade günü ve cinsiyet kriterleri bulunmadan frekansı 1000'e yakın ve 1000'den yüksek olan bölümler incelendiğinde son 10 yılda sözel alanda Türk Dili ve Edebiyatı, Sosyoloji ve Tarih bölümü öğrencileri kütüphaneyi en çok kullanan ilk 3'tür. Sayısal alanda ise Elektrik, Kimya, Gıda, Biyomedikal, Harita, İnşaat, Makine, Elektrik-Elektronik Mühendislik bölümleri ve Tıp öğrencileri kütüphaneyi en çok kullananlar sıralamasındadır. Kitap kodu, iade günü ve cinsiyet olarak bakıldığında kurumlara göre İşletme bölümü örnek olarak alınmış ve bu kriterlere göre sonuçlara bakılmıştır. Genel olarak öğrencilerin 15 günlük iade gününü zaman zarfını aştıkları tespit edilmiştir.

Kümeleme işlemleri için; Farthest First, Filtered Clusterer, Make Densit Based Clusterer ve Simple KMeans kümeleme analizlerinden faydalanılmıştır. Farthest First kümeleme algoritmasının sonucu olarak iki küme oluşturulmuş. Birinci kümede Sosyal Bilimler Enstitüsünde Özel Hukuk (YL) (TEZLİ) (ANADOLU UNV. ORTAK) bölümünde K: Hukuk alanındaki kitapları alan Erkek öğrenciler iade günü 15 günü geçen kitap ödünç aldıkları ve diğer kümede ise Afyon Meslek Yüksekokulu bölümlerine genel olarak bakıldığında G: Coğrafya. Antropoloji. Turizm alanındaki kitapları alan Kadın öğrenciler kitapları 13 günlüğüne ödünç aldıkları sonucuna varılmıştır.

Literatürde konuyla ilgili yapılan araştırmalarda kullanılan veri madenciliği yöntemlerine bakıldığında;

Cano vd. (2019) öneri olarak; kümeleme algoritmalarının kullanımı, belirli profilleri tanımlamak için kullanıcıların karakterizasyonunu verebileceğini ve birliktelik kuralları tekniği ile, kullanıcı davranışlarının ilişkisini gösterebileceğini sunmuştur. Daha fazla araştırma ile, farklı kütüphanelerde belirli kalıpları bulmak için panel verilerinin analizi incelenmesini ve kullanıcı profili ile ilgili daha fazla özellik olması durumunda, varsayılan model riskinin muhtemelen belirtilen ödünç kitap alma süresi içinde başarısız olacak öğrencileri tespit etmek için kullanılabileceğini belirtmiştir.

Luo (2017), öğretmen ve öğrencilerin kitaplarda ne aradıklarını bulmalarını kolaylaştırmak için kütüphaneden DBSCAN algoritmasıyla veri toplamış ve bu algoritmayı kullanarak; hedeflenen kitapların önerilebilmesi için, kullanıcıların ilgi alanlarını ve hobilerini, profesyonel geçmişlerini ve diğer profillerini çıkarmak için veri örnekleri kümesini analiz etmiştir. Çalışmasında veri madenciliği için birliktelik kuralları analizi ve sınıflandırma analizi gibi diğer analiz yöntemlerinin yanı sıra küme analizini de kullanmıştır.

Demiral vd. (2017), veriler üzerinde ayırıcı hiyerarşik kümeleme tekniği ile 11 adet küme oluşturulmuş ve kümeleri kendi içinde homojen, kümeler arasında ise heterojen şeklinde ayırmıştır. Elde edilen istatistiksel veriler ile veri madenciliği yöntemlerinden görselleştirmeyi uygulayarak birbirinden farklı grafikler ve tablolar oluşturulmuş ve bunlardan elde edilen çıkarımlar ile kitap ödünç alma davranışları incelemiştir.

Siguenza-Guzman vd. (2015) çalışmasında akademik kütüphanelere uygulanan veri madenciliği teknikleri için kapsamlı bir literatür taraması ve sınıflandırma yöntemi kullanılmış ve bunun için, 1998-2014 dönemi boyunca kırk bir pratik katkı tespit etmiş ve bunlarla doğrudan ilgisi olan nedenler gözden geçirmiştir. Her makale veri madenciliği yöntemlerinden kümeleme, birliktelik kuralları, sınıflandırma ve regresyona göre kategorize edilmiş ve bunlar hizmetler, kalite, koleksiyon ve kullanım davranışı şeklinde dört ana kütüphane alanına uygulanarak sonuç elde etmiştir.

Chen (2014) çalışmasının sonucu olarak; model oluşturma, üniversite kütüphanelerinin geliştirilmesi için daha iyi bir teorik temel sağlamış, böylece üniversite kütüphanelerinin uygulama becerisini geliştirmiştir.

Zhang (2011) çalışmasında uygulama üç açıdan incelemiş; uygulamanın anlamının ortaya çıkartılması, uygulamadaki temel problemlerin analiz edilmesi ve sonunda karar ağacı algoritmasının uygulanmasını araştırmıştır.

Sonuç olarak literatürdeki araştırmalara bakıldığında; birliktelik kuralları, sınıflandırma ve kümeleme yöntemlerinin yanı sıra bazı veri madenciliği teknikleride kullanılmıştır. Ayrıca elde edilen veriler üzerinde yapılan bu çalışmada, literatürdeki çalışmalar ile benzer seviyede başarı oranlarının elde edildiği görülmüştür.

Bu çalışma ile kütüphane sisteminin öğrencilerin ihtiyaçları doğrultusunda hangi bölümlerin kütüphanede araştırmalarını daha çok yaptığını ve öğrencilerin ihtiyaçlarına göre kitap getirilmesi ve popüler olan alanlarda sıralamalarını ona göre yapıp öğrencilerin kitaplara kolaylıkla ulaşmasını sağlayabileceklerdir. Böylece öğrenciler hem kitapların kullanım alanlarına göre daha kolay ulaşacaklar ve hem de zamandan tasarruf edeceklerdir.

Bir diğer açıdan bakarsak diğer üniversite kütüphaneleri de bu veri madenciliği yöntemlerini uygulayarak öğrencilerin kitaplara erişimini kolaylaştırabilir ve ödünç alma ile ilgili işlemleri daha yakından takip edebilir. Dolayısıyla öğrenciler hakkında bilgi sahibi olabilir ve tahminlerde bulunarak öğrencinin kitap alma davranışına yön verebilir ve katkı sağlayabilir. Bununla birlikte kitap öneri sistemi oluşturulabilir, öğrencilerin iade günleri alışkanlıkları tahmini ortaya çıkartılabilir, hangi kitabın diğer kitaplarla uyumlu olduğuna bakılabilir, öğrencilerin sayısına ve kütüphanedeki davranışlarına bakılarak kullanım oranları hakkında bilgi sahibi olunabilir, cinsiyete göre kullanım oranları karşılaştırılabilir, kitap ödünç alma davranışına göre öğrencinin genel not ortalaması başarısına sağlanan katkı ölçülebilir. Veri madenciliğini oluşturan disiplinleri ile öğrenmesi teknikleri ile öneri sistemleri geliştirilebilir, ödünç alma cezasına yakalananlar tespit etme geliştirilebilir, kitap kodlarına göre azlık/çokluk miktarı tespit edilebilir ve cinsiyete göre kütüphane içerisinde popüler olan kitap tespit edilebilir, bölümlere ve kurumlara göre yüzdelik dağılım yıllara göre görselleştirilebilir, şirketlerin korkusu olan müşteri kaybı düşünülerek kütüphane ilgi çekici hale getirilebilir.

6. KAYNAKLAR

- Adriaans, P. and Zantinge, D. (1996). Data Mining. Harlow: Addison Wesley Longman, England.
- Agrawal, R., Imielinski, T. and Swami, A. (1993). Mining Association Rules Between Sets Of Items In Large Databases. ACM SIGMOD Conference on Management of Data, DC: ACM Press, Washington.
- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, **29(1)**: 1-22.
- Altaş, D., Gülpınar, V. (2012). Karar Ağaçları ve Yapay Sinir Ağlarının Sınıflandırma Performanslarının Karşılaştırılması: Avrupa Birliği Örneği. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, **14(1)**: 1 - 22.
- Anane, R. (2001). Data Mining and Serial Documents. Computers and the Humanities, UK.
- Arslan, H. (2008). Sakarya Üniversitesi Web Sitesi Erişim Kayıtlarının Web Madenciliği ile Analizi. Yüksek Lisans Tezi, Sakarya Üniversitesi, Fen Bilimler Enstitüsü, Sakarya.
- Ayık, Y. Z., Özdemir, A., Yavuz, U. (2007). Lise Türü ve Lise Mezuniyet Başarısının Kazanılan Fakülte ile İlişkisinin Veri Madenciliği Tekniği ile Analizi. *Sosyal Bilimler Enstitüsü Dergisi*, **10(2)**: 441-454.
- Aydın, E., Keleş, M. (2017). Breast cancer detection using K-nearest neighbors data mining method obtained from the bow-tie antenna dataset. *International Journal of RF and Microwave Computer-Aided Engineering*, **1**: 149-155.
- Aydın, A., Aydın, E. (2017). Evaluation of Limestone Layer's Effect for UWB Microwave Imaging of Breast Models Using Neural Network. *Technical Journal*, **11**:1-2.
- Aynekin, G. (2006). İnternet İçerik Madenciliğinde Yapay Sinir Ağları ve Bir Uygulama. Yüksek Lisans Tezi, Uludağ Üniversitesi Fen Bilimler Enstitüsü, Bursa.
- Baykasoğlu, A. (2005). Veri Madenciliği ve Çimento Sektöründe Bir Uygulama. Akademik Bilişim Konferansları, 82-83.
- Bergeron B. (2002). Bioinformatics Computing. Prentice Hall PTR, United States of America.

- Bhambri, V. (2011). Application of Data Mining in Banking Sector. *International Journal of Computer Science and Technology (IJCST)*, **2(2)**: 199-202.
- Biçer, M. B., Aydın, E. A., Akdağlı, A. (2014). Meme Kanseri Tespitinde Mikrodalgaların Önemi ve Kanserli/Sağlıklı Meme Dokularının Yapay Zeka Algoritmaları ile Tanımlanabilmesi. *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, **29(2)**: 27-38.
- Bounsaythip, C., Rinta-Runsala, E. (2001). Overview of Data Mining for Customer Behaviour Modeling. VTT Information Technology Research Report TTEI, Finland.
- Bozkır, A. S. (2009). OLAP ve Veri Madenciliği Teknolojilerinden Yararlanılarak Web Tabanlı Bir Karar Destek Sisteminin Gerçekleştirilmesi. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen Bilimler Enstitüsü, Ankara.
- Bozkır, A. S., Mazman, S. G., Sezer, E. A. (2010). Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case. 2nd International Symposium on Information Management in a Changing World, Hacettepe University, Ankara, 145-152.
- Caldas, C., Soibelman, L., Han, J. (2002). Automated Classification of Construction Project Documents, *J. Comput. Civ. Eng.*, **16(4)**: 234-243.
- Cano et al. (2019). A Case Study of Library Data Management: A New Method to Analyze Borrowing Behavior, Springer International Publishing, Universidad del Pacífico, Lima, Peru.
- Chen, Y. (2014). Analysis on Approach of Data Mining in Personalized Information System of University Library. School of Educational Science, Zhoukou Normal University, Zhoukou, China, **10(21)**: 12804-12809
- Çiftçi, S. (2006). Uzaktan Eğitimde Öğrencilerin Ders Çalışma Etkinliklerinin Log Verilerinin Analiz Edilerek İncelenmesi. Yüksek Lisans Tezi, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Daş, B. ve Türkoğlu, İ. (2014). DNA Dizilimlerinin Sınıflandırılmasında Karar Ağacı Algoritmalarının Karşılaştırılması. Elektrik, Elektronik, Bilgisayar ve Biyomedikal Mühendisliği Sempozyumu (ELECO 2014), 27-29 Kasım, 381-383.

- Demiral, G., Soba, M., Armutlu, Ş. (2017). Kütüphane Veri Tabanında Veri Madenciliği: Uşak Üniversitesi Örneği. *Bartın Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, **8(16)**: 241-264.
- Doğan, O. (2017). Türkiye’de Veri Madenciliği Konusunda Yapılan Lisansüstü Tezler Üzerine Bir Araştırma. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, **19(3)**: 929-951.
- Dolgun, M. O., Özdemir, T. G., Deliloğlu, S. (2007). Öğrenci Seçme Sınavında (ÖSS) Öğrencilerin Tercih Profillerinin Veri Madenciliği Yöntemleriyle Tespiti. Bilişim 07 Kongresi, Ankara.
- Dunham, M.H. (2003). Data Mining Introductory and Advanced Topics. Prentice Hall, Pearson Education Inc., New Jersey.
- Durairaj, M., Ranjani, V. (2013). Data Mining Applications In Healthcare Sector: A Study. *International Journal Of Scientific & Technology Research*, **2(10)**: 29-35.
- Durmuşoğlu A. (2017). Veri Madenciliği Çalışmaları Üzerinde Bir Analiz: Türkiye Adresli Yayınlar. *Elektronik Sosyal Bilimler Dergisi*. **16(62)**: 1111-1122.
- Elomaa, T. and Kaariainen, M. (2001). An analysis of reduced error pruning. *Journal of Artificial Intelligence Research*, **15**: 163–187.
- Erten, H. (2015). Veri Madenciliği Teknikleri İle Organ Nakli İçin Uygun Donör Oranının Hesaplanması. Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Fayyad, U., Stolorz, P. (1997). Data Mining and KDD: Promise and Challenges. Future Generation Computer Systems, USA.
- Ganti, V., Gehrke, J. and Ramakrishnan, R. (1999). Mining Very Large Databases, IEEE Computer, USA.
- Getoor, L. (2003). Link Mining: A New Data Mining Challenge. ACM SIGKDD Explorations Newsletter, New York, NY, USA, 84-89.
- Goebel, M., Gruenwald, L. (1999). A Survey of Data Mining and Knowledge Discovery Software Tools. ACM SIGKDD Explorations, New York, NY, USA, 20-33.
- Han, J. And Kamber, M. (2001). Data Mining : Concepts and Techniques. Academic Press. 3. edition, USA.
- Han, J. and Kamber, M. (2006). Data Mining: Concepts and techniques (Second edition). San Francisco, 5, 10, 12-20, 21, 25, 47-49, 67, 86, 327, 649-654, 658-659.

- İşler, Y., Narin, A. (2012). WEKA Yazılımında K-Ortalama Algoritması Kullanılarak Konjestif Kalp Yetmezliği Hastalarının Teşhisi. *Süleyman Demirel Üniversitesi Teknik Bilimler Dergisi*, **2(4)**: 21-29.
- Karataş, A. (2004). Yapay Sinir Ağları ve Finans Alanına Uygulanması. *Atatürk Üniversitesi, İktisadi ve İdari Bilimler Dergisi*, **18(3)**: 337-349.
- Kaya, M., Keleş, A. E., Oral, E. (2013). Construction Crew Productivity Prediction by Using Data Mining Methods. Proceedings of the 4th World Conference on Learning, Teaching and Educational Leadership, Procedia - Social and Behavioral Sciences, **141**: 1249-1253.
- Kaya, M., Özel, S. A. (2014). Açık Kaynak Kodlu Veri Madenciliği Yazılımlarının Karşılaştırılması (Comparison of Open Source Data Mining Software). Akademik Bilişim Konferansı (AB2014), AB2014 Proceedings, 47-53.
- Keleş, A. E. (2016). İnşaat Projelerinde Şantiye Şeflerinin Liderliği ve Çalışan Motivasyonu İlişkisinin Veri Madenciliği ile Belirlenmesi. Doktora Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana.
- Keleş, A. E. (2016). İnşaat Sektöründe Veri Madenciliği Uygulamalarına Genel Bakış ve Ekonomik Etkilerinin Yorumlanması, International Congress of Management Economy And Policy Special Issue, İstanbul, 55-61.
- Keleş, A. E., Kaya, M. (2014). Duvar İnşa Edilmesinde Verimliliği Etkileyen Faktörlerin Apriori Veri Madenciliği Yöntemi Kullanılarak Analizi (The Analysis of the Factors Affecting the Productivity in the Wall Construction of the Using Apriori Data Mining Method). Akademik Bilişim Konferansları, AB 2014 Proceedings, 831-836.
- Keleş, M. (2017). An Overview: The Impact of Data Mining Applications on Various Sectors. *Technical Journal*, **11(3)**: 128-132.
- Keleş, M., Keleş, A. E. (2017). Veri Madenciliği Uygulamalarının ve Sezgisel Optimizasyon Algoritmalarının Yapım Yönetimindeki Yeri (The Place of Data Mining Applications and Heuristic Optimization Algorithms in Construction Management). *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, **32(1)**: 235-242.

- Kumar, Dharminder and Bhardwaj (2011). Rise of Data mining: Current and Future Application Areas. *IJCSI International Journal of Computer Science Issues*, **1**: 256-260
- Kurt, Ç., Erdem, A. (2012). Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi, *Politeknik Dergisi*, **15(2)**: 111-116.
- Liaoa, C. W., Perng, Y. H. (2008). Data Mining for Occupational Injuries in The Taiwan Construction Industry. *Safety Science*, **46(7)**: 1091-1012.
- Liew, A. (2007). Understanding Data, Information, Knowledge And Their Inter-Relationships. *Journal of Knowledge Management Practice*, **7(2)**: 1-10.
- Luo, L. (2017). Application of Data Mining in Library-Based Personalized Learning. Chongqing Radio & TV University, Chongqing, China. **12(12)**: 127-133.
- Onat, A. (2008). Veri Madenciliğinin Web Tabanlı Uygulamalarda İnsan Uyumluluklarının Tespiti Üzerine Bir Çalışma. Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya.
- Orhunbilge, N. (1996). Uygulamalı Regresyon ve Korelasyon Analizi. İstanbul Üniversitesi İşletme Fakültesi Yayınları, İstanbul.
- Özçakar, N. (1998). Genetik Algoritmalar. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, **1**: 69.
- Özel, C., Topsakal, A. (2014). Veri Madenciliği Kullanarak Beton Basınç Dayanımının Belirlenmesi. *Cumhuriyet Üniversitesi Fen Fakültesi Fen Bilimleri Dergisi*, **35(1)**: 43-57.
- Özkan, Y. (2008). Veri Madenciliği Yöntemleri. Papatya Yayıncılık Eğitim, İstanbul.
- Pulakkazhy, S., Balan R.V.S. (2013). Data Mining In Banking and Its Applications-A Review. *Journal of Computer Science*, **9(10)**: 1252-1259.
- Saruhan, H. (2004). Genetic Algorithms: An Optimization Technique. *Teknoloji*, **7(1)**: 105.
- Savaş, S., Topaloğlu, N., Yılmaz, M. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, **21**: 1-23.
- Seven, M. F. (2009). Veri Tabanlarında Bilgi Keşfi: Veri Madenciliği ve Bir Sağlık Uygulaması. Yüksek Lisans Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.

- Shahbaz, M., Rahman, M. (2008). Data mining for engineering sector in pakistan: Issues and implications. Proceedings of the World Congress on Engineering and Computer Science 2008 (WCECS 2008), San Francisco, USA.
- Siguenza-Guzman, L., Van Den Abbeele, A., Vandewalle, J., Verhaaren, H., Cattrysse, D. (2015). A holistic approach to supporting academic libraries in resource allocation processes. *The Library Quarterly: Information, Community, Policy*, **85(3)**: 295–318.
- Silahtaroglu, G. (2008). Kavram ve Algoritmalarıyla Temel Veri Madenciliği. Papatya Yayıncılık Eğitim, İstanbul.
- Thuarisingham, B. M. (2003). Web Data Mining and Applications in Business Intelligence and Counter Terrorism. Auerbach Publishers, Incorporated, Boca Raton, FL, USA.
- Tsiptsis, K., Chorianopoulos, A. (2009). Data Mining Techniques in CRM: Inside Customer Segmentation. John Wiley & Sons, Ltd..
- Vatansever, M., Büyüklü, A., H. (2009). Using Visual Data Mining Techniques in Clustering Analysis and An Application. *Mühendislik ve Fen Bilimleri Dergisi*, **27**: 86.
- Wang, J, Hu, X., Zhu, D. (2008). Data Mining in Public Administration. Handbook of Research on Public Information Technology, IGI Global, 556-567.
- Wilmot, C. G., Cheng, G. (2003). Estimating Future Highway Construction Costs. *J. Constr. Eng. Manage*, **129 (3)**: 272-279.
- Zhang, M. (2011). Application of Data Mining Technology in Digital Library. *Linyi, Shandong, China Journal of Computers*, **6(4)**: 761- 768.

Internet Kaynakları

1. <https://searchsqlserver.techtarget.com/definition/database-management-system> 05.03.2019
2. <https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd> 06.03.2019
3. <https://www.techopedia.com/definition/1184/data-warehouse-dw> 07.03.2019
4. <http://blog.emretopcuoglu.com/wp-content/uploads/2016/07/etl.png> 08.03.2019
5. <https://www.britannica.com/technology/data-mining> 04.04.2019
6. <https://mlf6w4ol0fdr.i.optimole.com/w:auto/h:auto/q:auto/https://www.lifeofmedica.com/wp-content/uploads/2019/02/tarihsel-1.png> 05.04.2019
7. <https://www.kdnuggets.com/2016/12/poll-analytics-data-mining-data-science-applied-2016.html> 06.04.2019
8. <http://tugbaguder.blogspot.com/2014/05/veri-madenciligi-ve-kullanilan.html> 07.04.2019
9. <https://docplayer.biz.tr/docs-images/67/57913215/images/32-0.jpg> 05.05.2019
10. <http://mgocenoglu.blogspot.com/2014/06/veri-madenciligi-asamalar.html> 10.05.2019
11. https://cdn-images-1.medium.com/max/1600/1*tWaaZX75oumVwBMcKN-eHA.png 14.05.2019
12. <https://tr.wikipedia.org/wiki/Weka> 12.01.2019

ÖZGEÇMİŞ

Adı Soyadı : Ahmet Gökhan GÜREL
Doğum Yeri ve Tarihi : İstanbul, 1990
Yabancı Dili : İngilizce
İletişim (Telefon/e-posta) : 554 862 40 41 ahmet.gokhan.gurel@gmail.com

Eğitim Durumu (Kurum ve Yıl)

Lise : Küçükçekmece Anadolu Lisesi, İstanbul (2004 - 2008)
Lisans : Kocaeli Üniv. Fen-Edebiyat Fak. Matematik, Kocaeli
(2009 - 2013)
Lisans : Karabük Üniv. Müh. Fak. Bilgisayar Müh., Karabük
(2013 - 2016)
Sertifika : Karabük Üniv. Eğitim Fak. Pedagojik Formasyon Eğitimi
Programı Sertifikası, Karabük (2014-2015)

Çalıştığı Kurumlar ve Yıl : Safranbolu Bilgi Evi, Matematik Öğretmenliği
Karabük (2014 – 2015)
Kurumsal Bilgi Sistemleri A.Ş. (Abas),
Yazılım Geliştirici, İstanbul (Kasım 2017 – Mart 2019)

EKLER

Ek 1. Kütüphane Resmi Yazı

Evrak Tarih ve Sayısı: 20/12/2018-E.55461



T.C.
AFYON KOCATEPE ÜNİVERSİTESİ
Kütüphane ve Dokümantasyon Daire Başkanlığı

Sayı : 35062616-302.08.01-
Konu : Ahmet Gökhan GÜREL'in Veri Talebi
Hk.

FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 05/12/2018 tarihli ve 95624635-302.08.01-E.53076 sayılı yazınız.

Enstitünüz İnternet ve Bilişim Teknolojileri Yönetimi Anabilim Dalı öğrencisi Ahmet Gökhan GÜREL'in "Üniversite Kütüphanesi Verileri Üzerinde Veri Madenciliği Yöntemlerinin Uygulanması" konulu tez çalışmasında kullanılmak amacıyla Birimimizden talep etmiş olduğu "2009-2018 yılları arasında öğrencilerin kitap ödünç alma bilgileri" elektronik olarak hazırlanmış olup, öğrencinin kendisi veya Tez Danışmanı tarafından Birimimizden teslim alınması gerekmektedir.

Bilgilerinizi ve gereğini arz ederim.

e-imzalıdır
Mustafa AKYÜZ
Daire Başkanı Vekili

20/12/2018 Mem.
20/12/2018 Kütüp.Şb.Müd.

Ayşe BAŞHAN
Fatma ÇAKIR

Evrak Doğrulamak İçin : <https://ebys.aku.edu.tr/enVision/Dogrula/L54SZIHH>

Adres: Afyon Kocatepe Üniversitesi Ahmet Necdet Sezer Kampüsü
Telefon: 444 03 03 | Faks: 0 272 228 1181
e-Posta: kutuphane@aku.edu.tr

Bilgi için: Ayşe Başhan
Unvanı: Memur



Bu belge 5070 sayılı Elektronik İmza Kanununun 5. Maddesi gereğince güvenli elektronik imza ile imzalanmıştır.