

T. C.
İstanbul Üniversitesi
Sosyal Bilimler Enstitüsü
Bilgi ve Belge Yönetimi Anabilim Dalı

Yüksek Lisans Tezi

Bilgi ve Belge Yönetiminde Veri Madenciliği

Ahmet AKÇAY

2501080646

Tez Danışmanı

Doç. Dr. Ümit KONYA

İstanbul, 2014



Y Ü K S E K L İ S A N S
T E Z O N A Y I

ÖĞRENCİNİN

Adı ve Soyadı : Ahmet AKÇAY Numarası : 2501050645
Anabilim/Bilim Dalı : Bilgi ve Belge Yön. A.B.D. Danışman Öğretim Üyesi : Doç. Dr. Ümit KONYA
Tez Savunma Tarihi : 16.01.2014 Tez Savunma Saati : 11.00 'da
Tez Başlığı : “ Bilgi ve Belge Yönetiminde Veri Madenciliği”.

TEZ SAVUNMA SINAVI, Lisansüstü Öğretim Yönetmeliği'nin 35. Maddesi uyarınca yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin KABULÜ'NE OYBİRLİĞİ / ~~OYÇOKLUĞUYLA~~ karar verilmiştir.

JÜRİ ÜYESİ	İMZA	KANAATI (KABUL / RED / DÜZELTME)
1-Prof. Dr. Tuba KARATEPE		Kabul
2-Doç. Dr. Ümit KONYA		Kabul
3-Yrd. Doç. Dr. Işıl SERT		Kabul

YEDEK JÜRİ ÜYESİ	İMZA	KANAATI (KABUL / RED / DÜZELTME)
1- Doç. Dr. Oğuz İÇİMSOY		
2- Yrd. Doç. Dr. Rifat GÜNALAN		

Bilgi ve Belge Yönetiminde Veri Madenciliği

Ahmet Akçay

ÖZ

Veri miktarının her geçen gün artması bu verilerin doğru analizi sorununu ortaya çıkarmaktadır. Verilerden ilginç ve potansiyel olarak verimli bilgilerin çıkarılması, araştırılması ve keşfi önemli bir faktördür. Bilgi keşfi olarak adlandırılan veriden, enformasyon ve bilgiye erişilen süreç içerisinde yer alan veri madenciliği, son yıllarda dünyada olduğu gibi Türkiye’de de hızla yaygınlaşmaya başlayan disiplinlerarası bir çalışma olarak göze çarpmaktadır. Veri madenciliği, anlamsız görülen büyük veri yığınları içindeki gizli ilişkileri, desenleri tespit etmek ve elde edilen bilgiyi yararlı bir sonuç üretmek için kullanmaktır.

Bu çalışmanın temel amacı; Bilgi ve Belge Yönetimi disiplini içerisinde yer alan veriden, bilgiye erişim süreci içinde veri madenciliğinin yerini belirlemek, konuyla ilgili kavramları incelemek, kullanım alanları hakkında bilgiler vermek ve anlaşılabilir düzeyde veri madenciliği yöntem ve tekniklerini anlatmaktır. Bu amaç doğrultusunda çalışmada veri, enformasyon, bilgi kavramları açıklanmış, veri hazırlama süreci detaylarıyla incelenmiş ve veri madenciliği kavramı anlatılmıştır. Son olarak da veri madenciliği yöntem ve tekniklerine değinilmiştir.

Anahtar Kelimeler: Veri, Enformasyon, Bilgi, Veri Madenciliği, Bilgi Keşfi

Data Mining in Information and Records Management

Ahmet Akçay

ABSTRACT

The exponential growth in amount of data causes reliable data analysis problems. The purpose of extracting, exploring, discovering interesting, and potentially beneficial knowledge from these data is a very critical factor. Data mining, within the process of transforming data into information and then into knowledge, called knowledge discovery, has become a more common and as well as an interdisciplinary study around the world in the last few years. Data Mining is discovering hidden dependencies, unknown patterns in huge amounts of seemingly meaningless data and producing useful results with the collected information.

The main purpose of this thesis is to study and bring up the place and importance of Data Mining in the process of knowledge discovery, to provide a point of view about the concepts related to Data Mining, to give information on application areas and to introduce the methods and techniques of data mining basically. In the light of this purpose, in this study first of all the concepts of data, information and knowledge explained theoretically, then analyzing the data preparation in detail, and the concepts of data mining are described. Finally methods and techniques of data mining has been examined.

Keywords: Data, Information, Knowledge, Data Mining, Knowledge Discovery

ÖNSÖZ

Son yıllarda bilgi teknolojilerindeki gelişmeler bilginin yapısal anlamda değişmesine de neden oldu. Tarihsel süreçte değişim yaşayan bilgi, eskiden taş, papirüse, kağıda yazılırken artık sayısal olarak kaydedilmeye başlandı. Bununla beraber bilginin niceliği de arttı. Bu artış sonucunda ilerleyen süreçte verinin kontrol edilememeye tehlikesi ortaya çıktı. Büyük çaptaki verilerin içinden nitelikli, işe yarar bilgiye erişimin yolları aranmaya başladı. Bu noktada öne çıkan veri madenciliği birçok disiplin için kullanılmaya başlandı. Özellikle ticari işletme ve kurumlar tarafından benimsenen veri madenciliği yöntem ve teknikleri, hızla yaygınlaşmaya başladı.

Bilginin nitelik ve niceliğinin gelişmesiyle birlikte, tarih boyunca en doğru bilgiye erişim sağlayan bilgi merkezleri de değişim içerisine girdi. Bilgi artık fiziksel olarak değil, bilgisayar teknolojilerinin geliştirdiği cihazlarda sayısal olarak yer almaya başladı. Bilgi ve bilgisayar teknolojileri bilgiyi düzenleyen, erişimini sağlayan bilgi ve belge yönetimi disiplininin de bakış açısını değiştirdi. Amacı genel olarak bilgi gereksinimlerine karşılık vermek olan bu disiplin, teknolojik gelişmeler doğrultusunda yeniden yapılandırıldı.

Bu çalışmada savunulan düşünce de bilgi ve bilgisayar teknolojileriyle değişen, gelişen bilgi ve belge yönetimi disipliniyle veri madenciliği kavramının ilişkilendirilmesidir. Veriden enformasyona, bilgiye ve bilgeliğe dönüşen süreç, bu disiplin içerisinde zaten yer almaktadır. Veri madenciliği ise bu süreçte önemli bir yere sahiptir. Bu noktadan yola çıkarak çalışmamda veri madenciliğini sade, teknik ayrıntıya girmeden, anlaşılabilir biçimde yazmaya çalıştım. Konuyu olabildiğince kısıtlayarak ve anlaşılması zor teknik terimlerden, formüllerden arındırarak en rahat anlaşılacak sadelikte işlemeye çalıştım. Kaynak bulma sırasında, konunun özellikle sosyal bilimler alanında henüz yeni yeni duyulması nedeniyle bazı sıkıntılar yaşadım. Daha çok mühendislik ve işletme/iktisat bilimlerinin kaynaklarından yararlandım. Konu hakkında verimli Türkçe kaynak bulmak da pek olanaklı olmadığından daha

çok yabancı kaynaklara başvurdum. EBSCOhost, JSTOR Art&Sciences, Elsevier veritabanlarını taradım. ULAKBİM Ulusal Veritabanları'ndan Türkçe kaynaklar bulmada yararlandım. Jiawei Han ve Micheline Kamber'in Data Mining Concepts and Techniques kitabını, konunun işleyiş sürecini belirlemek için ana kaynak olarak kullandım.

Konuyu üç bölümde inceledim. Veri madenciliğini anlatabilmek, yerini ve önemini ortaya koyabilmek için öncelikle veri, enformasyon ve bilgi kavramlarını tanımlamam gerekiyordu. İlk bölümde bununla birlikte bilgi keşfi sürecindeki önemli bir aşama olan veri hazırlama sürecini anlattım. İkinci bölümde veri madenciliğini detaylı biçimde anlatırken, veri madenciliğinin kullanıldığı yerlerle örneklendirdim. Son bölümde veri madenciliği yöntem ve tekniklerini inceledim. Tezin en teknik kısmını oluşturan bu bölümde, daha anlaşılır anlatabilmek için yararı olmayacağını ve daha çok karışıklığa neden olacağını düşündüğüm bazı teknik ve yöntemlere yer vermedim.

Teze başlarken danışmanım olan, üniversite hayatımın başladığı yıllardan itibaren öğrenim hayatım boyunca bana çok yardımcı olan, bu konuyu işleyiş biçimimde beni yönlendiren, üzerimde büyük katkısı olan değerli hocam Prof.Dr.Ayşe Üstün'e teşekkür ederim.

Tezimi başarıyla bitirmemi sağlayan, yönlendirmeleriyle bana çok yardımcı olan, tezimi titizlikle inceleyerek, eleştirileriyle daha iyiye yol almamı sağlayan danışman hocam Doç.Dr.Ümit Konya'ya teşekkür ederim.

Tezime konu olan düşüncenin oluşmasında yardımcı olan ve sadece tezi yazarken değil hayatımın her döneminde bana katkı sağlayan ve hayatımın yönlenmesinde çok büyük rolü olan abime, maddi manevi desteğini hiç esirgemeyen aileme ve hayatının zor zamanlarının bir kısmında yanında olmadığım ablama teşekkür ediyorum.

İÇİNDEKİLER

Öz	iii
Abstract	iv
Önsöz	v
İçindekiler	vii
Tablolar	ix
Kısaltmalar	x
Giriş	1
1. BİLGİ KAVRAMI VE BİLGİ KEŞFİ SÜRECİ.....	6
1.1. Tanımlar	6
1.1.1. Veri.....	7
1.1.2. Enformasyon	9
1.1.3. Bilgi.....	11
1.2. Bilgi Keşfi Süreci	16
1.2.1. Veri Hazırlama	23
1.2.1.1. Veri Temizleme	25
1.2.1.2. Veri Birleştirme	26
1.2.1.3. Veri Dönüştürme.....	27
1.2.1.4. Veri İndirgeme.....	28
1.2.2. Veri Madenciliği	31
1.2.3. Yorumlama ve Değerlendirme	31
2. VERİ MADENCİLİĞİ TANIMI, SÜRECİ VE UYGULAMA ALANLARI	32
2.1. Veri Madenciliği Tarihçesi.....	32
2.2. Veri Madenciliği Tanımı	34
2.3. VERİ MADENCİLİĞİ UYGULAMA SÜRECİ	40
2.3.1. Problemin Tanımlanması	41
2.3.2. Modelin Kurulması ve Değerlendirilmesi.....	42
2.3.3. Modelin Kullanılması ve İzlenmesi	44

2.4. VERİ MADENCİLİĞİ UYGULAMA ALANLARI.....	45
2.4.1. Pazarlama	49
2.4.2. Finans	51
2.4.3. Tıp ve Sağlık	53
2.4.4. Haberleşme ve Endüstri	55
2.4.5. Bilgi ve Belge Yönetimi.....	57
3. VERİ MADENCİLİĞİ YÖNTEM VE TEKNİKLERİ.....	61
3.1. Veri Madenciliği Yöntemleri	61
3.1.1. Tahmin Edici Yöntemler	62
3.1.1.1. Sınıflandırma	63
3.1.1.2. Regresyon	66
3.1.2. Tanımlayıcı Yöntemler	67
3.1.2.1. Kümeleme	68
3.1.2.2. Birliktelik Kuralı.....	70
3.2. Veri Madenciliği Teknikleri.....	78
3.2.1. Karar Ağaçları	78
3.2.2. Yapay Sinir Ağları	80
3.2.3. Genetik Algoritmalar.....	81
3.2.4. Veri Görselleştirme	83
Sonuç ve Değerlendirme	85
Kaynakça	90

TABLO VE GRAFİKLERİN LİSTESİ

Şekil 1. Bilgi Hiyerarşisi	14
Şekil 2. Bilgi Keşfi Süreci.....	22
Şekil 3. Veri Hazırlama (Önişleme) Biçimleri.....	30
Şekil 4. Veri Madenciliği Uygulama Süreci	40
Şekil 5. Veri Madenciliği Uygulama Süreci	42
Şekil 6. Veri Madenciliği Yöntemleri	61
Şekil 7. Doğrusal Regresyon Analizi	66
Şekil 8. Bir Kümeleme Yöntemi Örneği	69
Şekil 9. Market Sepet Analizi	73
Şekil 10. Örnek Bir Karar Ağacı Yapısı	79
Tablo 1. Bilgi Hiyerarşisi	15
Tablo 2. Bilgi Hiyerarşisine Kavramsal Bakışlar.....	16
Tablo 3. Sınıflandırma Yöntemiyle Hastalık Teşhisi.....	64
Tablo 4. Sebze reyonundan yapılan alışverişler.....	75
Tablo 5. Birlikte alınan ürünler	76
Tablo 6. Birlikte satın alınan ürünlerden üretilen kurallar	76

KISALTMALAR LİSTESİ

A.e. Aynı eser

A.g.e. Adı geçen eser

Bkz. Bakınız

KA Karar Ağaçları

OLAP Online Analytical Processing

SQL Structured Query Language

VM Veri Madenciliği

VTBK Veri Tabanlarında Bilgi Keşfi

YSA Yapay Sinir Ağları

v.b. Ve benzeri

v.s. Vesaire

GİRİŞ

Tarihin ilk zamanlarından günümüze kadar önemini kaybetmeyen ve sürekli değerlendirilen bir şey varsa, o da hiç şüphesiz bilgidir. Bilginin elde tutulması, korunması tarihin her döneminde farklı yöntemlerin kullanılması ile sağlanmıştır. Taş tabletlerden papirüse, oradan kağıda ve en son olarak sayısal ortama aktarılan bilginin korunması için o çağın gelişen teknolojisinden de yararlanılarak çeşitli sistemler geliştirilmiştir. Günümüzde bu teknolojik sistemler sayesinde veri toplamak ve saklamak kolaylaşınca, veriyi analiz etme ve bununla beraber sadece gerekli bilgiyi saklama düşüncesi doğmuştur.

İçinde bulunduğumuz çağda günlük hayatta yaptığımız her iş, her davranış neredeyse kontrol altında tutulmaktadır. İnsanlar bilerek veya bilmeyerek her yerde kendileri hakkında tanımlayıcı öğeler, belirli belirsiz veriler bırakmaktadır. Bilgisayar ve iletişim teknolojilerindeki gelişmeler, veri toplama araçlarının yardımıyla veri tabanlarında çok miktarda verinin hızlı toplanmasına, depolanmasına olanak sağlamıştır. Tutulması gereken verilerin miktarının ve çeşitliliğinin artması bu verilerin çözümlenmesini, işlenmesini zorunlu kılmıştır.

Yeni teknolojilerden önemli ölçüde etkilenen yeni dünya düzeninin kuralları gereği, bilgi ve zaman boyutlarının önemi çok daha artmış ve dolayısıyla karar verme süreçlerinde doğru ve anlamlı bilgiye dayalı hızlı karar alma her zamankinden daha fazla ön plana çıkmıştır. Verinin analizinin daha hızlı yapılması ve sonuçta anlamlı ve eyleme yönelik bilgiler ortaya çıkarılması, yani verinin nitelikli bilgiye dönüştürülmesi değişen düzendeki yoğun rekabet ortamının bir gereğidir.

Bilgi keşfi süreci olarak adlandırılan, veri yığınından anlamlı bilgi elde etme sürecinde ortaya çıkan kavramlardan biri de veri madenciliğidir. Büyük miktarda verinin çeşitli yöntemler ile analiz edilmesi ve ortaya çıkan sonuçların raporlanması, değerlendirilmesi ve yorumlanmasıyla geçmiş verilerden gelecek tahminleri yapmaya yarayacak bilgiler elde edilmesi olarak söz edeceğimiz veri madenciliği,

karar verici tarafından ortalama kararlar yerine özgün kararlar verilmesini destekleyen, satışları, kârlılığı, yenilikçiliği ve kaynak kullanımında etkinliği artıran önemli bir yönetim aracıdır.

Veri madenciliği, son yıllarda dünyada olduğu gibi Türkiye’de de hızla yaygınlaşmaya başlayan disiplinlerarası bir alan olarak göze çarpmaktadır.

Aslında bilgi yeni bir şey değildir, ancak bilgiyi başlı başına bir kurumun kaynaklarından biri olarak görme düşüncesi yenidir. Bilgi toplumlarında, üretimin temel unsurları olan sermaye, emek ve doğal kaynakların yanında artık bilgi ve teknoloji de yer almaktadır. Hatta bilgi bu üç unsuru yönlendiren, bir araya getiren önemli bir unsur olmuştur. Üretimin diğer ögesi olan girişimcinin beslenme kaynağıdır.

Artık insanlık bilgiye sahip olmakla yetinmeyip, bilgiye en hızlı şekilde erişme, doğru anlam kazandırma ve bilgiyi en doğru biçimde kullanma arayışlarına da başlamıştır. Bu açıdan bilgiye doğru anlam vermek onun kullanımını ve yönetimini daha etkin kılacaktır. Bu nedenle kurum ve işletmelerde bilginin yönetilmesi rekabet ortamının önemli unsurlarından biri durumuna gelmiştir. Günümüzde işletmeler bilgiyi daha iyi kullanma ve işleme yönünde adımlar atmaktadırlar.

Bilginin hem bireyler hem de örgütler için ulaştığı önemli ve etkili konum sonucunda sosyal, kültürel, ekonomik ve teknolojik bütün unsurlar doğrudan etkilenmektedir. Özellikle bilginin ekonomik bir değer taşıması günümüz tüketim toplumlarının gereksinimlerini karşılamak üzere kullanılmasının gerekliliği, bireylerin, örgütlerin veya işletmelerin bilgiyi elde etmelerini ve verimli olarak kullanmalarını bir zorunluluk olarak ortaya çıkarmıştır. İş ve örgüt yaşamında ilgi görmeye başlayan bilgi yönetimi uygulamaları, bugün yaygınlığını arttırarak devlet kurumlarında ve üniversitelerde de uygulanmaya çalışılan bir disiplin olarak kendisini göstermektedir.

Gelişen ve değişen çevre koşulları ile teknolojinin de yardımıyla bilgiye erişimde sınırların kalkması sonucu örgütler veya işletmeler küreselleşen ve sürekli canlı, hareketli nitelendirilebilecek bir çevreyle karşı karşıya kalmışlardır. Bu nedenle işletmeler rekabet ortamında bilgiyi örgütsel işleyişlerinde daha etkin kılacak farklı pazarlama ve araştırma geliştirme becerileri kazanıp bunları korumak ve sürdürmek amacıyla bilgi yönetim stratejileri geliştirmiştir.

Örgüt ve işletmelerin bilgi yönetim stratejileri belirlemeye yönelmesiyle son zamanlarda adından sıkça söz ettirmeye başlamış bir kavram olarak “iş zekası” ortaya çıkmıştır. Bilgi keşfi sürecinin teknik altyapısını oluşturan iş zekası, kurumsal verilerin bir takım analiz sürecinden geçirilerek, kurumun işleyişine yönelik yeni kararlar alınmasına veya alınacak kararlara destek olunmasına yarar sağlayacak bir süreçtir. İşletmelerin iş zekasına yönelmelerinin temelinde farklı kaynaklardan toplamış oldukları verinin çokluğu ama buna karşılık işe yarar bilginin azlığı gösterilebilir. Kurumsal öngörüler sonucunda ürün veya hizmet sunduğu kitlenin beklentilerini belirleme ve karşılama olanağı sunması başka bir neden olarak gösterilebilir. Ayrıca iş zekası araçlarının maliyetlerinin eskiye göre daha ucuz olması ve daha kullanışlı araçlar üretilmesi de önemli bir etkidir. Daha önce sadece akademik alanda kullanılan iş zekası için gerekli veri madenciliği yöntemlerinin artık yazılım araçları ile kolaylıkla kullanılabilir olması iş zekasının yaygınlaşmasında önemli rol oynamıştır.

Günümüz endüstrisinin ilgi odağı olan veri madenciliği, kurumsal iş zekası ürün ailesinin üyelerinden biridir. Bilgi keşfi sürecinin anahtar bileşeni olarak kullanılmaktadır. Kurumsal kararların alınmasında önemli bir yere sahiptir. Kamu kuruluşların veya özel işletmeler, “Müşteri İlişkileri Yönetimi” (CRM), “Kurumsal Kaynak Planlaması” (ERP), “İnsan Kaynakları” (İK) gibi çeşitli uygulamalar ve teknikler aracılığıyla veri madenciliği yapmaktadır. Artık bir banka müşterisinin aldığı krediyi geri ödeyip ödeyemeyeceğini, bir market müşterisinin bir sonraki alışverişinde hangi ürünü alacağını, bir hastalığa neden olan genlerin hangileri

olduğunu ve hastalığın nasıl gelişim göstereceğini veya kamu güvenliğini sağlama amacı ile olası güvenlik sorunlarını önceden bilmek şaşırtıcı değildir.

Bir telekomünikasyon kuruluşu, ürün ve servislerini kullandırdığı müşterilerinin geçmiş aramalarını ve telefon kullanımlarını inceleyerek onlara ürünlerini daha çok kullandırabileceği yeni kampanyalar önerebilir, böylece müşteri bağlılığını ve karını artırabilir, bir finans kuruluşu, müşterilerinin özellik ve davranışlarını inceleyerek batık kredi oranını azaltabilir, müşteri kaybını engelleyebilir, bir sağlık kuruluşu, uyguladığı tedavi yöntemlerinin başarısını irdeleyerek hangi hastalıklarda nasıl bir tedavi yöntemi izleyeceğini belirleyebilir, bir süpermarket, müşterilerinin satın alma eğilimlerini irdeleyerek kampanyalarını belli müşterilere yönlendirebilir, emniyet birimleri, suç istatistiklerine bakarak hangi kişilikteki insanların suça meyilli olduğunu tespit ederek, onların davranışlarını inceleyebilir, vergi kuruluşları, şirketler için risk modelleri kurarak vergi incelemelerini daha etkin yönlendirip, vergi kaçaklarını azaltabilir, internet üzerinden servis sunan bir web sayfası, ziyaretçilerinin site üzerindeki gezintilerini inceleyerek, onların ilgisini çekebilecek yeni bağlantılar ve ürünler tavsiye edebilir veya bir araştırma kütüphanesi, herhangi bir konuda araştırmacıların arama yaptığı ve yararlandığı yayınları inceleyerek, onlara konuyla ilgili yeni yayınların tavsiyesinde bulunabilir.

Bilgiyi, çeşitli araç ve yöntemler kullanarak ortaya çıkaran, oluşturan ve ona değer katan hiç kuşkusuz insandır. Her alanda, verilen kararların doğruluğu bilgiyi doğru çözümlenip kullanabilen ve bu kararı veren kişinin yeteneklerine ve deneyimlerine bağlıdır. Birçok örgüt ve işletme doğru bilgi teknolojisini uygulamadıklarından, bilgi yönetiminden gerektiği kadar yararlanamamıştır. Bu süreçte bilgiyi iş haline getiren, bilgi üreten, düzenleyen, sunan işletmelerin kurulması önem kazanmıştır.

“Bilgi ve Belge Yönetimi” disiplini bu alandaki insan gücü gereksinimini karşılayan önemli bir bölüm durumuna gelmiştir. Önceleri sadece kütüphane ve arşiv

gibi statik (durağan) bilginin yer aldığı kurumlardaki iş gücünü yönlendirirken, artık bilginin kullanıldığı her alanda söz sahibi olmaktadır. Bilgi keşfi sürecinin keşfedicilerini yetiştirmektedir.

Bu çalışma veri madenciliği konusunda yazılmış varolan literatüre bir alternatif değildir. Çalışmanın temel amacı; konuyu giriş seviyesinde incelemek, veriler ile neler yapılabildiğini göstermek, veri madenciliğinin, veri, enformasyon, bilgi denkleminde nerede yer aldığını saptamak, veri madenciliğinin nerelerde kullanıldığını aktarmak, anlaşılabilir düzeyde veri madenciliği yöntem ve tekniklerinden söz etmek ve en önemlisi veri madenciliğinin “Bilgi ve Belge Yönetimi” disiplini altında incelenmesi gerektiğini savunmaktır.

Tezin hipotezi, “veri madenciliği veri, enformasyon, bilgi hiyerarşisinde önemli bir yere sahiptir ve Bilgi ve Belge Yönetimi disiplini içerisinde incelenmelidir” şeklinde belirlenmiştir.

Veri madenciliğini anlamak için öncelikle bu kavramın temelini oluşturan veri, enformasyon ve bilgi terimlerinin ne olduklarının ve içeriklerinin iyi anlaşılması gerekmektedir. Çalışmanın birinci bölümünde kavramsal olarak bilgi hiyerarşisine değinilmektedir. Ayrıca verinin sayısal olarak depolandığı veritabanı terimi tanımlanmakta, bununla beraber bilgi keşfi süreci ve bu sürecin aşamaları incelenmektedir. Özellikle veri hazırlama süreci detaylarıyla anlatılmaktadır.

Çalışmanın ikinci bölümünde bilgi keşfi sürecinin en önemli aşaması olan veri madenciliğine yer verilmektedir. Veri madenciliği tanımı yapılırken farklı bakış açıları değerlendirilerek karşılaştırmalar yapılmıştır. Veri madenciliği sürecinden kısaca söz edilerek, veri madenciliğinin günümüzde yaygın olarak kullanıldığı alanlar örnekleriyle anlatılmıştır. Ayrıca bu bölümde, veri madenciliğinin henüz kullanımı az olsa da kütüphane ve arşivlerde nasıl kullanılacağına değinilmiştir.

Çalışmanın üçüncü bölümünde konu teknik boyutuyla incelenmekte, veri madenciliği yöntem ve teknikleri anlatılmaktadır. Daha iyi anlaşılması açısından mümkün olduğunca teknik terimlerden, sayısal anlatımdan kaçınılmış, konu sadeleştirilerek ve basitleştirilerek değerlendirilmeye çalışılmıştır.

1. BİLGİ KAVRAMI VE BİLGİ KEŞFİ SÜRECİ

1.1. Tanımlar

Yaşadığımız yüzyılı bilgi çağı olarak düşündüğümüzde günlük yaşamda, gerek basılı gerekse elektronik ortamda sıklıkla karşılaştığımız veri, enformasyon ve bilgi sözcüklerinin tanımlanması, aralarındaki hiyerarşinin belirlenmesi, bilgi keşfi sürecinin neresinde olduğunun bulunması ve ayrıca çalışmamızda yer alacak veri madenciliği kavramının anlaşılması bakımından önem taşımaktadır. Veri madenciliğinin tanımını yapabilmek, kavramsal olarak inceleyebilmek ve veri-enformasyon-bilgi döngüsü içerisinde nerede olduğunu belirleyebilmek için öncelikle bu kavramların arasındaki sistematik ilişkinin iyice anlaşılması gerekir.

Bu kavramların günlük yaşamda çevremizde salt birer sözcük gibi kullanılması, anlam daralmasına ve terminolojik açıdan sıkça karıştırılarak birbirlerinin yerine kullanılması anlam kaymasına neden olmaktadır. Enformasyon ve iletişim teknolojilerinin hızla gelişmesi ve bu gelişimin kaynağının daha çok yurtdışı olması, dilimize giren yabancı dilde sözcüklerin çoğalmasını da beraberinde getirmekte, bunları Türkçe'ye çevirmek ve yeni sözcükler türetmek zorunluluğu kavramlar arasında anlaşmazlığa yol açmaktadır.

Günümüze kadar birçok kişi tarafından veri, enformasyon ve bilgi kavramları farklı kategorilerde incelenmiş, tartışılmış, bu kavramlara doğru anlamı kazandırma arayışları içerisinde çeşitli değerler yüklenerek değerlendirilmiştir.

Geniş bir ölçekte karşılaştığımız kavramların, birbirleriyle bağlantılı ama farklı özelliklere sahip unsurlarla birlikte tanımlanması gerekir. Bu ayrıma dikkat edilmediğinde yanlış anlaşılmalara neden olan yorumlar yapılabilir.¹

¹ İsmet Barutçugil, **Bilgi Yönetimi**, İstanbul, Kariyer Yayıncılık, 2002, s.57.

İlgincidir ki bilgi hiyerarşisi kavramından ilk olarak ne “Bilgi Yönetimi” ne de “Bilgi Bilimi” disiplinlerinde söz edilmiştir. Cleveland,² bu hiyerarşinin ilk kez Amerikalı şair T. S. Eliot’un 1934 yılında sahnelenen ‘Choruses from The Rock’ isimli oyununda;

*“Yaşarken yitirdiğimiz yaşam nerede? Bilgide yitirdiğimiz bilgelik nerede? Enformasyonda yitirdiğimiz bilgi nerede?”*³

dizeleriyle hayat bulduğunu belirtir.

Günümüzde hala tartışılmaya devam eden, yeni yaklaşımlarla yeniden incelenerek tanımlanan ve bilgi keşfi sürecinin yapı taşları olan bu kavramların, terminolojik ve kavramsal açıdan tanımlanması veri madenciliğinin bu süreçte nerede olduğunu anlamak açısından yararlı olacaktır.

1.1.1. Veri

Veri, anlam kazanmamış harf, rakam, simge, işaretler gibi ham, birbirleriyle ilişkilendirilmemiş, özümlememiş, işlenmemiş gerçeklerdir,⁴ enformasyon parçacıklarıdır.⁵ Herhangi bir içerikten yoksun formlardır.⁶

Bazen fiziksel bir olaydır, yorumlanmamış gözlemlerdir.⁷ Örneğin esen rüzgar, yağan yağmur, uzaktan gelen bir ses veridir.

² Harlan Cleveland, “Information as Resource”, **The Futurist**, C.XVI, No:6, 1982, s. 34.

³ Şiirin orijinal dizeleri; “Where is the Life we have lost in living? Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?”

⁴ Barutçugil, **a.g.e.**, s.57.

⁵ Ali Akgün, Halit Keskin, “Sosyal Bir Etkileşim Süreci Olarak Bilgi Yönetimi ve Bilgi Yönetimi Süreci”, Gazi **Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, C.V, No:3, 2003, s.3.

⁶ D. C. Misra, Rama Harihan, Manie Khaneja, “E-Knowledge Management Framework for Government Organizations”, **Information Systems Management**, C.XX, No:2, 2003, s.39.

Yorum taşıyıcılar ancak işlenmek için hazır dırlar. Karar verme sürecinde tek başlarına etkili değillerdir, sürece sadece destek olurlar.⁸

Veri, bilgi hiyerarşisinin en alt basamağındadır.⁹

Bu tanımlardan farklı olarak Dervişođlu, veriyi tanımlamadan önce hiyerarşide yer almayan ancak veriden de önce gelen işareti tanımlar. İşaret, bir anlam ifade etmeyen harf veya rakamlardan oluşan, bazen de sadece özel bir belirtiden oluşan ifadelerdir. Veriden önce gelmesinin nedeni, işaretlerin birbirinden bağımsız olması veya herhangi bir şekilde birbirleriyle ilişkili olmamasından kaynaklanır. Veri, bir veya bir dizi işaretin yanyana gelmesi ve işlenmeye hazır olması ancak tek başına kullanım amacı taşımasıdır.¹⁰

Kurumsal bakış açısıyla ise, kurumun amacına bağılı olarak yararlı olacağı düşünülerek, henüz işlenmeksizin saklanan, depolanan kayıtlardır.¹¹

İşletmelerde veri, bilgi sisteminin ham malzemesi olmaktan daha fazla anlam taşır. Örgütle ilgili her türlü işlemlerin metin, görüntü, ses gibi kayıtlarından oluşan veri, karar verici uzmanlar tarafından kapsamlı olarak ele alınmaktadır.¹² Kurumsal veri işlemede, ham veri toplanır, temizlenir, var olan veri sistemleriyle bütünleştirilerek biçimi değiştirilir ve kolayca bulunabilecek sorgulanabilir, kullanılabilir biçimde veritabanlarında saklanır.

⁷ James A. O'Brien, **Introduction to Information Systems: Essentials for the Interworked EBusiness Enterprise**, 10.bs., USA, McGraw-Hill Higher Education, 2001, s.14.

⁸ Ufuk Durna, Yavuz Demirel, Bilgi Yönetiminde Bilgiyi Anlamak, **Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, No: 30, Ocak-Haziran 2008, s.129.

⁹ Mehmet Şahin, **Yönetim Bilgi Sistemi**, Eskişehir, Birlik Ofset, 2000, s.53.

¹⁰ H.Gökçe Dervişođlu, **Stratejik Bilgi Yönetimi**, Dışbank Kitapları-8, 2004, s.22.

¹¹ Barutçugil, **a.g.e.**, s.57.

¹² Şahin, **a.g.e.**, s.9.

Günümüzde kurumlar çok büyük miktarda ve giderek artmakta olan işlevsel ve etkileşimsel veriyi değişik formatlarda ve veritabanlarında biriktirmektedir.¹³

Veri kendisinin önemli olup olmadığı veya bir işe yarayıp yaramayacağı konusunda yol göstermez ancak enformasyon ve bilginin hammaddesi olduğundan önemlidir.¹⁴

Verinin değer kazanıp, kazanmayacağı önceden bilinemediği gibi, bir kurum için değerli olan veri, bir başka kurum için değerli olmayabilir.¹⁵ Bu durum bazen kurumlar için bir handicap oluşturmaktadır. Yararlı olacağı düşünülerek çok fazla veriyi elde tutmak yüksek kapasiteli veritabanları gerektirir ve bu da kuruma ayrıca maliyet getirir. Yararlı olduğu düşünülen verilerden de her zaman doğru kararlar çıkmayabilir. Bu bağlamda veri güvensizdir, istenilen sonucu vermeyebilir.

1.1.2. Enformasyon

Enformasyon sözcüğü, İngilizce'deki 'information' sözcüğünün Türkçe'ye uyarlanmış halidir. Basit tanımıyla veriye değer katılarak, verinin anlamlandırılmasıdır. Belli bir amaçla veya bir sorun çerçevesinde,¹⁶ birbiriyle ilişkili verilerin biraraya getirilmesi, düzenlenmesi¹⁷ sonucu oluşur ve bir mesaj taşır.¹⁸

Bu tanımla enformasyon bir iletişim kanalı içerisindedir ve haber niteliği taşır. Her haberin taşıdığı mesajın bir vericisi ve bir alıcısı vardır.

¹³ S. Sumathi, S. N. Sivanandam, **Introduction to Data Mining and its Applications**, New York, Springer, 2006, s.9.

¹⁴ Thomas H. Davenport, Laurence Prusak, **İş Dünyasında Bilgi Yönetimi: Kuruluşlar Ellerindeki Bilgiyi Nasıl Yönetirler**, Çev. Günhan Günay, İstanbul, Rota, 2001, s.22-23.

¹⁵ Amrit Tiwana, **Bilginin Yönetimi**, Çev. Elif Özsayar, İstanbul, Dışbank, 2003, s.84.

¹⁶ Dervişoğlu, **a.g.e.**, s.22.

¹⁷ Barutçugil, **a.g.e.**, s.57.

¹⁸ Davenport, Prusak, **a.g.e.**, s.24.

İletişim boyutuyla düşünüldüğünde enformasyon, alıcısındaki şüpheyi gideren, onda etki bırakan mesajın uygun bir kanal aracılığıyla iletilmesidir.¹⁹

Taşıdığı zengin içeriğe sahip yazılı, sözlü veya görsel mesajın hedefinde ulaştırdığı kişinin algısını, yargısını etkileme ve değiştirme vardır.²⁰

Enformasyondan kasıt, bir konu veya eylem hakkında alıcıyı görüş sahibi yapmak, oraya dikkatini çekmeye çalışmaktır. Enformasyon, mesajın alıcısı mesaj hakkında yorum yapabiliyorsa amacına ulaşmış olur, yoksa anlamsız kalmış olur. Bununla beraber enformasyonun zenginliği, mesajın içeriğine, güçlülük, zayıflık özelliğine ve iletiildiği kanala bağlı olarak değişir.²¹

Bir başka tanıma göre enformasyon; desenler, çağrışımlar, enformasyonu sağlayan veriler arasındaki ilişkililerdir. Örneğin, perakende satışların işlem hacimlerinin oluşturduğu verilerin analizi, hangi ürünün ne zaman satıldığını gösteren enformasyona ulaşmayı sağlayabilir.²²

Bu tanımın paralelinde kurumsal olarak bakıldığında enformasyon, anlamı olan ve yönetimin kontrolünde kullanılan bir veridir veya kurumsal veritabanıdır.²³

Verilere değer katarak enformasyona dönüştürme işlemi birkaç süreçten geçer. Bunlar; verilerin toplandığı amacı belirleme, amaca göre verileri sınıflandırma, birbirleriyle olan ilişkileri belirlemek için matematiksel veya istatistiksel olarak analiz etme, ilişkili olmayan verileri ayıklama, hataları, yanlışları

¹⁹ Oya Gürdal, **Tekstil Endüstrisinde Enformasyon Olgusu**, Ankara, Türk Kütüphaneciler Derneği, 2000, s.3.

²⁰ Robert A. Szymanski, Donald P. Szymanski, Donna M. Pulschen, **Computers and Information Systems**, USA, Prentice-Hall, 1995, s.12.

²¹ Durna, Demirel, **a.g.e.**, s.129.

²² Sumathi, Sivanandam, **a.g.e.**, s.9.

²³ Famil ŞAMILOĞLU, **Entelektüel Sermaye**, Ankara, Gazi Kitabevi, 2002, s.326.

düzeltilme, veriden en iyi şekilde yararlanılabilmesi için veriyi kısa, net, anlaşılır biçimde özetleyerek sunma şeklinde sayılabilir.²⁴

Kalseth ve Cummings'e göre ise bu süreci, veri; özetleme, düzeltme, hesaplama, sınıflandırma ve içerik işlemleriyle değer kazandırılarak enformasyona dönüştürülmektedir, biçiminde açıklamaktadır.²⁵

Enformasyon veriden doğmaktadır ve enformasyon da bilgiye dönüşmektedir.

1.1.3. Bilgi

Kullanımda veri ve enformasyondan daha popüler olan bilgi, geçmişten günümüze kadar tartışılmış, tanımlanmaya çalışılmıştır. Eski Yunan Uygarlığı'nda bilgiyi tanımlamak, anlayabilmek üzere çalışmalar yapıldığını biliyoruz. Modern felsefede bilgi kuramı (epistemoloji) temel konulardan biri olmuş, bilginin ne olduğu ve hangi kanallardan bilgiye ulaşıldığı tartışılmıştır.²⁶

Bilgi hiyerarşisinin en üst katmanında bulunan bilgi, belirli bir amaç için enformasyonun analiz ve yorumla zenginleştirilmesi, kişisel anlamda ise özümsemesi demektir.²⁷

Bilgi, insanın çevresinde olup biteni kavrayabilmesi için kişiselleştirdiği enformasyondur. Bilgi her bireyde kendini düşünceler, sezgiler, öngörüler,

²⁴ Davenport, Prusak, **a.g.e.**, s.25.

²⁵ Karl Kalseth, Sarah Cummings, "Knowledge Management: Development Strategy or Business Strategy?", **Information Development**, C.XVII, No:3, 2001, s.166.

²⁶ A. Kadir Çüçen, **Bilgi Felsefesi**, Bursa, Asa Kitabevi, 2001, s.30-31.

²⁷ Barutçugil, **a.g.e.**, s.57.

deneyimler, uygulamalar şeklinde gösterir.²⁸ Yani farklı bakış açılarıyla farklı bilgiye ulaşılabilir. Enformasyon bireylerin aklında işlendiği zaman bilgiye dönüşür.²⁹

Bilgi her ne kadar ikisiyle de ilişkili olsa da veri veya enformasyon demek değildir.³⁰

Enformasyonun bir biçimidir ve sadece bireylerin zihinlerindedir. Özne bir oluşumdur, genel olarak veri ve enformasyonun bireyler tarafından yorumlanmasıyla ortaya çıkar. Deneyimlerle kazanılır, zaman içinde yaşanan bazen başarılar bazen başarısızlıklardır.³¹ Kişinin kafasındadır, çoğu zaman metne veya söze dökmek bile olanaklı değildir.³² Bireyden bireye doğrudan transfer olamaz, ancak enformasyona dönüşerek iletilebilir. Zaten enformasyon bilginin iletilebilir ve kaydedilebilir biçimidir. Alavi'ye göre de bilgi; eğer metinler ve grafikler şeklinde sunulursa enformasyona dönüşür.³³

McDermott, bilgiyi enformasyondan ayıran altı özelliği şu şekilde sıralar:³⁴

1. Bilgi insan düşüncesinden kalanlardır.
2. Bilgi insan davranışdır.
3. Bilgi o anda oluşturulur
4. Bilgi toplumlara aittir.

28 A.e., s.10.

29 Maryam Alavi, "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issue", MIS Quarterly, C.XXV, No:1, 2001, s.109-111.

30 Davenport, Prusak, a.g.e., s.21.

31 Tiwana, a.g.e., 2003, s.84.

32 Gilbert Probst, Steffan Raub, Kai Romhardt, **Managing Knowledge: Building Blocks for Success**, New York, John Wiley & Sons Ltd. 2000, s.21-22.

33 Alavi, a.g.e., s.109-111.

34 Richard McDermott, "Why information technology inspired but cannot deliver knowledge management", **California Management Review**, C.XLI, No:4, 1999, s.105.

5. Bilgi toplum içinde çok çeşitli yollarla yayılır.
6. Yeni bilgi, eskisinin fonksiyonlarını yitirdiği yerde oluşturulur.

Bilgi, verinin karar verici için anlamlı ve kullanışlı biçime dönüşmesidir. Bu dönüşüm süreci aşağıdaki işlemlerden oluşur:³⁵

- Toplama
- Sınıflandırma
- Yeniden düzenleme
- Özetleme
- Saklama
- Yeniden elde etme
- İletme

Kurumsal anlamda bilgi, müşteriler, ürünler, süreçler hakkında sahip olunan enformasyon olarak belirtilmiştir.³⁶ Ancak yalnızca belgelerde veya arşivlerde değil rutin çalışmalarda, süreçlerde, uygulamalarda ve normlarda da kendini gösterir.³⁷

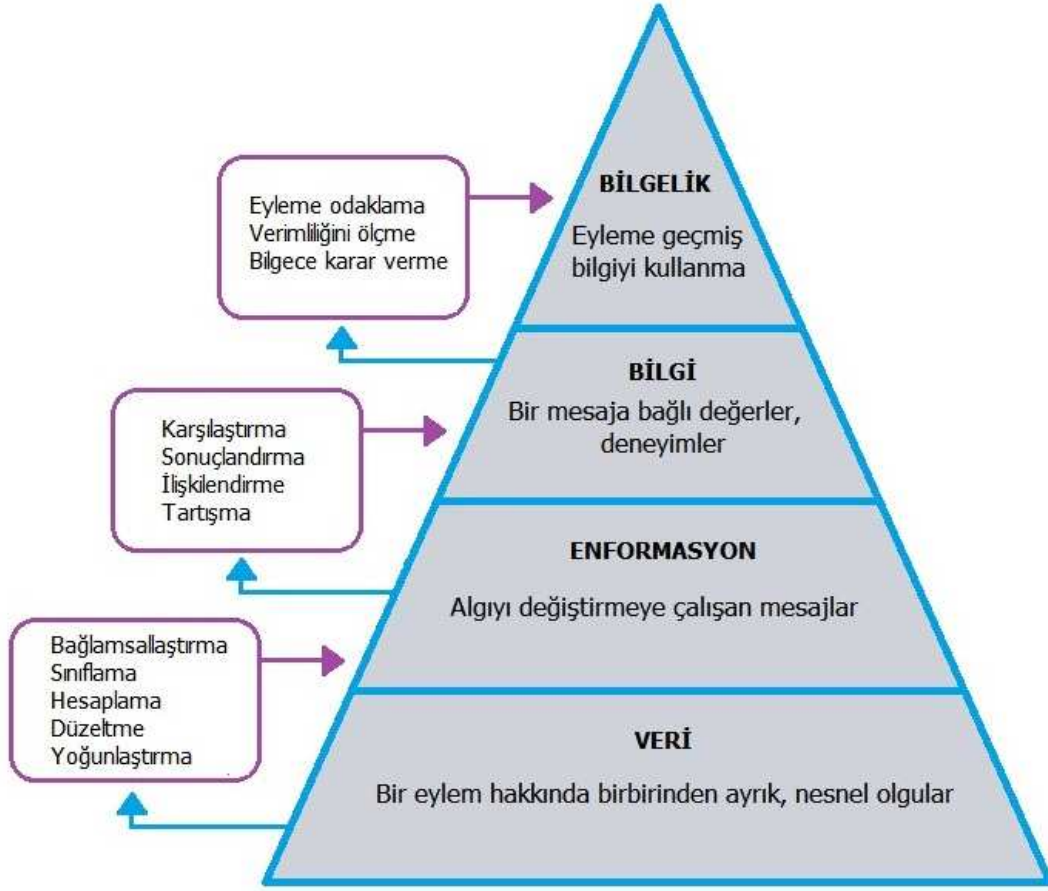
Certo, bilgi kavramını; veri olarak adlandırılan gerçeklerin, olayların ve istatistiklerin elde edilmesi ve bir örgütün işleyişi ile ilgili toplanan bu verinin bazı yöntemlerle incelenmesinden elde edilen sonuçlardır, biçiminde tanımlar.³⁸

³⁵ Uma G. Gupta, **Management Information System: A Managerial Perspective**, USA, West Pub. Co., 1996, s.3.

³⁶ Nalan K. Doğan, “Entellektüel Sermaye Yönetimi Yaklaşımı ve Entellektüel Varlıkların Korunmasına Yönelik Öneriler”, **İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadi Enstitüsü Dergisi**, No:47, 2004, s.16.

³⁷ Davenport, Prusak, **a.g.e.**, s.27.

³⁸ Samuel C. Certo, **Principles of Modern Management Function and Systems**, 4.bs., Boston, Allyn and Bacon, 1989, s.499.



Şekil 1. Bilgi Hiyerarşisi (Jay Leibowitz, The Knowledge Management Handbook, CRC Press LLC, 1999)

Davenport ve Prusak, bilgiye yönelik “veriler, kayıtlarda ve işlemlerde; enformasyon da mesajlarda bulunmaktadır. Buna karşılık bilgi, bireylerden veya bilenler grubundan veya bazı zamanlarda da kurumun rutin çalışmalarından elde edilmekte ve kitaplar ile belgeler gibi belli biçimlere sahip araçlarla (basılı ve elektronik enformasyon kaynaklarıyla) ve sohbetlerle, ustalık-çıraklık ilişkilerine kadar uzanan kişisel iletişim biçimleriyle aktarılmaktadır” açıklamasını getirir.³⁹

Veri, enformasyon ve bilgi hiyerarşisi üzerindeki tartışmalar devam etmekte, bu kavramlar farklı disiplinlerde farklı yaklaşımlarla incelenmektedir. Bunlar için

³⁹ Davenport, Prusak, **a.g.e.**, s.27-28.

kesinleşmiş tanımlar veya matematiksel olarak kanıtlanmış bağıntılar yoktur. Örneğin, Zeleny veriye sorulan nasıl sorusuna cevabı verenin enformasyon olduğunu iddia ederek bu hiyerarşiyi “know-nothing (veri), know-how (enformasyon), know-what (bilgi) , know-why (bilgelik)” biçiminde açıklarken,⁴⁰ Quigley ve Debons, nasıl sorusunun cevabının bilgi olduğunu söyler.⁴¹

Tablo 1. Bilgi Hiyerarşisi (Milan Zeleny, "Management Support Systems: Towards Integrated Knowledge Management," Human Systems Management)

	ÖRNEK	BİLME BİÇİMİ
VERİ	Elementler; H ₂ O, maya bakterisi, nişasta molekülleri	BİLİNMEZLİK (<i>Know-Nothing</i>)
ENFORMASYON	İçerik; un, şeker, su, baharat, çeşitli ekmek tarifleri	NASIL YAPACAĞINI BİLME (<i>Know-How</i>)
BİLGİ	Tariflerinden birinin seçimiyle ekmek yapma	NE YAPACAĞINI BİLME (<i>Know-What</i>)
BİLGELİK	Neden ekmek de ay çöreği değil?	NEDEN YAPACAĞINI BİLME (<i>Know-Why</i>)

Cleveland ise, bu hiyerarşiyi daha açık bir örnekle anlatır. İlkel bir mağara adamı tavşan, toprak, çimen, yağmur, bulut vs. nedir bilir. Bildiği bu enformasyonları seçer, düzenler ve bilgiye çevirir; tavşan çimeni yer, çimen toprakta yetişir, yağmur bulutlardan gelir, kurt tavşanı yer gibi. Ancak bu bilgileri birbirleriyle ilişkilendirip kendisi için daha yararlı duruma getirmediği bilgelğe ulaşamaz; eğer bütün kurtları öldürürsem, tavşanlar çimenleri yiyip bitirir ve toprak tertemiz olur.⁴²

⁴⁰ Milan Zeleny, "Management Support Systems: Towards Integrated Knowledge Management", Human **Systems Management** , C.VII, No:1, 1987, s.60.

⁴¹ Edward J. Quigley ve Anthony Debons, "Interrogative Theory of Information and Knowledge", **Proceedings of SIGCPR'99**, Los Angeles, ACM Press, New Orleans, 1999, s.4.

⁴² Cleveland, **a.g.e.**, s.34.

Veri, enformasyon, bilgi sürecine yönelik bazı yazarların farklı kavramsal bakış açılarını Tablo 2’de görebiliriz.

Tablo 2. Bilgi Hiyerarşisine Kavramsal Bakışlar (Dick Stenmark, “Information vs. Knowledge: The Role Of Intranets In Knowledge Management”, Proceedings Of The 35th Hawaii International Conference On System Sciences)

Yazar	Veri	Enformasyon	Bilgi
Wig	-	Bir durumu veya koşulu tanımlamak için düzenlenmiş olgular	Gerçekler ve inançlar, bakış açıları ve kavramlar, yargılar ve beklentiler, yöntemler ve nasıl yapacağını bilme
Nonaka ve Takeuchi	-	Anlamli bir mesajın dolaşımı	Mesajdan varılan kesin hükümler ve inanışlar
Spek ve Spijkervet	Yorumlanmamış simgeler	Anlamli olan veriler	Anlam verme yeteneği
Davenport	Basit gözlemler	İlgi ve amaç taşıyan veriler	İnsan zihninde oluşan değerli enformasyon
Davenport ve Prusak	Bir dizi birbirinden ayrı gerçekler	Alıcısının algısını değiştirmeye çalışan bir mesaj	Deneyimler, değerler, anlayışlar ve bağlamsal enformasyon
Quigley ve Debons	Herhangi bir soruyu yanıtlamak için oluşturulmamış metinler	Kim, ne zaman, nerede ya da ne sorularını yanıtlayan metinler	Neden ve nasıl sorularını yanıtlayan metinler
Choo, vd.	Olgular ve mesajlar	Anlam kazanmış veriler	Gerçekliği kanıtlanmış inanışlar

1.2. Bilgi Keşfi Süreci

İnsanlığın tarih boyunca en heyecan duyduğu şeyler şüphesiz keşif ve icatlardır. Her keşif ve icat, öncesinde yapılan sayısız deneylerle ortaya çıkmıştır. Bu deneyler sırasında elde edilen veriler yeni keşif ve icatlara yol göstermiştir. Teknolojinin gelişmesi ve bilgisayarın ortaya çıkmasıyla, deneme yanılma yöntemiyle veya belki de tesadüfen erişilen bilginin yerini artık hesaplanan ve ne olduğu bilinen veriler üzerinde yapılan çalışmalarla erişilen bilgi almıştır.

20. yüzyılın sonlarına kadar karar vericilerin yaşadığı en temel sorun veri kıtlığı, bilgi yetersizliği, var olan verilerin nerelerde depolanacağı ve bilgiye erişimdeki zorluklardı.⁴³ 21. yüzyılda ise sorunlar, bilgi fazlalığı ve bunun sonucunda yaşanan bilgi karmaşası, nitelikli bilginin seçilip saklanması, çok miktardaki bilgi içerisinden yararlı olana erişim zorluklarıdır.

Günümüz insanının yaptığı alışverişler, bankacılık işlemleri, telefon konuşmaları, kamera kayıtları, internet tanımlama bilgileri ve benzeri işlemlerin kayıtları tutulmakta ve saklanmaktadır. Örneğin, İngiltere üç milyonun üzerinde kapalı devre kamera sistemiyle izlenmekte, kentte yaşayan sıradan bir İngiliz günlük işlerini yaparken yaklaşık 300 defa kameralara yakalanmaktadır.⁴⁴ Bu tutulan kayıtlardan elde edilen veriler inanılmaz boyutlardadır. Sadece uydu ve diğer uzay araçlarından elde edilen anlık görüntülerin oluşturduğu bir saatlik verinin boyutu 50 gigabyte'ın üzerindedir.⁴⁵

Amerikan telekomünikasyon şirketi AT&T yıllık 70 milyarın üzerinde uzun mesafeli telefon konuşması sağlıyor. Petrol şirketi Mobil Oil, petrol araştırmalarıyla ilgili 100 terabaytın üstündeki verileri depolamayı hedefliyor. İngiltere'nin en büyük kredi kartı şirketi olan Barclaycard yılda 350 milyon işlem hacmine sahip olmasına rağmen, yılda 7 milyarın üzerinde işlem gören Amerikan perakende şirketi Wal-Mart ile karşılaştırıldığında bir hiç kalır.⁴⁶

Wal-Mart'ın bir saatte müşteri alışverişlerinden elde ettiği veri, Amerikan Kongre Kütüphanesi'ndeki 167 kitaba eşdeğer. 2000 yılında çalışmaya başlayan Sloan Digital Sky Survey teleskopunun, ilk haftasında topladığı veri bütün astronomi

⁴³ John Seely Brown, Paul Duguid, **Enformasyonun Sosyal Yaşamı**, Çev. İbrahim Bingöl, İstanbul, Türk Henkel Dergisi Yayınları, 2001, s.19.

⁴⁴ Jessica Williams, **Dünyada Değişmesi Gereken 50 Gerçek**, Çev. Yurdakul Gündoğdu, İstanbul, Aykırı Yayınevi, 2005, s. 251.

⁴⁵ Haldun Akpınar, "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", **İ.Ü.İşletme Fakültesi Dergisi**, C.XXIX, No:1, İstanbul, 2000, s.7.

⁴⁶ David J. Hand v.d., "Data Mining for Fun and Profit", **Statistical Science**, C.XV, No:2, 2000, s.111.

tarihinin o zamana kadar topladığı verinin üzerindeydi. 10 yılda toplanan veri ise 140 terabyte'tı.⁴⁷

Verilerin artmasıyla birlikte kitap üretimi de arttı. Türkiye Yayıncılar Birliği verilerine göre 2011 yılında kitap üretimi geçen seneye göre %21 arttı. Son yıllarda e-kitap kullanımının da yaygınlaşmasıyla birlikte kütüphanelerden e-kitap ödünç alan kullanıcı sayısı çoğaldı. Artık kullanıcıların bilgileri de kütüphane sistemlerinde birikmeye başladı.

Çok farklı alanlarda ortaya çıkan ve hızla artmakta olan bu büyük miktardaki veriler artık elektronik ortamda veritabanları veya veri ambarlarında toplanılmakta ve biriktirilmektedir.⁴⁸

Veritabanı, basit tanımıyla, verilerin depolanmasını, değiştirilmesini, silinmesini, erişilmesini kolaylaştırmak için sistematik olarak dosyalar biçiminde düzenlenmiş veri topluluklarını ifade eder.⁴⁹ Bir veritabanı sistemi veya veritabanı yönetim sistemi, birbirleriyle ilişkili ve benzer verileri biraraya getiren, verilere erişimi ve verilerin yönetilmesini sağlayan değişik tekniklerle tasarlanmış sistem ve yazılımlardır.⁵⁰

Veritabanı yönetimi sisteminin önemli bir parçası olan veritabanı yazılımlarına örnek olarak, kişisel veritabanı yazılımları; Microsoft Access, dBase, FoxPro, Paradox ve hatta Microsoft Excel, ilişkisel veritabanı yazılımları; Oracle, MySQL, PostgreSQL, Sysbase, Informix, Progress, Microsoft SQL Server, veritabanları üzerinde karmaşık analizler, raporlamalar yapabilen OLAP sistemi gibi çözümler verilebilir.

⁴⁷ "Data, Data everywhere: A Special Report on Managing Information", **The Economist**, 2010, (çevrimiçi) <http://www.economist.codm/node/15557443>, 16.12.2013.

⁴⁸ Alex A. Freitas, **Data Mining and Knowledge Discovery with Evolutionary Algorithms**, Almanya, Springer-Verlag, 2002, s.1.

⁴⁹ William J. Flawley, Gregory Piatetsky-Shapiro, Christopher J. Matheus, "Knowledge Discovery in Databases : An Overview", **AI Magazine**, C.XIII, No:3, 1992, s.57.

⁵⁰ Jiawei Han, Micheline Kamber, **Data Mining: Concepts and Techniques**, San Francisco, Morgan Kaufmann Publishers, 2001, s.10.

Veri ambarı, bir işletmenin veya kuruluşun farklı birimleri tarafından toplanan verilerden değerli olanlarının, gelecekte analiz işlemlerinde kullanılması amacıyla işletimsel sistem veritabanından farklı bir ortamda birleştirilmesinden oluşan büyük çaplı bir veri deposudur. İşletimsel sistemlerde var olan verilerin ayıklanması ve temizlenmesi, karar verme mekanizmalarına hizmet edecek şekilde hazırlanması, doğru şekilde saklanması, çeşitli yazılımlar aracılığıyla veriye erişilmesi ve belirleyici veri ilişkilerinin aranıp bulunması işlemlerinin tümünü içeren bir aktiviteler zinciridir. Veri ambarı kullanıldığında, günlük işletimsel görevlerle yeterince meşgul olan veritabanı kullanılmadan, analiz işlemleri farklı bir ortamda kolay, hızlı ve doğru biçimde yapılır.⁵¹

Veri ambarı, son yıllarda, bilgi üreticilerine (tepe yöneticiler, yöneticiler ve analistler) ve karar vericilere daha iyi ve daha hızlı karar alma imkanı tanıyan veri kaynaklarının entegre edilmesinde önemli bir teknoloji olmuştur.⁵²

Birçok işletme geçmiş deneyimlerini ve şimdiki eğilimlerini göz önüne alarak yapısını, kapasitesini geliştirmek istemektedir. Ancak işletmeler kendi etkinlik alanlarındaki hızlı değişimlere, gelişimlere ayak uydurabilme ve cevap verebilme bakımından baskı altındadır. Çalışmalarında doğru kararlar verebilmek için kendi veritabanlarında biriken verilere, kaynaklara en hızlı biçimde erişim sağlamaları gerekir.⁵³ Geleceğe yönelik sağlıklı, amaca yönelik kararlar verebilmek için geçmiş veriler üzerinde araştırma, inceleme yapmak ve bu verilerden gereksinim duyulan bilgileri çıkartabilmek artık işletme için yaşamsal önem taşımaktadır.

Veri toplama ve depolama alanında yaşanan teknolojik gelişme, karar verme süreçlerinde, geleneksel veri analizi yerine yeni eğilimlere yönelmiş, rekabet avantajı

⁵¹ Mehmet Ali ALAN, “Veri Madenciliği Ve Lisansüstü Öğrenci Verileri Üzerine Bir Uygulama”, **Dumlupınar Üniversitesi Sosyal Bilimler Dergisi**, No:33, 2012, s.165.

⁵² Rahman, Hakikur, **Data Mining Applications for Empowering Knowledge Societies**, Information Science Reference, New York, 2009, s.190.

⁵³ Pieter Adriaans, Dolf Zantinge, **Data Mining**, 2. bs., England, Addison-Wesley, 1997, s.25.

sağlayan, hızlı sonuç veren veritabanlı çözümlenin etkili olmasını sağlamıştır.⁵⁴ Güçlü veri analiz araçları olmaksızın verileri analiz etmek insanoğlunun yeteneklerini aşmış, karar vericiyi, veri zengini ancak bilgi fakiri konumuna sokmuştur.⁵⁵ Fayyad, verinin değerinin artık, ona ne kadar çok sahip olduğunuzla değil, onu ne kadar hızlı ve etkili keşfedip yönettiğinizle ölçüldüğünü söyler.⁵⁶

Verilerin toplanması, saklanması ve bunlardan bilgi çıkarılması gibi sorunlara öncelikle dosya sistemleri ve veritabanlarındaki bir takım gelişmelerle çözüm aranmıştır. Özellikle bilgisayar donanım ve yazılımlarının çoğalması ve fiyatlarının ucuzlaması bu çözümlere destek olmuştur. Toplanan verilerin çoğu zaman sadece bir kısmının yararlı olması, verilerin boyutlarının çok büyük olması, herhangi bir yazılımsal araç kullanmadan verilerin analizini ve karar destek aşamasında kullanılmasını olanaksız kılmıştır.⁵⁷ Ayrıca biriken veri miktarı çoğaldıkça, daha iyi veri analiz ve çözümlene tekniklerine gereksinim duyulmaktadır. Veritabanı analizi için SQL dili yetersiz kaldığı için daha gelişmiş programlama teknikleri yazılmaya başlanmıştır.

Bu noktada “Veritabanlarında Bilgi Keşfi” kavramı ortaya çıkmaktadır. Bu kavram ilk kez 1989 yılında, verinin işlenip bilgiye dönüşmesi süreci olduğu için “bilgi keşfi” olarak ortaya atılmıştır. Veriden anlamlı örüntüler çıkarma süreci literatürde, veri madenciliği, bilgi çıkarımı, bilgi keşfi, özbilgi keşfi,⁵⁸ veri

⁵⁴ Chidanand Apte v.d., “Business Applications of Data Mining”, **Communications of the ACM**, C.XLV, No:8, 2002, s.49.

⁵⁵ Han, Kamber, **a.g.e.**, s.4.

⁵⁶ Hamparsum Bozdogan, **Statistical Data Mining and Knowledge Discovery**, Boca Raton, London, New York, Chapman&Hall/CRC, 2004, s.1.

⁵⁷ Christopher Westphal, Teresa Blaxton, **Data Mining Solutions: Methods and Tools for Solving Real-World Problems**, U.S.A., Wiley & Sons, 1998, s.1.

⁵⁸ Ayşe Oğuzlar, **Veri Madenciliğine Giriş**, Bursa: Ekin Kitabevi, 2004, s.4.

arkeolojisi, veri örüntü işleme, veri tarama, veri analizi,⁵⁹ veri balıkçılığı⁶⁰ olarak da yer almıştır.⁶¹

Veritabanlarında veya veri ambarlarındaki saklı bilgiyi keşfedebilmek amacıyla gereksinim duyulan yeni nesil hesaplama teknikleri ve araçları, veritabanlarında bilgi keşfinin konusunu oluşturmuştur.⁶² Daha basit değerlendirmeyle bilgi keşfi, büyük verilerin analizindeki bilgisayar destekli işlemlerdir.⁶³

Bilgi keşfi birbirini takip eden şu basamaklardan oluşur:⁶⁴

1. Veri Temizleme (gürültülü ve tutarsız verileri temizleme)
2. Veri Bütünleştirme (birden fazla veri kaynağını birleştirme)
3. Veri Seçimi (veri tabanlarından analiz edilecek ilgili verilerin belirlenmesi)
4. Veri Dönüştürme (bir sonraki aşama için verileri uygun biçime dönüştürme)
5. Veri Madenciliği (veri desenlerini ortaya çıkartmak için uygulanan akıllı yöntemlerden oluşan temel bir süreç)
6. Desen Değerlendirme (bazı ölçümlere dayalı bilgiyi göstermek için ilginç desenleri belirleme)

⁵⁹ Ming-Syan Chen, Jiawei Han, Philip S. Yu, "Data Mining: An Overview from Database Perspective", **IEEE Transactions on Knowledge and Data Engineering**, C.VIII, No:6, 1996, s.866.

⁶⁰ David J. Hand, 'Data Mining: Statistics and More?', **The American Statistician**, C.LII, No:2, 1998, s.112.

⁶¹ Gregory Piatetsky-Shapiro, "Knowledge Discovery in Real Databases: A Report on the IJCAI- 89 Workshop", **AI Magazine**, C.XI, No:5, 1990, s.68-70.

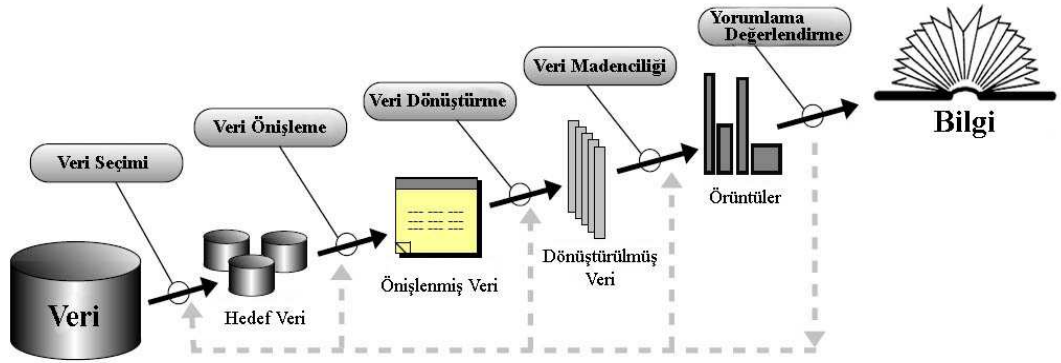
⁶² Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, "Knowledge Discovery and Data Mining: Towards a Unifying Framework", **Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)**, C.XXXII, AAAI Press, 1996, s.82.

⁶³ Sumathi, Sivanandam, **a.g.e.**, s.9.

⁶⁴ Han, Kamber, **a.g.e.**, s.7.

7. Bilgi Sunumu (keşfedilen bilgiyi kullanması için kullanıcının gözünde canlandırma ve bilgiyi sunma)

İlk dört aşama veri madenciliği aşamasına verinin hazırlanması için uygulanan veri ön işleme aşaması olarak değerlendirilir.



Şekil 2. Bilgi Keşfi Süreci (Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques, 2.bs., USA, The Morgan Kaufman Publishers)

Bazı kaynaklarda sürecin aşamalarının sayısı farklı olarak yer alsa da içeriği aynıdır. Sadece sınıflandırmada değişiklikler göze çarpmaktadır. Örneğin Fayyad, Piatetsky-Shapiro ve Smyth'e göre ise VTBK (Veri Tabanlarında Bilgi Keşfi) sürecinin aşamaları şunlardır:⁶⁵

1. Bilgi keşfi yapılacak alanı (veritabanı, veri ambarı) belirleme. Bu aşama uygulamayla ilgili önbilgiyi ve uygulamanın amacını da içerir.
2. Örnek veri kümesi oluşturma. Sorgu yapılacak örnek veri kümesi oluşturma üzerine odaklanma.
3. Veri temizleme ve veri ön işlemleri. Örnek veri kümesinde yer alan gürültülü ve tutarsız verileri belirleyerek çıkarma, hatalı verileri

⁶⁵ Usama Fayyad, Gregory Piatetsky-Shapiro ve Padhraic Smyth, "The KDD Process for Extracting Useful Knowledge from Volumes of Data.", **Communications of ACM**, C.XXXIX, No:11, 1996, s.30-31.

düzenleme, eksik ve bilinmeyen veri alanlarını doldurma, veritabanı yönetim sistemine karar verme, veri tipleri, veri taslakları oluşturma.

4. Veri indirgeme. Çalıştırılacak sorguların daha hızlı sonuç üretmesi için uygulamanın amacına göre verilerin nitelikleri belirlenerek, veri kümesindeki ilgisiz nitelikte ve tekrarlı verilerin çıkarılması, veri boyutunun azaltılması.
5. Veri madenciliğinin işlevini belirleme. Veri madenciliği yöntemlerinin (sınıflama, regresyon, kümeleme vb.) amaçlarını tanımlama, hangi yöntemlerle nasıl sonuçlara ulaşacağını öngörme.
6. Veri madenciliği yöntemini belirleme. Sorgulama için hangi veri madenciliği yöntemi ve tekniğinin uygun olduğu konusunda karar verme.
7. Veri madenciliği. Seçilen veri madenciliği yöntemini ve tekniğini uygulama, veriler arasındaki örüntüyü belirleme.
8. Yorumlama. Keşfedilen ilginç örüntüleri değerlendirme, aşamaların tekrar edilip edilmeyeceğine karar verme, ilgisiz ve gereksiz örüntüleri çıkarma ve kullanıcıların anlayacağı biçime getirme.
9. Keşfedilen bilgiyi kullanma. Bilgiyi çalışma alanıyla birleştirme, bilgiye dayalı eyleme geçme, bilgiyi belgeleyerek ilgili yere raporlama, daha önce keşfedilen, inanılan bilgiyle çelişkisini denetleme.

1.2.1. Veri Hazırlama

Verinin kalitesi, veri madenciliği için anahtar niteliği taşır. Veri madenciliğinden doğru, olumlu, işe yarar sonuçlar alınabilmesi için ham veriler önceden bir takım işlemlerden geçmelidir. Veri hazırlama bilgi keşfi sürecindeki

önemli bir aşamadır. Veri madenciliği aşamasında modelin kurulması sırasında çıkabilecek sorunlar, yeniden bu aşamaya dönülmesine neden olabilir.⁶⁶

Bu süreçteki ilk adım verinin toplanmasıdır. İlk olarak elde var olanlar gözden geçirilir ve toplanmasına gereksinim duyulan veri belirlenir. Gereksinim duyulacak verilerden bazıları daha önce hiç toplanmamış olacağından, veri toplama aşaması yeni verilerin elde edilmesini de içerebilir.⁶⁷ Eğer bu veriler farklı kaynaklardan toplanacak ise bu verilerin bütünleştirilmesi ve veriler arası tutarsızlıkların bulunup, düzeltilmesi yine bu aşamada yapılır. Kullanılacak veri belirlenirken her ne kadar uygulanacak veri madenciliği metodu bilinmese de, veri madenciliği hedefleri ve kurumun amaçları göz önünde bulundurulmalıdır.⁶⁸

Verinin toplanmasıyla başlayan bu süreçte, veri hakkında daha fazla bilgi sahibi olmak için yapılan işlemler, veri kalitesiyle ilgili sorunların belirlenmesi ve veri hakkındaki ilk izlenimi edinmeyle, verinin kavranması sağlanır.⁶⁹ Veri madenciliği analizi düşük kalitedeki veri seti üzerinde yapılırsa, yüksek kalitede bilgiye ulaşmak başarısız olur. Burada amaç, verinin kalitesini iyileştirmeye çalışmaktır.⁷⁰

Bu aşama analistin toplam zaman ve enerjisinin %80'ini harcamasına neden olabilir.⁷¹

⁶⁶ Akpınar, **a.g.e.**, s.7.

⁶⁷ Introduction to Data Mining, Two Crows Corporation, USA, 1999, (Çevrimiçi) <http://www.twocrows.com/intro-dm.pdf>, 16.12.2013.

⁶⁸ William A. Giovinazzo, **Internet Enabled Business Intelligence**, U.S.A., Prentice Hall PTR, 2002, s. 331.

⁶⁹ Irma Becerra-Fernandez, Stelios H. Zanakis, Steven Walczak, "Knowledge Discovery Techniques for Predicting Country Investment Risk", **Computers and Industrial Engineering**, No:43, No:4, 2002, s.790.

⁷⁰ Freitas, **a.g.e.**, s.65.

⁷¹ Selwyn Piramuthu, "Evaluating Feature Selection Methods for Learning in Data Mining Applications", **European Journal of Operational Research**, C.CLVI, No:2, Article In Press, 2004, s.483.

Veri hazırlama aşamasının önemi, hazırlanan veri seti üzerinde çeşitli veri madenciliği modelleri denendiğinde yaklaşık veya benzer olumlu sonuçlar alındığı görülerek kanıtlanmıştır.⁷²

Veri Madenciliği'nin başarılı sonuç vermesinde kaliteli verinin ne kadar önemli olduğu düşünülürse, bu aşamanın da önemi ortaya çıkacaktır. Veri ön işleme olarak da adlandırılan veri hazırlama aşaması, veri kaynağıyla ilgili işlemleri içerir ve veri temizleme, veri birleştirme, veri dönüştürme, veri indirgeme basamaklarından oluşur.⁷³

1.2.1.1. Veri Temizleme

Veri temizleme, veri hazırlamanın en önemli basamağını oluşturur. Veri madenciliği sistemlerinin kullandığı ham veri büyük miktarda olup veritabanlarında tutulur. Bu noktada veritabanlarının barındırdığı bazı sorunlara dikkat çekmek gerekir. Başlıca sorunlar, dinamik veri yapısı, geçersiz veri alanları, kayıp, gürültülü, eksik, artık, boş veri değerleri olarak sıralanabilir.⁷⁴

Verinin nitelik değerlerindeki eksiklik ve hatalar “gürültü” olarak adlandırılır. Veri kümesi içinde yer alan ancak bir anlam içermeyen verilerdir. Örneğin, kullanıcının doğum tarihinin 1200 olması gibi. Ortalama değerlere göre çok düşük veya yüksek değerlere sahip aşırı uç veriler de bu sınıftandır. Gürültü, yanlış nitelik değerleri, veri girişi veya iletimi sorunları, teknolojik kısıtlılık ve tutarsızlıktan olabilir.⁷⁵

⁷² Sven F. Crone, Stefan Lessmann, Robert Stahlbock, “The Impact of Preprocessing on Data Mining: An Evaluation of Classifier Sensitivity in Direct Marketing”, **European Journal of Operational Research**, C.CLXXIII, No:3, 2006, s.781.

⁷³ Han, Kamber, **a.g.e.**, s.105.

⁷⁴ Flawley, Piatetsky-Shapiro, Matheus, **a.g.e.**, s.62-63.

⁷⁵ Han, Kamber, **a.g.e.**, s.108.

Farklı kaynaklardan toplanan veri türlerinde eksik ve kayıplar olabilir. Örneğin, bir veritabanından alınan verilerde kişilerin medeni hali belliyken diğer veritabanında bu kayıtlar eksik veya hiç girilmemiş olabilir.⁷⁶

Eksik veya kayıp veriler, zamanla veritabanına girilebilir. Kullanılan veritabanı boyutu küçükse ve zaman sorunu yoksa kayıp veriler elle girilebilir. Başka bir yol ise bu verileri öngörü yöntemiyle doldurmaktır veya kayıp verilere bütün verilerin ortalama değeri verilebilir. Kayıp verilerin bulunduğu kayıt, veri kümesinden tamamen çıkartılabilir veya bu kayıt iptal edilebilir.⁷⁷

Yine bazı kayıtlar fazladan girilmiş olabilir. Eğer bir veritabanındaki kişilerin hem yaşları hem de doğum tarihleri girilmişse bu iki kayıttan biri fazladır. Bu iki kaydın birleştirilmesi, tek bir değişkenmiş gibi işleme sokulması gerekir.⁷⁸

Uzmanlar tarafından kaydedilmeyen gerçek zamanlı veriler genelde dinamik, eksik ve gürültülü olduğundan veri kalitesini sağlamak yoğun dikkat ve çaba gerektiren bir durumdur.⁷⁹

1.2.1.2. Veri Birleştirme

Veri madenciliğinde kullanılacak olan veriler sadece bir kaynaktan alınmış olmayabilir. Hatta daha iyi sonuç alınması için birçok kaynaktan toplanması gerekir. Her kaynaktaki veriler kendi niteliklerine göre depolanmışlardır. Bunların bir araya getirilmesi doğal olarak veri uyumsuzluğuna neden olacaktır. Bu uyumsuzluklar genelde verilerin farklı kodlama biçiminde olmasından kaynaklanır. Veri birleştirme

⁷⁶ A.e., s.106.

⁷⁷ Şule Özmen, “Veri Madenciliği Süreci”, Veri Madenciliği ve Uygulama Alanları Konferansı, İstanbul Ticaret Üniversitesi, İstanbul, 2003.

⁷⁸ Han, Kamber, a.g.e., s.108.

⁷⁹ Fayyad, Piatetsky-Shapiro, Smyth, a.g.e., s.27.

aşaması, toplanan bütün verilerin aynı biçimde tek bir veritabanında veya veri ambarında birleştirilmesini kapsamaktadır.⁸⁰

Örneğin bir ağırlık sisteminde bazı değerler kilogram ile girilmişken diğerinde paund ile girilmiş olabilir. Aynı birimi tanımlayan verilerin bu şekilde heterojen bir yapı oluşturması verinin bütünlüğü açısından doğru değildir.⁸¹

Bir kütüphane kullanıcı veritabanında girişler “kullanıcı-ID” şeklinde yapılmışken bir diğerinde “kullanıcı numarası” şeklinde yapılmış olabilir. Bu iki tabloyu aynı biçimde birleştirmek gerekir.

1.2.1.3. Veri Dönüştürme

Veri dönüştürme aşaması, verilerin anlam ve içeriklerinin değiştirilmeden veri madenciliğinde kullanılacak modele uygun olarak düzenlenmesidir. Bazı veri madenciliği modelleri sadece kategorik değerlerle çalışırken, bazıları sadece sayısal değerlerle çalışabilmektedir. Örneğin bir veritabanında “evet=1” , “hayır=0” biçiminde rakamsal değerler kullanılmış, farklı bir veritabanında ise “evet/hayır” biçiminde değerler girilmiş olabilir. Veritabanı üzerine uygulanacak metod hangisini destekliyorsa bütün girdilerin o formata dönüştürülmesi daha iyi sonuç için yararlı olacaktır.⁸²

Veriler özel bir şekilde düzenlenebilir veya varolan özelliklerinden yeni değerlerin türetilmesiyle zenginleştirilebilir.⁸³

Veri dönüştürme işlemlerinden bazıları şunlardır:⁸⁴

⁸⁰ Larissa T. Moss, **Business Intelligence Roadmap: The Complete Project Lifecycle for Decision- Support Applications**, Almanya, Addison Wesley, 2003, s.314.

⁸¹ Han, Kamber, **a.g.e.**, s.111.

⁸² Akpınar, **a.g.e.**, s.8.

⁸³ Rachid Anane, Data Mining and Serial Documents, **Computers and the Humanities**, C.XXXV, No:3, 2001, s.301.

- Düzeltme: Gürültülü verileri temizleme.
- Birleştirme: Verileri bazı sonuçlara ulaşmak için birleştirme işlemidir. Yıllık satış değerlerine ulaşmak için, günlük satış değerlerini birleştirmek gibi.
- Genelleştirme: Düşük düzeydeki ham veriyi daha üst düzeydeki değere dönüştürme işlemidir. Cadde ve sokak isimlerini şehir ve ülke isimleri altında toplamak gibi.
- Normalleştirme: En sık kullanılan veri dönüştürme işlemidir. Min-Max, Z skor ve Ondalık Ölçekleme gibi dönüşümlerdir. Örneğin, 900 maksimum değer ise, $n=3$ olarak alınırsa 900 sayısı 0,9 olarak normalleştirilir.⁸⁵

1.2.1.4. Veri İndirgeme

Üzerinde çalışılacak veri setleri çok büyük olduğunda veri madenciliği uygulamalarından verimli sonuç almak zorlaşmakta ve işlemler daha çok zaman almaktadır. Verilerin öznelikleri korunarak hacimce daha küçük biçime getirmeye veri indirgeme denir. Büyük hacimli bir veri setiyle verinin özellikleri korunarak hacimce daha küçük boyuta getirilen veri seti, aynı (veya hemen hemen aynı) sonucu üretir . Veri indirgeme için bazı teknikler kullanılır. Bunlar:⁸⁶

- Veri Küpü Birleştirme: Farklı tablolardaki aynı cins verilerin bir araya getirilmesidir. Örneğin, bir kütüphanedeki kullanıcı sayısı, her yıl iki

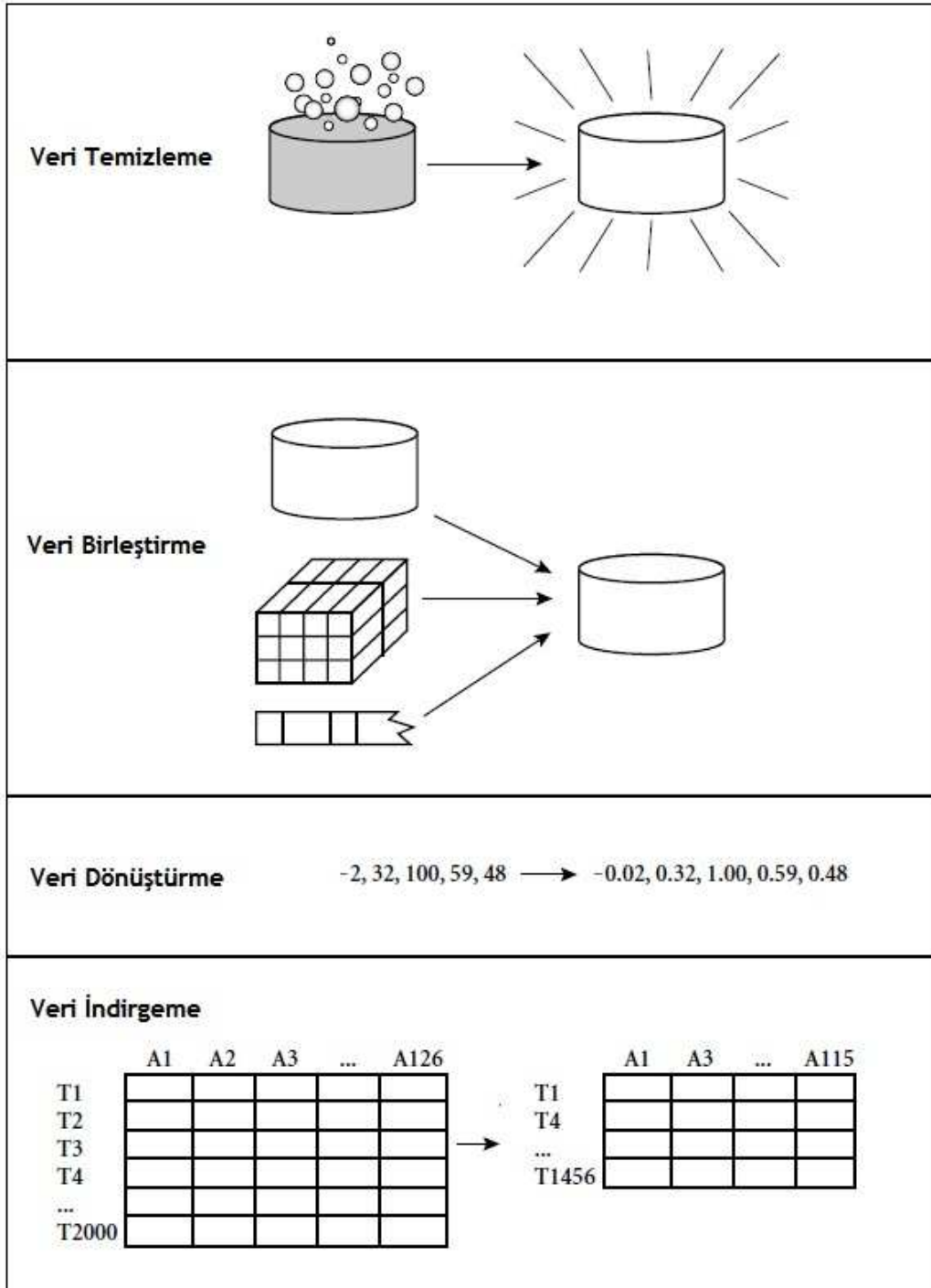
⁸⁴ Richard J. Roiger ve Michael W. Geatz, **Data Mining A Tutorial-Based Primer**, USA: Addison Wesley, 2003, s.156.

⁸⁵ Ayşe Oğuzlar, “Veri Ön İşleme”, **Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, No:21, 2003, s.73.

⁸⁶ Han, Kamber, **a.g.e.**, s.115.

dönem şeklinde bir tabloda kayıtlıysa, son üç yıla ait ikinci dönem kullanıcı sayısının bir tabloda birleştirilmesi gibi.

- Boyut İndirgeme: İlgisiz veya gereksiz değişkenlerin analizin uygulanacağı veri setinden kaldırılmasıdır. Örneğin, kütüphane kullanıcılarının cinsiyetini içeren kayıtların çıkarılacak sonuca etkisi olmayacağı bilindiğinde, bu kayıtların çıkarılması gibi.
- Veri Sıkıştırma: Analiz edilecek veri miktarını azaltmak için aynı cinsteki veri sıkıştırılır. Bu teknikte verinin yapısını bozmamak önemlidir. Bazen veri kayıplarına neden olabilir.
- Kesikleştirme: Sürekli verilerin kesikli değerlere dönüştürülmesi işlemidir. Bazı veri madenciliği yöntemleri sadece kategoriler üzerine uygulandığından verileri bu biçime sokmak gerekebilir. Örneğin, bir veritabanında kullanıcı yaş değerleri ayrıntılı biçimde yer alırken, bunları 15-30, 30-45 şeklinde kategorize etmek gibi.



Şekil 3. Veri Hazırlama (Önişleme) Biçimleri (Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques)

1.2.2. Veri Madenciliği

Veri madenciliği, veritabanlarından bilgi keşfi sürecinin en önemli aşamasıdır. Bazı kaynaklarda aynı süreci kapsadığı ileri sürülerek veri madenciliği ve veritabanlarında bilgi keşfi terimlerinin birbirlerinin yerine kullanıldığı görülmektedir.⁸⁷ Ancak VTBK süreci veriden yararlı bilgiyi keşfetmedeki bütün işlemleri kapsarken, veri madenciliği bu sürecin önemli ama sadece bir adımıdır.⁸⁸

VTBK, veri madenciliğini de içeren aşamalardan oluşan bir süreçtir. Bu süreçte sürekli tekrarlar, aşamalar arası ileri geri hareketler, atlamalar olabilir. Günümüzde genellikle sürecin veri madenciliği aşamasına eğilim gösterilmektedir, ancak diğer aşamalar sürecin en az VM kadar önemli bir parçasıdır.⁸⁹

Veri madenciliği konusu ikinci bölümde ayrıntılı olarak ele alınacaktır.

1.2.3. Yorumlama ve Değerlendirme

Veri tabanlarında bilgi keşfi sürecinin son aşaması olan yorumlama ve değerlendirme, veri üzerine uygulanan veri madenciliği modellerinin izlenmesi ve değerlendirilmesidir. Bilgi keşfi sürecinin ne kadar başarılı olduğunu, işletmenin amaçlarını karşılayıp karşılamadığını, karar verme sürecine katkı verip vermeyeceğini değerlendirme adımıdır. Ayrıca sürecin tekrar edilip edilmeyeceğine de karar verilir.

Bu aşamada çeşitli görselleştirme ve raporlaştırma araçları kullanılarak keşfedilmiş bilgiler ilgili kullanıcılara sunulur. Bu aşama ikinci bölümde veri madenciliği süreci içerisinde yer alacaktır.

⁸⁷ Karin Becker, Cinara Ghedini, "A Documentation Infrastructure for the Management of Data Mining Projects", **Information and Software Technology**, C.XLVII, No:2, 2005, s.95.

⁸⁸ Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, "Knowledge Discovery and Data Mining: Towards a Unifying Framework", **Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)**, C.XXXII, AAAI Press, 1996, s.83.

⁸⁹ Han, Kamber, **a.g.e.**, s.6.

2. VERİ MADENCİLİĞİ TANIMI, SÜRECİ VE UYGULAMA ALANLARI

2.1. Veri Madenciliği Tarihçesi

Bilginin öneminin artmasıyla, bilgiye ulaşımın kolaylaştırılması ve elde edilen bilgilerin kullanılabilir hale getirilerek analiz edilmesi konusunda gelişmeye ihtiyaç duyulmasına sebep olmuştur. Geçmişte bilgiye ulaşma ve zor ulaşılmış bu bilgiyi kayıt altında tutma işlemi vakit alırken, bu bilgiler ışığında kullanılan yetersiz karar mekanizmaları çabuk işliyor, bu da bilgiden maksimum derecede verim alınamamasına ve sonuçların etkin olamamasına sebep oluyordu. Günümüze yaklaştıkça ise bilgiye ulaşım hızlandığı halde elde edilmiş bilgilerin analizlerinin yapılması konusunda sayısız yöntem ve karar mekanizması geliştirildiğinden, çıktı alabilme safhası daha karmaşık hale getirilerek vaktin büyük bölümünde bilgiden verim alma üzerine yoğunlaşmaktadır. Veriden en yüksek faydayı sağlama amacıyla uygulanan yöntemlerin karmaşıklığı ve zorluğu bu işlemleri bilgisayarlarla yapma fikrini doğurmuş, böylelikle, çeşitli matematiksel ve istatistiksel hesaplamalara dayanan algoritmalar geliştirilmiş ve “Veri Madenciliği” kavramı doğmuştur.

Veri madenciliğinin tarihsel gelişimi:

1950’ler: Geliştirilen teknikler sayesinde sürekli gelişen veri madenciliği, günümüzde daha fazla bilgiye çok daha kısa sürede ulaşabilmeyi sağlayarak hayatımızı kolaylaştırmış, bir çok meslek grubunun yükünü de hafifletmiştir. Bu teknikler üzerinde çalışmalara ilk olarak 1950’li yıllarda başlanmış, mantık ve bilgisayar bilimleri konularında çalışmalar yaparak yapay zeka ve makine öğrenme konularında önemli gelişmeler yakalanmıştır. Bu yıllarda bilgisayarların sayım için kullanılmaya başlanması veri madenciliğinin ortaya çıkışına zemin hazırlayan ilk olaydır.

1960'lar: 1960'lı yıllarda istatistikçiler, regresyon analizi, en büyük olasılık kestirim, sinir ağıları vb. gibi yeni algoritmalar ve etkili yöntemler üzerinde çalışmışlardır. Yöntemlerdeki bu önemli gelişmelere paralel olarak veritabanı sistemleri giderek gelişmiş ve büyük sayıda metin dokümanlarının saklanması sağlanmıştır. Böylelikle veri tabanı ve verilerin depolanması kavramı teknoloji dünyasında yerini almıştır. 1960'lı yılların sonlarına yaklaştıkça bilim adamları basit öğrenmeli bilgisayarlar geliştirebilmişlerdir. Böylelikle verilerin toplanması ve veri tabanlarının yaratılması konularında büyük ilerlemeler yaşanmış, Veritabanı Yönetim Sistemi konusunda ilk adımlar atılmıştır.

1970'ler: İlişkisel Veri Tabanı Yönetim Sistemleri uygulamaları kullanılmaya başlandığı 1970 li yıllarda, bilim adamları basit kurallara dayanan uzman sistemler geliştirmişlerdir.

1980'ler: İlişkisel Veri Tabanı Yönetim Sistemleri kullanımının bir çok alanda yaygınlaşmaya başladığı 1980'li yıllarda, şirketler, müşterileri, rakipleri ve ürünleri ile ilgili verilerden oluşan veri tabanları oluşturmaya başlamışlardır. Bu veri tabanları çok büyük miktarda veriyi içinde tutmaktadır ve buradaki verilere veri tabanı sorgulama dilleri sayesinde ulaşılmaktadır. Bu yıllarda en yaygın kullanılan veri tabanı sorgulama dili SQL'dir.

1990'lar: Bu yıllarda artık katlanarak çoğalan veri miktarı, faydalı bilgiye nasıl ulaşılabilir sorusunun gündeme gelmesine sebep oldu ve 1990'ların başlarında bu konu hakkında çalışmalar başlandı. 1989, KDD (IJCAI)-89 Veri Tabanlarında Bilgi Keşfi Çalışma Grubu toplantısı ve 1991, KDD (IJCAI)-89'un sonuç bildirgesi sayılabilecek "Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop" makalesinin KDD (Knowledge Discovery and Data Mining) ile ilgili temel tanım ve kavramları ortaya koyması ile süreç daha da hızlanmış ve nihayet 1992 yılında veri madenciliği için ilk yazılım gerçekleştirilmiştir.¹

¹ Serkan Savaş, Nurettin Topalaoglu, Mithat Yılmaz, "Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri", **İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi**, No:21, 2012, s.5.

2000'ler: Veri madenciliği sürekli gelişmiş ve hemen hemen tüm alanlara uygulanmaya başlanmıştır. Alınan sonuçların faydaları görüldükçe, bu alana ilgi artmıştır. Bu yıllarda internet kullanımının artmasıyla internet kullanıcıları arasında bilgi paylaşımı çoğalmış ve internet ortamında biriken bu bilgiler devasa boyutlara ulaşmıştır. Bu da veri madencilerini çok büyük miktarda verinin yönetimi için yeni yöntemler aramaya yöneltmiştir.

2.2. Veri Madenciliği Tanımı

Veritabanlarında bilgi keşfi sürecinin en önemli aşaması olan veri madenciliği üzerine yapılmış birçok tanım vardır. Veri madenciliği yeni ortaya çıkan ve gelişmekte olan bir alan olduğu için her gün farklı kaynaklarda farklı tanımlar ortaya çıkmaktadır. Tanımların çoğu tanımlayanın arka planına, ilgi alanına ve bakış açısına dayanmaktadır. Yapılan tanımların hepsi doğru olmakla birlikte, konuyu kısıtlandırmak doğru olmayacağı için geneli kapsayan kesinleşmiş tek bir tanım yapılamamaktadır.

Öncelikle kavrama mecazi olarak baktığımızda veri madenciliği, veri dağlarının içinde kalmış, altın değerindeki, gizli bilgi hazinelerini keşfedilmek için yapılan çalışmalardır. Han, veri madenciliğini, büyük veri yığınlarından bilgi çıkarımı olarak tanımlarken kavramın yanlış kullanıldığını öne sürer. Örneğin, kumdan veya kayadan altın çıkarmaya “altın madenciliği” denir, “kaya madenciliği” veya “kum madenciliği” denmez. Bu yüzden terimin doğrusu “bilgi madenciliği” olmalıydı.²

Veri madenciliği, büyük verileri analiz ederek, verileri anlamlı hale getiren bilgisayar destekli işlemlerdir. Veri madenciliği terimi büyük veritabanları içerisindeki değerli bilgiyi arama ve değerli bir maden cevherine sahip dağdaki

² Jiawei Han, Micheline Kamber, **Data Mining Concepts and Techniques**, 2.bs., USA, The Morgan Kaufman Publishers, 2006, s.5.

madeni kazıp çıkartma arasındaki benzerlikten türemiştir. Her ikisi de ya uçsuz bucaksız miktardaki materyali eleme veya değerli olanı bulmak için derinlemesine araştırma gereksinimi duyan işlemlerdir.³

Veri madenciliği, veritabanlarında veya veri ambarlarında çeşitli biçimlerde depolanan büyük ölçekli verilerin içindeki gizli kalmış, önceden bilinmeyen ama potansiyel olarak kullanışlı olduğu düşünülen anlamlı bilginin keşfedilmesi sürecidir.⁴

Alpaydın'a göre veri madenciliği;⁵ büyük miktarda veri içinden, gelecekle ilgili öngörüde bulunmamızı sağlayacak bağıntı ve kuralların bilgisayar programları aracılığıyla aranması ve analizidir. Ayrıca, çok büyük miktardaki verilerin içindeki ilişkileri inceleyerek aralarındaki bağlantıyı bulmaya yardımcı olan veri analizi tekniğidir.

Veri madenciliği, verilerden örüntüleri keşfetmektir. Burada örüntü, veriler arasındaki potansiyel olarak yararlı ve ilginç ilişkiler ve eğilimler anlamına gelmektedir.⁶ Veri madenciliği ham verilerdeki örüntüleri ve ilişkileri bulma sürecini otomatikleştirir ve karar destek sistemine yararlı olabilecek veya karar vericiler tarafından değerlendirilebilecek sonuçlar ortaya çıkarır.⁷

³ S. Sumathi, S. N. Sivanandam, **Introduction to Data Mining and its Applications**, Springer, New York, 2006, s.9.

⁴ W.J. Flawley, G. Piatetsky-Shapiro, C.J. Matheus, "Knowledge Discovery in Databases : AnOverview", **AI Magazine**, C.XIII, No:3, 1992, s. 57.

⁵ Ethem Alpaydın, "Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri", Bilişim 2000 Eğitim Semineri, (Çevrimiçi) <http://www.cmpe.boun.edu.tr/~ethem/> , 16.12.2013.

⁶ Hüseyin Gürüler, Ayhan İstanbullu, Mehmet Karahasan, "A new student performance analysing system using knowledge discovery in higher educational databases", **Journal Computers and Education**, C.LV, No:1, 2010, s.247.

⁷ George Fernandez, **Data Mining Using SAS Applications**, Chapman and Hall/CRC, 2002, s.1.

Veri madenciliği, tek başına bir çözüm değil, çözüm için verilecek karar sürecini destekleyen, sorunu çözmek için gerekli olan verileri sağlamaya yarayan bir araçtır.⁸

Kurumsal bakış açısıyla baktığımızda veri madenciliği, kurumun sahip olduğu büyük miktardaki verilerde, karar verici yönetici veya analistin sormayı düşünemediği sorularla, kurumla ilgili cevapların aranmasıdır.⁹

Bu tanımlar ışığında genel olarak veri madenciliği; eyleme geçirilecek kararlar verebilmek için veritabanı veya veri ambarında saklanan büyük miktardaki verilerin içindeki gizli kuralları ve ilişkileri keşfederek, bu verilerden potansiyel olarak kullanışlı, yararlı olacağı düşünülen ve daha önce keşfedilmemiş bilgileri ortaya çıkarma sürecidir.

Veri madenciliği tanımlardan anlaşılacağı üzere farklı kişilere, farklı kuruluşlara, farklı kullanıcılara göre değişik anlamlar içerebilmektedir. Ancak bütün bu tanımlama çabaları incelendiğinde ortak birkaç nokta ön plana çıkmaktadır. Bunlar;

- Büyük ölçekteki veriler,
- Veriler içerisindeki ilişkiler,
- Daha önce keşfedilmemiş anlamlı ve yararlı bilgi çıkarımı,
- Bilgisayar ve bilgi teknolojilerinin kullanımınıdır.

Büyük ölçekli veriler, günümüzde veritabanlarında depolanmaktadır. Veritabanlarının gelişmesinde bilgisayar ve bilgi teknolojilerinin gelişmesi rol

⁸ Sanjay Kumar Madria v.d., "Research Issues in Web Data Mining", **DaWaK '99 Proceedings of the First International Conference on Data Warehousing and Knowledge Discovery**, Berlin, Springer-Verlag Berlin Heidelberg, 1999, s.303.

⁹ Paul Gray, Hugh J. Watson, **Decision Support in The Data Warehouse**, U.S.A., Prentice Hall, 1998, s.144.

oynamıştır. Veri madenciliği kavramının ortaya çıkmasında da veritabanı teknolojisindeki gelişmelerin önemi oldukça büyüktür. Veritabanı bir veri yığındır. Bu veri yığınının analiz edilmesi, sınıflandırılması, raporlanması gerekmektedir. Bu aşamada veri madenciliği gibi çeşitli teknikler devreye girmektedir. Var olan verilerden çıkarılan veya hesaplanan sonuçlar üreten bir sorgulamanın sonucu olan geleneksel veritabanlarının aksine, veri madenciliği, veriler içerisindeki geçerli, orijinal, işe yarayabilecek ve anlaşılabilir desenleri (veriyi tanımlayan değerler bütünü) tanıma işlemidir.¹⁰

Veri analizinde istatistik bilimi de önemli rol oynamaktadır. İstatistiğin amacı analitik yöntemlerle veriler hakkında anlamlı bilgiler üretmek ve yorum yapılmasına olanak sağlamaktır. Bu durumda veri madenciliğinin istatistikten farkının ne olduğu sorusu akla gelmektedir. İstatistiğin doğuşu, bilgisayarın icadından önceye dayanmaktadır. İstatistiksel yöntemler elle de uygulanabilir. Bilgisayar teknolojisinin doğuşu ve gelişimi özellikle büyük verilerin istatistiksel analizinde kolaylık sağlamış olsa da verinin içindeki gizli örüntülerin bulunması, çözümlenmesi ve yorumlanmasında yeterli olmamıştır. Bu noktada bazı modellere, algoritmalara gereksinim duyulmuştur. Bu da veri madenciliği kavramının ortaya atılmasına neden olmuştur.¹¹

Hand, veri madenciliğini, istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenimi ve diğer alanların ara yüzünde bulunan yeni bir disiplin olarak tanımlamıştır. VM, veritabanlarının ikincil analizidir.¹² Özmen'e göre veri

¹⁰ Rachid Anane, "Data Mining and Serial Documents", **Computers and the Humanities**, C.XXXV, No:3, 2001, s.300.

¹¹ Pieter Adriaans, Dolf Zantinge, **Data Mining**, 2. bs., England, Addison-Wesley, 1997, s. 25.

¹² David J. Hand, 'Data Mining: Statistics and More?', **The American Statistician**, C.LII, No:2, 1998, s. 112.

madenciliği, istatistik biliminin teknolojiyle bütünleşmesi sonucu oluşturulan bir araçtır.¹³

Veri madenciliği yeni bir istatistik yöntemi değildir. Bu terim, standart istatistiksel veri işlemenin genişletilmeye, değiştirilmeye ve eksikliklerinin giderilmesine gereksinim duyduğundan ortaya çıkmıştır. İstatistiksel verilerin birleştirilmesi, keşifsel veri analizi, örüntü tanıma ve veritabanı teknolojileri konularını kapsayan disiplinlerarası bir konudur. İstatistikçiler bir veya birkaç tür sonuç çıkarmayla ilgilenirler. Amaçları birbiriyle ilişkili olasılıklardan bir örnek sunma veya özet çıkarmaktır.¹⁴ Klasik istatistiksel uygulamalar ve veri madenciliği arasındaki başka önemli bir fark veri kümesinin büyüklüğüdür. İstatistikçi için veri kümesi yüz veya bin veri içerirken, veri madencisi için milyon veya milyar veri beklenmedik bir durum değildir.¹⁵ İstatistiksel veri analizinde, verilerin nerede ve nasıl analiz edileceği konusunda karar vericinin güçlü rehberliği gerekir. Ayrıca, istatistiksel analiz sonuçlarının yorumlanması oldukça güç ve göz korkutucu olabilmektedir.¹⁶

Veri madenciliğinin her ne kadar yeni bir analiz yöntemi, çalışma alanı olduğu doğrusa da yeni bir buluş olduğunu söylemek yanlıştır. Veri madenciliği birkaç disiplini bir araya getiren yeni bir açılamdır.¹⁷

Veri madenciliği yöntemleri, diğer veri analiz yöntemlerinden farklı olarak;¹⁸

¹³ Şule Özmen, “İş Hayatı Veri Madenciliği ile İstatistik Uygulamalarını Yeniden Keşfediyor”, **V. Ulusal Ekonometri ve İstatistik Sempozyumu**, Çukurova Üniversitesi, Adana, 19–22 Eylül 2001

¹⁴ David J. Hand v.d., “Data Mining for Fun and Profit”, **Statistical Science**, C.XV, No:2, 2000, s.111.

¹⁵ Selim Tüzüntürk, “Veri Madenciliği ve İstatistik”, **Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, C.XXIX, No:1, 2010, s. 73.

¹⁶ William J. Flawley, Gregory Piatetsky-Shapiro, Christopher J. Matheus, “Knowledge Discovery in Databases : An Overview”, **AI Magazine**, C.XIII, No:3, 1992, s.61.

¹⁷ Michael C. Lovell, “Data Mining”, **The Review of Economics and Statistics**, C.LXV, No:1, 1983, s.1.

- Büyük miktarda (milyarlarca kayıt) veri biriktirilip, analiz edilebilmekte,
- Bilgisayar teknolojileri kaynaklı karar ağaçları, yapay sinir ağları gibi çeşitli yöntemler kullanılabilen,
- Hedef kullanıcı kitlesi seçilip, sadece onlar için özel çözümler öngörmekte,
- Karar vericilere yönelik profesyonel analiz yapan, kullanıcı dostu, basit ara yüzü ama çok pahalı olmayan yeni yazılımlar kullanılmaktadır.

Veri madenciliğinin iki görevi vardır:¹⁹

- Doğru kararlar verebilmek için veriyi enformasyona veya bilgiye dönüştürmek,
- Doğru, uygun eylemlerin gerçekleşebilmesi için bilgiyi dağıtma işleyişini sağlamak.

Veri madenciliği yapabilmek için 3 önemli unsur ise şunlardır:

1. İnsan: Veri analisti, model tasarımcısı
2. Teknolojik altyapı (veritabanları, veri ambarı, vs.),
3. Veri madenciliği yazılımları.

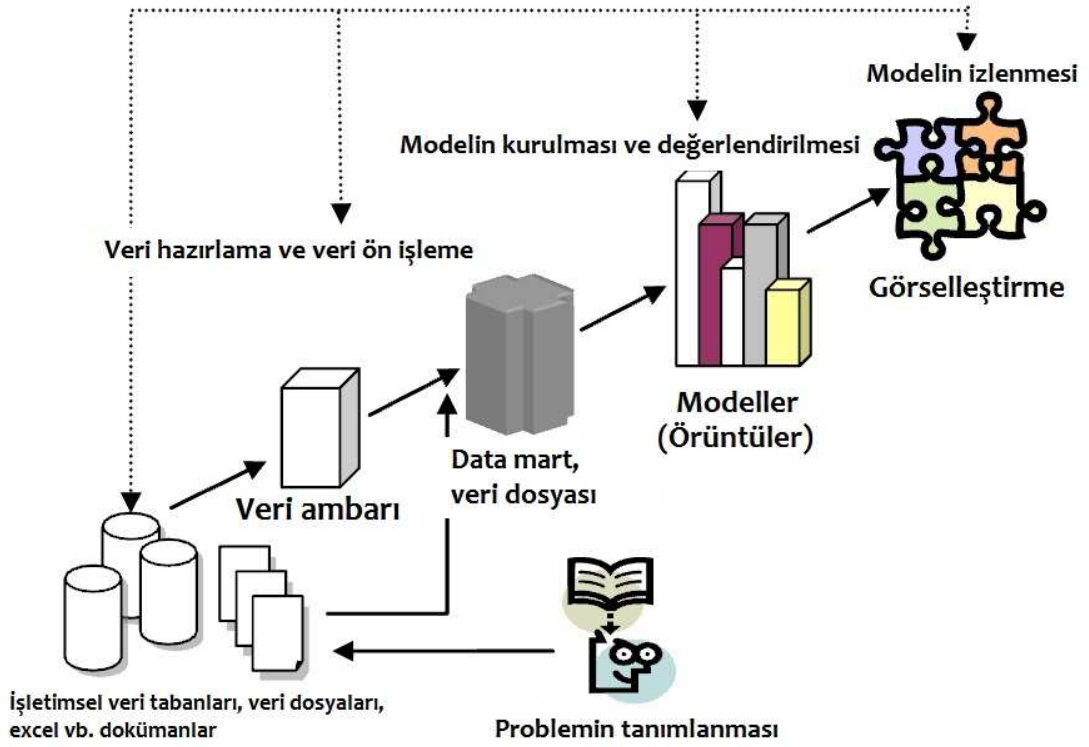
Veri madenciliği için insan ne kadar önemliyse de uygulamanın kendisi için en önemli koşul veri madenciliği araçlarıdır. Veri madenciliği projeleri otomatik araçlar yoluyla yürütüldüğünden, kullanılacak yazılımın seçilmesi de önem

¹⁸ Gilbert Saporta, "Data Mining and Official Statistics", **Quinta Conferenza Nazionale di Statistica**, ISTAT, Roma, 2000, s.15-17.

¹⁹ Ronald S. Swift, **Accelerating Customer Relationship: using CRM and Relationship Technologies**, Prentice Hall PTR, 2001, s.98.

taşımaktadır.²⁰ En yaygın kullanımlı veri madenciliği yazılımları arasında RapidMiner, WEKA, SPSS Clementine, SAS Enterprise Miner, IBM Intelligent Miner for Data, DBMiner, Statistica Data Miner ve XLMiner gösterilebilir.

2.3. VERİ MADENCİLİĞİ UYGULAMA SÜRECİ



Şekil 4. Veri Madenciliği Uygulama Süreci (Jiann-Cherng Shieh, “The Integration System For Librarians’ Bibliomining”, Electronic Library)

Veri madenciliğini bir süreç olarak değerlendirmek gerekir. Bu süreç birbirini izleyen problemin tanımlanması, uygun modelin kurulması, değerlendirilmesi, kullanılması ve izlenilmesi aşamalarından oluşur.

²⁰ Introduction to Data Mining, 3. bs., Two Crows Corporation, USA, 1999, (Çevrimiçi) <http://www.twocrows.com/intro-dm.pdf> , 16.12.2013.

2.3.1. Problemin Tanımlanması

Problemin tanımlanması aşaması, temel olarak kurumun hedeflerinin ve amaçlarının belirlenmesini içerir. Veri madenciliği uygulamasının sonucunu etkileyebilecek etkenlerin bulunması, tartışılması ve değerlendirilmesi aşamasıdır. Sürecin en önemli aşaması olarak görülebilir. Veri madenciliği analizinin başarılı sonuç vermesi için, kurumun problemini çözmeye yönelik, kesin, açık, gerçekleştirilebilir, sonuçları ölçülebilir bir hedefe sahip olma ve uygulama sürecinin dikkatlice planlanması gerekir.²¹

Veri madenciliği çalışmalarında başarılı olabilmek için öncelikle uygulamanın ne amaçla kullanılacağına açık ve net biçimde belirlenmesi gerekir. Amaç, problem üzerine odaklanmış, açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeyinin nasıl ölçüleceği tanımlanmalıdır. Bununla birlikte karşılaşılabilecek yanlış tahminlerde harcanacak maliyete ve doğru tahminlerde kazanılacak faydaya ilişkin öngörülere de bu aşamada yer verilmelidir.²²

Veri madenciliği yöntemini kullanacak bir perakende firması, yapacağı pazarlama kampanyasına cevap verme olasılığı yüksek müşterileri belirlemek, üreteceği yeni ürün için hedef kitle belirlemek, ürün için yeni stratejiler geliştirerek markanın pazar payını artırmak gibi hedefler belirleyebilir.²³ Belirlenecek hedeflerin, firmanın daha çok ürün satması, daha çok kâr etmek, markayı tanıtmak gibi genel geçer hedefler değil, daha belirgin, özgül bir problem üzerine yoğunlaşmış olması gerekir.

Bir kütüphane, uluslararası veritabanlarından birini satın almak istediğinde, hangisini alması gerektiğini belirlemek için kullanıcılarının yapmış olduğu yayın

²¹ William A. Giovinazzo, **Internet Enabled Business Intelligence**, U.S.A., Prentice Hall PTR, 2002, s. 331.

²² Haldun Akpınar, “**Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği**”, İ.Ü. İşletme Fakültesi Dergisi, 2000, C.XXIX, No:1, s.8.

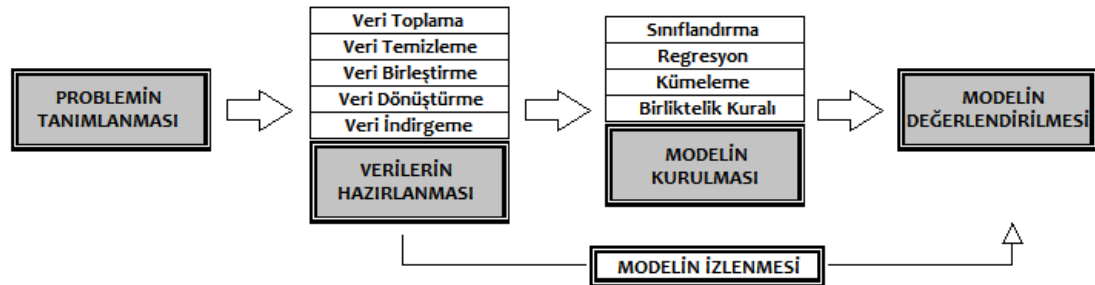
²³ Larissa T. Moss, **Business Intelligence Roadmap: The Complete Project Lifecycle for Decision- Support Applications**, Almanyada, Addison Wesley, 2003, s. 307-310.

taramalarını analiz etmeye karar verdiğinde, bu problemi açıkça tanımlanması gerekir. Yani problem, kullanıcıların en çok arama yaptıkları konuyu bulmak ve bu konuyla ilgili veritabanı satın almaktır, şeklinde kesin ve açıkça belirlenmelidir..

2.3.2. Modelin Kurulması ve Değerlendirilmesi

Tanımlanan problem için uygun model bulunması ve bu modelleme tekniğinin kurulması aşamasıdır. Veri madenciliğinde veri kaynaklarından en fazla verimin alınabilmesi için bu aşama önemlidir. İyi kurulan bir model analiz sonucunda elde edilecek sonuçların kalitesini belirleyecektir. Eğer uygun model kurulmazsa veriler arasında bulunabilecek ilişkiler doğru çözümlenemez ve önemli örüntüler saptanamaz. Böylece veri madenciliğinden başarılı sonuç elde edilemez.²⁴

Uygun model bulunabilmesi, çok sayıda tekniğin denenmesi ile olur. Bu yüzden bu aşama, en uygun olduğu düşünülen model bulununcaya kadar tekrarlanan bir süreçtir. Süreç tekrarlandıkça performans iyileşmekte, sonuçlar daha güvenilir olmaktadır.²⁵



Şekil 5. Veri Madenciliği Uygulama Süreci (H.Akpınar, "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", İ.Ü. İşletme Fakültesi Dergisi)

Bir modelin uygunluğunun test edilmesinde kullanılan en basit yöntem geçerlilik testi sürecidir. Bu süreçte öncelikle bütün verilerin %5 ile %33 arasındaki

²⁴ Michael J.A Berry, Gordon Linoff, **Data Mining Solutions**, U.S.A., Wiley Computer Publishing, 1998, s.25.

²⁵ Arlene Zaima, James Kashner, **Data Mining Primer for the Data Warehouse Professional**, (Çevrimiçi) <http://www.tdan.com/view-articles/5827>, 16.12.2013.

bir kısmı test verisi olarak ayrılır ve model bu veriler üzerinde denenir. Veriler üzerinde bazı varsayımlar yapılarak çıkan sonuçlar yorumlanır ve başarılı olan teknikler dikkate alınarak seçim yapılır.²⁶ Sadece modelin seçimine değil, modelle birlikte kullanılacak tekniklerden (karar ağaçları, yapay sinir ağları vb.) hangisinin uygun olduğuna da karar verilmeye çalışılır.²⁷

Modelin kurulması çalışmasına başlamadan önce hangi tekniğin en uygun olduğuna karar vermek güçtür. Bu nedenle farklı modeller kurularak, doğruluk derecelerine göre en uygun modeli bulmak ve sonrasında modelde en uygun tekniğin kullanılmasını belirlemek üzere sayısız deneme yapılmasında yarar vardır.

Kurumsal hedefe ve amaca yönelik problemlerin çözümünde en kullanışlı modelin bulunması için çeşitli alternatifler araştırılmalıdır.

Modelin uygunluk derecesinin değerlendirilmesinde önemli bir ölçüt modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar önemli olsa da, kararın niçin verildiğinin yorumlanması çok daha büyük önem taşır. Yorumlanamayacak kadar karmaşık modeller, karar için doğru verilere erişilebilse bile tercih edilmemelidir. Kurulan modelin belirlenmesinde diğer bir ölçüt, modelin uygulanmasında elde edilecek kazancın, bu modelin uygulanma maliyetine oranla yüksek olmasıdır.

Kurulan modelin uygunluk ve doğruluk derecesi ne kadar yüksek olursa olsun, gerçek dünyayı tam anlamıyla modellediğini garanti etmek olanaklı değildir.²⁸

²⁶ Haldun Akpınar, **a.g.e.**, s.11.

²⁷ Arlene Zaima, James Kashner, **Data Mining Primer for the Data Warehouse Professional**, (Çevrimiçi) <http://www.tdan.com/view-articles/5827>, 16.12.2013.

²⁸ Haldun Akpınar, **a.g.e.**, s.13.

2.3.3. Modelin Kullanılması ve İzlenmesi

Kurulan ve geçerliliği onaylanan modelin kullanılması aşaması, modelin zaman içerisinde izlenip ortaya çıkan değişikliklerin yeniden modele yansıtılması sürecidir.²⁹ Modelin izlenmesi ise veri madenciliği uygulama sürecinin en son aşamasıdır. Uygulanan modelin kurumun amaçlarına ve hedeflerine ne kadar ulaştığı, belirlenen probleme ne kadar yanıt verdiği ölçülür. Sürecin gözden geçirilmesiyle modelin doğru kurulup kurulmadığı, modelleme tekniklerinden ne kadar verim alındığı, gelecekte yapılacak analizlerde bunların kullanılabilirliği değerlendirilir. Değerlendirme sonuçlarına göre çalışmanın nasıl devam edeceğine, sürecin tekrar edilip edilmeyeceğine bu aşamada karar verilir.³⁰ Model sonuçlarının izlenmesinde, modelin kurulması öncesinde tahmin edilen ve sonrasında gözlenen değişkenler arasındaki farklılığı gösteren grafikler kullanılabilir.³¹

Modelin izlenmesi uygulama sürecinin son aşamasıdır ancak verilerin hazırlanmasından, modelin değerlendirilmesine kadarki sürecin tamamını içerebilir. İzleme aşamasında ilgili problemin çözümüne yönelik sonuç alınmadığı görüldüğünde ilk aşamaya dönülerek, bu aşamadaki iyileştirmeler sonucunda izleme devam edilebilir.

Veri madenciliği süreci başarılı olursa, keşfedilen yeni bilgiler kurumun karar verme sürecini iyileştirmede kullanılabilir.³²

²⁹ A.e.

³⁰ Pete Chapman, v.d., “**CRISP-DM 1.0 Step-by-Step Data Mining Guide**,” (Çevrimiçi) <http://www.crisp-dm.org/CRISPWP-0800.pdf>, 16.12.2013, s. 31.

³¹ Akpınar, a.g.e., s.14.

³² Alex A. Freitas, **Data Mining and Knowledge Discovery with Evolutionary Algorithms**, Almanya, Springer-Verlag, 2002, s.1.

2.4. VERİ MADENCİLİĞİ UYGULAMA ALANLARI

Veri madenciliği henüz yeni bir disiplin olmasına rağmen, birçok alanı ilgilendiren disiplinlerarası bir yaklaşımdır.³³

Teknolojinin gelişmesiyle kurumlarda veri artışına çözüm olarak getirilen veritabanı yönetim sistemleri, verinin toplanmasını ve saklanmasını kolaylaştırmış, veri işleme maliyetini azaltmıştır. Veri analizinde yeni geliştirilen yöntemlerin çoğalmasıyla birlikte veri madenciliği uygulamalarına olan ilgi de artmaktadır.³⁴ Bu nedenle her geçen gün uygulama alanları genişlemektedir.

Özellikle endüstriyel kurumlar, müşterilerinin davranışlarını izlemekte ve gelecek davranışları için öngöründe bulunup onların kuruma bağlılıklarını artırmaya çalışmaktadır. Bu bağlamda veri madenciliği öngörü amacıyla kullanılabilir olacak önemli bir yöntemdir.

Veri madenciliği herhangi bir disipline özgü geliştirilmemiştir. Verinin toplandığı, saklandığı ve analiz edildiği bütün alanlarda kendine yer bulmaktadır.³⁵

Veri madenciliği uygulamalarına; son yıllarda dünya piyasasında değişen ekonomik koşulların sonucu olarak rekabetin artmasıyla başta pazarlama olmak üzere, astronomi, biyoloji, finans, sigorta, tıp, güvenlik, milli güvenlik, spor, trafik, ulaşım, ulaştırma, lojistik, tedarik zinciri, meteoroloji ve daha birçok alanda rastlanmıştır.³⁶

³³ Chris Rygielski, Jyun-Chen Wang, David C. Yen, "Data Mining Techniques for Customer Relationship Management", **Technology in Society**, C.XXIV, No:4, 2002, s.488.

³⁴ Sang C. Park, Selwyn Piramuthu, Michael J. Shaw, "Dynamic Rule Refinement in Knowledgebased Data Mining Systems," **Decision Support Systems**, No:31, 2001, s. 205.

³⁵ Don Meyer, Casey Cannon, **Building a Better Data Warehouse**, USA, Prentice Hall, 1998, s. 186.

³⁶ H.Akpınar, "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", İ.Ü. İşletme Fakültesi Dergisi, 2000, C.XXIX, No:1, s.4.

Günümüzde yaygın olarak kullanıldığı alanlar şunlardır:³⁷

Pazarlama ile ilgili olarak;

- Müşterilerin satın alma örüntülerinin belirlenerek gruplanmasında,
- Müşterilerin demografik özellikleri arasındaki bağlantıların kurulmasında,
- Çeşitli pazarlama kampanyalarında,
- Mevcut müşterilerin elde tutulması için geliştirilecek pazarlama stratejilerinin oluşturulmasında,
- Pazar sepeti analizinde,
- Çapraz satış analizlerinde,
- Müşteri değerlemede,
- Müşteri ilişkileri yönetiminde,
- Çeşitli müşteri analizlerinde,
- Satış tahminlerinde kullanılabilir.

Bankacılık ile ilgili olarak;

- Farklı finansal göstergeler arasındaki gizli bağlaşımların bulunmasında,
- Kredi kartı dolandırıcılıklarının tespitinde,
- Müşteri gruplamasında,

³⁷ Abdullah BAYKAL, “Veri Madenciliği Uygulama Alanları”, **Dicle Üniv. Ziya Gökalp Eğitim Fakültesi Dergisi**, No:7, 2006, s. 97.

- Kredi taleplerinin deęerlendirilmesinde,
- Usulsüzlük tespitinde,
- Risk analizlerinde,
- Risk yönetiminde kullanılabilirlerdir.

Sigortacılık ile ilgili olarak;

- Yeni poliçe talep edecek müşterilerin tahmin edilmesinde,
- Sigorta dolandırıcılıklarının tespitinde,
- Riskli müşteri sınıfının belirlenmesinde kullanılabilirlerdir.

Perakendecilik ile ilgili olarak;

- Satış noktası veri analizlerinde,
- Alış-veriş sepeti analizlerinde,
- Tedarik ve mağaza yerleşimini uygun şekilde getirmede kullanılabilirlerdir.

Borsa ile ilgili olarak;

- Hisse senedi fiyat tahminlerinde,
- Genel piyasa analizlerinde,
- Alım-satım stratejilerinin uygulanmasında kullanılmaktadır.

Telekomünikasyon ile ilgili olarak;

- Kalite ve iyileştirme analizlerinde,

- Hisse tespitlerinde,
- Hatların yoğunluk tahminlerinde kullanılabilir.

Tıp ve Sağlık ile ilgili olarak;

- Test sonuçlarının tahmininde,
- Ürün geliřtirmede,
- Tıbbi teşhislerde,
- Tedavi sürecinin belirlenmesinde kullanılabilir.

Endüstri ile ilgili olarak;

- Kalite kontrol analizlerinde kullanılabilir.
- Lojistik ile ilgili olarak;
- Üretim süreçlerinin optimizasyonunda kullanılabilir.

Bilim ve Mühendislik ile ilgili olarak;

- Ampirik veriler üzerinde modeller
- Yeni virüs türlerinin keşfi ve sınıflandırılmasında,
- Gen haritasının analizi ve genetik hastalıkların tespitinde,
- Kanserli hücrelerin tespitinde,
- Gezegen yüzey şekillerinin, gezegen yerleşimlerinin ve yeni galaksilerin keşfinde kullanılabilir.

2.4.1. Pazarlama

Günümüzde endüstriyel kurumların müşterilere yaklaşımı değişmiştir. Bilgi ve haberleşme teknolojilerinin gelişmesi, internetin günlük hayatta etkinliğinin artması, piyasa rekabetinin sertleşmesi, ürün ve hizmet hakkında bilgiye erişim yollarının farklılaşması özellikle endüstriyel kurumları eskiye göre daha fazla müşteri davranışlarını inceleme ve bunlar için farklı pazarlama yöntemleri geliştirme zorunda bırakmıştır.³⁸ Kar amacı güden kurumlar müşteri bekleyen değil, müşteri arayan konuma gelmiştir.

Yaş aralığı, ilgi alanı, gelir düzeyi gibi geleneksel sınıflandırmalarla belirlenen hedef kitlelere dayandırılarak yapılan tanıtım, reklam çalışmaları düşük yanıt oranıyla sonuçlanmaktadır. Günümüzde çok farklı zevk ve tercihlere sahip müşterileri büyük gruplara ayırarak pazarlama yöntemi geliştirmek yeterli olmamaktadır. Müşteriler kendi bireysel ve eşsiz gereksinimlerine yönelik hizmet verilmesini beklemektedirler.³⁹

Veri madenciliği uygulamaları, pazarlama yönetiminde müşteri tercihlerinin belirlenmesinde kullanılır. Veri madenciliği yeni müşteriler kazanmayı olduğu gibi var olan müşteriyi korumayı da hedefleyebilir. Burada veri madenciliğinin amacı müşterilerin özelliklerini belirlemek, kurumla ilişkilerini takip etmek, rakip şirkete gidecek müşterileri saptamak ve karar vericilerin veri madenciliği sonuçlarından yararlanarak yeni stratejiler üretmesine destek olmaktadır.⁴⁰

Pazarlama sektöründe müşteri ilişkileri yönetimi büyük rol oynamaktadır. Müşteri ilişkileri yönetimi, endüstriyel kurumların müşteri odaklı pazarlama anlayışı geliştirmesidir. Müşteri ilişkileri göz önünde bulundurularak geliştirilen stratejiler,

³⁸ YongSeog Kim, W. Nick Street, "An Intelligent System for Customer Targeting: A Data Mining Approach," **Decision Support Systems**, No:37, 2004, s. 216.

³⁹ Michael J. Shaw, v.d., "Knowledge Management and Data Mining for Marketing," **Decision Support Systems**, No:31, s. 132-133.

⁴⁰ Bryan Bergeron, **Essentials of CRM: A Guide to Customer Relations Management**, USA, John Wiley&Sons, 2002, s. 80.

müşterilerle birebir iletişim içinde olan, onların satın alma davranışlarını takip eden ve bunlarla ilgili veri tabanları oluşturan, kurum ve müşterileri arasında iletişim kuracak teknolojilere sahip özellikler içerir.⁴¹ Böylece müşterilerle ilgili birçok kanaldan veri, enformasyon elde edilmiş olur. Veri madenciliği uygulamalarıyla bu verilerden anlamlı örüntüler çıkartılarak müşteri profilleri oluşturulur ve müşteriye yönelik kişiselleştirilmiş ürün, hizmet sunulur.⁴² Kümeleme ve sınıflandırma gibi teknikler kullanılarak benzer özellikler gösteren müşteri grupları ortaya çıkartılır.⁴³

Müşterinin ürünü, hizmeti ne kadar kullanacağı, ne zaman kullanmayı bırakacağı bile öngörülebilir.⁴⁴

Veri madenciliğiyle pazarlama maliyetlerinin düşürülmesi de sağlanabilir. Herhangi bir kampanyaya telefon, e-posta veya internet aracılığıyla geri bildirim yapması düşük olan müşteriler elimine edilerek maliyette tasarruf edilebilir.⁴⁵

Pazarlama analizinde veri madenciliği uygulamaları kullanılarak;⁴⁶

- Benzer davranış (satın alma, terk etme, vb.) gösteren müşteriler için bir örüntü belirlenmesi ve hedef pazarın oluşturulması,
- Çapraz satış⁴⁷ yöntemiyle ürünler arası ilişkilerin ortaya çıkartılması, yani birlikte satılabilecek ürünlerin ortaya çıkartılması ve raf sisteminin buna göre düzenlenmesi,⁴⁸

⁴¹ Filiz Otay DEMİR, Yalçın KIRDAR, “Müşteri İlişkileri Yönetimi: CRM”, **Review of Social, Economic & Business Studies**, C.VII, No:8, s.297.

⁴² Michael J. Shaw, v.d., “Knowledge Management and Data Mining for Marketing,” **Decision Support Systems**, No:31, s. 133.

⁴³ Sang C. Park, Selwyn Piramuthu, Michael J. Shaw, “Dynamic Rule Refinement in Knowledgebased Data Mining Systems”, **Decision Support Systems**, No:31, 2001, s.206.

⁴⁴ Chris Rygielski, Jyun-Cheng Wang, David C. Yen, “Data Mining Techniques for Customer Relationship Management,” **Technology in Society**, No:24, 2002, s. 494.

⁴⁵ Michael J.A Berry, Gordon Linoff, **Data Mining Solutions**, U.S.A., Wiley Computer Publishing, 1998, s.14.

⁴⁶ **Perakende Sektöründe Veri Madenciliği**, (Çevrimiçi) http://www.spss.com.tr/pdfs/SPSSPerakendeRaporu_web2010.pdf, 16.12.2013.

- Alışveriş olasılığının arttığı özel günlerde, alışveriş potansiyeli yüksek olan müşterilerin tespit edilmesi ve bunlara uygun duyuruların gerçekleştirilmesi,
- Müşterilerin gereksinimlerinin belirlenmesi ve farklı müşteriler için en iyi ürünlerin neler olduğunun belirlenmesi,
- Müşterinin fiyat artışı ile değişen satın alma alışkanlıklarının belirlenmesi,
- Yeni müşterileri çekmede hangi faktörlerin etkili olacağının ön görülmesi gibi saptamalarda bulunulabilir.

2.4.2. Finans

Finans sektörü günümüz ekonomi düzeninde her geçen gün büyük ölçekte veriler üretmektedir. Karar vericiler için bu verilerden verimli bilgilerin çıkarılması ve bu bilgilerin sektöre yeniden dönüşümü oldukça önemlidir. Son yıllarda bilgi teknolojilerinin de gelişmesi üzerine bankacılık ve sigortacılık işlemleri artık internette yapılabilmektedir. Bu durum kullanıcılar için kolaylık sağlamış ancak sektördeki veri artışını hızlandırmıştır. Ayrıca dolandırıcılık, hilekârlık gibi bazı sorunlar ve riskler de ortaya çıkmıştır.

Bilgiye dayalı yönetimlerde verinin sağlıklı ve doğru akışını sağlamak için analiz edilmesi gereklidir. Bu gereklilik veri madenciliğinin bu alanda yer almasına neden olmuştur.

⁴⁷ Çapraz satış, bir ürün veya hizmet alan müşteriye, bir başka ürün veya hizmeti kendisi talep etmeden satış önerisinde bulunmaktır.

⁴⁸ Shu-Hsien Liao, Yin-Yu Chen, "Mining Customer Knowledge for Electronic Catalog Marketing", **Expert Systems with Applications**, C.XXVII, No:4, 2004, s. 522.

Finans kurumları veri madenciliğini etkin biçimde;⁴⁹

- Kredi kartlarında borç takibi, limit belirleme gibi işlemlerde,
- Ev, araba, yatırım kredilerinde,
- Çapraz satış, dikey satış, kar analizlerinde,
- Risk yönetiminde,
- Dolandırıcılığın saptamasında kullanılmaktadır.

Veri madenciliği uygulamaları kullanılarak bir banka için yapılan müşteri kaybı analiziyle (*churn analizi*), banka müşterilerinin gelecek altı ay içerisinde kurumu terk edip etmeyeceği öngörülme çalışılabilir. Böylece banka çeşitli kampanyalarla müşterisinin ilgisini yeniden çekip onu elde tutmaya çalışacaktır.

Bankacılıkta uygulanan kredi skorlama/puanlama analiziyle kredi başvurusunda bulunan kişinin veya kurumun ön görülen sürede krediyi geri ödeme olasılığı yorumlanmaktadır. Bu analizle başvuru sahibinin ödemelerde kaç kere gecikmeye düşebileceği bile öngörülebilir.⁵⁰ Kredi skorlarının belirlenmesinde, klasik istatistik teknikleri yerine veri madenciliği tekniklerinin kullanılmasıyla daha iyi verim alındığı görülmüştür.⁵¹

Davranışsal skorlama analizi ise müşterilerin geçmiş satın alma davranışlarını inceleyerek gelecekteki davranışlarını öngörmeye çalışır.⁵² Örneğin kredi kartını

⁴⁹ Önder Halisdemir, “Bireysel Kredi Riskinin Ölçümünde Kredi Skorlaması Tekniğinin Kullanımı”, **Active Bankacılık ve Finans Dergisi**, No:25, 2002, s. 77.

⁵⁰ Önder Halisdemir, “Bireysel Kredi Riskinin Ölçümünde Kredi Skorlaması Tekniğinin Kullanımı”, **Active Bankacılık ve Finans Dergisi**, No:25, 2002, s. 66.

⁵¹ Hui Wang, Andreas S. Weigend, “Data Mining for Financial Decision Making,” **Decision Support Systems**, No:37, 2004, s. 459.

⁵² Nan-Chen Hsieh, “An Integrated Data Mining and Behavioral Scoring Model for Analyzing Bank Customers”, **Expert Systems with Applications**, C.XXVII, No:4, s.3.

yurtdışında kullandığı görülen müşterinin, çeşitli yurtdışı kampanyalarıyla bilgilendirilerek, yeniden kartla harcama yapması sağlanır.

Finans sektöründe, bankacılık ağırlıklı olarak uygulama yaygınlığına sahip olan veri madenciliğinin sigortacılık ve sermaye piyasası uygulamaları da gün geçtikçe artış göstermektedir.⁵³ Özellikle banka bağlantılı sigorta işletmeleri, banka veritabanında analizler yaparak yeni müşteriler kazanmak için çeşitli kampanyalar yapmaktadır.

Bir sigorta şirketinin müşteri veritabanına veri madenciliği teknikleri uygulanarak kaza, konut, emeklilik gibi sigortaları hangi müşterilerin tercih edebileceği öngörülmektedir.

Son yıllarda önemli markalara sahip işletmelerin finansal zararlara uğraması nedeniyle işletmelerde muhasebe denetiminin önemi artmıştır. Finansal kayıpları azaltmanın en etkin yolu, hile ve usulsüzlüklerin önlenmesi ile olmaktadır. Özellikle günümüz teknolojik ortamında geniş bir veritabanına sahip olan işletmelerde verilerin analiz edilerek olağan olmayan hareket ve işlemlerin tespit edilmesinde veri madenciliğinden yararlanılmaktadır. Bu amaçla çok sayıda veri madenciliği teknikleri hile tespiti için finans şirketlerinde kullanılmaktadır. Veri madenciliği ile sadece hile ve usulsüzlükler tespit edilmemekte, ayrıca hile ve usulsüzlüklerin önlenmesi de mümkün olmaktadır.⁵⁴

2.4.3. Tıp ve Sağlık

Tıp ve sağlık sektörünün gelişmesiyle artan insan yaşamının süresi bazı sorunları da beraberinde getirmiştir. Örneğin birçok insan kalp, diyabet ve astım

⁵³ Ali Serhan Koyuncugil, Veri Madenciliği Ve Sermaye Piyasalarına Uygulanması, **Sermaye Piyasası Kurulu Araştırma Raporu**, 2007, s.1.

⁵⁴ Serkan Terzi, “Hile ve Usulsüzlüklerin Tespitinde Veri Madenciliğinin Kullanımı”, **Muhasebe ve Finansman Dergisi**, No:54, 2012, s.56.

hastalıkları gibi kronik hastalıklarla yaşamak zorunda kalmıştır. Bu hastalıkların hem tıbbi açıdan hem de hastanenin kaynak ve maliyeti açısından ele alınarak doğru yönetilmesi gerekir. Hastane bilgi sisteminin klasik sorgulamalarla analizi yetersiz kalmakta, veri madenciliği gibi yöntemlerle daha yararlı ve anlamlı olacak gizli kalmış bilgilerin keşfedilmesi gerekmektedir.⁵⁵

Hasta veritabanını oluşturan elektronik tıbbi kayıt verilerinde veri madenciliği teknikleri uygulanarak çeşitli öngörüler yapılabilmektedir. Örneğin;⁵⁶

- Aynı hastalığa sahip kişilerin ortak nitelikleri çözümlenerek hastalığa karşı yeni tedaviler üretilmesi,
- Tıbbi tedavinin sonuçları incelenerek daha etkili çözümler üretilmesi,
- Ölüm oranları ve salgın hastalıkların öngörülmesi,
- Hastane ve ilaç maliyetlerinin öngörülmesi.

Hasta veritabanında veri madenciliği kullanılarak yapılan bir analizde hastaların muayene için farklı doktorları tercih etmesinin nedeni araştırılmış ve yaşın, cinsiyetin, hastaneye gidiş sıklığının, kronik rahatsızlıkların farklı doktorlara gitmelerinde etkili olduğu belirlenmiştir.⁵⁷

Veri madenciliğinden, bir ilacın hangi yaş gruplarındaki hastalarda daha etkili olacağının öngörülmesinde, kanser tedavisinde hasta için en iyi tedavi yönteminin ne olduğunun belirlenmesinde yararlanılır.⁵⁸

⁵⁵ Pınar YILDIRIM, Mahmut ULUDAĞ, Abdülkadir GÖRÜR, “Hastane Bilgi Sistemlerinde Veri Madenciliği”, **Akademik Bilişim 2008 Konferansı**, 2007.

⁵⁶ Stephan Kudyba, **Managing Data Mining: Advice from Experts**, USA, CyberTech Publishing, 2004, s.146.

⁵⁷ Yu-Chun Chen, Shiao-Chi Wu, “Exploring Out-Patient Behaviors in Claim Database: A Case Study Using Association Rules”, **AMIA Annu Symposium Proceedings**, 2003, s.811.

⁵⁸ Daniel T. Larose, **Discovering Knowledge in Data: An Introduction to Data Mining**, John Wiley & Sons Inc., 2005, s. 42.

Hastaların anormal davranışları üzerine yapılan bazı araştırma ve incelemelerde; koleranın tedavisi, şizofreni ve kişinin doğduğu ay arasındaki bağ, yüzsek dozaj uyuşturucu alıp ölenler ve içinde buldukları ay arasındaki bağıntılar veri madenciliği yöntemleriyle tespit edilebilir.⁵⁹

Biyolojik veritabanları çok çeşitli, zengin ilişkisel yapıya sahip veri türlerini barındırır. Biyolojik veri tabanlarında gen yapıları incelenerek, RNA ve DNA moleküllerinin sırasının belirlenmesinde, protein yapılarının ve biyokimyasal reaksiyonların öngörüsünde veri madenciliği kullanılmaktadır.⁶⁰

Hastaların tedavi kayıtları veri tabanlarında saklanabilir ve bunlar üzerine veri madenciliği teknikleri uygulanarak tedavide uygulanacak yöntem belirlenebilir.

- Bir kanser hastası için yalnızca kemoterapi veya radyoterapi mi uygulanmalı yoksa ikisi beraber mi uygulanmalı?
- Doktorlar bir diyaliz hastasının tedavisinin etkinliğini artırmak için ne yapabilir?
- İnsan DNA veritabanları genetik kodlama modelleri oluşturmak üzere hastalıklara karşı örneklenebilir mi?

2.4.4. Haberleşme ve Endüstri

Haberleşme ve iletişim sektöründe günümüzde büyük rekabet yaşanmaktadır. Bu da kurumları sürekli müşteri kaybetme riskiyle baş başa bırakmaktadır. Kurumlar, müşterileri iletişim veritabanını analiz ederek yeni yöntemler,

⁵⁹ John Roddick, "Exploratory Medical Knowledge Discovery: Experiences and Issues", *ACM SIGKDD Explorations Newsletter*, C.V, No:1, 2003, s. 94.

⁶⁰ David Page, Mark Craven, "Biological Applications of Multi-Relational Data Mining," *ACM SIGKDD Explorations Newsletter*, C.V, No:1, 2003, s.69.

kampanyalar geliřtirmek zorunluluęundadır. Veri madencilięi analiz için bařvurulacak en etkili yöntemdir.

Haberleřme kurumları müřterilerinin detaylı arama kayıtlarına sahiptir. Bu veriler üzerinde veri madencilięi kullanarak analizler yapmakta ve benzer özellik taşıyan müřterilerini sınıflandırıp yeni fiyatlandırma, tarifelendirme ve kampanya seçenekleri oluřturmaktadır.⁶¹ Veri madencilięi uygulamaları kullanılarak yapılacak müřteri kaybı analiziyle (*churn analizi*) ayrılma eğilimi gösteren müřteriler belirlenerek onlara özel kampanyalarla müřteri kaybı engellenebilir. Örneęin, kurum aylık belirli bir konuřma kotasını ařan müřterilerine ücretsiz konuřma olanaęı verebilir, hatta cep telefonu gibi hediyeler önerebilir. Aynı řekilde belirli saatlerde telefonu kullanmadıęı görölen müřteriyi o saatlerde konuřturabilmek için çeřitli kampanyalar sunulabilir. Yoęun konuřma saatleri belirlenerek o saatlerde daha fazla iřgücü ve altyapı saęlanarak olası aksamaların önüne geçilebilir.

Haberleřme ve iletiřim kurumları veri madencilięini;

- Var olan müřterilerden en yüksek kazancı saęlamak,
- Karlı müřterilerin kuruma baęlılıęını arttırmak ve zarar edeceęini düřündüęü müřterilerini uzaklařtırmak,
- Rakip řirketlerden müřteri çalmaya çalıřmak için kullanabilir.⁶²

Enerji üretimi veya daęıtımı yapan endüstriyel kurumların, tüketicilerin tüketim alışkanlıklarının belirlenmesinde ve bunların analiz edilip tüketicilere buna

⁶¹ Chris Rygielski, Jyun-Chen Wang, David C. Yen, "Data Mining Techniques for Customer Relationship Management", **Technology in Society**, C.XXIV, No:4, 2002, s.490.

⁶² Umman Tuęba řimřek Gürsoy, "Customer churn analysis in telecommunication sector", **İstanbul Üniversitesi İřletme Fakóltesi Dergisi**, C.XXXIX, No:1, 2010, s. 38.

bağlı tarifeler sunmasında veri madenciliği yöntemleri olumlu sonuçlar vermektedir.⁶³

2.4.5. Bilgi ve Belge Yönetimi

Veri madenciliğinin kütüphane ve arşivlerde kullanımı henüz yaygın değildir. Yaygın olmamasının en büyük nedeni, veri madenciliği yazılımlarının satın alınmasının ve kullanılmasının maliyetli olmasıdır. Ancak açık kaynak kodlu ve kullanıcı dostu arayüzlü yazılımlar geliştirildikçe, bu yazılımlar kütüphane ve arşiv gibi bilgi merkezlerinde kullanılmaya başlanacaktır.

İşletmelerde, kullanıcı veya müşterilerin ürün ve hizmetlerden ne kadar yararlandıkları, ne kadar memnun kaldıkları ancak onların davranışları izlenerek belirlenebilir. Bilgi merkezleri de kullanıcılarının davranışlarını izleyen ve elde ettiği verilere göre kaynak ve hizmetlerini geliştiren kurumlardır. Kullanılan kütüphane ve arşiv otomasyon yazılımları sayesinde oluşan verinin incelenmesi, bilgi merkezi kullanıcıları hakkında anlamlı bilgiler sunacaktır. Oluşan bu verilerden anlamlı bilgiye ulaşabilmek için bazı özel yöntem ve teknikler kullanılması gerekmektedir. Veri madenciliği, verilerin analiz edilmesi, yorumlanması, amaç ve hedeflere ulaşmada gerekli adımları önermesi sürecidir.⁶⁴

Kütüphanelerde veri madenciliği yaklaşımındaki amaç; hizmet kalitesini, kütüphane performansını, okuyucu memnuniyetini artırmak ve gereksiz materyallerin satın alınmasını engelleyerek maliyeti düşürmektir.

Kütüphane otomasyon yazılımlarından, yerel ağdan veya internetten kütüphaneyi kullanan kullanıcılar hakkında veri toplanabilmektedir. Veri madenciliği

⁶³ Marino Sforna, "Data Mining in a Power Company Customer Database," **Electric Power Systems Research**, C.LV, No:1, 2000, s. 201.

⁶⁴ Sacit Arslantekin, "Veri Madenciliği ve Bilgi Merkezleri", **Türk Kütüphaneciliği**, C.XVII, No:4, 2003, s.369.

teknikleri sayesinde bu verilerin analiz edilmesiyle “kütüphaneyi kimler kullanıyor”, “hangi aralıkta kullanıyorlar”, “hangi hizmeti kullanmışlar”, “hangi kaynakları ödünç almışlar” gibi soruların cevaplarını alabiliriz. Bu soruların cevapları sayesinde; kütüphanenin hangi bölümlerinin ne sıklıkta kullanıldığı, nasıl daha verimli kullanıma sunulabileceği gibi sonuçlar elde edebiliriz. Otomasyondan veya internetten yapılan taramalarda tutulan bilgiler (tarama yapılan anahtar kelimeler vs.) bir sonraki tarama sırasında yardımcı olabilmektedir. Bir önceki taramada kullanılan anahtar kelimeler yeni araştırmada bulunulan nokta konusunda bilgi vermektedir. Burada önemli olan nokta, geleneksel taramalarda kullanılan anahtar kelimeyle ilişkilendirilerek yeni önerilerde bulunma değil, eski taramalar kullanılarak ilişkili yeni öneriler sunma olmaktadır. Örneğin, kütüphane veritabanında “veri madenciliği” konusunu araştıran bir kullanıcıya, daha önce bu konuda araştırma yapmış diğer kullanıcıların yararlandığı kaynaklar sistem tarafından önerilebilir.

Uluslararası veritabanları bilgi merkezlerinin büyük yatırım yaptığı kaynaklardır. Bu yatırım, kullanıcıların veritabanlarını ne kadar, nasıl kullandığını, ne derecede yararlandığını öğrenmeyi zorunlu kılmaktadır. Bu zorunluluk kullanıcıların veritabanları üzerindeki işlemlerini analiz etmeyi gerektirmektedir. Veri madenciliği ile kullanıcı davranışları incelenip her kullanıcı için ayrı ayrı profil belirlenerek veya kullanıcılar gruplanarak, kütüphane hizmetleri otomatik olarak detaylandırılabilir.⁶⁵

Veri madenciliği tekniklerinin kütüphane otomasyon sistemi ve web sitesinde uygulanması sonucunda verilen hizmetlerin kalitesi ve performansı artacak ve kullanıcı hizmetleri amacına ulaşacaktır. Bu tekniklerinin uygulanmasıyla kullanıcıya özel hizmet verilmesi olanaklı hale gelmiştir. Örneğin, bir kullanıcı kütüphane üyeliğine kayıt olurken ilgi alanı olarak bilgisayarı seçtiğinde bilgisayar ile ilgili yeni bir kitap kütüphaneye geldiğinde veya içinde bilgisayar bilimleri bulunan bir çevrimiçi veritabanına üyelik yapıldığında otomatik olarak bu kullanıcı

⁶⁵ A.e., s.380.

bilgilendirilebilecektir.⁶⁶ Aynı şekilde tahmin edici veri madenciliği yöntemleri kullanılarak, bir materyali ödünç almak isteyen kullanıcının o materyali zamanında geri getirip getirmeyeceği tahmin edilebilir ve buna göre önlem alınabilir.

Kitapların sayısal ortamda depolanabilir olması sonucunda oluşan elektronik kütüphanelerin sayısı günümüzde hızla artmaktadır. E-Kütüphanelere yerel ağlar veya internet aracılığı ile erişmek mümkün oluyor. Bilgiye erişmek kolaylaştıkça, ona erişme sayısı artıyor ve her erişim ardında büyük veri yığınları bırakıyor. Bu verilerden yararlı bilgiler çıkarılması için de veri madenciliği gibi tekniklere gereksinim duyulacaktır.

Kütüphanelerde veri madenciliği kullanımı, geleneksel kataloglamaya göre kütüphane kaynaklarına daha hızlı ve kapsamlı erişim sağlar ve bilgisayar veya teknolojik bilgisi olmayan kullanıcılara herhangi bir yardım almadan aradıkları materyale erişebilme olanağı verir.

Arşivlerde, veri madenciliğinin bir alt kolu olarak gelişen doküman madenciliği kullanılabilir. Doküman madenciliği arşivde belirli bir dokümana benzer dokümanların bulunması sonucunda, “arşivde bu dokümana benzer hangi dokümanlar var?” gibi soruların yanıtlarını bulacak şekilde analiz yapmaktadır. Bu çalışmada amaç dokümanlar arasında ayrıca elle bir ayırım gerekmeden benzerlik hesaplayabilmektir. Bu da genelde otomatik olarak çıkarılan anahtar sözcüklerin tekrar sayısı ile yapılır.⁶⁷

Veri madenciliğinin bilgi merkezlerinde kullanılması araştırmalarında ortaya çıkan başka bir kavram da “bibliomining”tir. “Bibliomining” terimi ilk olarak 2003’te Nicholson ve Stanson tarafından kütüphaneler için veri madenciliğini

⁶⁶ Hidayet Takçı, "Tam Otomatik Kütüphane Web Sitesi", ÜNAK'02: I. ÜNAK Genel Konferansı, Samsun, 2002, s.3.

⁶⁷ Yılmaz Argüden, Burak Erşahin, **Veri Madenciliği**, ARGE Danışmanlık, İstanbul, 2008, s.34.

tartışmak üzere kullanılmış, “veri madenciliği” ve “bibliyometrik” terimlerinin birleşiminden oluşmuştur.

Kütüphane sistemlerinde büyük yığınlardan oluşan veriyi düzenleme tekniklerinin belirlenmesi amacıyla, istatistiksel araçların kullanılmasıyla ortaya çıkan “bibliomining” süreci; odaklanılacak alanın belirlenmesi, iç ve dış veri kaynaklarının belirlenmesi, verinin veri ambarı içinde toplanması ve veri kirliliğinin yok edilmesi, uygun analiz araçlarının seçilmesi, raporların oluşturulması için gerekli geleneksel araçların belirlenmesi, analizin yapılması ve sonuçların uygulanması şeklinde ortaya çıkar. Bu sayede kütüphane kullanıcıları için yararlanılmak veya araştırılmak istenen sahaya daha kolay erişim, kütüphaneciler için de kaynakların hitap ettiği hedef kitle için hizmet kolaylığı sağlanmış olur. Ayrıca “bibliomining” sayesinde kütüphane yöneticilerinin bütçelerini daha fazla gereksinim duyulan kaynaklara yönlmesine yardımcı olur. Diğer bir yararı ise, kütüphanelerin daha sık kullanıldığı zaman dilimlerinin gözlemlenerek personel miktarının gereksinime göre belirlenip hizmet kalitesinin artırılmasıdır.⁶⁸

Kütüphane otomasyon sistemleri sayesinde, kütüphane dermesi ve kullanıcıları hakkında veriler toplanır. Ancak bu veriler çoğu zaman kütüphane hakkında daha iyi karar verme amacıyla kullanılmamaktadır. Veri madenciliği uygulamalarıyla eldeki verilere daha etkin bir yaklaşım sağlanır.

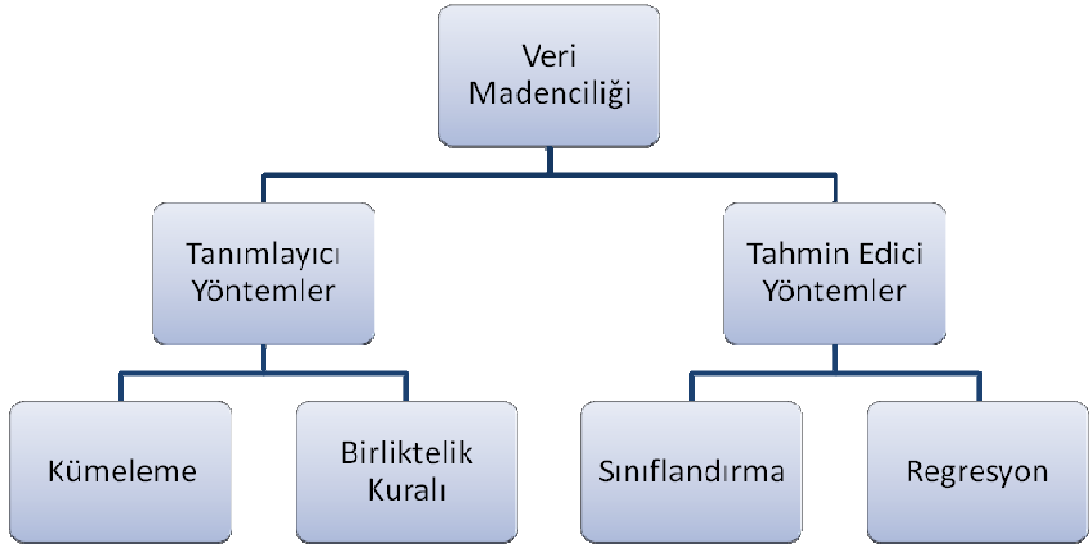
Veri madenciliğinin etkin kullanımıyla kütüphaneler, çevrimiçi satış yapan kitap siteleriyle, internette var olan indirilebilir e-kitaplarla, sesli kitaplarla, videolarla, farklı kalitedeki uçsuz bucaksız ücretsiz bilgiyle rekabet edebilir duruma gelecektir.

⁶⁸ Scott Nicholson, The Bibliomining Process: Data Warehousing and Data Mining for Library Decision-Making, **Information Technology and Libraries**, C.XXII, No:4, 2003, s.146.

3. VERİ MADENCİLİĞİ YÖNTEM VE TEKNİKLERİ

3.1. Veri Madenciliği Yöntemleri

Veri madenciliği temel olarak, yazılım teknikleri kullanılarak verilerin analiz edilmesiyle ilgilidir. Bu analiz doğrulamaya dayalı yöntemle veya keşfetmeye dayalı yöntemle yapılabilir. Bu yöntemlerden kastedilen, verinin bilgiye dönüştürülmesinde nasıl bir yol izleneceğini açıklamaktır.



Şekil 6. Veri Madenciliği Yöntemleri (Oded Maimon, Lior Rokach, Data Mining and Knowledge Discovery Handbook)

Doğrulamaya dayalı yöntem, yeni bilgi üretmez. Geleneksel veritabanlarıyla yakından ilgili, çoğu kez basit istatistiksel analizlerle birlikte, temel olarak sorgulara ve raporlamaya dayalı bir yöntemdir. Amacı, bir varsayımı veritabanlarındaki ilişkileri kullanarak desteklemektir.

Keşfetmeye dayalı yöntem, veritabanlarındaki verilerden yeni bilgilerin üretilmesine olanak sağlar. Bu yöntem yeni bilgileri tanımlama veya tahmin etmede

kullanılır. Bu doğrultuda keşfetmeye dayalı yöntemler, tahmin edici (*predictive*) ve tanımlayıcı (*descriptive*) olmak üzere iki ana başlık altında incelenmektedir.

Tahmin edici yöntemler bilinen verilerden yararlanarak, bilinmeyen bir değeri tahmin etmeye çalışırlar. Tahminde amaç, bazı değişkenlerin (kullanıcı, müşteri gibi) davranışlarının gelecekte ne olacağını bilmek için verilerdeki desenlerin belirlenmesiyle ilgilidir. Bu amaç, sınıflama (*classification*), regresyon (*regression*) ve zaman serileri (*time series*) gibi yöntemlerle gerçekleşir.

Tanımlayıcı yöntemler ise verilerdeki gizli ortak özellikleri ve ilişkileri araştırırlar. Bunlar da kümeleme (*clustering*), özetleme (*summarization*) ve birliktelik kuralı (*association rule mining*) gibi yöntemlerdir.¹

Bu yöntemler bazı kaynaklarda ortak noktaları olmakla birlikte farklı biçimde tanımlanmalarına rağmen, genel bir kabul olarak yukarıdaki şekilde iki temel başlık altında incelenmektedir.²

Hangi yöntemin kullanılacağına veriye bakılarak karar verilir.³

Tahmin edici ve tanımlayıcı yöntemlerin her ikisi de veri madenciliği teknikleriyle desteklenir.

3.1.1. Tahmin Edici Yöntemler

Sonuçları bilinen veriler incelenerek yeni bir yöntem geliştirilir. Kurulan bu yöntemden yararlanarak sonuçları bilinmeyen veri kümeleri için yeni sonuçlar

¹ Rachid Anane, "Data Mining and Serial Documents", **Computers and the Humanities**, C.XXXV, No:3, 2001, s.300.

² Haldun Akpınar, "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", **İ.Ü. İşletme Fakültesi Dergisi**, 2000, Cilt 29, No:1, s. 1.

³ Catherine Bounsaythip, Esa Rinta-Runsala, "Overview of Data Mining for Customer Behaviour Modeling", **VTT Information Technology Research Report TTEI**, 2001, s.14.

öngörmeye çalışmaktır.⁴ Bu yöntemde öngörme, bir yargıya varma, sınıflandırma gibi işlevler önplandadır.⁵ Örneğin bir kütüphane, kendisinden daha önce herhangi bir kitap almış bir kullanıcısıyla ilgili bütün verilere sahiptir. Burada kullanıcının özellikleri, daha önce kaç tane kitap, dergi, materyal aldığı ve aldıklarını ne zaman getirdiği ile ilgili veriler bilinenlerdir. Yeni alacağı kitabı ne kadar sürede geri getireceği ise bilinmeyen veridir. Bu verilere uygun olarak kurulan yöntemle kullanıcının aldığı ödünç kitabı ne zaman getireceği tahmin edilebilir. Yine aynı kullanıcının aldığı materyalleri zamanında geri getirme verileri düşükse, kütüphane yönetiminin kullanıcıya karşı gelecekteki davranışı belirlenebilir.

Bankalar bu yöntemi kredilendirme sisteminde sıklıkla kullanırlar. Bir banka, müşterisine daha önce vermiş olduğu kredinin geri ödenip ödenmediğini, veritabanındaki verilerden yararlanarak bulabilir ve buna göre yeni kredinin miktarı veya ödenme zamanı hakkında tahminde bulunur.⁶

3.1.1.1. Sınıflandırma

Veritabanlarında işlemler değişken tanımlanarak yapılır. Değişken, bir niceliği veya ölçülebilir niteliği belirten ve değişkenlik gösteren sembolik değerlerdir. Belli bir türe ait verileri saklarlar. Bir sistemin işletilmesinde veya bir deneyin gidişatı doğrultusunda değişebilecek değerlerin yerlerini tutan varsayımlardır. Örneğin, bir kütüphane veritabanında kullanıcı adı, soyadı şeklinde girilen veriler, sistem tabanında “kullanıcı” diye bir değişkenle tanımlanır. Bu değişken, kütüphaneyi kullanan herhangi bir kullanıcı olarak değerlendirilebilir.

⁴ Ning Zhong, Lizhu Zhou, **Methodologies for Knowledge Discovery and Data Mining: Third Pacific-Asia Conference, Pakdd-99, Beijing, China, April 26-28, 1999** : Proceedings, Springer Verlag, 1999.

⁵ Mehmed M. Kantardzic, “General Data Analysis and Data Mining”, **Wiley-IEEE Press**, 2002, s. 1.

⁶ Serhat ÖZEKES, “Veri Madenciliği Modelleri Ve Uygulama Alanları”, **İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi**, C.III, No:3, 2003, s.67.

Bağımsız değişken, bizim atadığımız değişkendir. Bağımlı değişken ise bizim atadığımız değişkene, yani bağımsız değişkene bağlı olarak değişen değişkendir. Bir analizde bağımsız değişkenin bağımlı değişken üzerinde etkisi incelenir. Örneğin; “x” bir bağımsız değişkendir, “y” ise bağımlı değişkendir ve aldığı değer “x”in aldığı değere göre değişir. “x” açıklayıcı, tahmin edici, “y” ise açıklanan, tahmin edilen değişkendir.

Veri madenciliği uygulamalarında çok kullanılan bir yöntem olan sınıflandırma,⁷ bir verinin niteliğini inceleme ve bu veriyi önceden belirlenmiş bir sınıfa atamaktır. Burada önemli olan her bir sınıfın özelliklerinin daha önceden açıkça belirlenmiş olmasıdır. Önceden belirlenmiş sınıflar, veri ambarından veya veritabanından alınan verinin sınıflandırılması için yöntem geliştirmede kullanılır.⁸

İstenilen bir değişken bağımlı değişken olarak atanır, diğerleri ise bağımsız değişkenlerdir. Bağımsız değişkenler tahmin edici değişkenlerdir. Sınıflandırmada amaç, tahmin edici değişkenlerin yer aldığı modelde, bağımlı değişkenin değerinin bulunduğu anlamlı bir yöntem kurmaktır.⁹

Tablo 3. Sınıflandırma Yöntemiyle Hastalık Teşhisi

Vaka	Özellikler			Grip Teşhisi
	Ateş	Baş Ağrısı	Bulantı	
1	37.8	Evet	Hayır	Evet
2	39.2	Evet	Evet	Evet
3	37.8	Hayır	No	No
4	37.5	Evet	Evet	Evet
5	37.6	Evet	Evet	No
6	35.7	Evet	No	No
7	35.8	Hayır	Evet	No
8	39.2	Evet	Evet	Evet

⁷ Michael Goebel, Le Gruenwald, “A Survey of Data Mining and Knowledge Discovery Software Tools,” *ACM SIGKDD Explorations*, 1999, C.I, No:1, s. 23.

⁸ Bryan Bergeron, *Bioinformatics Computing*, U.S.A., Prentice Hall PTR, 2002, s. 120.

⁹ V. Ganti, J. Gehrke ve R. Ramakrishnan, “Mining Very Large Databases”, *IEEE Computer*, C.XXXII, No:8, 1999, s.38.

Sınıflandırma, bir ürünün ve müşterinin özelliklerini belirleyerek, bu iki değişkenin birbirleriyle eşleşmesini sağlar. Bu durumda müşteri için ideal ürün, ürün için de ideal müşteri belirlenir. Örneğin, bir otomobil satıcısı, geçmiş müşteri hareketlerinin analizi ile “genç kadınlar küçük araba, yaşlı ve zengin erkekler büyük ve lüks araba satın alır” gibi bir kural bulursa, genç kadınların okuduğu bir dergiye küçük model bir otomobil reklamı verecektir.¹⁰

Sınıflandırmaya örnek olarak kredi kartı başvurularını düşük, orta ve yüksek risk grubu olarak ayırmak gösterilebilir.

Bu yöntemin en önemli işlevi, sınıflandırma sonrasında varolan kayıtlardaki kişilerin, alanların, nesnelerin, kurumların niteliklerini ortaya çıkarmasıdır. Bu sınıflar belirli kullanıcı veya müşteri grupları için oluşturulmuşsa, burada öncelikli amaç her grubun profilini çıkarmak olmalıdır. Örneğin, internet bankacılığında yaptıkları elektronik fon transferi sıklıklarına göre internet müşterileri; “seyrek” , “orta sıklıkta” ve “sık” kullanıcı olarak sınıflandırılabilir. Müşteriler bu şekilde gruplandıktan sonra amaç ise, bu grupların tutum ve davranışlarını değiştirecek yöntemler geliştirmektir.¹¹

Bir halk kütüphanesi, kütüphaneye ilgi çekebilme ve kütüphanenin daha fazla kullanılmasını sağlamak için okuyucularına kütüphaneye yeni gelen yayınlarla ilgili elektronik posta gönderebilir. Ancak her gelen yayının duyurusunun bütün okuyuculara gönderilmesi ilgi çekmekten çok caydırıcı olur. Ancak okuyucular, geçmiş ödünç alma ve arama kayıtlarından oluşan verilerin analiziyle belirli sınıflara ayrılırsa, sadece o sınıfla ilgili kategorideki kitapların duyurusunun gönderilmesi okuyucunun dikkatini çekebilir. Halk kütüphanelerinin kullanımının gittikçe azaldığı günümüzde okuyucu, bu tür yöntemler kullanılarak kütüphaneye çekilebilir.

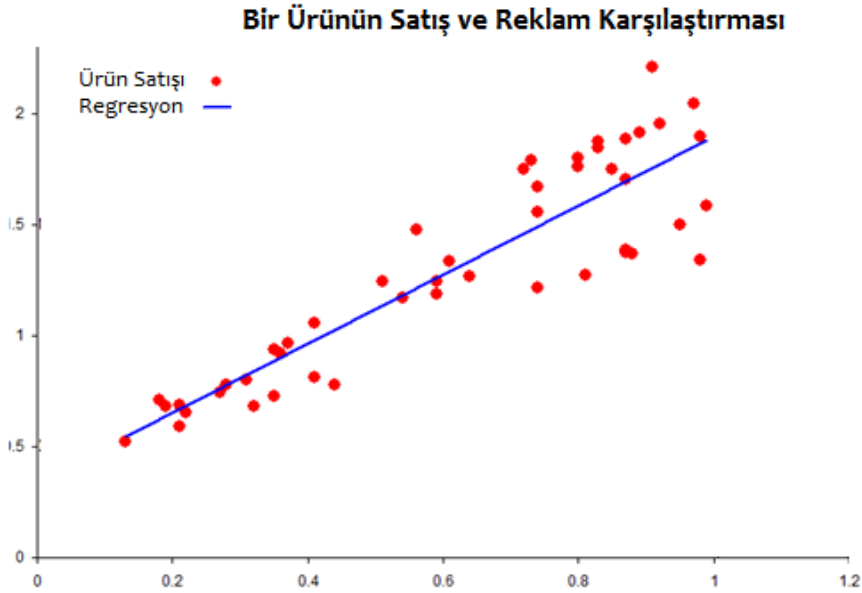
¹⁰ Ethem Alpaydın, “Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri”, Bilişim 2000 Eğitim Semineri, (Çevrimiçi) <http://www.cmpe.boun.edu.tr/~ethem/> , 16.12.2013.

¹¹ Sang C. Park, Selwyn Piramuthu, Michael J. Shaw, “Dynamic Rule Refinement in Knowledgebased Data Mining Systems”, **Decision Support Systems**, No:31, 2001, s. 206.

Veri madenciliği tekniklerinden karar ağaçları, yapay sinir ağları ve genetik algoritmalar sınıflama amaçlı olarak oldukça yaygın şekilde kullanılmaktadır.¹²

3.1.1.2. Regresyon

Regresyon analizi, bağımlı bir değişkenin, bir veya birden fazla bağımsız değişkenle kurduğu ilişkinin bir fonksiyon biçiminde yazılması ve bu fonksiyon yardımıyla bağımlı değişkenin ulaşabileceği değerlerin tahmin edilmeye çalışılmasıdır.¹³



Şekil 7. Doğrusal Regresyon Analizi (Yılmaz Argüden, Burak Erşahin, Veri Madenciliği, ARGE Danışmanlık)

Örneğin, bir banka kredi başvurusunda banka müşterilerinden “kredi kartı borçlarını zamanında ödeyen, işinde 5 yıldan fazladır çalışan, evli bir kişinin kredi skoru 900 iken, aynı şartlarda olan ancak işe yeni başlamış bir kişinin ise kredi skoru

¹² Şule Özmen, **Ağ-Ekonomisinde Yeni Ticaret Yolu: e-Ticaret**, İstanbul Bilgi Üniversitesi Yayınları, İstanbul, 2003, s. 188.

¹³ Neyran Orhunbilge, **Uygulamalı Regresyon ve Korelasyon Analizi**, İstanbul, İ.Ü.İşletme Fakültesi Yayınları, 1996, s.9.

650'dir" sonucu bir regresyon ilişkisidir. Bu tür bir ilişkiyle müşterinin krediyi zamanında ödeyip ödeyemeyeceği tahmin edilmeye çalışılır.

Sınıflama ve Regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Tahmin edilecek alan eğer sayısal (sürekli) bir değişken ise bir regresyon problemidir. Kategorik bir değişken ise sınıflama problemidir.¹⁴ Her iki yöntemde kullanılan teknikler hemen hemen aynı olduğundan, iki yöntem birbirlerine oldukça yakındır.¹⁵

3.1.2. Tanımlayıcı Yöntemler

Karar vermeye yardımcı olacak verilerdeki desenlerin tanımlanmasıdır.¹⁶ Tanımlayıcı modellerde amaç, büyük veri kümelerindeki desen ve ilişkileri belirlemek, verileri ilişkilendirerek anlamlandırmaktır.¹⁷

Perakende sektöründe, hangi ürünlerin birlikte satıldığı, hangi müşteri grubunun hangi zaman aralıklarında bir hizmeti kullandığını ve buna göre en iyi müşteri grubunun hangisi olduğunun belirlenmesinde bu yöntem kullanılır.¹⁸

Örneğin, belli bir aralıkta geliri, iki veya daha fazla arabası olan çocuklu aileler ile geliri aynı aralıktan düşük çocuksuz ailelerin satın alma davranışlarının

¹⁴ Usama Fayyad, Paul Stolorz, "Data Mining and KDD: Promise and Challenges", **Future Generation Computer Systems**, No:13, 1997, s. 104.

¹⁵ B. M. Thuarisingham, **Web Data Mining and Applications in Business Intelligence and Counter Terrorism**, Auerbach Publishers, Incorporated, Boca Raton, 2003, FL, USA, s. 35.

¹⁶ N. Zhong, L. Zhou, Methodologies for Knowledge Discovery and Data Mining : Third Pacific-Asia Conference, Pakdd-99, Beijing, China, April 26-28, 1999 : Proceedings, Springer Verlag, 1999.

¹⁷ Kantardzic, **a.g.e.**, s.1.

¹⁸ M.Helen Moshkovich, Alexander I. Mechitov ve David L. Olson., 'Rule Induction in Data Mining: Effect of Ordinal Scales', **Expert Systems with Applications**, C.XXII, No:4, 2002, s.303.

birbirine benzerlik gösterdiğinin belirlenmesi tanımlayıcı yöntemle gerçekleşir.¹⁹ Bu bilgiye geleneksel istatistiksel verilerle ulaşmak olanaklı değildir.

Bir konu üzerinde araştırma yapan belli bir kütüphane kullanıcı grubundan elde edilen, o konuyla ilgili hangi yayınlardan birlikte yararlandığı bilgisi belirlenebilir.

3.1.2.1. Kümeleme

Veritabanlarında veriler çok büyük boyutlarda olduğu gibi çok karmaşık bir yapıya da sahip olabilir. Bunlar üzerinde uygulanacak veri madenciliği tekniklerinin başarılı olması, bu karmaşık verilerden anlamlı sonuçlar çıkarması zordur. Bu tür sorunlarda izlenilecek yöntem verileri parçalara ayırmak, alt bölümlere ayrıştırmak ve öncelikle alt bölümlerden başlayarak çözüm üretmeye başlamaktır. Bu dağınık verilerin nasıl bölümleneceği ve gruplara ayrılacağını belirlemek için bazı yöntemler geliştirilmiştir.

Bölümleme, demetleme olarak da bilinen kümeleme yöntemi, öngörülen alanda heterojen yapıda bulunan verilerin belirlenmesi ve birbirine benzeyen verilerin bir araya getirilerek homojen yapıda alt kümeler oluşturulmasıdır.²⁰ Kümeleme yönteminde amaç, dağınık verileri değişkenler arası benzerlik veya uzaklıklara dayalı olarak gruplayıp işlenebilir duruma getirmektir.²¹

Kümeleme analizi, veri setinde önceden belirlenmeyen, doğal olarak oluşan alt sınıfları bulur.²² Bu anlamıyla kümeleme bir denetimsiz öğrenme türüdür.²³

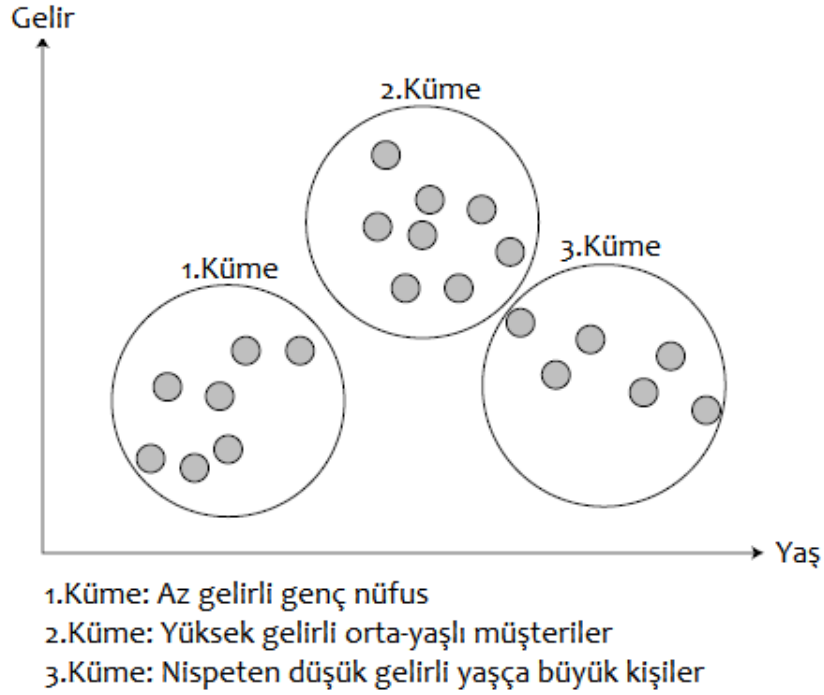
¹⁹ Tuncay Sevindik, Korhan Kayışlı, Orhan Ünlükahraman, “Web Tabanlı Eğitimde Veri Madenciliği”, **Turkish Journal of Computer and Mathematics Education**, C.III, No:3, 2012, s.188.

²⁰ Lise Getoor, “Link Mining: A New Data Mining Challenge”, **ACM SIGKDD Explorations Newsletter**, C.V, No:1, 2003, s. 85.

²¹ Akpınar, **a.g.e.**, s. 6.

²² Lise Getoor, “Link Mining: A New Data Mining Challenge”, **ACM SIGKDD Explorations Newsletter**, C.V, No:1, 2003, s. 85.

Birbirine benzer gibi gözükken sınıflandırma ve kümeleme yöntemleri arasındaki fark, kümelemenin sınıflandırmada olduğu gibi önceden tanımlanmış sınıflara göre değil de, veri içerisinde saklı bulunan ancak önceden tanımlanmamış sınıflara göre ayırım yapmasıdır.²⁴ Bu açıdan bakıldığında sınıflandırma tahmin edici bir yöntemken, kümeleme tanımlayıcı bir yöntemdir.



Şekil 8. Bir Kümeleme Yöntemi Örneği (ZhaoHui Tang, Jamie MacLennan, Data Mining with SQL Server 2005)

Kümeleme analizinde örnek sınıflar bulunmaz. Verilerin kümelmesi işlemi analiz sırasında varolan verilerin birbirine olan benzerliklerine göre yapılmaktadır. Verilerin hangi kümelere ayrılacağı, oluşan kümelerin hangi anlamı taşıdığı veya hangi değişkenin özelliklerine göre sınıflandığı tamamen analizi yapan kişiye

²³ Jiawei Han, v.d., "DBMiner: A System for Data Mining in Relational Databases and Data Warehouses", CASCON '97 Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research, IBM Press, 1997, s. 9.

²⁴ Jiawei Han, Micheline Kamber, **Data Mining : Concepts and Techniques**, Academic Press, 2001, s.335.

bağlıdır. Bu analiz konunun uzmanı olan bir kişi tarafından yapılacağı gibi, bu alanda geliştirilmiş bilgisayar programları aracılığıyla da yapılabilir.

Sınıflandırmada olduğu gibi önceden kaç tane kümenin olduğu bilinmediği için, kümeleme işlemi iki aşamalı olarak gerçekleşir. Olası küme sayısını belirlemek için dıştan bir döngü ve belirlenen sayıdaki kümelerde veriler için en iyi kümelemeye ulaşmaya çalışan içsel bir döngü gerçekleştirilir.²⁵

Kümeleme analizi birkaç adımdan oluşan bir çözüm sürecidir. İlk aşamada veri girişi yapılır. Doğal olarak oluşan sınıflamalarla ilgili kesin bilgilerin bulunmadığı verilerin, incelenen değişkene ilişkin gözlem sonuç değerleri elde edilir. Böylece veri şeması oluşturulmuş olur. Daha sonra uygun kümeleme tekniği seçilir ve uygulanır. Tekniğin uygulanmasıyla veriler kümelere ayrılmış olur. Kümeleme sonuçlarının anlamlılığının yorumlandığı aşama, analizin son aşamasıdır.²⁶

Veri madenciliği araştırmalarında genişçe yer bulan kümeleme analizi, bunun dışında istatistik, biyoloji, psikoloji, tıp, arkeoloji, sosyoloji gibi pek çok alanda kullanım olanağı bulmaktadır.²⁷

3.1.2.2. Birliktelik Kuralı

Veri madenciliğinin en yaygın kullanılan yöntemlerinden biri olan birliktelik kuralı, belirli türlerdeki veri yapıları arasında olası ilişkinin biçimini tanımlama yönelik bir yöntemdir. Birbirleriyle ilişkili olan değişkenlerin aralarındaki ilişkinin

²⁵ Usama Fayyad, Paul Stolorz, "Data Mining and KDD: Promise and Challenges," **Future Generation Computer Systems**, No:13, 1997, s. 104.

²⁶ Bryan F.J. Mainly, **Multivariate Statistical Methods**, 2. bs, Londra: Chapman Hall, 1994, s.280.

²⁷ Metin Vatansever, Ali Hakan Büyüklü, "Using Visual Data Mining Techniques in Clustering Analysis and An Application", **Mühendislik ve Fen Bilimleri Dergisi**, Sigma 27, 2009, s.86.

büyükliğini ve yakınlığını saptamaya çalışır. Daha basit bir tanımla verinin potansiyel ilişkilerini tanımlar.²⁸

Örneğin, bir alışveriş sırasında müşterinin hangi ürün veya hizmetleri satın almaya eğilimli olduğunu saptamak, o müşteriye daha fazla ürün veya hizmetin satılarak şirketin kar oranının artmasında rol oynar.

Birliktelik kuralı önce gelen ve sonra gelen olarak adlandırılan değişkenlerden oluşur. Kural genel olarak önce gelenden sonra gelene yönelik bir ok biçiminde gösterilir.²⁹

Yöntem, “X nesnesini alan bir kişinin, olasılıkla Y nesnesini de alması” türündeki kuralların tanımlanmasını amaçlamaktadır.³⁰

Örneğin, “bir müşteri market alışverişinde süt alıyorsa, aynı alışverişte sütün yanında ekmeğe alma olasılığı nedir?” sorusunun yanıtına ulaşan market yöneticileri, bu bilgi ışığında market raflarını düzenleyerek satışlarını artırabilirler. Eğer bu markette süt ile ekmeğin beraber alınma oranı yüksekse, süt ile ekmeğe rafları yan yana konarak, müşterilerin alışveriş eğilimleri yönlendirilebilir.³¹

Birliktelik kuralı analiziyle, müşterilerin alışveriş davranışları keşfedilmeye çalışılır. Bir ürün satın alındığında eş zamanlı olarak gerçekleşen veya birbirini izleyen alışverişlerde müşterinin hangi mal veya hizmeti satın alma eğiliminde

²⁸ Dunham Margareth, **Data Mining Introductory and Advanced Topics**, Prentice Hall, USA, 2003, s. 8.

²⁹ R. Agrawal, T. Imielinski, ve A. Swami, “Mining Association Rules Between Sets Of Items In Large Databases”, **ACM SIGMOD Conference on Management of Data**, Washington DC, ACM Press. 1993 ss. 207.

³⁰ S. Brin, R. Motwani, ve C.Silverstein, “Beyond Market Baskets: Generalizing Association Rules to Correlations, Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data, New York, USA, 1997, s. 265.

³¹ Judea Pearl, **Probabilistic reasoning in intelligent systems: Networks of plausible inference**, Morgan Kaufman, 1992, s.130.

olduđu belirlenerek, müşteriye daha fazla ürün ve hizmet sağlanılmaya çalışılır.³² Eğer müşteriye sunulan yeni ürün veya hizmet önerisinin kabul edilip edilmediğinin sonucu, müşterinin kişisel bilgileriyle veya önceki tutum ve davranışlarıyla ilişkilendirilebilirse önemli bir bilgiye ulaşılmış olur.

Birliktelik kuralıyla önceden öngürülebilir bir sonuca varılabilir. Örneğın; Tatil dolayısıyla tüm aile bireyelerine uçak bileti alan bir müşteri, %95 olasılıkla tatil beldesinde araba da kiralayacaktır. Daha önce öngörülmesi mümkün olmayan bir sonuca da ulaşılabilir. Örneğın; Hazır yemek alan kadın müşterilerin %60'ı kozmetik ürünlerinden de satın almaktadır.

Veritabanının büyüklüğü ile birliktelik kuralının başarılı sonuç vermesi arasında doğru orantı vardır. Bu yüzden büyük veritabanlarında birliktelik kurallarını bulmak zor değildir. Ancak buradaki zorluk yakın veya uzak ilişkili birçok kuralın ortaya çıkmasıdır. Bulunabilecek birliktelik kuralı sayısı sonsuzdur. Kuralın daha verimli olabilmesi için öncelikle bir eşik değerin (herhangi bir olayın gerçekleşmesi için gereken en küçük şey) bulunması gerekir. Gürültülü, kayıp, eksik veriden değerli, gerekli veriyi ayırabilmek ve bu eşik değerini bulabilmek oldukça zordur. Bu bakımdan ilişkileri tanımlamak için kullanılan bazı ölçütlerin belirlenmesi gerekir. Bu ölçütler, istatistiksel sayısal kavram olan destek ve güven değerleridir. Güçlü bir kural çok büyük destek ve yüksek seviyeli bir güvenliğe sahiptir.^{33 34}

Birliktelik kuralları, veri madenciliğı arařtırmalarında çok büyük yatırımlar yapılan özel bir veri madenciliğı yöntemidir. Ticaret, mühendislik, fen ve sağık sektörlerinin içinde bulunduğı birçok alanda uygulanmaktadır.

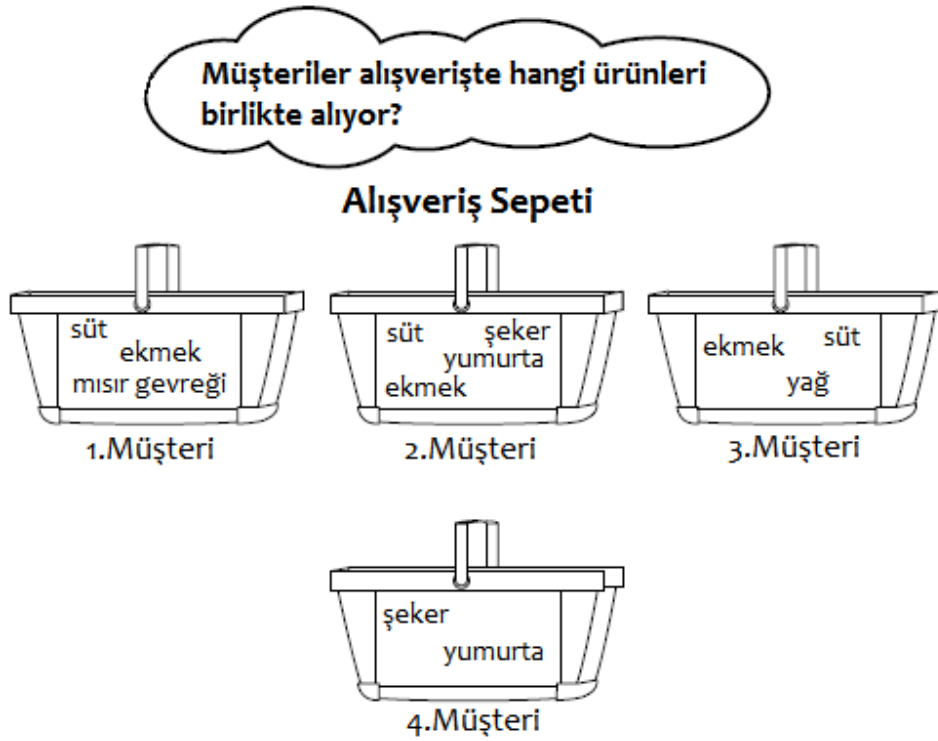
³² Show-Jane Yen, Yue-Shi Lee, "An Efficient Data Mining Approach for Discovering Interesting Knowledge from Customer Transactions", **Expert Systems with Applications**, C.XXX, No:4, 2006, s.650.

³³ P. Adriaans ve D. Zantinge, **Data Mining**, Longman, Harlow: Addison Wesley, 1996, s.63.

³⁴ R. Agrawal, T. Imielinski ve A. Swami, "Mining Association Rules Between Sets Of Items In Large Databases", **ACM SIGMOD Conference on Management of Data**, Washington, DC: ACM Press, 1993, s. 207.

Sepet Analizi:

Birliktelik kurallarının uygulandığı en bilinen örnek market sepet analizidir. Bu analiz, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri veya yakınlıkları bularak, müşterilerin alışveriş eğilimlerini, satın alma alışkanlıklarını, tüketim davranışlarını belirler ve hangi ürünlerin birlikte alındığını ortaya çıkarır. Elde edilen bu bilgi market yöneticileri tarafından daha etkili satış stratejisi, kampanya veya müşteriye özel teklif geliştirmek için kullanılabilir.³⁵



Şekil 9. Market Sepet Analizi

Sepet analizinde amaç değişkenler arasındaki ilişkileri bulmaktır. Eğer X ve Y ürününü alanların genelde Z ürününü de aldığı biliniyorsa ve bir müşteri X ve Y ürününü alıp Z ürününü almıyorsa, bu müşteri potansiyel bir Z ürünü alıcısıdır. Bu müşteriye Z ürününün reklamı yapılabilir.

³⁵ Han, Kamber, a.g.e., s.45.

Analizden elde edilen bilgi mağaza raf düzenlemelerinde de kullanılabilir.³⁶ Eğer X, Y ve Z ürününün birlikte alındığı biliniyorsa, bu ürünler market raf düzeninde yanyana konularak müşterinin satın alma eğilime katkıda bulunulabilir.

Sepet analizi her ne kadar daha çok pazarlama alanında kullanılsa da, bunun dışında birçok alanda kullanılmaktadır.³⁷

Kredi kartlarıyla yapılan alışverişlerde müşterilerin yapacakları potansiyel harcamanın bulunması,

Telekomünikasyon şirketlerinin isteğe bağlı hizmetlerinin (internet, telesekreter, vb.) müşteriler tarafından kullanılmasını sağlamak için, hangi hizmetlerin birlikte kampanyaya girmesi gerektiğinin belirlenmesi,

Sigortacılıkta yapılan işlemin dolandırıcılık olma olasılığının saptanması,

Hastaların sağlık kayıtlarından, yeni hastaların tedavisi için kullanılacak yöntemin belirlenmesi ve önerilen çeşitli tedavi birleşimlerinden doğabilecek yan etkilerin öngörülmesi.

Apriori Algoritması:

Veri madenciliğinin en önemli işlevlerinden biri kural çıkarmaktır. Bu amaçla farklı yöntemler geliştirilmiştir. Bu yöntemlerden en bilinenlerinden biri Apriori algoritmasıdır.

Apriori Algoritması, bilgileri bir önceki adımdan aldığı için “prior”(önceki) kelimesinden türetilmiştir. Algoritma çalışma prensibi olarak eldeki bütün verileri birçok kez tarar. İlk taramadan elde edilen sonuçlar üzerinde ikinci bir tarama yapar. Bu taramalar veriler arasında ilişkiler kalmayınca kadar devam eder. İlişkileri

³⁶ Larissa T. Moss, **Business Intelligence Roadmap: The Complete Project Lifecycle for Decision- Support Applications**, Almanya, Addison Wesley, 2003, s. 320.

³⁷ G. Linoff, ve M.J.A. Berry, **Data Mining techniques For Marketing Sales and Customer Relationship Management**, New York: Wiley Publishing, 2004, s.148.

ortaya çıkarmak için destek ve güven gibi iki ölçütten yararlanılır. Destek ölçütü, bir ilişkinin hangi oranda tekrar ettiğini belirler. Güven ölçütü ise bir ilişkinin doğru sonuçlanma olasılığını ortaya koyar.³⁸

Büyük boyuttaki veritabanlarında birliktelik kuralları bulunurken, öncelikle işlem hareketlerinde sık tekrarlanan öğeler bulunur. Burada önemli olan ele alınacak herbir öğenin en az eşik değeri kadar tekrarlanmış olmasıdır. Eşik değeri, verinin büyüklüğüne ve karar mekanizmasına göre değişebilir. Bir ilişkinin iki kez tekrarlanması da yüz kez tekrarlanması da eşik değeri olarak tanımlanabilir. Daha sonra tekrarlanmış öğeler arasından destek ve güven değerleri göz önünde bulundurularak güçlü bağıntılarla kurallar oluşturulur.

Örnek olarak, bir marketin sebze reyonundan satın alınan ürünleri inceleyerek, basit bir birliktelik kuralı oluşturalım:

Tablo 4. Sebze reyonundan yapılan alışverişler

Alışveriş	Satın Alınan Ürün
1	Brokoli, yeşil biber, mısır
2	Salatalık, kabak, mısır
3	Mısır, domates, fasulye, kabak
4	Yeşil biber, domates, mısır, fasulye
5	Fasulye, salatalık, brokoli
6	Kabak, salatalık, fasulye, domates
7	Domates, mısır
8	Brokoli, domates, yeşil biber
9	Kabak, salatalık, fasulye
10	Fasulye, mısır
11	Yeşil biber, brokoli, fasulye, kabak
12	Salatalık, fasulye, kabak
13	Kabak, mısır, salatalık, fasulye
14	Mısır, yeşil biber, domates, fasulye, brokoli

³⁸ Semra Erpolat, "Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması", **Anadolu Üniversitesi Sosyal Bilimler Dergisi**, C.XII, No:2, 2012, s.137.

Uygulanacak Kuralın Adımları:

1- Birlikte çok satılan ürünleri bul,

2- Bulunan ürünlere destek ve güven ölçütü kullanarak kuralı üret,

Tablo 5. Birlikte alınan ürünler

Birliktelik	Satın Alınma	Birliktelik	Satın Alınma
Salatalık, fasulye	5	Brokoli, mısır	2
Salatalık, brokoli	1	Brokoli, yeşil biber	4
Salatalık, mısır	2	Brokoli, kabak	1
Salatalık, yeşil biber	0	Brokoli, domates	2
Salatalık, kabak	5	Mısır, yeşil biber	3
Salatalık, domates	1	Mısır, kabak	3
Fasulye, brokoli	3	Mısır, domates	4
Fasulye, mısır	5	Yeşil biber, kabak	1
Fasulye, yeşil biber	3	Yeşil biber, domates	3
Fasulye, kabak	6	Kabak, domates	2
Fasulye, domates	4		

Destek ölçütü= İki ürünü birlikte içeren alışveriş sayısı / Toplam alışveriş sayısı

Güven ölçütü= İki ürünü birlikte içeren alışveriş sayısı / Bir ürünü içeren alışveriş sayısı

Tablo 6. Birlikte satın alınan ürünlerden üretilen kurallar

Kurallar	Destek	Güven
Salatalık ve fasulye alanlar, kabak da alırlar	4/14=28.6%	4/5=80%
Salatalık ve kabak alanlar, fasulye de alırlar	4/14=28.6%	4/5=80%
Fasulye ve kabak alanlar, salatalık da alırlar	4/14=28.6%	4/6=66.7%
Brokoli alanlar, yeşil biber de alırlar	4/14=28.6%	4/5=80%

Uyguladığı veri madenciliği yöntemiyle Tablo 4'teki sonuçlara ulaşan bu market sahibi, sebze satışlarını arttırmak için çeşitli stratejiler geliştirebilir. Örneğin, salatalık, fasulye ve kabakları reyonda aynı rafa dizerek müşterisine farkındalık oluşturabilir. Bu üç üründen birini almayı unutan müşteri diğer ürünleri almak için rafa yaklaştığında unuttuğu ürünü görerek hatırlayacaktır.

Sık tekrarlanan öğeleri bulmak için en çok kullanılan yöntem Apriori Algoritması'dır. Apriori Algoritması, eşik değerini tanımlayarak oluşabilecek birliktelik kümesi sayısını azaltmak için kullanılan bir çözüm yoludur.³⁹

Ardışık Zamanlı Örüntüler:

Ardışık zamanlı örüntüler, art arda gelen dönemler içerisinde birbirleriyle ilişkili olan değişkenlerin aralarındaki ilişkiyi ortaya çıkarmak için kullanılan yöntemdir. Örneğin;⁴⁰

- X ameliyatı yapıldıktan 15 gün sonra %45 olasılıkla Y enfeksiyonu oluşacaktır.
- Çamaşır makinesi alan bir müşteri 6 ay içerisinde %70 olasılıkla kurutma makinesi de alacaktır. Satış yöneticisi bu olasılığı artırmak için 4 ay sonra müşteriye kurutma makinesi için indirim uygulayabilir.
- İlk üç taksidinden iki veya daha fazlasını geç ödemiş olan müşteriler %60 olasılıkla krediyi geriye ödeyemiyor. (Behavioral scoring, Churning)

Birliktelik kuralıyla ardışık zamanlı örüntüleri birbirinden ayıran temel unsur zaman kavramıdır. Belli bir dönem içerisinde nesnelere arasındaki birlikteliklerin analizi ardışık zamanlı örüntü olarak adlandırılır.⁴¹

³⁹ Mohammed J. Zaki, "Parallel and Distributed Association Mining: A Survey", **IEEE Concurrency**, C.VII, No:5, 1999, s. 14.

⁴⁰ Haldun Akpınar, "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", **İ.Ü.İşletme Fakültesi Dergisi**, No:1, 2000, İstanbul, s. 7.

⁴¹ M. Goebel, L. Gruenwald, "A Survey of Data Mining and Knowledge Discovery Software Tools", **ACM SIGKDD Explorations Newsletter**, C.I, No:1, 1999, s.20.

3.2. Veri Madenciliği Teknikleri

Veri madenciliği yöntemlerinin teknik içeriğini oluşturan veri madenciliği teknikleri, istatistiksel ve matematiksel formüller, algoritmalar, örüntü tanıma teknolojileri ve makine öğrenimini birlikte kullanan süreçlerden oluşur.

İşletme veya kurumların problemlerini eldeki verilerle çözmek için uygun model ve algoritmanın seçilmesi gerekir. Seçilecek teknik, probleme göre farklı sonuçlar üretir.⁴²

3.2.1. Karar Ağaçları

Karar alma, kurumların işleyişleri sırasında sıklıkla karşılaştıkları bir durumdur. Dışarıdan bakıldığında sıradanmış gibi gözüken birçok işlem aslında bir dizi karar alma işlemi içermektedir.⁴³ Karar verici, kaç tane seçeneğin gerçekleşeceğinin belirsiz olduğu bir problemle ilgili en doğru kararı verebilmek için bir takım yöntemlere gereksinim duyar. Karar Ağacı yöntemi, kolay anlaşılabilir kurallar üretmesi ve kuralların görselleştirilebilmesi sebebiyle en sık kullanılan VM tekniklerindedir. Etkili bir sınıflandırma yöntemidir.⁴⁴ Karar ağacı analizi, özellikle seçenek sayısının fazla olduğu, birden fazla kararın ardışık olarak verilmesi gibi daha karmaşık bir problemin çözüm analizinde kullanılan bir tekniktir.⁴⁵

Sınıflandırma ve kümeleme yöntemlerinde oldukça sık kullanılan, tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları;

⁴² Ayşe Çınar, Gökhan Silahtaroglu, “Veri Madenciliği Teknikleri İle Müşteri Memnuniyetine Etki Eden Gizli Nedenlerin Keşfi”, **Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi**, C.XIII, No:2, 2012, s.311.

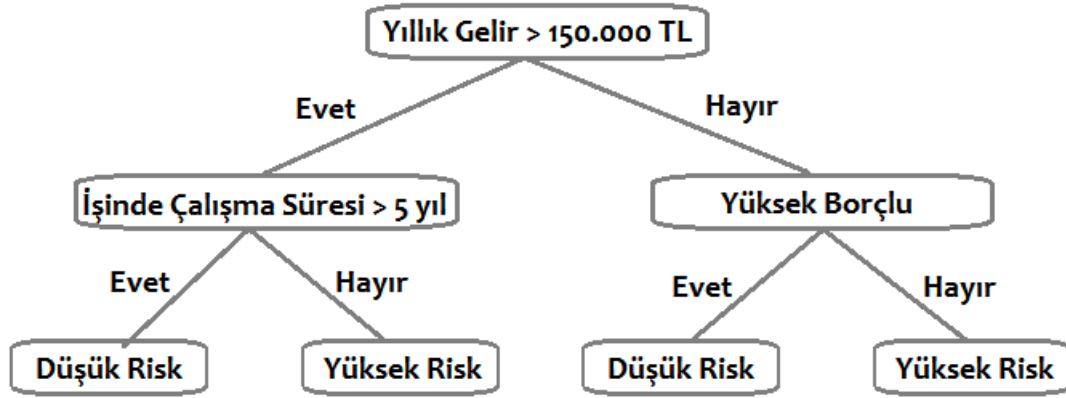
⁴³ Alain Decrop, **Vacation Decision Making**, Londra, CABI Publishing, 2006, s.1.

⁴⁴ Dilek Altaş, Vildan Gülpınar, “Karar Ağaçları Ve Yapay Sinir Ağlarının Sınıflandırma Performanslarının Karşılaştırılması: Avrupa Birliği Örneği”, **Trakya Üniversitesi Sosyal Bilimler Dergisi**, C.XIV, No:1, 2012, s.4.

⁴⁵ Hayrettin Kemal Sezen, **Yöneylem Araştırması**, Bursa, Ekin Kitapevi, 2004, s.5.

- Oluşturulmasının ucuz olması,
- Yorumlanmasının kolay olması,
- Veritabanı sistemleriyle kolayca bütünleştirilmesi,
- Güvenilirliğinin yüksek olması nedenlerinden dolayı veri madenciliğinde yaygın olarak kullanılmaktadır.⁴⁶

Karar ağaçları verileri hiyerarşik olarak alt sınıflara ayırmaktadır.⁴⁷ Basit bir altyapıya sahiptir. Diğer tekniklerle karşılaştırıldığında daha kolay anlaşılabilir ve yorumlanabilir sonuçlara ulaşır. Bu yüzden daha çok tercih edilir.⁴⁸



Şekil 10. Örnek Bir Karar Ağacı Yapısı (Serkan Terzi, “Hile ve Usulsüzlüklerin Tespitinde Veri Madenciliğinin Kullanımı”, Muhasebe ve Finansman Dergisi)

KA'nın sağladığı avantajlar yanında dezavantajları da vardır. Genellikle karar vermek bu yaklaşım için ciddi bir sorun oluşturur. Sorun şundan kaynaklanır: Ağaç, ayrılan bölümleriyle genişledikçe, yapılan sınıflandırma veriyi çok fazla sayıda ayrıma böler. Bu ayrımların içeriği küçülmeye başlar. İncelenmesi gereken farklı

⁴⁶ Akpınar, **a.g.e.**, s. 12.

⁴⁷ Ercan Öztemel, **Yapay Sinir Ağları**, İstanbul, Papatya Yayıncılık, 2003, s.36

⁴⁸ Agrawal Rakesh, Tomasz Imielinski, Arun Swami, “Database Mining: A Performance Perspective”, **IEEE Transactions on Knowledge and Data Engineering**, C.V, No:6, 1993, s. 915.

durumların sayısı attıkça, kümelerinin her biri daha da küçülür. Bu durum karar ağacını yönetmeyi zorlaştırır ve uygulamada sorunlara yol açabilir.⁴⁹

Karar ağacı yapısı oluşturulduktan sonra, budama adı verilen, problemin çözümüne bir etkisi olmadığı düşünülen ve sınıflamada etkisi olmayan dalların ağaç yapısından alınma işlemi yapılabilir.⁵⁰

3.2.2. Yapay Sinir Ağları

Yapay sinir ağları, insan beyнинin özelliklerinden olan öğrenme yolu ile yeni bilgiler üretebilme veya keşfedebilme gibi yeteneklerden esinlenerek geliştirilmiş, bu yetenekleri otomatik olarak gerçekleştirmeye çalışan bilgisayar yazılımlarıdır.⁵¹ İnsanlığın doğayı araştırma ve öykünme çabasının en son ürünlerinden biri olan teknolojidir. İnsanlar tarafından gerçekleştirilen örnekleri (gerçek beyin fonksiyonlarının ürünü olan örnekler) kullanarak olayları öğrenebilen, çevreden gelen olaylara karşı nasıl tepkiler üreteceğini belirleyebilen bilgisayar sistemleridir.⁵² Son yıllarda yaygınlaşan yapay sinir ağları adı verilen yazılımlar, biyolojik sinir sisteminin çalışma biçimini canlandırmak için tasarlanmışlardır. Canlı organizmaların karşılıklı bağlantılı sinir hücrelerinin ürettiği karmaşık algoritmalarından esinlenerek geliştirilmiştir.⁵³ Biyolojik yetenekleri geleneksel programlama yöntemleriyle gerçekleştirmek oldukça zor olduğundan, yapay sinir ağları, programlanması zor olan veya mümkün olmayan sistemler için geliştirilmiş bağdaşık bilgi işlemeyi konu alan bilgisayar bilim dalıdır.⁵⁴

⁴⁹ Altaş, Gülpınar, **a.g.e.**, 2012, s.5.

⁵⁰ Peter Cabena, v.d., **Discovering Data Mining : From concept to Implementation**, USA, International Business Machines Corporation, 1998, s. 73

⁵¹ Öztemel, **a.g.e.**, s.29.

⁵² Altaş, Gülpınar, **a.g.e.**, 2012, s.7.

⁵³ Haykin S., **Neural Networks: A Comprehensive Foundation**, USA, Prentice Hall, 1994, s.138.

⁵⁴ Öztemel, **a.g.e.**, s.29.

Veri madenciliğinde en çok bilinen ancak en az anlaşılan bir teknik olan YSA, insan beyninin çalışma ilkelerine benzer özelliklere sahiptir. Kendilerine girilen verilerden elde ettikleri bilgiler ile kendi deneyimlerini oluştururlar ve benzer konulara kurallar üretirler. Veri yorumlama, sınıflandırma, ilişkilendirme, genelleme gibi çözümlerde kullanılabilir. Günümüzde yapay sinir ağlarının kullanım alanlarından bazıları şunlardır:⁵⁵

- Denetim
- Sistem modelleme
- Ses tanıma
- El yazısı, parmak izi tanıma
- Meteorolojik tahminler
- Otomatik araç denetimi.

Yapay sinir ağları, veri madenciliğinde hem tanımlayıcı hem tahmin edici yöntemlerde birçok amaç için kullanılmaktadır.⁵⁶ Bankacılıkta kredi risklerinin değerlendirilmesinde, kredi kartı hilelerinin saptanmasında, pazarlama sektöründe ürünün performansını öngörme gibi işlemlerde kullanılabilir.⁵⁷

3.2.3. Genetik Algoritmalar

Canlı varlıkların biyolojik yapısından esinlenerek geliştirilen bilgisayar teknolojilerinin gelişmesi yakın zamanda daha güvenilir ve etkili eniyileme (optimizasyon) yöntemlerinin ortaya çıkmasını sağladı. Bu yöntemlerden biri olan

⁵⁵ Çetin Elmas, **Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama)**, Seçkin Yayıncılık, Ankara, 2003, s. 25.

⁵⁶ Paolo Giudici, **Applied Data Mining: Statistical Methods for Business and Industry**, John Wiley & Sons, 2003, s.107.

⁵⁷ Öztemel, **a.g.e.**, s.36.

genetik algoritmalar, teknik, mühendislik, tıp ve siyaset bilimi gibi birçok alanda diğer arama tekniklerine alternatif olarak genel bir çözüm sunma yöntemi olarak tanındı. Genetik algoritmalar basit tanımıyla etkili sonuçlara ulaşmaya çalışan rastgele arama teknikleridir.⁵⁸ Karmaşık düzenli problemlerin çözümünü gerçekleştirmek amacıyla geliştirilmiş, kromozomların üretme esasına dayanan sezgisel bir araştırma yöntemidir. Biyolojik evrimi temel alır. İşlem sonucunda en iyi kromozoma ulaşmaya çalışır. En iyi kromozoma ulaştığında ise genlerin yerini değiştirerek yeni sonuçlar üretir.⁵⁹

Genetik algoritmalar, geleneksel yöntemlerle çözümü zor veya imkansız olan problemlerin çözümünde kullanılmaktadır.⁶⁰ Farklı sonuçları eşzamanlı değerlendirerek arama eğilimini daha iyi çözüm alanlarına yönlendirir. Genellikle bir çözüme odaklanılan problemin en büyük ve etkili sonucuna ulaşmaya çalışır.⁶¹

Genetik algoritmaların özellikleri şunlardır:⁶²

- Uygun çözümler için birden çok popülasyon olabilir,
- Önceden bilinen çözümlerin özelliklerini biraraya getirerek veya bu özellikleri rastgele değiştirerek yeni uygun çözümler üretebilir,
- Popülasyon içinden öncelik vererek, değişik çözümleri seçebilir ve bazı çözümleri gözardı edebilir.

⁵⁸ Hamit Saruhan, "Genetic Algorithms: An Optimization Technique", **Teknoloji**, C.VII, No:1, 2004, s.105.

⁵⁹ Necdet Özçakar, "Genetik Algoritmalar", **İ.Ü. İşletme Fakültesi Dergisi**, C.XXVII, No:1, 1998, s.69.

⁶⁰ Mustafa Kurt, Cumali Semetay, **Mühendis ve Makine Dergisi**, TMMOB Makina Mühendisleri Odası, Ankara, No: 501, 2011, s.2.

⁶¹ YongSeog Kim, W. Nick Street, "An Intelligen System for Customer Targeting: A Data Mining Approach," **Decision Support Systems**, No:37, 2004, s. 216.

⁶² David M. Tatei Alice E. Smith, "A Genetic Approach to the Quadratic Assignment Problem", **Computers&Operations Research**, C.XXII, No:1, 1994, s.74.

Genetik algoritmalar, veri madenciliği uygulamalarında başarılı sonuçlar vermektedir.

3.2.4. Veri Görselleştirme

Veri madenciliği uygulamalarında verilerin birbirleriyle olan ilişkilerin anlaşılması büyük önem taşır. Geleneksel yöntemlerle bulunan karmaşık ilişkileri kolay algılanabilir, rahatça anlaşılabilir biçimde sunmaya çalışmak gerekir. Veri görselleştirme, verinin algılanabilirliğini arttırmak için bilgisayar desteği ile görsel olarak temsil edilmesidir. Veriden yeni yorumlanabilir örüntüler elde etmek amacıyla, veriyi alışlagelmiş sayısal ve istatistiksel biçimden soyutlayarak grafik arayüzler ile tasarlar.⁶³

Bu teknikle verilerin grafik yardımıyla gösterimi yapılmakta ve ikiden fazla boyutlu grafikler kullanılmakta olup veriler içerisinde gizlenmiş olan anormallikler tespit edilebilmektedir.⁶⁴

Veri görselleştirme insanın algı ve yorumlama yeteneğini dikkate alarak analiz yapmaya olanak verir. İnsanı algılama sistemi yalnızca 3 boyut ile sınırlı olduğundan daha fazla boyut içeren veriler insan algısının dışında kalır. Veri görselleştirme teknikleri çok boyutlu veriyi 2 veya 3 boyuta indirir. Görselleştirmenin temel amacı kuralların, kavramların daha iyi anlaşılmasıdır. Grafikler ve resimlerle yeni düşünceler oluşturmak, yeni ilişkiler kurmak ve yeni yapılar keşfetmek veya bu yapıları düzenlemektir.⁶⁵

Karar verme sürecinde görsellik gereksinimi fazla olan karar verici açısından kullanışlı bir yöntemdir. Bazı durumlarda verinin en iyi anlaşılabilceği yöntemdir.

⁶³ Metin Vatansever, Ali Hakan Büyüklü, "Using Visual Data Mining Techniques in Clustering Analysis and An Application", **Mühendislik ve Fen Bilimleri Dergisi**, Sigma 27, 2009, s.84.

⁶⁴ Serkan Terzi, "Hile ve Usulsüzlüklerin Tespitinde Veri Madenciliğinin Kullanımı", **Muhasebe ve Finansman Dergisi**, No:54, 2012, s.51.

⁶⁵ Tugay Bilgin, A. Yılmaz Çamurcu, Çok Boyutlu Veri Görselleştirme Teknikleri, **Akademik Bilişim Semineri**, Çanakkale, 30 Ocak-01 Şubat, 2008, s.108.

Grafik ve haritalar gibi görselleştirme araçlarının yardımıyla, verideki eğilim, değişkenlik veya benzerlikleri tanımlamak, verileri gruplandırarak kümelere ayırmak olanaklıdır. Örneğin, yüksek suç oranına sahip bir bölgeyi gösteren veriler haritayla görselleştirilerek sunulabilir. Coğrafi bölgelere göre tutum ve davranışları değişen müşteriler, yoğunluk derecesi, satış rakamları, belirli ekonomik göstergeler açısından değerlendirilerek görselleştirme tekniğiyle çok daha etkin ve kolay biçimde analiz edilebilir.⁶⁶

Kutu, çizgi, histogram gibi bilinen çeşitli görselleştirme yöntemlerinin bu alanda yapılan çalışmalar incelendiğinde 5 temel sınıfa ayrılacağı görülmektedir:⁶⁷

- Standart 2 ve 3 boyutlu gösterimler (Kutu grafikleri, Histogram)
- Geometrik olarak dönüştürülmüş gösterimler (Andrews eğrileri, PolyViz grafikleri)
- Simgesel gösterimler (Chernoff yüzleri, Star)
- Yoğun piksel gösterimler (Matris grafikleri)
- İstiflenmiş gösterimler (Dünyalar içinde dünyalar -worlds within worlds-, Treemap)

Özellikle kümeleme yönteminde kullanılan veri görselleştirme, bununla beraber genelde veri madenciliği sürecinin son aşamasında kullanılmaktadır.

⁶⁶ Şule Özmen, **Ağ-Ekonomisinde Yeni Ticaret Yolu: e-Ticaret**, İstanbul Bilgi Üniversitesi Yayınları, İstanbul, 2003, s.189.

⁶⁷ Metin Vatansever, Ali Hakan Büyüklü, "Using Visual Data Mining Techniques in Clustering Analysis and An Application", **Mühendislik ve Fen Bilimleri Dergisi**, Sigma 27, 2009, s.85.

SONUÇ VE DEĞERLENDİRME

Kurum ve işletmeler etkinlik gösterdikleri her alanda, organizasyonel anlamda pek çok problemle karşılaşır. Bu tür durumların üstesinden gelebilmek için ya bilinen çözüm yollarından birine yönelir veya yeni çözümler üretmeye çalışarak, bunlara göre davranış gösterir, yani karar verirler. Bir problemin tek bir çözüm yolu olduğunda orada karar verme sürecinden söz edilemez. Karar, çoğunlukla bir seçme durumu olduğunu gösterir. Bir seçim yapıldığında ise aslında devam edilecek yol belirlenmiş olur. Bu yönden bakılınca karar verme sadece o anı etkileyen bir durum olmaktan çıkar, işleyişin geri kalanına da yön verir.

Bilgi toplumlarında bilginin giderek önem kazandığı göz önüne alındığında, bilgi teknolojilerindeki gelişmeleri takip etmek rekabet dünyasında başarılı olmak için kaçınılmazdır. Her gün yeni bir teknolojik ürünün ortaya çıktığı varsayıldığında karşılaşılan problemi çözmek için seçilen yöntem ve tekniğin doğru olması gerekir. Bu durumda karar verme süreçleri de oldukça önem kazanmaktadır.

Karar verme eyleminin başarıyla sonuçlanması ancak doğru kararın verilmesine bağlıdır. Kurum ve işletmeler için doğru karar, onu belirlediği kurumsal amaçlara ulaştıran, o anki gereksinimlerini karşılayan veya onlara yeni olanaklar oluşturan seçimlerdir. Doğru kararı vermek zorlu bir süreçtir. Bu süreçte karar vericiye yardımcı olması için temelinde bilgi olan pek çok teknik geliştirilmiştir. Karar verme sürecine etki eden ve bu süreci kolaylaştıran bilgi teknolojilerinin günümüzde geldiği son nokta “iş zekası” olarak adlandırılan bilimsel teknik ve teknolojilerdir.

Bilgi keşfi sürecinin teknik altyapısını oluşturan iş zekası teknolojileri ile kullanıcı ve müşterilerin karakteristik özellikleri belirlenerek veya bir ürünü ve hizmeti kullanma veya satın alma davranışları incelenerek, onları gereksinimlerine göre gruplayıp o hedef kitlenin bütününe görülmesi sağlanır. Bu sayede kurum veya işletmenin durumunu ve performansını anlamak, geleceğe yönelik kurumsal

öngörülerde bulunmak, maliyetleri azaltırken kazancı artırmak ve en önemlisi de karar vermeyi etkin hale getirmek kolaylaşır.

İş zekası sistemlerinin önemli bir ögesi de veri madenciliğidir. Verilerin ve veri işleme araçlarının gelişmesi sonucunda sayısal ortamda toplanan verilerin analizi ve değerlendirilmesi gerekliliği fark edilmeye başlamış ve bu verilerden özellikle gelecek tahmini yapabilmek için farklı bilim dalları bir araya gelerek çalışmalar yapmışlardır. Veri madenciliği kavramı bu noktada ortaya çıkmış ve birçok alanda uygulanabilir olması nedeniyle hızla yaygınlaşmaya başlamıştır.

Bu çalışmada veri madenciliği; veri, enformasyon, bilgi kavramlarını içeren bilgi keşfi süreci çerçevesinde ele alınarak incelenmiş ve bu kavramlar sistematik olarak değerlendirilmiştir.

Çalışmadan çıkarılan sonuca göre veri madenciliği için veri olmazsa olmazdır. Verinin bilgiye dönüşmesi için öncelikle kaliteli, güvenilir veriye gereksinim vardır. Bu yüzden veri önışleme olarak adlandırabileceğimiz veri hazırlama aşaması, veri madenciliği için gerekli veriyi hazırlar. Veri hazırlama; veri temizleme, veri birleştirme, veri dönüştürme ve veri indirgemeyi içerir. Eksik değerleri tamamlamak, gürültülü veriyi düzeltmek, veri tutarsızlıklarını yok etmek, tutarlı bir veri deposu oluşturmak üzere çeşitli kaynaklardan gelen veriyi birleştirmek, veriyi madencilik için uygun biçimlere dönüştürmek, verinin öznitelikleri koruyarak hacimce daha küçük boyuta indirgemek veri hazırlama aşaması olarak veri madenciliğinden önce uygulanmalıdır.

Veri madenciliği ile farklı ortamlardan gelen kurumsal veriler düzenlenir, çözümlenir, işlenir ve birbiriyle ilgili veriler bir araya getirilerek raporlanır. Bu raporlar karar vermek ve eylem planını gerçekleştirmek için kullanılır. Veriye dayalı kararların kalitesi, güvenilirliği, doğruluğu artar. Veriye dayalı kararların verildiği kurumlarda kaynakların kullanım etkinliği artar ve üretim potansiyeli gelişir. Bu anlamıyla veri madenciliği tek başında bir çözüm değil, çözüme ulaşmak için verilmesi gereken kararları destekleyici bilgileri sağlayan araçtır.

Veri madenciliğinde üç önemli unsur rol almaktadır. Veri analisti, veri uzmanı olarak insan gücü, teknolojik altyapıyı oluşturan veritabanı, veri ambarı sistemleri ve bu sistemlerin üzerinde kurulacak veri madenciliği yazılımları. Bir veri madenciliği uygulamasının başarılı olması için alanındaki konuyu bilen analistin, veritabanı sistemlerinde uzman veritabanı yöneticisinin ve veri madencisi uzmanının beraber çalışması gerekir. Teknolojik altyapıyı oluşturacak veritabanı ve veri ambarı yönetim sistemlerinin veritabanı motoru, veri depolama, veri sorgulama, veri tanımlama, veri işleme alt sistemlerine sahip olması gerekir. Bunları sağlayan birçok veritabanı yönetim sistemi yazılımları bulunmaktadır. Veri madenciliği uygulamaları için gerekli yazılımlarının geliştirilmesi ise son zamanlarda artmış, satın alma maliyetleri düşmüş ve hatta açık kaynak kodlu yazılımlar ortaya çıkmıştır. Bu yazılımlar içerisinde veri kümeleme, karar ağaçları, apriori yöntemi gibi birçok veri madenciliği yöntem ve teknikleri kullanılabilir.

Veri madenciliği uygulama süreci öncelikle problemin tanımlanmasıyla başlamalıdır. Uygulandığı kurumun hedef ve amaçları olarak tanımlanan problemin kesin, açık, ölçülebilir olması gerekir. Sürecin en önemli aşamasıdır. Problem belirlendikten sonra uygun model bulunmalıdır. Doğru model seçimi analizden elde edilecek sonuçların kalitesini belirler. Daha sonra model kurulmalı ve öncelikle test verisi üzerinde denenmelidir. Geçerliliği onaylanan model izlenir ve problemin ne kadarına çözüm getirdiği değerlendirilerek süreç tamamlanabilir.

Veri madenciliği araçları genel olarak veriyi analiz eden ve sonuçlar çıkaran bilgi teknolojileridir. Bu analizlerde farklı yöntemler kullanılmaktadır. Bu yöntemlere ait pek çok teknik vardır. Bu tekniklerden hangisinin daha yararlı ve kullanışlı olduğu analizin uygulanacağı kurum tarafından belirlenmelidir. Herbir tekniğin diğerine göre avantajları dezavantajları vardır. Örneğin tanımlayıcı yöntem verilerdeki gizli ortak özellikleri, ilişkileri araştırır ve tahmin edilebilir sonuçlar çıkarabilir. Ayrıca geleceğe yönelik öngörülerde bulunmaz. Ancak veri madenciliği araçlarının asıl yeteneği veri yığınları içerisinde gizli kalmış, önceden tahmin edilemeyen bilgileri açığa çıkarmasıdır. Tahmin edici yöntem olarak söz ettiğimiz bu

yöntem, karar verme süreci içerisindeki problemlere çözüm ararken, sorulması düşünülmeyen sorulara bile yanıt verilebilir. Veri madenciliğini diğer istatistiksel analizlerden ayıran en belirleyici özellik budur.

Çalışmada savunulan düşünce veri madenciliğinin Bilgi ve Belge Yönetimi Bilimi içerisinde yer alması gerektiğidir. Verinin bilgiye dönüştüğü süreç Bilgi Yönetimi disiplini içerisinde teorik olarak genişçe yer almaktadır. Sürecin içerisinde yer alan veri madenciliğini bu yapının dışında tutmak doğru olmayacaktır. Bilindiği üzere teknolojik gelişmeler doğrultusunda Bilgi ve Belge Yönetimi disiplini yeniden yapılandırıldı. Bilgi ve bilgisayar teknolojilerinin bütün disiplinler üzerinde etkisi artmaya başlayınca, bilginin kaynağını yöneten ve düzenleyen bilgi yönetiminin bakış açısı da değişti. Artık bilgi sadece basılı ortamda erişilebilen bir olgu olmaktan çıktı ve sayısal ortamda depolanmaya başladı. Belge denildiğinde artık sadece bir kağıt üzerine kayıtlı metinler düşünülüyor. Veritabanı sistemleri ve bunlar üzerinde işlem yapmaya yarayan veritabanı yazılımları gelişti. Bilgi artık yazılarak kayıt altına alınan bir harf, rakam, simge olmaktan çıktı, veri tabanlarındaki tablolarda saklanan ve hesaplanan bir forma dönüştü.

Bilgi ve belgedeki fiziksel değişiklik bilgi yöneticilerinin ve karar vericilerinin bakış açısını da değiştirdi. Daha önce bilgi arşivlerde, kütüphanelerde saklanan ve sadece gereksinim duyulduğunda erişilebilen kayıtlardı. Bununla yetinilmedi ve verilerin işlenmesi, içlerinde sakladıkları gizli örüntülerin bulunması ve kullanılması yani verinin değerlendirilmesi düşüncesi ortaya atıldı. Bununla beraber bilgi keşfi diye adlandırılan bir süreç başladı. Veri madenciliği bu sürecin en önemli ögesi olarak ön plana çıktı.

Bilgi ve Belge Yönetimi disiplini de bilgiye erişimde ve bilgi yönetimiyle ilgili bireysel veya kurumsal karşılaşılan sorunları çözmek için uğraşmaktadır. Bu disiplinin amacı genel olarak bilgi gereksinimlerine karşılık verilmesi, bilginin derlenmesi, saklanması, korunması, paylaşılması ve erişilebilir olmasıdır. Bilgi keşfi sürecinin ortaya çıkmasıyla bu amaca ek olarak verinin saklanacağı ve işleneceği

bilgi sistemlerinin tasarlanması, kurulması ve yönetilmesi, verilerin işlenmesi, birbiriyle ilişkili olanların gruplanarak beraber depolanması ve bütün bunlara olanak sağlayacak altyapının kurulması gibi amaçlar yer almalıdır. Böylece Bilgi ve Belge Yönetimi, sadece bilgiye erişilmesini sağlayan bir disiplin olmaktan çıkıp yeni bilgi üreten bir disipline dönüşebilir.

KAYNAKÇA

Adriaans, P., D. Zantinge:	Data Mining , ABD, Addison Wesley Longman, 1996.
Agrawal, R., T. Imielinski, A. Swami:	“Mining Association Rules Between Sets Of Items In Large Databases”, ACM SIGMOD Conference on Management of Data , Washington, DC. ACM Press, 1993, s. 207-216.
Akgün, Ali, Halit Keskin:	“Sosyal Bir Etkileşim Süreci Olarak Bilgi Yönetimi ve Bilgi Yönetimi Süreci”, Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi , C.V, No:3, 2003, s.175-188.
Akpınar, Haldun:	“Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği”, İ.Ü. İşletme Fakültesi Dergisi , C.XXIX, No:1, 2000, s.1-22.
Alan, Mehmet Ali:	“Veri Madenciliği Ve Lisansüstü Öğrenci Verileri Üzerine Bir Uygulama”, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi , No:33, 2012, s.165-174.
Alavi, Maryam:	“Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issue”, MIS Quarterly , C.XXV, No:1, 2001, s.109-111.
Alpaydın, Ethem:	“Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri”, Bilişim 2000 Eğitim Semineri , (Çevrimiçi) http://www.cmpe.boun.edu.tr/~ethem/ , 16.12.2013.

Altaş, Dilek, Vildan Gülpınar:	“Karar Ağaçları Ve Yapay Sinir Ağlarının Sınıflandırma Performanslarının Karşılaştırılması: Avrupa Birliği Örneği”, Trakya Üniversitesi Sosyal Bilimler Dergisi , C.XIV, No:1, 2012, s.1-21.
Anane, Rachid:	“Data Mining and Serial Documents”, Computers and the Humanities , C.XXXV, No:3, 2001, s.300-301.
Apte, Chidanand v.d.:	“Business Applications of Data Mining”, Communications of the ACM , C.XLV, No:8, 2002, s.49-53. (Chidanand Apte, Bing Liu, Edwin P. D. Pednault, Padhraic Smyth)
Argüden, Yılmaz, Burak Erşahin:	Veri Madenciliği , İstanbul, ARGE Danışmanlık, 2008.
Arslantekin, Sacit:	“Veri Madenciliği ve Bilgi Merkezleri”, Türk Kütüphaneciliği , C.XVII, No:4, 2003, s.369-380.
Barutçugil, İsmet:	Bilgi Yönetimi , İstanbul, Kariyer Yayıncılık, 2002.
BAYKAL, Abdullah :	“Veri Madenciliği Uygulama Alanları”, Dicle Üniv. Ziya Gökalp Eğitim Fakültesi Dergisi , No:7, 2006, s. 95-107.
Becerra-Fernandez, Irma, Stelios H. Zanakis, Steven Walczak:	“Knowledge Discovery Techniques for Predicting Country Investment Risk”, Computers and Industrial Engineering , No:43, No:4, 2002, s.787-800.

Becker, Karin, Cinara Ghedini:	“A Documentation Infrastructure for the Management of Data Mining Projects”, Information and Software Technology , C.XLVII, No:2, 2005, s.95-111.
Bergeron, Bryan :	Bioinformatics Computing , U.S.A., Prentice Hall PTR, 2002.
Bergeron, Bryan :	Essentials of CRM: A Guide to Customer Relations Management , USA, John Wiley&Sons, 2002.
Berry, Michael J.A, Gordon Linoff,:	Data Mining Solutions , U.S.A., Wiley Computer Publishing, 1998.
Bilgin, Tugay, A. Yılmaz Çamurcu:	“Çok Boyutlu Veri Görselleştirme Teknikleri”, Akademik Bilişim Semineri , Çanakkale, 30 Ocak-01 Şubat, 2008, s.107-112.
Bounsaythip, Catherine, Esa Rintarunsala:	“Overview of Data Mining for Customer Behaviour Modeling”, VTT Information Technology Research Report TTEI , 2001, s.1-49.
Bozdogan, Hamparsum:	Statistical Data Mining and Knowledge Discovery , Boca Raton, London, New York, Chapman&Hall/CRC, 2004.

Brin, S., R. Motwani, C.Silverstein:	“Beyond Market Baskets: Generalizing Association Rules to Correlations”, Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data , New York, USA, 1997, s. 265-276.
Brown, John Seely, Paul Duguid:	Enformasyonun Sosyal Yaşamı , Çev. İbrahim Bingöl, İstanbul, Türk Henkel Dergisi Yayınları, 2001.
Cabena, Peter, v.d.:	Discovering Data Mining : From concept to Implementation , USA, International Business Machines Corporation, 1998. (Peter Cabena, Pablo Hadjinian, Rolf Stadler, Jaap Verhees,Alessandro Zanasi)
Certo, Samuel C.:	Principles of Modern Management Function and Systems , 4.bs., Boston, Allyn and Bacon, 1989.
Chapman, Pete, v.d.:	“CRISP-DM 1.0 Step-by-Step Data Mining Guide,” (Çevrimiçi) http://public.dhe.ibm.com/common/ssi/ecm/en/ytw03084usen/YTW03084USEN.PDF , 16.12.2013, s. 1-44. (Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, Rüdiger Wirth)
Chen, Ming-Syan, Jiawei Han, Philip S. Yu:	“Data Mining: An Overview from Database Perspective”, IEEE Transactions on Knowledge and Data Engineering , C.VIII, No:6, 1996, s.866-883.

Chen, Yu-Chun, Shiao-Chi Wu:	“Exploring Out-Patient Behaviors in Claim Database: A Case Study Using Association Rules”, AMIA Annu Symposium Proceedings , Taiwan, 2003, s.811.
Cleveland, Harlan:	“Information as Resource”, The Futurist , C.XVI, No:6, 1982, s. 34-39.
Crone, Sven F. , Stefan Lessmann, Robert Stahlbock,:	“The Impact of Preprocessing on Data Mining: An Evaluation of Classifier Sensitivity in Direct Marketing”, European Journal of Operational Research , C. CLXXIII, No:3, 2006, s.781-800.
Çınar, Ayşe, Gökhan Silahtaroglu:	“Veri Madenciliği Teknikleri İle Müşteri Memnuniyetine Etki Eden Gizli Nedenlerin Keşfi”, Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi , C.XIII, No:2, 2012, s.309-330.
Çuçen, A. Kadir:	Bilgi Felsefesi , Bursa, Asa Kitabevi, 2001.
Davenport, Thomas H. , Laurence Prusak:	İş Dünyasında Bilgi Yönetimi: Kuruluşlar Ellerindeki Bilgiyi Nasıl Yönetirler , Çev. Günhan Günay, İstanbul, Rota Yayınları, 2001.
Decrop, Alain :	Vacation Decision Making , Londra, CABI Publishing, 2006.
Demir, Filiz Otay, Yalçın Kırdar:	“Müşteri İlişkileri Yönetimi: CRM”, Review of Social, Economic & Business Studies , C.VII, No:8, s.293-308.

Dervişoğlu, H.Gökçe:	Stratejik Bilgi Yönetimi , Kelebek Matbası, Dışbank Kitapları-8, 2004.
Doğan, Nalan K.:	“Entellektüel Sermaye Yönetimi Yaklaşımı ve Entellektüel Varlıkların Korunmasına Yönelik Öneriler”, İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadi Enstitüsü Dergisi , No:47, 2004, s.15-25.
Durna, Ufuk, Yavuz Demirel:	“Bilgi Yönetiminde Bilgiyi Anlamak”, Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi , No: 30, Ocak-Haziran 2008, s.129-156.
Elmas, Çetin:	Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama) , Seçkin Yayıncılık, Ankara, 2003.
Erpolat, Semra:	“Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması”, Anadolu Üniversitesi Sosyal Bilimler Dergisi , C.XII, No:2, 2012, s.137-146.
F.J. Mainly, Bryan:	Multivariate Statistical Methods , 2. bs, Londra: Chapman Hall, 1994.
Fayyad, Usama, Gregory Piatetsky-Shapiro, Padhraic Smyth:	“The KDD Process for Extracting Useful Knowledge from Volumes of Data.”, Communications of ACM , C.XXXIX, No:11, 1996, s.30-31.

Fayyad, Usama, Paul Stolorz:	“Data Mining and KDD: Promise and Challenges” Future Generation Computer Systems , C.XIII, No:2-3, 1997, s.99-115.
Fayyad, Usama, Gregory Piatetsky-Shapiro, Padhraic Smyth:	“Knowledge Discovery and Data Mining: Towards a Unifying Framework”, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96) , C.XXXII, AAAI Press, 1996, s.82-83.
Fernandez, George:	Data Mining Using SAS Applications , Chapman and Hall/CRC, 2002.
Flawley, W.J., G. Piatetsky-Shapiro, C.J. Matheus:	“Knowledge Discovery in Databases : AnOverview”, AI Magazine , C.XIII, No:3, 1992, s. 57-70.
Freitas, Alex A. :	Data Mining and Knowledge Discovery with Evolutionary Algorithms , Almany, Springer-Verlag, 2002.
Ganti, V. , J. Gehrke ve R. Ramakrishnan:	“Mining Very Large Databases”, IEEE Computer , C.XXXII, No:8, 1999,s.38-45.
Getoor, Lise:	“Link Mining: A New Data Mining Challenge”, ACM SIGKDD Explorations Newsletter , C.V, No:1, 2003, s. 84-89.

Giovinazzo, William A.:	Internet Enabled Business Intelligence , U.S.A., Prentice Hall PTR, 2002.
Giudici, Paolo:	Applied Data Mining Statistical Methods for Business and Industry , İngiltere, Wiley, 2003.
Goebel, M., L. Gruenwald:	“A Survey of Data Mining and Knowledge Discovery Software Tools”, ACM SIGKDD Explorations Newsletter , C.I, No:1, 1999, s.20-33.
Gray, Paul, Hugh J. Watson:	Decision Support in The Data Warehouse , U.S.A., Prentice Hall, 1998.
Gupta, Uma G.:	Management Information System: A Managerial Perspective , USA, West Pub. Co., 1996.
Gürdal, Oya:	Tekstil Endüstrisinde Enformasyon Olgusu , Ankara, Türk Kütüphaneciler Derneği, 2000.
Gürüler, Hüseyin, Ayhan İstanbullu, Mehmet Karahasan:	“A New Student Performance Analysing System Using Knowledge Discovery İn Higher Educational Databases”, Journal Computers and Education , C.LV, No:1, 2010, s.247-254.
Hakikur, Rahman:	Data Mining Applications for Empowering Knowledge Societies , Information Science Reference, New York, 2009.

Han, Jiawei, Micheline Kamber:	Data Mining : Concepts and Techniques , Academic Press, 2001.
Han, Jiawei, Micheline Kamber:	Data Mining Concepts and Techniques , 2.bs., USA, The Morgan Kaufman Publishers, 2006.
Han, Jiawei, v.d.:	“DBMiner: A System for Data Mining in Relational Databases and Data Warehouses”, CASCON '97 Proceedings of the 1997 Conference of the Centre for Advanced Studies on Collaborative Research , IBM Press, 1997, s. 1-12.
Hand, David J. :	‘Data Mining: Statistics and More?’, The American Statistician , C.LII, No:2, 1998, s. 112-118.
Hand, David J., v.d.:	“Data Mining for Fun and Profit”, Statistical Science , C.XV, No:2, 2000, s.111 (David J. Hand, Gordon Blunt, Mark G. Kelly ve Niall M. Adams)
Haykin S.:	Neural Networks: A Comprehensive Foundation , USA, Prentice Hall, 1994.
Kantardzic, Mehmed M.:	“General Data Analysis and Data Mining”, Wiley-IEEE Press , 2002, s. 1-18.

Kim, YongSeog , W. Nick Street:	“An Intelligen System for Customer Targeting: A Data Mining Approach”, Decision Support Systems , No:37, 2004, s. 215-228.
Koyuncugil, Ali Serhan:	Veri Madenciliği Ve Sermaye Piyasalarına Uygulanması, Sermaye Piyasası Kurulu Araştırma Raporu , 2007
Kudyba, Stephan:	Managing Data Mining: Advice from Experts , USA, CyberTech Publishing, 2004.
Larose, Daniel T. :	Discovering Knowledge in Data: An Introduction to Data Mining , John Wiley & Sons Inc., 2005.
Leibowitz, Jay:	The Knowledge Management Handbook , CRC Press LLC, 1999
Liao, Shu-Hsien, Yin-Yu Chen:	“Mining Customer Knowledge for Electronic Catolog Marketing”, Expert Systems with Applications , C.XXVII, No:4, 2004, s. 521-532.
Linoff, G., M.J.A. Berry:	Data Mining techniques For Marketing Sales and Customer Relationship Management , New York: Wiley Publishing, 2004.
Lovell, Michael C.:	“Data Mining”, The Review of Economics and Statistics , C.LXV, No:1, 1983, s.1-12.

Madria, Sanjay Kumar, v.d.:	“Research Issues in Web Data Mining”, DaWaK '99 Proceedings of the First International Conference on Data Warehousing and Knowledge Discovery , Berlin, Springer-Verlag Berlin Heidelberg, 1999,s.303-312. (Sanjay Kumar Madria, S. Bhowmick, W. -K. Ng, E. P. Lim)
Maimon, Oded, Lior Rokach:	Data Mining and Knowledge Discovery Handbook , Springer, 2.bs., 2010.
Margareth, Dunham :	Data Mining Introductory and Advanced Topics , Prentice Hall, USA, 2003.
McDermott, Richard:	“Why information technology inspired but cannot deliver knowledge management”, California Management Review , C.XLI, No:4, 1999, s.103-117.
Meyer, Don, Casey Cannon:	Building a Better Data Warehouse , USA, Prentice Hall, 1998.
Misra, D. C., Rama Harihan, Manie Khaneja:	“E-Knowledge Management Framework for Government Organizations”, Information Systems Management , C.XX, No:2, 2003, s.38-48.
Moshkovich, M.Helen, Alexander I. Mechitov, David L. Olson:	‘Rule Induction in Data Mining: Effect of Ordinal Scales’, Expert Systems with Applications , C.XXII, No:4, 2002, s.303-311.

Moss, T. Larissa:	Business Intelligence Roadmap: The Complete Project Lifecycle for Decision- Support Applications , Almanya, Addison Wesley, 2003.
O'Brien, James A.:	Introduction to Information Systems: Essentials for the Interworked EBusiness Enterprise , 10.bs., USA, McGraw-Hill Higher Education, 2001.
Oğuzlar, Ayşe :	Veri Madenciliğine Giriş , Bursa: Ekin Kitabevi, 2004.
Oğuzlar, Ayşe:	“Veri Ön İşleme”, Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi , No:21, 2003, s.67-76.
Orhunbilge, Neyran:	“Uygulamalı Regresyon ve Korelasyon Analizi”, İstanbul, İ.Ü.İşletme Fakültesi Yayınları , 1996, s.1-264.
Özmen, Şule:	“İş Hayatı Veri Madenciliği ile İstatistik Uygulamalarını Yeniden Keşfediyor”, V. Ulusal Ekonometri ve İstatistik Sempozyumu , Çukurova Üniversitesi, Adana, 19–22 Eylül 2001.
Özmen, Şule:	“Veri Madenciliği Süreci”, Veri Madenciliği ve Uygulama Alanları Konferansı, İstanbul Ticaret Üniversitesi , İstanbul, 2003.
Özmen, Şule:	Ağ-Ekonomisinde Yeni Ticaret Yolu: e-Ticaret , İstanbul Bilgi Üniversitesi Yayınları, İstanbul, 2003.

Öztemel, Ercan:	Yapay Sinir Ağları , İstanbul, Papatya Yayıncılık, 2003.
Page, David, Mark Craven:	“Biological Applications of Multi-Relational Data Mining,” ACM SIGKDD Explorations Newsletter , C.V, No:1, 2003, s.69-79.
Park, Sang C., Selwyn Piramuthu, Michael J. Shaw:	“Dynamic Rule Refinement in Knowledgebased Data Mining Systems”, Decision Support Systems , No: 31, 2001, s. 205-222.
Pearl, Judea:	Probabilistic reasoning in intelligent systems: Networks of plausible inference , Morgan Kaufman, 1992.
Piatetsky-Shapiro, Gregory:	“Knowledge Discovery in Real Databases: A Report on the IJCAI- 89 Workshop”, AI Magazine , C.XI, No:5, 1990, s.68-70.
Piramuthu, Selwyn:	“Evaluating Feature Selection Methods for Learning in Data Mining Applications”, European Journal of Operational Research , C.LVI, No:2, Article In Press, 2004, s.483-494.
Probst, Gilbert, Steffan Raub, Kai Romhardt:	Managing Knowledge: Building Blocks for Success , New York, Jhon Wiley & Sons Ltd. 2000.

Quigley, Edward J., Anthony Debons:	“Interrogative Theory of Information and Knowledge”, Proceedings of SIGCPR’99 , Los Angeles, ACM Press, New Orleans, 1999, s.4-10.
Rakesh, Agrawal, Tomasz Imielinski, Arun Swami:	“Database Mining: A Performance Perspective”, IEEE Transactions on Knowledge and Data Engineering , C.V, No:6, 1993, s. 914-925.
Roddick, John:	“Exploratory Medical Knowledge Discovery: Experiences and Issues”, ACM SIGKDD Explorations Newsletter , C.V, No:1, 2003, s. 94-99.
Roiger, Richard J., Michael W. Geatz:	Data Mining A Tutorial-Based Primer , USA: Addison Wesley, 2003.
Rygielski, Chris , Jyun-Chen Wang, David C. Yen:	“Data Mining Techniques for Customer Relationship Management”, Technology in Society , C.XXIV, No:4, 2002, s.488-494.
Saporta, Gilbert:	“Data Mining and Official Statistics”, Quinta Conferenza Nazionale di Statistica , ISTAT, Roma, 2000, s.15-17.
Saruhan, Hamit :	“Genetic Algorithms: An Optimization Technique”, Teknoloji , C.VII, No:1, 2004, s.105-114.

Sevindik, Tuncay, Korhan Kayışlı, Orhan Ünlükahraman:	“Web Tabanlı Eğitimde Veri Madenciliği”, Turkish Journal of Computer and Mathematics Education , C.III, No:3, 2012, s.183-193.
Sezen, Hayrettin Kemal:	Yöneylem Araştırması , Bursa, Ekin Kitapevi, 2004.
Sforna, Marino:	“Data Mining in a Power Company Customer Database,” Electric Power Systems Research , C.LV, No:1, 2000, s. 201-209.
Shaw, Michael J., v.d.:	“Knowledge Management and Data Mining for Marketing”, Decision Support Systems , No:31, s. 132-133. (Michael J. Shaw, Chandrasekar Subramaniam, Gek Woo Tana, Michael E. Welge)
Shieh, Jiann-Cherng:	“The Integration System For Librarians’ Bibliomining”, Electronic Library , C.XXVIII, No:5, 2010, s.709-721.
Sumathi, S., S. N. Sivanandam:	Introduction to Data Mining and its Applications , New York, Springer, 2006.
Swift, Ronald S.:	Accelerating Customer Relationship: using CRM and Relationship Technologies , Prentice Hall PTR, 2001.

Szymanski, Robert A., Donald P. Szymanski, Donna M. Pulschen:	Computers and Information Systems , USA, Prentice-Hall, 1995.
Şahin, Mehmet:	Yönetim Bilgi Sistemi , Eskişehir, Birlik Ofset, 2000.
Şamiloğlu, Famil:	Entelektüel Sermaye , Ankara, Gazi Kitabevi, 2002.
Şimşek Gürsoy, Umman Tuğba:	“Customer churn analysis in telecommunication sector”, İstanbul Üniversitesi İşletme Fakültesi Dergisi , C.XXXIX, No:1, 2010, s. 35-49.
Takçı, Hidayet:	"Tam Otomatik Kütüphane Web Sitesi", ÜNAK'02: I. ÜNAK Genel Konferansı , Samsun, 2002, s.1-4.
Tang, ZhaoHui, Jamie MacLennan:	Data Mining with SQL Server 2005 , Wiley Publishing, Indianapolis, 2005.
Tatei, David M. , Alice E. Smith:	“A Genetic Approach to the Quadratic Assignment Problem”, Computers&Operations Research , C.XXII, No:1, 1994, s.73-83.
Terzi, Serkan:	“Hile ve Usulsüzlüklerin Tespitinde Veri Madenciliğinin Kullanımı”, Muhasebe ve Finansman Dergisi , No:54, 2012, s.51-64.

Thuarisingham, B. M. :	Web Data Mining and Applications in Business Intelligence and Counter Terrorism , Auerbach Publishers, Incorporated, Boca Raton, FL, USA, 2003.
Tiwana, Amrit :	Bilginin Yönetimi , Çev. Elif Özsayar, İstanbul, Dışbank, 2003.
Tüzüntürk, Selim:	“Veri Madenciliği ve İstatistik”, Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi , C.XXIX, No:1, 2010, s. 65-90.
Vatansever, Metin, Ali Hakan Büyüklü:	“Using Visual Data Mining Techniques in Clustering Analysis and An Application”, Mühendislik ve Fen Bilimleri Dergisi , Sigma 27, 2009, s.83-104.
Westphal, Christopher, Teresa Blaxton:	Data Mining Solutions: Methods and Tools for Solving Real-World Problems , U.S.A., Wiley & Sons, 1998.
Williams, Jessica:	Dünyada Değişmesi Gereken 50 Gerçek , Çev. Yurdakul Gündoğdu, İstanbul, Aykırı Yayınevi, 2005.
Yen, Show-Jane, Yue-Shi Lee:	“An Efficient Data Mining Approach for Discovering Interesting Knowledge from Customer Transactions”, Expert Systems with Applications , C.XXX, No:4, 2006, s.650-657.

Yıldırım, Pınar, Mahmut ULUDAĞ, Abdülkadir GÖRÜR:	“Hastane Bilgi Sistemlerinde Veri Madenciliği”, Akademik Bilişim 2008 Konferansı , 2007.
Zaima, Arlene, James Kashner:	Data Mining Primer for the Data Warehouse Professional, (Çevrimiçi) http://www.tdan.com/view-articles/5827 , 16 Aralık 2013.
Zaki, Mohammed J.:	“Parallel and Distributed Association Mining: A Survey”, IEEE Concurrency , C.VII, No:5, 1999, s. 14-25.
Zeleny, Milan:	"Management Support Systems: Towards Integrated Knowledge Management", Human Systems Management , C.VII, No:1, 1987, s.59-70.
Zhong, N., L. Zhou:	Methodologies for Knowledge Discovery and Data Mining, Third Pacific-Asia Conference , Pakdd-99, Beijing, China, April 26-28, 1999 : Proceedings, Springer Verlag, 1999.