



VERİ MADENCİLİĞİNİN ÖNEMİ VE KÜTÜPHANELERDE KULLANIMI*

THE IMPORTANCE OF DATA MINING AND ITS USAGE IN LIBRARIES

Korcan DOĞAN 

Öğr. Gör. Dr., Ankara Üniversitesi, Bilgi ve Belge Yönetimi Bölümü, dogank@ankara.edu.tr

Sacit ARSLANTEKİN 

Prof. Dr., Ankara Üniversitesi, Bilgi ve Belge Yönetimi Bölümü, arslantekin@ankara.edu.tr

Makale Bilgisi

Türü: Araştırma makalesi
Gönderildiği tarih: 3 Mayıs 2023
Kabul edildiği tarih: 1 Haziran 2023
Yayınlanma tarihi: 20 Haziran 2023

Article Info

Type: Research article
Date submitted: 3 May 2023
Date accepted: 1 June 2023
Date published: 20 June 2023

Anahtar Sözcükler

Veri Madenciliği; Büyük Veri; Kütüphane Hizmetleri; Veri Madenciliği Modelleri; Veri Madenciliği Süreçleri

Keywords

Data Mining; Big Data; Library Services; Data Mining Models; Data Mining Processes

DOI

10.33171/dtcfjournal.2023.63.1.21

Öz

Veri madenciliği farklı kaynaklardan toplanan büyük ölçekli verilerden örüntüler bulmak ve anlamlı sonuçlar çıkarabilmek için en önemli yöntemlerden biridir. Kütüphanelerin de farklı kaynaklardan veri toplayabilmesi ve bu verilerden veri madenciliği ile anlamlı sonuçlar çıkarabilmesi önemlidir.

Bu noktadan hareketle çalışmada "Kütüphanecilerin veri madenciliği tekniklerini kullanarak, işlem ve hizmetlerinde yeni örüntüler elde etmesi ve bunları karar destek süreçlerine yansıtarak yeni hizmet modelleri geliştirmek için kullanabilmeleri mümkündür" ana hipotez olarak belirlenmiştir.

Araştırmada kuramsal temelin oluşturulması amacı ile betimsel analiz ve literatür taraması yapılmıştır. Bu aşamada veri madenciliği, veri madenciliği modelleri, veri madenciliği süreçleri vb. kavramlar, yapılan ulusal ve uluslararası çalışmalar doğrultusunda incelenmiş, kütüphanelerde veri madenciliğinin kullanım alanlarına ve uygulamalarına yer verilmiştir.

Araştırma sonucunda veri madenciliği süreciyle elde edilen bulgulara ve değerlendirmelere yer verilmiş, yapılan betimsel analiz ve literatür taramasıyla elde edilen verilerle ana hipotez ve alt hipotezler doğrulanmıştır.

Abstract

Data mining is one of the most important methods to find patterns and draw meaningful conclusions from big data collected from different sources. It is important for the libraries to collect data from different sources and to draw meaningful results from this data by data mining.

In this context, the main hypothesis in this study is: "It is possible for libraries to obtain new patterns in library operations and services by using data mining techniques and to use them to develop new service models by reflecting them in decision support processes."

A literature review and descriptive analysis is conducted to establish the theoretical basis of the research. At this stage, nationwide and international studies including the concepts related to data mining, data mining models, data mining processes, etc. are examined. The areas of utilization and applications of data mining are included in libraries.

As a result of the study, the findings and evaluations obtained through the data mining process are presented, and the main hypothesis and sub-hypotheses are confirmed with the data obtained through descriptive analysis and literature review.

* Bu çalışma Korcan Doğan'ın 2022 yılında Sacit Arslantekin danışmanlığında yaptığı "Bilgi Merkezi ve Hizmetlerinde Veri Madenciliğinin Kullanılabilirliği: Üniversite Kütüphaneleri Örneği" başlıklı doktora tezinden üretilmiştir.

Giriş

Veri toplama ve depolama teknolojisindeki hızlı ilerlemeler, verilerin üretilme, dağıtılma, işleme ve analiz edilme kolaylığı ile birleştiğinde, veri miktarının (hacim), karmaşıklığının (çeşitliliği) ve toplanma ve işleme hızının (hız) büyümesini tetikleyerek içinde bulunduğumuz büyük veri çağına yol açmıştır. Bu büyük veri kümelerinden içgörüler elde etmek, pek çok alanda karar vericiler için giderek daha önemli hale gelmektedir. Bu nedenle, büyüklüğünün ve çeşitliliğinin getirdiği zorluklara rağmen, büyük veriden yararlı bilgiler çıkarmak için otomatikleştirilmiş araçlara gereksinim duyulmaktadır. Bu noktada veri madenciliği, büyük miktardaki veriyi işlemek için geleneksel veri analizi yöntemlerini farklı algoritmalarla harmanlamak için geliştirilmiş bir yöntem olarak öne çıkmaktadır (Tan, Steinbach, Karpatne ve Kumar, 2019, s. 21).

Veri madenciliğinin kütüphanelerde kullanım alanları, yararları, hangi hizmetler için kullanılabileceği, hangi veri madenciliği yöntemlerinin uygulanabileceği ise dünyada olduğu gibi Türkiye’de de araştırılması gereken önemli bir konudur. Nitekim yurtdışı literatürde bu konuda yeterli yayın bulunmakla birlikte ülkemizdeki çalışmalara ilişkin yaptığımız literatür incelemesinde kütüphanelere yönelik yeterli miktarda çalışma bulunmaması bu çalışmanın yapılmasına neden olmuştur. Kütüphaneler için veri madenciliği, daha önceden yapılmayan raporlama imkânları yaratılabilmesi, kütüphane kullanıcılarına daha iyi kullanıcı deneyimiyle hizmetlerin verilebilmesi, kütüphanelerin hizmetlerinde ve iş süreçlerinde yeni öngörüler elde edebilmesi açısından son derece önemlidir.

Kütüphanecilerin veri madenciliği yapabilmeleri için çeşitli kaynaklardan verileri toplayarak depolayabilmeleri ve üzerinde işlem yapabilmeleri gereklidir. Bu durum gerekli veri ambarlarını oluşturabilmek için çalışan personelin ne gibi yeterliliklere ve bilgilere sahip olması da üzerinde durulması gereken önemli bir konudur.

Gerek yerli gerekse yabancı literatür incelendiğinde veri madenciliği kullanımının daha çok kuramsal boyutları ile ele alınmış olduğu ve teknolojik alt yapının yetersizliği nedeniyle pratik düzlemde bir model önerisi sunulmasına olanak bulunmadığı görülmektedir. Günümüzde ise bilgi ve iletişim teknolojilerinde yaşanan büyük dönüşümle birlikte konuyu mesleki açıdan uygulama ekseninde modelleyebilmek olanaklıdır. Özellikle büyük veri ve bulut teknolojilerinde yaşanan değişimle birlikte veri madenciliği programları bulut sunucular üzerinde çalışabilir hale gelmiştir. Kütüphane sunucularında yapılamayan analitik işlemler için veriler

bulut üzerine taşınarak bu işlemler yapılabilmektedir. Analitik işlemlerin bulut üzerinde yapılmasının en önemli nedenlerinden biri, bulut sunucuların kiralama olanakları ile erişilebilir olması, ayrıca yüksek maliyetli donanım ve yazılım yatırımı gerektirmemesidir. Böylece bu yeni teknolojiler, kütüphanelere veri madenciliğini kullanarak işlem ve hizmetlerini geliştirmeleri açısından önemli fırsatlar yaratmaktadır.

Çalışmada veri madenciliğinin günümüzdeki büyük veri teknolojileri gibi teknolojik gelişmelerle gelmiş olduğu noktanın ortaya konularak süreç ve modellerinin açıklanmaya çalışılmasının yanı sıra kütüphanelerin bu gelişmeler ışığında yapması gerekenler saptanmaya çalışılacaktır. Bu nedenle çalışmada aşağıdaki sorulara yanıt aranacaktır:

- Veri madenciliğinde kullanılan süreç ve modeller nelerdir?
- Kütüphaneler hangi hizmetlerinde veri madenciliği kullanılabilir?
- Kütüphanelerde hangi veri madenciliği yöntem ve/veya teknikleri kullanılabilir?

Bu noktadan hareketle çalışmada “Kütüphanecilerin veri madenciliği tekniklerini kullanarak, kütüphane işlem ve hizmetlerinde yeni örüntüler elde etmesi ve bunları karar destek süreçlerine yansıtarak yeni hizmet modelleri geliştirmek için kullanabilmeleri mümkündür” ana hipotez olarak belirlenmiştir.

Mevcut literatürün taranması ve betimsel analiz yöntemleri ile yapılan çalışmanın amacı; veri madenciliğinin genel işlevlerinin ve kütüphanelerin işlem ve hizmetlerinde veri madenciliğinin kullanım alanlarının anlaşılmasını sağlamaktır. Çalışmanın kütüphanecilere ve bilgi profesyonellerine uygulamalarında veri madenciliğinin kullanımı ile ilgili fikir vermesi beklenmektedir.

VERİ MADENCİLİĞİ

Veri madenciliğine ilişkin literatürde pek çok farklı tanım bulunmaktadır. Çalışmanın bu bölümünde veri madenciliği kavramı için kullanılan terimlerin kronolojik olarak nasıl değiştiği ortaya konulmuş, önemli bulunan tanımlar incelenmiş ve çalışma için gerekli görülen tanımlar içeriğe eklenerek farklı yazarların konuya yaklaşımları değerlendirilmiştir.

1960’larda istatistikçilerin kullandığı “veri avlama” (data fishing) veya “veri tarama” (data dredging) terimlerinin, veri madenciliği kavramının ataları olduğu söylenebilir. Bu iki terim de denizcilik literatüründen türetilmiş olup “veri tarama” fiili balık avlanması için denizi ağ ile tarama veya deniz tabanını tarama anlamına

gelen özel bir terimdir. İstatistikçiler ve veri analizi ile uğraşanlar, bu terimleri detaylı analizlere başlamadan önce analiz için uygun olmayan verilerin bulunup ayıklanması için kullanmışlardır (Pektaş, 2013, s. 99).

2000’li yılların başında “veri tabanı madenciliği” (database mining) kavramı, San Diego’ da HNC isimli veri tabanı firması tarafından patentlenmiş ve veri madenciliği üzerine geliştirdikleri bir program grubunu piyasaya sürmek için kullanılmıştır. Sonraki süreçte ise zamanla “veri tabanı madenciliği kavramı”, “veri madenciliği” terimi olarak kullanılmaya başlanmıştır (Pektaş, 2013, s. 99).

Roiger ise çalışmasında, veri madenciliği kavramının ilk olarak 1995 yılında akademik topluluklarda ortaya çıktığını; zaman zaman veri madenciliği yerine de kullanılan VBK’nın (Veri Tabanında Bilgi Keşfi’nin) ise 1989 yılında kullanılmaya başlandığını belirtmiştir (Roiger, 2017, s. 5). Ancak günümüzde bu iki kavram birbirinden farklı anlamları ifade etmektedir.

Veri madenciliği, büyük veri depolarında (veri ambarları) biriken verilerden otomatik olarak yararlı bilgi keşfetme sürecidir. Veri madenciliği teknikleri, bilinmeyen ama yeni ve yararlı olabilecek kalıpları bulmak için büyük veri kümelerinin taranmasını ve gözlemlerin sonucu olarak gelecek ile ilgili kestirimde bulunma yeteneğini sağlamaktadır. Tüm bilgi keşfi görevleri veri madenciliği olarak tanımlanmayabilir. Örnek olarak, veri tabanında bireysel kayıtları arama veya belirli bir anahtar kelime kümesi içeren web sayfalarını bulma gibi sorgular veri madenciliği kapsamında değerlendirilemez. Bu işlem veri tabanı yönetim sistemiyle veya bilgi erişim sistemleriyle gerçekleştirilebileceği için böyle görevler veri madenciliği olarak kabul edilmezler. Bu sistemler, büyük veri depolarından bilgiyi depolamak, organize etmek ve erişmek için karmaşık indeksleme yapıları ve sorgu işleme algoritmaları gibi geleneksel bilgisayar bilimi tekniklerine dayanmaktadır. Bununla birlikte, veri madenciliği teknikleri, girdi sorguları ile ilişkilerine ve ilgilerine dayanarak arama sonuçlarının kalitesini ve bu sistemlerin performansını artırmak için kullanılabilir (Tan vd., 2019, s. 24).

Veri madenciliği, büyük ölçekli veriler arasından bilgiye ulaşma, bilgiyi bir madenden değerli bir cevheri çıkarıyormuşçasına gün yüzüne çıkarma ya da büyük veri yığınları içerisinde gelecekle ilgili kestirimde bulunabilmeyi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranması şeklinde tanımlanabilir (Pektaş, 2013, s. 100).

Şeker (2013, s. 25) ise veri madenciliğini, “çeşitli şekillerde ve çeşitli kaynaklardan toplanan verilerin üzerinde işlem yapılarak anlamlı bilgilerin çıkarılması” olarak tanımlamıştır.

Sankur’un 2004 yılında yayınladığı İngilizce-Türkçe Ansiklopedik Bilişim Sözlüğü’nde “Veri Madenciliği”, “Doğal dillerin semantik yapısına dayanarak elektronik metin belgeleri içinde saklı kalmış ilintileri, örüntüleri, stratejik bilgileri, modelleri vb. bulup ortaya çıkarmayı amaçlayan araştırma tekniği” (Sankur, 2004, s. 831) şeklinde tanımlanmaktadır.

Cackett (2013, s. 10) ise veri madenciliğini, büyük miktarda veriden, otomatik ve yarı otomatik yöntemler kullanmak suretiyle bilinmeyen enformasyonun çıkarılması şeklinde tanımlamış ve bazı yayınlarda veri madenciliğinin geniş veri kümelerindeki bilgi keşfi olarak anıldığını ve bu tanımın, veri madenciliğinin geçmişi şeklinde düşünülebileceğini belirtmiştir.

Bir başka kaynakta veri madenciliği, veri ambarlarında saklanan yararlı olabilecek, aralarında bilinmeyen ilişkilerin olduğu verilerin keşfedilerek, bu verilerin hem anlaşılır hem de kullanılabilir bir şekle dönüştürülmesine yönelik geliştirilmiş yöntemler topluluğudur (Oğuzlar, 2011, s. 5-6).

Arslantekin (2003, s. 372-373), veri madenciliğini “büyük miktarda veriden anlamlı bilgi çıkarma sanatıdır” şeklinde tanımlamış, konuya “toplanan büyük yığın halindeki veriler arasında örnek kalıpların tanımlanması, eğilimlerin belirlenmesi ve gerekli ilişkilerin kurulması işlemlerine ait bir süreçtir” şeklinde devam etmiştir.

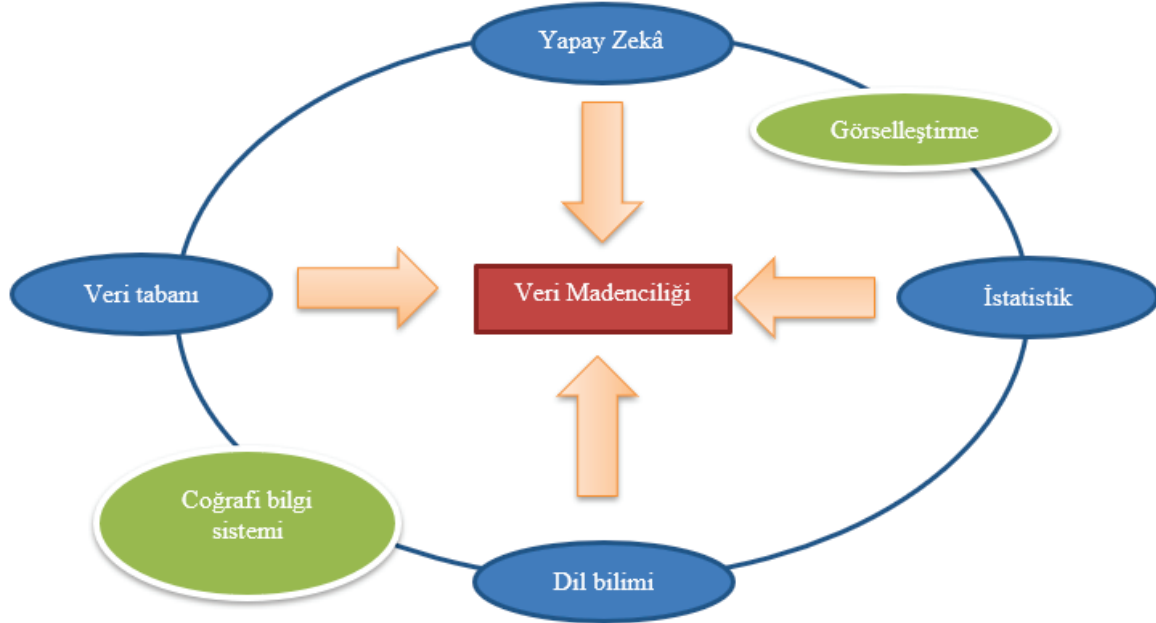
Prytherch (2005, s. 195) ise veri madenciliğini, ilişkileri ve modelleri analiz etme amacı ile gizli enformasyonu keşfetmek için ayrıntılı teknikler kullanarak, enformasyonun veri tabanından ve veri kümelerinden çıkarılması işlemi olarak tanımlamıştır.

Veri madenciliğine pek çok farklı bilim dalı katkıda bulunmaktadır. Bu nedenle veri madenciliğinin farklı terim ve kavramlarla ilişkisi bulunmaktadır.

Veri Madenciliği ile İlişkili Terim ve Kavramlar

Roiger, veri madenciliğinin kökenlerinin istatistik, matematik, makine öğrenimi, yapay zekâ ve işletme alanlarında bulunduğunu belirtmekte; VBK’nın ise sıklıkla, veri madenciliği ile birbirinin yerine kullanıldığını vurgulamaktadır (Roiger, 2017, s. 5).

Veri madenciliği yöntemleri genel olarak istatistik, yapay zekâ ve yapay zekânın bir uzantısı olan makine öğrenimi olmak üzere iki ana kökten beslenmektedir. Bu iki ana disiplinin yanında Şekil 1’de görüldüğü gibi veri tabanı yönetim sistemleri, dilbilim, görselleştirme ve coğrafi bilgi sistemleri de veri madenciliğinde önemli rol oynamaktadır (Akpınar, 2014, s. 50).



Şekil 1. Veri madenciliğine etki eden terim ve kavramlar (Akpınar, 2014, s. 50)

Şekil 1’de yer alan veri madenciliği ile ilişkili terim ve kavramlara aşağıda kısaca değinilmiştir.

İstatistik: Modern istatistik, 20. yüzyıl başlarından beri veri bilimi olarak tanımlanmasına rağmen günümüzde popüler olarak kullanılan ve bilgi teknolojilerinin eklenmesiyle içeriği değiştirilmiş olup, veri bilimden bazı alanlarda ayrılmaktadır (Altunkaynak, 2022, s. 5).

İstatistik ve makine öğrenimi, veri madenciliğinin temelinde yer alan en önemli iki disiplindir. Bu iki akademik disiplinde de çeşitli sınıflandırma yöntemleri ve algoritmaları bulunmaktadır. İstatistik içerisinde yer alan sınıflandırma yöntem ve algoritmalarından lineer regresyon analizi, lojistik regresyon analizi ve diskriminant analizi, bu yöntemlere örnek olarak verilebilir. Makine öğrenimi kapsamında geliştirilen yöntemlerden bazıları ise karar ağacı (decision tree), yapay sinir ağı (artificial neural network) ve destek vektör makineleridir (support vector machines). Bu yöntemlerden özellikle diskriminant analizi, istatistik ve makine öğrenimi arasında bir arakesit oluşturmaktadır. Makine öğrenimi ise yapay zekâ çalışmalarının bir uzantısıdır. Yapay zekâ, insana özgü zekâ davranışlarının

otomasyonunu araştıran, bilginin saklanması ve işlenmesinde veri yapıları, algoritmalar, programlama dilleri ve teknikleri gibi bilgi işlem yöntemlerini kullanan, bilgisayar biliminin bir alt disiplini (Akpınar, 2014, s. 51).

Veri madenciliğinin bir kısım yöntemleri istatistiksel yöntemlere dayanmaktadır. Ancak veri madenciliği yöntemleri klâsik istatistiksel yöntemlerden aşağıdaki özelliklerle farklılık göstermektedir (Altunkaynak, 2022, s. 5).

1. Veri madenciliği yöntemleri istatistiğe göre çok daha büyük veri yapılarında kullanılmaktadır.

2. Veri madenciliğinde, klâsik istatistiksel yöntemlerin gerektirdiği birçok varsayım (dağılım varsayımları, varyans homojenliği varsayımı, dağılım yakınsamaları için gerekli varsayımlar vb.) gerekli değildir.

3. Çıkarılacak sonuçlar için klâsik istatistikte hipotezler oluşturulur ve bunları test etmek için uygun veriler toplanır. Veri madenciliğinde ise faydalı bilgiler gerekli/gereksiz toplanmış veri yığınlarından çıkarılır.

4. Genellikle veri madenciliği açıklayıcı, istatistik ise doğrulayıcı yöntemlerden oluşmaktadır.

5. Yeni veriler elde edildikçe veri madenciliğinde mevcut modellerin veya kuralların eğitilmesine devam edilebilmektedir.

6. Veri madenciliği genelde çok büyük hacimli veriye dayalı modeller kullanırken istatistik yeterli bir örneklem büyüklüğüne dayalı olasılık modelleri kullanmaktadır (Altunkaynak, 2022, s. 5).

Verilen açıklamalardan da anlaşılacağı üzere makine öğrenimi, bilgisayarların veri üzerinde öğrenme ve kendini geliştirme kabiliyetini kazanması için kullanılan bir dizi algoritma ve teknikleri içeren bir yapay zekâ alt dalı olup bu algoritma ve teknikler, özellikle istatistik, veri analizi, matematik, bilgisayar bilimi ve diğer disiplinlerden geliştirilmiştir.

Yukarıda sayılan özelliklerden dolayı çalışmada istatistiksel modeller yerine aşağıda verilen veri madenciliğine dayalı modeller açıklanmaya çalışılmıştır.

Dilbilimi: Günümüzde dilbilim alanı ile veri madenciliği arasında yakın ilişki kurulabilmektedir. Kelime-komşuluk ağları ile yalnızca metin içindeki kelimelerin bir arada bulunma durumları değil, aynı zamanda birbirlerine yakınlık durumları da incelenebilmektedir (Zweig, 2016, s. 156).

Kelime-komşuluk ağları ile (word-adjacency networks), kelimelerin karşılıklı ortak oluşumları dikkate alınarak, belirli bir metin temsili olarak değerlendirilebilir. Böylece metinler, kelime yapılarını nicel olarak tanımlayan bazı parametrelerle karakterize edilebilirler. Bu parametrelerin belirlenmesi, sayı kümelerinin elde edilmesiyle sonuçlanmaktadır. Bu anlamda bir metinden bir ağ oluşturma ve terimlerin karakteristiklerini hesaplama süreci, dil örneğinin temel özelliklerini çıkarmak ve bunları yeni bir forma dönüştürmek olarak görülebilir. Elde edilen verilerin veri madenciliği algoritmalarına tabi tutulması ile de metinlerin içindeki örüntülerin ortaya çıkarılması ve metinlerle ilgili kestirimde bulunulması sağlanabilir. Örnek olarak farklı yazarların metinlerini temsil eden ağların parametrelerinden, metinlerin birbirinden ayırt edilebileceği ortaya koyulabilmektedir. Böylelikle hangi metnin hangi yazara ait olduğu belirlenebilmekte; dilbilim ve veri madenciliği, metin madenciliği altında birleşerek, bireysel dil yapılarının tanımlanmasına ilişkin sonuçlar çıkarılabilmektedir (Akimushkin vd., 2017; Mehri vd. 2012'den aktaran Stanisiz vd., 2019, s. 302).

Veri tabanları: Veri tabanları, veri madenciliği için gerekli olan verileri elde etmede ve veri madenciliği sürecinde gerekli olan verilerin saklanması veri ambarları ile birlikte kullanılabilir. Veri madenciliği uygulamalarında veri tabanları, veri kaynaklarının hazırlanması ve veri temizliği gibi önemli işlemlerin gerçekleştirilmesinde, veri madenciliği modellerinin oluşturulması ve test edilmesi için kullanılan veri setlerinin saklanması da kullanılabilir.

Görselleştirme: Veri madenciliğinin ilişkili olduğu bir diğer kavram ise görselleştirme. Görselleştirme, veri madenciliği süreçlerinde yapılan işlemleri desteklemek için son derece önemlidir. Veri madenciliğinin ön işleme sürecinde veriler arasındaki ilişkileri incelemek, verileri daha iyi anlayabilmek, içlerindeki aykırı değerleri saptamak vb. konularda veri görselleştirme kullanılabilir (Olson D. L., 2018, s. 10,32). Veri madenciliği sonucu elde edilen sonuçların gösterilmesi ve sunumu için de görselleştirme kullanılmaktadır (Olson D. L., 2018, s. 34).

Coğrafi bilgi sistemleri: Coğrafi bilgi sistemlerinde de veri madenciliğinin önemli bir yeri vardır. Örnek olarak ulaşım, ekoloji, kamu güvenliği, halk sağlığı, iklim bilimi, vb. alanlarda konuma dayalı veri işleyen coğrafi bilgi sistemlerinde aykırı değerlerin tespiti veri madenciliği ile mümkün olabilmektedir (Oded ve Rokach, 2010, s. 117, 126, 841).

Yapay zekâ: 1950'lerin ortalarında yapay zekâ bir disiplin olarak ilk defa tanındığında, makine öğrenimi de yapay zekâ disiplinin merkezinde yer alan bir araştırma alanı olmuştur (Quinlan, 1986, s. 81). Bu kadar önem verilmesinde temel olarak şu neden ileri sürülebilir:

“Öğrenebilme yeteneği”, akıllı davranışın ayırt edici bir özelliğidir. Dolayısıyla zekâyı bir fenomen olarak anlama çabası, bir öğrenme anlayışını içermelidir. Daha somut olarak ise öğrenme, yüksek performanslı sistemler oluşturmak için potansiyel bir metodoloji sağlamaktadır (Quinlan, 1986, s. 81).

Bu noktada profesyonel bir alan olarak gelişmekte olan veri madenciliğinde kullanılan istatistiksel modellemenin, geleneksel anlamdaki veri tabanlarından bilgi keşfinden daha geniş aktivite alanlarından kavramsal olarak ayrılması gerektiği ortadadır. Nisbet vd. (2018, s. 22) bu kavramları aşağıdaki şekilde tanımlamışlardır:

- İstatistiksel modelleme: Kestirici değişkenlere (predictor variables) dayanarak bir sonucu veya olayı gruplamak veya kestirimde bulunabilmek için parametrik istatistiksel algoritmaların kullanılmasıdır.

- Veri madenciliği: Büyük, gürültülü ve dağınık veri kümelerindeki veri elemanları arasındaki belirsiz ilişki kalıplarını saptamak için makine öğrenme algoritmalarının kullanılmasıdır.

- Bilgi keşfi: Tüm veri erişimi, veri araştırma, veri hazırlama, modelleme, model dağıtımı ve model izleme süreci olarak tanımlanabilir. Veri madenciliği faaliyetlerini de içermektedir.

- Veri bilimi: Bilginin keşfini gerçekleştirebilmek için bir yandan karmaşık görüntü, konuşma ve metin analizi vb. içeren analitik veri ambarlarının veri mimarisi ile diğer yandan da çok gelişmiş makine öğrenme algoritmaları ile işlenmesi şeklinde tanımlanabilir.

Roiger ise veri madenciliği ile ilgili kavramlara daha farklı yaklaşmaktadır. Roiger'a göre hem veri biliminin hem de VBK'nin temel amacı mevcut verilerden yararlı bilgiler elde etmektir. Bununla birlikte, iki kavramın iki belirgin ayrımı olduğunu belirtir. İlk olarak, veri bilimi genellikle büyük hacimli verilerle ilişkilendirilirken, VBK'nin kökenleri veri tabanı topluluğundadır. İkinci olarak, veri madenciliği VBK süreç modeli içinde gerekli bir adımdır. Veri madenciliğinin veri bilimi uygulamalarında sıkça görülmesine rağmen gerekli olmadığını, tek gerekliliğin, analiz süreci için kullanılan modellerin verilerden gelmesi olduğunu ifade etmektedir (Roiger, 2017, s. 5-6).

VBK, yapılandırılmış bir veri topluluğundan çeşitli istatistiksel yöntemler ve veri madenciliği ile yapılan bir işlemdir. Buna karşılık veri ambarlarında bulunan yapılandırılmamış verilerin bilgi keşfi amacıyla kullanılması tamamıyla veri madenciliğine dayanmaktadır.

Bu noktada üzerinde durulması gereken bir diğer kavram da veri analitiği ve veri analizidir.

Veri analizi bir şeyi keşfetmek veya karar vermeye yardımcı olmak amacıyla özellikle bir bilgisayar kullanarak bilgiyi inceleme süreci olarak tanımlanmıştır (Data analysis, t.y.). Veri analizi, verileri toplama, sınıflandırma ve çözümlene yöntemlerini kullanarak verilerden anlamlı bilgiler elde etme sürecidir. Bu süreç, verilerin anlamını ortaya çıkarmak ve bu verileri kullanarak karar verme süreçlerini desteklemek için kullanılır. Finans, sağlık, mühendislik ve pazarlama gibi çeşitli alanlarda veri analizi yöntemleri kullanılır.

Veri analitiği ise veri analizi için kullanılan daha modern bir terimdir. Günümüzde genellikle, veri analitiği, modern veri bilimi yöntemlerini kullanarak yapılan araştırma ve değerlendirmeleri ifade etmekteyken, veri analizi geleneksel istatistiksel yöntemlere dayalı analitik faaliyetleri ifade etmektedir (Yu, 2022, s. 22).

Veri madenciliğine dayalı olarak karar alırken geleceği öngörmek ya da geleceğe yönelik çıkarımlarda bulunmak karar alıcıya önemli avantajlar sunmaktadır. Dolayısıyla bu kavramların ayrımının ayırt edilmesinde fayda bulunmaktadır.

Geleceği öngörebilmek geçmişte de insanoğlunun en büyük hayallerinden biri olduğu için kehanet terimi bu alanda ilk karşımıza çıkan terimdir. Nitekim bu terim günümüzde de varlığını sürdürmektedir. “Kehanet” (divination), standart ritüeller eşliğinde, gizemli bilgiler kullanılarak bir soru veya durum için geleceği öngörmektir (Akpınar, 2014, s. 48).

Veri madenciliği ve kestirim analizleri, veri modelleri yardımıyla olası olaylar hakkında kehanette bulunmaya çalışan bir alan olarak karşımıza çıkmaktadır (Sathishkumar vd., 2020, s. 971) ve kehanetin günümüzdeki bilimsel yöntemi haline gelmiştir.

Veri analizi alanında sıkça kullanılan diğer üç terim ise estimation, forecasting ve prediction'dır. Bu kavramların Türkçe karşılıkları; estimation için “tahmin”, forecasting için “öngörü”, prediction için “kestirimdir” (T.C. Başbakanlık Devlet İstatistik Enstitüsü Başkanlığı, s. 32, 36, 70).

Bu kavramlar birbirlerine yakın anlamlar içermesine rağmen aslında çok farklıdır. Bu kavramların çalışmada hangi anlam ve kapsam doğrultusunda kullandığına kısaca göz atmakta fayda bulunmaktadır.

Tahmin, kitlenin bir parametresi için önerilen tahmin edicinin değeridir (Akdi, 2018). Bir örnekten toplanan bilgilere dayanarak bir evren parametresi için sayısal bir değer sağlama sürecidir. Bilinmeyen parametre için tek bir rakam hesaplanırsa, işleme nokta tahmini denir. Parametreyi içermesi muhtemel bir aralık hesaplanırsa, aralık tahmini olarak adlandırılır (Everitt ve Skrondal, 2010, s. 154).

Kestirim, bir rastgele değişken için seçilen bir modelin parametrelerinin tahmin değerleri yerine konulduğunda elde edilen değerdir (Akdi, 2018). Kestirim, görünmeyen verilerin sonuçlarının tahmin edilmesi ile ilgilidir (Döring, 2018). Bu amaçla eğitim veri kümesi ile bir model oluşturulur ve bu modelde yeni x örnekleri için kestirimde bulunabilen bir kestirimci $f(x)$ kullanılır (Döring, 2018). Kestirim uygulamalarına örnek olarak konuşma tanıma, makine öğrenimi, örüntü tanıma uygulamaları verilebilir (Dunham, 2002, s. 7). Kestirim, ayrıca regresyon analizinde bağımsız (açıklayıcı) değişkenler tarafından açıklanmaya çalışılan bağımlı (açıklanan) değişkeni tahmin etmede kullanılmaktadır.

Öngörü (forecasting), zaman serisi verilerine dayanarak gelecek hakkında kestirimde bulunulan bir kavram olarak, kestirimin bir alt disiplinini ifade etmektedir (Döring, 2018). Bu nedenle kestirim ve öngörü arasındaki en önemli fark zaman boyutunun dikkate alınıp alınmamasıdır (Döring, 2018). Öngörü bir modele göre, model parametrelerinin tahmin edilmesi ile rastgele değişkenin gelecekte alacağı değer için önceden kestirim anlamında kullanılmaktadır (Akdi, 2018).

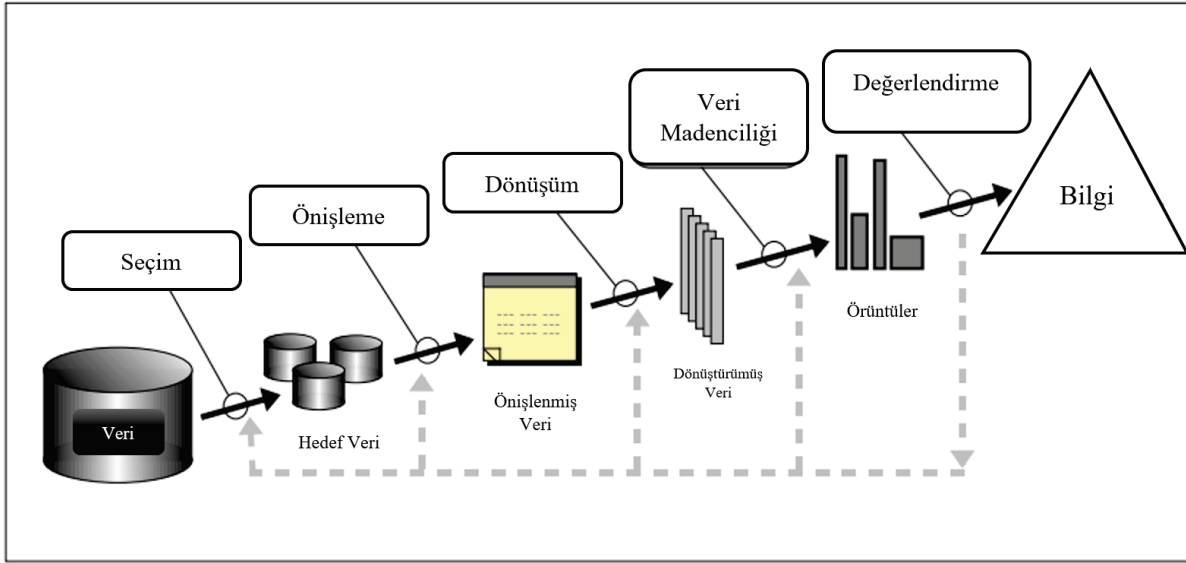
Bir diğer değinilmesi gereken kavram ise “iş zekasıdır”. Veri madenciliği, işletme uygulamalarında kullanıldığında ise “iş zekâsı” adını almaktadır (Yu, 2022, s. 22). Dolayısı ile bu kavramın kestirim ve öngörü ile daha ilişkili olduğu söylenebilir.

Veri madenciliği uygulamaları kimi zaman uzunca bir süreç almaktadır ve bu süreçte hangi veri madenciliği modelinin seçilmesi gerektiğine karar verilmesi gerekmektedir.

Veri Madenciliği Süreç ve Modelleri

Veri madenciliği konusunda literatürde önerilen çeşitli süreçler bulunmaktadır.

Çalışmada ilk olarak Fayyad vd. (1996, s. 41) önerdiği süreç incelenecektir.



Şekil 2. Fayyad vd. (1996, s.41) tarafından önerilen veri madenciliği süreci

Şekil 2 'de verilen bu süreç beş ana aşamadan oluşmaktadır:

- Seçim aşamasında önemli olduğu düşünülen veri seçilir veya oluşturulur ve bu yeni veri artık hedef veri (target data) adını alır.

- Bu aşamadan sonra önişleme (preprocessing) aşamasına geçilir. Veri madenciliğinde bütün süreçte iyi sonuç elde edebilmek için başlangıçta hazırlanan veri uygun duruma getirilmelidir. Veride gürültü (noise) miktarı azaldıkça, elde edilebilecek sonuçların kalitesi artar. Bu aşama, sorunlu verinin temizlenmesi, eksik verinin bir şekilde tamamlanması ve yeni özniteliklerin oluşturulması adımlarını kapsamaktadır.

- Dönüştürme (transformation) aşamasında verinin farklı veri madenciliği yöntemleri için uygun formata dönüştürülmesi hedeflenir.

- Veri madenciliği aşaması, belirlenen amaçlara uygun veri madenciliği yöntemlerinin kullanılması aşamasıdır. Bu aşama örüntüleri elde ettiğimiz aşamadır.

- Yorumlama (interpretation) aşamasında ise elde edilen örüntülerin yeterli bilgi içerip içermediği yorumlanır. Eğer inceleme sonucunda elde edilen örüntüler yeterli bulunmazsa önceki adımlara geri dönmelidir (Akpınar, 2014, s. 72-73).

CRISP-DM (Cross Industry Standard Process for Data Mining) süreç modeli veri madenciliği için kullanılan önemli bir modeldir (IBM SPSS Modeler CRISP-DM Guide, 2000, s. 1). Bu modele Pektaş'ın (2013, s. 104) ve Akpınar'ın (2014, s. 74) çalışmasında da yer verilmiştir.



Şekil 3. CRISP-DM Süreç Diyagramı (IBM SPSS Modeler CRISP-DM Guide, 2000, s. 1), (Akpınar, 2014, s. 74)

CRISP-DM'de veri madenciliği süreci 6 ana aşamadan oluşmaktadır. Şekil 3'de verilen aşamaların sırası kesin olmamakla birlikte farklı aşamalar arasında ileri ve geri geçişler söz konusu olabilmektedir. Bu aşamalar (IBM SPSS Modeler CRISP-DM Guide, 2000, s. 1):

- İşin anlaşılması (Business understanding): Başlangıç aşamasında veri madenciliğinden beklentilerin keşfedilebilmesi gereklidir. Bu tartışmalara mümkün olan en çok sayıda ilgili kişi veya birim dahil edilmeye çalışılmalıdır ve sonuçların belgelendirilmesi önemlidir. Bu aşama kimi zaman gereksiz gibi görülebilir ancak veri madenciliğinin hangi nedenlerle yapıldığını öğrenmek, değerli kaynakları harcamadan önce kişilerin ve birimlerin aynı fikirde olduğundan emin olunmasına yardımcı olacaktır. Bu aşamanın alt başlıkları ise yapılacak işe genel bakış, işin hedeflerini belirleme, durumu değerlendirme, veri madenciliği hedeflerini belirleme, bir proje planı üretme adımlarından oluşmaktadır.

• Verinin anlaşılması (Data understanding): Bu aşamada veri madenciliği yapılabilmesi için mevcut verilerin daha yakından incelenmesi gerekir. Bu adım, genellikle projenin en uzun bölümü olan veri hazırlama aşamasında beklenmeyen sorunları önlemek için son derece önemlidir. Verinin anlaşılması, veriye erişmeyi ve düzenlenecek tablolar-grafikler ile veriyi keşfetmeyi içermektedir. Bu aşama, verinin kalitesinin belirlenmesini ve bu adımların sonuçlarının proje belgelerinde açıklamasını sağlamaktadır. Bu aşamanın alt başlıkları ise başlangıç verilerinin toplanması, verilerin tanımlanması, verilerin keşfi, veri kalitesinin doğrulanmasıdır.

• Veri hazırlama (Data preparation): Veri hazırlama, veri madenciliğinin en önemli ve genellikle en fazla zaman alan aşamalarından biridir. Veri hazırlama aşamasının, genellikle bir projedeki toplam harcanan zamanın ve çabanın %50-70'ini aldığı tahmin edilmektedir. Önceki aşamalara daha fazla özen gösterilmesi bu aşamanın iş yükünü en aza indirebilecektir, ancak yine de bu aşamada yoğun emek harcanması gerekmektedir. Organizasyona ve hedeflere bağlı olarak, veri hazırlama aşaması genellikle şu görevleri içerir: Veri kümelerinin ve/veya kayıtların birleştirilmesi, verinin bir örnek alt kümesinin seçilmesi, yeni değişkenlerin türetilmesi, modelleme için verinin sıralanması, boş veya eksik değerlerin kaldırılması veya değiştirilmesi, eğitim ve test veri kümelerine bölünmesidir.

• Modelleme (Modeling): Bu aşamada elde ettiğimiz sonuçlar için anlaşılması aşamasında ortaya konulan iş sorununa ışık tutmaya başlayacaktır. Öncelikle uygun modelleme yöntemi seçilir ve uygulanır. Modelleme genellikle bir sürecin veya algoritmanın aynı adımların tekrar tekrar uygulanması yoluyla yapıldığı bir döngü (iterasyon) olarak gerçekleştirilmektedir. Genel olarak, veri madencileri varsayılan parametreleri kullanarak birkaç model çalıştırır ve daha sonra parametreleri değiştirerek ayarlar veya seçtikleri model için gerekli manipülasyonlar için veri hazırlama aşamasına geri dönerler. Veri madenciliği sorusunun tek bir model ve tek bir yürütme ile tatmin edici şekilde cevaplanması çok nadir olarak gerçekleşir. Bu aşama; modelleme tekniklerinin seçilmesi, bir test tasarımının oluşturulması, modellerin oluşturulması, modelin değerlendirilmesi alt adımlarından oluşmaktadır.

• Değerlendirme (evaluation): Bu aşamada, veri madenciliği projesinin büyük çoğunluğu tamamlanmıştır. Modelleme aşamasında, önceden tanımlanan veri madenciliği başarı kriterlerine göre, oluşturulan modellerin teknik açıdan ne kadar doğru ve etkili olduğunu belirlenir. Ancak devam edilmeden önce, projenin başlangıcında belirlenen iş başarı kriterleri kullanılarak verilen emeğin sonuçlarının değerlendirilmesi gerekir. Yapılan bu işlem, elde edilen sonuçların

kullanılabilmesinin sağlamanın anahtarıdır. Veri madenciliği ile modeller ve modellerden elde edilen herhangi bir sonuç veya kestirim şeklinde (bulgular) iki tür sonuç üretilebilir.

- Kullanıcılarla paylaşım (deployment): Modelin oluşturulması sonrasında elde edilen bilginin son kullanıcıların faydalanabileceği şekilde düzenlenmesi ve erişime açılması gereklidir. Yeni içgörülerin organizasyonda iyileştirmeler yapmak için kullanılması da bu aşamada gerçekleşir. Bu aşamaya kadar yapılan uygulanma iş süreçlerine entegre edilebilir. Ayrıca, veri madenciliğinden elde edilen içgörülerle yapılan iş ile ilgili değişiklikler yapılabilir.

Veri madenciliğinde Crisp-DM ve Fayyad vd. tarafından önerilen veri madenciliği süreçlerinden başka süreçlerde bulunmaktadır. Tablo 1'de de görülebileceği gibi Cios vd. (2007, s. 17-18) hazırlamış oldukları çalışmada, Crisp-DM ve Fayyad vd. tarafından önerilen veri madenciliği sürecini de ekleyerek beş farklı veri madenciliği sürecini detaylı olarak karşılaştırmışlardır.

Tablo 1. Veri Madenciliği Süreçleri İle İlgili Beş Farklı Yaklaşımın Karşılaştırması (Cios vd., 2007, s. 17-18)

Model	Fayyad vd.	Anand ve Buchner	Cios vd.	Cabena vd.	CRISP-DM
Ortaya Çıktığı Alan	Akademik	Akademik	Akademik/Endüstri	Endüstri	Endüstri
Adım sayısı	9	8	5	4	6
Adımlar	1. Uygulama alanının anlaşılması ve geliştirilmesi	1. İnsan kaynaklarının belirlenmesi 2. Problemin Tanımlanması	1. Problem alanının anlaşılması	1. İş Hedeflerinin Saptanması	1. İşin anlaşılması
	2. Hedef Veri Setinin Oluşturulması	3. Veri araştırması	2. Verinin anlaşılması		2. Verinin Anlaşılması
	3. Verinin temizlenmesi ve ön işlenmesi	4. Alan bilgisinin ortaya çıkarılması	3. Verinin hazırlanması	2. Verinin Hazırlanması	3. Verinin hazırlanması
	4. Verinin temizlenmesi ve indirgenmesi	6. Verinin ön işlenmesi			
	5. Veri madenciliği işinin seçilmesi				
	6. Veri madenciliği algoritmasının seçimi	5. Metodolojinin tanımlanması			
	7. Veri madenciliği	7. Örüntü Keşfi	4. Veri Madenciliği	3. Veri Madenciliği	4. Modelleme
	8. Örüntülerin Yorumlanması	8. Bilginin son işlenmesi	5. Keşfedilmiş bilginin değerlendirilmesi	4. Sonuçların Analizi	5. Değerleme
	9. Keşfedilmiş bilginin birleştirilmesi				6. Kullanıcılarla Paylaşım

Tablo 1’de veri madenciliği süreçleri ile ilgili beş farklı yaklaşımın karşılaştırması görülmektedir. Bu beş farklı yaklaşım da içerikleri farklı gibi görünmesine rağmen, tablo incelendiğinde birbirlerine yakın oldukları görülebilmektedir.

Örnek olarak CRISP-DM sürecinde ilk adım olan “işin anlaşılması” adımı diğer süreçlerde farklı isimlerle bulunmakta, yalnızca Anand ve Buchner’in sürecinde “problemin tanımlanması” ve “insan kaynaklarının belirlenmesi” şeklinde iki adımda yer almaktadır. Aynı şekilde CRISP-DM sürecinin üçüncü adımı olan “verinin hazırlanması” adımı, Anand ve Buchner’in sürecinde “verinin ön işlenmesi”, Fayyad vd.’lerinin sürecinde “verinin ön işlenmesi ve temizlenmesi” ile başlayan 3 adım olarak ele alınmıştır. Cios vd. ve Cabena vd.’nin süreçlerinde ise aynı isimle yer almaktadır.

Fayyad vd. Tablo 1’de verilen bu model ile VBK sürecine odaklanmış ve VBK sürecini veri madenciliğini içeren şemsiye bir terim olarak tanımlamıştır. VBK sürecinin veri madenciliği adımı ile verilerden ilgili örüntüleri bulma amacıyla yapıldığını belirtmiştir (Fayyad vd., 1996, s. 40).

Görüldüğü üzere veri madenciliği ya da bilgi keşif süreçleri ile ilgili pek çok farklı yaklaşım bulunmaktadır. Adımların, birbirleri arasındaki geçişlerde kısmi farklılıklar olsa da önemli olan çalışmada da görüleceği üzere herhangi bir süreç modelini referans almaktır. Sonrasında o modeli takip etmek, süreçte beklemeyen durumlar, çeşitli aksaklıklar vb. problemler ortaya çıktığında, modelden çıkarak tekrar başa ya da herhangi bir aşamaya dönmek mümkün olabilecektir.

Tablo 1’de görülen verinin hazırlanması ve veri ön işleme aşamaları, veri madenciliği sürecinin pek çok çalışmada en çok zaman alan aşamalarından biridir. Veri analizi sürecinin tamamlanabilmesi ve veri analizi sonuçlarının kalitesinin artırılabilmesi amacını taşımaktadır.

Verinin ön işleme süreci tamamlandıktan sonra veri madenciliği modellerinden çalışmaya uygun olan bir tanesi seçilerek veri madenciliği uygulamasına geçilir. Bu aşamada da veri madenciliği modelleri konusunun açıklanmasında fayda görülmektedir.

Şekil 3’te verdiği gibi veri işleme aşamasından sonra modelleme aşamasına geçilmelidir. Modelleme, veri madenciliği sürecinde elde edilen verileri kullanarak gelecekteki olayları tahmin etmek için ve veri içindeki örüntüleri ortaya çıkarabilmek için yapılır. Veri madenciliğinde modelleme, veri madenciliği sürecini tamamlayan ve

verilerden anlamlı bilgiler elde ederek bu bilgileri kullanarak karar verme süreçlerini destekleyen bir aşamadır.

Veri madenciliği modelleri ile mevcut veri türüne bağlı olarak farklı türde analizler yapılabilmektedir. Kümeleme ve birliktelik kuralları betimleyici modeller olarak kullanılmaktadır. Kestirim amacı ile ise sınıflandırma modelleri kullanılmaktadır (Olson ve Araz, 2023, s. 2).

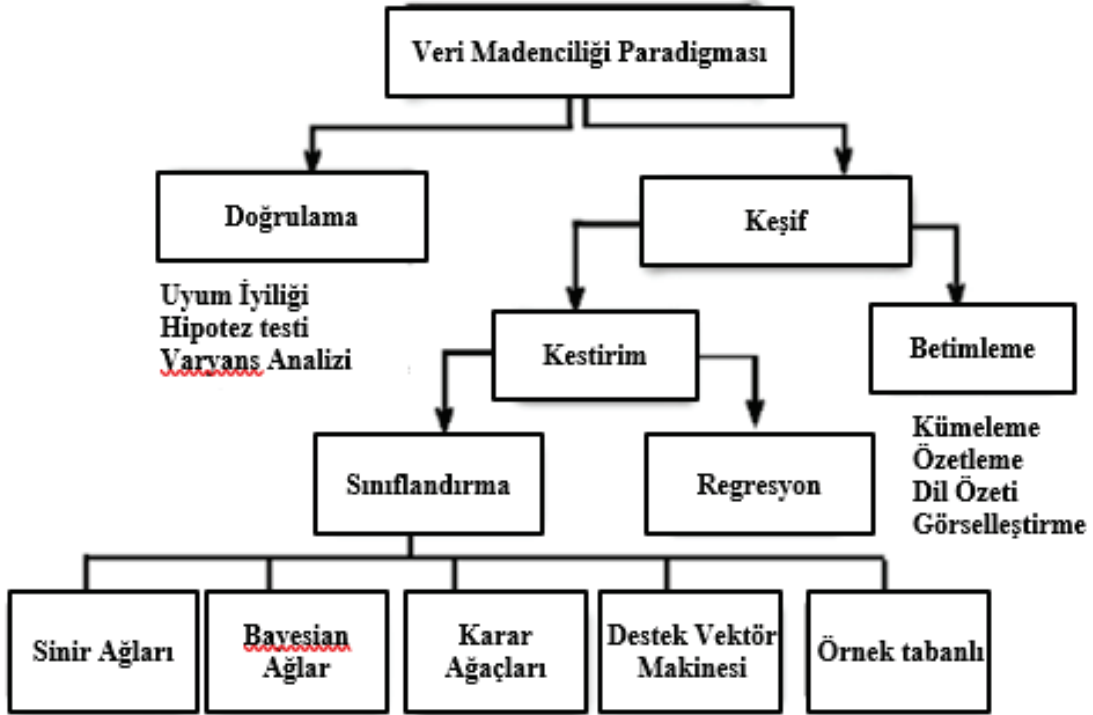
Sınıflandırma, sınıflandırılmak istenen her bir nesnenin, tam olarak birbirinden farklı ve tamamen ayrı bir sınıfa atanması, yani asla birden fazla sınıfa ve yine asla hiçbir sınıfa atanmamasını ifade etmektedir (Zaki ve Meira, 2020, s. 467). Veri madenciliğindeki sınıflandırmanın makine öğreniminde karşılığı “denetimli öğrenme” olarak kullanılmaktadır.

Kümelemede ise verilerin birbirinden farklı kümelere gruplaması amaçlanmaktadır. Küme içindeki veri noktalarının birbirine benzer veya veri uzayında birbirine yakın olması istenir. Diğer yandan, kümelerin farklı olması veya aralarında büyük mesafeler olması istenir. Kümeleme ise makine öğreniminde “denetimsiz öğrenme” olarak kullanılmaktadır (Olson ve Araz, 2023, s. 53).

Birliktelik kuralları, birlikte sık sık meydana gelen nesnelere kombinasyonlarını (ilgi analizi) tanımlamaya çalışır (Olson ve Araz, 2023, s. 36). Bir başka ifade ile daha fazla veri nesnesinin birbiriyle olan ilişkisini gösterir. Örneğin, bir mağazada satılan ürünlerin satış verilerini inceleyerek, bir ürünün satışının diğer ürünlerin satışı üzerinde bir etkisi olup olmadığını görmek için birliktelik analizi yapılabilir. Veri madenciliğinde birliktelik, veriler arasındaki ilişkileri ortaya çıkarmak ve bu ilişkileri kullanarak, karar verme süreçlerini desteklemek için kullanılır.

Veri Madenciliği Taksonomisi

Veri madenciliğinin çeşitli amaçlar ve hedefler için kullanılmasının pek çok farklı yöntemi vardır. Şekil 4’te verilen taksonomi, metodların çeşitliliğini, karşılıklı ilişkilerini ve gruplamasını daha rahat görebilmemiz için yardımcı olabilecektir. Veri madenciliği iki ana başlık altında incelenebilir. Birincisi doğrulama amaçlı (verification), ikincisi ise keşif amaçlı olarak kullanımınıdır. Doğrulama amaçlı kullanımında sistem kullanıcı hipotezini doğrulamaktadır, keşif amaçlı kullanımında ise ilgili sistem yeni kuralları ve örüntüleri bağımsız olarak bulmaktadır (Oded ve Rokach, 2010, s. 5).



Şekil 4. Veri Madenciliği Taksonomisi (Oded ve Rokach, 2010, s. 6)

Keşif metotları veri içindeki örüntüleri otomatik olarak saptamaktadır. Keşif metotları, kestirim metodu ve betimleme metodu olarak ikiye ayrılmıştır. Betimleme metodu veri yorumlamayı amaçlamaktadır. Bunu yaparken temel verinin, ilgili bölüm ile olan ilişki yolunu anlamaya odaklanmaktadır (görselleştirme de olduğu gibi). Kestirim amaçlı yöntemde ise otomatik olarak bir model oluşturulmaktadır ve bu model yeni ve gizli örnekleri bulmakta, ilgili örnekle ilgili bir ya da daha fazla değişkenin değerleri ile ilgili bir kestirimde bulunulmasını sağlamaktadır. Ayrıca keşfedilmiş bilgiyi oluşturan örüntüleri oluşturmaktadır. Bazı kestirim amaçlı yöntemler verinin anlaşılmasını sağlamaktadır. Kestirim amaçlı veri madenciliği yöntemlerinin geneli tümevarımsal öğrenmeye dayanmaktadır. Tümevarımsal yaklaşımın temel sanısı, eğitilmiş modelin gelecekteki gizli örneklere uygulanabilmesidir (Oded ve Rokach, 2010, s. 5).

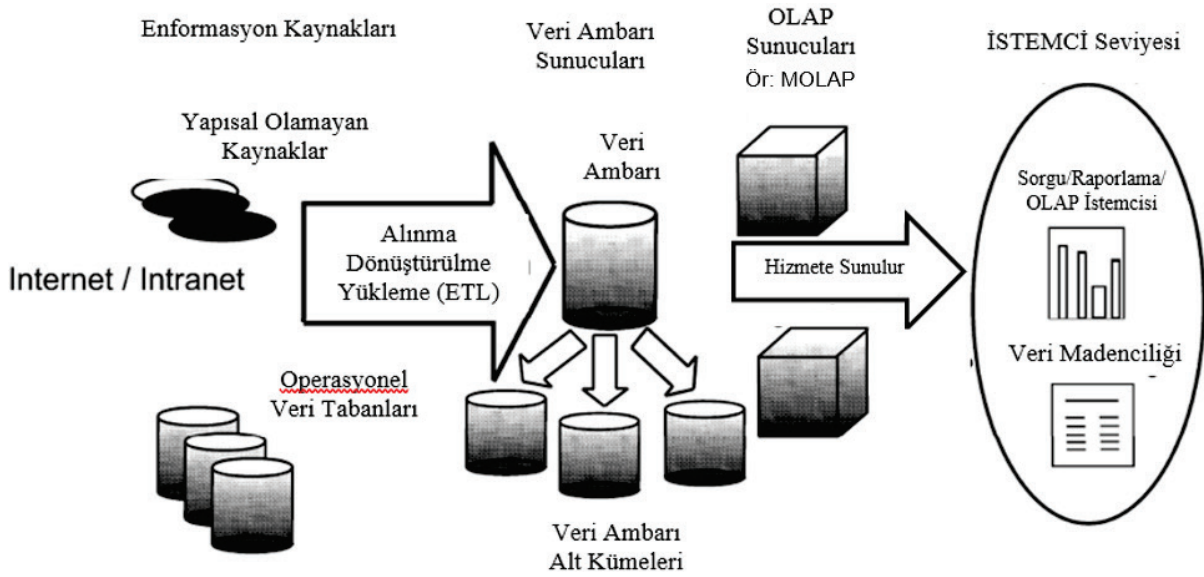
Doğrulama amaçlı yöntem, yabancı kaynak tarafından önerilen hipotezin değerlendirilmesi ile ilgilenmektedir. Bu yöntem uyum testi, hipotez testleri (ortalamaya ilişkin t testi gibi), varyans analizi (ANOVA gibi) istatistik alanındaki genel yöntemleri içermektedir. Bu yöntemler, veri madenciliği ile daha az ilgilidir. Çünkü pek çok veri madenciliği problemi, bilinen bir hipotezin test edilmesi yerine hipotezin keşfedilmesi ile ilgilenmektedir (Oded ve Rokach, 2010, s. 5-6).

Oded ve Rokach (2010, s. 5-6) bu taksonomide literatürde genellikle istatistik alanında gösterilen yöntemleri, veri madenciliğinde doğrulama yöntemleri arasında göstererek veri madenciliği kavramına daha geniş bir anlam katmıştır. Pek çok yazar ise Oded ve Rokach'a göre veri madenciliği kavramını yalnızca veri içinde örüntü keşfetme ve kestirimde bulunma yöntemleri olarak ve bu kavramı daha dar bir kalıpla vermişlerdir. Veri madenciliğinin kapsamı konusu ise burada belirtildiği gibi apayrı bir tartışma konusu olarak görülmektedir.

Veri Madenciliğinin Karar Destek Sistemleri İçindeki Yeri

Karar Destek Sistemleri (KDS), bir karar aşamasında, toplanmış bilgilerden faydalanarak karar vermeyi kolaylaştıran sistemlerdir. Günümüzde karar destek sistemleri bilgisayar tabanlı, insan tabanlı veya ikisinin karışımı şeklinde tasarlanmaktadır. KDS genel olarak bir karar uzmanının (stratejist, üst düzey yönetici, CEO, vs.) kararını daha rahat verebilmesi için, saf veri kaynaklarından (veri tabanları, metin kaynakları, bilgi depoları vs.) veriyi alıp, işleyip, karar için gereken hale getirerek işe başlar. Ardından bu veri üzerinde, daha önceden hazırlanmış olan şablonlar, veri işleme yöntemleri, veri madenciliği çalışmaları vs. kullanarak veri üzerinde ilkel bazı kararlar verebilir veya karar verilmesine destek olabilecek olası senaryoları hazırlayabilir (Şeker, MISSözlük).

Veri madenciliği yöntemleri entegre enformasyon teknolojilerinin (information technology) bir parçası olmaya başlamıştır. Şekil 5, enformasyon teknolojilerinin karar destek yönünün 3 aşamasını göstermektedir (Oded ve Rokach, 2010, s. 7).



Şekil 5. Enformasyon teknolojileri karar destek aşamaları (Oded ve Rokach, 2010, s. 7)

Veri kaynağından başlarken ilk aşama veri ambarıdır ve bunu OLAP (Online Analytical Processing) (Çevrimiçi Analitik İşleme Süreci) sunucu takip eder ve analiz araçlarını toplarlar. İncelendiğinde bunların öncüsünün veri madenciliği araçları olduğu görülmektedir (Oded ve Rokach, 2010, s. 7). Veri Ambarı verilerini analiz etmekte kullanılan iş zekâsı sistemlerinin temel bileşeni Çevrimiçi Analitik İşleme Sürecidir. OLAP karar süreçlerini destekleyen bilgiler veren, çok boyutlu verileri analiz etmeyi sağlamaktadır. OLAP için yapılandırılan veriler genellikle çok boyutlu bir veri modeli kullanmaktadır. Ürün satış bölgesi ve zaman gibi veri öznitelikleri, kullanıcıların karmaşık analiz raporları gerçekleştirmesine olanak tanıyan ayrı boyutlar olarak değerlendirilmektedir. Yapılan işlemleri depolamak ve analiz etmek için kullanılan geleneksel veri tabanlarının aksine, Veri Ambarı, Çevrimiçi Hareket İşleme (online transaction processing - OLTP) kullanılarak analiz edilen geçmiş ve özet bilgilerdir. OLAP, kuruluşların verilere daha iyi bir bakış açısı kazanmasını sağlayan iş problemlerini modelleme becerisi sağlamaktadır (Patil ve Gangadhar, 2019, s. 119).

Şekil 5'te de verilen 3 aşama kolayca şu şekilde ayrılabilir. Aşamalardan ilki rapor üreticilerine ulaşmaktır (örnek olarak, araba hırsızlığı gibi sorunların ortaya çıktığı anda sunulması). Ardından çok düzeyli (multi-level) analize başlanır (örnek olarak on şehirde son ayda ve bir önceki ayda ki araba hırsızlıklarının artışının karşılaştırılması). Son olarak örüntülerin keşfedilmesini içeren karmaşık analizler tamamlanır. Bu süreç bu şehirlerdeki araba hırsızlıklarını tahmin edecek ve hırsız önleyici sistemlerin yerleştirilmesini sağlayacak olan süreçtir. Son süreç olgunun matematiksel olarak modellenmesine dayanmaktadır. İlk iki süreç ise veri toplamanın ve hızlı manipülasyonunun bir aracıdır (Oded ve Rokach, 2010, s. 7-8).

Veri madenciliği, daha önce tanımlanmamış bilgilerin büyük veri tabanlarından çıkarılmasını amaçlamaktadır. Gizli kalıpları keşfetmek ve veriden bilgi çıkarmak için algoritmaların otomatik bir uygulaması olarak görülebilir. Öte yandan, Çevrimiçi Analitik İşleme (OLAP) Sistemleri, büyük veri kümelerini keşfetmek ve etkileşimli biçimde sorgulamaya izin vermektedir. OLAP Sistemleri veri ambarı ortamlarında kullanılan önde gelen ön uç araçlardır. Geçmişteki pek çok yayın, OLAP ve veri madenciliğinin bütünleşmesini vurgulamıştır. Daha yakın zamanlarda ise OLAP ile birlikte veri madenciliği teknikleri, büyük veri setlerini etkili bir şekilde analiz etmek için karar destek uygulamalarında uygulanmıştır (Usman vd., 2010).

VERİ MADENCİLİĞİNİN KÜTÜPHANELERDE KULLANIMI

Veri madenciliğinin kütüphanelerde kullanımına geçmeden önce Bilgi Yönetimi, Veri Madenciliği, Büyük Veri ilişkisi konusuna değinmek gerekir.

Bilgi yönetimi, bilgiyi tanımlama, saklama ve erişim becerisine atıfta bulunan kapsayıcı bir terimdir. Tanımlama, ihtiyaç duyulan verileri toplamak, tanımlamak ve kurumun yaptığı işte etkili kararlar alabilmek için mevcut verileri analiz etmeyi gerektiren aşamadır. Verilerin depolanması ve erişimi ise bilgisayar bilimi tarafından geliştirilen birçok araç kullanılarak veri tabanı yönetimini içermektedir. Böylelikle bilgi yönetimi, organizasyon için hangi bilgilerin önemli olduğunu anlama, organizasyonel karar verme için önemli sistemleri anlama, veri tabanı yönetimi ve veri madenciliğinin analitik araçlarını içermektedir (Olson D., 2017, s. 1).

Olson (2017, s. 1), bulunduğumuz dönemi, büyük veri çağı olarak değerlendirmektedir. Davenport (Davenport 2014'ten aktaran Olson D. , 2017, s. 1) büyük veriyi aşağıda ki gibi tanımlamıştır:

- Veri tek bir sunucuya sığmayacak kadar büyüktür (günümüzde farklı sunucular kullanılmaktadır),
- Satırlar ve sütunlar halinde bir veri tabanına aktarılacak şekilde yapılandırılmamıştır,
- Bir veri ambarına uyabilecek şekilde sürekli akmaktadır,
- Yetersiz yapının özelliklerine sahiptir. Bir başka deyişle yapılandırılmamış durumdadır.

Bilgi yönetimi, kurumlardaki bilgi kaynaklarını tanımlayarak ve yöneterek büyük verilerle başa çıkmak zorundadır. Bilgi yönetimi, süreç odaklı olup, bilginin nasıl edinilebileceğini, karar vermede yardımcı olacak araçlar açısından değerlendirir. Büyük verinin amacı analiz etmek ve veriyi içgörüyeye (kurumun neyi, neden, nasıl yaptığını anlayabilme), inovasyona ve iş değerine dönüştürebilmektir. Gerçek zamanlı performans ölçümleri, verinin tamamına dayanan, daha gerçek zamanlı analizler ve daha sağlıklı kararlar ile kurum ve kuruluşlara değer katabilir (Manyika et al. 2011'dan aktaran Olson D., 2017, s. 1). Bu noktada örneklem kümeleriyle değil, evrenin tamamına yakını ile çalışmak önem kazanmaktadır. Ancak verinin sürekli aktığı düşünülüğünde hiçbir zaman bütün evren yakalanamayacaktır. Bunun yanında son verilerle çalışabilmek de önemlidir.

Özellikle son 20 yılda çevrimiçi araştırma materyallerinde ve birincil kaynak miktarında büyük bir artış yaşanmıştır. Bununla birlikte, bilginin bulunabilirliğindeki gelişmeler ve zorluklar doğal olarak sadece bir dizi araştırmacı için değil, aynı zamanda kütüphaneler için de sorunlara yol açmaktadır. Verideki artış ile birlikte büyük veriden en yüksek değeri nasıl alınacağı konusu önem kazanmıştır. Veri madenciliği ya da çoğu zaman veri madenciliği tarafından atıfta bulunan metin analizi, büyük veri ile yakından ilişkili olan ve gün geçtikçe önemi daha da artan kavramlardır. Veri madenciliği ile bilgisayar yazılımları aracılığıyla büyük veri içinde örüntülerin, eğilimlerin veya gerekli enformasyonun çıkarılması sağlanabilir (Porritt, 2015, s. 18) . Porritt bu ifadeyle veri madenciliğinin, büyük veri ile ilişkili bir kavram olduğunu göstermek istemiştir.

Gelecekte yayınların ve özellikle de akademik yayınların üreticilerinin, veri ve metin madenciliği kullanarak uygun verilerin ulaşılabilir olması için bilgi ve iletişim profesyonelleri ile birlikte çalışmaları gerekmektedir. Kütüphaneler ve araştırmacılar, büyük veriye erişme sorumluluğunun bilincini artırarak sürecin pratik unsurlarına güven duyulması konusunda yardımcı olabilirler. Veri madenciliğinin ve alt konusu olarak metin analizinin, büyük veride bulunan araştırma değerini serbest bırakmanın ve birlikte çalışarak geliştirilebilmenin temel faktörü olduğu göz ardı edilmemelidir (Porritt, 2015, s. 18).

Veri madenciliği pek çok alanda olduğu gibi kütüphane işlem ve hizmetlerini geliştirmek içinde kullanılabilir.

Kütüphanelerde veri madenciliğinin kullanımı yönelik çeşitli örnekler verilebilir. Bunlardan bazıları şöyledir:

- Derme geliştirme: Kütüphaneler dermelerini geliştirmek için kullanıcıların demografik bilgilerini ve kütüphane kullanımı ile ilgili diğer bilgilerini, ödünç alma verileri ile birlikte kullanarak veri madenciliği ile yeni kaynakların alınması için bir model oluşturabilirler ve kullanıcılarının ilgi alanları ile ilgili kaynakların kestiriminde bulunabilirler.
- Kullanıcı analizi: Kullanıcıların kütüphane kullanımına ilişkin verileri incelenerek veri madenciliği ile ihtiyaç ve tercihleri daha iyi anlaşılabilir ve kütüphaneler hizmetlerini elde edilen sonuçlara göre iyileştirebilir.
- Öneri sistemleri: Kütüphaneler, veri madenciliği algoritmalarını kullanarak, geçmiş kullanım verilerine dayalı olarak kullanıcılar için kişiselleştirilmiş öneriler sunabilir.

- Metin analizi: Kütüphaneler, veri madenciliğinin bir alt dalı olan metin madenciliği tekniklerini kullanarak, koleksiyonlarındaki materyallerin içeriğini analiz edebilir, içeriğin içindeki örüntülerini ve eğilimleri ortaya çıkarabilir.
- Bilgi hizmetlerinin pazarlaması: Kütüphaneler kullanıcı verilerini analiz ederek, hizmetlerini ve kaynaklarını tanıtmak için kişiye özel tanıtım faaliyetleri oluşturabilir.
- Teknik hizmetler: Kütüphane dermesinin önceki kayıtlarından geliştirilen model ile kütüphane kaynaklarının bibliyografik kaydı sırasında kütüphaneciye öneri sistemi oluşturulabilir.

Görüldüğü gibi teorik olarak kütüphanelerde pek çok alanda veri madenciliğini kullanmak mümkündür. Konu ile ilgili olarak literatür tarandığında ise bugüne kadar yapılmış çalışmalarla ilgili aşağıdaki örnekler verilebilir.

Girija ve Srivatsa, veri madenciliğinin web sayfalarında kullanılmasını web madenciliği, multimedya veri tabanlarında kullanılmasını multimedya madenciliği, metinler için kullanılmasını metin madenciliği, bu çalışmanın konusu ile ilgili olarak dijital kütüphanelerde uygulanmasını ise bibliyomadencilik olarak tanımlamıştır (Girija ve Srivatsa, 2006, s. 593). Fakat literatür incelendiğinde bu çalışmada da görülebileceği üzere kütüphanelerde veri madenciliği kullanımı çoğunlukla bibliyomadencilik olarak geçmemekte, bibliyomadencilik bu konuda özel bir alan olarak karşımıza çıkmaktadır.

Konu ile ilgili ülkemizde üç farklı lisansüstü tezine ulaşılmıştır.

Doğan'ın, doktora tezi çalışmasında kütüphanelerde veri madenciliği modellerinden sınıflandırma ve birliktelik kuralı yöntemlerinin kullanılabilirliğini göstermek için simülasyon tekniği ile iki farklı hipotetik veri kümesi oluşturulmuştur. Bu veri kümeleri ile ilk olarak kütüphane kullanıcılarının veri tabanı kullanımına ilişkin birliktelik kuralı oluşturulmuştur. İkinci olarak ise kullanıcılara ilişkin oluşturulan hipotetik verilere dayanarak veri tabanı öneri modeli kurulmuş ve model ile kullanıcılara ilişkin model kestiriminde bulunulabileceği gösterilmiştir (Doğan, 2022).

Kurt, doktora tezi çalışmasında veri madenciliğini iki farklı şekilde kullanmıştır. Araştırmanın ilk analizinde, yayınlanan içeriklerin okunma sayılarını etkileyen faktörleri ortaya çıkarmak amacıyla, içerik metinlerini metin madenciliği yaklaşımıyla çözümlenmiş ve çoklu doğrusal regresyon analizi uygulanarak okunma sayılarını

etkileyen faktörler ortaya çıkarılmıştır. İkinci analizde ise, okunma sayılarının ilgili organizasyonun gelirleri üzerindeki etkisi araştırılmıştır. Bu amaçla, aylık bazda şirket gelirleri ve okunma sayıları, basit doğrusal regresyon problemi olarak ele alınarak bir tahmin modeli oluşturulmuştur (Kurt, 2023).

Akçay yapmış olduğu yüksek lisans tez çalışmasında bilgi keşfi sürecini, veri madenciliği tanımını, veri madenciliği uygulama sürecini, veri madenciliği kullanım alanlarını, veri madenciliği yöntem ve tekniklerini incelemiştir (Akçay, 2014). Akçay'ın çalışmasında Bilgi ve Belge Yönetimi'nde veri madenciliği konusu da kavramsal olarak işlenmiştir (Akçay, 2014, s. 57-60).

Nicholson (2006) kütüphanelerde veri madenciliği tekniklerinin uygulanabilmesi için gerekli olan verilerin elde edilmesinde kullanılacak veri ambarındaki veri türleri için Şekil 6'da bulunan kavramsal çerçeveyi önermiştir.

Bibliyografik Vekil		Bibliyometrik Veri (Kümelenmiş Elektronik Kaynak Verisi Gereklidir) Atıflar/Linkler/Sosyal Ağlar vs. Disiplinler Bağlantılar
Kütüphane İşlemleri Personelle İlgili Bilgi Talebi Tedarikçi Planı Paketleme Giderleri Geçen Süre Personel Düzenleri	MARC, Dublin Core, Yazar, Yayıncı, Konu Başlıkları/Sınıflama, Özet, Anahtar Kelimeler Materyal Türü, Yer(Fiziksel ve Sanal), Erişilebilirlik	
Kütüphane Hizmetleri Temel Öğeler: Tarih, Zaman, Yer, Format, Müşterek Kullanım		
Tarama İçerik Tarama Yolu (Önce/Sonra) Tarama Netliği Materyallerin Müşterek Kullanımı	Ödünç Verme / Kullanım OPAC Bilgisi (Dolaşım Uzunluğu) Hareket Günlüğü Analizi Tıklama Linki / Vekil Bilgisi Sayaç Verisi	
Danışma Metot Soru Soru Konusu Geçen Süre Cevap Başvuru Yolu	Destek ve Eğitim Program Türü Program Başlığı Program Konusu Katılımcıların Demografisi Başvuran Bilgisi	ILL / Diğer Talepler Taleplerin Akışı Son Tedarikçi Zamanın İşlenmesi Tedarik Zamanı Maliyetler
Demografik Vekil Ayarlara Bağlıdır: Üniversite / Bölüm / Düzey / Posta Kodu / Grup Üyeliği / İlgili Bölüm Alanları / Kademe / Öğretmen		

Şekil 6. Nicholson'ın (2006) kütüphanelerde veri madenciliği tekniklerinin uygulanabilmesi için kullanılabilecek veri türlerine ilişkin önerdiği kavramsal çerçeve

Weng makalesinde üniversite öğrencilerinin kitap abonelik (book subscription) süreçlerini anlamayı ve kütüphane yöneticilerine kütüphanedeki kitaplar için etkin kütüphane yönetim planları yapmalarını sağlamayı amaçlamaktadır. Tek bir veri kümesinden örüntüler bulmaya yarayan geleneksel birliktelik kuralı madenciliği (association rule mining) yöntemlerinden farklı olarak, bu makale, kütüphane kullanıcılarının kitap abonelik özelliklerini analiz etmek için “yenilik-frekans-üniversite” (RFC) modelini önermektedir. Çalışmada ayrıca eşdeğer sınıf RFC bölümlerinden (equivalence-class RFC segments) ilişki kuralları keşfedilmiştir (Weng, 2016).

Kovacevic vd. ilgili çalışmalarında veri madenciliği tekniklerinin (kümeleme ve sınıflandırma) dijital kütüphane hizmetlerinde kullanılmasını öneren bir yaklaşımda bulunmuşlardır. Çalışmada benzer kullanıcılar, profillerine ve arama davranışlarına dayanarak birlikte kümelenmişlerdir. Kullanıcılara uygun hizmetleri tavsiye etmek için öngörülü (predictive) sınıflandırma kullanılmıştır. Çalışmanın sonucunda aynı kümedeki kullanıcıların benzer hizmetleri veya kalıpları kabul etme ihtimalinin yüksek olduğu gösterilmiştir (Kovacevic vd., 2010, s. 829). Bu çalışma Goodall ve Pattern'in (2011) ve Renaud vd. lerinin (2015) ilgili yayınlarında da atıf almıştır.

Preza (2016, s. 2-3) "Kütüphanelerde Veri Bilimi ve Analitik" isimli çalışmasında, veri biliminin kütüphanelere yardımcı olabileceği somut alanları aşağıdaki gibi tanımlamıştır:

- Dijital nesne (object) sınıflaması/semantik/arama,
- Resim tanıma ve sınıflandırma,
- İçerik kümeleme ve segmentasyon,
- Raporlama,
- Kestirim analitikleri (predictive analytics),
- Makine çevirisi,
- Sesten metine dönüştürme,
- Metinden sese dönüştürme,
- İntihal,
- Kurumsal kaynaklar için analitik platform.

Preza (2016, s. 4-5) kütüphanelerde hangi enformasyonun analiz edilebileceği ile ilgili aşağıdaki maddeleri sıralamıştır.

1. Aramalar: Kullanıcıların veri tabanında ne aradıklarını, ihtiyaçlarının ne olduğunu ve en çok neleri aradıklarını bilmek önemlidir. Kullanıcılar aramaları ile uyumsuz içerik arıyorlar mı?

2. Bant genişliği: Kütüphane otomasyon sisteminin hangi dönemlerde daha sık kullanıldığı bilmek daha hızlı bir erişim ve kullanıcı memnuniyeti için önemlidir.

3. Depolama: Depolama ve bant genişliği verileri, Depo Yöneticisi ve Kurum'a ek bütçeleri doğrulama, büyüme ve kullanım planlama, maliyetleri tahmin etmede yardımcı olabilir.

4. Kullanıcı ve kullanım: Kullanıcılar veri havuzunda ne yapıyorlar ve kim neyin sahibi?

5. Trafik: Tüm web trafiğini günlüklere kaydetmek önemlidir. İzlenecek yollar arasında ziyaretler, ziyaretlerin süresi, yönlendiren URL'ler, etkinlikler (yükleme, indirme, vb.), tarayıcı vb.

6. Dijital nesnelere: veri havuzunun içinde neler var?

7. Sınıflama: Genellikle, dijital nesnelere bir veri havuzuna yüklendiğinde, sistem nesne için bir "etiket" veya "sınıflandırma" istemektedir. Bir nesne birden fazla etiket veya sınıflandırma içerebilir.

8. Ses, video, metin ve resim analizi: Dijital nesnelere içinde neler var? Bu görev, bilişsel hizmetler kullanılarak otomatik (cognitive services) olarak yapılabilir.

9. En iyi 10: İlk 10'daki en çok yüklenen dosyalar, ilk 10'daki en çok indirilen dosyalar, ilk on yükleyici, ilk on arama, vb. fikir sahibi olunabilir.

10. Diğer: Hiç indirilmemiş veya görülmemiş dosyalar, hiç oturum açmamış kullanıcılar, vb.

Cullen, 2005 yılında yayınladığı makalesinde işletmelerin operasyonlarında yıllarca veri madenciliğini kullandığından bahsetmiş; kütüphanelerin ve kütüphanelerle ilgili işletmelerin de bu doğrultuda veri madenciliğini ne şekilde kullandıklarına değinmiştir (Cullen, 2005).

Gibson, yayınında incelediği tıp kütüphanesinin çevrimiçi materyallerinin kullanımında belirli kullanım kalıplarının olup olmadığını araştırmak için teorik ve pratik olarak uygun yöntemler bulmayı amaçlanmıştır. Sunucu erişim kütük (log) dosyalarında keşfedilebilecek kullanıcı örüntüleri yardımı ile bilinen ve istatistiksel olarak önemi olan kullanıcı arama davranışlarına yönelik çıkarımların türetilmesinin mümkün olup olmadığı sorusunu incelemiştir. Bu çalışmayı web kullanım madenciliği (web usage mining) konusu altında yapmıştır (Gibson, 2001, s. 15-16).

Noh ve Kim, toplumun genel olarak halk kütüphaneleri hakkındaki algısını anlamak amacıyla, halk kütüphaneleriyle ilgili belirli blog ve web sitelerinde yer alan metinleri analiz ederek, semantik bağlantı ağı ve duygu analizi yöntemlerini kullanmıştır (Noh ve Kim, 2022, s. 41).

Yukarıda literatürde karşılaşılan örnekleri vermenin yanı sıra, yine literatürde kütüphanelerin ve diğer bilgi merkezlerinin belirli hizmet ve işlemlerine yönelik örneklerle de karşılaşılması mümkündür. Bunlar aşağıda kısaca örneklenmeye çalışılmıştır.

Guzman, veri madenciliğinin kütüphanelerde uygulanmasına ilişkin kapsamlı bir literatür taraması yapmıştır. Çalışmasında 1998 ile 2014 yılları arasında altı farklı dergide konu ile ilgili yapılan çalışmaları incelemiştir (Guzman vd., 2015, s. 499).

Brettschneider, Alman telif hakkı hakları yasasında metin ve veri madenciliği konusunu ele almıştır. Bu çalışmada hem kütüphane hem de hukuk açısından ortaya çıkan engellere değinilmektedir (Brettschneider, 2021, s. 104). Benzer şekilde Papadopoulos'ta konunun hukuki boyutu ile ilgilenmektedir. Çalışmasında Yunanistan Milli Kütüphanesi tarafından uygulanan metin ve veri madenciliği uygulamalarını ve genel veri koruma yönetmeliği açısından gerekliliklerini dikkate alan uygulamalı internet güvenliği çözümlerini ele almaktadır (Papadopoulos vd., 2020, s. 441).

Katsurai, çalışmasında veri madenciliği ile ilgili oluşturduğu sözlükte yer alan terimleri, kütüphane ve bilgi bilim alanında tarayarak bunların metin analizlerini yapmıştır (Katsurai ve Joo, 2021).

Khademizadeh, çalışmasında üniversite kütüphanecileri ve veri bilimcilerinin işbirliği önemine, veri madenciliğinin üniversite kütüphanelerinde kaynak seçimi ve dolaşım sürecini etkileyebileceğine, veri madenciliği tekniklerinin özellikle derme geliştirme alanında üniversite kütüphanelerinde kullanılabilmesine, birliktelik kuralları algoritmalarının kullanımının kütüphane kaynaklarının kullanıcılar için tahmin edilmesine yardımcı olabileceğine değinmiştir (Khademizadeh vd., 2022, s. 1).

Chan-Chine Chang, Ruey-Shun Chen ise inceledikleri kütüphanenin ödünç verme verilerini kullanarak kütüphane kullanıcılarını kümeleme tekniği ile her bir kümenin kendi özelliklerine sahip olduğu beş farklı kümeye ayırmışlardır (Chang ve Chen, 2006, s. 307).

Kao vd. çalışmalarında veri madenciliği yöntemi ile derme geliştirme bütçesi tahsis modelini önermişlerdir. Bu modeli oluşturmak için kütüphane ödünç verme verilerini kullanmışlardır (Kao vd., 2003, s. 133).

Sun vd., veri madenciliği kullanarak üniversite kütüphaneleri kullanıcıları için kitap ödünç alma konusunda kişiselleştirilmiş hizmet stratejisi ortaya koymuşlardır (Sun vd., 2020, s. 515).

Irfiani, çalışmasında belirtilen halk kütüphanesinde 17-25 yaş aralığındaki okuyucuların kitap türlerine göre ödünç alma davranışlarına ilişkin örüntülerini veri madenciliğinde birliktelik kurallarını kullanarak saptamıştır (Irfiani, 2020, s. 137).

Duan ve Wang, veri madenciliği kullanarak üniversite kütüphanelerinde kişiselleştirilmiş hizmetlerin nasıl verileceğine yönelik bir çalışma yapmışlardır. Bu amaçla okuyucu özellikleri ile ödünç alınan kitaplar arasında bir ilişki modeli kurulabilmesi için birliktelik kuralı madenciliği kullanılmıştır (Duan ve Wang, 2021, s. 1-5).

Xu, çalışmasında öncelikle bilgi merkezinde öğretmen olarak kayıtlı olan kullanıcıları kümeleme yöntemi ile gruplara ayırmıştır (Xu, 2011, s. 35), sonrasında öğretmenlerin kişisel özelliklerinin, web kaynakları kullanımlarını etkileme durumlarını araştırmıştır (Xu, 2011, s. 44).

Literatürde pek çok kaynakta kütüphane kullanımı ile akademik başarı arasındaki ilişkinin ölçülmesi konusu işlenmiş ve bu bölümde bu kaynaklardan bazılarına yer verilmiştir.

Bu konuda yapılan çalışmalardan birinde Cox ve Jantti kütüphane kullanımı ile başarı notları arasında güçlü bir korelasyon bulmuştur (Cox ve Jantti, 2012).

Soria, Fransen ve Nackreud, ilgili çalışmasında 13 üniversitede birinci sınıflar arasındaki kütüphane kullanımının akademik başarılarını pozitif yönde etkilediği sonucuna varmıştır (Soria, Fransen ve Nackerud, 2013).

Goodal ve Pattern'de akademik başarı ile kütüphane kullanımını ilişkilendirmek için veri madenciliği tekniklerini kullanmıştır. Bu çalışmada kütüphanenin öğrenci tarafından ziyaret verileri, kitap ödünç verme verileri ve elektronik kaynak kullanımı verileri vb. veriler kullanılmıştır. (Goodall ve Pattern, 2011).

Renaud, Britton, Dinding ve Mitsouri, ilgili çalışmalarında, kütüphanelerde veri madenciliği tekniklerini kullanarak kullanıcı davranış analizi, kitap kullanım analizi, kütüphane kullanımı ve öğrenci başarısı arasındaki korelasyon gibi konulara eğilmişler ve bu konularda ilgili örüntüleri bularak kütüphane performansını ölçmek için yeni bir yöntem denemişlerdir. Çalışmada kullanıcı ve kitap kullanım analizi bölümlerinde ödünç alınan kitapların LC sınıflandırma kodlarıyla ve akademik birimlerle ilişkileri ortaya konmuştur (Renaud, Britton, Dingding ve Mitsunori, 2015).

Ayrıca Shieh, bibliyomadencilik sürecinin, konuların tanımlanması, veri ambarının oluşturulması, verinin arıtılması, verinin keşfedilmesi ve sonuçların değerlendirilmesi gibi aşamalardan oluştuğunu belirtmiştir. İlgili çalışmada

kütüphaneciler için daha kolay erişilebilir bir bibliyomadencilik uygulama modeli ve bibliyomadencilik uygulamalarında yardımcı olması için bir prototip sistem uygulaması önerilmiştir (Shieh, 2010).

İncelenen bu sonuçlara dayanarak kütüphanelerde pek çok farklı alanda veri madenciliği uygulamasının yapılabileceği sonucuna varılmıştır.

Sonuç ve Öneriler

Bilgi ve iletişim teknolojilerinde yaşanan gelişmeler ve bu teknolojilerin yaygınlaşması ile birlikte üretilen verinin aşırı derecede artması ile birlikte büyük bir veri patlaması olmaktadır. Kurum ve kuruluşlar ise kullandıkları operasyonel veri tabanları dışında üretilen bu verileri kendileri için katma değer yaratacak bir şekilde kullanmak istemektedir. Kütüphanelerin de bu dönüşümün dışında kalması söz konusu değildir.

Kütüphaneciler taksonomi alanına, kullandıkları sınıflandırma sistemi nedeniyle alışık oldukları için, edinebilecekleri istatistiksel bir sınıflandırma modeliyle kütüphane verilerini işleyebilirler. Kütüphaneciler için veri madenciliği programları satın alıp, kurabilme ve kullanabilme bilgisi, verilecek küçük çaplı eğitimlerle elde edilebilir. Bunun yanında ileri düzeyde veri madenciliği uygulamaları ve analiz için disiplinler arası bir çalışma gerekecektir. Bu nedenle kütüphaneciler daha ileri düzeyde karşılaşabilecekleri problemlerle ve uygulamalarla ilgili olarak istatistikçiler ve bilgisayar programcıları ile birlikte ortak çalışmalıdırlar.

Araştırma sonucunda; Kütüphanecilerin veri madenciliği tekniklerini kullanarak, kütüphane işlem ve hizmetlerinde yeni örüntüler elde ederek bunları karar destek süreçlerinde kullanabilmesinin uygun veri kümelerinde mümkün olabileceği sonucuna varılmış ve araştırma hipotezi doğrulanmıştır.

Bu çalışma aynı zamanda kütüphanelerin daha iyi bir kullanıcı deneyimi sunabilmesi için kullanıcı verilerinin ve davranışlarının kütüphanelerde toplanması ve tutulması gereğini ortaya koymaktadır. Bunun ötesinde kullanıcıların tarama ve benzeri davranışlarını kullanıcı adları ve şifreleri ile giriş yaptıktan sonra gerçekleştirmesinin zorunlu kılınması değerlendirilebilir.

Araştırma kapsamında elde edilen bu sonuçlara göre aşağıdaki öneriler sıralanabilir:

- Kütüphanecilerin daha fazla değer yaratabilmeleri için kullanıcıları ile özel ve kişisel deneyimler yaratmaları gerekmektedir.

• Kütüphane hizmetlerinin ve kullanıcı deneyiminin geliştirilmesinde veri madenciliği yöntemleri kullanılabilir.

• Kütüphanecilerin kullanıcıları ile anlamlı bir ilişki kurmaları için hangi verilere ihtiyaçları olduğu saptanmalıdır.

• Kütüphane kullanıcılarının çevrimiçi kütüphane kullanımlarında kullanıcı adı ve şifre kullanması sağlanmalıdır. Bu şekilde hukuki ve etik kurallar çerçevesinde kullanıcı hareketlerinden gelecek veriler toplanarak bir veri tabanında veya veri ambarında daha sonra veri madenciliği süreçlerinde kullanılmak üzere yapılandırılması gereklidir.

• Bu hizmetlerin ve ilişkinin geliştirilmesinde hangi veri tabalarından faydalanabileceği, bu veri tabanları dışında hangi tür verilerin toplanabileceği araştırılmalıdır.

• Toplanan bu verilerin saklanması ve analitik işlemlere tabi tutulmasında kullanılacak donanım ve yazılımlar belirlenmelidir.

• Kütüphaneciler yaptıkları bu tarzda çalışmalarını diğer kütüphanecilerle paylaşmalı ve onlara yol göstermelidir.

• Kütüphaneciler veri madenciliğini kendi meslekleri içinde değerlendirmeli, onlara bu konuda gerekli eğitimi almaları için gerekli ortam sağlanmalı, kütüphanecilik alanında eğitim veren bölümler ise müfredatlarını bu doğrultuda düzenlemelidir.

• Günümüz şartlarında araştırmacıların veri madenciliği uygulamaları yapılabilmesi için yüksek maliyetli donanım ve yazılımlara ihtiyaçları bulunmaktadır. Kütüphaneciler başka bilimlerde de veri madenciliği çalışmalarının yapılabilmesi için öncü olmalı, gerekli makineler, bilgisayar programları ve uzmanları ile kütüphane içinde hizmet vermelidir.

• Bu çalışmada görüldüğü üzere bu tarzda çalışmalar istatistik bilimi, bilgi teknolojileri vb. farklı disiplinlerin ortak çalışması ile gerçekleştirilmelidir.

Kaynakça

Akçay, A. (2014). *Bilgi ve Belge Yönetiminde Veri Madenciliği*. Yayımlanmamış yüksek lisans tezi, İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Akdi, Y. (2018). *İST308 Zaman Serileri Analizi, 8. Hafta Ders Notları*. Ankara Üniversitesi Açık Ders Malzemeleri. Erişim adresi: <https://acikders.ankara.edu.tr/mod/resource/view.php?id=47751>

- Akpınar, H. (2014). *Data Veri Madenciliği Veri Analizi*. İstanbul: Papatya Yayıncılık Eğitim.
- Altunkaynak, B. (2022). *Veri Madenciliği Yöntemleri ve R Uygulamaları* (3. bs.). Ankara: Seçkin Yayıncılık.
- Arslantekin, S. (2003). Veri madenciliği ve bilgi merkezleri. *Türk Kütüphaneciliği*, 17(4), 369-380. <http://www.tk.org.tr/index.php/TK/article/view/303/295>
- Brettschneider, P. (2021). Text und data-mining – juristische Fallstricke und bibliothekarische Handlungsfelder. *Bibliotheksdienst*, 55(2), 104-126.
- Cackett, D. (2013). *Information Management and Big Data, A Reference Architecture*. Redwood Shores: Oracle Corporation. Erişim adresi: <https://citeseerx.ist.psu.edu/pdf/15a6c422db770da6eaecde3fd1630c3eec3880e5>
- Chang, C. C. ve Chen, R. S. (2006). Using data mining technology to solve classification problems: A case study of campus digital library. *The Electronic Library*, 24(3), 307-321.
- Cios, K., Pedrycz, W., Swiniarski, R. ve Kurgan, L. (2007). *Data Mining, A Knowledge Discovery Approach*. New York: Springer.
- Cox, B. ve Jantti, M. (2012). Discovering the impact of library use and student performance. *Educause Review*, s. 1-9. Erişim adresi: <http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1507&context=asdpapers>
- Cullen, K. (2005). Delving into data: Businesses have used data mining for years. Now libraries are getting into the act. *Library Journal*, 30(13), s. 30-32.
- Data analysis. (t.y.) *Cambridge dictionary içinde*. Erişim adresi: <https://dictionary.cambridge.org/dictionary/english/data-analysis>
- Doğan, K. (2022). *Bilgi Merkezi ve Hizmetlerinde Veri Madenciliğinin Kullanılabilirliği: Üniversite Kütüphaneleri Örneği*. Yayımlanmamış Doktora Tezi, Ankara: Ankara Üniversitesi Sosyal Bilimler Enstitüsü.
- Döring, M. (2018). *Prediction vs forecasting*. Data Science Blog. Erişim adresi: https://www.datascienceblog.net/post/machine-learning/forecasting_vs_prediction/#:~:text=Prediction%20is%20concerned%20with%20estimating%20the%20outcomes%20for%20unseen%20data.&text=Forecasting%20is%20a%20sub%2Ddiscipline,we%20consider%20the%20temporal%20

- Duan, S. ve Wang, Z. (2021). Research on the service mode of the university library based on data mining. *Scientific Programming*, 1-9. doi:<https://doi.org/10.1155/2021/5564326>
- Dunham, M. (2002). *Data Mining Introductory and Advanced Topics*. New Jersey: Pearson Education Inc.
- Everitt, B. ve Skrondal, A. (2010). *The Cambridge Dictionary of Statistics* (4. bs.). New York: Cambridge University Press.
- Fayyad, U., Piatetsky-shapiro, G. ve Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 37-54. Erişim adresi: <https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1230/1131>
- Gibson, I. E. (2001). *Data mining analysis of digital library database usage patterns as a tool facilitating efficient user navigation* (Doktora Tezi). Erişim adresi: <https://www.proquest.com/dissertations-theses/data-mining-analysis-digital-library-database/docview/251131313/se-2?accountid=8319>
- Goodall, D. ve Pattern, D. (2011, 02 22). Academic library non/low use and undergraduate student achievement : A preliminary report of research in progress. *Library Management*, 32(3), 159-170. Erişim adresi: <http://www.emeraldinsight.com/doi/pdfplus/10.1108/01435121111112871>
- Girija, N. ve Srivatsa, S. K. (2006). A research study: Using data mining in knowledge base business strategies. *Information Technology Journal*, 5(3), 590-600.
- Guzman, L. S., Saquicela, V., Ordóñez, E. A., Vandewalle, J. ve Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship*, 41(4), 499-510.
- IBM SPSS Modeler CRISP-DM Guide*. (2000). IBM Web Sitesi. Erişim adresi: <https://www.ibm.com/docs/en/spss-modeler/18.1.1?topic=spss-modeler-crisp-dm-guide>
- Irfiani, E. (2020). Determination of book loan association pattern using apriori algorithm in public libraries. *Techno Nusa Mandiri: Journal of Computing and Information Technology*, 17(2),137-142.
- Kao, S. C., Chang, H. C. ve Lin, C. H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing & Management*, 39(1), 133-147.

- Katsurai, M. ve Joo, S. (2021). Adoption of data mining methods in the discipline of library and information science. *Journal of Library and Information Studies*, 19(1), 1-17.
- Khademizadeh, S., Nematollahi, Z. ve Danesh, F. (2022). Analysis of book circulation data and a book recommendation system in academic libraries using data mining techniques. *Library and Information Science Research*, 44(4), 1-9.
- Kovacevic, A., Devedzic, V. ve Pocajt, V. (2010). using data mining to improve digital library services. *The Electronic Library*, 28(6), 829-843. doi: 10.1108/02640471011093525
- Kurt, L. (2023). *Bilgi Yönetiminde Veri ve Metin Madenciliği: Bir Dijital İçerik Analizi Uygulaması*. Yayınlanmamış Doktora Tezi, Ankara: Ankara Üniversitesi Sosyal Bilimler Enstitüsü.
- Nicholson, S. (2006). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. *Information Processing and Management*, 42(01), 785-804. doi:10.1016/j.ipm.2005.05.008
- Nisbet, R., Miner, G. ve Yale, K. (2018). *Handbook of Statistical Analysis and Data Mining Applications* (2. bs.). London: Elsevier.
- Noh, Y. ve Kim, D. (2022). A study on social perceptions of public libraries utilizing the sentiment analysis. *International Journal of Knowledge Content Development & Technology*, 12(4), 41-65.
- Oded, M. ve Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (2. bs.). New York: Springer.
- Oğuzlar, A. (2011). *Temel Metin Madenciliği*. Bursa: Dora.
- Olson, D. (2017). *Descriptive Data Mining*. Singapore: Springer.
- Olson, D. L. (2018). *Data Mining Models, Second Edition*. New York: Business Expert Press.
- Olson D. L. ve Araz Ö. M. (2023). *Data mining and analytics in healthcare management: Applications and tools*. Springer. <https://doi.org/10.1007/978-3-031-28113-6>

- Papadopoulos, M., Gerolimos, M., Vavousis, K. Xenakis, C. (2020). Text and data mining for the national library of Greece in consideration of internet security and GDPR. *Qualitative and Quantitative Methods in Libraries (QQML)*, 9 (3), 441-460.
- Patil, A. ve Gangadhar, N. (2019). Olaas: OLAP as a service. *2016 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Cloud Computing in Emerging Markets (CCEM)* (s. 119-124). Bangalore, India: IEEE. Erişim adresi: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7819682>
- Pektaş, A. O. (2013). *SPSS ile Veri Madenciliği*. İstanbul: Dikeyksen.
- Porritt, G. (2015). Data mining in the humanities and social sciences. *Information Today*, 18.
- Preza, J. L. (2016). *Data Science and Analytics in Libraries*. Erişim adresi: <https://zenodo.org/record/375809/files/Data%20Science%20and%20Analytics%20in%20Libraries.pdf>
- Prytherch, R. (2005). *Harrod's librarians' glossary and reference book : a dictionary of over 10,200 terms* (10 bs.). Hampshire: Ashgate Publishing Limited.
- Quinlan, J. (1986). Induction of Decision Trees. *Machine Learning*, 81-106. Erişim adresi: https://link.springer.com/content/pdf/10.1007%2F978-1-4613-1162-5_1.pdf
- Renaud, J., Britton, S., Dingding, W. ve Mitsunori, O. (2015). Mining library and university data to understand library use patterns. *The Electronic Library*, 33(3), 355-372. <https://search.proquest.com/docview/1683340136?accountid=8319>
- Roiger, R. J. (2017). *Data mining: A tutorial-based primer*. Florida: CRC Press, Taylor & Francis Group.
- Sankur, B. (2004). *İngilizce - Türkçe Ansiklopedik Bilişim Sözlüğü*. İstanbul: Pusula.
- Sathishkumar, S., Devi Priya, R. ve Karthika, K. (2020). Survey on data mining and predictive analytics techniques. G. Ranganathan, J. Chen, ve Á. Rocha (Ed.), *Inventive Communication and Computational Technologies. Lecture Notes in Networks and Systems* (s. 971-981) içinde. Singapore: Springer.
- Shieh, J. C. (2010). The integration system for librarians' bibliomining. *The Electronic Library*, 28(5), 709-721. Erişim adresi: <http://www.emeraldinsight.com/doi/pdfplus/10.1108/02640471011081988>

- Soria, K., Fransen, J., ve Nackerud, S. (2013). Library use and undergraduate student outcomes: New evidence for students' retention and academic success. *Libraries and the Academy*, 13(2), 147-164. Erişim adresi: <http://resolver.ebscohost.com/openurl?sid=EBSCO%3aconedsqd7&genre=article&iissn=15312542&ISBN=&volume=13&issue=2&date=20130101&spage=147&pages=147-164&title=portal%3a+Libraries+and+the+Academy&atitle=Library+Use+and+Undergraduate+Student+Outcomes%3a+New+Ev>
- Stanisz, T., Kwapien, J. ve Drozd, S. (2019). Linguistic data mining with complex networks: A stylometric-oriented approach. *Information Sciences*, 482, 301-320. Erişim adresi: <https://doi.org/10.1016/j.ins.2019.01.040>
- Sun, Y., Sun, J., Du, Q., Zhao, H. ve Liu, J. (2020). Research on personalized service strategy of university library based on big data mining system. *2020 5th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, (s. 515-518). doi:10.1109/ICSGEA51094.2020.00117
- Şeker, Ş. E. (2013). *İş Zekâsı ve Veri Madenciliği*. İstanbul: Cinius Yayınları.
- Şeker, Ş. E. (t.y.). *MISSözlük*. <http://mis.sadievrenseker.com/2014/06/metin-madenciligi-text-mining/>
- Tan P. N., Steinbach M., Karpatne, A. ve Kumar, V. (2019). *Introduction to data mining* (2. bs.). Pearson Education Limited.
- T.C. Başbakanlık Devlet İstatistik Enstitüsü Başkanlığı. (t.y.). *İstatistik Terimleri Sözlüğü*.
- Usman, M., Asghar, S. ve Fong, S. (2010). Data mining and automatic OLAP schema generation. *2010 Fifth International Conference on Digital Information Management (ICDIM)* içinde, (s. 35-43). Thunder Bay. doi:10.1109/ICDIM.2010.5664622
- Weng, C.-H. (2016). Knowledge discovery of digital library subscription by RFC itemsets. *Electronic Library*, 35(5), 772-788. doi:10.1108/EL-06-2015-0086
- Xu, B. (2011). *Understanding teacher users of a digital library service: A clustering approach* (Doktora tezi). <https://www.proquest.com/dissertations-theses/understanding-teacher-users-digital-library/docview/862643655/se-2?accountid=8319>

Yu, C. H. (2022). *Data mining and exploration from traditional statistics to modern data science*. CRC Press. <https://doi.org/10.1201/9781003153658>

Zaki, M. J. ve Meira W. (2020). *Data mining and machine learning : fundamental concepts and algorithms* (2. bs.). Cambridge University Press.

Zweig, K. A. (2016). Are Word-Adjacency Networks Networks?. A. Mehler, A. Lücking, S. Banisch, P. Blanchard ve B. Job (Ed.). *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks. Understanding Complex Systems* içinde (s. 153-163). Berlin: Springer. doi:https://doi.org/10.1007/978-3-662-47238-5_7

Summary

Data mining is one of the most important methods to find patterns and draw meaningful conclusions from big data collected from different sources. It is important for libraries to collect data from different sources and to draw meaningful results from this data by data mining.

The usage areas and benefits of data mining in libraries, which services it can be used for, and which data mining methods can be applied are important topics that need to be researched in Turkey, as well as in the world. Although there are sufficient publications on this topic in foreign literature, the lack of sufficient studies on libraries in our country has led to the need for this study. Data mining for libraries is extremely important in terms of creating reporting opportunities that have not been previously available, providing better user experience for library users, and obtaining new insights in library operations.

The study reveals the current state of data mining with technological developments such as big data technologies. An effort has been made to determine what libraries should do in the light of these developments.

In this context, the main hypothesis in the study is: "It is possible for libraries to obtain new patterns in library operations and services by using data mining techniques and to use them to develop new service models by reflecting them in decision support processes."

A literature review is conducted to establish the theoretical basis of the research. At this stage, nationwide and international studies including the concepts related to data mining, data mining models, data mining processes, etc. are examined. The areas of utilization and applications of data mining are included in libraries.

In the literature review, it has been observed that terms and concepts related to data mining are not used correctly in many studies. Some of the terms and concepts mentioned in the study are; artificial intelligence, machine learning, statistics, data visualization, databases, geographic information systems, linguistics, statistical modeling, information discovery from databases, data science, data analysis, prediction, forecasting, estimation, etc.

In many sources, it has been observed that different processes related to data mining are used. Two of these processes have been explained in detail in the study. A detailed comparison of the examined five processes has also been included.

Data mining models such as clustering, association rules, and classification have been explained. Then, the taxonomy of data mining and the place of data mining in decision support systems have been examined.

Before delving into the use of data mining in libraries, the relationship between knowledge management, data mining, and big data is examined in the study.

When looking at the literature and applications in our country, the use of data mining in libraries has been mainly approached from its theoretical aspects. It seems that it is not possible to propose a model on a practical level due to the inadequacy of technological infrastructure. Nowadays, with the great transformation in information and communication technologies, it is possible to model the subject on a practical level. Especially with the changes in big data and cloud technologies, data mining programs have become capable of working on cloud servers. Data can be moved to the cloud for analytical operations that cannot be performed on library servers. One of the most important reasons for performing analytical operations on the cloud is that cloud servers can be accessed through leasing options, and they do not require expensive hardware and software investments. Thus, these new technologies create significant opportunities for libraries to improve their processes and services using data mining. At this point, examples have been given from various fields for the use of data mining in libraries. Some of these fields include collection development, user analysis, recommendation systems, text analysis, marketing of information services, and technical services.

In the literature review on the use of data mining in libraries, firstly, postgraduate theses conducted in our country on the subject were given. Then, data sets that can be used in data mining applications in libraries were examined. This was mainly examined in theoretical studies.

In addition to these examples encountered in the literature, studies on specific services and processes of libraries and other information centers were also included. Specifically, sources were found on measuring the relationship between library usage and academic achievement in the literature."

Librarians can process library data with a statistical classification model that they can acquire, since they are familiar with the taxonomy field due to the classification system they use. Librarians can acquire the knowledge to purchase, install and use data mining programs with small-scale trainings. In addition, interdisciplinary work will be required for advanced data mining applications and analysis. Therefore, librarians should collaborate with statisticians and computer programmers regarding the problems and applications they may encounter at a more advanced level.

As a result of the research, it has been concluded that it is possible for libraries to use data mining techniques to obtain new patterns in library operations and services and use them in decision support processes, provided that suitable data sets are available, and the research hypothesis has been confirmed.

This study also highlights the need for libraries to collect and maintain user data and behaviors in order to provide a better user experience. As seen in this study, such applications should be carried out through the collaboration of different disciplines such as statistics, information technology, etc.

Based on the results obtained at the end of the study, various recommendations have been made regarding the use of data mining in libraries.