



Burdur Mehmet Akif Ersoy Üniversitesi
Sosyal Bilimler Enstitüsü / Yönetim Bilişim Sistemleri Anabilim Dalı

YÜKSEK LİSANS TEZİ

VERİ MADENCİLİĞİ YÖNTEMLERİ İLE BİLGİ MERKEZİ VERİLERİNDEN BİLGİ KEŞFİ

SEFA BAYRAKTAR

Danışman
Prof. Dr. İSMAİL KIRBAŞ

T.C.
BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI

**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE BİLGİ
MERKEZİ VERİLERİNDEN BİLGİ KEŞFİ**

Sefa BAYRAKTAR

YÜKSEK LİSANS TEZİ

DANIŞMAN
Prof. Dr. İsmail KIRBAŞ

Üye: Prof. Dr. Adnan KALKAN
Üye: Dr. Öğr. Üyesi Turgay AYDOĞAN

BURDUR - 2024

T.C.
BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

ETİK BEYANI

Burdur Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Lisansüstü Eğitim- Öğretim ve Sınav Yönetmeliği'ne göre hazırlamış olduğum “Veri Madenciliği Yöntemleri ile Bilgi Merkezi Verilerinden Bilgi Keşfi” adlı tezin hazırlanması sürecinde akademik ve etik kuralları ihlal etmediğimi, tezimin özgün olduğunu, başvurduğum kaynakları metin içinde, dipnotlarda ve kaynakçada eksiksiz ve bilimsel kurallara uygun olarak gösterdiğimi taahhüt ederim.

Sefa BAYRAKTAR

26/06/2024

TEŞEKKÜR METNİ

Çalışmanın analizleri için kullanılan veri setinin temininde desteğini esirgemeyen Burdur Mehmet Akif Ersoy Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı'na teşekkürlerimi sunarım. Destekleri için danışmanım Prof. Dr. İsmail KIRBAŞ'a teşekkür ederim.

Ayrıca yüksek lisans eğitimim boyunca zaman ayıramadığım canım oğlum Ömer Tuğra ve biricik kızım Miray'a, sürekli yanımda olan anne ve babama teşekkür ederim.

Bu süreçte fikirleri ve bakış açısıyla destek olan, kendisine yeterince ilgi gösteremediğim sevgili eşim Şeyda BAYRAKTAR'a süreç boyunca gösterdiği sabır ve özveri için sonsuz teşekkürler, sensiz yapamazdım.

Sefa BAYRAKTAR

Ankara, 2024

(BAYRAKTAR, Sefa, *Veri Madenciliği Yöntemleri ile Bilgi Merkezi Verilerinden Bilgi Keşfi*, Yüksek Lisans Tezi, Burdur, 2024)

ÖZET

Bilgi teknolojilerinin neden olduğu büyük verinin etkisiyle hem bireyler hem de kurumlar yoğun bir bilgi bombardımanıya karşı karşıya kalmaktadır. Bu yoğun bilgi karşısında ihtiyaç duyulan bilgiye erişim hem zahmetli hem de maliyetli olabilmektedir. Buna karşın, büyük hacimli veri kümelerinden, uygun teknik ve yöntem kullanılarak yapılan uygulamalar sonucunda kritik öneme sahip değerli bilgiler elde edilebilmektedir. Bu çalışmanın amacı, kütüphane otomasyon sistemlerinde tutulan verilerden veri madenciliği yöntemleri ile kullanıcı davranışları, kullanıcı örüntüleri, koleksiyon örüntüleri ve trendleri ortaya çıkarmak; bilgi merkezlerinde koleksiyon geliştirme, koleksiyon yönetimi ve bilgi kaynaklarının etkin kullanımı için bir yol haritası ortaya koymaktır. Araştırmanın iki motivasyon odağı vardır: (i) kullanıcıların bilgi kaynaklarını kullanma eğilimleri üzerinden her bir kullanıcıya ilgi odağına yönelik bilgi kaynaklarını tavsiye etmek; (ii) kullanıcıların beklenti ve eğilimlerine yönelik ürün ve hizmetlerin geliştirilmesinde kütüphane yönetimine ve üst yönetime karar desteği sağlamaktır. Araştırmanın yürütülmesi için Burdur Mehmet Akif Ersoy Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı tarafından kullanılan kütüphane otomasyon sisteminde (Yordam Kütüphane Bilgi Belge Otomasyonu) tutulan kullanıcılara ilişkin basılı kitap ödünç kayıtları kullanılmıştır. Temin edilen veri seti, veri temizleme, veri dönüştürme ve veri indirgeme işlemleri ile veri ön işlemeden geçirilmiş, RapidMiner Studio Educational 10.3 ve Spyder IDE v5.5.4 paket programı aracılığıyla analiz edilmiştir. Araştırmanın amacına uygun olarak tanımlayıcı istatistikler verilmiş, yıllık bazda en fazla ödünç alınan kitaplar açıklanmış, ödünç verilen kitapların konu dağılımları ve yıllık ödünç sayıları verilmiştir. Öte yandan cinsiyet ve yaş gruplarına göre kitap tercihleri incelenmiştir. Ardından kişiselleştirilmiş kitap öneri sistemi için k-means kümeleme analizi ve SVDs işbirlikçi filtreleme analizi uygulanmıştır. Analizlerin sonunda kitap önerileri sıralanmış ve sistemin RMSE performans değerleri paylaşılmıştır.

Anahtar Kelimeler: *Kitap Tavsiye Sistemi, K-Means, Singular Value Decomposition, Veri Madenciliği, Bibliomining*

(BAYRAKTAR, Sefa, *Discovering Knowledge in Information Centers Data With Data Mining Methods*, Master Thesis, Burdur, 2024)

ABSTRACT

With the impact of big data caused by information technologies, both individuals and institutions are facing an intense information bombardment. Accessing the needed information amidst this abundance can be both laborious and costly. However, through the application of appropriate techniques and methods on large data sets, valuable information of critical importance can be extracted. The aim of this study is to uncover user behaviors, user patterns, collection patterns, and trends from the data stored in library automation systems using data mining methods; and to provide a roadmap for collection development, collection management, and effective use of information resources in information centers. The study has two main motivational focuses: (i) to recommend information resources relevant to each user's area of interest based on their tendencies to use information resources; (ii) to provide decision support to library management and top management in the development of products and services that meet users' expectations and tendencies. For the execution of the research, loan records of printed books related to users held in the library automation system (Yordam Library Information and Document Automation) used by the Library and Documentation Department of Burdur Mehmet Akif Ersoy University were utilized. The obtained data set was preprocessed through data cleaning, data transformation, and data reduction processes, and analyzed using RapidMiner Studio Educational 10.3 and Spyder IDE v5.5.4 software packages. Descriptive statistics were provided in accordance with the aim of the study, the most borrowed books on an annual basis were explained, and the subject distributions and annual loan numbers of the borrowed books were presented. Moreover, book preferences were examined according to gender and age groups. Subsequently, k-means clustering analysis and SVD collaborative filtering analysis were applied for the personalized book recommendation system. At the end of the analyses, book recommendations were listed, and the system's RMSE performance values were shared.

Key Words: *Book Recommendation System, K-Means, Singular Value Decomposition, Data Mining, Bibliomining*

İÇİNDEKİLER

ETİK BEYANI.....	i
TEŞEKKÜR METNİ.....	ii
ÖZET.....	iii
ABSTRACT.....	iv
KISALTMALAR.....	vii
ŞEKİLLER DİZİNİ.....	viii
TABLolar DİZİNİ.....	ix
GİRİŞ.....	1

BİRİNCİ BÖLÜM

BİLGİ ve BİLGİ MERKEZLERİ

1.1. Bilgi ile İlgili Temel Kavramlar	3
1.2. Bilgi Merkezleri ve Türleri	7
1.3. Otomasyon Sistemleri.....	15
1.4. Kütüphane Otomasyon Sistemi Standartları.....	26

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ

2.1. Büyük Veri.....	34
2.2. Yapay Zekâ	38
2.3. Veri Madenciliği	46
2.4. Veri Madenciliği Algoritmaları	54
2.5. Veri Madenciliği Süreci.....	65
2.6. Kütüphanelerde Veri Madenciliği- Bibliomining.....	68

ÜÇÜNCÜ BÖLÜM
VERİ MADENCİLİĞİ YÖNTEMLERİ İLE BİLGİ MERKEZİ
VERİLERİNDEN BİLGİ KEŞFİ

3.1. Araştırmanın Amacı ve Önemi	73
3.2. Literatür Taraması.....	75
3.3. Araştırmanın Varsayımı ve Sınırlılıkları	77
3.4. Araştırmanın Evreni.....	78
3.5. Veri Toplama Yöntemi ve Aracı	78
3.6. Araştırmanın Modeli ve Soruları	80
3.7. Veri Ön İşleme.....	81
3.8. Bulgular ve Değerlendirme.....	83
3.8.1. Tanımlayıcı İstatistikler	83
3.8.2. Kümeleme Analizi	92
3.8.3. İşbirlikçi Filtreleme	102
SONUÇ.....	109
KAYNAKÇA.....	117
EKLER.....	137
EK 1. <i>k</i> -means Kümeleme Analizi Python Kodu.....	138
EK 2. SVDs İşbirlikçi Filtreleme Analizi Python Kodu.....	141
EK 3. Veri Seti Kullanım İzni.....	142
ÖZGEÇMİŞ.....	143

KISALTMALAR

OPAC	: Online Public Access Catalog
OCLC	: Online Computer Library Center
IFLA	: The International Federation of Library Associations and Institutions
EKUAL	: Elektronik Kaynaklar Ulusal Akademik Lisansı
ANKOS	: Anadolu Üniversitesi Kütüphaneleri Konsorsiyumu
MEDLARS	: Medical Literature Analysis and Retrieval System
MARC	: Machine-Readable Cataloging
ILL	: Interlibrary Loan
ISO	: International Organization for Standardization
ANSI	: American National Standards Institute
XML	: Extensible Markup Language
HTML	: HyperText Markup Language
NLP	: Natural Language Processing
KDD	: Knowledge Discovery in Databases
OLAP	: Online Analytical Processing
SVD	: Singular Value Decomposition
SSE	: Sum Square Error
RMSE	: Root Mean Square Error

ŞEKİLLER DİZİNİ

Şekil 1: Bilgi Piramidi	4
Şekil 2: ISO 2709 Yapı Modeli	27
Şekil 3: MARC Kaydı Ham Veri	27
Şekil 4: MARC Kaydı	28
Şekil 5: XML Örneği.....	30
Şekil 6: MARC Kaydı XML Formatı.....	32
Şekil 7: Bibliyografik Katalog Kaydı.....	33
Şekil 8: CRISP-DM Veri Madenciliği Süreci	65
Şekil 9: Araştırma Modeli	80
Şekil 10: Ödünç Veri Seti.....	82
Şekil 11: Yıllık Ödünç Sayısı Dağılımı.....	89
Şekil 12: Silhouette Grafiği	94
Şekil 13: Elbow Grafiği.....	95
Şekil 14: k-means Küme Yoğunlukları	95
Şekil 15: Kümelerin Cinsiyet Dağılımı	96
Şekil 16: Kümelerin Yaş Aralığı Dağılımı	97
Şekil 17: Kümelerin Grup Dağılımı	99
Şekil 18: Kullanıcı-Kitap Etkileşim Matrisi Isı Haritası	105
Şekil 19: k Latent Sayısı Belirleme	106
Şekil 20: Tahmin-Gerçek Karşılaştırması	108

TABLÖLAR DİZİNİ

Tablo 1: Üniversite Kütüphanelerinde Kullanılan Otomasyon Sistemleri	24
Tablo 2: Dünya Geneli Üniversitelerde Kullanılan Otomasyon Sistemleri.....	25
Tablo 3: Bibliomining Veri Akışı	70
Tablo 4: Veri Seti	79
Tablo 5: Nihai Veri Seti	83
Tablo 6: Cinsiyet Dağılımı.....	84
Tablo 7: Yaş Dağılımı.....	84
Tablo 8: Kullanıcı Grubu Dağılımı	85
Tablo 9: Akademik Birim Dağılımı	86
Tablo 10: Yıl Bazında En Çok Okunan Kitaplar	87
Tablo 11: Konu Dağılımı	88
Tablo 12: Cinsiyete Göre Kitap Tercihleri	90
Tablo 13: Yaş Gruplarına Göre Kitap Tercihleri	91
Tablo 14: Kümelerin Konu Dağılımı	98
Tablo 15: Kümelerin En Fazla Ödünç Aldığı Kitaplar	100
Tablo 16: Kullanıcı Kitap Tavsiyeleri (k-means)	101
Tablo 17: Tekil Değer Ayrışımı Performans Karşılaştırma.....	104
Tablo 18: Kullanıcı Kitap Tavsiyeleri (SVDS).....	107

GİRİŞ

Bilgi teknolojilerinde yaşanan gelişim bilgi ortamı format ve miktarında artışlara sebep olmaktadır. İnsanlar her an tüketebileceğinden fazla bilgi ile muhatap olmakta, ihtiyaç duyduğu veya nitelikli bilgiye erişim için fazladan çaba sarf etmektedir. Tüketicilerin karşı karşıya olduğu aşırı bilgi yüklemesi yalnız web ve sosyal medyada değil bilakis nitelikli, filtrelenmiş bilginin bulunması gereken bilgi merkezlerinde de meydana gelmektedir. Basılı kitap, süreli yayın, magazin, dergi dizini ve bilgisayar teknolojilerindeki yenilikler mevcut bilgi miktarına ayak uydurmanın imkânsız olduğuna dair şikayetleri beraberinde getirmektedir (Bawden ve Robinson, 2009: 183). Öyle ki maruz kalınan bilgi potansiyel olarak faydalı dahi olsa aşırı bilgi yüklenmesi kullanıcı tarafından bir engel olarak algılanabilmektedir.

Aşırı bilgiye maruz kalma problemi yakın dönemin değil uzunca zamandır çözüm aranan bir meselesidir. 1852 yılında The Smithsonian Institution tarafından yayınlanan yıllık raporda her yıl binlerce cilt yeni bilgi kaynağının yayınlandığını ve bilgi kaynaklarının düzgün yönetim ve kullanımının bunların depolanmasından daha önemli bir konu olduğu belirtilmekte, eğer bilgi kaynakları doğru düzenlenip kullanıma sunulmaz ise edebiyat ve akademinin bu 'kütle-yığın' altında ezileceği belirtilmektedir (Smithsonian Institution, 1851: 108). 1948 yılında Royal Society's Scientific Information Conference'ta aşırı bilgi yüklemesinin açıkça bir sorun olduğu ve bilim insanlarının bunalmış olacağı, bilgi kaynaklarından gelen muazzam miktardaki potansiyel olarak da alakadar olan materyalin artık kontrol edilemeyeceğini ve bilimin kendisinin tehdit altında olduğu belirtilmiştir (Bawden ve Robinson, 2020: 8). Aşırı bilgi yüklenmesi karşısında arama motorları ve çevrimiçi kütüphane katalogları ile bilgi ve bilgi kaynakları filtrelenmekte, ilgi duyulan kaynaklara erişim kolaylaştırılmaktadır. Yakın zamana kadar arama motorları ve kütüphane katalogları etkin bir çözüm getirmiş olsa bile değişen kullanıcı beklentileri salt arama sonuçlarından fazlasını talep etmektedir. Amazon, Netflix, Spotify ve Alibaba gibi platformlar kullanıcıların geçmişteki davranış ya da tercihlerini analiz ederek ilgi duyacağı ürün, hizmet ya da içerikleri kendisine sunmaktadır. Kütüphaneler de e-ticaret, dijital medya ve sosyal medya platformlarının uyguladığı gibi kişiselleştirilmiş kütüphane deneyimini sağlayarak ürün ve hizmetlerini kullanıcının profiline uygun şekilde sunabilir. Çoğu kullanıcı için kütüphane

katalogundan ihtiyaç duyduğu bilgi kaynağını anahtar kelime, yazar/eser adı sorguları ile bulup ilgili rafta kaynağa erişmek zahmetli bir çözümdür. Ancak Netflix gibi binlerce içerik ve milyonlarca kullanıcıya sahip platformda yayınlanan içeriğin 2/3'ü öneri sisteminin tavsiyelerinden oluşmaktadır (Gomez-Urbe ve Hunt, 2016: 2). Kütüphaneler de binlerce nitelikli bilgi kaynağına sahip olmanın ötesinde içeriklerini kişiselleştirilmiş şekilde hizmete sunarak kullanıcı deneyimini artıracaktır.

Bilgi merkezleri günlük rutin iş süreçleri içerisinde otomasyon sisteminde pek çok işlem yürütmekte ve bu işlemlere ait kayıt tutulmaktadır. Büyük verinin farkında olan işletmeler veri madenciliği uygulamalarına önem vermekte ve tutulan veriden işletmenin ürün veya hizmetlerinin performansını artıracak faydalı örüntüler elde etme gayretindedir. Veri madenciliği, verilerden örtük, daha önce bilinmeyen ve potansiyel olarak faydalı bilgilerin çıkarılmasıdır (Witten vd., 2017: xxiii). Bu çalışma kütüphane otomasyon sistemlerinde kayıtlı veri setleri üzerinden kütüphane koleksiyon ve kullanıcılarının etkileşim ve örüntülerini ortaya koymaktadır.

Bu çalışmanın amacı, kütüphane otomasyon sistemlerinde tutulan verilerden veri madenciliği yöntemleri ile kullanıcı davranışları, kullanıcı örüntüleri, koleksiyon örüntüleri ve trendleri ortaya çıkarmak; bilgi merkezlerinde koleksiyon geliştirme, koleksiyon yönetimi ve bilgi kaynaklarının etkin kullanımı için bir yol haritası ortaya koymaktır. Araştırmanın iki motivasyon odağı vardır: (i) kullanıcıların bilgi kaynaklarını kullanma eğilimleri üzerinden her bir kullanıcıya ilgi odağına yönelik bilgi kaynaklarını tavsiye etmek; (ii) kullanıcıların beklenti ve eğilimlerine yönelik ürün ve hizmetlerin geliştirilmesinde kütüphane yönetimine ve üst yönetime karar desteği sağlamaktır. Araştırmanın birinci bölümünde bilgi kavramından ve bilgi merkezlerinden bahsedilmiş; kütüphane otomasyon sistemleri ve standartları irdelenmiştir. İkinci bölümde büyük veri, yapay zekâ ve veri madenciliği hakkında bilgiler verilmiş, veri madenciliği algoritmalarına değinilmiştir. Üçüncü bölümde ise araştırmanın amacı, yöntemi ve soruları açıklanmış; araştırmanın amacına uygun olarak kişiselleştirilmiş kitap öneri sistemi için *k*-means kümeleme analizi ve SVDs (Singular value decomposition) işbirlikçi filtreleme analizi uygulanmıştır. Analizlerin sonunda kitap önerileri sıralanmış ve sistemin RMSE performans değerleri paylaşılmıştır.

BİRİNCİ BÖLÜM

BİLGİ ve BİLGİ MERKEZLERİ

1.1. Bilgi ile İlgili Temel Kavramlar

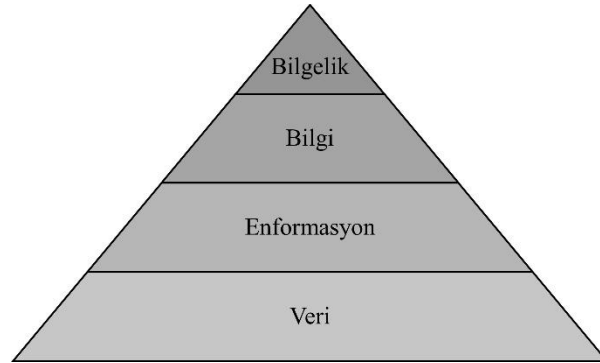
Veri madenciliği, veri tabanları, veri ambarları, web ve diğer bilgi depoları gibi büyük veri yığınlarından ilginç desenler ve bilgi keşfetme sürecidir (Han, Pei, Tong, 2022: 2). Bu süreç, organizasyonlar için stratejik kararlar alırken kritik bir rol oynar ve ham verileri anlamlı bilgilere dönüştürmeyi amaçlar (Fayyad, Piatetsky-Shapiro, Smyth, 1996: 39). Anlamsız gibi görünen her bir veri parçası, doğru teknik ve araçlar kullanılarak kişi veya organizasyonların kullanabileceği etkin analizlere dönüşür. Veri madenciliği tekniklerinin etkin analizler yapmak için nasıl kullanıldığını anlamak, bu tekniklerin temel kavramlarını irdelemeyi gerektirir. Tek başına anlamsız olan veri parçacıkları, hayati öneme sahip tecrübe niteliğindeki bilgiye evrilir. Bu bölümde, tek başına anlamsız olan veri parçacıklarının hayati öneme sahip tecrübe niteliğindeki bilgiye evrimin yapı taşları irdelenmiştir.

İnsanoğlunun karşılaştığı “şeyleri” anlama veya öğrenme arzusu yani merak (Akalin, 2019: 1654), bugünkü teknolojik ve kültürel gelişmenin ana sebeplerinden biridir (Çelik, 2004: 15). Rönesans öncesinde özellikle antikçağlarda merak konusunda olumlu ve olumsuz düşüncelerin olduğu, merakın insan hayatına katkılarının olduğunu savunanların yanı sıra merak etmenin bir musibet, bir tehlike olarak görüldüğü bilinmektedir (Burke, 2017: 34). Buna karşın İslam’da merak, araştırma ve bilmek bir üstünlük ölçütü olarak gösterilmiş (Yeşilyurt, 2004: 3), “bilenlerle bilmeyenlerin bir tutulamayacağı” zikredilmiştir (Zümer 39). İnsanoğlunun bilme arzusu, yaratıldığı günden itibaren nesiller boyunca biriken, yığılan, genişleyen ve gelişen *bilgi* birikimini var etmiştir.

Bilgi teknolojilerinin gelişimi ve yayılması ile literatüre pek çok kavram da yerleşmiştir. Temelde aynı noktadan türemiş olan ‘data, information, knowledge ve wisdom’ kavramları Türkçe’ye henüz tam olarak oturtulabilmiş kavramlar değildir. *Information* ve *knowledge* Türkçe’de yaygın olarak ‘bilgi’ kavramı ile karşılanmakta; kavramlar arasında anlam farklılığı olmasına karşın dilimizde belirgin bir ayırma gidilememektedir (Külcü, 2019: 35). Bu dört kavram hiyerarşik olarak bir piramide oturtulmuş ve bilgi piramidi (Şekil 1) olarak kullanılagelmiştir. Bu piramit, veri (data) en

alt basamakta yer alırken, enformasyon (information), bilgi (knowledge) ve en üstte bilgelik (wisdom) şeklinde yukarı doğru ilerlemektedir. Bilgi piramidinin oluşumu konusunda çokça çalışma yapılmış olsa da genel kanı Rowley (2007: 164) tarafından, Ackoff'un (1989) çalışmasına yapılan atıfla birlikte geliştirilmiş olduğu yönündedir (Başkarada ve Koronios, 2013: 6; Ardolino vd., 2018: 2117; Frické, 2008: 131; Bartlett, 2021: 19). Rowley (2007: 164) bu çalışması ile bir dizi ders kitabından yola çıkarak DIKW (bilgi piramidi) hiyerarşisini yeniden ele alarak veri, enformasyon, bilgi ve bilgelik kavramlarını analiz etmeye çalışmıştır.

Öte yandan Hey (2004: 2) orijinal bilgi piramidinin bilgi bilimi ve bilgi yönetimi disiplinleri için Sharma (2008) tarafından ortaya atıldığını belirtmektedir. Sharma (2008) çalışmasında bilgi piramidinin doğuşunda Zeleny (1987) ve Ackoff'un (1989) çalışmalarından esinlendiğini belirtmektedir. Ayrıca birçok araştırmacı hiyerarşinin ilk kez bilgi yönetimi veya bilgi bilimi alanında değil Eliot'a (1934) ait olan The Rock isimli şiirde geçtiğini vurgulamaktadır (Rowley, 2007: 166; Sharma, 2008; Hey, 2004: 2; Bartlett, 2021: 19). Şiirde geçen *"Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?"* ifadeleri enformasyon, bilgi ve bilgelik kavramlarını peşi sıra hiyerarşi içinde ortaya koymaktadır.



Şekil 1: Bilgi Piramidi

Diğer bir yandan Frické (2008) bilgi piramidinde bazı mantık hataları olduğunu ve modelin bilgi bilimi alanında kullanılmaması gerektiğini vurgulamaktadır. Bilgi piramidi 80'lerden bu yana sürekli iyileştirmeye tabi tutulmuş olsa da hiyerarşinin unsurları ve yapısı hakkında sürekli bir tartışmanın olduğu bilinmektedir (Aven, 2013: 31). TDK Türkçe Sözlük'te veri için şu tanım yapılmıştır: *Bir araştırmanın, bir tartışmanın, bir muhakemenin temeli olan ana öge, muta, done; olgu, kavram veya*

komutların, iletişim, yorum ve işlem için elverişli biçimli gösterimi; bilgi, data (Akalin, 2019: 2480).

Literatür incelendiğinde ise bu dört kavram için yabancı orijinine yönelik olarak belli bir hiyerarşi içinde kavramların sıralandığı görülmektedir. Akpınar (2014: 15) veri (*data*) kelimesinin latince “vermek” anlamındaki **dare** ve onun çoğulu olan **datum** kelimesinden geldiğini ifade etmektedir. Veri, herhangi bir aksiyon üzerinden elde edilen ve doğrudan olmasa da dolaylı olarak belli bir anlam içeren, işlenememiş gerçeklikleri ifade etmektedir (Çağiltay, 2010: 9). Bir diğer tanıma göre ise veri, gözlem veya ölçümlerle elde edilmiş ve herhangi bir işlem görmemiş olan her tür değere karşılık gelmektedir (Şeker, 2015: 22). Veri, nesneler, olaylar ve çevrenin özelliklerini temsil eden sembollerdir (Aven, 2013: 31). İngilizce karşılığında ‘data’ olarak konumlandırılan veri gerek bilgi teknolojilerinde gerekse insan hayatında en temel olguyu temsil etmektedir. Herhangi bir nesneden belli ölçüm ve etkileşimler elde edilebiliyorsa ‘bilgi’ edinme yolunda ilk adımın atıldığı söylenilebilir. Veri, belli bir sayı dizisinden oluşabildiği gibi hava tahmin sensöründen elde edilen sıcaklık değerleri de olabilmektedir.

Veri madenciliği, ham verilerin analiz edilerek değerli bilgiler ve desenler ortaya çıkarılması sürecidir (Fayyad, Piatetsky-Shapiro, Smyth, 1996: 40). Bu noktada veri kavramının doğru şekilde oturtulmuş olması yapılan bilgi keşfi aşamasına bir ön hazırlık olacaktır.

Ham değerlerin işlenerek belli bir anlam yüklenmiş hali ise Türkçe’de enformasyon olarak kullanılmaktadır. Enformasyon (*information*), verilerin işlenerek herhangi bir karar için değer kazandırılan ve düzenlenmiş gerçekliklerdir (Çağiltay, 2010: 9). Veri, ilgili bir forma dönüştürülene kadar hiçbir değere sahip değildir (Aven, 2013: 31). Enformasyon somut bir olgu olarak sesli, görüntülü, yazılı vs. biçimlerde bilgi kayıt ortamlarında yer almaktadır (Odabaş, 2007: 99). Enformasyon gönderici ile alıcı arasında bir etkileşime sebebiyet vermeli, alıcısını biçimlendirmeli, bakış açısında ya da anlayışında bir fark yaratmalı; bu bağlamda da enformasyon fark yaratan veridir (Davenport ve Prusak, 2000: 24).

İşlenen verilerden elde edilen enformasyon, bilgi olma yolunda yorum ve analize tabi tutulabilecek bir forma sokulmuştur. Enformasyon üzerinden elde edilen yorumlar

ve analizler gerek kişisel gerekse organizasyonel karar almada hayati bir rekabet avantajı sağlamaktadır. Bilgi (*knowledge*) “*insan aklının erebileceği olgu, gerçek ve ilkelerin bütünü, bili, malumat; öğrenme, araştırma veya gözlem yolu ile elde edilen gerçek, malumat, vukuf*” olarak tanımlanmakta (Akalın, 2019: 338); bilginin elde edilmesinde insan unsurunun etkinliğine vurgu yapılmaktadır. Bilgi, teknik bilgidir (know-how) yani neyi nasıl yapacağını bilme durumudur (Ackoff, 1989: 4; Rowley, 2006: 256). Çağıltay (2010: 9) ise insan faktörüne vurgu yaptığı tanımında bilgiyi, enformasyon haline dönüştürülmüş gerçeklerin analiz edilmesi ve sentezlenmesi sonucu, karar vermeye yönelik olarak elde edilen daha üst seviyeli gerçekler olarak tanımlamaktadır. Diğer bir deyişle de bilgi, fikirlerimizin uyumu veya uyumsuzluğunun algılanabilmesidir (Locke, 2018: 211). Bilgi, bilen tarafından içselleştirilir ve bu haliyle kişinin mevcut algı ve deneyimleri sonucunda şekillendirilir (Hey, 2004: 9).

Bilgi piramidinin en üst basamağında konumlandırılan en büyük değerin verildiği bilgelik (*wisdom*), ham verinin işlenmesinin ardından büyük bir yorum ve tecrübe eklenerek elde edilen değer olarak karşımıza çıkmaktadır. Bilgelik, insanoğlunun elde ettiği verilerden çıkarımlar yapabilmesinin son raddesi; uzmanlığın ötesinde tecrübe ile sabit kanılar bütününe ifade etmektedir. Bilgelik doğru ile yanlışın, iyi ile kötünün ayırt edildiği veya yargılandığı bir süreç olması bakımından bilgisayarların/ bilgi sistemlerinin *bilgelik* yeteneğine sahip olamayacağı vurgulanmaktadır (Bellinger, Castro ve Mills, 2004). Bellinger, Castro ve Mills’in *bilgelik* kavramında bilgisayar ve insanları birbirinden ayıran *ruh ve kalp* olgularına parmak basarak bilgeliğin akılda olduğu kadar ancak kalpte var olabileceğine dikkat çekilmiştir. Rowley (2006: 257) de bilgeliğin ahlaki, sosyal ve etik unsurları barındıran bir muhakeme yeteneği olduğunu vurgulamış; “*bilinen (knowledge) ile neyin iyi olduğunu (etik ve sosyal düşünceler) hesaba katarak en uygun davranışı eyleme geçirme*” tanımını yapmıştır. Bilgelik, şimdiki zaman ya da geçmiş değil; bilakis gelecek ile ilgilenmekte (Bozkurt, 2014: 512), geleceği şekillendirmek için geçmiş veri, bilgi ve tecrübelerden faydalanmaktadır. Şeker (2015: 22) wisdom kavramını bilgeliğin yanında irfan olarak da nitelendirmiş; veriden bilgeliğe giden yolu belirsiz (nondeterministic) olarak ifade etmiştir.

1.2. Bilgi Merkezleri ve Türleri

İnsanoğlunun gerek yaşamsal faaliyetleri gerek entelektüel girişimleri sebebiyle merak ve öğrenme arzusu bilgi edinimine, bilgi birikimine ve bilgi üretimine olanak sağlamaktadır. Diğer varlıklardan farklı olarak insan her daim bireysel ve toplumsal olarak edindiği bilgi sayesinde var olduğu konumdan ileriye gitmiş, sınırlarını aşmış ve nihai hedeflerine ulaşmıştır. Mağaralarda duvarlara çizilen figürlerden, kil tabletlere, papirüs ve parşömenden kâğıda kadar bilgi kayıt ortamları bilginin muhafazası iç güdüsüyle kullanılagelmiştir. Elde edilen her yeni bilgi birey veya toplum için birer üstünlük çarpanı haline gelmiştir. Bilginin kayıt altına alındığı ortam her ne olursa olsun ihtiva ettiği unsur değişmemiş, insanoğlunun bilgi birikimini diğer toplumlara, gelecek nesillere ulaştırmaya vesile olmuştur. Bilgi merkezleri (kütüphaneler) sadece bireyin değil aynı zamanda okur-yazar toplumların kültürel hafıza kurumlarıdır (Mittler, 2016: 235).

Elde edilen bu bilgi birikimi sonrası özellikle de bilgi kayıt ortamları üzerinde tutulan bilgilerin bir araya getirilerek korunması, düzenlenmesi ve kullanılması ihtiyacı çeşitli yapıların ortaya çıkmasına vesile olmuştur. *Bünyesinde barındırdığı bilgiyi belirli kurallar ve standartlar dahilinde düzenleyen, koruyan, kullandıran ve yarınlara aktarılmasına aracılık eden toplumsal kurumlara bilgi merkezi* adı verilmiştir (Akkaya, Odabaş, 2019: 8). Gelişen bilgi kayıt ortamları ve kullanıcı kitlesi bakımından bilgi merkezlerinin türleri ortaya çıkmış; kütüphane, arşiv, dokümantasyon ve enformasyon, müze ve arşivler bilgi merkezlerinin birer türü olarak şekillenmiştir (Yontar, 1995: XI; Akkaya ve Odabaş, 2019: 8). Öte yandan özellikle kullanıcı odaklı bilgi hizmeti verilmesi fikrinden hareketle bilhassa üniversitelerde bilgi merkezi kavramının kütüphane yerine kullanıldığı, pek çok kütüphane teşkilatının bilgi merkezi ismiyle nitelendirildiği bilinmektedir (Üstün, 2017: 7). Bu çalışmada bilgi merkezi ve kütüphane eş anlamlı olarak kullanılmış, arşiv, müze, dokümantasyon ve enformasyon merkezlerinden ziyade kütüphane kurumu üzerinde durulmuştur.

Bilgi merkezlerinin türü ne olursa olsun ana odak noktasında bilgi ve bilgi kaynaklarının toplanması, düzenlenmesi ve kullanıma sunulması amacı vardır. Kütüphane de *bilgi kaynaklarını toplayan, düzenleyen belirli sistemlere tabi tutarak onları kullanıma hazır hale getiren ve koruyan kurumlar* (Üstün, 2017: 5) olarak

tanımlanmaktadır. İngilizce *library*, Almanca *bibliothek*, Türkçe’de ise Arapça ve Farsça iki kelimenin birleşiminden *kütüphane* (Ar. *kutup* + Far. *hâne*) (Akalin, 2019: 1565) kelimesi kullanılmaktadır. Kütüphane kavramı tüm dillerde yazılı materyallerin depolandığı yer olarak nitelendirilmekte; tablet evi, papirüs evi gibi (Keseroğlu, 2019: 11) dönemin bilgi kayıt ortamlarına göre farklı tabirler yer almaktadır. Bilginin kayıt altına alındığı ilk zamanlardan bu yana binlerce yıldır insanoğlu her daim “kitaplarını” saklamanın çeşitli yollarını bulmuş; bugün kitap diye görmediğimiz “metinleri” dahi sistemli bir şekilde bir araya getirerek muhafaza etmiştir (Pyne, 2019: 10).

Tarihi olarak bilgi merkezlerinin ortaya çıkışı yazının icadına kadar dayandırmak mümkün olsa da Suriye’de Halep şehri yakınlarındaki Ebla Kraliyet Sarayı’nda ve Irak’taki Tell Fara (antik Shuruppak) bulunan ve sistematik olarak tabletlerin muhafaza edildiği yapılar kütüphanelerin (M.Ö. 2600) başlangıcı olarak kabul edilmektedir (Potts, 2014: 38; Pyne, 2019: 12; Casson, 2001: 3). Daha öncesine ait yazılı belgelerin bulunduğu arkeolojik çalışmalar da olmakla birlikte tutulan kayıtların senet, mahkeme kaydı, sözleşme, makbuz, envanter türü bilgiler içermesi bakımından bu yapıların daha çok arşiv olarak nitelendirildiği bilinmektedir (Casson, 2001: 3). Ebla sarayında bulunan binlerce kil tablet sistematik olarak raflara dizildiği ve ufak odalarda saklandığı ortaya çıkarılmıştır. Öte yandan rafların ve tabletlerin raflara yerleşimi kütüphanenin “kullanıcılarının” olduğuna işaret etmekte ve belli bir kullanım senaryosunun varlığını kanıtlamaktadır (Pyne, 2019: 12). Kütüphanenin dermesine bakıldığında ise Sümerce ve Ebla dilinde büyüler; meslekler, coğrafi konumlar, hayvanların isimleri; Sümer mitleri; saray tarafından yapılan materyal dağıtımlarını ihtiva eden tablet kopyaları tespit edilmiş; ayrıca saray kayıtlarının bulunması bu yapının saray katipleri tarafından kullanıldığına işaret etmektedir (Casson, 2001: 3).

Asur hükümdarları I. Tiglath-Pileser (M.Ö. 1115-1077) ve Asurbanipal (M.Ö. 668-627) kütüphane kurucusu olarak adı bilinen ilk kişilerdir (Casson, 2001: 9). Her iki hükümdar da kütüphaneye büyük önem vermiş ve yalnızca kendisine ait olan bilgi kaynaklarını değil farklı şehirlerden ganimet olarak elde edilen tabletler de toplanarak kütüphaneye kazandırılmıştır (Potts, 2014: 42). Özellikle hükümdar Asurbanipal okuma-yazmaya büyük bir önem vermiş, tablet yapımı ile ilgilenmiş, kutsal kitaplara ve taş tabletlere ilgi duymuş, kütüphaneler kurmuş ve bir yazıcı olarak “*atalarım krallar arasında hiçbiri böyle bir sanat öğrenmemişti*” diyerek bu özelliğiyle övünmüştür

(Tekin, 2019: 51-52). Kütüphane dermesindeki tabletlerin sayısı konusunda ihtilaflı çalışmalar olsa da 30.000 tabletin bu kütüphanede var olduğu bilinmektedir (Potts, 2014: 40; Tekin, 2019: 63). Dermenin büyük bir kısmı 19. yüzyılda British Museum'a aktarılmış ve hükümdar Asurbanipal'in Babil ve Asur eserlerini bir araya getirme gayreti ile ilk ulusal kütüphane olma özelliği kazanmıştır (Barnes, 2014: 83).

Antik zamanların en ihtişamlı ve efsanevi kütüphanesi ise İskenderiye Kütüphanesidir. Büyük İskender'in ölümünden (M.Ö. 323) ve sonrasında generalleri tarafından imparatorluğun parçalanması sonucu kendini Mısır kralı olarak ilan eden I. Ptolemaios Soter' in entelektüel kişiliği ve bilime, bilim adamlarına olan düşkünlüğü Yunan aydınlarını Mısır'a toplamasıyla kütüphanenin oluşumu başlamıştır (Barnes, 2014: 84). Yapının tam olarak bir tasviri olmasa da ilk başlarda bir müze, okul ve araştırma enstitüsü yapma fikri olduğu, gelişen zaman içinde saraya yakın bir (veya iki) yapının tesis edildiği tahmin edilmektedir (Barnes, 2014: 85). Sarayda doğrudan müze üyelerine hizmet veren bir kütüphane ve saraya yakın bir noktada (mabet içinde olabilir) ikinci bir kütüphane (harici kütüphane veya kız kütüphanesi) olduğu düşünülmektedir (Casson, 2001: 34; Barnes, 2014: 86). Kütüphane kralın, alimlerin, filozoflar ve sanatkarların ihtiyaç duyduğu/ duyabileceği her bilgi kaynağını toplamayı amaç edinmiştir (Tunay, 1970: 25). Hükümdar Asurbanipal de kütüphanesine kaynak toplama amacını gütmüş ancak İskenderiye Kütüphanesi'nde bu gayret oldukça kapsamlı olarak işletilmiş ve dünyanın her yerinden kitapları (tomar) buraya toplamak için uğraş verilmiştir. Gerek kralın kütüphaneye ayırdığı geniş bütçe ile satın alınan kitaplar gerekse özellikle limana gelen gemilerde yapılan aramalarla ele geçirilen kitapların kopyaları çıkarılarak aslının kütüphaneye alınması sonucu kütüphane dermesi zamanın en büyük koleksiyonunu ortaya çıkarmıştır. Koleksiyondaki kitap sayısı konusunda çeşitli bilgiler olsa da bu sayının 200.000 ila 700.000 arasında olduğu tahmin edilmektedir (Barnes, 2014: 86-88). M.Ö. 48 yıllarında Julius Caesar tarafından İskenderiye' ye başlatılan savaş sonrası kütüphane büyük yara almış ve koleksiyonun bir kısmı çıkan yangınlarda kül olmuştur. Kütüphanenin ve dolayısıyla dermenin yanması konusunda ihtilaflı düşünceler bulunmakta; tek seferde olmasa bile geçen zaman içinde tüm İskenderiye Kütüphanesi dermesi yok edilmiş ve günümüze gelebilen bir kaynak olmamıştır (Casson, 2001: 47; Barnes, 2014: 98). Öte yandan geçen zaman içinde dermenin kalan kitaplarının Araplar (Halife Ömer) tarafından yakıldığı hususunda zayıf bir düşünce var olsa da hem kronoloji

bakımından hem de kaynakların güvenilirliği bakımından buna pek itibar edilmemektedir (Barnes, 2014: 99). Bu konuda İlber Ortaylı verdiği bir röportajda Mısır'ı fetheden Amr İbnül As'ın kütüphane koleksiyonunun yakılması emrini verdiği iddiasının yalan olduğunu ifade etmektedir (Ortaylı, 2006: 85).

İslam dünyasında kitap toplama ve kütüphane teşkil etme faaliyetleri Asr-ı Saadet dönemine kadar gitmektedir (Erünsal, 2018: 326). İslamiyet öncesinde cahiliye döneminde Araplar arasında okuma yazma bilenlerin olduğu ve edebiyata düşkünlükleri bilinmekte, dönemin bilgi kayıt ortamlarında yazılı eserler verildiği görülmektedir. Arapların İslamiyet öncesinde verdiği en büyük sanat eseri şiirdir (Çetin, 2011: 1). Şiir Araplar arasında önemli bir yer tutmakta, her yıl düzenli olarak panayırarda şiir müsabakaları düzenlenmekte ve başarılı olan şiirler “el-Muallakatu’s-Seb’a (Yedi Askı)” Kabe’nin duvarına asılmaktaydı (Ürün, 2015: 19). İlk vahyin “Oku” (Alak 1) emri ile nazil olması sonrası Hz. Muhammed’in peygamberliği süresince Kur’an-ı Kerim’in tüm ayetleri tamamlanmış, sahabe içerisinde bu ayetleri ezberleyenlerin olduğu kadar papirüs, deri, tahta tablet, kemik ve hurma dallarına da yazılmıştır (Erünsal, 2018: 33). Peygamberin vefatı sonrası Halife Hz. Ebu Bekir zamanında ayetlerin bir araya toplatılması ve Kur’an-ı Kerim’in kitap haline getirilmesi, sonrasında da Halife Hz. Osman zamanında çoğaltılması ilk kitap koleksiyonlarının oluşmasını sağlamıştır. Ancak ilk kütüphane örneklerinin ortaya çıkması Hz. Muhammed’in vefatından uzun zaman sonra Emeviler dönemine (M.S. 661-750) denk geldiği tahmin edilmektedir. Miladi 7-8. yüzyılda beyt’ül hikme veya darü’l hikme adındaki araştırma kurumlarında kütüphanelerin ortaya çıktığı görülmektedir (Erünsal, 2018: 327). Bu kurumlarda hadis, tarih ve biyografi alanlarında kitapların olduğu ve kütüphaneden sorumlu bir kütüphanecinin bulunduğu da ifade edilmektedir.

Gelişen zaman içinde tüm disiplinlerde olduğu gibi kütüphanecilik ve kütüphane kavramı üzerinde de çeşitlenmeler yaşanmış, önceleri kitapların bulunduğu bir kurum olan kütüphaneler zamanla ihtiva ettiği derme çeşidi ve hizmet verdiği kullanıcı grubuna göre farklı isimlendirmelere ve alt dallara ayrılmıştır. Kütüphanelerin genel olarak amacı sabit olsa da bağlı olduğu kurum, fon sağlayıcı, derme geliştirme politikası, kullanıcı grubu vs. gibi belli bazı ayırıştırıcı unsurları bulunmaktadır. Aşağıda yaygın olarak hizmet veren kütüphanelerden milli kütüphane, üniversite kütüphanesi ve halk kütüphanesi irdelenmiştir.

Milli kütüphane, adı ne olursa olsun bir yasa ya da diğer yönetmeliklerle ülke içinde yayınlanan tüm önemli yayınları derleyen ve muhafaza eden kütüphanelerdir (UNESCO, 1971: 145). UNESCO yaptığı tanımda milli kütüphanelerin ulusal bibliyografya üretme, ülke hakkında temsil yeteneğine sahip bilgi kaynaklarına sahip olma ve bunları güncel tutma, ulusal bibliyografik bilgi merkezi olarak hareket etme ve toplu katalog çalışmalarına sahip olma yükümlülüklerini de vurgulamaktadır. IFLA ise ülkede yayınlanan tüm bilgi kaynaklarının derlenmesi ve muhafaza edilmesinden sorumlu kütüphane olarak tanımlamakta ve bir ülkede birden fazla milli kütüphane var olabileceğini belirtmektedir (IFLA, 2023). Milli kütüphaneleri diğer kütüphanelerden ayıran üç özellik vardır; *miras* (ulusal kültürün korunup gelecek kuşaklara aktarılması için bilgi kaynaklarının muhafaza edilmesi), *altyapı* (ulusal koordinasyon, diğer kütüphanelere önderlik etme), *kapsamlı ulusal hizmet* (diğer kütüphane türlerinin aksine ülke vatandaşlarının tümüne hizmet verme) (Lor, 1997: 5). Kütüphaneler genel olarak belli bir kullanıcı kitlesine belli niteliklerde koleksiyon ile hizmet verme gayreti içindedir. Ancak milli kütüphanelerde hem derme açısından hem de kullanıcı profili açısından odak çok geniştir. Koleksiyon ülkedeki tüm bilgi kaynaklarını bir araya getirme ve ülke dışında yayınlanan kaynaklarla da zenginleşirken kullanıcı profili olarak öğrenci, akademisyen, kurum personeli, sivil vatandaş vs. ayrıma gitmeden tüm son kullanıcıları içine almaktadır. Öte yandan milli kültür mirası gibi bir amacı olması nedeniyle de genellikle bilgi kaynakları kütüphane dışına ödünç verilmez (Konya, 2019: 21). Kütüphanenin amacı gereği koleksiyondaki bilgi kaynakları uzun yıllar boyunca saklandığı, diğer mecralarda bulunmasa bile milli kütüphane koleksiyonunda en az bir nüshasının temin edileceği bir yapı söz konusudur. Bu bakımdan ulusun en önemli kültür hazineleri olarak hizmet vermektedir. IFLA verilerine göre 2022 yılı itibarıyla dünyada 337 adet milli kütüphane bulunmaktadır (IFLA, 2022).

Türkiye’de milli kütüphane kuruluş çalışmaları Adnan Ötüken önderliğinde 15 Nisan 1946 yılında gayri resmi olarak “Milli Kütüphane Hazırlık Bürosu” adı ile başlamış, burada bir süre bazı bilgi kaynaklarının temin edilmesi ve ön hazırlık çalışmaları yapılmış ve 8000 kitaplık bir koleksiyon oluşturulmuştur. 14 Ağustos 1946 yılında ise Başvekil Recep Peker’in hükümet programında “Ankara’da bir milli kütüphane kurulması çalışmalarının planlandığını” duyurması üzerine çalışmalar resmi olarak başlatılmış ve 16 Ağustos 1948’de kütüphane halka açılmıştır. Milli Kütüphane 23

Mart 1950’de 5632 sayılı kanun ile Milli Eğitim Bakanlığı’na bağılı bir kurum olarak resmiyet kazanmış, 5 Ağustos 1983’ ise Ankara’da bugün halen hizmet verdiği binaya taşınmıştır (Ötüken, 1955: 12-25; Milli Kütüphane, 2023). 29.02.2012 tarih ve 6279 sayılı “Çoğaltılmış Fikir ve Sanat Eserlerini Derleme Kanunu” ve 18.08.2012 tarih ve 28388 sayılı “Çoğaltılmış Fikir ve Sanat Eserlerini Derleme Yönetmeliğı” gereğı Türkiye Cumhuriyeti sınırları içerisinde çoğaltılmış ve yayımlanmış her türlü eserin ve ikili ya da çok taraflı anlaşmalarla yurt dışında çoğaltılmış ve yayımlanmış fikir ve sanat eserlerinin nüshaları, Milli Kütüphane, Türkiye Büyük Millet Meclisi Kütüphane ve Arşiv Hizmetleri Başkanlığı, İstanbul Beyazıt Devlet Kütüphanesi, İzmir Milli Kütüphane Vakfı Kütüphanesi, İstanbul Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı ve Cumhurbaşkanlığı (Cumhurbaşkanlığı Millet Kütüphanesi)’na (Değişik: 03/07/2017- 30085/9 md.) gönderilmek suretiyle derlenmektedir.

Üniversite kütüphanesi, bir veya daha fazla bilim ve sanat dalında eğitim ve araştırma yapılan ve derece, diploma veya sertifika vermeye yetkili bir yükseköğretim kurumunun (kolej, fakülte veya üniversite) ayrılmaz bir parçası olan (Gelfand, 1968: 15) ve bağılı olduğı kurumun fakülte, öğrenci ve personelinin araştırma ve müfredat ihtiyaçlarını karşılamak üzere üniversite tarafından kurulan, yönetilen ve finanse edilen kütüphane veya kütüphane sistemidir (Reitz, 2004: 743; Levine-Clark ve Carter, 2013: 263). Bir eğitim öğretim kurumu olmasının yanı sıra bilimsel araştırma ve geliştirme, bilim insanı yetiştirme gibi hedeflerinin var olması üniversite kütüphanelerini de kültürel bir organizasyon olmanın ötesinde araştırma odaklı bir yapıya büründürmektedir. Yapılan araştırma ve geliştirme faaliyetleri bilim insanlarının daha fazla bilgi kaynağına ihtiyaç duymasına sebebiyet vermekte bu da kütüphaneyi bilimin doğabileceğı ortamın merkezine çekmektedir. Farklı kişilere atfedilmesi bir kenara ünlü “*Kütüphane üniversitenin kalbidir*” sözü (Simpson, 2016: 503; Leupp, 1924: 193; Himmelfarb, 1997: 197; Greider, 1956: 93) kütüphanenin, üniversitenin olmazsa olmaz yapılarından biri olduğuna işaret etmektedir.

Milli kütüphaneden farklı olarak üniversite kütüphaneleri belli bir kullanıcı kesimi ve kullanım senaryosuna sahiptir. Bu sayede kütüphane dermesi daha özel bir içerik barındırır. İçerik, üniversite mensuplarının bilgi ve entelektüel ihtiyacına yönelik olarak geliştirilir. Çoğı otorite üniversite kütüphanelerini araştırma kütüphanesi kategorisinde değerlendirmekte (IFLA, 2023; Association American Library, 2016) ve eğitim-öğretim,

araştırma kavramları ile birlikte anılmaktadır (Polat, 2019: 227). Üniversite kütüphanelerinin ana görevleri şu şekilde sıralanabilir;

- *Kullanıcıların ihtiyaç duyduğu bilgi ile bilgi kaynaklarının seçimi ve temini,*
- *Bilgi kaynaklarının kolay erişimi için organize edilmesi ve katalog kaydının tutulması,*
- *Bilgi kaynaklarının temin edilme şeklinin belirlenmesi için işaretlenmesi ve etiketlenmesi,*
- *Bilgi kaynaklarının gelecek nesillere aktarılmasını sağlamak için gerekli koruma yöntemlerini kullanmak,*
- *Bilgi kaynaklarının kurum mensupları tarafından maksimum fayda sağlayacak şekilde erişilebilir hale getirmek,*
- *Farklı öğretim araçları ile bilgi kaynaklarının kullanılması konusunda kullanıcılara yardımcı olmak,*
- *Bilimsel çalışmalara elverişli mekanların tesis edilmesi,*
- *Diğer kütüphanelerle kaynak paylaşımı konusunda iş birliği sağlamak (R. D. Rogers, Weber, 1971: 3).*

Üniversite kütüphaneleri, üniversitelerin yapı ve kültüründen doğrudan etkilenmektedir. Türkiye’de merkezi kampüse sahip nispeten daha yeni üniversiteler kütüphane olarak “merkez kütüphane” anlayışını benimsemiş; daha eski üniversitelerde kurulan kütüphaneler ise merkez kütüphanenin yanında “fakülte kütüphanesi” anlayışını da sürdürmektedir. Coğrafi olarak farklı semtlerde kurulu fakültelerden müteşekkil üniversitelerin kütüphaneleri her fakültede bir kütüphane/ kitaplık kurulması şeklinde anlayış benimsemiştir. Bu yapı koleksiyonların büyümesi ile birçok sorunu beraberinde getirmiştir (Üstün, 2017: 175).

Türkiye’de 4/11/1981 tarih ve 2547 sayılı “Yükseköğretim Kanunu” ile kurulan üniversitelerde kütüphane ve kütüphane hizmetleri 21/11/1983 tarih ve 124 sayılı “Yükseköğretim Üst Kuruluşları ile Yükseköğretim Kurumlarının İdari Teşkilatı Hakkında Kanun Hükmünde Kararname” gereği “Kütüphane ve Dokümantasyon Daire Başkanlığı” sorumluluğunda yürütülmektedir. Kütüphaneler üniversitenin sağlamış olduğu maddi imkanlar doğrultusunda koleksiyon, mekân ve altyapıyı geliştirmekte;

diğer kütüphanelerle kaynak, personel paylaşımı yapabilmekte; özellikle elektronik bilgi kaynakları konusunda TÜBİTAK'ın sağlamış olduğu EKUAL programlarından yararlanarak kullanıcılarına hizmet vermektedir.

Halk kütüphanesi, *her türlü bilgi ve enformasyonu kullanıcılarına hazır hale getiren yerel bilgi merkezi* (Krass vd., 2022); *tamamen veya kısmen kamu tarafından fonlanan, belli bir coğrafi bölgenin tüm sakinlerine ücretsiz kütüphane kaynak ve hizmetlerinden sınırsız erişim imkanı sağlayan kütüphane veya kütüphane sistemidir* (Reitz, 2004: 578). Halk kütüphaneleri temel hizmet ve koleksiyonlarını, yasal hizmet alanındaki kullanıcılara ücretsiz olarak sunabildiği gibi yasal hizmet alanı dışındaki kullanıcılardan da ürün ya da hizmetleri karşıda ücret talep edebilmektedir (Levine-Clark ve Carter, 2013: 205). Milli kütüphaneler ile benzer özellikler taşıyan halk kütüphaneleri, toplumun (özellikle yerel ölçekte) her kesimine hizmet verme amacı gütmektedir. Üniversite, kamu ya da özel kurum mensupları haricinde kalan ve bu kurumlara ait kütüphane hizmetlerinden yararlanamayan kütüphane kullanıcılarının bilgi ve bilgi kaynağı ihtiyaçlarını karşılamada önemli bir yer tutmaktadır. Toplumun her kesimine kütüphane hizmeti sağlamayı amaçlayan halk kütüphaneleri, il/ ilçe kütüphaneleri, gezici kütüphane, bebek ve çocuk kütüphanesi olarak farklı varyasyonlarda hizmet verebilmektedir. Türkiye'de Kültür ve Turizm Bakanlığı'na bağlı olarak 2021 yılı itibariyle 1252 halk kütüphanesi (çocuk kütüphaneleri dahil), 66 gezici kütüphane ile 15.683.134 kullanıcıya hizmet vermektedir (Kütüphaneler ve Yayınlar Genel Müdürlüğü, 2021).

Halk kütüphanelerinin kullanıcı kitlesinin geniş olması koleksiyon ve hizmet çeşitliliğini de artırmaktadır. Yalnız bilgi ve bilgi kaynağı sunmanın dışında yaşam boyu öğrenme konusunda da destek verebilmektedir. Halk kütüphanelerini yalnızca bilgi ve bilgi kaynaklarının erişilebildiği mekanlar olarak nitelendirmemek, demokrasiyi, eşitliği ve sosyal adaleti sağlayan, bilgiye erişimi kolaylaştıran, kültür ve bilginin yayılmasını sağlayan, nitelikli ve bilgilendirici boş zamanlara katkı sağlayan bir sosyal buluşma mekanı olarak görmek gerekmektedir (Aabø, 2005: 210). Halk kütüphaneleri, toplumda okuryazarlığın artırılması için destek verebilmekte, her yaştan kullanıcılar için geniş bir bilgi kaynağı çeşitliliği sağlayabilmekte ve toplumsal bilgi hizmetleri noktasında toplumun vazgeçilmez bir parçası olarak kabul edilmektedir (Haider vd., 2022). UNESCO tarafından halk kütüphaneleri, *"bilgiye evrensel erişim ve tüm insanlar için*

bilginin etkin kullanımının sağlanması adına sürekli olarak yeni iletişim araçlarına uyum sağlayan bilgi toplumlarının gerekli bir bileşeni, bilgi üretimi, bilgi ve kültür paylaşımı/ değişimi ve sivil katılımın teşviki için kamuya açık alan sağlayan merkezler” olarak nitelendirilmektedir (Krass vd., 2022).

Milli kütüphane ve üniversite (araştırma) kütüphanelerinin kullanıcı kitleleri daha çok bilgi kaynaklarına yoğun şekilde ihtiyaç duyan öğrenci ve akademisyenlerden oluştuğu için kütüphane kullanımları halk kütüphanelerine nispeten daha etkin olabilmektedir. Fakat halk kütüphanelerinin kullanıcı kitlesi göz önüne alındığında, özellikle akıllı telefon ve kişisel bilgisayarlarla birlikte internet kullanımının toplumun hemen her kesimine yayılmış, kişinin ihtiyaç duyacağı bilgiye (doğruluğu tartışma konusudur) her an ulaşabileceği bir ortamda halk kütüphanelerinin nasıl bir rol üstleneceği önem arz etmektedir. Bu konuya işaret eden Aabø (2005) halk kütüphanelerinin sosyal etkisini *toplum*, *beceriler* ve *ekonomi* olmak üzere üç başlık altında toplamıştır. *Toplum* üzerinde; yerel kimliklerin sürdürülmesi, iş gücü piyasası dışında kalan insanları desteklemek, kültürel zenginleşme ve çeşitliliği teşvik, kriz zamanlarında bilgi sağlama, yeni bilgi kaynaklarının kullanımına destek olmak gibi etkilerinin olduğunu ifade etmektedir. *Beceriler* üzerinde; okuryazarlık ve bilgi yeterliliği, yaşam boyu öğrenme gibi kazanımlarla artan istihdam ve yaşam kalitesinin artırılmasına fayda sağladığı belirtilmiştir. *Ekonomi* üzerinde ise; yerel refahın artırılması, bireysel ve toplumsal yoksullukla yüzleşme, eğitim ile boş zaman arasındaki ayrım üzerine bir köprü kurmak bakımından etkisinin olduğu ortaya koyulmuştur (Aabø, 2005: 207).

1.3. Otomasyon Sistemleri

Otomasyon, bir işlem ya da iş sürecinin otomatik araçlarla kendi kendini etkinleştirmesi, denetlemesi veya gerçekleştirmesidir (Levine-Clark ve Carter, 2013: 24). Otomasyon, insan tarafından gerçekleştirilen iş süreçlerinin veya insan tarafından yapılabilmesi artık mümkün olmayan işlemlerin makineler tarafından yerine getirilmesidir (Encyclopædia Britannica, 2020). Üretimin insan elinden kısmen ya da tamamen makinelere geçmesi otomasyonun temelini oluşturmaktadır. Sanayi devrimi sonrası üretim maliyetlerinin düşürülmek istenmesi, buna karşın üretimin artırılması ve sarf edilen emeğin azaltılması düşüncesi iş akışlarının gözden geçirilmesi ve insan

emeğinin organize edilmesini gerekli kılmıştır. Bu yeni üretim mantığı *seri üretim* kavramını ortaya çıkarmış, pek çok öncülü olmasına karşın seri üretim Henry Ford (Ford Motor Company) ile özdeşleşmiş ve fabrikada kurduğu hareketli montaj hattı büyük yankı uyandırmıştır. Ford'a göre seri üretim, yalnızca ürünün miktar olarak fazlaca üretilmesi değil aynı zamanda güç, doğruluk, ekonomi, sistem, süreklilik ve hız ilkeleri ile bir üretim projesine odaklanması hali olarak nitelendirmektedir (Encyclopædia Britannica, 2023). Otomasyon seri üretim kavramı ile büyük bir ivme kazanmış ve bugün sadece otomasyonun organize ettiği robotlardan müteşekkil üretim bantları ve fabrikalar ortaya çıkmıştır. Otomasyon hem insan-makine birlikteliğini hem de makine-makine birlikteliğini üretim alanında mümkün kılmaktadır.

Kütüphane otomasyonu, satın alma, kataloglama ve dolaşım gibi geleneksel kütüphane işlemlerini gerçekleştirmek için otomatik ve yarı otomatik veri işleme makinelerinin kullanılmasıdır (Salmon, 1975: 1). Kütüphane otomasyon sistemi, satın alma, muhasebe, dolaşım, kütüphaneler arası ödünç (ILL), referans, açık erişim katalog, kataloglama, süreli yayın yönetimi, raporlama ve yönetim gibi kapsamlı işlemleri bir arada yürütebilmektedir (Frontz, Alvey, 1982: 197). Kütüphane otomasyonu genel olarak kütüphanenin yürütmekte olduğu iş ve işlemleri bilgisayar yazılım ve donanımları veya makinelerce yürütülmesini ifade etmektedir. Endüstriyel sahada otomasyon sistemleri daha çok makineler ve robotlarla birlikte seri üretimi gerçekleştiriyor olsa da kütüphanelerde durum daha çok yazılım tarafında şekillenmektedir. Öte yandan otomasyona bağlı olarak çevre birimleri, kiosk birimler ve otomatik kütüphane sistemleri de kütüphanecinin müdahalesi olmadan iş süreçlerini yürütebilmektedir. Ancak ağırlıklı olarak yapılan işlemler kütüphane yazılımı üzerinden kütüphane paydaşlarının iş süreçlerini yerine getirmesine olanak sağlamaktadır.

Bilgisayarlar ve bilgi sistemlerinin erken kullanım örneklerinin verildiği sektörlerin başında kütüphaneler gelmektedir. Kütüphaneler ortaya çıktığı dönemden bu yana sahip olduğu bilgi ve belgelerin yönetim ve organizasyonu için çeşitli yöntem ve sistemler kullanmıştır. Bilgisayarın ortaya çıkması ile de kütüphane hizmetlerinin birer birer bu yeni sisteme geçişi sağlanmıştır. Bilgisayar ve kütüphane birlikteliğinin diğer sektörlerle nispeten erken başlaması her iki kavramın birbirine olan yakınlığı ile de ilişkilidir. Kütüphane, gerekli olduğunda tanımlama ve geri alma süreçlerini kolaylaştıran bir sistem altında bilgiyi depolama yeri olarak tanımlanabilir, ki bu aynı zamanda

bilgisayarın da tanımıdır (Martin, 1968: 711). Bu kavram benzerliği bilgisayarların kütüphanelere hızlıca uyarlanması ve hizmetlerin dijital ortama aktarılmasına bir nevi vesile olmuştur.

Herman Hollerith'in icadıyla 1890'lı yıllarda nüfus sayımında kullanılmaya başlanan delikli kartlar (Austrian, 1982: 45) 1930'larda University of Texas Library dolaşım birimine kurulan IBM makine ile ödünç iade işlemlerinin takibinde kullanılmaya başlanmıştır (Becker, 1964: 1007). Delikli kartlar dolaşım birimi dışında da rağbet görmüştür. Delikli kartlar Library of Congress tarafından 1950'de süreli yayın kataloğunda; 1951'de ise King County Public Library tarafından kitap kataloğunda kullanılmaya başlanmıştır (Dewey, 1959: 36). Daha sonraları delikli kartlar çeşitli üniversite ve halk kütüphanesi tarafından ödünç iade ve katalog olarak kullanılmıştır. Ancak analog bilgisayarların sınırlarının aşıldığı ve dijital bilgisayarların kullanılmaya başlanması ile kütüphane otomasyon sistemlerinin ilk örnekleri ortaya çıkmaya başlamıştır. National Library of Medicine kütüphane personelinin, tıp literatür listesinin otomatikleştirilmesi düşüncesinden hareketle başlattıkları proje ile dijital bilgisayar üzerinde çalışan ilk kütüphane otomasyon sistemi MEDLARS (Medical Literature Analysis and Retrieval System) 1964 yılında ortaya çıkmıştır (Rogers, 1964: 150). MEDLARS, 1879 yılından beri basılı olarak yayınlanan bibliyografik tıp dizini Index Medicus için yeni bir soluk getirmiş, tıp alanındaki bibliyografik verilerin işlenmesi süreçleri açısından kolaylık sağlamıştır (Adams, 1964: 173). MEDLARS, aramalarda Boolean işlemlerini (and, or, not) de desteklemekteydi (W. L. Miller, 1971: 261). Bu dönem bilgisayarların kütüphane hizmetlerinde kullanılması yönünden çeşitlilik kazanmıştır. University of California süreli yayınların kontrolü için; Illinois University 1961'de dolaşım hizmetlerinde kullanılması için; University of Toronto ise 1963'te katalogların bilgisayarda kullanılması için çalışmalara başlamıştır (Salmon, 1975: 4). Kansas City'deki The Shawnee Mission Public Schools 1968 yılında okulun bilgi işlemi tarafından gerçekleştirilen teşebbüsler ile otomatik kütüphane faaliyetlerine başlamıştır (Miller ve Hodges, 1971: 13). Geliştirilen sistem, toplu sipariş ve kataloglama modülleri ile hizmete girdikten sonra dolaşım modülü de eklenmiştir. Bölgedeki diğer okul kütüphanelerinin sisteme dahil edilmesi ve yeni ihtiyaçlar çevrimiçi sistem tasarlanmasını gerekli kılmış, 1970 yılında iki okul kütüphanesi ile pilot çalışmalara başlanmıştır. Kataloglama ve dolaşım hizmetlerinin çevrimiçi olarak yürütülmesi anlayışı bu proje ile

ilk kez ortaya atılmıştı (Kilgour vd., 1972: 158; Miller ve Hodges, 1971: 14) ve daha sonrasında geliştirilecek bütün kütüphane otomasyon sistemleri için çığır açıcı bulunmuştur.

Kütüphane otomasyon sistemlerinin bu erken dönem örneklerinin ardından ortaya çıkan ihtiyaçlar doğrultusunda bugün halen kullanılan kütüphane standartları da geliştirilmeye başlanmıştır. Library of Congress'e yönelik olarak Council on Library Resources, Inc. tarafından 1961 yılında kütüphane kaynaklarının kataloglama, arama, dizinleme ve iade süreçlerinin otomatize edilmesi amacıyla 100.000 dolarlık hibe programı başlatılmıştır (King vd., 1963: 1; McCallum, 2002: 37). Council on Library Resources, Inc. daha önce MEDLARS programını ve bir dizi kütüphane otomasyon projelerini de 1 milyon dolara yakın bütçe ile desteklemiştir (Council on Library Resources, 1973: 15). 1963 yılında Library of Congress kütüphanecileri ve bilgisayar uzmanlarından oluşan ekibin başlattığı çalışma ile aralarında University of Chicago, University of Toronto, Washington State Library ve Yale University'nin de bulunduğu 16 kamu, özel ve üniversite kütüphanesi pilot çalışmalara dahil edilmiştir (Avram, 1975: 5). Yürütülen çalışmanın amacı kütüphane işlemlerinin koordineli bir şekilde bilgisayarlar tarafından yürütülmesine olanak sağlayacak bir altyapının oluşturulmasıydı. Bu anlayışla ortaya çıkan projeye MARC (Machine-Readable Cataloging) ismi yani "makinelerce okunabilir kataloglama" verilmiş ve önceki projelerden farklı olarak yalnızca bir kütüphane içindeki bilgi sistemi üzerinden işlemlerin yürütülmesi değil aynı zamanda farklı kurumların ve dolayısıyla bilgi sistemlerinin birlikte çalışmasının önünü açacak proje üzerinde durulmuştur. Council on Library Resources MARC projesinin kütüphane otomasyon sistemlerinin başarılı bir şekilde kullanılmasında en önemli faktörlerden biri olduğunu vurgulamaktadır (Council on Library Resources, 1973: 16). Kütüphaneler temel bilgi arama işlemleri için çok uzun zamandan beri kart katalogları kullanmaktaydı. Fakat bilgisayarlar üzerinden kataloglama işlemlerinin yapılabilmesi ve ilerleyen yıllarda ortaya çıkacak kütüphane bilgi sistemlerine ön ayak olacak ve temel bileşen haline gelecek bir formatın ortaya çıkması oldukça önemli bir yer tutmaktadır.

Library of Congress'te düzenlenen ve komite, dernek, üniversite ve araştırma kurumlarının katıldığı konferansta MARC ile ilgili üç karar çıkmıştır;

- Library of Congress tarafından üretilen ve dağıtılan makinelerce okunabilir katalog kayıtları otomasyon sistemine sahip kütüphanelere yardımcı olacaktır,
- Makinelerce okunabilir kayıt Library of Congress'in kart kataloğunda halihazırda mevcut tüm verilerin yanı sıra çok amaçlı bir kayıt oluşturmak üzere ek bilgileri de içermelidir,
- Kütüphane camiasının ekseriyetinin, kayda dahil edilecek unsurlar konusunda mutabık kalması ve kaydın Library of Congress tarafından tasarlanması standardizasyonu sağlamanın en iyi yoludur (Avram, 1975: 6).

Pilot projenin devam ettiği 1967 yılında ortaya çıkan yeni isterler doğrultusunda projenin revizyonu gündeme gelmiş; özellikle uluslararası çok yönlü veri değiş tokuşunu mümkün kılan bir standardın önemi ortaya çıkmıştır. Bu sebeple proje devam ederken kitap, dergi, harita, müzik vb. tüm materyalleri kapsayacak şekilde bibliyografik bilgileri ve ilgili kayıtları içerebilen makinece okunabilir bir format olarak MARC II üzerinde çalışmalar başlatılmıştır (Avram, 1975: 10). Yapılan çalışmalar sonucunda Library of Congress ve American Library Association (ALA) tarafından 1969 yılında ilk MARC kılavuzu yayınlanmıştır (Library of Congress, 1969).

Kütüphane otomasyon sistemleri kurumsal, ticari ya da açık kaynak kodlu uygulamalardan meydana gelebilmektedir. Bir otomasyon sistemi kullanılmaya başlandığı andan itibaren kurumun sahip olduğu iş süreçleri bu sistemin kapasiteleri doğrultusunda şekillenmektedir. Sistem üzerindeki geliştirme ve değişiklikler de genelde kullanılan uygulamanın geliştiricisinin tasarrufundadır. Ancak kütüphanenin otomasyon sistemi üzerinde oluşturduğu kayıtlardan oluşan veri tabanının sahibi kurumun kendisidir. MARC formatı kütüphanenin hangi otomasyon sistemine sahip olduğunu veya kullandığını önemsemeksizin ortak bir veri kayıt formatı ortaya koymakta ve zaman içinde kurumun sistem değişikliğine ihtiyaç duyduğu anda sahip olduğu veri tabanını yeni otomasyon sistemine aktarımı noktasında önemli bir avantaj sağlamaktadır. MARC formatı kurulu sistemler arasında kütüphanelerin kayıt paylaşımına imkân vermesinin yanı sıra kurumun bir otomasyon sistemine ömür boyu bağımlı kalmasının da önüne geçmektedir. MARC kullanılmaya başlandığı yıllardan itibaren kütüphane camiasında evrensel bir format hüviyeti kazanmış ve USMARC, OCLC-MARC, UKMARC,

CANMARC, UNIMARC (Byrne, 1998: 11) gibi pek çok varyant ortaya çıkmış; buna karşın Library of Congress tarafından da kullanılan MARC21 en sık kullanılan format olmuştur.

MARC formatının kullanılmaya başlaması ile katalog kayıtlarının hızlıca bilgisayar ortamına aktarılıyor olması kütüphaneler arası etkileşim ve kaynak paylaşımını da yoğunlaştırmıştır. Ohio Eyaleti tarafından 1967 yılında kâr amacı gütmeyen bir kuruluş olarak yetkilendirilen The Ohio College Library Center (OCLC)'a (daha sonra adı Online Computer Library Center olarak değişmiştir) 1971-1972 yılında kırk dokuz kurum katılmış; akademik kütüphanelerdeki kaynakların kullanılabilirliğini artırmak ve kullanıcıların kütüphaneye olan maliyetini azaltmak üzere çalışmalarına başlamıştır (Kilgour vd., 1972: 157). OCLC tarafından 1970 yılında geliştirilen ilk sistem çevrimdışı katalog üretimine olanak sağlıyordu (Hopkins, 1973: 310). Otuz beş OCLC üye kütüphane bu katalog kartı üretim sistemine dahil olmuş ve IBM tarafından sağlanan katalog kartlarına kitaba ilişkin LC yer numarası veya yerel yer numarası işlenmiştir. Kayıtlar MARCII formatında veri tabanında tutulmuş ve iş birliği içindeki kütüphanelere gönderilmiştir. Çevrim dışı katalog sisteminin bir yılı aşkın kullanımının ardından 1971 yılında ulusal ve uluslararası standartlara uygun olarak Ohio akademik kütüphanelerine otomasyon desteği vermek üzere katot ışın tüpü terminaline sahip bir merkezi bilgisayar üzerinde çevrim içi katalog sistemi geliştirilmiştir (Kilgour vd., 1972: 157). Geliştirilen yeni sistem, katalog paylaşımı, süreli yayın kontrolü, teknik işlem, uzaktan katalog erişimi, dolaşım (ödünç-iade) ve konu-başlık erişiminin çevrimiçi olarak yürütülmesi mantığına dayanmaktaydı (Schoenung, 1974: 4). Daha önceden The Shawnee Mission Public Schools tarafından çevrimiçi katalog hizmeti sunulmuştu fakat bu sistem MARC formatında tasarlanmamıştı (Miller ve Hodges, 1971: 19). OCLC ise tüm sistemi MARC formatı üzerine kurgulamıştır. OCLC katalog sistemi Ohio Eyaleti'nin dışına hatta ülke dışına taşan bir popülerite kazanmış ve OCLC, Inc. şu an keşif aracı, kaynak paylaşımı, proxy, meta data hizmeti ve yönetim gibi dünya çapında pek çok otomasyon sistemi bulunan kurum haline gelmiştir. OCLC'nin en önemli ürünlerinden biri olan ve 1971'de ilk katalog kaydının sisteme eklenmesinin ardından ortaya çıkan WorldCat, 100 ülkeden 16.000'e yakın kütüphanenin bibliyografik verilerinin harmanlandığı ve kayıtlar üzerinden tarama yapılabildiği bir arama motoru olarak kitap, dergi, gazete, harita, dvd, müzik notası gibi materyallere ait yaklaşık 3,3 milyar kayıt içermektedir (OCLC, 2023).

Gelişmiş ülkelerde ve özellikle ABD’de 1930’lu yıllarda kütüphane otomasyon sistemlerinin ilk örnekleri verilmeye ve 1960’lı yıllarda kullanım çeşitliliği kazanmış olmasına karşın Türkiye’de bilgisayar sistemlerinin kütüphanelerde uygulanabilmesi 1970’li yılları bulmuştur. Soysal (1965) kütüphane otomasyon sistemlerinin gerekliliği konusunda yaptığı araştırmada kitap ve dergilerin artan basım sayılarına (öncesine nispeten) vurgu yapmakta ve bu “muazzam dalga” ile kütüphanelerin geleneksel kataloglama yöntemleriyle başa çıkamayacağını belirtmektedir. Öte yandan araştırmanın yapıldığı yıllarda uluslararası arenada hizmete giren kütüphane bilgi sistemleri hakkında bilgi vermekte ve araştırmacılar açısından kart kataloglar yerine bilgisayar üzerindeki bir katalogdan tarama yapmanın kolaylığını örneklerle açıklamaktadır. Kütüphanelerde otomasyon sistemlerinin tartışıldığı bu yıllarda Çelik (1972: 231) bilhassa Türkiye’de kütüphanelerin otomasyon kullanımının önündeki kritik etkenlere vurgu yapmaktadır; maddi olanaklar, kullanım, personel. 2020’li yıllarda bir bilgi teknolojisine bireysel veya kurumsal olarak sahip olmak çok büyük külfetler gerektirmiyor fakat bilgisayar sistemlerinin yeni yeni ortaya çıktığı ve geliştiği 1970’li yıllarda bırakın bireysel olarak bilgisayara sahip olmayı bir kurumun dahi yatırım yapıp bilgisayarı envanterine sokması ciddi bir bütçe gerektiriyordu. Kütüphanelerin de otomasyona geçişi için hem bilgisayara sahip olması hem de koleksiyonu otomasyon için elden geçirmesi maddi olanakları gerekli kılmaktaydı. İkincil kısıt olarak kullanım olasılığını ön plana çıkarmıştır. Kurumun maddi zorluklar ile yatırım yapıp yeni bir ürün ve hizmet ortaya koyduktan sonra kütüphane kullanıcılarının (okuyucu, müşteri) bu yeni teknolojiye hazır olmayışı veya ihtiyaç duymayacak olması yatırımın boşa gideceği anlamına gelmektedir. Çelik çalışmasında en önemli sorun olarak personeli göstermiştir. Gerek ülkede yetişmiş teknik personelin (kütüphaneci, programcı) bulunmaması gerekse var olan personelin kütüphanenin otomasyon sistemine geçişi sonrası işsiz kalacağı düşüncesi bilgi teknolojilerine yönelik karşıt görüş oluşturduğunu vurgulamaktadır.

Bu ortamda Türkiye’de ilk kütüphane otomasyon sistemi girişimleri Hacettepe Üniversitesi Tıp Merkez Kütüphanesi ve Bilgi İşlem Merkezi’nin ortak teşebbüsleri sonucu 1970 yılında kitap sipariş ve izleme işlemleri; 1971 yılında ise süreli yayınlar kataloglama işlemleri bilgisayar üzerinden gerçekleştirilmeye başlanmış; 1972 yılında ise Hacettepe Üniversitesi Kütüphaneleri Kitap ve Süreli Yayın Dolaşım İzleme Sistemi geliştirilmiştir (Tunçkanat vd., 1988: 110; Baysal, 1987: 173). Büyük beklentilerle ortaya

koyulan bu sistemin ömrü kısa sürmüş, daha kapsamlı yeni bir kütüphane otomasyon sistemi 1987 yılında duyurulmuştur (Dikeç, 1988: 26). Hacettepe Üniversitesi tarafından gerçekleştirilen bu atılımlar konusunda Gürsoy'un yaptığı iki çalışma (Gürsoy, 1975, 1978) oldukça önemlidir. Gürsoy yüksek lisans tezi çalışmasında süreli yayınlar kataloğunun bilgisayar üzerinden yürütülmesinin gerekçelerini, yöntemini ve geliştirilen program hakkında yazılım örneklerini vermiştir. Doktora tezi çalışmasında ise kitap ve süreli yayınlar için dolaşım, arama, süre uzatımı, ayırtma, ceza gibi iş süreçlerinin gerçek zamanlı olarak bilgisayar üzerinden gerçekleştirilmesinin faydaları ve uygulaması hakkında detaylar vermiştir. Öte yandan bu çalışmada gelişmiş ülkelerde kütüphane otomasyon sistemlerinin geldiği nokta hakkında kapsamlı bilgiler vermiştir. Hacettepe Üniversitesi'nin gerçekleştirdiği çalışmaların dışında Orta Doğu Teknik Üniversitesi 1970 yılında süreli yayınlar kataloğunu; Türkiye Bilimsel ve Teknik Dokümantasyon Merkezi (TÜRDOK) 1968 yılında çalışmalara başlamış ve 1977 yılında Ankara Süreli Yayınlar Kataloğunu; Çukurova Üniversitesi 1982 yılında kitap kataloğunu bilgisayar üzerinden hizmete sunmuştur (Baysal, 1987: 189–190; Dikeç, 1988: 26–27; Gürsoy, 1978: 22–23; Sağdıç, 1987: 49–50). Özellikle 1980'li yılların sonu itibariyle üniversite kütüphanelerinde bilgisayar ve otomasyon kullanımı açısından büyük bir gayret gösterildiği aşıkardır. Ancak burada göze çarpan husus gelişmiş ülkelerdeki örneklerin aksine her üniversitenin kendi iç imkanları doğrultusunda bireysel iştiraklerin ortaya koyulması ve kurumlar arası iş birliğinin gerçekleştirilememiş olmasıdır. Belki bu sebeptendir ki geliştirilen sistemler uzun ömürlü olmamıştır. 1987 yılında bu noktaya parmak basmak üzere Yükseköğretim Kurulu Dokümantasyon ve Uluslararası Bilgi Tarama Merkezi üniversite kütüphane ve dokümantasyon daire başkanlarının katılımıyla kütüphane otomasyon sistemleri konusunda iş birliğine dayalı toplantılar yapılmıştır (Dikeç, 1988: 31). 1980'li yılların sonunda ortaya çıkan ve daha önceki örneklerine göre çok daha geniş kullanım alanı ve süre yakalayan otomasyon sistemi Bilkent Üniversitesi tarafından geliştirilmiştir. Şubat 1988 yılında başlatılan proje ile cobol programlama dilinde geliştirilen BLISS (Bilkent Library Information Services System) kataloglama, OPAC, dolaşım, sayım, istatistik gibi modüllere sahipti (Ünal Orhan, 2006: 73) ve MARC formatına uygun olarak hazırlanmıştı (Erdoğan, 2011: 90). Gelişim sürecinde PC, Unix ve Linux sürümleri de hizmete giren BLISS 18'i üniversite olmak üzere 36 kütüphane tarafından kullanılmıştır. Bugün Bilkent Üniversitesi dahil BLISS kullanan kütüphaneler

otomasyon sisteminin güncelliğini yitirmesi ve yeterli gelişim göstermemesi sebebiyle farklı otomasyon sistemlerine geçiş yapmıştır. 1986 yılında Anadolu Üniversitesi'nin IBM 4381 mainframe makine üzerinde başlattığı çalışmalar sonucu KYBELE (Kütüphane Yönetim ve BELge Erişim) Sistemi hayata geçirilmiştir (Yılmaz, Aslan, 1992: 11). PL/I programlama dilinde ve USMARC formatına uygun olarak tasarlanan sistem kataloglama, sağlama, dolaşım, OPAC, süreli yayın ve yönetim gibi modüllerden meydana gelmekteydi (Ünal Orhan, 2006: 76). Başta geliştiricisi Anadolu Üniversitesi olmak üzere İnönü Üniversitesi, Uludağ Üniversitesi, Ondokuz Mayıs Üniversitesi, Haliç Üniversitesi, MEB Eğitim Teknolojileri Genel Müdürlüğü, TRT Genel Müdürlüğü, İslam Ülkeleri İstatistiksel Ekonomi ve Sosyal Araştırma ve Eğitim Merkezi (M. Yılmaz, 2006: 149) gibi pek çok devlet ve vakıf üniversitesi kütüphanesi tarafından kullanılan KYBELE 2010 yılından itibaren LIBRA Kütüphane Otomasyon sistemi olarak hayatına devam etmektedir (LIBRA, 2023; Aslan, 2022). KYBELE'nin LIBRA'ya dönüşümü yalnız isim üzerinde olmamış, geliştirici ekibin Anadolu Üniversitesi'nden ayrılması ile bağımsız bir firma altında köklerine bağlı ve güncel kütüphane isterleri çerçevesinde MARC21, z39.50, rfid, sip gibi formatları destekleyen yeni bir ürün olarak geliştirilmiş.

1990'lı yıllardan itibaren Türkiye'de Milas, İskenderiye, Kâşif, Ufuk, Librid, Millenium, SirsiDynix, Koha, Yordam gibi pek çok yerli ve yabancı kütüphane otomasyon sistemi kullanılmıştır. Her otomasyon sisteminin kendine göre avantaj ve dezavantajları olmasıyla birlikte açık kaynak kodlu ve açık erişimli otomasyon sistemi Koha özellikle Kültür ve Turizm Bakanlığı'nın yapmış olduğu çalışma sonucu bakanlığa bağlı halk kütüphanelerinde kullanılması ile ciddi bir kullanım oranına ulaşmıştır. Halk kütüphanelerinin dışında devlet ve vakıf üniversiteleri, kurum kütüphaneleri Koha kullanmaktadır. Bunun yanı sıra ticari bir ürün olan Yordam Kütüphane Otomasyon Sistemi üniversite kütüphaneleri tarafından yaygın bir kullanım oranına sahiptir (Tablo 1). Bugün itibarıyla Türkiye'de devlet ve vakıf üniversite kütüphanelerinin toplam 146 tanesi Yordam kütüphane otomasyon sistemini kullanırken 21 tanesi de açık kaynak kodlu Koha kütüphane otomasyon sistemini kullanmaktadır. Salihoğlu'nun (2012: 19) yaptığı çalışmada Türkiye'deki 164 üniversiteden 95'inin Yordam, 2 tanesinin Koha kullandığını; 30'unun ise herhangi bir otomasyon sistemi kullanmadığını belirtmiştir. Aradan geçen zaman içinde açık kaynak kodlu otomasyon sistemi Koha'nın hem devlet hem de vakıf üniversitelerinde belirgin bir yükselişi olduğu göze çarpmaktadır. Öte

yandan Milas, BLISS, Ufuk, Symphony gibi otomasyon sistemlerinin artık kullanım alanı bulamadığı ortaya çıkmaktadır.

Tablo 1: Üniversite Kütüphanelerinde Kullanılan Otomasyon Sistemleri

Otomasyon Sistemi	Devlet Üniversiteleri		Vakıf Üniversiteleri		Toplam	
	N	%	N	%	N	%
Yordam	98	77,17	49	65,33	147	72,77
Koha	8	6,3	13	17,33	21	10,4
SirsiDynix	9	7,09	7	9,33	16	7,92
Libra	6	4,72	0	0	6	2,97
Millenium	3	2,36	2	2,67	5	2,48
Sierra	2	1,57	2	2,67	4	1,98
Librid	1	0,79	1	1,33	2	0,99
Üniversite Otomasyonu	0	0	1	1,33	1	0,5
Toplam	127	100	75	100	202	100

Üniversite kütüphaneleri tarafından kullanılan otomasyon sistemlerinden Yordam, Libra ve Librid yerli sistemler olmakla birlikte Koha, SirsiDynix, Millenium ve Sierra yabancı menşeli ithal sistemlerdir. Tablo 1 incelendiğinde Türkiye’deki üniversite kütüphanelerinin yerli sistemleri kullanmaya özen gösterdiği anlaşılmaktadır. Burada kütüphanelerin bütçeleri ve ithal sistemlerinin yüksek kurulum ve işletim maliyetleri kurumların kararını önemli ölçüde etkilediği düşünülmektedir. İthal bir sistem olmasına karşın açık kaynak kodlu ve açık erişimli Koha otomasyon sistemi hem özelleştirilebilir yapısı hem de ücretsiz kullanım imkânı ile öne çıkmaktadır. Kurumlar Koha’yı iç imkanları kapsamında ücretsiz kullanabilmelerinin yanı sıra dış kaynaklardan hizmet alımı ile kullanım imkânı da bulabilmektedir. 2012 yılındaki durum ile karşılaştırıldığında otomasyon sistemi kullanmayan üniversite kütüphane sayısının 30’dan 1’e düşmüş olması bir diğer önemli gelişmedir.

Yükseköğretim Kurulu’nun sayfasında(Yükseköğretim Kurulu, t.y.) yer alan 208 üniversiteden 202 üniversite kütüphanesinin kullandığı otomasyon sistemi tespit edilmişken 1 üniversite kütüphanesinin otomasyon sistemi kullanmadığı belirlenmiş; 5 üniversite kütüphanesinin ise web sitesi ve iletişim adresleri üzerinden bilgi temin edilememiştir. Bu sebeple 202 üniversite kütüphanesine ilişkin bilgiler tabloya yansıtılmıştır.

Kütüphane otomasyon sistemleri Türkiye’de üniversite ve halk kütüphanelerinin haricinde belediye, bakanlık, kurum, okul ve özel kütüphaneler tarafından da kullanılmaktadır. Bu tür kütüphaneler merkezi bir organizasyona bağlı olmadığı için Türkiye geneli toplam otomasyon sistemi kullanımı hakkında net bilgi verilememekte; her kurum kendi ihtiyaçları doğrultusunda ücretli veya ücretsiz sistemleri kullanabilmektedir.

Dünya genelinde üniversite kütüphanelerinde kullanılan kütüphane otomasyon sistemlerine bakıldığında (Breeding, 2023) ExLibris firmasının ürünü Alma en fazla kullanılan kütüphane otomasyon sistemi olarak göze çarpmaktadır. Koha ise ikinci en fazla kullanılan sistem olarak ortaya çıkmaktadır (Tablo 2).

Tablo 2: Dünya Geneli Üniversitelerde Kullanılan Otomasyon Sistemleri

Otomasyon Sistemi	N	%	Otomasyon Sistemi	N	%
Alma	1696	29,41%	Verso	33	0,57%
Koha	733	12,71%	Spydus	28	0,49%
Worldshare Man. Ser.	498	8,64%	Slima-St	27	0,47%
Sirsidynix Symphony	461	7,99%	Libraryworld	25	0,43%
Sierra	409	7,09%	Totals II	23	0,40%
Aleph	211	3,66%	Cobiss	22	0,38%
Folio	141	2,44%	Cds/Isis	20	0,35%
Yordambt	116	2,01%	Mandarin	20	0,35%
Virtua	103	1,79%	Sebina Openlibrary	20	0,35%
Iliswave	102	1,77%	Ilas	17	0,29%
Horizon	96	1,66%	Prolib	17	0,29%
Library.Solution	69	1,20%	Bibliovation	16	0,28%
Neo Sirius	55	0,95%	Adis/Bms	15	0,26%
Voyager	51	0,88%	Hylib	15	0,26%
Destiny	47	0,81%	V-Smart	15	0,26%
Opals	47	0,81%	Absys	14	0,24%
Carin-İ	45	0,78%	Liberty	14	0,24%
Atrium	41	0,71%	Library Management Cloud	14	0,24%
Eos.Web	41	0,71%	M7 Bhc Library System	14	0,24%
Libsys	41	0,71%	Sowa	14	0,24%
E-Cats Library	40	0,69%	Locally Developed	13	0,23%
Polaris	37	0,64%	Olib	13	0,23%
Solars	37	0,64%	Volcano-I	13	0,23%
Millennium	36	0,62%	Lbs Lokaal Bibliotheek Sys	12	0,21%
Nalis	36	0,62%	Toread	11	0,19%
Xmlas	36	0,62%	Abcd	10	0,17%
Limedio	35	0,61%	Cybertools For Libraries	10	0,17%
Evergreen	33	0,57%	Alexandria	9	0,16%

Koha'nın, pazar içinde köklü ve hatırı sayılır ürünlerin arasından sıyrılarak rekabetçi bir kullanım oranı yakalaması açık kaynak kodlu sistemlerin geleceği açısından umut verici bir durumdur. Ayrıca yine Folio açık kaynak kodlu bir sistem olarak önemli bir kullanım miktarına sahiptir. Folio, Library of Congress tarafından kütüphanenin fiziksel ve dijital koleksiyonlarının yönetilmesi için yeni otomasyon sistemi olarak belirlenmiştir (Library of Congress, 2022a). Tablodaki veriler 1994 yılından bu yana kütüphaneler hakkında kapsamlı bilgiler barındıran Library Technology Guides sayfasından temin edilmiştir. Verilen sayılar kesin olarak dünyadaki durumu yansıtmasa da çeşitli coğrafyalardan 5767 üniversite kütüphanesinin kullandığı otomasyon sistemi hakkında bilgi vermektedir.

1.4. Kütüphane Otomasyon Sistemi Standartları

Bilgi sistemlerinin ve de kütüphane otomasyon sistemlerinin ortaya çıkışı ve gelişimine bakıldığında sistemlerin küçük ölçekli kullanım senaryosuna bağlı meziyetler ile geliştirildiği ve temel ihtiyaçların göz önüne alındığı görülmektedir. Aradan geçen zaman ile gelişen teknoloji, küreselleşen dünya ile bilgi teknolojileri, birbirinden bağımsız sistemlerin birlikte çalışabilirliğini gerektirmiştir. Eğitim, bankacılık, sanayi, bilim, teknoloji gibi hayatın her alanında kullanılan bilgi teknolojilerinin aynı dili konuşması, aynı altyapıya sahip olması sistemlerin sürdürülebilirliği açısından elzemdir. Kütüphane otomasyon sistemlerinin de diğer yazılım ve çevre birimleri ile çalışabilirliği açısından bazı temel standartları desteklemesi beklenmektedir. Bunların bazıları aşağıda irdelenmiştir.

Daha önce bahsedildiği üzere kütüphane otomasyon sistemleri arasındaki uyumu sağlayan en belirgin standartlardan birisi MARC formatıdır. MARC, kurulu sistemler arasındaki veri alışverişini kolaylaştırdığı gibi kurumun sistem değişikliği yapmak istediğinde verilerinin transferini de mümkün kılmaktadır. MARC aynı zamanda ISO 2709 standartıdır (Committee on Technical Standards for Library Automation, 1974: 134; Hopkinson, 1984: 13). ISO 2709 ile ANSI Z39.2 (ANSI, 2016) muadil standartlardır (IFLA, t.y.; Library of Congress, 2000). ISO 2709, bibliyografik tanımlamaya uygun tüm materyallerin kayıtlarını genel bir değişim formatı için zorunlu gereksinimleri ortaya koymaktadır (International Organization for Standardization, 2008). Bu standart kayıtların uzunluğu veya içeriğini tanımlamamakla birlikte etiketlerin, göstergelerin anlamını da vermez; bu özellikler ilgili format tarafından belirlenmektedir. MARC

formatı ISO 2709'a dayanarak geliştirilmiş ve kütüphane otomasyon sistemlerinin isterleri doğrultusunda bibliyografik kayıtların tanımlamaları gerçekleştirilmiştir. ISO 2709, geliştirilmiş bir yapıyı, özellikle de veri işleme sistemleri arasındaki iletişim için tasarlanmış ve sistemlerin içinde işleme formatı olarak kullanılmayan çerçeveyi tanımlamaktadır (Galabova, Trencheva ve Trenchev, 2009: 231; Morton, 1986: 26). Standardın formüle edildiği 1970'li yıllarda depolama maliyetlerinin pahalılığını yansıtacak şekilde manyetik bantlar referans alınmış ve manyetik bandın işlenmesi için özel olarak geliştirilmiştir (Hopkinson, 1984: 14). ISO 2709 kaydında dört bölüm vardır (Şekil 2); *record label* (kayıt etiketi), *directory* (dizin), *data field* (veri alanları), *record separator* (kayıt ayırıcı) (International Organization for Standardization, 2008).

Record Label
Directory
Data Fields
Record Separator

Şekil 2: ISO 2709 Yapı Modeli

Kayıt etiketi, kaydın uzunluğunu ve kayıta yer alan verilerin temel adreslerini içerir. Dizin, kayıttaki alanlara giriş yapılan konumları sağlar. Veri alanları, kayıttaki tüm alan ve alt alanları içeren bölümdür. Kayıt ayırıcı, kaydın son karakterini ifade etmekle birlikte kaydın son veri alanının alan ayırıcısını takip eder (Galabova, Trencheva, Trenchev, 2009: 232). MARC bir ISO 2709 ve ANSI Z39.2 uygulaması olduğu için kayıt bölümleri benzerdir.

```
01142cam 2200301 a 4500
00100130000000300040001300500170001700800410003401000170007502000250009204000180
01170420009001350500026001440820016001701000032001862450086002182500012003042600
05200316300004900368500004000417520022800457650003300685650003300718650002400751
650002100775650002300796700002100819- 92005291 -DLC-19930521155141.9-920219s1993
caua j 000 0 eng - a 92005291 - a0152038655 :c$15.95- aDLCcDLCdDLC- alcac-00-
aPS3537.A618bA88 1993-00a811/.52220-1 aSandburg, Carl,d1878-1967.-10aArithmetic /cCarl
Sandburg ; illustrated as an anamorphic adventure by Ted Rand.- a1st ed.- aSan Diego :bHarcourt
Brace Jovanovich,cc1993.- a1 v. (unpaged) :bill. (some col.) ;c26 cm.- aOne Mylar sheet included
in pocket.- aA poem about numbers and their characteristics. Features anamorphic, or distorted,
drawings which can be restored to normal by viewing from a particular angle or by viewing the
image's reflection in the provided Mylar cone.- 0aArithmeticxJuvenile poetry.- 0aChildren's poetry,
American.- 1aArithmeticxPoetry.- 1aAmerican poetry.- 1aVisual perception.-1 aRand, Ted,eill.-
```

Şekil 3: MARC Kaydı Ham Veri

Bir MARC kaydı üç ana bölümden oluşur (Şekil 3, Şekil 4); *leader (başlık)*, *directory (dizin)* ve *variable fields (değişken alanları)* (Library of Congress, 1996). Başlık (Şekil 3-4 yeşil), kaydı işlemek için gerekli parametreleri içerir. Dizin (Şekil 3 kırmızı), kayıttaki her bir alanın etiketini, başlangıç konumu ve uzunluğunu barındırır. Değişken alanları (Şekil 3-4 mavi), kayda ilişkin veri içeriğinin bulunduğu bölümdür. Değişken kontrol ve veri alanları ile pek çok alt alan ve etiket barındırır.

```
01142cam 2200301 a 4500
000
001 92005291
003 DLC
005 19930521155141.9
008 920219s1993 caua j 000 0 eng
010 |a 92005291
020 |a0152038655 :|c$15.95
040 |aDLC|cDLC|dDLC
042 |alcac
050 00|aPS3537.A618|bA88 1993
082 00|a811/.52|220
100 1 |aSandburg, Carl,|d1878-1967.
245 10|aArithmetic /|cCarl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.
250 |a1st ed.
260 |aSan Diego :|bHarcourt Brace Jovanovich,|cc1993.
300 |a1 v. (unpaged) :|bill. (some col.) ;|c26 cm.
500 |aOne Mylar sheet included in pocket.
520 |aA poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.
650 0|aArithmetic|xJuvenile poetry.
650 0|aChildren's poetry, American.
650 1|aArithmetic|xPoetry.
650 1|aAmerican poetry.
650 1|aVisual perception.
700 1 |aRand, Ted,|eill.
```

Şekil 4: MARC Kaydı

MARC kaydının html gösterimi (Şekil 4) son kullanıcı olmasa da kütüphane profesyonellerinin sıkça karşılaştığı bir yapıdır. Bu yapı, bibliyografik kaydın tüm alanlarını ihtiva ettiği gibi alanların başında olan “245” şeklinde ifade edilen etiketleri, “|a” şeklinde ifade edilen alt alanları ve göstergeleri barındırmaktadır. Şekil 4’te verilen MARC yapısı Library of Congress’in de kullandığı MARC21 varyasyonuna aittir. MARC formatında oluşturulan bir kayıt diğer tüm uyumlu sistemler tarafından okunabilir ve dönüştürülerek kendi format yapısına uygun hale getirilebilir. Bu yapı sayesinde ortak bir dil kullanılmakta, veri alışverişi mümkün hale gelmektedir.

ISO 2709 standardına dayanan bir diğer bilgi erişim standardı ANSI Z39.50'dir (ANSI, 2015). ANSI Z39.50, editoryal farklılıklar dışında teknik olarak ISO 23950 (International Organization for Standardization, 1998) ile aynı standarttır (Library of Congress, 2015). 1980'lerde OCLC, Library of Congress, Washington Library Network (WLN) ve Research Libraries Group (RLG)'un ortak girişimi olarak hayata geçmiş; bilgi alışverişi için ağ üzerinde birbirine bağlı bilgisayar sistemleri tarafından kullanılan uluslararası bir standart protokoldür (National Information Standards Organization, 2002: 3-5). Z39.50 standardı, bilgisayar sistemi içindeki uygulamayı ifade etmediği gibi kısıtlamamakta; veri değiş tokuşu için kural ve protokollerin uygulanması için gereksinimleri ortaya koymakta; bağlantı yönelimli programdan programa iletişimi ele almaktadır (ANSI, 2015: 1). Bir istemci makinenin sunucu makinedeki veri tabanında sorgu çalıştırmasına ve sorgu sonucu tanımlanan kayıtları istemcinin almasına izin veren veri yapılarını ve değişim kurallarını belirleyen protokoldür (Lynch, 1997). Bilgi merkezleri, üniversiteler, kütüphaneler ve ortak katalog merkezlerini destekleyen sistemler için tasarlanmıştır. Kütüphanecilerin, bibliyografik kayıt oluştururken en sık faydalandığı araçlardan biridir. Z39.50 sayesinde uyumlu otomasyon sistemleri arasında basitçe veri transferi mümkün olmaktadır. Özellikle otorite kütüphanelerce oluşturulmuş katalog kayıtları diğer kütüphaneler tarafından çekilerek sisteme dahil edilmektedir. Böylelikle hem kayıtlarda standardizasyon sağlanmış hem de veri girişinde zaman tasarrufu elde edilmiş olur. Kütüphaneler farklı otomasyon sistemi kullanabilmekte ve bu otomasyon sistemlerinin her biri kendine özgü arayüzlere sahip olabilmektedir. Kütüphane profesyonelleri bu sistemler arasında ihtiyaç duyduğu veriyi elde edebilmek için sunucu sistemin arayüzüne hâkim olması ve nasıl kullanıldığını bilmesi gerekmektedir. Z39.50'nin amacı da kayıt sunucusu sistemin nasıl çalıştığını bilmek zorunda kalmadan verileri arayabilme ve transfer etme kolaylığı sağlamaktır. Kullanıcı, farklı komut veya arama yöntemi öğrenmek zorunda kalmadan kendi sistemi üzerinden birden fazla uzak sisteme erişim sağlar. Otomasyon sistemlerinin, işletim sistemlerinin veya donanımların aynı olması zorunluluk değildir. Farklı bilgisayar sistemlerinin z39.50 altyapısına sahip olması sistemler arasında veri transferi için yeterlidir. Başlangıçta Open System Interconnection (OSI) protokollerini kullanan bilgisayar iletişim ağlarında kullanılmak üzere geliştirilen z39.50, artık internet ortamında başarıyla çalışmaktadır (Haravu, 2006: 10).

Kütüphane otomasyon sistemleri, diğer kütüphane otomasyon sistemleriyle veri alışverişi yaptığı gibi web kaynakları veya kütüphane harici sistemlerden de veri takası yapabilmektedir. ISO 8879 (SGML) standartına (International Organization for Standardization, 1986) dayanan XML (Extensible Markup Language- Genişletilebilir İşaretleme Dili), web ve diğer bilgi sistemleri arasında veri alışverişini sağlayan basit ve esnek bir metin biçimidir (Quin, 2016).

```
<?xml version="1.0"?>
<catalog>
  <book id="bk101">
    <author>Gambardella, Matthew</author>
    <title>XML Developer's Guide</title>
    <genre>Computer</genre>
    <price>44.95</price>
    <publish_date>2000-10-01</publish_date>
    <description>An in-depth look at creating applications with XML.</description>
  </book>
</catalog>
```

Şekil 5: XML Örneği

XML, yapılandırılmış verileri web uygulamalarına iletmek için özel olarak tasarlanmış standart bir metin biçimidir (Khare ve Rifkin, 1997: 80). Kendisi gibi bir SGML uygulaması olan HTML’de olduğu gibi XML’de de öğeler etiketlerle ayrılarak sıralanmaktadır. Şekil 5’te bir kitap kataloğu uygulaması için yazılmış kodda kitaba ait yazar, başlık, fiyat, tür, yayın tarihi ve tanım bölümlerinin bulunduğu basit bir XML örneği verilmiştir. Etiketleri kitap kataloğunu hazırlayan geliştirici projenin ihtiyaçlarına uygun olarak bağımsız bir şekilde belirlemiştir. Etiketler incelendiğinde “catalog” ve “book id” etiketlerinden bu kodun bir kitap kataloğuna ait olduğunu, Matthew Gambardella’ ya ait XML Developer's Guide adlı kitabın 2000-10-01 tarihinde yayınlandığını, bilgisayar türünde ve 44,95 birim fiyata sahip olduğu anlaşılmaktadır. HTML’in aksine isteğe bağlı bir etiket yapısı bulunduğu için her uygulama için tanımlanabilir bir format olarak her biri birkaç nitelik-değer çifti içerebilen, iç içe geçmiş açma-kapama etiket kümelerinden oluşan sistematik bir hiyerarşi oluşturur (Decker vd., 2000: 65). Bu özellik XML’i HTML’den ayıran ve aslında XML’in doğuşuna sebep olan uyarlanabilme ihtiyacından ileri gelmektedir. Geliştiriciler, XML’de projenin ihtiyaçları doğrultusunda kendi yapılarını ve etiketlerini oluşturarak ana yapıdan bağımsız bir format oluşturabilmektedir. World Wide Web Consortium (W3C) tarafından 1996 yılında geliştirilen XML için şu tasarım hedefleri belirlenmiştir (Bray vd., 2008):

- İnternet üzerinden doğrudan kullanılabilir olacaktır
- Çeşitli uygulamaları destekleyecektir
- SGML ile uyumlu olacaktır
- XML belgelerini işleyen program geliştirmek kolay olacaktır
- XML'deki isteğe bağlı özellik sayısı, mümkün olan en az sayıda, ideal olarak da sıfırda tutulmalıdır
- XML belgeleri insan tarafından okunabilir ve net olmalıdır
- Tasarımı hızlı bir şekilde hazırlanmalıdır
- Tasarımı muntazam ve kısa-öz olacaktır
- Belgelerin oluşturulması kolay olacaktır
- İşaretlemedeki vecazet az bir öneme sahiptir.

XML'in gelişimi ve kendine geniş çaplı kullanım alanı bulması kütüphanecilik uygulamalarında da kendine yer bulmuştur. MARC formatının haricinde yeni bir yönelime gidilmesinin temel nedeni MARC'ın bilgisayar teknolojisinin gelişmeye başladığı ilk dönemde ortaya çıkmış olmasıdır. MARC, bellek, depolama ve işlem gücünün sınırlı ve pahalı olduğu bir dönemde geliştirilmiş, bu dönemin imkanları doğrultusunda şekillenmiştir (Tennant, 2002: 26). Tennant (2002) makalesinde MARC'ın modasının geçmiş olduğunu ve artık o eski kullanışlılığının bulunmadığını vurgulamıştır. XML ise kütüphanecilerin içgüdüsel olarak dikkatini çeken güç, hassasiyet, kontrol ve esneklik özelliklerini barındırır (Miller, 2000: 17). XML'e olan ilgi, eldeki bibliyografik kayıtların daha geniş bir dolaşımı ve kullanımını sağlamak, bibliyografik verilerin görüntüleme formatlarında (html, pdf, text) esneklik sağlanması talebi ve ISO 2709'un kullanım dışı kalıyor olması gibi nedenlere de dayanmaktadır (Ramos de Carvalho vd., 2004: 131). MARC ve XML kavramları kütüphanecilik camiasında çokça tartışıldıktan sonra (Carvalho, Cordeiro, 2002; B. C. Johnson, 2001; Miller, 2000) Library of Congress' in gerçekleştirdiği çalışmalar sonucunda 2002 yılında (Radebaugh, 2007: 15) XML ortamında MARC verileriyle çalışmaya yönelik kullanıcıların kendi ihtiyaçları doğrultusunda çalışmasına olanak sağlayacak şekilde esnek ve genişletilebilir bir çerçeve ortaya koymuştur (Library of Congress, 2008). Library of Congress Network Development and MARC Standards Office, MARCXML tasarımında dikkat edilecek hususları altı başlık altı toplamış ve ortaya atılan XML çerçevesinin bu bileşenlere sahip olacağını belirtmiştir (Library of Congress, 2022b):

- Basit ve esnek bir MARC XML şeması
- MARC'ın XML'e kayıpsız şekilde dönüştürülmesi
- XML'den MARC'a geri dönülebilirlik
- Veri sunumu
- MARC alanlarının düzenlenebilmesi
- Veri dönüşümü
- MARC verilerinin doğrulanması
- Genişletilebilirlik

Şekil 6'da XML yapısının bir hiyerarşi oluşturduğunu ve katmanlar altında istenen alanların verildiği anlaşılmaktadır. Daha önce de orijinal MARC kaydı olarak verilen ve Carl Sandburg'e ait olan Arithmetic adlı esere ait bibliyografik kaydın XML gösteriminde MARC kayıt alanları "datafield" etiketleri ile verilmiş, ardından alt alanlar "subfield" etiketini ile katalog verilerini içerecek şekilde sıralanmıştır.

<pre> <collection xmlns="http://www.loc.gov/MARC21/slim"> <record> <leader>01142cam 2200301 a 4500</leader> <controlfield tag="001"> 92005291 </controlfield> <controlfield tag="003">DLC</controlfield> <controlfield tag="005">19930521155141.9</controlfield> <controlfield tag="008">920219s1993 caua j 000 0 eng </controlfield> <datafield tag="010" ind1=" " ind2=" "> <subfield code="a"> 92005291 </subfield> </datafield> <datafield tag="020" ind1=" " ind2=" "> <subfield code="a">0152038655 :</subfield> <subfield code="c">\$15.95</subfield> </datafield> <datafield tag="040" ind1=" " ind2=" "> <subfield code="a">DLC</subfield> <subfield code="c">DLC</subfield> <subfield code="d">DLC</subfield> </datafield> <datafield tag="042" ind1=" " ind2=" "> <subfield code="a">lcac</subfield> </datafield> <datafield tag="050" ind1="0" ind2="0"> <subfield code="a">PS3537.A618</subfield> <subfield code="b">A88 1993</subfield> </datafield> <datafield tag="082" ind1="0" ind2="0"> <subfield code="a">811/.52</subfield> <subfield code="2">20</subfield> </datafield> <datafield tag="100" ind1="1" ind2=" "> <subfield code="a">Sandburg, Carl,</subfield> <subfield code="d">1878-1967.</subfield> </datafield> <datafield tag="245" ind1="1" ind2="0"> <subfield code="a">Arithmetic </subfield> <subfield code="c">Carl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.</subfield> </datafield> <datafield tag="250" ind1=" " ind2=" "> <subfield code="a">1st ed.</subfield> </datafield> <datafield tag="260" ind1=" " ind2=" "> <subfield code="a">San Diego :</subfield> </pre>	<pre> <subfield code="b">Harcourt Brace Jovanovich,</subfield> <subfield code="c">c1993.</subfield> </datafield> <datafield tag="300" ind1=" " ind2=" "> <subfield code="a">1 v. (unpaged) :</subfield> <subfield code="b">ill. (some col.) :</subfield> <subfield code="c">26 cm.</subfield> </datafield> <datafield tag="500" ind1=" " ind2=" "> <subfield code="a">One Mylar sheet included in pocket.</subfield> </datafield> <datafield tag="520" ind1=" " ind2=" "> <subfield code="a">A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.</subfield> </datafield> <datafield tag="650" ind1=" " ind2="0"> <subfield code="a">Arithmetic</subfield> <subfield code="x">Juvenile poetry.</subfield> </datafield> <datafield tag="650" ind1=" " ind2="0"> <subfield code="a">Children's poetry, American.</subfield> </datafield> <datafield tag="650" ind1=" " ind2="1"> <subfield code="a">Arithmetic</subfield> <subfield code="x">Poetry.</subfield> </datafield> <datafield tag="650" ind1=" " ind2="1"> <subfield code="a">American poetry.</subfield> </datafield> <datafield tag="650" ind1=" " ind2="1"> <subfield code="a">Visual perception.</subfield> </datafield> <datafield tag="700" ind1="1" ind2=" "> <subfield code="a">Rand, Ted,</subfield> <subfield code="e">ill.</subfield> </datafield> </record> </collection> </pre>
---	---

Şekil 6: MARC Kaydı XML Formatı

Daha açıklayıcı olması ve kodlardan bağımsız bir gösterim için Şekil 7’de “Arithmetic” adlı kitabın bibliyografik katalog kaydı verilmiştir. MARC kaydında bulunan ve anlamsız gibi görünen “100, 245, 260” gibi sayılar aslında yazar adı, eser adı ve yayın yeri gibi materyale ait olan bibliyografik öğeleri ifade etmektedir. MARC ile standart bir formata kavuşan bibliyografik künyeler XML ile programlama dili ve platformdan bağımsız olarak bilgisayar sistemleri arasında kolaylıkla aktarım kabiliyeti elde etmektedir. Kütüphane otomasyon sisteminin işleyişi açısından XML, kütüphaneler arası ödünç verme (ILL), MARC, Z39.50 ve Z39.83 uygulamalarında kendine yer bulmaktadır (Tesendic, Krsticev ve Surla, 2011: 132).

Arithmetic / Carl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.
Yazar: Sandburg, Carl, 1878-1967
Katkıda Bulunan(lar): Rand, Ted [ill.]
Materyal Türü: Metin
Yayın Ayrıntıları: San Diego : Harcourt Brace Jovanovich, c1993.
Baskı: 1st ed
Tanım: 1 v. (unpaged) : ill. (some col.) ; 26 cm
ISBN: 0152038655 :
Konu(lar): Arithmetic -- Juvenile poetry | Children's poetry, American | Arithmetic -- Poetry | American poetry | Visual perception
DDC Sınıflandırma: 811/.52
LOC Sınıflandırması: PS3537.A618 | A88 1993
Özet: A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.

Şekil 7: Bibliyografik Katalog Kaydı

Kütüphane otomasyon sistemleriyle ilgili tüm standartları burada irdelemek zor olsa da bir kütüphane otomasyon sistemi yukarıda bahsedilenlerin dışında Dublin Core, RDA, SIP2, NCIP, BIBFRAME gibi standartları; REST veya SOAP gibi API protokollerini desteklemelidir (Breeding, 2016: 8). Kütüphane otomasyon sistemi kavramı kütüphanelerin basılı materyalleri öncelediği dönemin birer ürünü olarak halen satılmakta ve kütüphaneler tarafından kullanılmaktadır (Breeding, 2009: 58). Kullanıcıların artan elektronik kaynak ihtiyaçları karşısında ürün ve hizmetlerini web ve elektronik kaynak tarafına aktarmaya çalışan kütüphaneler mevcut otomasyon sistemlerinin yetersizliğinden kaynaklı bir memnuniyetsizlik de taşımaktadır. Bu noktada daha yenilikçi çözümler barındıran ve bilgi kaynaklarının değişimine ayak uydurabilecek aynı zamanda da kullanıcıların özellikle Google ve sosyal medya kullanımı ile edinmiş olduğu alışkanlıkları karşılayabilecek ürünlere yönelmek karlı bir yatırım olacaktır.

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ

2.1. Büyük Veri

Birinci bölümde veri kavramı üzerinde durulmuş ve insanın kendisi bir veri üreticisi olduğu gibi kainattaki elektronik veya değil her nesnenin veri kaynağı olabileceğinden bahsedilmişti. Evrenin var olduğundan bu yana da verinin üretildiğini söylemek yanlış olmayacaktır. Bu bölümde bilgisayar sistemlerinin kullanılmaya başlamasından bu yana kayıt altına alınan veri kastedilmektedir. Bilgisayarın insan hayatına girmesi dolayısıyla dünya üzerinde üretilen verinin miktarı ve büyüklüğü pek çok araştırmada sıklıkla zikredilmekte, her geçen yıl bir önceki yıldan daha da fazla veri üretildiği karşılaştırmalı olarak verilmektedir. İklim bilgilerini toplamak için kullanılan sensörler, sosyal medya gönderileri, dijital resimler ve videolar, çevrimiçi satın alma işlem kayıtları ve cep telefonu GPS sinyalleri artan verinin başlıca sebeplerindendir (Ohlhorst, 2012: 2). Büyük veri kavramının kullanılmaya başladığı yıllar olan 1990'ların sonunda kişisel bilgisayarlar hızla kullanılmaya başlanmış ve şirketler de iş akışlarını yavaş yavaş dijital ortama kaydırmaya başlamıştı. Günümüzde bir fırsat ve iş kolu olarak görülüyor olsa da büyük veri, veri kümelerinin bilgisayar sistemlerini bellek ve depolama kapasiteleri açısından zorlaması noktasında “büyük veri sorunu” (*problem of big data*) (Cox ve Ellsworth, 1997) olarak ifade edilmiş ve üstesinden gelinmesi gereken bir konu olarak nitelendirilmiştir. 1990'lı yıllarda bilgisayar bileşenleri bugüne oranla çok daha maliyetliydi. Bununla birlikte büyüklük göreceli bir kavramdır, kişisel bilgisayar kapasiteleri için büyük olan veri setleri nispeten küçük ya da orta ölçekli bir kurumun kapasiteleri için bile oldukça küçük kalabilmektedir. Tam bu noktada Davenport (2018: 8) veri miktarının büyüklüğünün önemli olmadığını; bu veriyi analiz edebilme ve inovasyona, katma değere dönüştürülebilmenin önemli olduğunu vurgulamaktadır. Büyük veri, aslında verilerle ve daha fazla verinin mevcut olmasıyla ilgili değil; veriden çıkarılacak analitiklerle (Zikopoulos vd., 2013: 4), sayı yığınlarının eyleme geçirilebilmesiyle ilgilidir (King, 2016: vii).

Yapay zekâyı ateşleyen hammadde ve yapay zekânın başarılı olmasını sağlayan etkenlerden biri olarak veri, yapay zekâ çağının petrolü olarak nitelendirilmektedir (Lee, 2018: 72). Eğer ortalama bir algoritmaya sahip yapay zekâya çok fazla veri akışı

sağlanırsa uygulamanın toplam gücü ve karar aşamasındaki isabet oranı büyük oranda artacaktır. Özellikle mobil uygulamalar marifetiyle toplanan web aramaları, fotoğraf yüklemeleri, izlenen videolar, gönderilere yapılan beğeniler, alışveriş verileri, yemek tercihleri, ulaşım verileri, gün içindeki lokasyon verileri gibi veri zenginliği yapay zekâ uygulamaları için paha biçilmez hammaddeledir. Şirketler, kullanıcıların günlük alışkanlıklarını belirleyerek onlara özelleştirilmiş hizmetler sunmak üzere elde edilen verileri derin öğrenme algoritmalarında kullanmaktadır. Ürün ve hizmetlerin pazarlanmasından siyaset ve politikaya kadar geniş bir yelpazede kullanılan büyük veri, bu bağlamda bir hammaddenin ötesinde Lee'nin (2018: 72) ifadesiyle Arap ülkelerinin zenginliğinin ana kaynağı olan petrol gibi bir çarpan etkisinin olduğu ve yapay zekâ çağında şirket ve ülkelere büyük bir hazine sunduğu belirtilmektedir.

Yapılan araştırmaya göre 2022 yılında bir internet kullanıcısı saniyede yaklaşık 1,7 mb veri üretmekte; dakikada 5,9 milyon Google araması yapılmakta, 231 milyon eposta ve 16 milyon mesaj gönderilmekte, 347 bin tweet, 500 saatlik YouTube videosu, 1,7 milyon Facebook içeriği, 66 bin Instagram fotoğrafı gönderilmekte; 5 milyar internet kullanıcısı bulunmakta; 97 zetta bayt veri hacmi oluşmaktadır (Jayaraman, 2023). Toplam veri hacmine bakıldığında 2010 yılında 2 zetta bayt iken 2014 yılında 12 zetta bayta, 2018'de 33 zetta bayta yükselmiştir; 2025'te ise 180 zetta bayt seviyesi öngörülmektedir (Taylor, 2023). Sayılar her ne kadar büyüleyici olsa da nicelik konusunda aynı şeyi söylemek zordur. Zira oluşan bu veri hacminin %90'ı kopyalanmış verilerden; %10'u ise tekil verilerden oluşmaktadır (Djuraskovic, 2023). Yani ortaya çıkan veri yığını sanılanın aksine katlanan bir bilgi birikimi oluşturmamakta, var olanın sürekli tekrar etmesinden ileri gelmektedir. Öte yandan veri hacmi satır-sütun formatında yapılandırılmış da değildir, bu sebeple de potansiyel olarak kıymetli olduğu değerlendirilen %25'lik bir veri olmasına karşın toplam hacmin ancak %0,5'lik bir kısmı analiz edilebilmiştir (Davenport, 2018: 8). Aslında büyük verinin iştah kabartan kısmı da bu problemten ileri gelmektedir. NewVantage Partners'in Fortune 1000 içinde listelenen 116 şirketten yönetici ile yapmış olduğu araştırmaya göre (Davenport ve Bean, 2023: 12-14) katılımcıların %87,8'i veri ve veri analitiğine yatırım yapmaya devam ettiğini, %91,9'unun ise veri ve veri analitiği yatırımlarından ölçülebilir bir katma değer elde ettiklerini ifade etmiştir. Aynı araştırmaya göre 2017 yılında bu yatırımlardan katma değer elde ettiğini belirtenlerin oranı %48,4 olarak vurgulanmaktadır ki beş yıl içinde oluşan bu fark çarpıcıdır. Başa

dönecek olursak büyük veri, anlamlandırılabilirdiği, analiz, inovasyon ve eyleme dönüştürülebildiği sürece değerlidir.

Büyük veri, geleneksel veri işleme araçlarıyla yönetebilmenin ve analiz edebilmenin mümkün olamadığı, son derece büyük veri kümeleridir (Ohlhorst, 2012: 1; Chen ve Zhang, 2014: 314). Kavramın kimin tarafından ortaya atıldığı ile ilgili bir takım söylentiler olsa da Diebold'un (2012: 3) yaptığı araştırmaya göre bugünkü kullanımı ve net ifade edilişi bakımından Silicon Graphics (SGI) bilim adamı Mashey'in (1998) hazırlamış olduğu sunum büyük veri ifadesinin kullanımı açısından belirleyici olmuştur. Mashey konuşmasında verilerin artan büyüme hızı karşısında depolama, bant genişliği, işlemci gücü, işletim sistemleri, dosya sistemlerinin yetersiz kalması ve bilgi işlem altyapısında oluşması muhtemel streslerden bahsederek geleceğe yönelik potansiyel çözümlere dikkat çekmektedir. Moore Yasası'na (Moore, 1965) göre mikro işlemcilerin üzerine yerleştirilen transistör sayısının her 18 ila 24 ayda iki katına çıkacağı ve buna paralel olarak bilgisayar bileşenlerinin daha uygun fiyatlı olacağı yönündeki tahmin büyük veriye çare olamamıştır. Verinin aranması, elde edilmesi, depolanması, analizi, paylaşımı ve görselleştirilmesi aşamaları o kadar zor bir hale gelmiş ki verinin yönetimi veya veriden katma değer elde edilmesi imkansızlaşmıştır (Ohlhorst, 2012: 1). Bilginin büyüme hızı üstel olarak artarak CPU ve depolama aygıtlarının performans hızını ve Moore Yasasını aşmakta; dolayısıyla büyük veri uygulamalarında en gelişmiş teknik ve teknolojiler gerçek zamanlı analizlerde yaşanan sorunları ideal şekilde çözememektedir (Chen ve Zhang, 2014: 318).

Büyük veriyi daha iyi tanımlamak ve karakteristiğini ortaya koymak için bazı nitelendirmeler yapılmıştır. Laney (2001) büyük veriyi 3 özellik ile tanımlamış ve kendisi 3D olarak nitelendirse de (data volume, data velocity, data variety) literatürde 3V (volume-hacim, velocity-hız, variety-çeşitlilik) olarak anılmıştır. Daha sonra diğer araştırmacılar bu tanıma *veracity*-doğruluk (Ohlhorst, 2012: 3; Goes, 2014: iv), *value*-değer (Davenport, 2018: 314), *variability*-değişkenlik, *virtual*-sanal (Chen ve Zhang, 2014: 314) gibi daha fazla V eklemiş ancak burada orijinal 3V'ye bağlı kalınarak volume, velocity ve variety kavramları incelenmiştir. Hacim veri kümesinin büyüklüğünü ifade eder; hız verinin giriş ve çıkış hızını gösterir; çeşitlilik veri türleri ve kaynakları arasındaki farklılığı tanımlar (Chen ve Zhang, 2014: 314).

Hacim (*volume*), genellikle verinin kaydedilip depolanması için gereken depolama alanını ifade eder (Kitchin ve McArdle, 2016: 6). Akıllı saatlerden otomobil sensörlerine, gps takip cihazlarından beyaz eşyalara, bilimsel araştırma laboratuvarlarından uzay istasyonlarına kadar her an her yerde veri üretilmekte, üretilen veriler sistematik veya değil bir yerlerde depolanmaktadır. Büyük verinin çıkış noktası da bu çarpıcı boyutlara ulaşan veri kümesinin yönetiminden ileri gelmektedir. Daha önce internet ortamında üretilen veri miktarından örnek verilmişti ve bu hacmin geleneksel yönetim ve analiz araç ve yöntemleriyle yürütülemeyeceği vurgulanmıştı. Bir başka örnek de CERN (Conseil Européen pour la Recherche Nucléaire) üzerinden vermek gerekirse Büyük Hadron Çarpıştırıcısı'nda (Large Hadron Collider LHC), parçacıklar dedektörlerde saniyede yaklaşık 1 milyar kez çarpışır ve saniyede yaklaşık 1 petabaytlık çarpışma verisi üretilir (Gaillard, 2017). Üretilen tüm çarpışma verisi mevcut bilgi işlem altyapısı ile depolanıp analiz edilemediği için bazı filtrelerden geçirildikten sonra araştırmacılar için ilginç olduğuna karar verilen veriler arşivlenmekte ve bu haliyle bile günde ortalama 1 petabaytlık veri işlenmektedir. Bu iş sürecinde veriler toplandıktan sonra filtrelenir, depolanır ve arşivden çevrimdışı olarak analiz edilir. CERN'in 2023 yılında yapmış olduğu çalışmalar ve bilgi teknolojilerinin geldiği aşama itibarıyla saniyede 10 gb veri kalıcı olarak saklanabilecek ve fizik analistlerinin kullanımına sunulacak şekilde gerçek zamanlı analiz yeteneğine kavuşmuştur (LHCb, 2023). Önceleri sistematik bir filtre sistemi kullanan bilim adamları gerçek zamanlı yapılanma ile dedektörlerden gelen ham veriyi inceleyebildiği gibi araştırma için ilgi çekici veriyi seçme konusunda esnekliğe sahip olmaktadır. CERN'in üretmiş olduğu bu hacimdeki veriyi depolaması ve aslında doğrudan merkezin amacı gereği ortaya çıkan veri kümesinden analiz elde edebiliyor olması büyük verinin hacminin somutlaştırılması açısından önem arz etmektedir.

Hız (*velocity*), verilerin kuruma ulaşma, işlenme veya iyi anlaşılabilme hızıdır (Zikopoulos vd., 2013: 10). Büyük verinin işletmeye kattığı değeri maksimize etmek için verinin hem kurum tarafından depolanırken hem de depolandığı arşivden erişiminde hızlı bir akış sağlamalıdır (Ohlhorst, 2012: 3). Hacim konusunda da bahsi geçti üzere gerçek zamanlı analiz büyük verinin kullanım senaryosu açısından giderek önemli bir katma değer unsuru haline gelmektedir. Büyük veri açısından üretim hızı ve işleme, kaydetme, yayınlama hızı olarak iki farklı hız çeşidi vardır (Kitchin ve McArdle, 2016: 7). Bu hızı

sağlamak için hem veri yakalama ve depolama hem de aktarım aşamasında ileri teknolojilere ihtiyaç duyulmaktadır. Büyük veri konusunda en çok gözden kaçırılan ve ikinci plana atılan alan olarak hız, kurumun iş sahası ne olursa olsun ister bir trafik sisteminin durumu, ister hasta verileri, isterse mali portföy verilerini içeriyor olsun verileri doğru ve hızlı analiz edip karar verirken kuruma büyük bir güç katacaktır (Zikopoulos vd., 2013: 11).

Çeşitlilik (*variety*), veri kümesindeki farklı türlerden yapılandırılmış, yarı yapılandırılmış veya yapılandırılmamış formattaki heterojenliği ifade eder (Gandomi ve Haider, 2015: 138). Veri kavramı ilk bakışta düz metin içerikleri veya sayılarla dolu tabloları çağrıştırıyor olabilir fakat bugün veri üreticisi binlerce çeşit kaynaktan farklı formatlarda tablo, resim, ses, video, metin, sensör verisi gibi içerikler elde edebilmekte ve analiz etmek üzere depolayabilmektedir. Yapılandırılmış verilere ilişkisel veri tabanlarındaki; yarı yapılandırılmış verilere XML, JSON formatındaki; yapılandırılmamış verilere de ses, video, metin dosyaları örnek olarak verilebilir (Terzi, Sağiroğlu, Demirezen, 2017: 20). Verileri üretilme biçimlerine göre sınıflandırmak gerekirse çok kullanıcıli uygulamalarda (sosyal medya) üretilen içerikler; bilgisayar sistemlerinin ürettiği işlem verileri; deney ve uygulamalardan toplanan bilimsel veriler; web aramaları ve madencilik uygulamalarında kullanılan web verileri; bilgi düğümleri ve düğümler arasındaki bağlantılardan oluşturulan grafik verileridir (J. Chen vd., 2013: 158).

2.2. Yapay Zekâ

Yapay zekâ alanındaki ilk önemli çalışma matematikçi Alan Mathison Turing tarafından 1950 yılında yayımladığı ve daha sonraları manifesto olarak da anılan “*Computing machinery and intelligence*” adlı makaledir (Turing, 1950; Görz ve Nebel, 2002: 10; Copeland, 1998). Makale “makinelere düşünebilir mi?” sorusu ile başlamakta; şimdilerde Turing Testi olarak bilinen “*Taklit Oyunu*” diye adlandırılan bir dizi insan-makine diyalogu içermekte; insanoğlunun ve özellikle entelektüel kişilerde daha fazla olacak şekilde düşünce gücüne binaen diğer yaratılanlardan üstün olduğu fikri karşısında makinenin insan davranışları sergileyebileceği ve makine ile iletişim kuran insanların karşısındakinin insan olduğunu sanacağını savunmaktadır. Taklit oyununda birbirinden izole odalarda bir sorgucu, bir erkek ve bir kadın bulunur. Sorgucu karşı taraftaki kadın ve erkeğe yazılı olarak soru ve cevaplar gidecek şekilde sorular sorar ve karşıdakilerin

cinsiyetini belirlemeye çalışır. Oyunun makine kısmı ise erkeğin yerine makine ile değiştirilmesi üzerine kurgulanmıştır. Makine, sorgucuya kendisinin bir kadın olduğuna ikna edebilirse testten başarılı olacağı kabul edilir. Turing, makinenin bir çocuğun eğitim hayatına paralel ve benzer bir şekilde eğitimi tamamlayacağını ve eninde sonunda tüm entelektüel alanlarda insanlarla rekabet edeceğini ifade etmiştir (Turing, 1950: 460). Böylelikle şu an kullanıldığı haliyle “yapay zekâ” terimini kullanmamış olsa bile John McCarthy’dan önce 1940-1950 yılları arasında İngiltere’de yaptığı çalışmalarla Alan Mathison Turing’in yapay zekânın öncüsü olduğu kabul edilmektedir (Dennett, 2004: 329).

Yapay zekâ terimi, Dartmouth College Matematik Bölümünde görevli bilim adamı John McCarthy öncülüğünde New Hampshire – Hannover’da küçük bir ekibin Dartmouth Yapay Zekâ Yaz Araştırma Projesi kapsamında 1956 yılında Dartmouth’ta bir araya gelmesi sonucu ortaya atılmıştır (Lee ve Qiufan, 2022: 13; Trustees of Dartmouth College, 2023). John McCarthy ve arkadaşları 1955 yılında hazırladıkları proje teklifinde 10 kişilik bir ekip ile 2 ay boyunca insanî özellikler olan öğrenme ve zekâyı simüle edebilecek bir makinenin yapılabileceği varsayımında, makinelerin insan hayatındaki sorunları nasıl çözebileceği ve kendilerini nasıl geliştirebileceği hususlarına kafa yorulacağı belirtilmiştir (McCarthy vd., 1955: 2). McCarthy Dartmouth atölyesiyle ilgili yazdığı bir yazıda “yapay zekâ” teriminin Rockefeller Vakfı’ndan 1955 yılında talep edilen fon için özel olarak seçildiğini, bu terim ile dikkat çekebileceklerini düşündüğünü belirtmiştir (McCarthy, 2006). 1956 yılında Dartmouth’ta yapay zekâyı yönelik büyük bir buluş elde edilmese de Dartmouth Yapay Zekâ Yaz Araştırma Projesi yapay zekânın bir kavram olarak belirlenmesi, insanlar tarafından üzerinde çalışılabilir bir konu olarak benimsenmesi ve o günden bu yana yapay zekânın geldiği nokta açısından büyük bir atılım olarak nitelendirilmektedir.

Yapay zekâ, bilgisayar veya bilgisayar kontrollü robotun akıl yürütme, keşfetme, genelleme, deneyimlerden öğrenme gibi insanoğluna özgü akılcı kabiliyetlere yönelik görevleri yerine getirme yeteneğidir (Copeland, 1998). Russel ve Norvig (2010: 2) yapay zekâyı tarihsel yaklaşımlara ışık tutacak şekilde insani ve rasyonalist bakış açısına göre insanî düşünen, insanî davranan, akılcı düşünen, akılcı davranan sistemler olarak tanımlamıştır. Russel ve Norvig, insan davranışı ile akılcı davranışı birbirinden ayırarak insanoğlunun irrasyonel (dengesiz veya deli) olduğunu ima etmediğini ancak insan doğası

gereği tüm insanların da mükemmel olmadığını ifade etmektedir. Harari'nin (2018) ifadesine göre de insanoğlu (aslında tüm memeliler) karşılaştığı sorunları zekâ ve bilinçlerini beraber kullanarak çözmektedir. Zekâ sorunları çözüme yeteneği kazandırırken bilinç insana acı, sevinç, sevgi, öfke gibi hissiyatlar sağlamaktadır ve bu hisler insanda bilgisayardan farklı olarak hata yapma olasılığına sebebiyet vermektedir. Bilgisayarlar ise süper zeki ama tamamen bilinçsiz bir yaşam formuna sahip olabilecektir. Başka bir tanıma göre yapay zekâ, *hem bilişsel sistemleri simüle etmeyi (akılcı/insan gibi düşünme) hem de akıllı sistemler yapılandırmayı (akılcı/insan gibi davranma) kendine amaç edinmiş bir bilimsel disiplindir* (Görz ve Nebel, 2002: 11).

Yapay zekâ, pek tabii bilim kurgu film ve romanlarında olduğu gibi bir sabah ansızın insan hayatına girmeyecektir. Hemen her sosyal, ekonomik veya teknolojik değişim gibi belli bir birikim ve alt yapı sonucu belli bir sürece yayılarak aşama aşama ve aslında çoğu insanın da farkına bile varmadan kanıksanmış olacaktır. Esasen yapay zekâ uygulamaları uzak bir gelecekte değildir, sanılanın aksine bugün pek çok günlük rutin içinde zaten kullanılmaktadır. Yapay zekâ devriminin aşamalarını adlandırmak gerekirse, internet yapay zekâ, iş yapay zekâsı, algısal yapay zekâ ve otonomi yapay zekâsı (Lee, 2018: 125) olarak dört dalga halinde gerçekleşeceğini söylemek mümkündür.

İnternet yapay zekâsı, bugün sosyal medya ve alışveriş platformlarının hemen hepsinde kullanılan algoritmayı ifade etmektedir. İnternet ortamında içerik tüketirken sıklıkla karşılaşılan ve çoğu kez kullanıcıların ilgisini çeken (veya çekmesi muhtemel) içeriklerin peşi sıra önünde akıtıldığı, eğer bir alışveriş yapacak ise kendi tarzına uygun ürünlerin listelendiği malumdur. Üzerinde vakit geçirilen platformlar aynı olsa da her kullanıcının gördüğü içerik birbirinden bağımsızdır ve kişiye özgü bir deneyim sunmaktadır. Sunulan içerik kullanıcının sayfada geçirdiği süreyi veya yaptığı harcamayı maksimize etme üzerine kuruludur. Sistemin altında yatan algoritma ise tavsiye motoruna dayanmaktadır (Lee, 2018: 126). Tavsiye, kullanıcının geçmiş tercihlerinden (içerik bazlı), benzer kullanıcıların tercihlerinden (işbirlikçi) veya geçmiş tercihlerle benzer kullanıcıların tercihlerinin beraber değerlendirildiği (hibrit) üç kategoride sağlanır (Adomavicius ve Tuzhilin, 2005: 735; Basu, Hirsh ve Cohen, 1998). Kullanıcı internet ortamında ne kadar fazla zaman geçirirse (iz bırakırsa) algoritma için o kadar fazla done oluşturacak, algoritma da kullanıcıya şahsına münhasır bir kullanım deneyimi sunacaktır. Her ne kadar kişiye özel içerik ve ürün mantığı iyi bir fikir gibi görünse de algoritmaların

manipölasyona açık yapısı kullanıcılara platformun kendi arzu ettiği içeriği sunabileceği kuşkusunu da beraberinde getirmektedir.

İş yapay zekâsı, geleneksel şirketlerin büyük hacimli etiketlenmiş verileri inceleyip insan gözünden kaçan gizli korelasyonları tespit etmesi, kurumun neden-sonuç ilişkisi içinde etiketlenen verileriyle algoritmayı eğitmesi ve geleceğe yönelik çıkarımda bulunması mantığına dayanmaktadır (Lee, 2018: 130). Yapay zekâ algoritmalarının vurucu noktası, sadece merkezde yer alan güçlü ilişkiler ağı değil bunun ötesinde zayıf ilişki olarak görünen özelliklerin de belirleyici bir anlama sahip olabileceği düşüncesidir. Algoritmanın isabetli karar verebilmesi -daha önce de bahsi geçtiği üzere- büyük hacimde kaliteli veri ile eğitilmesinden geçer (Öztemel, 2020: 22). Şirketler de bu verileri sosyal medyadan, arama sorgularından, kurum iş süreçlerinden, diğer işletmelerden veya kamuya açık ticari veri tabanlarından sağlayarak maliyet azaltma, isabetli karar alma ve maksimum karlılık için gerçek bir rekabet avantajı sağlamaktadır (Akerkar, 2019: 11).

Algısal yapay zekâ, beş duyu organı vasıtasıyla çevresel farkındalık sağlayan insani duyulara benzer şekilde duyuşal verileri algılayıp yorumlayabilen algoritmaları ifade etmektedir (Vincze, Wachsmuth ve Sagerer, 2014: 168). Bilgisayarlar, ses ve görsel içerikli duyuşal verileri önceden beri saklayabiliyorlardı ancak multimedya içerikler bilgisayar tarafından içeriğin ne olduğu önemsenmeksizin salt 0,1 ikilisinden ibaret olan veri kümesi olarak bilinmekte ve diğer veri setleriyle aynı muameleyi görmektedir. Algısal yapay zekâda durum bunun ötesindedir. Algoritma, bir görsel veya videonun içinden anlamlı kümeler çıkarabilir, görselin belli bir noktasındaki nesneleri algılayabilir, ses dosyasındaki farklı kaynağa ait sesleri ayırt edebilir veya içerdiği metinleri çözümleyebilir (Lee, 2018: 139). Bugün kullanılan pek çok mobil akıllı cihaz bu algoritmaları belli düzeylerde zaten kullanmaktadır. Akıllı cihazlarda bulunan kameralar sahnede gördüğü nesnelere uygun filtreler uygulayabilmekte, yüz tanıma sistemi ile cihaz sahibini ayırt edebilmekte, sesle komut özelliği ile metinler yazdırılabilmektedir. Burada algoritmanın zeki davranış sergilemesi yalnızca karşılaştığı bir duruma önceden kurgulanmış bir aksiyon verilmesi değildir. Çünkü algının zeki sayılabilmesi için zekâ sahibinin algıladığının bilinçli olarak farkında olması ve elde ettiği deneyimden de tecrübe kazanması gerekir (Kendrick, 1998: 213). Araştırmacılar da bu noktada beyin mekanizmasının etkili bir şekilde simüle edebilecek akıllı algılama sistemlerini tasarlama ve uygulama gayreti içerisinde (Tian vd., 2017: 59). Optik kamera, mikrofon, dokunsal

sensör, LIDAR sensör, kızılötesi tarayıcı, ivmeölçer, jiroskop gibi algılayıcılar ile görüntü işleme, ses analizi, dokunma duyuları simüle edilebildiği gibi koku ve tat duyuları da artık uygulanabilir hale gelmiştir (Wayner, 2022).

Otonomi yapay zekâsı, kendisine verilen görevi insan müdahalesi olmadan yerine getiren sistemlerdir (Zgurovsky, Kasyanov ve Levenchuk, 2023: 763). Verilen görevi yerine getiren makineler yeni değil zira onlarca yıldır çeşitli sektörlerde (özellikle sanayi) kullanılmakta, pek çok üretim süreçleri kısmen veya tamamen robotlar tarafından yürütülmektedir. Ancak otonom yapay zekâda durum biraz daha farklıdır. Geleneksel robotik üretim sistemlerinde işlemler insan nezareti altında makineye tanımlı rutinler üzerinden yürütülür, makine sürekli olarak aynı işlemi tekrar eder. Makine yaptığı işlemde bir tecrübe kazanmaz, karşılaştığı aksi duruma tepki vermez, rutinin dışına çıkmaz. Otonomi yapay zekâda ise internet yapay zekâ, iş yapay zekâsı, algısal yapay zekânın birbirine entegrasyonu sonucu makineler, karmaşık veri setlerini optimize etme becerilerini sensör temelli teknolojiler ile birleştirecek, dünyayı anlama ve onu şekillendirme yeteneğine sahip olacaktır (Lee, 2018: 151). Otonomi yapay zekânın günlük hayata en bariz yansıması otonom arabalar olarak karşımıza çıksa da pek çok sektörde otonom teknolojiler hayata geçirilmektedir. 1926 yılında Linriccan Wonder marka radyo kontrollü otomobil ile başlayan otonom araç çalışmaları (Bimbrow, 2015: 191) ve Carnegie Mellon University'nin 1984 yılında Navlab projesi olarak tanıttığı değişken koşullar altında kendi kendine giden araç fikri (Thorpe vd., 1991) ve en nihayetinde özellikle Tesla'nın kullanıma sunduğu ve otopilot kabiliyetine sahip modellerinde yakın zamanda sürücünün müdahalesinin en aza indirilmesi ve aracın müdahale olmadan seyrini tamamlaması beklenmektedir (Lambert, 2023). Karayolu motorlu otonom taşıtları için 0 ile 5 arasında hiyerarşik altı seviye tanımlanmıştır (SAE International, 2021). Seviye 0 herhangi bir makine yardımı olmaksızın insan kontrolündeki araçları, seviye 5 ise insan yardımı veya nezareti olmaksızın tüm rotanın makine vasıtasıyla tamamlandığı araçlardır. Otonom otomobillerin dışında insansız hava araçları konsepti de oldukça rağbet görmektedir. Bazı yabancı örnekleri olmakla birlikte yerli üretim Anka, Akıncı, TB2-TB3 ve Kızılelma gibi askeri sınıf insansız hava araçları yarı otonom veya otonom uçuş, iniş-kalkış kabiliyetlerine sahiptir. Otomobiller ve askeri sınıf İHA'lar dışında sivil ve ticari amaçlı İHA (Tomic vd., 2012), insansız su altı aracı (Liu vd., 2014), bilimsel, ticari veya askeri insansız yüzey araçları (Liu vd., 2016), keşif,

arama-kurtarma, tarım, madencilik ve askeri amaçlı insansız kara aracı (Papadakis, 2013) ve birden çok insansız aracın müşterek görev icra ettiği sürü İHA (Schilling vd., 2019) bugün kullanılan otonom araç yaklaşımlarından bazılarıdır.

Yapay zekânın başarılı olabilmesi, doğal dil işleme (NLP), bilgi temsili, otomatik akıl yürütme, makine öğrenimi, bilgisayarlı görü ve robotik disiplinlerinde başarılı olmasına bağlıdır (Russell ve Norvig, 2010: 3). Russel ve Norvig (2010) bu önermeyi Turing testine dayanarak izah etmiş ancak Turing testinde fiziksel etkileşimden kaçınılması sebebiyle göz ardı edilen bilgisayarlı görü ve insanın fiziksel simülasyonun (robotik) bugün için gerekli ölçütler olduğunu ve yapay zekâda bulunması gereken özellikler olarak ifade etmiştir. Doğal dil işleme (NLP), günlük kullanılan insan dilinin bilgisayar tarafından doğal dildeki kelime, karakter veya bayt dizileri üzerinde olasılık dağılımını tanımlayan dil modellerine dayanarak kullanılmasıdır (Goodfellow, Bengio ve Courville, 2016: 448). Son zamanlarda sıklıkla adından söz ettiren ChatGPT ve türevleri doğal dil işleme modeli kullanılmaktadır (Kuroiwa vd., 2023: 2). Modern NLP sistemleri bilgiyi işlerken derin öğrenme tekniklerini ve insani tepkiler üretmek için de olasılığı kullanılmaktadır (Encyclopædia Britannica, 2024). Algoritmanın olasılığa dayalı yapısı bazı durumlarda modelin kullanıcıya konu hakkında bilgi sahibi olmadığını belirtmek yerine iyi huylu, potansiyel olarak tehlikeli veya art niyetli olmak üzere “halüsinasyon” olarak tanımlanan yanlış bilgiler verebilmektedir (McIntosh vd., 2023: 1). Bilgi temsili, bilginin bilgisayarlar tarafından depolanmak üzere kolayca işlenebilecek şekilde yapılandırılarak temsil edilmesidir (Wang, Tao ve Wu, 2011: 3568). Otomatik akıl yürütme, bir sistem ya da algoritmanın ne yapacağı veya yapmayacağı konusunda matematiksel kanıta dayalı güvence sağlamaya çalışan bilgisayar bilimidir (Amazon, 2024). Makine öğrenimi, verilerdeki örüntüleri otomatik olarak tespit edebilen ve daha sonra ortaya çıkarılan örüntüleri gelecekteki verileri tahmin etmek veya belirsizlik altında diğer karar verme türlerini gerçekleştirmek için kullanabilen bir dizi yöntem olarak tanımlanmaktadır (Murphy, 2012: 1). Makine öğrenimi, karmaşık fonksiyonları istatistiksel olarak tahmin etmek için bilgisayar kullanımına önem veren uygulamalı bir istatistik biçimidir (Goodfellow, Bengio ve Courville, 2016: 95). Bilgisayarlı görü, mühendislik sistemlerinin görsel algıyı kullanarak çevreyi modellemesini ve manipüle etmesini sağlamak amacıyla üç boyutlu dünyanın görsel görüntülerdeki bilgileri çıkartmak, karakterize etmek ve yorumlamak için dijital bilgisayar tekniklerinin

kullanılmasıdır (Zhihui, 2008: V). Robotik, yekpare fiziksel yapı üzerinde sensörler, aktüatörler ve hesaplama platform ve süreçlerini bütünleştiren (McKee, 2006: 693), insanlar tarafından verilen görevleri gerçekleştirmek için makinelerin tasarım inşa ve kullanımıdır (Hosch, 2009). Robotların yapımı yapay zekânın tüm teknik ve yöntemlerinin bir arada kullanımını gerektirmesi bakımından yapay zekâ alanındaki en çok çaba ve mücadele gerektiren olgudur (Görz ve Nebel, 2002: 85). Robotlar temel olarak, algı (iç-dış sensörler, kameralar), hareket (tekerlek, palet, mekanik bacak) ve manipülasyon aparatı (mekanik kol, yapay el) bileşenlerinden oluşmakta (Siciliano vd., 2009: 2), bulunduğu mekanı idrak edebildiği gibi yapay zekâ kabiliyetleri ile mantık yürütebilmekte, mekanik aksamaları ile çevresine müdahale edebilmektedir. Robotların bu yeteneği insan-robot etkileşiminde belli sınırların oluşmasına da sebebiyet vermiştir. İlk kez 1942 yılında Isaac Asimov tarafından kaleme alınan Kovalamaca adlı kısa öyküde dile getirilen Robotbilimin Üç Kanunu insan-robot etkileşimini ve bu etkileşimden doğabilecek zarara vurgu yapılmıştır. Yasalar şu şekildedir: *1-Robotlar, insanlara zarar veremez ya da eylemsiz kalarak onlara zarar gelmesine göz yumamaz. 2-Robotlar, Birinci Kanun’la çakışmadığı sürece insanlar tarafından verilen emirlere itaat etmek zorundadır. 3-Robotlar, Birinci ve İkinci Kanun’la çakışmadığı sürece kendi varlıklarını korumak zorundadır* (Asimov, 2016: 49-50). Asimov ortaya koyduğu yasalarda insan hayatını ve verilen emrin yerine getirilmesini öncelemiş son seçenek olarak da makinenin kendi varlığını koruması gerektiğini belirtmiştir. Asimov 1985 yılında Kurtarıcı/ Robotlar ve İmparatorluk (Robots and Empire) kitabının Düello bölümünde robotların çok yönlülüğünün artması ile daha öncelikli bir kanuna ihtiyaç hissettiğini ve Robotbilimin Sıfırıncı Kanunu (Asimov, 2016: 237) olarak adlandırdığı yasada tek bir insanın değil tüm insanlığın zarar görmesini önleyecek şekilde “*Bir robot insanlığa zarar veremez ya da harekete geçmeyerek insanlığın zarar görmesine göz yumamaz.*” maddesi eklemiştir (Asimov, 1986: 290).

Yapay zekânın ortaya çıktığı ilk dönemlerden bu yana neredeyse bir trilyon kat daha fazla bilgi işlem kapasitesi elde edilmişken depolama maliyetleri on beş milyon kat ucuzlamıştır (Lee, Qiufan, 2022: 14). Bu kapasite ve işlem gücü artışı yapay zekâ için isabetli sonuçlar verebilmesine imkân tanıyan eğitici verinin birikmesine, sistemin kendi kendine öğrenmesine olanak sağlamaktadır. Her ne kadar bilgi teknolojileri insan hayatına şu ya da bu şekilde dokunuyor, belli kazanımlar sağlıyor olsa da bu avantajlar,

sistemin hiç dezavantaj barındırmadığı anlamına gelmiyor. Özellikle, kontrolden çıkan yapay zekâya karşı korku ve endişe duyma hali olarak ifade edilen “yapay zekâ kaygısı” (Johnson ve Verdicchio, 2017), kişisel mahremiyet, güvenlik ve gizlilik (Ahmad vd., 2023), etik (Siau ve Wang, 2020) gibi konularda yapay zekâ uygulamaları bir takım soru işaretleri barındırmaktadır (Vardi, 2012: 5).

Yapay zekânın harcadığı enerji hesaba katılarak iklim üzerinde yapay zekâ ve uygulamalarının olumsuz bir etkisi olduğu yönünde de çalışmalar ortaya çıkmıştır (Naughton, 2023). 2019’da yapılan bir araştırmaya göre doğal dil işleme (NLP) uygulamalarının doğruluğu, büyük miktarda hesaplama kaynakları ile modellerin yoğun şekilde eğitilmesini zorunlu kıldığı, bu işlemlerin de hem donanım, elektrik ve bulut bilişim maliyetleri hem de modern tensör işleme donanımını beslerken ortaya çıkan karbon ayak izi bakımından oldukça maliyetli olduğu ortaya koyulmuştur (Strubell, Ganesh ve McCallum, 2019). 2021 yılında yapılan bir araştırmaya göre GPT-3 doğal dil işleme (NLP) uygulaması eğitim modülünün kullandığı enerji miktarının 1287 MWh, karbon emisyonun ise 552 tCO₂e olduğu tahmin edilmekte ve bu emisyon oranı, standart bir yolcu uçağının San Francisco (SFO) ile New York (JFK) arasında gidiş-dönüş için karbon emisyonunun (~180 tCO₂e) yaklaşık 3 katı olduğu hesaplanmıştır (Patterson vd., 2021: 7). Yine aynı araştırmada Meena, T5, GShard ve Switch Transformer gibi modellerin eğitim için emisyon ve harcadığı enerji miktarına bakıldığında Meena 232 MWh enerji, 96 tCO₂e emisyon; T5 86 MWh enerji, 47 tCO₂e emisyon; GShard 24 MWh enerji, 4,3 tCO₂e emisyon; Switch Transformer ise 179 MWh enerji harcadığı ve 59 tCO₂e emisyonu görülmektedir. Yapay zekâ modellerinin maliyeti sadece eğitim aşamasıyla sınırlı değildir. Modeller eğitildikten ve kullanıma hazır hale geldikten sonra sisteme gönderilen her bir sorguda sistemin bağlı olduğu kaynaklar yeniden devreye girmekte ve diğer sistemleri devreye sokarak yeni maliyetler oluşturmaktadır. 2023 yılında yapılan bir araştırmaya göre yapay zekâ çıkarım aşamasında görüntü altyazısı oluşturma, görüntü oluşturma gibi üretken görevlerde, metin ve görüntü sınıflandırma gibi nispeten az enerji gerektiren görevlere göre 10 kata kadar daha fazla enerji kullandığı; metin oluşturma modelinin ortalama bir akıllı telefonun şarjının %16’sı kadar, görüntü oluşturma modelinin ise en az 950 akıllı telefonun şarjı kadar enerji kullandığı saptanmıştır (Luccioni, Jernite ve Strubell, 2023: 6).

2.3. Veri Madenciliği

Önceki bölümlerde bahsi geçtiği üzere bugün isteyerek veya istem dışı, bilinçli veya değil her an bir yerlerde veri üretiyor, üretilmesine sebebiyet veriyor veya kişisel cihazlar marifetiyle bu verilerin depolanmasına vesile oluyoruz. İnsanlar tarafından üretilen veriler bir yana artık makinelerin kendi aralarında kurduğu iletişimden (IoT), yerine getirdiği iş süreçlerinden, doğrudan makinenin kendisinden de sürekli olarak veri üretiliyor. Her an üretilen veri yığınları büyük veri, bilgi patlaması gibi kavramların ortaya çıkmasına neden olmuş ve veriden bilgi elde etmek isteyenler için birer maden yatağı; veri setleri ise hammadde değerine erişmiştir. Ancak her maden yatağında olduğu gibi veri madenlerinde de cevher elinizin altında değildir, cevheri yani bilgiyi çıkartmak için belli yöntem ve teknikler kullanmak, soyut ya da somut (zaman, emek, para vs.) harcamalar yapmak gerekir. Dünya üzerinde devasa bir veri hacmi bulunmakta fakat ünlü bir sözde vurgulandığı gibi *bilgi içinde boğuluyoruz fakat bilgiye aç kalıyoruz* (Naisbitt, 1984: 17). Naisbitt (1984), kontrolsüz ve organize olmayan bilginin bilgi toplumunda bir değerinin olmadığını aksine bilgi uzmanlarının düşmanı olduğunu ve bilgi teknolojilerinin bilgi kirliliği içinde işe yaramayacak verilere değer kattığını ifade etmektedir. Bilimsel ve mühendislik uygulamalarından, tıp ve sağlık endüstrisinden, arama motorlarından, sosyal medya ve web topluluklarından, işletmeler ve müşteri işlemlerinden gelen muazzam miktardaki veri içerisinden değerli bilgileri (information) otomatik olarak ortaya çıkarmak ve bu tür verileri düzenlenmiş bilgiye (knowledge) dönüştürmek için güçlü ve çok yönlü araçlara ihtiyaç duyulmaktadır (Han, Kamber ve Pei, 2012: 2). İş, bilim ve mühendislik alanlarının çoğunluğunda bulunan büyük, karmaşık ve bilgi (information) açısından zengin veri kümelerini anlama ihtiyacı, veri madenciliğini, neyin ilginç bir sonuç doğuracağı bilinmediği durumlarda keşif amaçlı analiz senaryolarının en maharetlisi yapmaktadır (Kantardzic, 2019: 2).

Tarihsel olarak veri setinden faydalı örüntüler bulma kavramı, veri madenciliği, bilgi çıkarma, bilgi keşfi, bilgi toplama, veri arkeolojisi, veri örüntüsü işleme gibi isimlerle ifade edilmiş (Fayyad, Piatetsky-Shapiro ve Smyth, 1996: 39), bilginin (knowledge) veriye bağlı bir keşfin son ürünü olduğunu vurgulayan “veri tabanından bilgi keşfi” (Knowledge Discovery in Databases-KDD) literatürde uzunca zaman kullanılan bir kavram olarak 1989 yılında Michigan’da gerçekleştirilen 11. International Joint Conference on Artificial Intelligence çalıştayında ortaya atılmıştır (Piatetsky-Shapiro,

1990: 70). Veriden bilgi keşfi veya veri tabanından bilgi keşfi şemsiye bir terim olarak salt keşif algoritmalarını kullanma işlemi olan veri madenciliğini de içine alacak şekilde tüm veri anlamlandırma işlemlerini ifade eden bir kavram olarak kullanılmış (Piatetsky-Shapiro vd., 1996: 89; Fayyad, Piatetsky-Shapiro ve Smyth, 1996: 39) ancak zamanla endüstri, medya ve araştırma camiasında tüm bilgi keşif sürecini ifade etmek için “veri madenciliği” kavramı kanıksanmıştır (Han, Kamber ve Pei, 2012: 6). Önceleri veri madenciliği tekniklerinin ilgilendiği veri setleri satır sütun formunda tabloda tutulan verilerden oluşsa da bilgi işlem gücündeki sıçrama ve verinin hacim ve format bağlamında yaşadığı büyüme veri madenciliğini standart dışı veri kümeleri (metin, görüntü, video, grafik ve ağ verileri) ile çalışmaya itmiştir (Coenen, 2011: 26). Alanın büyümesi ve uzman sistemlerden keşif, genetik algoritma, makine öğrenme, sinir ağları gibi yapay zekâ uygulamalarının bilgi keşfine nasıl uyarlanabileceği sorusu atölye ve konferanslar düzenleme ihtiyacını doğurmuş, ilki 1995 yılında Montreal’de olmak üzere The First International Conference on Knowledge Discovery and Data Mining (AAAI, 2024); 1997 yılında Norveç’te European PKDD ve yine 1997 yılında Singapur’da Pacific/Asia PAKDD düzenlenen konferanslar veri madenciliği popülaritesini önemli ölçüde artırmış; ilk veri madenciliği şirketleri ve ürünleri de bu dönemde ortaya çıkmıştır (Clifton, 2009).

Veri madenciliği, daha önceden **bilinmeyen, geçerli ve uygulanabilir** bilgilerin (information) büyük veri tabanlarından çıkarılması ve enformasyonların kritik **işletme kararları** alırken kullanılması sürecidir (Cabena vd., 1998: 12). Veri madenciliği, büyük miktardaki veriden **ilginç** örüntüler ve bilgi (knowledge) keşfetme sürecidir (Han, Pei ve Tong, 2022: 2). Veri madenciliği, büyük hacimli verilerde **yeni, değerli ve sıradan-önemsiz olmayan** bilginin (information) aranmasıdır (Kantardzic, 2019: 2). Veri madenciliği, büyük bir veri koleksiyonundan değerli ancak aşikâr olmayan (**gizli**) bilgilerin verimli şekilde keşfedilmesidir (Bigus, 1996: 9). Tanımlarda vurgulanan **bilinmeyen, ilginç, sıra dışı ve gizli** kavramları önemlidir zira bir veri tabanı üzerinden SQL tarzı sorgulamalar ile (ne kadar komplike olursa olsun) elde edilen çıkarımlar veri madenciliği değildir (Coenen, 2011: 25). Elde edilen bulgular daha önce tahmin edilememiş, bilinmeyen, işletme için geçerli ve kritik derece değerli, yeni bilgilerden müteşekkil olmalıdır. Veri madenciliği, önceden bilinmeyen veya tahmin edilmeyen sonuçlara odaklanmakta; bir sebeple bile olsa öngörülen veya haricen zaten çıkarılmış

sonuçların ispatı ile ilgilenmemektedir (Silahtaroglu, 2013: 12). Veri madenciliği uğraş alanı bilginin önceden tahmin edilme ihtimalinin düşük olduğu, sezgisel olmayan hatta sezgilere aykırı olanı aramaktır (Cabena vd., 1998: 12). Cabena vd. (1998: 12) kitabında bilginin gizli, bilinmeyen olmasını veri madenciliği konusunda anlatılan meşhur bir araştırma ile açıklamaktadır: *Bir zincir market satın alma modellerini analiz etmek için veri madenciliğini kullanmış ve cuma akşamları çocuk bezi ile bira satışları arasında güçlü bir ilişki olduğu fark edilmiş. Bu ilişki irdelendiğinde hafta sonu için bebek ihtiyaçlarını stoklayan erkek müşteriler markete uğradığında kendi ihtiyaçlarını da gidermektedir. Bira ile çocuk bezi arasında tespit edilen bu ilişki o kadar mantıksızdır ki ilişki tespit edile kadar hiç kimsenin aklına böyle bir ihtimal gelmemiştir ve market çalışanları bu tespiti karlı bir şekilde kullanacaktır. Ancak bir otomobil bayi geçmiş satış kayıtlarından veri madenciliği yöntemlerini kullanarak büyük sedan kasa otomobillerin nispeten yaşlı ve varyetli müşteriler tarafından tercih edildiğini tespit etmesi büyük bir ticari fayda getirmeyecektir. Zira bu satın alma modeli herkesin malumdur. Veri madenciliği herkes tarafından bilinen veyahut tahmin edilen ilişkilerle ilgilenmemektedir.*

Veri madenciliğini farklı tür ve sınıflara ayırmak mümkün olsa da temelde *tanımlayıcı veri madenciliği* (descriptive data mining) ve *öngörüselsel veri madenciliği* (predictive data mining) olarak iki kategoride (Kantardzic, 2019: 3; Tan vd., 2019: 29), operasyonel olarak da *çok boyutlu veri özetleme* (multidimensional data summarization); *sıklık örüntüleri, birliktelik ve korelasyon* (frequent patterns, associations, correlations); *sınıflama ve öngörü* (classification, regression); *bölütleme* (clustering); *derin öğrenme* (deep learning); *aykırılık* (outlier) olmak üzere 6 başlıkta incelemek mümkündür (Han, Pei ve Tong, 2022: 6). *Tanımlayıcı veri madenciliği*, veri seti analiz edilerek mevcut durum hakkında yeni bilgilerin elde edildiği analiz çeşididir ve doğası gereği keşif niteliğinde olup sonuçları doğrulamak ve açıklamak işlem sonrası tekniklere ihtiyaç duyar (Tan vd., 2019: 29). *Öngörüselsel (tahmin) veri madenciliği*, gelecekte yaşanacak muhtemel sonuçlara ilişkin tahmin ve öngörülerde bulunmak üzere büyük veri setlerinden bilgi çıkarma sürecidir (Larose ve Larose, 2015: 4). *Çok boyutlu veri özetleme*, büyük miktardaki veri içerisinde ilgi duyulan veri kümesinin otomatik özetlenmesi ve diğer veri kümeleriyle karşılaştırmalarının çıkarılması amacıyla pasta grafik, çubuk grafik, eğriler, veri küpleri, çapraz tablolar gibi biçimlerde sunulmasıdır (Han, Pei ve Tong, 2022: 6). *Sıklık örüntüleri*, ilk olarak 1993 yılında (Agrawal, Imieliński ve Swami, 1993)

ortaya atılmış ve yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış veri setlerinden sık görülen (belirlenen bir eşik değerden az olmayan) örüntüleri keşfetmeye odaklanır; birliktelik kuralı, sınıflandırma, kümeleme ve değişiklik tespiti, sıralı örüntü, yapılandırılmış örüntü, buzdağı küp hesaplama, küp gradyan analizi, ilişkisel sınıflandırma (Giannella vd., 2004: 105) gibi görevlerde temel araç olarak da kullanılır (Jin ve Agrawal, 2007: 61). *Sınıflama ve öngörü*, veri sınıflarını veya kavramları tanımlayan ve ayırt eden, bir dizi eğitim verisinin analizine dayanarak türetilmiş model bulma (sınıflandırma kuralları, karar ağaçları, sinir ağları veya matematiksel formüller) sürecidir (Han, Pei ve Tong, 2022: 8). *Bölütleme (kümeleme)*, sınıf etiketinin bilinmediği, tahmin edilecek bir sınıfın bulunmadığı, verilerin birbirlerine olan benzerliklerine göre doğal gruplara bölüneceği durumlarda uygulanır (Witten vd., 2017: 141). *Derin öğrenme*, insan beyni, istatistik ve uygulamalı matematik alanlarından büyük ölçüde yararlanan, zekâ gerektiren görevleri yerine getirebilen bilgisayar sistemlerine odaklanmış bir makine öğrenme yaklaşımıdır (Goodfellow, Bengio ve Courville, 2016: 26). *Aykırlık*, veri setinde her nokta çifti arasındaki mesafenin hesaplanarak diğer tüm noktalardan uzaktaki noktaların tespit edilmesi (Parthasarathy, Ghoting ve Otey, 2007: 291), çoğu durumda bu anomalinin gürültü olarak istisna tutulması bazen ise (dolandırıcılık tespiti gibi) özellikle öne çıkarılması sürecidir (Han, Pei ve Tong, 2022: 10).

Uygulama odaklı bir disiplin olarak veri madenciliği verinin olduğu her yerde bir veri madenciliği uygulaması sağlayarak büyük başarılar elde etmektedir (Han, Kamber ve Pei, 2012: 27). Cabena vd. veri madenciliği uygulama alanlarını, pazar yönetimi (hedef pazarlama, müşteri ilişkileri yönetimi, pazar-sepet analizi, çapraz satış, pazar segmentasyonu), risk yönetimi (tahmin, müşteri tutma, geliştirilmiş sigortalama, kalite kontrol, rekabet analizi) ve dolandırıcılık yönetimi olarak üç başlıkta toplamıştır (Cabena vd., 1998: 26). Öte yandan web madenciliği, muhakeme gerektiren kararlar, görüntü işleme, yük tahmini, teşhis, pazarlama ve satış (Witten vd., 2017: 21) uygulama alanları arasında sayılabilmektedir. Sektörde kullanılan tüm veri madenciliği uygulamalarını sıralamak mümkün olmasa da ticari olarak başarılı bazı uygulamalara vurgu yapacak şekilde *iş zekâsı*; *arama motorları* (web madenciliği); *sosyal medya ve sosyal ağlar*; *biyoloji, tıp ve sağlık hizmetleri* (biyomedical madenciliği) uygulamaları (Han, Pei ve Tong, 2022: 17) irdelenmiştir.

İş zekâsı, kurum yönetiminin kâr ve performansını artırmak için ihtiyaç duyduğu önemli bilgileri ürün, teknoloji ve yöntemleri bir araya getirerek karar ve eylemlere yönelik işletme bilgisi ve işletme analizleridir (Williams ve Williams, 2007: 2). İş zekâsı, karmaşık karar alma süreçleri için faydalı bilgi ve enformasyon üretmek üzere mevcut verilerden yararlanan bir dizi matematiksel model ve analiz metodolojisidir (Vercellis, 2009: 3). İşletmelerin müşteri, pazar, tedarikçi ve rakipleri daha iyi değerlendirmek için iş süreçlerinin geçmişi, bugünü ve geleceğe yönelik tahminlerinde raporlama, çevrimiçi analitik işleme, iş performansı yönetimi, rekabete yönelik istihbarat, kıyaslama ve tahmin gibi araçlarla destek sağlanabilmektedir (Han, Pei ve Tong, 2022: 17). İş zekâsı, kurumsal raporlama, OLAP (online analytical processing-çevrimiçi analitik işleme), sorgu ve tahmine dayalı analitik üretmek için veri ambarında tutulan içeriğe erişen işletme kullanıcıları ve uygulamalar bütünüdür (Watson ve Wixom, 2007: 96). İş zekâsı terimi 1980’li yılların sonunda ortaya çıkmış ve çalışanların verileri analiz etmek üzere çıkarma (extract), dönüştürme (transform) ve yükleme (load) sürecine (ETL) ve bu analizleri rapor, uyarı ve puantajlarda kullanılabilir hale getirmesine olanak tanıyan, böylelikle daha iyi bir karar verme imkanı sağlamak üzere veriyi toplamak, analiz etmek ve yayınlamak için kullanılan çeşitli süreç ve yazılımları kapsamaktadır (Davenport, 2006: 105).

1989 yılında Tim Berners-Lee ve arkadaşlarının CERN’de geliştirmeye başladığı WWW projesi ve 1990 yılı sonlarında CERN laboratuvarında bulunan NeXT bilgisayarında çalıştırdığı ilk web sitesini oluşturmalarının ardından (CERN, t.y.) bugün dünyada 1 milyardan fazla web sitesi bulunmakta (Tambe ve Jain, 2023) ve web sitelerindeki bu hacim web sayfalarından gerekli bilgilerin çıkarılmasında **web madenciliği** tekniklerini kullanan **arama motorlarını** gerekli kılmaktadır (Bhatia, 2019: 368). Web madenciliği, web içerik ve verilerinden, web yapısından ve web kullanım verilerinden bilgi (knowledge) çıkarmak için veri madenciliği tekniklerinin uygulanmasıdır (Srivastava, Desikan ve Kumar, 2004: 406). Web madenciliği, popüler iki alan olan internet ve veri madenciliğini bir araya getirmekte; veri madenciliği ile internetteki kullanıcı davranışlarına ilişkin içgörüler geliştirilmekte; kullanıcıların ziyaret geçmişi analiz edilerek ilgili siteler dinamik olarak önerilebilmektedir (Cabena vd., 1998: 36). Webte bulunan veriler veri madenciliği için çok büyük bir uygulama alanı sağlamakta; kazancının önemli bir kısmını veriden elde eden Google gibi arama motoru şirketleri web siteler için “prestij” olarak da ifade edilen sıralama ölçümü yapmakta ve

sayfalar ne kadar fazla prestije sahipse o kadar yoğun trafik elde edebilmektedir (Witten vd., 2017: 21). Arama motorları, büyük veri, sorgu sayısının azlığı yanında en büyük problemi çevrimiçi verilerle uğraşmak zorunda olması veri madenciliğine zorluklar yaşatmaktadır (Han, Pei ve Tong, 2022: 17). Bu zorluk, veri madenciliğinin geleneksel olarak çevrimdışı verilerle çalışma alışkanlığının dışında akan veri biçimleri ile de ilgilenmesini gerektirmektedir. Akan veri, ağ cihazları, IoT sensörleri, sunucu günlükleri, finansal işlemler, coğrafi konum verileri, mobil cihazlar ve makine sensörleri gibi çeşitli kaynaklar tarafından üretilen; başı ve sonu olmadan sürekli akan veridir (Raj vd., 2022: 1).

Sosyal medya madenciliği, sosyal medya kullanımı, çevrimiçi sosyal davranışlar, bireysel bağlantılar, çevrimiçi alışveriş gibi büyük miktardaki sosyal medya verisini örüntü ve trendleri belirlemek için incelemek; sosyal ağ madenciliği ise ağ, grafik teorisi ve veri madenciliği yöntemlerini kullanarak sosyal ağ yapılarını ve bu ağlara yönelik bilgileri araştırmaktır (Han, Pei ve Tong, 2022: 18). Sosyal medya madenciliği, devasa sosyal medya verilerini araştırmak için temel kavram ve algoritmaları belirler; bilgisayar bilimi, veri madenciliği, makine öğrenmesi, sosyal ağ analizi, ağ bilimi, sosyoloji, etnografya, istatistik, optimizasyon ve matematik gibi disiplinleri bir araya getirerek anlamlı örüntüleri çıkaran, ölçen ve modelleyen araçları kapsamaktadır (Zafarani, Abbasi ve Liu, 2014: 2). Daha önce büyük veri başlığında sosyal medya kullanımı ile ilgili güncel veriler paylaşıldığı üzere, sosyal medya dünyanın her yerinden milyarlarca bireyin etkileşim, paylaşım, gönderi ve günlük aktivite yürüttüğü; gazetecilerin paylaştığı haber ve yayınların tüketildiği; insanların birbiriyle bağlantı ve etkileşim kurduğu; hemen her meslekten insanların görev icra ettiği; muazzam boyutta çoğu gürültülü ve yapılandırılmamış verilerden oluşan özellikle insan davranışlarını gözlemlemek için veri madenciliği adına altın fırsatlar sunmaktadır. Sosyal medya ve ağ madenciliği ile gizli toplulukları tespit, sosyal ağların evrimi ve dinamiklerinin ortaya koyulması, ağ ölçümleri, bilginin sosyal medyada yayılması, homofili (benzerlik sevgisi) ölçmek, konuma dayalı sosyal ağ analizi (Han, Pei ve Tong, 2022: 18) yürütülebildiği gibi sahte bilgi tespiti için de kullanılmaktadır (Shu ve Liu, 2019: 1).

Biyomedical madenciliği, veri madenciliği ve makine öğrenme kabiliyetleri ile karar destek, sinyal-görüntü analizi, epidemiyoloji, biyoinformatik, biyoloji uygulamalarının tıp ve sağlık bilimlerinde kullanılmasıdır (Peek, Combi ve Tucker, 2009:

225). Genomik ve proteomik dizi madenciliği, alt grafik örüntü madenciliği, düzenleyici ağ madenciliği, protein etkileşimi tahmini, tıbbi görüntü sınıflandırma ve tahmini, biyolojik bilgi ağı, sağlık kayıtları madenciliği ve biyomedikal ağ madenciliği olmak üzere pek çok veri madenciliği görevi icra edilmektedir (Han, Pei ve Tong, 2022: 19). Tıp ve sağlık sektöründe bilim insanları tarafından gerçekleştirilen araştırmalar neticesinde büyük bir veri seti ortaya çıkması bir yana dünya üzerinde sağlık merkezlerine müracaat eden hastaların oluşturmuş olduğu devasa bir veri hacmi bulunmaktadır. Sağlık sektörünün birebir insan hayatı ile alakadar olması ortaya çıkan verinin kullanımını kelimenin tam anlamıyla “hayati derecede önemli” yapmaktadır. Sağlık merkezleri kayıt aldığı DNA, genomik dizi, protein ve molekül yapıları, biyomolekül etkileşimleri, gen ontolojisi, hastalık sebepleri, medikal görüntü, elektronik sağlık kayıtları gibi biyolojik ve klinik verilerin neden olduğu veri bombardımanı sektörün verileri çıkarma ve analiz etme kabiliyetini aşmış; odak nokta veri madenciliği yoluyla etkili biyolojik madencilik, yeni ilaç keşfi, bilinçli karar destek mekanizmalarına kaymıştır (Li, Ng ve Wang, 2014: v). Veri madenciliği sağlık sektöründe, sigortacılığa ilişkin dolandırıcılık ve suistimal tespiti, etkili tedavi ve uygulamaların belirlenmesi, hastalar için en uygun fiyatlı sağlık hizmeti tespiti, daha hızlı rapor ve analiz, düşük işletme maliyeti gibi amaçlarla kullanılmaktadır (Nalinipriya vd., 2021: 156).

Veri madenciliği, büyük veri tabanlarından bilgi çıkarma sorununu çözmek için makine öğrenimi, istatistik, veri tabanı (Bhatia, 2019: 25), örüntü tanıma ve görselleştirme (Cabena vd., 1998: xi), yapay zekâ, uzman sistemler (Fayyad, Piatetsky-Shapiro ve Smyth, 1996: 39), paralel hesaplama, dağınık hesaplama, (Tan vd., 2019: 28), doğal dil işleme, insan-bilgisayar etkileşimi, algoritmalar, yüksek başarımlı hesaplama, sosyal bilimler tekniklerini bir araya getiren disiplinler arası bir alandır (Han, Pei ve Tong, 2022: 12). Burada veri madenciliğine katkıda bulunan tüm disiplinleri irdelemek gereksiz olsa da veri madenciliğinin arkasındaki lokomotif alanlar olan istatistik, makine öğrenme ve veri tabanı (Fayyad, Piatetsky-Shapiro ve Smyth, 1996: 40; Kargupta vd., 2004: x) kavramlarını açıklamak yerinde olacaktır.

İstatistik, verileri toplama, analiz etme, sunma ve yorumlama bilimidir (Anderson, Sweeney ve Williams, 1998). Veri madenciliği ile istatistik arasında doğal bir bağ olmakla birlikte istatistiksel modeller veri ve veri sınıflarını modellemek; veri ve istatistiksel modelleri kullanarak tahmin ve öngöründe bulunmak; verilerden çeşitli

örüntülerin çıkarılması ve örüntülerin irdelenmesi; veri madenciliği ile elde edilen sonuçların doğrulamak için istatistik kullanılabilir (Han, Pei ve Tong, 2022: 13). Verilerden bilgi keşfetmek temelde istatistiksel bir çabadır çünkü istatistik genel popülasyonun belirli bir örnekleminde genel örüntüler çıkarmaya çalışıldığında ortaya çıkan belirsizliği ölçmek için bir dil ve çerçeve sağlamaktadır (Fayyad, Piatetsky-Shapiro ve Smyth, 1996: 40). İstatistik, veri madenciliğinde olmazsa olmaz bir bileşen olarak önemli bir rol oynamakta, örüntü bulma, modelleme, analiz etme gibi veri madenciliğine atfedilen bazı özellikler istatistiğin de temelini oluşturmaktadır (Hand, Mannila ve Smyth, 2001: 19). Veri analizinin bilgisayar tabanlı ilk uygulamaları istatistikçilerin desteğiyle geliştirilmiş ve istatistik bilimi, tek boyutlu/ çok boyutlu veri analizi, farklı regresyon ve diskriminant analizi türleri olmak üzere veri madenciliğine çeşitli yöntemler sunmaktadır (Kantardzic, 2019: 166).

Makine öğrenme, bilgisayarların örnek veri ya da geçmiş deneyimi kullanarak başarımlarını artıracak biçimde programlanmasıdır (Alpaydın, 2017: 3). Makine öğrenme doğası gereği multidisipliner bir alan olarak yapay zekâ, olasılık ve istatistik, hesaplamalı karmaşıklık teorisi, kontrol teorisi, bilgi teorisi, felsefe, psikoloji, nörobiyoloji alanlarından elde edilen sonuçlardan yararlanır (Mitchell, 1997: 2). Bilgisayar programları günlük hayatta pek çok işlevi yerine getirmektedir fakat makine öğrenmesinde odak nokta sadece elektronik bir aletin sorunları çözüyor olması değildir. Makinenin, insana bahşedilen (bir yönden de hayvanlara) “öğrenme” yetisine sahip olması, deneyimlerden birikimli bir öğrenme sürecinin sağlanması, bu süreç sonunda da yeni durum karşısında öğrendiği ile uyum göstermesi beklenir. Makine öğrenme ile veri madenciliği birbirlerine çok benzemekte, makine öğrenmesinin büyük veri tabanlarına uygulanmasına da veri madenciliği denmektedir (Alpaydın, 2017: 2). Ancak veri madenciliği makine öğrenmeden, çok büyük veri kümeleri (hatta sonsuz), yüksek kaliteli eğitim verisi ve uygulama çeşitliliği bakımından farklılık göstermektedir (Han, Pei ve Tong, 2022: 14). Çoğu makine öğrenme algoritması denetimli öğrenme ve denetimsiz öğrenme olarak iki kategoriye ayrılmaktadır (Goodfellow, Bengio ve Courville, 2016: 98).

Veri tabanı, birbiriyle ilişkili verilerin tutulduğu, kullanım amacına uygun şekilde düzenlenmiş veriler topluluğunun mantıksal ve fiziksel tanımlarının olduğu bilgi depolarıdır (Şentürk, 2006: 4). Veri madenciliği hakkında yapılan tanımlara yeniden

bakıldığında veri madenciliğinin verinin tutulduğu koleksiyon, depo yani veri tabanından bilgi keşfetme süreci olduğu anımsanacaktır. Bir yerlerde veri barındırılıyor ve veri madenciliği uygulanmak isteniyorsa veri tabanlarından faydalanılacak demektir. Temelde bir veri tabanı ilgi duyulan veriler koleksiyonudur (Halpin ve Morgan, 2008: 2). Bu bir kütüphane veri tabanı ise derme ve kullanıcılara ait verilerin tutulduğu bir koleksiyon olması beklenir. Kütüphane mevcut dermeye ilişkin verileri veri tabanına ekler ve sonrasında yeni gelen kitaplar için veri tabanı her seferinde güncellenir. Benzer şekilde kullanıcılara ait veriler her yeni kullanıcı eklendiğinde veya bir kitap ödünç alındığında/ iade edildiğinde güncellenir. Bu işlemler manuel veya çevrimiçi sistemler vasıtasıyla otomatik olarak icra edilebilir. Veri tabanlarını yönetmek için kullanılan yazılımları ifade eden veri tabanı yönetim sistemleri (VTYS) ise birbiriyle ilişkili veriler ve bu verilere erişimde kullanılan bir dizi programdan oluşan koleksiyonlardır (Silberschatz, Korth ve Sudarshan, 2020: 1). Yalnız veri madenciliği uygulamalarında değil veri ve bilgisayar uygulamalarının bulunduğu her yerde yoğun olarak kullanılan veri tabanlarında, veri içeriğine ilişkin basit sorgu ve güncellemelerin yapıldığı Çevrimiçi İşlem Gerçekleştirme (OLTP) (Hand, Mannila, Smyth, 2001: 417); verileri özetleyerek değer tahmini üreten ve veri madenciliğinde de önemli bir yer tutan Çevrimiçi Analitik İşleme (OLAP) (Bhatia, 2019: 420; Halpin ve Morgan, 2008: 888) olmak üzere iki geniş uygulama sınıfı yer alır (Silberschatz, Korth ve Sudarshan, 2020: 1231). Birçok veri madenciliği görevinde büyük veri kümeleri, gerçek zamanlı veriler ve hızlı akan verilerin işlenmesi gerekirken son zamanlarda veri tabanı sistemleri veri ambarı ve veri madenciliği imkanlarını bünyesinde barındırarak veriler üzerinden sistematik veri analizi yeteneği sunmaktadır (Han, Pei ve Tong, 2022: 15). Veri tabanları, veri madenciliğinin arkasındaki itici güçtür ve verimli veri erişimi sağlamak, veriye erişirken gruplama ve sıralama işlemleri, sorguları optimize etmeye yönelik veri tabanı teknikleri, algoritmaları büyük veri kümelerine ölçeklendirmede temel oluşturur (Fayyad, Piatetsky-Shapiro ve Smyth, 1996: 40).

2.4. Veri Madenciliği Algoritmaları

Algoritma, bir problemin çözümü, belirli bir görevin yerine getirilmesi için gereken sonlu talimatlar kümesidir (Sherine vd., 2023: 1). Algoritma kelimesi, M.S. 800'lü yıllarda Bağdat'ta yaşayan ve Beytülhikme kütüphanesinde de görev yapan Müslüman matematik alimi Ebû Ca'fer Muhammed b. Mûsâ el-Hârizmî'nin (Fazlıoğlu,

1997) adından geldiği kabul edilmektedir (Crossley ve Henry, 1990: 105). Veri madenciliği alanında pek çok algoritma geliştirilmiş ve kullanılmıştır. Buna karşın veri madenciliği camiasında yaygın olarak kullanılan ve en etkili algoritmaları belirlemek amacıyla 2006 yılında The IEEE International Conference on Data Mining (ICDM) konferansında yapılan bir çalışma (Wu vd., 2008) en iyi 10 algoritmayı belirlemiştir. Çalışmada belirlenen 10 algoritma (C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, CART), sınıflandırma, kümeleme, istatistiksel öğrenme, birliktelik analizi ve web madenciliği olarak 5 kategoriye ayrılmıştır (Wu vd., 2008; Witten vd., 2017: 504). Burada adı geçen algoritmalar ilgili kategorileriyle birlikte irdelenmiştir.

Sınıflandırma, makine öğrenimi araştırmacıları ve istatistikçiler tarafından bilinmeyen örneklerin sonucunu tahmin etmek için, nesnelerin (veya şeylerin) verilen belirli sayıda sınıfa kategorize edilmesidir (Bhatia, 2019: 65). Sınıflandırma, tahmin edici bir model ve en popüler veri madenciliği tekniği olarak resim ve örüntü tanıma, tıbbi teşhis, kredi onayı, hata tespiti, finansal eğilimleri belirleme alanlarında bolca kullanılmaktadır (Dunham, 2003: 75). İstatistik alanında lineer regresyon analizi, lojistik regresyon analizi, diskriminanz analizi, Bayes; makine öğrenimi alanında ise karar ağaçları, en yakın komşu, yapay sinir ağları, destek vektör makineleri en bilinen sınıflandırma yöntemleridir (Akpınar, 2014: 69). Sınıflandırma kategorisinde C4.5, kNN, Naive Bayes, CART ve AdaBoost algoritmaları incelenmiştir.

C4.5, J. Ross Quinlan tarafından ID3 algoritmasına dayanarak geliştirilen (Quinlan, 1993: 1), denetimli öğrenmeyi hedefleyen, karar ağacı sınıfından (decision trees) makine öğrenme ve veri madenciliği algoritmasıdır (Wu ve Kumar, 2009: 1). C4.5, ID3 aracılığıyla aslında CLS (Concept Learning Systems) ailesinin özelliklerini de bünyesinde barındırarak en iyi bölen öznelilik seçiminde entropi ve enformasyon kazancı tekniklerini kullanmaktadır (Akpınar, 2014: 221). C4.5 algoritması, ID3 algoritmasını *eksik veri* (eksik veriler göz ardı edilir); *sürekli veri* (eğitim örneğinde verinin nitelik değerine göre bölme); *budama* (alt ağaç değiştirme/ alt ağaç yükseltme); *kurallar* (karar ağacı kurallarına göre sınıflandırma); *bölme* (ortalamanın üstünde bilgi kazancı) konularında geliştirmekte ve üstünlük sağlamaktadır (Dunham, 2003: 100). ID3 algoritması sadece kategorik değişkenlerin incelenmesine olanak verirken C4.5 hem kategorik hem de nümerik değişkenlerin analizine olanak vermektedir (Sarul, 2016: 249).

Algoritmanın en önemli noktası eğitim örneklerinden karar ağacı oluşturma sürecidir ve oluşturulan ağaç kökten başlayarak yaprakla karşılaşana kadar ilerleyerek yeni örneği sınıflandırmak için kullanılır (Kantardzic, 2019: 201).

kNN (*k*-nearest neighbor – *k*-en yakın komşu), tahmin ve kestirim için kullanılabilmesinin yanı sıra sınıflandırma algoritmaları arasında en sık tercih edilen; eğitim veri setinin saklandığı ve sınıflandırılmamış yeni verinin sınıfı, eğitim setindeki en benzer veriler karşılaştırılarak bulunduğu algoritmadır (Larose, 2005: 96). Diğer bir tanıma göre, her yeni veri bir mesafe ölçüsü kullanılarak ve mevcut veriler karşılaştırılarak en yakın verinin sınıfı yeni veriye atanmak üzere kullanıldığı, örnek tabanlı ve tembel öğrenme tekniğini içeren bir algoritmadır (Witten vd., 2017: 85). kNN ilk olarak 1950’li yılların başında geliştirilmiş ancak eğitim setinin işlenmesi büyük bir bilgi işlem gücü gerektirdiği için 1960’lı yıllara kadar popülerlik kazanamamıştır (Han, Pei ve Tong, 2022: 266). Tembel öğrenenlerden ezberci sınıflandırıcılar (rote classifier) tüm eğitim setini ezberleyen ve sınıflandırmayı yalnız test verileri ile eğitim verileri eşleştğinde gerçekleştirdiği ve bu sebeple eğitim setinden herhangi bir veri ile eşleşmeyen test verilerinin sınıflandırılmaması problemi, en yakın komşu yaklaşımı ile test örneklerinin nispeten de olsa benzeyen eğitim örneklerini belirleyerek çözüm sunmaktadır (Tan vd., 2019: 410). Eğitim ve test verilerinin birbiriyle tamamen örtüşmesi mecbur değildir. Eğitim verilerinden elde edilen örnekler ile test verilerinde en yakın komşular bir araya toplanır ve aynı sınıfa yerleştirilir. Kavramın içinde geçen yakınlık olgusu matematiksel bir mesafe ölçümünü gerektirmektedir. Sayısal niteliklerle tanımlanan verilerin birbirinden farklılığını ortaya koymak ve hesaplamak için *Öklid*, *Manhattan* (X. Wu, Kumar, 2009: 154), *Minkowski*, (Han, Pei ve Tong, 2022: 48), *Chebyshev* (Bhatia, 2019: 160), *Hamming*, *Mahalanobis* (Tan vd., 2019: 96-116) gibi çeşitli uzaklık hesaplama ölçütleri kullanılmaktadır. “*k* en yakın komşu” kavramında yer alan *k* parametresi, uzaklık hesabı yapılırken en yakın kaç komşunun ölçüleceğini ifade etmekte ve 1’den büyük ve tek sayı olmak üzere genellikle *k*=3 veya *k*=5 tercih edilmektedir (Kantardzic, 2019: 136). Çift sayı seçilmemesinin sebebi en yakın komşular arasında eşitlik olma olasılığını bertaraf etmektir.

Naive Bayes, 18. yüzyılda yaşayan Thomas Bayes tarafından ortaya atılan Bayes teoremine dayanan ve yapay zekâ, makine öğrenimi gibi alanlarda yer bulan olasılığa dayalı sınıflandırma algoritmasıdır (Bhatia, 2019: 110). Bayes, eldeki sınıflandırılmış

hazır veriler kullanılarak yeni verinin mevcut sınıflar içerisinde birine girme olasılığını hesaplayan bir yöntem (Silahtaroglu, 2013: 97) olarak Thomas Bayes'in ölümünün ardından yayınlanan çalışmasına (Bayes ve Price, 1763) dayanmaktadır (Bishop, 2006: 21). Bayes teoremi yapay zekâ sistemlerinde belirsiz muhakemeye yönelik modern yaklaşımların hemen hepsine temel oluşturmaktadır (Russell ve Norvig, 2010: 9). Naive Bayes, koşullu olasılıkları tahmin etme yöntemi dışında Bayes sınıflandırıcı ile aynı prosedürü izlemekte; hesaplamaları basitleştirmek için bir öznitelik değerinin bir sınıf üzerindeki etkisinin diğer öznitelik değerlerinden bağımsız olduğunu kabul etmekte ve bu anlamda da “naif” (basit, bağımsız, aptal (Wu ve Kumar, 2009: 163)) olduğu kabul edilmektedir (Han, Pei ve Tong, 2022: 260). Naive bayes, önceden sınıflandırılmış hazır verilerle eğitilen algoritmanın, sonradan verilen koşulların her birinin ayrı ayrı ve mevcut sınıfların olasılıklarının beraber hesaplanması sonucu olasılık değeri en yüksek olana göre sınıflandırma mantığında çalışmaktadır. Bu hesaplamada verilen değişken ya da koşulun gerçekleşme sayısının 0 (sıfır) olması yani olmaması durumunda matematiksel bir problem oluşmaktadır. Bayes mantığı koşulların gerçekleşme sayıları üzerinden çarpma işlemi gerçekleştirdiği için hiç gerçekleşmemiş bir değişkenin çarpımı algoritmayı bozabilmektedir. Olasılık değerlerinin 0 (sıfır) olması durumunda Laplace Düzeltmesi ile tüm koşullara 1 eklemek suretiyle tahmin edilen olasılık değerinde göz yumulabilir farkla sorun giderilmiş olur (Han, Pei ve Tong, 2022: 265). Naive Bayes, karmaşık parametre tahmin sistemine ihtiyaç duymadığı için geliştirilmesi kolaydır; büyük veri setlerine uygulanabilir; yorumlanması kolaydır, uzman olmayan kullanıcılar tarafından dahi algılanabilir; sınıflandırma algoritmaları arasında mükemmel olmasa bile genellikle başarılı ve güvenilir sonuçlar verir (Wu ve Kumar, 2009: 163).

CART (classification and regression trees – sınıflandırma ve regresyon ağaçları), Breiman vd. (1984) tarafından geliştirilen, karar ağacı algoritmalarından, kök düğümden başlayarak her seviyede iki alt diziye ayrılan, ikil olarak büyüyen ve bu sayede daha homojen düğümler elde etmeyi amaç edinmiş sınıflandırma ve regresyon algoritmasıdır (Akpınar, 2014: 225). Kategorik bağımlı değişken veya sürekli bağımlı değişkenlerden sınıflandırma problemlerini analiz etmek için geliştirilmiş parametrik olmayan bir algoritma olarak CART, bağımlı değişken kategorik olduğunda sınıflandırma ağacı oluştururken, bağımlı değişken sürekli olduğunda regresyon ağacı oluşturur (Yohannes ve Webb, 1999: 4). CART algoritması istatistik alanında karar ağacı uygulamalarının

öncüsü olmuştur (Alpaydın, 2017: 170). Bir diğer karar ağacı algoritması olan ID3'te olduğu gibi CART, en iyi dal ayırma özellik ve kriterini seçmek için entropiden yararlanır ancak ID3'te her alt kategori için bir alt öge oluşturulurken CART'ta iki alt öge oluşturulur (Dunham, 2003: 102). CART algoritmasının diğerlerine göre kritik avantajlarını şu şekilde sıralamak mümkündür:

- En uygun ağacı oluşturmak için ağaç bölme durdurma kuralını rafa kaldırır ve tahmin ağacı bölünme imkânsız hale gelene kadar devam eder. Maksimum ağaç oluşturulduktan sonra CART hem tahmin doğruluğu ve cezaya dayalı olarak mükemmel ağacı bulana kadar dalları budar.
- Sonuç doğruluğunu değerlendirmek için ayrı bir test veri seti veya verinin az olduğu durumlarda çapraz doğrulama yöntemlerini kullanır.
- Diğer ağaç büyütme yöntemlerinden farklı olarak belirli değişken etkilerinin bağlam bağımlılığını ortaya çıkarmak için aynı değişkeni ağacın farklı bölümlerinde kullanabilir.
- Geleneksel istatistik modellerini iyileştirmek için kullanılabilir (Steinberg ve Colla, 1995: 26).
- CART'ın en büyük avantajı büyük veri kümeleri ve yüksek boyutluluk problemleri ile başa çıkabilmekte, birkaç önemli değişkeni kullanarak analiz için sunulan çok sayıdaki değişkenden yararlı sonuçlar üretebilmektedir (Yohannes ve Webb, 1999: 10).

AdaBoost, tanıtımı ilk olarak 1995 yılında (Freund ve Schapire, 1995) yapıldıktan sonra 1996 yılında uygulama detayları ile birlikte açıklanan algoritma (Freund ve Schapire, 1996), sınıflandırma algoritmalarının hatasını önemli ölçüde azaltma amacıyla birden fazla algoritmanın bir arada kullanıldığı *boosting* mantığına dayanmaktadır. Boosting, her bir modelin öncekinin hataları üzerinden eğitildiği, üç zayıf modeli birleştirerek güçlü bir model oluşturulduğu yöntem (Alpaydın, 2017: 371) olarak 1990 yılında ortaya atılmıştır (Schapire, 1990). AdaBoost, “*adaptive boosting*” (uyarlanabilir iteleme/ güçlendirme) kısaltması ve en popüler *boosting* algoritması olarak zor problemleri çözmek için zayıf modelleri yapay sinir ağları, karar ağaçları veya kNN dahil en doğru herhangi bir sınıflandırıcıda birleştirir (Kantardzic, 2019: 288). AdaBoost'ta hep aynı eğitim kümesi kullanılır ve bölünmediği için çok büyük eğitim kümesi gerektirmez ancak sınıflandırıcıların veriyi ezberlememesi için basit sınıflandırıcı seçilmelidir.

(Alpaydın, 2017: 371). Boosting yaklaşımında var olan birleştirilecek sınıflandırıcı sayısının üç tane olma kısıtı AdaBoost'ta yoktur, isteğe bağlı sayıda sınıflandırıcı birleştirilebilir. Algoritmanın başarısını kenar payını arttırmaktan ileri geldiği ve bu sayede kenar payı arttıkça eğitim örneklerinin birbirinden daha iyi ayrılıp hata olasılığının düştüğü vurgulanmaktadır (Schapire vd., 1998: 1663; Alpaydın, 2017: 372). Öte yandan algoritmanın başarısı, bilgisayar bilimleri alanında en prestijli ödüllerden Gödel Prize'ı 2003 yılında kazanmasıyla tescillenmiştir (EATCS, 2003; Freund ve Schapire, 1997).

Kümeleme, benzer özellikler taşıyan nesnelerin bir arada gruplanması işlemidir (Murphy, 2012: 875). Kümeleme işleminde bir küme içindeki nesnelerin yüksek düzeyde birbirine benzerlik göstermesinin yanı sıra diğer kümelerdeki nesnelere benzememesi için veri setinin birden fazla kümede gruplama işlemidir (Han, Pei ve Tong, 2022: 379). Buradaki benzeşme durumu önemlidir zira aynı küme içindeki nesnelerin birbirine maksimum seviyede benzemesi ancak nesnelerin diğer kümelerdeki örneklerle minimum benzerlik göstermesi kümeleme algoritmasının da bir anlamda başarısını gösterir. Kümeleme teknikleri tahmin edilmesi gereken bir sınıfın olmadığı, bunun yerine veri setinin doğal gruplara ayrılması gerektiği durumlarda uygulanır ve bu gruplar muhtemelen veri setinin elde edildiği alanda işleyen ve bazı örneklerin birbirine benzemesine sebep olan bazı mekanizmaları yansıtır (Witten vd., 2017: 141). Kümeleme analizi de verileri anlamlı, faydalı veya her ikisine sahip gruplara böler ve istatistik, veri madenciliği, makine öğrenme, bilgi erişimi, örüntü tanıma, biyoloji ve sosyal bilimler gibi alanlarda uzun süredir önemli rol oynamaktadır (Tan vd., 2019: 379). Uygulama sahasında farklı ihtiyaçlara cevap verebilecek geniş bir kümeleme algoritma yelpazesi olmakla birlikte çoğu algoritma temelde *hiyerarşik kümeleme (hierarchical clustering)* ve *bölümlemeli kümeleme (partitional clustering)* olarak iki popüler yaklaşıma dayanmaktadır (Kantardzic, 2019: 298; Akpınar, 2014: 286; Silahtaroglu, 2013: 163). Kümeleme kategorisinde bölümlemeli algoritmalarından K-means incelenmiştir.

K-means, eğitim setini birbirine yakın örneklerden oluşan ve kullanıcı tarafından belirlenen sayı kadar farklı kümeye bölen algoritmadır (Goodfellow, Bengio ve Courville, 2016: 150). K-means, örneklerin istenen kümeye ulaşmaya kadar kümeler arasında taşındığı yinelemeli bir algoritmadır (Dunham, 2003: 140). K-means, belirli bir veri setini kullanıcı tarafından verilen k sayıda kümeye bölen, basit, yinelemeli, uygulanması ve çalıştırılması basit, nispeten hızlı, uyarlanması kolay, yaygın kullanıma sahip bir

algoritmadır (X. Wu ve Kumar, 2009: 22). K ifadesi kullanıcı tarafından verilir ve veri setinin kaç farklı kümeye bölünmek istendiğini belirtmektedir. K-means üzerine yapılan ilk çalışmalar farklı disiplinlerden araştırmacılar (Lloyd, 1982; Forgy, 1965; Friedman ve Rubin, 1967; MacQueen, 1967) tarafından yürütülmüştür (Wu, 2012: 7). K-means algoritması, küme merkezini kümenin içindeki noktaların ortalama değeri olarak tanımlar ve ilk olarak veri setinden her biri başlangıçta bir küme merkezini temsil eden k adet rastgele nesne seçer; kalan veri setindeki nesneler için seçilen ortalamalar arasındaki Öklid mesafesine bağlı olarak her nesne en çok benzer olduğu kümeye gönderilir; her seferinde küme içindeki nesneler ile yeni ortalama hesaplanır ve yeni atamalar yapılır; yinelemeler, atama kararlı olana kadar devam eder en iyi kümeler oluşturulur (Han, Pei ve Tong, 2022: 388). K-means algoritmasının hızlı çalışmasının yöntemlerinden biri, nispeten de olsa oluşturulan kümelerin kalitesinden ödün vermek pahasına keyfi hiper düzlem bölümlerini kullanmak yerine veri setini yansıtılmasına ve seçilen eksenler boyunca kesim yapılmasına izin vermesine bağlıdır (Witten vd., 2017: 144). K-means algoritması, kategorik verilerle çalışamama gibi çeşitli sınırlamaların üstesinden gelebilmek için *k-Modes*, *k-Medoids*, *k küme sayısı tahmini*, *öznitelik dönüşü* gibi farklı uygulamalar barındırmaktadır (Han, Pei ve Tong, 2022: 392).

İstatistiksel öğrenme teorisi, Vapnik–Chervonenkis (VC) Teorisi olarak da bilinen tümevarımsal öğrenmeyle ilgili tüm kavramları titizlikle tanımlayan, çoğu tümevarımsal öğrenme çıktıları için matematiksel kanıt sunan, diğer yaklaşımların aksine teorik kanıt ve formalizasyona ihtiyaç duymadan pratik uygulamaya vurgu yapan, büyük örneklemeler için geliştirilen istatistiksel yöntemleri içinde barındıran tümevarımsal öğrenme teorisidir (Kantardzic, 2019: 104). İstatistiksel öğrenme, eğitim hatası ile genelleme hatası arasındaki tutarsızlığın model kapasitesi arttıkça büyüyen fakat eğitim örneklerinin sayısı arttıkça küçülen miktarla sınırlandığını göstermektedir ve bu makine öğrenme algoritmalarının çalışabileceğine dair gerekçe sağlar (Goodfellow, Bengio ve Courville, 2016: 114). İstatistiksel öğrenme teorisi ilk olarak 1964 yılında Vapnik ve Chervonenkis'in yaptığı çalışma (Vapnik ve Chervonenkis, 1964) ile ortaya çıkmıştır. İstatistiksel öğrenme bilim, finans ve endüstrinin pek çok alt alanında kalp krizi tahmini, ekonomik fiyat tahmini, el yazısından karakter tanımlama, diyabet tahmini, kanser risk faktörlerinin belirlenmesi gibi pek çok konu ve problemde kullanılmaktadır (Hastie,

Tibshirani ve Friedman, 2009: 1). İstatistiksel öğrenme kategorisinde SVM ve EM algoritmaları irdelenmiştir.

SVM (Support Vector Machines), sınıfları ayırmak için öznitelik uzayında doğrusal veya doğrusal olmayan karar sınırlarını öğrenen ayırt edici bir sınıflandırma algoritmasıdır (Tan vd., 2019: 478). Denetimli öğrenmeye yönelik en etkili yaklaşımlardan biri olan SVM, doğrusal regresyona benzetmekle birlikte lojistik regresyonun aksine olasılık sağlamaz, yalnızca bir sınıf kimliği çıkarır (Goodfellow, Bengio ve Courville, 2016: 141). İlk olarak 1992 yılında (Boser, Guyon ve Vapnik, 1992) ortaya çıkan SVM'nin temelleri 1960'lı yıllarda geliştirilen istatistiksel öğrenme teorisine dayanmaktadır (Han, Pei ve Tong, 2022: 319; Akpınar, 2014: 268). SVM, çalışma mantığı açısından istatistiksel öğrenme teorisine dayanmasına karşın uygulanma amacı bakımından sınıflandırma algoritmasıdır zira SVM daha önce belirtildiği gibi veri setini birden fazla sınıfa ayırmaktadır. Algoritma, nesnelerin arasından geçen bir çizgi ile sınıflar oluşturmak üzere nesnelerin birbirine en uzak olduğu noktadan sınır çizgileri çekerek başarıyı yüksek sınıflar oluşturma gayreti güder. Sınıfları ayırmak için en basit durumda iki sınıf doğrusal olarak ayrılabilceği gibi sınıfların doğrusal olarak ayrılamadığı durumlarda SVM, *polinom çekirdeği*, *Gauss radyal*, *Sipmoid çekirdeği* gibi farklı fonksiyonlarla sınıfları ayırma yeteneğine sahiptir (Han, Pei ve Tong, 2022: 326). SVM, bir dizi önceden etiketlenmiş eğitim verisi vasıtasıyla denetimli öğrenen bir algoritma olarak sağlam bir teorik temele sahiptir ve eğitim için nispeten az sayıda örneğe ihtiyaç duymaktadır (Kantardzic, 2019: 117). SVM, sayısal tahmin ve sınıflandırma, el yazısı rakam, nesne, duygu ve konuşmacı tanıma, kıyaslama zaman serisi tahmin görevlerinde kullanılabilmektedir (Han, Pei ve Tong, 2022: 319).

EM (Expectation Maximization) algoritması, en büyük olasılırlık kestiriminde, biri gözlenen diğeri saklı olmak üzere iki küme değişken bulunduğunda gözlenen değişkenlerin olasılırlığını en büyük yapan yöneyi bulma amacı güder ve bunu doğru yapamadığında saklı değişkenleri de işleme sokarak modeli hem gözlenen hem de saklı değişkenleri kullanarak tanımlar, her iki değişken kümesi için birleşik dağılımın olasılırlığını gösteren tüm olasılırlık işlevini en büyük yapan yöneyi arar (Alpaydın, 2017: 123). Algoritma ilk olarak 1977 yılında yapılan bir çalışma (Dempster, Laird ve Rubin, 1977) ile ortaya çıkmış ve eksik veri sorunu görüldüğü durumlarda maksimum olasılık tahminlerini yinelemeli olarak hesaplanmasına yönelik bir yaklaşım sunulmuştur.

Algoritmanın her yinelemesi bir beklenti ve bir büyütme (maksimizasyon) adımı içermektedir. Algoritma gözlenen değişkenler için bilinen değerler ve parametrelerin mevcut tahmini göz önüne alındığında, saklı değişkenler için dağılımı bulan bir *beklenti* adımından ve beklenti adımında bulunan dağılımın doğru olduğu varsayımı altında parametreleri maksimum olabilirliğe sahip olanlar şeklinde yeniden tahmin eden bir *büyütme* adımından oluşur (Neal ve Hinton, 1998: 356). Yinelemeler, parametrelerin tahminleri değişmeyene kadar veya çok az değişiklik gösterene kadar devam eder (Tan vd., 2019: 652). EM algoritması, veri madenciliği, makine öğrenme, örüntü tanıma gibi çeşitli alanlarda sayısal kararlılığı, uygulama basitliği ve güvenilir küresel yakınsama gibi pek çok özelliği ile geniş kullanım alanı bulmaktadır (X. Wu, Kumar, 2009: 93). Genel olarak EM algoritması bulanık kümeleme (*fuzzy clustering*) ve olasılıksal model tabanlı kümeleme (*probabilistic model-based clustering*) işlemlerinde kullanılmaktadır (Han, Pei ve Tong, 2022: 441).

Birliktelik analizi, ilişkisel veri tabanları, işlemsel veri tabanları ve diğer veri depolarındaki nesneler arasındaki sık görülen örüntü, korelasyon, ilişki veya nedensel yapıların tanımlanmasıdır (Bhatia, 2019: 232). Birliktelik kuralı, X öncel ve Y ardıl olduğunda $X \rightarrow Y$ biçiminde bir gerektirmedir ve $X - Y$ iki nesne arasındaki ilişkinin tespit edilmesi problemine odaklanmaktadır (Alpaydın, 2017: 44). Yani X nesnesi varken Y nesnesinin var olma olasılığını bulmaya çalışır. Birliktelik kuralında yoğun olarak *destek* (*support*), *güven* (*confidence*) ve *etki* (*lift*) olmak üzere üç ölçüt (özellikle destek ve güven) kullanılmaktadır (Alpaydın, 2017: 44). Bu ölçütlerle birliktelik kuralı yorumlanmaktadır (Karahana Adalı, 2016: 222). Literatürde market (pazar) sepet analizi, ilişki analizi, yakınlık analizi olarak da bilinen birliktelik analizi, veri madenciliğinin ana tekniklerinden biridir ve denetimsiz öğrenme sistemlerinde yerel model keşfinde yaygın olarak kullanılır (Kantardzic, 2019: 336). Birliktelik analizinin en basit örneklendirilmesi, (daha önce detayları verildiği üzere) bira - bebek bezi çıkarımıdır. Bira ve bebek bezi satışları arasındaki bağı tespit edilmesi ve müşterilerin satın alma eğilimlerine yönelik daha fazla ürün satışının sağlanması en bariz örnektir. Birliktelik kuralındaki amaç, alışveriş esnasında müşterilerin satın aldıkları ürünler arasındaki birliktelik ilişkisini tespit etmek, bu ilişki sonucunda da müşterilerin satın alma alışkanlıklarını ortaya koymaktır (Karahana Adalı, 2016: 221). Birliktelik analizi nesneler arasındaki ilişkiyi tanımlar fakat nedensellik gerektirmez yani ilişkiye sebebiyet veren *saklı değişkenler*

olabilir (Dunham, 2003: 9; Alpaydın, 2017: 45). Bu saklı değişkenlerin bulunması nesneler arasındaki ilişki yapısının daha kolay tanımlanmasına imkân verir. Birliktelik analizi ilk olarak 1993 yılında Agrawal, Imielinski ve Swami'nin yaptığı çalışma ile ortaya atılmıştır. (Agrawal, Imieliński ve Swami, 1993; Hand, Mannila ve Smyth, 2001: 447). Birliktelik analizi kategorisinde Apriori algoritması irdelenmiştir.

Apriori algoritması, ilk olarak 1994 yılında Agrawal ve Srikant'ın yaptığı ve büyük satış işlemlerine ait verilerin tutulduğu veri tabanlarındaki öğeler arası ilişki kurallarını keşfetme problemini çözmeye yönelik çalışma ile ortaya koyulmuştur (Agrawal, Srikant, 1994). Apriori, birliktelik analizi konusunda en çok bilinen ve en yoğun kullanılan geniş veri setlerindeki nesneler arası ilişkileri ortaya çıkaran algoritmadır (Dunham, 2003: 170). Algoritma, sık öğe kümelerini birkaç yineleme yaparak hesaplamaktadır. Her yineleme iki adımdan oluşur: (1) sık öğe kümelerini bulma ve (2) kümeleri öncel ve ardıl nesneler olarak ikiye bölerek kurala çevirme (Alpaydın, 2017: 45). Algoritmanın da adı, sık öğe kümelerin öncel "*prior*" bilgisini kullanmasından ileri gelmektedir (Han, Pei ve Tong, 2022: 150; Bhatia, 2019: 247). Apriori algoritması, AIS ve SETM algoritmalarından aday öğe kümelerinin sayılma ve oluşturulma şekli bakımından farklılık göstermektedir (Agrawal ve Srikant, 1994: 488). AIS ve SETM, aday öğe kümeleri veri okunurken geçiş esnasında oluşturulmakta, gereksiz yere çok sayıda aslında küçük olan aday öğe kümesi oluşturulmasına sebebiyet vermektedir. Apriori ise veri tabanındaki işlemleri dikkate almadan önceki yinelemede geniş olan öğe kümelerini kullanarak aday öğe kümelerini oluşturur. Algoritmanın temel motivasyonu geniş öğe kümesinin alt kümesinin de geniş olması gerektiğidir. Böylece k sayıda öğeye sahip aday öğe kümeleri, $k-1$ öğeye sahip geniş öğe kümelerinin birleştirilmesi ve küçük olan alt küme içerenlerin silinmesiyle oluşturulur. Bu prosedür de çok az sayıda aday öğe kümesi üretilmesine imkân verir.

Web madenciliği (link mining/ web mining), web ortamında bulunan veriler veya web etkinliğiyle ilgili World Wide Web üzerinde gerçekleştirilen veri madenciliğidir (Dunham, 2003: 195). Bir diğer tanıma göre, web belge ve hizmetlerinden otomatik olarak bilgi keşfi ve çıkarımı için veri madenciliği tekniklerinin kullanılmasıdır (Kantardzic, 2019: 358). Web madenciliği, metin madenciliğine benzerlik gösterir fakat web madenciliği konu dizinleri ve webteki farklı bilgilerden yararlanarak ürettiği sonuçları iyileştirir (Witten vd., 2017: 519). Daha önce bahsi geçtiği üzere web ortamında

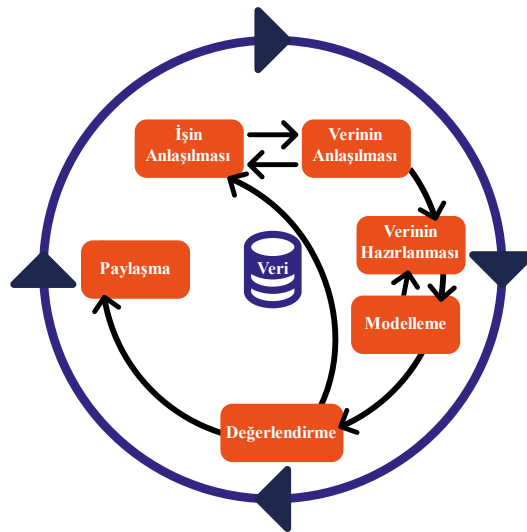
mevcut bulunan veri yoğunluğu veri madenciliği için ciddi bir fırsat sunmaktadır. Bu hem webte bulunan verinin miktarı hem de webte işlem yapan insan veya nesnelerin oluşturduğu trafik sebebiyle gerçekleşmektedir. Web madenciliği uygulamalarında kullanılacak verileri, *içerik*- web sayfalarındaki asıl veriler; *yapı*- içeriğe ilişkin HTML ve XML gibi yapısal veriler; *kullanım*- sayfaların kullanım verileri; *kullanıcı profili*- kullanıcılara ilişkin demografik veriler olarak dört kategoriye ayırmak mümkündür (Srivastava vd., 2000: 12). Web ortamından ihtiyaca uygun olarak elde edilen bu veriler uygun veri madenciliği yöntemleri ile işlenip kullanılabilir. Web madenciliği uygulamalarını da sınıflandırmak gerekirse *içerik (content) madenciliği*, *yapı (structure) madenciliği*, *kullanım (usage) madenciliği* olmak üzere üç sınıf bulunmaktadır (Bhatia, 2019: 369; Scime, 2005: vii). Web madenciliği kategorisinden yapı madenciliği uygulaması olan PageRank algoritması irdelenmiştir.

PageRank, Lawrence Page tarafından 1998 yılında web bağlantı yapısını kullanarak web sayfalarının görece önemini hesaplamak üzere arama, göz atma ve trafik tahmini uygulamalarından oluşan bir yöntem ve Google arama motorunun arkasında yatan algoritma olarak tanıtıldı (Page vd., 1998). PageRank, içerikten ziyade web link yapısının web sayfalarını sıralamada önemli olduğunu vurgulamaktadır (Işık, 2013: 135). PageRank, akademik atıf yaklaşımını web siteleri için uygulamış ve web sayfalarının kalitesini (önem) tespit etmek için o sayfaya verilen bağlantıların sayısını ve bağlantıların kalitesini hesaplamaktadır (Brin ve Page, 1998: 109). Algoritma, bir web sayfasının kalitesini belirlemek için temelde A ve B iki web sayfası olmak üzere A sayfasından B sayfasına giden her bir bağlantıyı (link) A'dan B'ye verilen bir oy olarak yorumlamakta ancak her oy verenin oyu birbirine eşit olmayacak şekilde ayrıca oy verenin de kalitesi hesaba katılarak bir önem/kalite ölçümü yapılmaktadır. Algoritma, bir sorguyla alakalı yalnızca bir sayfanın barındırdığı kelimelerin eşleşip eşleşmediğine bakmaz, aynı zamanda hangi sayfaların kaliteli veya yetkin olduğuna dair bir ipucu olarak sitelerin birbirine nasıl bağlantı verdiğine bakar (Reid, 2023). Yapılan sorguyla eşleşen kayıtların bulunmasının yetersiz gelmesi aslında webte çok büyük miktarda içerik olması ve içerik benzerliği ile spam riskinin artmasından dolayıdır (Wu ve Kumar, 2009: 117). Bir örnekle açıklamak gerekirse, bir işe alım sürecinde saygın bir iş insanı veya siyasetçinin referansı, çok sayıda sıradan insanın verdiği onaydan çok daha yararlıdır. Ancak eğer referansı veren saygın iş insanı benzer şekilde binlerce kişiye referans vermişse bu onun verdiği

tavsiyenin ağırlığını düşürecektir. Yani verilen her oyun ağırlığı (kalitesi), oyu verenin toplam verdiği oy sayısı ile dengelenmektedir (Langville, Meyer, 2006: 27).

2.5. Veri Madenciliği Süreci

Veri madenciliği tanımları yapılırken büyük veri kümelerinden gizli ve faydalı bilgilerin keşfedilme ve çıkarılma **süreci** olduğu hususu üzerinde durulmuştur. Her ne kadar veri madenciliğinde madencilik ve keşif yönleri öne çıksa da veri madenciliğinin tüm genel süreç içerisinde sadece bir adım olduğunu bilmek gerekir (Cabena vd., 1998: 42). Veri madenciliği sürecinde tek bir doğru yolun var olduğunu belirtmek imkansızdır zira literatürde uygulanan birbirinden farklı fakat temelde benzer görevi yerine getiren süreçler bulunmaktadır. Bunlardan bazılarını sıralamak gerekirse, en eski veri madenciliği süreçlerinden biri olan 1996 yılında Fayyad vd. tarafından hazırlanan *KDD Süreci* (Fayyad, Piatetsky-Shapiro ve Smyth, 1996: 41); SPSS önderliğinde geliştirilen CRISP-DM (IBM, 2011); SAS Industry tarafından geliştirilen ve Örnekle, İncele, Dönüştür, Modelle, Değerle (Sample, Explore, Modify, Model, and Assess) adımlarının baş harfinden oluşan SEMMA (SAS, 2006); Cabena tarafından hazırlanan iş süreci (Cabena vd., 1998: 42); Han, Pei ve Tong tarafından geliştirilen ve Veri Hazırlama, Veri Madenciliği, Model Değerlendirme ve Bilgi Sunumu adımlarından oluşan süreç (Han, Pei ve Tong, 2022: 3) örnek olarak verilebilir. Burada sektörde sıkça kullanılan ve standart olarak da benimsenen (Kantardzic, 2019: 9; Witten vd., 2017: 28) CRISP-DM veri madenciliği yaşam döngüsü (Şekil 8) genel hatlarıyla irdelenmiştir.



Şekil 8: CRISP-DM Veri Madenciliği Süreci (IBM, 2011: 1)

CRISP-DM (Şekil 8) 1999 yılında istatistiksel yazılım üreticisi SPSS'in önderliğinde otomobil üreticisi Daimler-Benz, sigorta sağlayıcı OHRA ve bir diğer yazılım üreticisi NCR Corp. firmalarının çalışmaları sonucu geliştirilmiştir (Kantardzic, 2019: 9). CRISP-DM (Cross-Industry Standard Process for Data Mining - Veri Madenciliği için Sektörler Arası Standart Süreç) yaşam döngüsü, bir **yöntem olarak** bir projenin tipik aşamalarını, bu aşamalarla ilgili görevleri ve görevler arasındaki ilişkilerin açıklamasını; bir **model** olarak, veri madenciliği yaşam döngüsüne genel bir bakış sunar (IBM, 2011: 1). Yaşam döngüsü içindeki aşamalar oklarla birbirine bağlı 6 adımdan oluşmaktadır. Adımların sırası kesin olmamakla birlikte çoğu proje için gerektiğinde adımlar arasında esneklik sağlanmaktadır.

İşin anlaşılması, ilk aşamada problem belirleme, fikir edinme, projenin amaç ve ihtiyaçlarının belirlenmesi gerekmektedir. Veri madenciliği uygulanmadan önce madenciliği uygulayarak neyi başarmayı murat edildiği anlaşılmalı, işe yönelik hedef ve gereksinimler araştırılarak hedefleri karşılamak için veri madenciliği uygulanıp uygulanmayacağına karar verilmelidir (Witten vd., 2017: 29). Sorunlar, hedefler ve kaynaklar tespit edilerek çözüme ilişkin muhtemel çıktılar tanımlanmalıdır. Bu aşamada, paydaşlara sunulabilecek bir çıktı için ne tür verilerin toplanabileceği belirlenir. Bu aşamada üç görev verilmiştir; mevcut iş hakkında geçmiş bilgisi topla; karar vericilerin iş hedeflerini belgele; iş açısından veri madenciliği başarısını belirlemek için kullanılan kriterler üzerinde mutabakat sağla (IBM, 2011: 4).

Verinin anlaşılması, bu aşama eldeki veriye daha yakından bakmayı gerektirir. Bir önceki aşamada problemin doğru belirlenmesi, gereksiz veri temini ve işleme zorluklarının önüne geçecektir. Her ne kadar büyük bir veri yoğunluğu içinde yaşıyor olsak da bazı verilerin temini maliyetli ya da zor olabilmektedir. Öte yandan eldeki verinin tümüyle gereksiz verilerle birlikte işlenmesi, zaman, maliyet ve emek kaybına neden olabilecektir. Bu aşamada bir başlangıç veri seti oluşturulur; sonraki aşamalar için uygunluğu incelenir; veri kalitesi düşükse daha sıkı kriterlerle yeni veriler toplanabilir; veriye yönelik iç görüler ve yeniden değerlendirmeler yapılır (Witten vd., 2017: 29).

Verinin Hazırlanması, ham veriden başlayarak nihai veri setine ulaşmaya kadar gerekli tüm işlemleri kapsayacak şekilde, önceden belirlenemeyen sırada çoğu zaman da defalarca gerçekleştirilir (Akpınar, 2014: 75). Hazırlık aşaması, madencilikte

kullanılacak algoritmanın bir örüntü üretebilmesi için ham verilerin ön işlenmesini içerir (Witten vd., 2017: 29). Hazırlık aşaması veri madenciliğinin en önemli ve çoğu durumda en çok zaman alan (zaman ve emeğin yaklaşık %50-70'i) aşamasıdır ve veri setleri birleştirme; örnek bir alt veri kümesi seçme; kayıtların toplanması; yeni niteliklerin türetilmesi; verileri modelleme için sıralama; eksik verileri silme ya da değiştirme; eğitim ve test veri setlerinin ayrılması görevlerini içerir (IBM, 2011: 19).

Modelleme, sürecin başında belirlenen problemin çözümlenmeye başladığı aşamadır. Modelleme genelde çok defa yinleme, varsayılan parametreler üzerinde ince ayar, modelin gerektirdiği veri manipülasyonları gerektirir (IBM, 2011: 25). Bu aşamada uygun modelleme teknikleri seçilir ve uygulanır; sonuçları optimize etmek için model ayarları kalibre edilir; aynı problemler için birkaç farklı teknik kullanılabilir; gerektiğinde veri madenciliği tekniğinin gereksinimlerine uygun hale getirmek için hazırlık aşamasına geri dönülebilir (Larose, 2005: 7).

Değerlendirme, bu aşama problemin belirlenmesinden bu yana gerçekleştirilen sürecin genel değerlendirmesini ve veri madenciliğinden murad edilen sonucun elde edilip edilmediğinin tespitini kapsar. Veriden elde edilen yapısal tanımlar bir tahmin değerine sahip olup olmadığı değerlendirilir şayet değerlendirme modelin zayıf olduğunu gösteriyorsa tüm proje gözden geçirilir ve hedef belirleme, veri toplama dahil önceki adımlara geri dönülebilir (Witten vd., 2017: 29). Üretilen model paylaşılmadan önce kalite ve fayda bakımından test edilir; modelin projenin başında belirlenen hedeflere uygunluğu değerlendirilir; araştırma probleminin tüm yönlerinin hesaba katılıp katılmadığı belirlenir; veri madenciliği sonuçlarının kullanımına ilişkin kararlar belirlenir (Larose, 2005: 7).

Paylaşma adımı madencilikle elde edilen çıktıların (örüntü, model, kalıp) bir ürün olarak hayata geçirilmesini kapsar. Bu bir sonuç raporu olabileceği gibi kurum bilgi sistemine entegre edilerek kullanıma sokulacak uygulama da olabilir. Üretilen model, harici bir yazılıma entegre edilmek üzere ilgililere teslim edilir (Witten vd., 2017: 29). Paylaşma aşaması, çıktıların paylaşımının planlanması; nihai raporun hazırlanması ve projenin gözden geçirilmesi olarak iki aktiviteden oluşur (IBM, 2011: 35). Oluşturulan model kullanıma sunulur; basit bir rapor hazırlanır; gerekirse farklı bir departmanda paralel bir veri madenciliği süreci yürütülür (Larose, 2005: 7).

2.6. Kütüphanelerde Veri Madenciliği- Bibliomining

Bibliomining, kütüphane bilgi sistemlerinden davranışa bağlı eser ve bilgi kaynaklarına bağlı meta veri örüntülerini çıkarmak için veri madenciliği, bibliyometrik, istatistik ve raporlama araçlarının bir bütünüdür (Nicholson vd., 2003: 478). Bir diğer tanıma göre veri madenciliği ve bibliyometrik araçlarının kütüphane hizmetlerinden elde edilen verilere uygulanmasıdır (Nicholson, 2003: 146). Kavram ilk olarak Nicholson ve Stanton tarafından 2003 yılında kullanılmıştır (Nicholson ve Stanton, 2004). Nicholson ve Stanton çalışmasında, kütüphanelerin sadece birer bina olarak algılanmaması gerektiğini, bağlı oldukları üniversite, meclis, şirket ya da toplumun birer parçası olduğu ve bilgi kaynağı sağlayıcısı olarak bu topluluğun entelektüel, öğrenim ve bilgi yönetimi faaliyetlerini küçük bir ölçekte temsil ettiğine vurgu yapmış; kullanıcı beklentilerinin karşılanmasında veri madenciliği uygulamalarının önemine işaret etmiştir.

Veri madenciliğinin kütüphanecilik alanında kullanılması şaşırtıcı bir gelişme değildir zira verinin bulunduğu hemen her sektörde veri madenciliği uygulanmaktadır. Bibliomining kavramı da esasen masaya yeni bir araç koymamaktadır. Her sektörde olduğu gibi veri madenciliği veri setinde hangi nitelikte veri bulunuyorsa ona uygun örüntüler ortaya koyacaktır. Kütüphanecilik camiasında da koleksiyon, kullanıcı ve personele ilişkin verilerin tutuluyor olması madencilik faaliyetinin de bu aktörlere yönelik olarak gerçekleşeceğini göstermektedir. Bibliomining teriminin kullanımından çok önce de veri madenciliği uygulamalarının kütüphanelerde kullanımına yönelik çalışmalar bulunmaktadır (Banerjee, 1998). Kütüphane bilgi sistemlerindeki kayıtlı işlemler üzerinden örüntü, davranış değişikliği ve trendleri izlemek için kullanılan kavram yeni olmamakla birlikte kütüphane ve veri madenciliği terimlerinin bilgisayar bilimlerinde kullanılan “software libraries” (yazılım kütüphanesi, kod koleksiyonu) kavramından ziyade kütüphanecilik bağlamında değerlendirilmesini kolaylaştırmak amacıyla türetilmiştir (Nicholson, 2003: 146).

Veri madenciliği uygulama alanlarına bakıldığında (pazarlama, sağlık, finans, otomotiv vs.) genelde kullanıcı kurumların finansal bazı beklentilerinin olduğu anlaşılmaktadır. Yani firma elindeki ürün her ne ise bunun daha fazla satılması ya da daha fazla kar bırakması maksadıyla müşterilerin bazen davranışlarını bazen alışkanlıklarını tahmin etme veya yönlendirme gayretindedir. Ancak kütüphaneler yapısı gereği (büyük

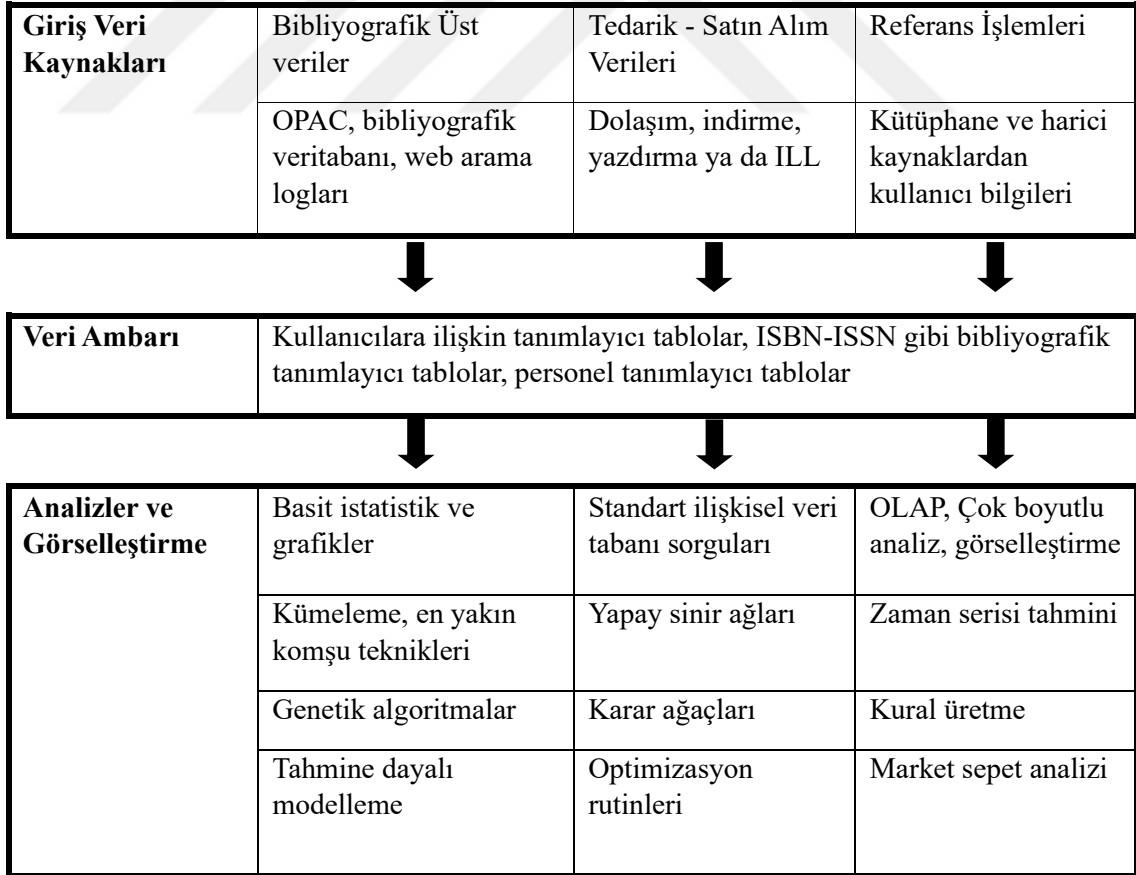
oranda) kâr amacı gütmeyen kurumlardır. Her ne kadar kar amacı gütmeyen kurum olsalar da kütüphaneler hizmet ve politikalarını günün şartlarına uygun olarak güncellemek ve kurumsal değerini, misyon ve hedeflere olan katkısını ispata ihtiyaç duymaktadır (Oakleaf, 2010: 6). Belki bugün kütüphanelerin kalitesini ölçen bir araştırma bulunmamakta fakat kütüphanelerin bağlı olduğu kurumlar, özellikle de üniversiteler çeşitli ölçüm yöntemleriyle sıralamaya tabi tutulmaktadır. Üniversite kütüphaneleri daha önce bahsi geçtiği üzere “*bağlı olduğu kurumun fakülte, öğrenci ve personelinin araştırma ve müfredat ihtiyaçlarını karşılamak*” (Reitz, 2004: 743; Levine-Clark ve Carter, 2013: 263) gibi misyonu olması hasebiyle üniversitenin akademik başarısındaki önemli aktörlerden birisidir. Bu bakımdan da koleksiyon ve kullanıcıların -terim user, patron ve customer olarak da kullanılabiliyor (Bell, 2019)- muhtemel örüntülerinin ortaya çıkarılması verilen hizmet ya da ürünün değerinin yükseltilmesi için kıymetlidir. Nicholson yaptığı çalışma ile kütüphanelerin kullanıcılar ve hizmetleri hakkında kapsamlı bir betimleme kazanmaları için dört başlıktan oluşan ölçüm matrisi geliştirmiştir (Nicholson, 2004):

- Kütüphane sisteminin iç perspektifi (süreç, hizmet analizi), bibliyografik koleksiyonun tüm yönleri, örgütsel akışlar, iş akışları, personel, bilgi teknolojileri, kaynaklar ve imkanlar gibi kütüphanenin sunduğu süreç ve hizmetlerin analizini içerir. Kütüphane sisteminin nelerden oluştuğu sorusuna cevap aranır.
- Kütüphane sisteminin dış perspektifi (kalite analizi), kütüphane hizmetlerinin kullanıcılar tarafından değerlendirilmesini içerir. Kullanıcı algıları keşfedilerek hizmetlerin bilinirliği, ilgililiği, uygunluğu ve kullanılabilirliği değerlendirilir. Kütüphane sistemi ne kadar etkilidir sorusuna cevap aranır.
- Kütüphane koleksiyonunun dış perspektifi (koleksiyon analizi), kütüphane koleksiyonunun kullanıcılar üzerindeki etkisinin ölçülmesini sağlar ve bu ölçüm koleksiyon geliştirme aşamasında daha nitelikli bilgi kaynağı temini için kriterler sunar. Kütüphaneciler mevcut koleksiyonu değerlendirerek eksikleri tespit edebilir ve koleksiyon gelişimini daha iyi yönetebilir (Agee, 2005: 92). Kütüphane sistemi ne kadar fayda sağlar sorusuna cevap aranır.

- Kütüphane koleksiyonunun iç perspektifi (kullanım analizi), kullanıcıların kütüphane sistemi ile olan etkileşimi, kütüphane hizmetlerini kişiselleştirme tercihlerini analiz eder. İşlem logları, web kullanımı, ödünç iade istatistikleri ile kullanıcıların hangi kaynaklar için sorgu gönderdiği, hangi kaynaklara erişebildiği veya hangi konu ve kaynaklara erişemediği bilgileri temin edilir. Kütüphane sistemi nasıl kullanılıyor sorusuna cevap aranır (Nicholson, 2004: 174).

Nicholson tarafından önerilen bu dört ayaklı değerlendirme matrisi kütüphanelerin, koleksiyon, kullanıcı ve hizmetler bakımından kapsamlı değerlendirmesini içermektedir. Veri madenciliği uygulamaları ile birliktelik analizleri, web madenciliği, tavsiye motoru, kümeleme veya sınıflandırma teknikleriyle koleksiyon gelişiminden kullanıcı beklenti analizine, sahtekarlık tespitinden süreç analizine kadar pek çok değerlendirme yapılabilecektir.

Tablo 3: Bibliomining Veri Akışı (Nicholson ve Stanton, 2004: 252)



Tablo 3'te verilerin toplanmasından veri ambarı oluşturulması ve analizlerin yapılmasına kadar kütüphanelerde uygulanabilecek veri madenciliği iş akışının genel bir örneği verilmiştir. Çalışmanın 2004 yılında gerçekleştirildiği göz önüne alındığında hem veri kaynaklarında hem de analiz yöntemlerinde gelişmeler olduğunu unutmamak gerekir. Zira kütüphaneler tabloda verilenlerin dışında -örneğin sosyal medya- farklı kaynaklardan gelen veriler maharetiyle madencilik faaliyeti yürütebilmektedir. Tabloda koleksiyon, dolaşım ve kullanıcı verilerinden oluşan üç girdinin madencilik sürecinde akışını göstermektedir. Buna göre diğer sağlayıcılar tarafından üretilen frekans tablolarının çok ötesinde veri madenciliği ile personel, kütüphane kullanımı, koleksiyon yönetimi, koleksiyon geliştirme, kullanıcı beklenti, davranış ve eğilimleri hakkında çok daha fazla bilgi edinebilir; geliştirilmiş kütüphane hizmetleri ile kullanıcılara, karar desteği ile kütüphane yöneticilerine, kullanıcı davranışlarının raporlanması ile üst kuruma potansiyel faydalar sağlar (Nicholson ve Stanton, 2004: 253).

Veri madenciliği uygulanabilecek kütüphane verileri, kütüphanenin mevcut verileri, koleksiyon kullanımından elde edilen veriler, kütüphane otomasyonuna dahil olmayan harici veriler olmak üzere üç kategoriye ayırmak mümkündür (Nicholson ve Stanton, 2009: 154). Bunları açıklamak gerekirse:

Mevcut kütüphane verileri, koleksiyona ait bibliyografik veriler ve tedarik verilerinden oluşmaktadır. *Bibliyografik veriler*, OPAC (Online Public Access Catalog-Çevrimiçi Genel Erişim Kataloğu) üzerinden kullanıcıların eriştiği basılı kaynak kataloğuna ilişkin veriler, elektronik kaynakların veri tabanlarında tutulan meta verilerini ihtiva etmektedir. *Tedarik verileri*, bilgi kaynaklarının sipariş aşamasından kütüphane kataloğuna işlendiği aşamaya kadar takip edilen süreç içerisinde üretilen verilerden oluşur. Benzer şekilde elektronik kaynakların sipariş ve temininde üretilen veriler de buna dahildir.

Kütüphane otomasyonunun kullanımından oluşan veriler, kullanıcı verileri, dolaşım/ kullanım verileri, arama ve gezinme verilerinden oluşmaktadır. *Kullanıcı verileri*, kütüphaneye üye kullanıcıların tanımlayıcı verilerinin tutulduğu tablolardır. Farklı kütüphane türlerinde farklı veriler tutulmakla birlikte genel olarak kimlik, iletişim ve eğitim bilgileri otomasyonda tutulmaktadır. Web 2.0 ile kullanıcılar kişisel tercihlerine yönelik bilgileri de kullanıcı tablolarına ekleyebilmektedir. *Dolaşım/ kullanım verileri*,

kütüphanenin kullanıcılara sağladığı bilgi kaynağı ve kitap dışı materyallerin ödünç-iadesiyle ilgili verilerin tutulduğu tablolardır. Kullanıcı davranış, eğilim ve trendlerine yönelik örüntülerin çıkarılabileceği en zengin veriler dolaşım verileridir. Yasal ve etik kısıtlar dahilinde bireysel mahremiyete dikkat edilerek dolaşım verilerinden yararlanmak veri madenciliği için faydalı olacaktır. Benzer şekilde elektronik kaynakların kullanımı da örüntü çıkarımı için faydalı veriler sağlamaktadır. Ancak elektronik kaynaklarda, tam metin erişimi için çoğu platformun oturum açma zorunluluğu bulundurmaması, tam metni görüntülemenin her durumda kaynağın gerçekten kullanıldığını göstermemesi gibi özel durumları bulunduğunu hesaba katmak gerekir. *Arama ve gezinme verileri*, kullanıcıların OPAC üzerinde gerçekleştirdiği sorguları içermektedir. Kullanıcının OPAC üzerinde yaptığı aramaları, koleksiyondan eriştiği veya erişemediği içeriklerin takibi mümkündür. Buna ek olarak elektronik kaynakların sunulduğu platformlar kullanıcı loglarını tutmakta ve loglardan tıklama akış analizi gerçekleştirilebilmektedir.

Harici veri kaynakları, danışma kütüphanecisi etkileşimleri, materyal kullanımı ve kütüphaneler arası ödünç verilerinden oluşmaktadır. *Danışma kütüphanecisi*, genel kütüphane otomasyonunda tutulmayan fakat yüz yüze veya uzaktan (telefon, chat, eposta) kütüphane hizmeti vermektedir. Herhangi bir kayıt tutulmadığında yüz yüze ve telefonla verilen hizmetler için madencilik uygulanamasa da dijital ortamda verilen hizmetlerin verileri kullanılabilir. Referans hizmetlerinin kütüphane otomasyonuna entegre hata/hizmet takip modülleri veya harici yazılımlar üzerinden verilmesi kayıt dışı sorununu çözecektir. *Materyal kullanımı*, dolaşım tablolarında tutulmayan, ödünç alınmadan kütüphane içi kullanım verilerini ifade etmektedir. Bilgi kaynağı veya diğer materyallerin her türlü kullanımı için veri tutulması kurum içi kullanımın izlenmesi ve kullanıcı davranışlarının keşfedilmesi için önemlidir. *Kütüphaneler arası ödünç (ILL-Inter Library Loan)*, kullanıcıların kütüphanenin sahip olduğu bilgi kaynaklarının dışında, kütüphanenin iş birliği yürüttüğü diğer kurumlardan kaynak temini sürecinde oluşan verileri ifade eder.

Kütüphane bilgi sistemleri, davranış tabanlı tavsiye hizmeti için uygun bir veri madenciliği alanıdır ve OPAC, dolaşım verileri, web logları kullanılarak Amazon, Netflix gibi kullanıcı odaklı hizmet portalları geliştirilebilir; kullanıcılar için kaynak temin ve kullanımında, kütüphaneciler için de kullanıcı desteği ve koleksiyon yönetimi açısından faydalar sağlayacaktır (Geyer-Schulz, Neumann ve Thede, 2003: 165).

ÜÇÜNCÜ BÖLÜM

VERİ MADENCİLİĞİ YÖNTEMLERİ İLE BİLGİ MERKEZİ VERİLERİNDEN BİLGİ KEŞFİ

Kurum, kuruluş ve organizasyonlarda yürütülen iş ve işlemler neticesinde her an çeşitli formatlarda veriler üretilmekte ve depolanmaktadır. Hem kurum içi hem kurum dışı kaynaklar vasıtasıyla üretilen veriler organizasyonların amaç ve hedeflerine yönelik faydalı ve ilginç örüntülerin ortaya çıkarılması için kullanılmaktadır. Bilgi merkezleri de hem organizasyonel iş süreçlerinde üretilen veriler hem de kurum kullanıcılarının yaptığı işlemler sonucu pek çok veri üretmektedir. Günün koşullarına ayak uydurabilmek ve kullanıcıların ihtiyaçlarını karşılayabilmek adına kütüphane ürün ve hizmetlerinin geliştirilebilmesi için kullanıcı kitlesinin iyi tanınması, davranış ve eğilimlerinin belirlenmesi hem maliyetler hem de doğru yatırımların yapılabilmesi için kritik öneme sahiptir.

3.1. Araştırmanın Amacı ve Önemi

Kütüphane, bilgiye ihtiyaç duyan her bilgi arayıcısının ihtiyaçlarına yönelik olarak gereksiz gecikmeler olmaksızın kaynakları sağlayan geniş kitap koleksiyonudur (Garrett, 1999: 116). Kütüphane biliminde temel düşünce, her bilgi kaynağı için benzersiz bir tanımlayıcı, iyi bir katalog ve katalog kaydından bilgi kaynağının bulunduğu konuma bağlantı vererek bilgiye ihtiyaç duyan kullanıcıların, kütüphanecinin yardımına ihtiyaç duymadan koleksiyondan tatmin edici şekilde yararlanabilmesidir (Buckland, 2017: 6). Kullanıcıların ihtiyaç duyduğu bilginin karşılanabilmesi için hem bilgi kaynaklarının temin edilmesi hem de bilgi kaynaklarının etkin şekilde kullanıcıların hizmetine sunulması gerekmektedir. Pek tabi her kullanıcının talep ettiği bilgi kaynaklarını istisnasız şekilde temin etmek fiziksel, maddi, tarihsel koşullar sebebiyle mümkün değildir ancak kütüphaneler bilgi hizmetinin maksimize edilmesini amaç edinmiş ve eldeki imkanlar doğrultusunda olabildiğince fazla nitelik ve nicelikte bilgi kaynağını kullanıcıya ulaştırmanın gayesi içindedir. Bu konuda Ranganathan'ın ileri sürdüğü Kütüphane Biliminin 5 Yasası'nda, her kitabın bir okuyucusu olduğu, her okuyucunun bir kitabı olduğu ve okuyucunun zamanından tasarruf edilmesi gerektiği (Ranganathan, 2006); bilgi merkezlerinin kullanıcıların ihtiyaçlarına yönelik kaynakların temin edilmesi ve bunların kullanıma sunulurken zamandan tasarruf edilmesi gerektiğini

vurgulamaktadır. Hem bireyler, hem şirketler hem de devletler için yararlı bilgiye hızla erişim sağlamak, ulusal ve uluslararası pazarlarda rekabet edebilirliğin temel koşullarından biri haline gelmiştir (Tonta, 1999: 366).

Bu araştırmanın amacı, kütüphane otomasyon sistemlerinde tutulan verilerden veri madenciliği yöntemleri ile kullanıcı davranışları, kullanıcı örüntüleri, koleksiyon örüntüleri ve trendleri ortaya çıkarmak; bilgi merkezlerinde koleksiyon geliştirme, koleksiyon yönetimi ve bilgi kaynaklarının etkin kullanımı için yol haritası ortaya koymaktır. Araştırmanın iki motivasyon odağı vardır: (i) kullanıcıların bilgi kaynaklarını kullanma eğilimleri üzerinden her bir kullanıcıya ilgi odağına yönelik bilgi kaynaklarını tavsiye etmek; (ii) kullanıcıların beklenti ve eğilimlerine yönelik ürün ve hizmetlerin geliştirilmesinde kütüphane yönetimine ve üst yönetime karar desteği sağlamaktır.

Daha önce bahsi geçtiği üzere tüketiciler büyük bir bilgi bombardımanı ile karşı karşıyadır. Her an her yerde bilgi kaynakları ile muhatap olmakta ve ihtiyaç duyduğu bilgi için zaman, emek ve ücret harcamaktadır. Bunun yanında içerik sağlayıcılar kullanıcı beklenti ve ihtiyaçlarını karşılamak için rekabet halinde çeşitli ürünler ortaya koymaktadır. Kullanıcı eğilimlerine yönelik olarak kişiselleştirilmiş öneriler, kullanıcı deneyimini üst seviyeye çıkartmak için Amazon ve Netflix başta olmak üzere pek çok içerik sağlayıcı ve e-ticaret platformunun ayrılmaz ve fark yaratan birer parçası haline gelmiştir (Koren, Bell ve Volinsky, 2009: 30). Dijital içerik sağlayıcı platformlarda kullanıcıların kişisel öneri sistemlerini sıklıkla kullanıyor olması, bilgi merkezleri için de cazip bir uygulama alanıdır. Netflix içerik kütüphanesinde olduğu gibi kütüphane OPAC platformunda da kullanıcıya özel kişiselleştirilmiş önerilerle bilgi kaynaklarının sunulması kullanıcı deneyimini üst seviyeye çıkaracaktır. Kütüphane kullanıcıları çoğu zaman hangi bilgi kaynağına ihtiyaç duyduğunun farkında değildir. Buna karşın bilgi merkezleri, kullanıcıların kitaplarla yaptığı etkileşimden (ödünç işlemi) hareketle Netflix'in öneri sisteminin bir parçası olan PVR (Personalized Video Ranker - Kişiselleştirilmiş Video Sıralayıcısı) (Gomez-Urbe ve Hunt, 2016: 3) benzeri bir öneri sistemi ile kitap kataloğunu kullanıcı profiline göre özelleştirilmiş şekilde sunabilir.

Çalışmada veri madenciliği ile bilgi merkezlerinde kullanıcı eğilimlerinin belirlenerek hem kullanıcı deneyimini yükseltmek hem de kurum yöneticilerine koleksiyon ve kullanıcının mevcut durumu veya geleceği hakkında fikir vermesi

amaçlanmıştır. Bu çalışmada kütüphane otomasyon sistemlerinde tutulan ve koleksiyon, kullanıcı ve ödünç kayıtlarından elde edilen veriler ile kullanıcı beklenti, davranış ve eğilimleri veri madenciliği ile irdelenmekte; kullanıcıların eğilimlerine yönelik kitap öneri sistemi sunulmaktadır.

3.2. Literatür Taraması

Veri madenciliği, veri kavramından söz edilebilen her disiplinde başarılı bir uygulama sahası bulmakta (Han, Kamber ve Pei, 2012: 27), pazar yönetimi, risk yönetimi, dolandırıcılık tespiti (Cabena vd., 1998: 26), web madenciliği, görüntü işleme, pazarlama, satış (Witten vd., 2017: 21), sosyal medya, biyoloji, tıp ve sağlık (Han, Pei ve Tong, 2022: 17) gibi sayısız örnek sıralanabilmektedir. Kütüphanecilik alanında veri madenciliği uygulamaları *bibliomining* kavramı altında toplansa da kavramın ortaya çıkışından önce de veri madenciliği alanının gelişimine paralel olarak uygulamalar gerçekleştirilmiştir (Banerjee, 1998). Literatürde gerek kütüphane koleksiyonuna yönelik gerekse kullanıcılara yönelik yapılan pek çok araştırma vardır. Aşağıda bu araştırmalardan öne çıkanlar irdelenmiştir.

Hollanda’da bulunan 5 halk kütüphanesinin ödünç kayıtlarının kullanıldığı bir çalışmada kullanıcı no, eser adı ve numarası, ödünç tarihi, yaş ve cinsiyet değişkenleri dahil edilerek ultrametrik ağaç (ultrametric trees) yöntemi kullanılmış ve kullanıcılar 4 kümeye ayrılmıştır (Boter ve Wedel, 2005). Çalışmada üye id, eser adı, ödünç tarih ve saati, yaş ve cinsiyet değişkenleri kullanılması nedeniyle bu araştırma ile benzerlik göstermektedir. Ancak bu çalışmada analizin çeşidine göre akademik birim, grup, konu başlığı gibi daha fazla değişken de kullanılmıştır.

Chen ve Chen (2007) tarafından yapılan çalışmada karınca kolonisi kümeleme algoritması ve birliktelik kuralları kullanılarak kullanıcıların ilgi alanlarına yönelik kitap öneri sistemi tasarlanmıştır. Araştırma hem kullanıcılar için kitap önerileri sunma hem de kütüphane yöneticileri için etkin bütçe yönetimi için öneriler sunmaktadır.

2010 yılında Akdeniz Üniversitesi’nde yapılan çalışmada (Uçan, 2010) apriori algoritması kullanılarak birliktelik analizi yapılmış ayrıca two-step algoritması ile de kümeleme analizi uygulanmıştır. Kümeleme analizinde bölüm, cinsiyet, grup ve toplam ödünç sayısı değişkenleri kullanılmış ve veri seti 4 kümeye ayrılmıştır.

2014 yılında Japonya’da bir üniversite kütüphanesinde yapılan çalışmada SVM (support vector machine) algoritması ve kütüphane ödünç kayıtları ile kitap başlıkları kullanılarak yapılan araştırma sonucunda kitap önerileri oluşturulmuştur (Tsuji vd., 2014). Araştırmada birliktelik kuralları, kitap başlıklarına dayalı benzerlik, sınıflama sistemindeki kategoriler arasındaki uyum ve kitap özetlerinin benzerlikleri kullanılmış; kullanıcılara kitap önerilerine yönelik değerlendirmeleri sorulmuştur. Deneklerin kitap önerilerini genel olarak olumlu buldukları, ancak Amazon e-ticaret platformunun kitap önerilerini daha olumlu değerlendirdikleri saptanmıştır.

2014 yılında yapılan bir diğer çalışmada New York University kütüphaneleri tarafından kullanılan sanal referans hizmetlerinden toplanan veriler ile *k*-means algoritması kullanılmış ve kullanıcılar iki farklı uygulama ile 3 ve 4 kümeye ayrılmıştır (Tempelman-Kluit ve Pearce, 2014). 2019 yılında Hollanda’da yapılan çalışmada bir halk kütüphanesi kullanıcılarına yönelik kümeleme analizi yapılmıştır (Glind, 2019). 2018 yılına ait veriler işlenerek kullanıcıların kütüphaneyi kullanma alışkanlıkları *k*-means algoritması ile irdelenmiş ve kullanıcılar 6 kümeye ayrılmıştır. 2019 yılında yapılan bir diğer çalışmada Erzincan Binali Yıldırım Üniversitesi Merkez Kütüphanesi otomasyon sisteminden elde edilen veriler ile veri madenciliği uygulaması yapılmış; two-step algoritması ile kümeleme analizi yapılmış, ağ grafiği yöntemiyle ise birliktelik analizi uygulanmıştır (Fırat, 2019). Kümeleme analizinde kullanıcılara ilişkin fakülte, grup, cinsiyet, ödünç sayısı ve konu başlığı değişkenleri kullanılmış ve veri seti 3 kümeye ayrılmıştır. Benzer bir çalışma 2020 yılında Trakya Üniversitesi Merkez Kütüphanesi’ne yönelik yapılmıştır (Taşkın, 2020). Çalışmada FP-Growth algoritması ile birliktelik analizi yapılmış, *k*-means algoritması ile de kümeleme analizi yapılmıştır. Kümeleme analizinde 23227 kayıt ile cinsiyet, eğitim programı, bölüm, yayın teslim durumu gibi kategorik değişkenler kullanılmış ve kullanıcılar 2 kümeye ayrılmıştır.

İşbirlikçi filtreleme yöntemlerinin kullanıldığı çalışmalara bakıldığında 2023 yılında University of Gondar’da öğrenci bilgi sistemi ve çevrimiçi katalogdan temin edilen veri seti ile kNN (k-nearest neighbors) ve SVD (singular value decomposition) algoritmaları kullanılarak kitap öneri sistemi geliştirilmiştir (Ahmed ve Letta, 2023). Çalışma bu araştırma ile benzerlik gösterse de kullanılan veri seti bakımından ayrılmaktadır. Zira araştırma kullanıcıların kitaplara verdikleri değerlendirme puanları üzerinden çalışmaktadır. Çalışma sonunda her iki algoritma için MAE ve RMSE

performans deęerleri paylařılmış; SVD algoritmasının kNN algoritmasına gre daha yksek performans gsterdięi saptanmıřtır. Benzer bir alıřma 2019 yılında Finlandiya’da yapılmıř ve Vantaa City Library otomasyon sisteminden temin edilen kullanıcı ve dn kayıtlarından oluřan veri seti ile SVD, SVD++, kNN ve slopeone gibi farklı neri algoritmalarını uygulayarak RMSE (Root mean square deviation - kk ortalama kare hatası) ve MSE (mean squared error - ortalama kare hatası) performans deęerlerini hesaplamıřtır (Satyal, 2019). Zaman faktr gz ardı edildięinde SVD++ algoritmasının dięer algoritmalara gre daha iyi performans gsterdięi belirlenmiřtir.

3.3. Arařtırmanın Varsayımı ve Sınırlılıkları

Ktphane kullanıcılarının davranıř, eęilim ve beklentilerinin veri madencilięi ile irdelendięi bu alıřmada, arařtırmanın genel yapısı ve evreni hakkında varsayım ve sınırlılıklar bulunmaktadır. alıřmanın varsayımları:

- Ktphane otomasyon sisteminde bulunan tm kayıtların ilgili yeye ait olduęu ve kayıtlar zerinde maniplasyon yapılmadıęı varsayılmıřtır. Kullanıcıların farklı kiřiler zerinden kaynakları dn aldıęı, sipariř verdięi veya sorguladıęı gibi durumlar gz ardı edilmiřtir.
- Dolařım kaydında bulunan kayıtlara gre kullanıcı tarafından dn alınan tm kaynakların kullanıldıęı varsayılmıřtır. dn alınan her kaynaęın tamamen okunup okunmadıęı, kullanılıp kullanılmadıęı irdelenmemiřtir.
- Ktphane otomasyon sistemin tutulan tm verilerin doęru ve gerek veri olduęu varsayılmıřtır.

alıřmanın sınırlılıkları:

- alıřma, ktphaneye ye kullanıcıların (akademisyen, idari personel ve ęrenci) verileri ile sınırlıdır. Ktphaneye ye olmayan niversite mensupları alıřmanın kapsamı dıřındadır.
- Analizler, ktphane otomasyon sistemine iřlenen dolařım verileriyle sınırlıdır. Kullanıcıların ktphane ierisinde dn almadan kullandıęı bilgi kaynakları alıřma kapsamı dıřındadır.

- Çalışma, kullanıcıların basılı kaynaklara yönelik kullanımları ile sınırlıdır. Elektronik bilgi kaynaklarının kullanımı (e-kitap- e-dergi, e-tez, e-gazete vs.) çalışma kapsamı dışındadır.
- Çalışma, basılı kitaplarla sınırlıdır. Basılı süreli yayın, tez ve kitap dışı materyaller ödünç verilmeden kullanıldığı için kapsam dışı bırakılmıştır.
- Çalışmada basılı kitapları temsil etmek için demirbaş, ISBN ve eser adı değişkenleri irdelenmiş ve eser adının baz alınması çalışmanın özelliği açısından daha uygun bulunmuştur. Zira aynı esere ait farklı demirbaş almış kopyalar bulunduğu gibi aynı adlı eserin farklı ISBN ile yayınlanması bu kararda etkili olmuştur.
- Çalışma, Burdur Mehmet Akif Ersoy Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı kütüphane otomasyon sisteminde 2008-2024 yılları arasında tutulan veriler ile sınırlıdır.

3.4. Araştırmanın Evreni

Bir araştırma için evren (population), problemlerin çözümü için kullanılacak verilerin elde edildiği canlı/ cansız varlıklardan müteşekkil büyük gruptur (Büyüköztürk vd., 2015: 80). Evren, analiz edilmek üzere içinden yansız örneklem alınabilen gruptur (Tabachnick ve Fidell, 2015: 7). Araştırmanın evreni, Burdur Mehmet Akif Ersoy Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı tarafından kullanılan kütüphane otomasyon sisteminde (Yordam Kütüphane Bilgi Belge Otomasyonu) tutulan kütüphane üyelerine ait verilerden oluşmaktadır. Otomasyon sistemi 2008 yılından itibaren gerçekleştirilen kütüphane işlemlerine ait verileri barındırmaktadır. Veri madenciliği tüm veri setine uygulandığı için çalışmada örneklem seçimi yapılmamıştır.

3.5. Veri Toplama Yöntemi ve Aracı

Kütüphane kullanıcılarının basılı kitap kullanım davranış, eğilim ve trendlerine ilişkin örüntüleri ortaya koymayı amaçlayan araştırmada veri madenciliği algoritmalarından *k*-means ve SVDs kullanılmıştır. Araştırmada kullanılan veri seti Burdur Mehmet Akif Ersoy Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı tarafından kullanılan Yordam Kütüphane Bilgi Belge Otomasyonundan temin edilmiştir. Veri tabanında 2008 yılından itibaren kayıt altına alınan koleksiyon, üye, ödünç-iade başta olmak üzere sipariş, web aramaları gibi koleksiyon ve kütüphane kullanıcılarına

ilişkin pek çok veri tutulmaktadır. Kütüphanenin varlığı, Burdur Mehmet Akif Ersoy Üniversitesi'nin kuruluşundan önce Süleyman Demirel Üniversitesi'ne bağlı fakülte kütüphanesine dayanması bakımından koleksiyona ait kayıtlar 2000 yılına kadar gitmektedir. Bu yıllarda kullanılan otomasyon sistemindeki veriler mevcut otomasyon sistemine aktarılarak bugüne gelmiştir. Üye ve dolaşıma ait veriler ise 2008 yılından bugüne uzanmaktadır. Araştırma kapsamında kütüphane otomasyon sisteminden Mart 2024 tarihine kadar var olan veriler temin edilmiştir.

Yordam Kütüphane Bilgi Belge Otomasyonu, Claris International Inc. ürünü Filemaker veri tabanı uygulaması temelli olarak YordamBT tarafından geliştirilen MARC uyumlu kütüphane otomasyon sistemidir. Otomasyon sistemi, merkezi bir sunucu ve ona bağlı istemcilerde kurulu Filemaker yazılımı üzerinden çalışmaktadır. İçeriğinde kütüphane iş süreçlerinin pek çoğu yürütülebilecek modüller bulunmakta ve özellikle Türkiye'de hatırı sayılır kullanım oranına sahiptir (Tablo 1). Otomasyon sistemi, verileri .tab, .csv, .xlsx, .htm gibi formatlarda dışarıya aktarabilmektedir. Uygulama kolaylığı açısından veriler .xlsx formatında dışarıya aktarılıp işlemler yapılmıştır. Temin edilen veri setleri MS Excel uygulamasında ön incelemeden geçirilmiş, gereksiz ve tekrar eden sütunlar için temizleme ve dönüştürme işlemi yapılmıştır. Sonrasında araştırmanın amacına uygun olarak kişiselleştirilmiş kitap öneri sistemi için *k*-means kümeleme analizi ve SVDs işbirlikçi filtreleme analizi uygulanmıştır. Kitap öneri sistemi için Spyder IDE v5.5.4; tanımlayıcı istatistikler içinse RapidMiner Studio Educational 10.3 uygulamaları kullanılmıştır. Veri setine ilişkin tanımlayıcı bilgiler aşağıda verilmiştir (Tablo 4).

Tablo 4: Veri Seti

Veri Seti	Nitelik Sayısı	Kayıt Sayısı	İçerik
Katalog	125	84849	Koleksiyonda bulunan basılı kitaplara ilişkin bibliyografik veriler
Üye	112	75233	Kütüphane üyelerine ait tanımlayıcı bilgileri
Dolaşım	99	222540	Kitap ödünç-iade işlemlerine ait bilgiler

Kütüphane otomasyon sisteminden temin edilen veri setinde birbirini tekrar eden ve boş kayıtlarla birlikte dolaşım (ödünç) tablosunda toplam 99 sütun bulunmaktadır. Benzer şekilde üye tablosunda 112, katalog tablosunda ise 125 nitelik sütunu

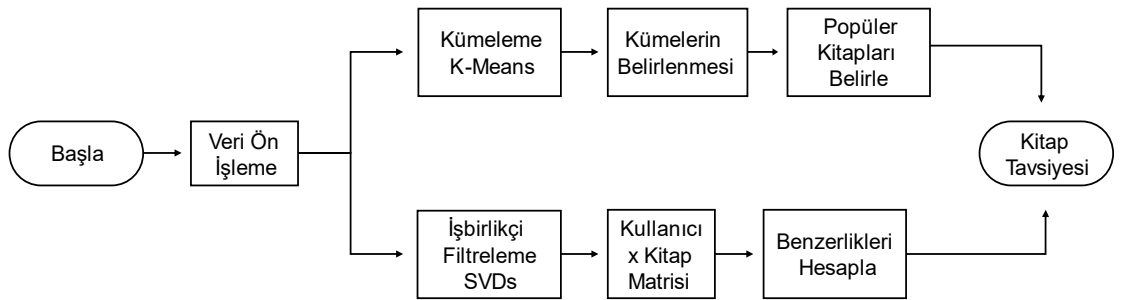
bulunmaktadır. Temin edilen veriler ön incelemeden geçirildikten sonra tüm alanların iş bu veri madenciliği uygulaması için gerekli olmadığı değerlendirilmiş ve araştırmanın amacına uygun nitelikler ön işleme süreci sonunda temizlenerek analizler yapılmıştır. Kullanıcı kimlik bilgileri gizlilik ihlaline mahal vermeyecek şekilde korunmuş ve her bir üye için tekil id verilerek analizler anonim olarak yürütülmüştür.

3.6. Araştırmanın Modeli ve Soruları

Araştırma, kütüphane kullanıcılarının eğilimlerini ortaya koymak ve benzer kitap okuma eğilimine sahip kullanıcılara kitap önerileri sunmak maksadıyla kullanıcı segmentlerinin ortaya koyulmasını ve kullanıcı-kitap etkileşiminin belirlenmesini kapsamaktadır. Çalışmanın amacı daha önce bahsedildiği üzere kütüphane kullanıcılarının davranış ve eğilimlerini; koleksiyona ilişkin örüntü ve trendleri ortaya çıkarmaktır. Bu amaç doğrultusunda araştırma soruları şu şekilde belirlenmiştir:

- Kütüphane kullanıcı kümeleri nelerdir?
- Kullanıcı kümelerinin özellikleri nedir?
- Kullanıcı kümelerinin en çok ödünç aldığı kitaplar hangileridir?
- Kullanıcıların demografik özellikleri kitap okuma davranışını etkiler mi?
- Kişiselleştirilmiş bir kitap öneri sistemi oluşturmak mümkün müdür?
- Oluşturulan kitap öneri sistemlerinin performans değerleri nedir?

Araştırma soruları kapsamında kullanıcıların davranışsal benzerliklerine yönelik kümelerin oluşturulması için veri madenciliği yöntemlerinden kümeleme ve kullanıcıların kitap tercihleri bakımından benzerliklerini ortaya koymak için işbirlikçi filtreleme analizi kullanılmış; veri seti ve kullanıcı kitlesine yönelik tanımlayıcı istatistikler verilmiştir.



Şekil 9: Araştırma Modeli

3.7. Veri Ön İşleme

Veri ön işleme, veri analizi süreçlerinin hızlıca gerçekleştirilebilmesi ve analiz sonuçlarının kalitesinin artırılması amacıyla toplanan ham veri üzerinde gerçekleştirilen *veri entegrasyonu, veri temizleme, veri dönüştürme ve veri indirgeme* işlemlerine verilen isimdir (Akpınar, 2014: 88). Araştırma kapsamında analizlerde kullanılmak üzere temin edilen veri seti tek bir kaynaktan (Burdur Mehmet Akif Ersoy Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı) sağlandığı için farklı kaynaklardan temin edilen verilerin entegrasyonuna ihtiyaç duyulmamıştır. Ancak kütüphane otomasyon sisteminin farklı modüllerinden alınan veri setleri analizlerin içeriğine göre birleştirilmiştir. Veri temizleme işlemi en çok zaman alan uygulamalardan biri olmuştur. Kütüphane otomasyon sisteminin 2017 yılında getirilen bir dizi güncellemeden önce var olan veri giriş alanlarına manuel müdahale imkânı, kayıtlardan çokça standart dışı verinin oluşmasına olanak sağlamıştır. Örneğin üye modülünde bulunan cinsiyet, bölüm, unvan, grup, doğum tarihi gibi kullanıcılara ait tanımlayıcı nitelikler standart dışı kayıt barındırmaktadır. Erkek-kadın dışında cinsiyet verisi, bölüm-fakülte verilerinin birbirine karıştırılması veya bunların tamamen boş bırakılması veri setinde parazit oluşumuna sebebiyet vermektedir. Öte yandan üye kodu (id) alanında numerik girişlerin yanı sıra harflere de izin veriliyor olması bu alanda “123x” formunda alfanümerik kullanıcı numarası barındırılmasına sebebiyet vermektedir. Otomasyon sistemi kod yapısında özellikle salt numerik veya metin girişi yapılacak alanlarda girdi maskesi veya geçerlilik kurallarının kullanılması bu tür hatalı verilerin ortaya çıkmasını engelleyecektir.

Öncelikle kütüphane otomasyon sistemi üzerinden yapılan gözlem sonucu analizlere dahil edilmesi muhtemel öznitelikler tespit edilmiş ve Filemaker işlemlerinden “=” ile boş veriler belirlenmiştir. Cinsiyet ve üye grubu boş kayıtlar manuel olarak otomasyon üzerinden tamamlanmıştır. Üye grubu alanına 01 akademik personel, 02 doktora öğrencisi, 03 idari personel, 04 öğrenci, 08 sözleşmeli personel ve 09 yüksek lisans öğrencisi ön tanımları yapılmış; fakat bahsi geçen manuel müdahale yetkisinin kısıtlanmasından önce bu alanda belirtilen kodların dışında pek çok kod ile giriş yapıldığı tespit edilmiştir. Bu kodlar MS Excel uygulamasında filtre özelliği ile süzölmüş ve ilgili grubun kodu ile topluca değiştirilmiştir. Kullanıcıların doğum tarihleri yıl bazında alınmış olup, 1997 yerine 1197 şeklinde hatalı girişlerin olduğu da göze çarpmıştır. Mantıksal olarak elle düzeltilmesi mümkün kayıtlar düzeltilmiş; ancak doğru yılı tahmin

edilemeyenler kayıp veri olarak bırakılmıştır. Kullanıcıların yıl bazında doğum tarihinden ödünç tarihinin farkını hesaplayarak ödünç işleminin yapıldığı andaki yaşı tespit edilmiş ve analizlere dahil edilmiştir. Öte yandan veri giriş alanlarının birbirine girdiği hatalı kayıtlar da mevcuttur. Otomasyon sisteminden veriler xls dosya formatında dışa aktarıldığında tablonun belli satırlarında tüm veri satırının boş kaldığı kayıtlar da göze çarpmıştır (Şekil 10). Yine bu boş satırlar MS Excel üzerinden filtrelenerek veri setinden çıkarılmıştır.

1	ÖdemeVuru	mayoBr	KD	KDT	KDZ	KG	KGT	kGün	KGZ	KKN	kÖdeme	kÖdemeTarih	Tutatarıny	Demirbasi	kunyeEserAdıYazarlar	inyeSınıflama
266	0		gülenyusuf	07.06.2022	10.57.12	banko1	22.09.2008	-4	15.28.45	609892F-40AC-46D7-9E32-3EF7C9AF2AE4	0013422	Benim hala umudum var / İbrahim Gürpınar / PL 248/G77				
267	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	15.50.48	F9A30F1A-2E99-4FAA-A921-73DA5A05CFE1	0013247	Yalnız / Peyami Safa / MAKÜ Prof.Dr. İlhan PL 248/S24				
268	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-3	15.51.36	9A3288C7-1992-4A1E-99F6-F66CD2670379	0003486	Ana / Maksim Gorki, çeviren İsmail Müstak PG 3325/G6519				
269	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-3	16.12.49	A8788017-0958-48C8-8963-7405E499095E	0013212	Dişi kurdun rüvaları / Cengiz Aytmatov, çeviri PG 3478/A9819				
270	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-7	16.18.31	96C02403-1148-4F73-8F19-8060F0F01173	0009151	Damga / Reşat Nuri Güntekin / MAKÜ Prof.Dr. PL 248/G68				
271	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	16.20.08	478D5CE7-CD35-4697-AACD-EC057612536A	0003412	Suç ve ceza / Fyodor Mihaylovic Dostoyevski PG 3325/D6719				
272	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	16.32.06	A31C1A19-61F4-4740-98C9-3A4337E37CA9	0013509	İnce Memed 1 / Yaşar Kemal / MAKÜ Prof.Dr. PL 248/G33				
273	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	0	16.46.25	0C2131C9-236C-49D4-89F1-4EE5402D7E6D	0006197	İyi düşün doğru karar ver : Etikali yaşam üzeri BF 637/C3				
274	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-5	16.53.30	B6C11102-91D0-4B6F-A231-A2CE61115084	0011540	Serdalye sokağı / John Steinbeck / MAKÜ Pr. P5 3537/S7419				
275	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-6	16.57.30	F5238535-03FF-435A-A304-70DE7699EB70	0013213	Beyaz gemi / Cengiz Aytmatov, çeviren Refik PG 3478/A9819				
276	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	17.47.16	F4ABAD80-13C2-409A-9E1E-04F147CC3A7B	0013284	Mai ve siyah / Halid Ziya Uşaklıgil, yayına PL 248/U83				
277	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-3	18.04.48	BA4D07D7-2028-48C5-AF4D-A029992A7084	0012641	Geniş zamanlar / Ayşe Kulın / MAKÜ Prof.Dr. PL 248/K85				
278	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-6	18.05.35	EEF9C721-83CB-4513-A189-5A397E890108	0012735	Halla güzel / Hablaade Nazim, naazlayan Ke PL 248/K39				
279	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-3	18.28.13	9E8D8238-8086-4321-937C-C866244E2A9	0013301	Metres / Hüseyin Rahmi Gürpınar / MAKÜ P. PL 248/G77				
280	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-3	18.47.28	912B656A-1C4F-4386-9C42-4BCF64F8F5F2	0012930	Gurbet kuzları / Orhan Kemal / MAKÜ Prof. Dr. PL 248/G74				
281	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-7	09.27.26	FD9F5983-59C1-4E53-A392-770F89D703F8	0014081	Safahat / Mehmed Akif Ersoy, hazırlayan M. PL 210/E77				
282	0		ortak	28.09.2020	12.25.24	banko1	22.09.2008	-1	09.39.23	8F62B206-4E87-4124-A331-8D7EA3DFE80B	0000832	İlkokuma yazma öğretimi / Necati Cemaloğlu LB 1525.24/C32				
283	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-1	09.39.34	048AD754-8006-450C-8766-9E20776E48C1	0000846	İlkokuma yazma programı ve öğretimi / Süli LB 1525.24/C45				
284	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-1	09.39.45	219F88A3-7562-4993-99CD-A55D5E2EE88E	0013183	Ses temelli cümle yöntemi ve zihinsel yapı LB 1525/G64				
285	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-12	09.39.57	1C10849F-C504-44D8-89F7-56061D778737	0000916	Kuramdan uygulamaya eğitimde program g LB 1570/D46				
286	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-1	09.40.08	063C3861-D47C-467E-9715-8F0B738AC6C1	0011032	Uygulamalı ilkokuma yazma öğretimi / M. F LB 1525.24/D9				
287	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-5	09.48.25	9988EAE2-1857-4392-4659-FF87478EE85E	0005694	Vahşetin çağrısı / Jack London, çeviren Mus PS 3537/L6619				
288	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-6	10.46.10	9025403C-FE8B-4EAD-8013-1B69C414A551	0009082	Sarıkamış : Beyaz hüznü / İsmail Bilgin / MA PL 248/B55				
289	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-6	10.47.04	BCFF8523-2F79-4819-AB88-416F8D91671C	0013074	Yorgun savaşçı / Kemal Tahir / MAKÜ Prof.Dr. PL 248/K46				
290	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	6	11.08.21	7C7F841A-5689-45CF-8764-C580BADE3A6	0009087	Semerikant / Amin Maalouf, çeviren Esin Tal PG 3979.2/M33				
291	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-1	11.11.01	788A0ED2-D089-4EE4-8ED3-8F56C0C79F90	0003397	Karamazov kardeşler / Fyodor Mihaylovic, C PG 3325/D6719				
292	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	11.30.22	6A564809-A160-4614-8949-89947F33F272	0010552	Genel fizik laboratuvar deney kitabı -II / e.d.c PG 3326/G77				
293	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-2	11.30.49	AA4C308E-1566-4F3A-8055-4C207DC9C307	0012456	Genel fizik (I-II) / Kamili Temizöz / MAKÜ OC 23746				
294	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	11.32.27	10023643-4C8E-44E0-850C-ABC668482E35	0009448	Saddam'ın gelini / Sevim Torun / MAKÜ Prof.Dr. PL 230/T67				
295	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	11.33.28	6D78A393-8946-44C5-B241-9D469DF837C8	0012265	Bir ses böler geceyi / Ahmet Ümit / MAKÜ Pr. PL 248/U48				
296	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	11.37.47	2870CDD0-0F54-4A48-B904-DFADC1422FAC	0009019	Ramses : İşğin oğlu / Christian Jacq, çevirei PG 2670/J3319				
297	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	12.16.17	7E1F5E02-9C90-48F7-8E6D-B1846A00AF25	0013028	Sönmez yıldızlar / Reşat Nuri Güntekin / MA PL 248/G68				
298	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	12.45.10	52FF28D3-F858-4A6A-9062-D78998143467	0007526	Şair / Michael Connelly, çeviren Özlem Ezer PS 3537/C6619				
299	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	13.09.31	8983F82A-F04D-4510-B195-1CC88484C50A	0004953	Bir koltuk nasıl devrilir / Aziz Nesin / MAKÜ PL 248/N47				
300	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-5	14.11.57	15395179-9AC5-45D4-AA66-480819F13884	0010169	Aşk köpekleri / Ahmet Ümit / MAKÜ Prof.Dr. PL 248/J48				
301	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-7	14.15.37	15395179-9AC5-45D4-AA66-480819F13884	0004379	Ya tutarsa / Tamer Levett / MAKÜ Prof.Dr. PL 220/L46				
302	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-6	14.20.33	859F3DE2-22C5-4678-B1E7-FF55828871C5	0012909	İstanbulular / Buket Uzuner / MAKÜ Prof.Dr. PL 248/U98				
303	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-6	14.21.11	CAE95A52-5AE2-4F14-8F64-807C648E9C35	0012917	Kumral ada mavi tuna / Buket Uzuner / MA PL 248/U98				
304	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-6	14.21.57	238C8E11-FEF8-418A-AE4D-419954827CA0	0012913	Gümüş yaz gümüş kız / Buket Uzuner / MAKÜ Prof.Dr. PL 248/U98				
305	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	14.40.33	922905D4-5289-4683-8BD0-C878B40C47AE	0013272	Handan / Halide Edib Adıvar / MAKÜ Prof.Dr. PL 248/A35				
306	0		yordambt	25.01.2018	11.06.23	banko1	22.09.2008	-4	14.40.33	922905D4-5289-4683-8BD0-C878B40C47AE	0013272	Handan / Halide Edib Adıvar / MAKÜ Prof.Dr. PL 248/A35				

Şekil 10: Ödünç Veri Seti

Katalog veri seti, üye ve dolaşım veri setine nispeten daha az hatanın bulunduğu bir tablodur. Katalog modülü hem kütüphane bünyesinde kullanılan hem de diğer kütüphanelerin entegre olduğu toplu-kataloglar sebebiyle daha standartlaşmış ve gözetimli bir yapı teşkil etmektedir. Veri setleri üzerinden analizlere katkı sağlamayacağı düşünülen öznitelikler veri indirgeme işlemi gereği veri setinden çıkarılmıştır. Ancak analizlerin yapılmasına donanımsal olarak (işlemci, bellek, depolama) mâni bir durum olmaması sebebiyle veri miktarını azaltacak örneklem seçimi veya nesne azaltımı yoluna gidilmemiştir.

Veri ön işleme süreci sonucunda elde edilen nihai veri setine ilişkin bilgiler Tablo 5'te verilmiştir. Buna göre, dolaşım veri seti araştırmada yer alan analizlerin temelini oluşturmaktadır. Kullanıcıların kitap okuma davranışlarının kestirimi için gerekli kitap dolaşım hareketleri bu tabloda tutulmaktadır. Dolaşım veri setinde hangi kullanıcının

hangi gün hangi demirbaş, eser adı ve konu başlığına sahip kitabı ödünç aldığı, hangi tarihte iade ettiği verilerine ulaşmak mümkündür. Kullanıcı veya kitapla ilgili detay verilere ulaşmak içinse üye ve katalog veri setine müracaat edilmiş, analizin içeriğine göre bu tablolar dolaşım tablosuyla birleştirilmiştir.

Tablo 5: Nihai Veri Seti

Veri Seti	Nitelik	Kayıt	İçerik
Katalog	16	84849	Koleksiyonda bulunan basılı kitaplara ilişkin bibliyografik veriler
Üye	8	25093	Dolaşım işlemi bulunan kütüphane üyelerine ait tanımlayıcı bilgileri
Dolaşım	16	221592	Kitap ödünç-iade işlemlerine ait bilgiler

Katalog veri setinde demirbaş, eser adı, yazar adı, yayın tarihi, yayınlayan, ISBN, konu başlığı, yer numarası, alt tür, ödünç sayısı, dil ve yayın yeri bulunmaktadır. Dolaşım tablosunda üye id, ödünç ve iade tarihi, ödünç verilen gün sayısı, fakülte, bölüm, grup, cinsiyet, doğum tarihi (yıl), demirbaş, eser adı, konu başlığı, yazar adı, yayın tarihi ve yer numarası bulunmaktadır. Üye tablosunda üye id, grup, unvan, fakülte, bölüm, cinsiyet, doğum tarihi (yıl) bulunmaktadır.

3.8. Bulgular ve Değerlendirme

Araştırmanın bu bölümünde veri setini daha yakından tanımak adına tanımlayıcı istatistikler verilmiş, kütüphane kullanım eğilim ve trendleri irdelenmiş, araştırma sorularına ilişkin kümeleme ve işbirlikçi filtreleme ile analizler yapılmış ve kişiselleştirilmiş kitap tavsiye sistemi önerilmiştir.

3.8.1. Tanımlayıcı İstatistikler

Dolaşım veri seti üzerinden yapılan incelemeye göre kütüphane üzerinden kitap ödünç iade işlemi yapmış olan kullanıcıların % 65,07'sı kadın, %34,93'ü ise erkeklerden oluşmaktadır (Tablo 6). Kütüphane üye tablosunda 75233 kullanıcı kaydı bulunuyor olsa da herhangi bir işlem yapmamış kullanıcıları bu tabloda belirtme gereği duyulmamıştır. Öte yandan kütüphane kullanıcılarının çoğunluğunun kadınlar tarafından oluşuyor olması dikkat çekmektedir.

Tablo 6: Cinsiyet Dağılımı

Cinsiyet	<i>f</i>	%
Kadın	16327	65,07
Erkek	8766	34,93
Toplam	25093	100

Kütüphane üyelerinden ödünç işlemi yapan aktif kullanıcıların yaş dağılımı, doğum tarihi ile ödünç işleminin yapıldığı tarih baz alınarak hesaplanmıştır. Zira kullanıcıların bugün itibarıyla yaşına bakılarak kitap okuma tercihlerinin yanıltıcı bilgi oluşturulacağı kanısına varılmıştır. Ödünç işleminin yapıldığında kullanıcı kaç yaşında ise bu yaş aralığı ile kitap tercihi arasında bir ilişkinin olabileceği varsayılmıştır. Otomasyon sisteminde tüm kullanıcıların doğum tarihi verisi bulunmamaktadır. Bu sebeple yaş bilgisine ulaşılabilen kayıtlar tabloya aktarılmıştır. Buna göre kütüphaneden ödünç kitap alanların %85,6'sı 16-25 yaş aralığındadır (Tablo 7).

Tablo 7: Yaş Dağılımı

Yaş	<i>f</i>	%
16-25	158208	85,6
26-35	19482	10,54
36-45	5657	3,06
46-55	1286	0,7
56 ve üzeri	179	0,1
Toplam	184812	100,0

Bu yaş aralığına giren kullanıcıların dağılımına bakıldığında da tahmin edileceği üzere %99 gibi büyük oranda öğrencilerden (ön lisans, lisans, yüksek lisans) meydana geldiği anlaşılmaktadır. Bu aksi veya anlaşılabilir bir durum değildir çünkü üniversite kütüphanelerinin büyük oranda genç üniversite öğrencileri tarafından kullanılması beklenen bir durumdur. Kitap ödünç alan kullanıcıların %10,54'ü 26-35 yaş aralığında ve bu yaş aralığında olanların %29'u yüksek lisans, %24'ü ön lisans ve lisans, %20'si de akademisyen, %13'ü de doktora öğrencilerinden oluşmaktadır. Ödünç alan kullanıcıların %3,06'sı 36-45 yaş aralığında bulunmakta ve bu aralığına giren kullanıcıların %47'si

akademisyen, %16'sı idari personel ve %10'u da yüksek lisans öğrencilerinden oluşmaktadır. Ödünç işlemi yapan diğer kullanıcılardan 46-55 yaş aralığındakiler toplam yaş aralığının %0,7'sini ve 56 yaş ve üzerindeki ise %0,1'ini oluşturmaktadır.

Tablo 8: Kullanıcı Grubu Dağılımı

Grup	<i>f</i>	%
Öğrenci (Lisans-Ön Lisans)	22431	89,39
Yüksek Lisans Öğr.	1463	5,83
Akademisyen	701	2,79
İdari Personel	271	1,08
Sözleşmeli Personel	116	0,46
Doktora Öğr.	111	0,44
Toplam	25093	100,0

Kütüphaneden en az bir kez ödünç kitap alan kullanıcıların grup dağılımına bakıldığında (Tablo 8) %89,39'u ön lisans ve lisans öğrencilerinin oluşturduğu “öğrenci” grubuna mensuptur. %5,83'ü yüksek lisans öğrencilerinden, %2,79'u akademisyenlerden, %1,08'i idari personellerden, %0,46'sı sözleşmeli personel ve %0,44'ü de doktora öğrencilerinden oluşmaktadır. Üniversitenin 2024 yılı güncel öğrenci ve personel sayılarına bakıldığında (Burdur Mehmet Akif Ersoy Üniversitesi, 2024) fakülte, yüksekokul ve meslek yüksekokullarda toplam 32530, yüksek lisans programlarında 2620, doktora programlarında 370 öğrenci bulunurken 1120 akademisyen ve 408'i kadrolu 485'i sözleşmeli olmak üzere 893 idari personel bulunmaktadır. Üniversitenin 37 bin kadar mevcut kullanıcı sayısı ile kütüphaneden 2024 yılına kadar en azından bir kez kitap ödünç alan kullanıcı sayıları karşılaştırıldığında kütüphanenin potansiyel kullanıcılarının küçük bir kesimine hizmet verdiği anlaşılmaktadır. Analizlerin başlangıç yılını oluşturan 2008 yılından bu yana görev yapan tüm akademisyen sayısı bir yana mevcut akademisyen sayısı ile dahi mukayese edildiğinde kütüphaneden basılı kitap alan kullanıcı sayısı yarısını temsil etmektedir.

Kullanıcıların akademik birim dağılımına bakıldığında (Tablo 9) en fazla ödünç kitap alan kullanıcılar eğitim fakültesine mensup olduğu anlaşılmaktadır (%40,68). Eğitim fakültesi, Burdur Mehmet Akif Ersoy Üniversitesi'nin en eski fakültelerinden biri

olmasının da verdiği avantaj ile en yakın takipçisi iktisadi ve idari bilimler fakültesinin üç katı kadar farkla ilk sırada yer almaktadır.

Tablo 9: Akademik Birim Dağılımı

Akademik Birim	<i>f</i>	%	Akademik Birim	<i>f</i>	%
Eğitim Fak.	9382	40,68	Burdur MYO	34	0,15
İİBF	2918	12,65	Diş Hekimliği Fak.	28	0,12
Fen-Edebiyat Fak.	2613	11,33	Bucak EGTBMYO	21	0,09
Müh.-Mim. Fak.	1368	5,93	Yeşilova İA MYO	19	0,08
Sağlık Bil. Fak.	1317	5,71	Göhlisar MYO	16	0,07
Sağlık Yüksekokulu	1007	4,37	Bucak Sağlık YO	14	0,06
Veteriner Fak.	876	3,8	Tefenni MYO	13	0,06
Meslek Yüksekokulu	570	2,47	Çavdır Meslek YO	13	0,06
Sağlık Hiz. MYO	546	2,37	Ağlasun MYO	12	0,05
İlahiyat Fak.	541	2,35	Bucak İşletme Fak.	12	0,05
Turizm İş.ve Ot. YO	454	1,97	Bucak Teknoloji Fak.	10	0,04
Sosy. Bil. MYO	296	1,28	Eğitim Bilimleri Ens.	9	0,04
BESYO	152	0,66	Sosyal Bilimler Ens.	8	0,03
Gıda Tar. ve Hay. MYO	148	0,64	Yabancı Diller YO	7	0,03
Tek. Bil. MYO	144	0,62	Fen Bilimleri Ens.	4	0,02
Bucak ZTYO	132	0,57	Göhlisar UB YO	4	0,02
Spor Bilimleri Fak.	119	0,52	Sağlık Bilimleri Ens.	4	0,02
Sanat ve Tasarım Fak.	116	0,5	Göhlisar SH MYO	3	0,01
Bucak HKMYO	69	0,3	Altınyayla MT MYO	1	0
Türk Müz. Dev. Kons.	60	0,26	Sağlık Hizmetleri MYO	1	0
Genel Toplam		23061		100	

İktisadi ve idari bilimler fakültesi %12,65, fen-edebiyat fakültesi %11,33, mühendislik mimarlık fakültesi %5,71, sağlık bilimleri fakültesi %4,37, veteriner fakültesi mensupları ise %3,8 oranında sıralanmaktadır. Akademik birim dağılımı, üye tablosunda akademik birim verisi ulaşılabilen kullanıcılar üzerinden oluşturulmuştur. Buna göre eğitim fakültesi mensuplarının kütüphane kullanımı noktasında belli bir üstünlüğünün olduğu anlaşılmaktadır. Diğer taraftan özellikle şehir merkezine ve dolayısıyla üniversite kampüsüne fiziksel uzaklığı bulunan ilçe yüksekokul ve meslek yüksekokullarının kütüphanenin basılı kitaplarını kullanımı noktasında ciddi bir eksiklik

olduğu anlaşılmaktadır. Bu durumun farkında olan kütüphane yönetimi, özellikle elektronik Türkçe kitap temini ile kullanıcıların ihtiyaçlarını gidermeye çalışmaktadır.

Tablo 10: Yıl Bazında En Çok Okunan Kitaplar

Yıl	Eser Adı (1)	f	Eser Adı (2)	f
2009	Aşk-ı Memnu	38	Sevdalinka	37
2010	İlköğretim programı (1-5. Sınıflar)	136	Öğretim ilke ve yöntemleri	92
2011	İlköğretim programı (1-5. Sınıflar)	145	Öğretim ilke ve yöntemleri	107
2012	İlköğretim programı (1-5. Sınıflar)	159	Öğretim teknolojileri ve materyal tasarımı	91
2013	Öğretim teknolojileri ve materyal tasarımı	52	Bilimsel araştırma yöntemleri	48
2014	Eğitim bilimine giriş	78	Öğretim ilke ve yöntemleri	72
2015	Öğretim ilke ve yöntemleri	70	Bilimsel araştırma yöntemleri	52
2016	Eğitim bilimine giriş	89	Bilimsel araştırma yöntemleri	52
2017	Bilimsel araştırma yöntemleri	67	Öğretim ilke ve yöntemleri	67
2018	Eğitim psikolojisi	71	Bilimsel araştırma yöntemleri	59
2019	Öğretim ilke ve yöntemleri	106	Eğitim psikolojisi	91
2020	Bilimsel araştırma yöntemleri	23	Çalığışu	20
2021	Olağanüstü bir gece	28	Şeker portakalı	26
2022	Eğitim sosyolojisi	52	Eğitim psikolojisi	39
2023	Anna Karenina	34	Eğitim sosyolojisi	19

Yıllık analizlerde 2008 ve 2024 yılları tam bir yıllık veri bulunmaması sebebiyle analizlerde kullanılmamıştır. 2009 yılından itibaren kronolojik olarak yıllık bazda en fazla okunan iki kitaba bakıldığında (Tablo 10) dönem dönem edebi eserler öne çıkmakla birlikte ekseriyetle eğitim bilimlerine yönelik akademik kitaplar ön plana çıkmaktadır (Tablo 10). Kümülatif orana bakıldığında da 2008'den 2023 yılına kadar en çok okunan iki kitap “Öğretim ilke ve yöntemleri” ve “Bilimsel araştırma yöntemleri” kitapları ilk iki sırayı almaktadır.

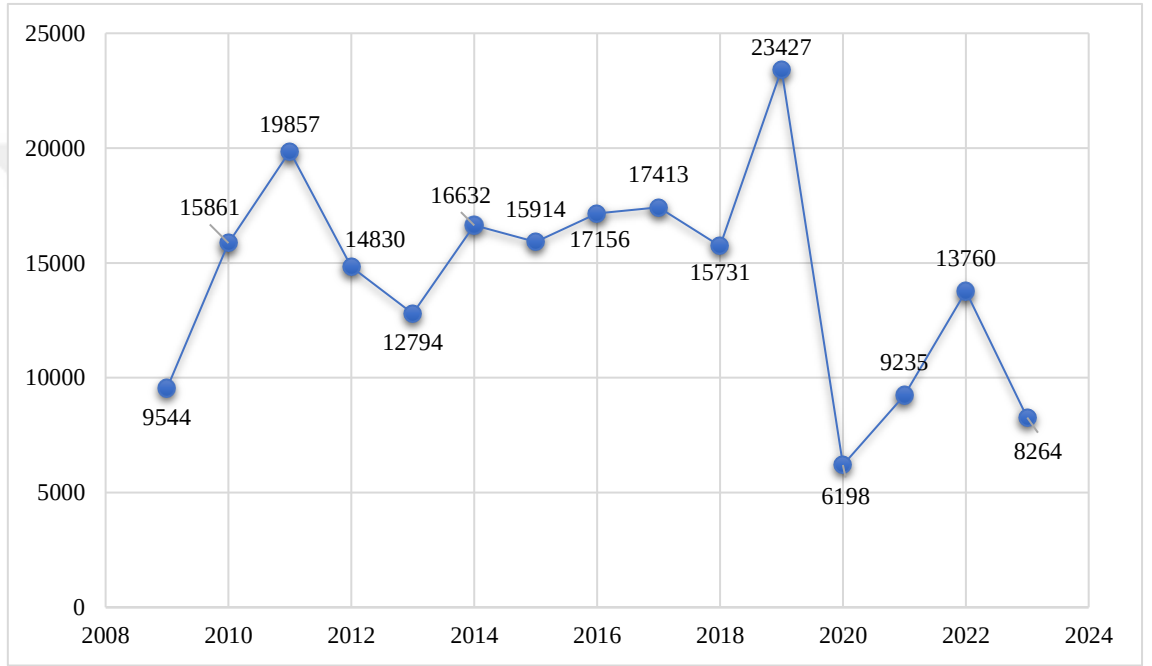
Ödünç verilen kitapların konu dağılımına bakıldığında (Tablo 11) dil ve edebiyat %41,06 ile açık ara en fazla ödünç alınan kitap türü olurken bilim konulu kitaplar %11,69, eğitim %10,74, sosyal bilimler %8,94, felsefe-psikoloji-din %7,44, tarih %5,91, coğrafya %2,71, tıp %2,15, teknoloji-mühendislik %2,13 ödünç alınma oranına sahiptir. Burdur Mehmet Akif Ersoy Üniversitesi'nin sosyal bilimler ağırlıklı akademik yapılanması göz

önüne alındığında ortaya çıkan tablo yadırganmayacaktır. Ancak bilim, tıp ve teknoloji-mühendislik konulu kitapların da azımsanmayacak kullanım oranı göze çarpmaktadır.

Tablo 11: Konu Dağılımı

Konu Başlıkları	<i>f</i>	%
Dil ve Edebiyat	90981	41,06
Bilim	25898	11,69
Eğitim	23789	10,74
Sosyal Bilimler	19803	8,94
Felsefe-Psikoloji-Din	16484	7,44
Tarih	13091	5,91
Coğrafya	6016	2,71
Tıp	4775	2,15
Teknoloji ve Mühendislik	4722	2,13
Güzel Sanatlar	4442	2
Siyaset Bilimi	3863	1,74
Tarım	3389	1,53
Hukuk	1625	0,73
Müzik	1431	0,65
Tarihe Yardımcı Bilimler	781	0,35
Bibliyografya ve Kütüphanecilik	178	0,08
Genel Konular	111	0,05
Tarih: Amerika	105	0,05
Askerlik	82	0,04
Denizcilik	22	0,01
Tıp ve Sağlık Bilimleri	4	0,00
Toplam	221592	100

Yıllık ödünç sayılarının dağılımına bakıldığında (Şekil 11) üniversitenin kuruluşunun ardından öğrenci ve personel sayısındaki artış kütüphanenin ödünç işlemlerinde de göze çarpmaktadır. 2009 yılında 9544 ödünç kitap verilmişken, 2010 yılında 15861, 2011 yılında grafiğin ikinci pik noktası olan 19857 ödünç kitap sayısına ulaşmış, 2012 yılında düşüş trendi ile 14830 ve 2013'te 12794 kitap ile bu dönemdeki en düşük seviyeye ulaşmıştır. 2014 yılında 16632 kitap ödünç verilirken 2015 yılında 15914, 2016 yılında 17156, 2017 yılında 17413, 2018 yılında 15731 kitap ödünç verilmiştir.



Şekil 11: Yıllık Ödünç Sayısı Dağılımı

2011-2018 yılları arasında dalgalı bir iniş çıkış gösteren ödünç sayıları 2018 yılı sonunda kütüphanenin mevcut yeni binasına taşınmasının ardından 2019 yılında 23427 kitap ile pik noktaya ulaştığı görülmektedir. Ardından 2020 yılında Covid-19 pandemisi ile yüz yüze eğitime ara verilmesiyle ödünç sayıları da hızla düşmüş ve 2020 yılı için 6198, 2021 yılında 9235, 2022 yılında 13760 ve 2023 yılında 8264 kitap ödünç verilmiştir.

Kullanıcıların kitap seçimi tercihleri üzerindeki demografik özelliklerin etkisi irdelendiğinde erkek ve kadınların en fazla tercih ettiği 10 konu başlığı Tablo 12'de verilmiştir. Buna göre erkekler %16,97 oranında en fazla Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya, dilleri konu başlığından kitaplar tercih ederken %7,56 oranında

eğitim teorisi ve pratiği, %5,33 oranında matematik, %5,25 oranında Balkan yarımadası tarihi ve %4,5 oranında da psikoloji, %3,88 oranında Amerikan edebiyatı, %2,86 İngiliz edebiyatı, %2,57 hayvan kültürü, %2,53 Fransız, İtalyan, İspanyol, Portekiz edebiyatı, %2,23 ise ticaret konulu kitaplar tercih etmiştir.

Tablo 12: Cinsiyete Göre Kitap Tercihleri

Cinsiyet	Konu Başlığı	<i>f</i>	%
Erkek	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri	12476	16,97
	Eğitim teorisi ve pratiği	5555	7,56
	Matematik	3916	5,33
	Balkan yarımadası tarihi	3860	5,25
	Psikoloji	3307	4,5
	Amerikan edebiyatı	2856	3,88
	İngiliz edebiyatı	2106	2,86
	Hayvan kültürü	1887	2,57
	Fransız, İtalyan, İspanyol, Portekiz edebiyatı	1862	2,53
	Ticaret	1639	2,23
Kadın	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri	36468	24,63
	Eğitim teorisi ve pratiği	14438	9,75
	Amerikan edebiyatı	12580	8,5
	İngiliz edebiyatı	7577	5,12
	Psikoloji	7530	5,09
	Matematik	6558	4,43
	Fransız, İtalyan, İspanyol, Portekiz edebiyatı	4751	3,21
	Balkan yarımadası tarihi	4238	2,86
	Endüstriler	2224	1,5
	Bilim	1998	1,35

Kadınlar ise %24,63 oranında Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya, dilleri konusunda, %9,75 eğitim teorisi ve pratiği, %8,5 Amerikan edebiyatı, %5,12 İngiliz edebiyatı, %5,09 psikoloji, %4,43 matematik, %3,21 Fransız, İtalyan, İspanyol, Portekiz edebiyatı, %2,86 Balkan yarımadası tarihi, %1,5 endüstriler, %1,35 oranında da bilim konulu kitaplar tercih etmiştir.

Tablo 13: Yaş Gruplarına Göre Kitap Tercihleri

Yaş Aralığı	Konu Başlığı	<i>f</i>	%
16-25	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri	38917	44,11%
	Eğitim teorisi ve pratiği	14203	16,10%
	Amerikan edebiyatı	12495	14,16%
	Matematik	7575	8,59%
	Psikoloji	7557	8,57%
26-35	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri	2614	33,58%
	Eğitim teorisi ve pratiği	2446	31,42%
	Ticaret	1004	12,90%
	Endüstriler	896	11,51%
	Psikoloji	824	10,59%
36-45	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri	837	37,94%
	Eğitim teorisi ve pratiği	440	19,95%
	Matematik	379	17,18%
	Amerikan edebiyatı	324	14,69%
	Balkan yarımadası tarihi	226	10,24%
46-55	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri	327	52,91%
	Psikoloji	101	16,34%
	Endüstriler	79	12,78%
	Amerikan edebiyatı	60	9,71%
	Eğitim teorisi ve pratiği	51	8,25%
56+	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri	29	38,16%
	Türkiye hukuku	14	18,42%
	Fransız, İtalyan, İspanyol, Portekiz edebiyatı	12	15,79%
	Amerikan edebiyatı	11	14,47%
	Siyasi teori	10	13,16%

Kütüphaneden ödünç kitap alan kullanıcıların, işlem yaptıkları tarihteki yaşlarına göre kitap tercihlerine bakıldığında (Tablo 13) tüm yaş grupları için en fazla tercih edilen kitap kategorisi Türkçe roman, hikâye ve şiirlerin de bulunduğu Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri konu başlığıdır. Bu konu başlığı 16-25 yaş aralığında %44,11, 26-35 yaş aralığında %33,58, 36-45 yaş aralığında %37,94, 46-55 yaş

aralığında %52,91 ve 56 yaş üstü kullanıcılarda da %38,16 oranında tercih edildiği görülmektedir. Yaş gruplarının detaylarına bakıldığında ise 16-25 yaş aralığındaki kullanıcılar Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri konu başlığındaki kitaplardan sonra en fazla %16,10 oranında eğitim teorisi ve pratiği, %14,16 oranında Amerikan edebiyatı, %8,59 oranında matematik ve %8,57 oranında da psikoloji konulu kitaplar tercih etmiştir. 26-35 yaş aralığındaki kullanıcılar Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri konu başlığından sonra %31,42 oranında eğitim teorisi ve pratiği, %12,90 ticaret, %11,51 endüstriler, %10,59 psikoloji konulu kitapları tercih etmiştir. 36-45 yaş aralığındaki kullanıcılar ikinci olarak en fazla %19,95 oranında eğitim teorisi ve pratiği, %17,18 matematik, %14,69 Amerikan edebiyatı ve %10,24 Balkan yarımadası tarihi konulu kitapları ödünç almıştır. 46-55 yaş aralığındaki kullanıcıların en fazla ödünç aldığı kitap kategorileri ise Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dillerinin ardından %16,34 oranında psikoloji, %12,78 oranında endüstriler, %9,71 oranında Amerikan edebiyatı, %8,25 oranında ise eğitim teorisi ve pratiği konulu kitaplar tercih edilmiştir. 56 yaş üstü kullanıcılar ise Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri konulu kitapları en fazla tercih edilmekle birlikte %18,42'lik oranla Türkiye hukuku ikinci sırada, %15,79 oranla Fransız, İtalyan, İspanyol, Portekiz edebiyatı üçüncü sırada, %14,47'lik oranla Amerikan edebiyatı dördüncü sırada, %13,16'lık oranla siyasi teori beşinci sırada en fazla tercih edilen kitap kategorisi olmuştur.

3.8.2. Kümeleme Analizi

Kütüphane kullanıcılarının eğilimlerini tespit etmek ve kullanıcı segmentasyonu ile benzer eğilimlere sahip kullanıcıların ihtiyaç duyabileceği muhtemel bilgi kaynaklarını belirlemek amacıyla kümeleme analizi uygulanmıştır. İkinci bölümde Veri madenciliği Algoritmaları başlığında detayları verildiği üzere kümeleme analizinde veri setinden birbirinden maksimum düzeyde ayrılan kümeler oluşturmak ve küme içindeki öğelerin maksimum düzeyde birbiriyle benzeşmesi amaç edinir. Böylelikle birbirine en çok benzeyen öğeler bir arada tutulurken kümelerin de birbirinden olabildiğince uzak olması sağlanır. Çalışmada kullanıcıların demografik özelliklerinin yanı sıra ödünç aldıkları kitap konuları bakımından birbirine en çok benzeyen kullanıcıların bir araya getirildiği kümeler oluşturulma gayesi güdülmüştür. Bu amaçla kümeleme algoritmaları

arasında en sık kullanılan *k*-means algoritması kullanılmış ve ortaya çıkan kümelerde en fazla ödünç alınan bilgi kaynakları ilgili kümedeki kullanıcılara tavsiye edilmiştir.

Süreç, Python dilinde yazılmış, veri analizi alanında sıklıkla kullanılan ve NumPy, SciPy, Matplotlib, pandas, IPython, SymPy ve Cython gibi pek çok kütüphaneye entegre olabilen Spyder IDE üzerinde yürütülmüştür. *xlsx* dosya formatında tutulan veri seti Spyder tarafından okunmuş ve Python dilinde pandas, scikit-learn kütüphaneleri çağrılarak analiz kodları çalıştırılmıştır. Kümeleme analizine ‘Bölüm’, ‘Cinsiyet’, ‘Grup’, ‘Fakülte’, ‘Konu’ ve ‘Yaş’ değişkenleri dahil edilmiştir. Veri setinde bulunan diğer değişkenler bilgi kaynaklarını nitelendirdikleri için kümeleme analizine dahil edilmemiştir. ‘Konu’ değişkeni her ne kadar bilgi kaynağını nitelendirse de kullanıcıların tercih ettiği konu başlıkları ile kullanıcılar arasında ilişki olacağı varsayımıyla kümeleme analizine dahil edilmiştir. Veri setine ön bakış yapıldıktan sonra eksik veri kontrolü yapılmış ve 221592 kaydın bulunduğu veri setinde 36780 kaydın yaş bilgisi olmadığı anlaşılmıştır. Analizlerde yaş verisi kullanılacağı için eksik verilerin tamamlanması gerekmektedir. SimpleImputer modülü ile veri setindeki yaş değerinin medyanı eksik verilere atanmıştır. Yaş değerlerini gruplandırmak için “cut” fonksiyonu ile 16-25, 26-35, 36-45, 46-55, 56+ olarak aralıklara bölünmüştür. Ardından LabelEncoder ile kategorik değişkenler (‘Fakülte’, ‘Bölüm’, ‘Konu’, ‘Yaş_Grubu’) sayısallaştırılarak *k*-means algoritmasına uygun hale getirilmiştir. ‘Cinsiyet’ değişkeni 1-erkek, 2-kadın; ‘Grup’ değişkeni 1-akademisyen, 2-doktora öğrencisi, 3-idari personel, 4-öğrenci (lisans-önlisans), 5-sözleşmeli personel, 6-yüksek lisans öğrencisi olacak şekilde veri ön işleme sürecinde sayısallaştırılmıştır. Tüm değişkenlerin sayısal verilere dönüştürülmesinden sonra yeni veri çerçevesi oluşturulmuştur. Oluşturulan yeni veri çerçevesinde ilk beş kaydın görünümü aşağıdaki şekildedir.

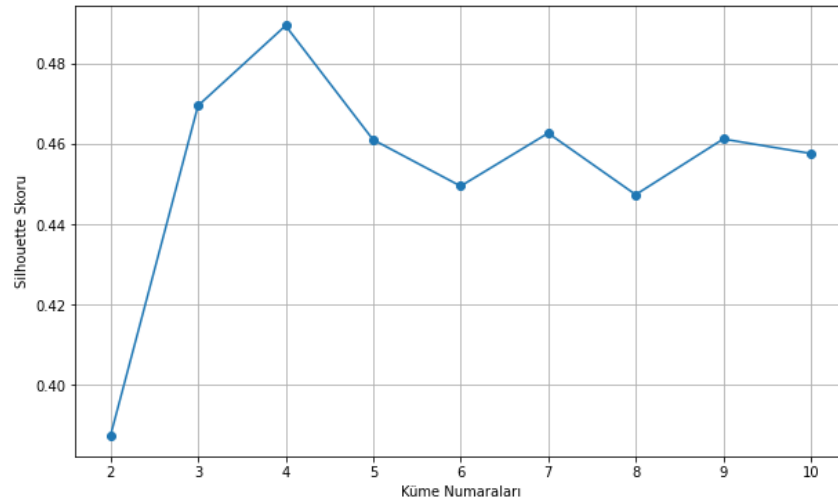
BÖLÜM	CINSİYET	GRUP	FAKÜLTE	KONU	YAŞ_GRUBU
174	2	4	14	157	1
174	2	4	14	157	1
134	1	4	14	178	2
174	2	4	14	157	2
174	2	4	14	178	2

Verilerin *k*-means algoritması için uygun formata dönüştürülmesinin ardından bölünecek küme sayısının belirlenmesi gerekmektedir. *k*-means algoritması için başlangıçta belirlenen *k* adet merkez (centorid), araştırmacı tarafından keyfi olarak belirlenebildiği gibi (Akpınar, 2014: 311) *k*-means++ (Shetkar, Fernandes, 2016), Dunn index (Dunn, 1974), Silhouette index (Rousseeuw, 1987), Davies-Bouldin (Davies ve Bouldin, 1979) ve Elbow (Merliana, Ernawati ve Santoso, 2015) gibi kurallara dayanarak da belirlenebilir. Araştırmada Silhouette index ve Elbow yöntemi ile *k*-means algoritmasında bölünecek optimum küme sayısı belirlenmiştir.

Silhouette index, küme içindeki ve kümeler arasındaki noktaların ortalama farklılığını hesaplamakta ve $-1 \leq S_i \leq 1$ arasında değer almaktadır (Mehar, Matawie ve Maeder, 2013: 52). Elde edilen değer +1'e ne kadar yakınsa kümeler arasındaki fark o kadar fazladır. Aşağıda verilen Silhouette index formülünde $a_{(i)}$, *i* numaralı nesnenin aynı kümedeki tüm nesnelere olan ortalama uzaklığını; $b_{(i)}$ ise *i* numaralı nesnenin farklı kümedeki tüm nesnelere olan uzaklığını ifade eder.

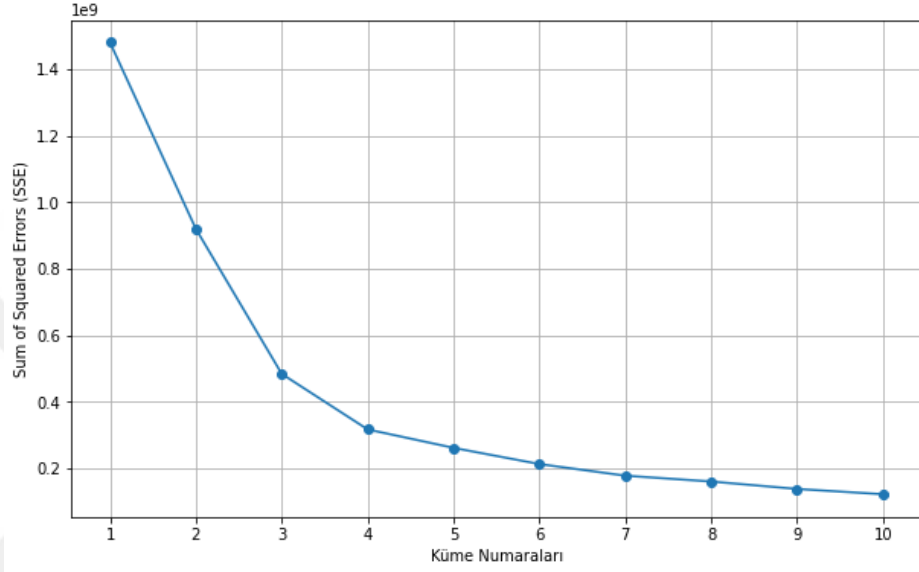
$$S_{index} = \frac{b_{(i)} - a_{(i)}}{\max \{a_{(i)}, b_{(i)}\}}$$

Spyder IDE'de sklearn.metrics ve matplotlib.pyplot kütüphaneleri çağrılarak yapılan analizde *k*=2'den 10'e kadar küme sayısı denenmiş ve tüm kümelerin Silhouette skoru incelenmiştir. Silhouette skoruna bakıldığında (Şekil 12) en yüksek değer 0,489 ile 4. kümede olduğu belirlenmiş ve *k*-means algoritması için optimum küme sayısı ortaya çıkmıştır.



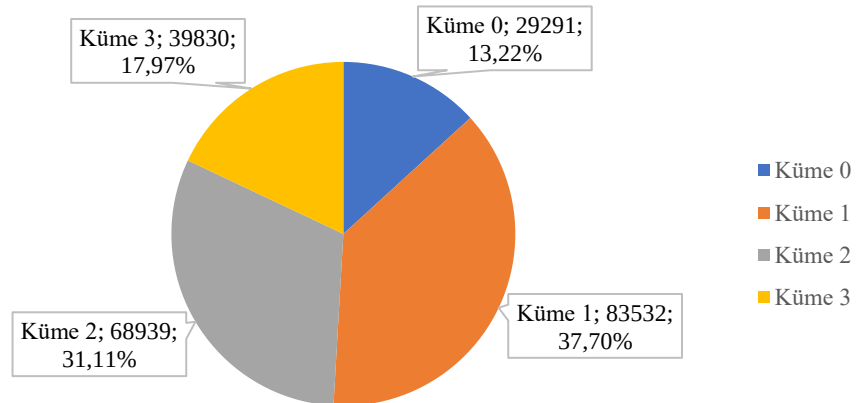
Şekil 12: Silhouette Grafiği

Elbow yöntemi, grafik tabanlı bir gösterge oluşturmakta ve veri seti için önce $k=1$ 'den başlayarak SSE (Sum Square Error) değeri hesaplanmakta ve her seferinde k değeri birer artırılarak SSE değeri hesaplanmaktadır. SSE değerindeki ciddi düşüşün yaşandığı ve artık bu noktadan sonra önemli değişikliğin olmadığı nokta optimal k değerini ifade eder (Marutho vd., 2018: 533). Bu k değeri k -means algoritmasında kullanılacak olan küme sayısını belirtmektedir.



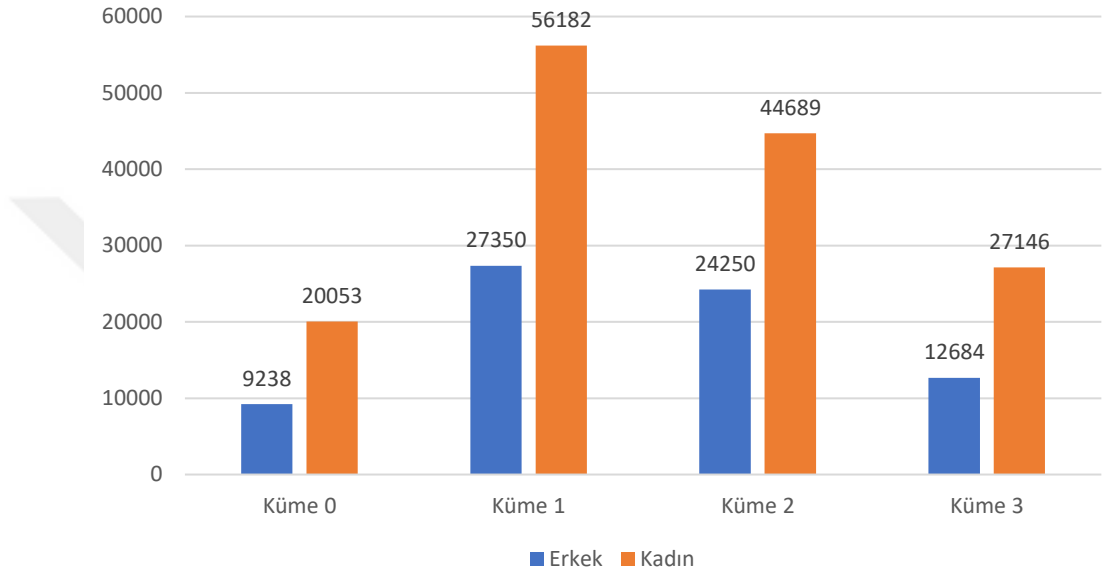
Şekil 13: Elbow Grafiği

Veri setine ilişkin Elbow grafiğine bakıldığında (Şekil 13) $k=3$ ve $k=4$ noktasında kırılma olduğu grafiğin bu iki nokta itibarıyla dirsek oluşturduğu anlaşılmaktadır. Silhouette skoru da dikkate alınarak Elbow grafiğinin işaret ettiği 4 küme sayısı k -means algoritmasında kullanılmıştır.



Şekil 14: k-means Küme Yoğunlukları

Spyder IDE programında sklearn.cluster modülü çağrılarak yapılan k-means kümeleme analizinde veri seti Silhouette skoru ve Elbow grafiği dikkate alınarak 4 kümeye ayrılmıştır (Şekil 14). 0 numaralı ilk kümede 29291 (%13,22), 1 numaralı kümede 83532 (37,7), 2 numaralı kümede 68939 (%31,11) ve 3 numaralı kümede 39830 (17,97) öge bulunmaktadır. En yoğun küme 83532 ile 1 numaralı küme olurken en seyrek küme 29291 öge ile 0 numaralı küme olmuştur.

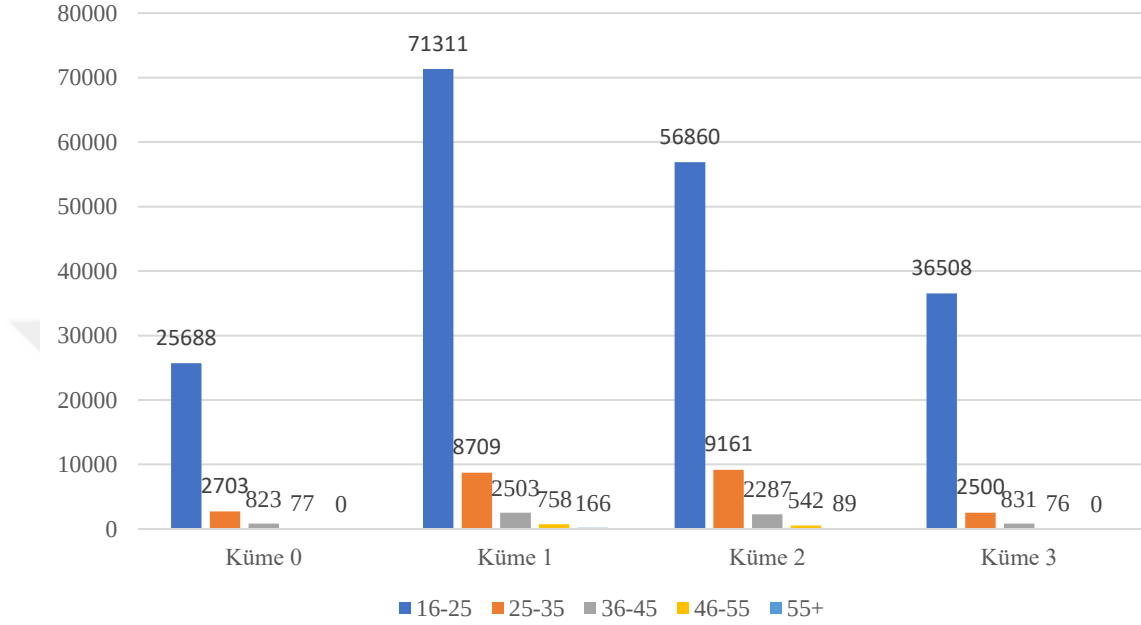


Şekil 15: Kümelerin Cinsiyet Dağılımı

Kümelerin cinsiyet bakımından yoğunluklarına bakıldığında (Şekil 15) veri setiyle paralel şekilde tüm kümelerin iç dağılımında kadınlar daha baskın taraftır. Küme 0’da erkekler 9238 (%31,54), kadınlar 20053 (%68,46); Küme 1’de erkekler 27350 (%32,74), kadınlar 56182 (%67,26); Küme 2’de erkekler 24250 (%35,18), kadınlar 44689 (%64,82); Küme 3’te ise erkekler 12684 (%31,85), kadınlar 27146 (%68,15) olarak dağılmıştır. Yüzdelik olarak cinsiyet dağılımları kümeler içerisinde diğer kümelerle benzerlik gösterse de erkeklerin en yoğun olduğu küme, Küme 2 (%35,18), kadınların en yoğun olduğu küme ise Küme 1 (%68,46) olarak ortaya çıkmaktadır.

Yaş aralığı dağılımına bakıldığında ise (Şekil 16) veri setinin en yoğun grubunu oluşturan genç kullanıcılar yani 16-25 yaş aralığındaki kullanıcılar çok büyük oranda tüm kümeleri domine etmektedir. Küme 0’da 16-25 yaş aralığındaki kullanıcılar 25688 (%87,7), 26-35 yaş aralığındakiler 2709 (%9,23), 36-45 yaş aralığı 823 (%2,81), 46-55

yaş aralığı 77 (%0,26) ve 56 yaş üstü 0 kullanıcı bulunmaktadır. Küme 1’de 16-25 yaş aralığı 71311 (%85,46), 26-35 yaş aralığı 8709 (%10,44), 36-45 yaş aralığı 2503 (%3), 46-55 yaş aralığı 758 (%0,91) ve 56 yaş üstü kullanıcılar 166 (%0,2) olarak dağılmaktadır.



Şekil 16: Kümelerin Yaş Aralığı Dağılımı

Küme 2’de 16-25 yaş aralığı 56860 (%82,48), 26-35 yaş aralığı 9161 (%13,29), 36-45 yaş aralığı 2287 (%3,32), 46-55 yaş aralığı 542 (%0,79) ve 56 yaş üstü kullanıcılar 89 (%0,13) şeklinde küme içine yayılmıştır. Küme 3’te ise 16-25 yaş aralığı 36508 (%91,46), 26-35 yaş aralığı 2500 (%6,26), 36-45 yaş aralığı 831 (%2,08), 46-55 yaş aralığı 76 (%0,19) ve 56 yaş üstü 0 kullanıcı küme içerisinde dağılmaktadır.

Kümelerin konu dağılımına bakıldığında (Tablo 14) Küme 0’da en fazla ödünç alınan kitapların konu başlığı 5445 kez ile Amerikan edebiyatına (%18,59) ait kitaplar olurken 4103 kez Eğitim teorisi ve pratiği konulu kitaplar (%14,01), 2038 kez ise Fransız, İtalyan, İspanyol, Portekiz edebiyatı (%6,96) konu başlığındaki kitaplar ödünç alınmıştır. Küme 1’de 36258 kez Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri ve edebiyatı (%43,45), 7741 kez psikoloji (%9,28), 6610 kez matematik (%7,92) konu başlığındaki kitaplar ödünç alınmıştır. Küme 2’de 15890 kez eğitim teorisi ve pratiği (%23,05), 9991 kez Amerikan edebiyatı (%14,49), 6101 kez Balkan yarımadası tarihi (%8,85) konu başlığındaki kitaplar ödünç alınmıştır. Küme 3’te ise 12686 kez Türk dili

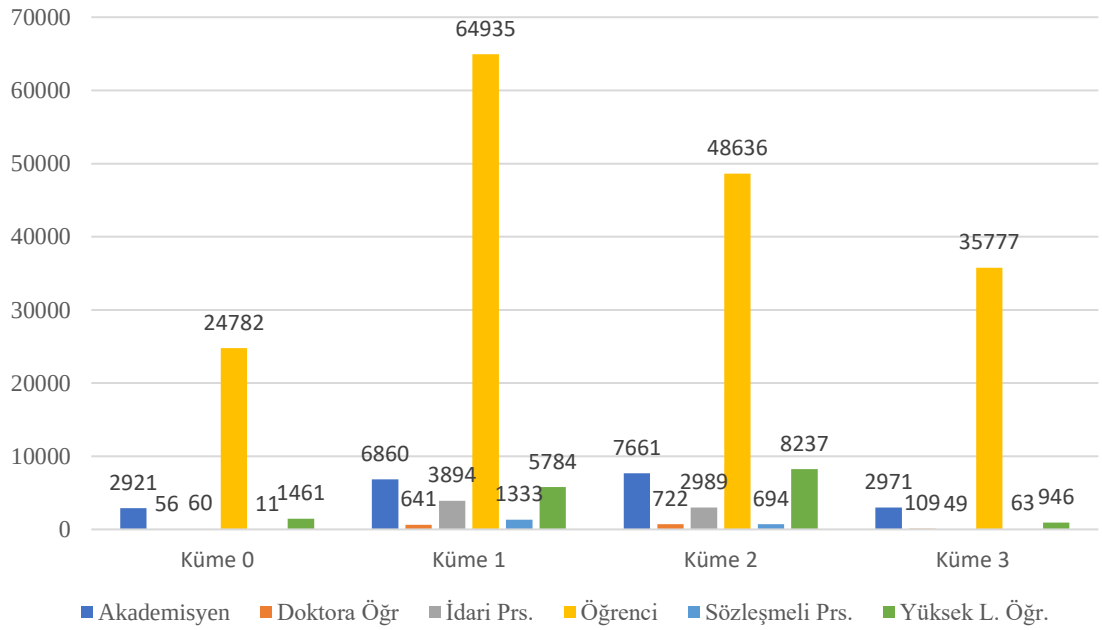
ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri ve edebiyatı (%31,78), 3864 kez matematik (%9,68) ve 3340 kez İngiliz edebiyatı (%8,37) konulu kitaplar ödünç alınmıştır.

Tablo 14: Kümelerin Konu Dağılımı

Küme	Konu	f	%
Küme 0	Amerikan edebiyatı	5445	18,59%
	Eğitim teorisi ve pratiği	4103	14,01%
	Fransız, İtalyan, İspanyol, Portekiz edebiyatı	2038	6,96%
Küme 1	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri ve edebiyatı	36258	43,45%
	Psikoloji	7741	9,28%
	Matematik	6610	7,92%
Küme 2	Eğitim teorisi ve pratiği	15890	23,05%
	Amerikan edebiyatı	9991	14,49%
	Balkan yarımadası tarihi	6101	8,85%
Küme 3	Türk dili ve edebiyatı, Doğu Asya, Afrika, Okyanusya dilleri ve edebiyatı	12686	31,78%
	Matematik	3864	9,68%
	İngiliz edebiyatı	3340	8,37%

Konu dağılımı, kümelerin kendi içerisinde ödünç alınan kitapların ait olduğu konu başlıklarından en fazla bulunan üç konu başlığının belirlenmesiyle oluşturulmuştur. Küme içerisindeki tüm kitapların konu başlıklarını burada sıralamak 190 konu başlığının bulunduğu veri seti için efektif olmayacağı ve buna karşın kümelerdeki kullanıcıların konu başlığı tercihlerini ortaya koymak maksadıyla en fazla tercih edilen üç konu başlığı tabloda verilmiştir.

Kullanıcı gruplarının kümeleme analizi sonucu oluşan kümelere dağılımı incelendiğinde (Şekil 17) Küme 0'da akademisyenler 2921 (%9,97), doktora öğrencileri 56 (%0,19), idari personeller 60 (%0,20), öğrenciler 24782 (%84,61), sözleşmeli personeller 11 (%0,04), yüksek lisans öğrencileri ise 1461 (4,99) olarak kümeye yerleşmektedir. Küme 1'de akademisyenler 6860 (%8,22), doktora öğrencileri 641 (%0,77), idari personeller 3894 (%4,67), öğrenciler 64935 (%77,82), sözleşmeli personeller 1333 (%1,6), yüksek lisans öğrencileri ise 5784 (6,93) şeklinde kümede temsil edilmektedir.



Şekil 17: Kümelerin Grup Dağılımı

Küme 2’de akademisyenler, doktora öğrencileri ve yüksek lisans öğrencileri için diğer kümelere nazaran en yoğun temsiller sağlanmış ve akademisyenler 7661 (%11,11), doktora öğrencileri 722 (%1,05), idari personeller 2989 (%4,34), öğrenciler 48636 (%70,55), sözleşmeli personeller 694 (%1,01), yüksek lisans öğrencileri ise 8237 (11,95) olarak kümeye dağılmıştır. Küme 3’te öğrencilerin diğer kümelere göre en yoğun yer aldığı küme olarak ortaya çıkmış ve akademisyenler 2971 (%7,44), doktora öğrencileri 109 (%0,27), idari personeller 49 (%0,12), öğrenciler 35777 (%89,63), sözleşmeli personeller 63 (%0,16), yüksek lisans öğrencileri ise 946 (2,37) şeklinde kümeye dağılmıştır.

Kümelerin belirlenmesinin ardından her kümenin en fazla ödünç aldığı kitapları bulmak için eser adlarına bağlı frekans hesaplaması yapılmıştır. Buna göre k-means algoritması ile belirlenen 4 küme içinde en fazla ödünç alınan 5 kitap tespit edilmiştir (Tablo 15). En fazla ödünç alınan kitapların tespit edilmesi kümelerdeki kullanıcıların kitap tercihlerini ortaya koymakta ve daha sonra yapılacak olan kitap tavsiyeleri için ön işlem oluşturmaktadır. Tablo 15 incelendiğinde Küme 0 Öğretim ilke ve yöntemleri adlı kitabı 190 kez, Eğitim bilimine giriş adlı kitabı 164 kez, Öğretim teknolojileri ve materyal tasarımı adlı kitabı 157 kez, Olasılıksız adlı kitabı 147 kez ve Eğitim psikolojisi adlı kitabı

142 kez ödünç almıştır. Bu gruptaki kullanıcılar daha çok akademik kitaplar tercih etmekle birlikte popüler romanları da ödünç aldığı anlaşılmaktadır.

Tablo 15: Kümelerin En Fazla Ödünç Aldığı Kitaplar

Küme	Eser Adı	f
Küme 0	Öğretim ilke ve yöntemleri	190
	Eğitim bilimine giriş	164
	Öğretim teknolojileri ve materyal tasarımı	157
	Olasılıksız	147
	Eğitim psikolojisi	142
Küme 1	Suç ve ceza	288
	Genel kimya	269
	Aşk	264
	Piraye	228
	Çalıkuşu	218
Küme 2	Öğretim ilke ve yöntemleri	607
	Bilimsel araştırma yöntemleri	597
	Eğitim psikolojisi	536
	Eğitim bilimine giriş	481
	İlköğretim programı (1-5. sınıflar)	457
Küme 3	Suç ve ceza	182
	Genel kimya	161
	Doğum ve kadın sağlığı hemşireliği	129
	Piraye	114
	Anatomi	111

Küme 1 incelendiğinde Suç ve ceza 288, Genel kimya 269, Aşk 264, Piraye 228, Çalıkuşu ise 218 kez ödünç alınmıştır. Genel olarak edebi eserlerin ödünç alındığı bu kümede Aşk (Elif Şafak), Suç ve ceza (Dostoyevski), Piraye (Canan Tan) ve Çalıkuşu (Reşat Nuri Güntekin) gibi Türk ve dünya romanları arasında çokça okunan kitaplar öne çıkmakta ayrıca Genel kimya adlı akademik kitap en fazla ödünç alınan ikinci kitap olmuştur.

Küme 2 Öğretim ilke ve yöntemleri 607, Bilimsel araştırma yöntemleri 597, Eğitim psikolojisi 536, Eğitim bilimine giriş 481, İlköğretim programı (1-5. Sınıflar) 457 kez ödünç verilmiştir. Bu küme en çok ödünç alınan ilk 5 kitaba bakıldığında genellikle akademik kitapların ve özellikle de eğitim bilimine yönelik kitapların ödünç alındığı bir küme olarak göze çarpmaktadır.

Küme 3’te Suç ve ceza 182, Genel kimya 161, Doğum ve kadın sağlığı hemşireliği 129, Piraye 114 ve Anatomi adlı kitap 111 kez ödünç verilmiştir. Kümenin genel yönelimi akademik ve edebi eserlerin ödünç alındığı ve daha çok sağlık bilimleri alanında kitapların tercih edildiği anlaşılmaktadır.

Kümeleme analizinin ardından kullanıcıların kitap tercihlerine yönelik kitap tavsiyeleri yapmak üzere kullanıcıların en sık yer aldığı kümeler tespit edilmiştir. Bunun için UYE-ID değeri temel alınarak hangi kümede daha fazla yer alıyorsa o kümeye dahil edilmesi sağlanmıştır. Böylelikle her kullanıcının yalnızca bir kümeden kitap önerisi alması amaçlanmıştır. Ardından her kullanıcıya ait olduğu kümenin daha önce belirlenmiş olan en fazla ödünç alınan 5 kitabı tavsiye listesi haline getirilmiştir. Tavsiye kitap listesi satırlarda UYE-ID olacak şekilde her bir sütunda bir kitap adı yer almak üzere dönüştürülmüş ve liste harici bir xlsx dosyasına yazdırılarak analiz tamamlanmıştır.

Tablo 16: Kullanıcı Kitap Tavsiyeleri (k-means)

ÜyeID	Küme	Tavsiye 1	Tavsiye 2	Tavsiye 3	Tavsiye 4	Tavsiye 5
0	2	Öğretim ilke ve yöntemleri	Bilimsel araştırma yöntemleri	Eğitim psikolojisi	Eğitim bilimine giriş	İlköğretim programı (1-5. sınıflar)
1	1	Suç ve ceza	Genel kimya	Aşk	Piraye	Çalıkuşu
2	2	Öğretim ilke ve yöntemleri	Bilimsel araştırma yöntemleri	Eğitim psikolojisi	Eğitim bilimine giriş	İlköğretim programı (1-5. sınıflar)
3	1	Suç ve ceza	Genel kimya	Aşk	Piraye	Çalıkuşu
4	3	Suç ve ceza	Genel kimya	Doğum ve kadın sağlığı hemşireliği	Piraye	Anatomi
5	1	Suç ve ceza	Genel kimya	Aşk	Piraye	Çalıkuşu
6	1	Suç ve ceza	Genel kimya	Aşk	Piraye	Çalıkuşu
7	2	Öğretim ilke ve yöntemleri	Bilimsel araştırma yöntemleri	Eğitim psikolojisi	Eğitim bilimine giriş	İlköğretim programı (1-5. sınıflar)
8	2	Öğretim ilke ve yöntemleri	Bilimsel araştırma yöntemleri	Eğitim psikolojisi	Eğitim bilimine giriş	İlköğretim programı (1-5. sınıflar)
9	3	Suç ve ceza	Genel kimya	Doğum ve kadın sağlığı hemşireliği	Piraye	Anatomi

Tablo 16 incelendiğinde veri setindeki ilk 10 kullanıcı için kullanıcıları ifade eden UYE-ID ile başlayan satırda kullanıcının *k*-means algoritması ile belirlenen küme numarası verilmiş ardından da kümeleme analizinden elde edilen modelleme ile 5 farklı kitap tavsiyesi verilmiştir. Aynı kümedeki kullanıcılara, aynı kitapları tercih edebileceği düşüncesiyle küme içerisinde en çok ödünç alınan kitaplar önerilmiştir. Ortaya koyulan kitap tavsiye modelinin Precision değeri 0,0208, Recall değeri 0,0162, F1 Score 0,1530 ve RMSE değeri 0,9895 olarak meydana gelmiştir.

3.8.3. İşbirlikçi Filtreleme

Kütüphane kullanıcılarının geçmiş kitap ödünç alma davranışlarından model oluşturarak ihtiyaç duyulacak kitaplardan her bir kullanıcı için öneri oluşturmak üzere işbirlikçi filtreleme analizi yapılmıştır. İşbirlikçi filtreleme, tavsiye sistemlerinin (recommender systems) bir türü olarak genellikle kullanıcı-içerik etkileşim verilerini kullanarak model geliştiren, kullanıcı tercihlerini tahmin etmek için birden fazla kullanıcının tercihlerini kullanan yöntemdir (Kantardzic, 2019: 375). Kullanıcı tercihleri, kullanıcı-öge matrisi üzerinden veri odaklı olarak belirlenebilir ve elde edilen model ile hedef kullanıcı(lar) için tahmin yapılabilir. İşbirlikçi filtreleme, belirli bir kullanıcının ilgi alanlarına uygun şekilde otomatik olarak seçilen kümenin gelişmiş görüş ve tavsiyelerini ortaya koyma çabasıdır (Hand, Mannila ve Smyth, 2001: 471). İşbirlikçi filtreleme algoritmaları genellikle hedef kullanıcıya benzeyen kullanıcı profillerinin alt kümesini tespit eder ve bu kümenin tercihlerine göre bir öneri hesabı yapar. Sistem tavsiye oluşturmak için hedef kullanıcı ve veri setindeki kullanıcıların özellik ve tercihlerine bağlı olarak çalıştığı için oluşturulan önerilerin kalitesi modele dahil edilen veri miktarı ve kalitesiyle doğru orantılıdır. İşbirlikçi filtreleme, 1992 yılında Xerox Palo Alto Research Center’da geliştirilen ve kullanıcılara artan eposta yoğunluğu karşısında ilgi duydukları konulardaki epostaları filtreleyebilme imkanı veren Tapestry sistemi ile ortaya çıkmıştır (Goldberg vd., 1992). Tapestry sistemi, büyük hacimli eposta yönetiminde kullanıcılara ilgi alanına yönelik eposta listelerine abone olma ve filtreleme çözümünün ötesinde insanları da filtreleme sürecine dahil ederek okunan epostalar hakkında tepkilerini kaydederek kullanıcıların birbirleri için filtreleme yapmasına yardımcı olması mantığına dayanır. İşbirlikçi filtreleme ile ilgili ikinci önemli çalışma 1994 yılında Tapestry sistemine benzer bir çalışma olarak yapılmıştır. GroupLens adındaki işbirlikçi filtreleme sistemi, NetNews makaleleri için okuyucuların geniş makale akışından

beğenecekleri makalelere erişmelerini sağlamak üzere tasarlanmış ve okuyucuların verdiği değerlendirme puanları üzerinden benzer tercihe sahip okuyucuların aynı içerikleri tercih edeceği düşüncesine dayanmaktadır (Resnick vd., 1994).

İşbirlikçi filtrelemenin komşuluk metodu ve gizli faktör metodu olmak üzere iki ana çalışma alanı vardır (Koren, Bell ve Volinsky, 2009: 31). Komşuluk metodu, öğeler veya kullanıcılar arasındaki ilişkileri hesaplamaya odaklanırken gizli faktör metodu kullanıcıların içeriğe vermiş olduğu derecelendirme veya puanlara açıklama getirmek için öğeler ve kullanıcıları puanlama kalıplarından çıkarılan faktörler üzerinden karakterize etmeye çalışan sistemdir. Gizli faktör modellerinin en başarılı uygulamalarından biri bu çalışmanın da gerçekleştirildiği matris faktörizasyon yöntemine dayanmaktadır. Matris faktörizasyonu, gözlemlenen verilerin gizli boyutlarını keşfetmek için kullanılan ve öneri sistemleri, metin madenciliği, bilgisayarlı görü, tıbbi teşhis, boyut azaltma ve veri sıkıştırma gibi pek çok alanda ilgi gören bir tekniktir (Bhavana ve Padmanabhan, 2021: 4016). Matris faktörizasyon modeli ile bilgi erişiminde gizli anlamsal faktörleri tanımlamak için *tekil değer ayrışımı* (singular value decomposition – SVD) bilinen en iyi tekniklerden biridir (Koren, Bell ve Volinsky, 2009: 32). SVD, bir çok alanda uygulaması olan matris ve matrisleri içeren problemleri analiz etmek için güçlü bir hesaplama aracı (Forsythe ve Malcolm, 1977: 204), terim ve dokümanları seçilebilir boyutlu bir uzayda vektörler olarak temsil eden ve uzaydaki noktalar arasındaki nokta çarpımı veya kosinüs benzerliklerini veren (Deerwester vd., 1990: 395), bir matrisi birden fazla bileşen matrise ayırmak ve matrisin duyarlılığını ölçebilen, doğrusal bir sistemin sayısal hataya duyarlılığını ölçebilen veya matris için optimal düşük dereceli yaklaşımını elde edebilen tekniktir (Chengwang, 2010: 422). $m \times n$ sütunlarına sahip karmaşık bir A matrisi verildiğinde, $U (m \times n)$ ve $V (n \times n)$ ortogonal matrisler, Σ ana köşegende negatif olmayan elemanlar olmak üzere $A = U \Sigma V^*$ formülü A matrisi için tekil değer ayrışımıdır. Araştırmanın özelinde A kullanıcı-kitap matrisini, Σ kullanıcı-kitap matrisinin önem derecesini gösteren tekil değer matrisini, U kullanıcıların latent (gizli) özelliklerini, V kitapların latent (gizli) özelliklerini, $*$ matrisin transpozunu ifade etmektedir.

Kullanıcıların kitap ödünç alma eğilimlerine yönelik kitap öneri sistemi uygulaması için işbirlikçi filtreleme yöntemlerinden SVDs (Propack) algoritması ile kitap önerileri verilmiştir. SVDs algoritması, büyük ve seyrek matrislerin tekil değer ayrışımı için bir dizi fonksiyon içeren Propack yazılım paketine dahil olan bir SVD algoritması

uyarlamasıdır (Larsen, 2012). SVD algoritması büyük veri kümelerinde hesaplama maliyeti (Yang, Ma ve Buja, 2014: 925), seyreklik problemi (Golub ve Loan, 2013: 597; Koren, Bell ve Volinsky, 2009: 32), büyük bellek hacmi gerektirmesi ve işlemlerin uzun zaman alması (Zhang ve Thurber, 2007: 337) gibi problemler bu araştırma kapsamında da ortaya çıkmış ve scipy.sparse.linalg kütüphanesine dahil SVDs algoritması kullanılmıştır. Aşağıdaki tabloda hem orijinal SVD algoritması hem de SVDs algoritması ile veri setinde yapılan zaman, seyreklik ve RMSE incelemesi verilmiştir (Tablo 17). Yapılan karşılaştırma bu çalışmanın tüm analizlerinde kullanılan i7-13700H işlemci, 32 GB DDR4 bellek, Nvidia GeForce RTX 4060 ekran kartı ve Windows 11 işletim sistemine sahip bilgisayarda gerçekleştirilmiş; optimum gizli faktör sayısı $k=60$ olarak hesaplanmıştır.

Tablo 17: Tekil Değer Ayırışımı Performans Karşılaştırma

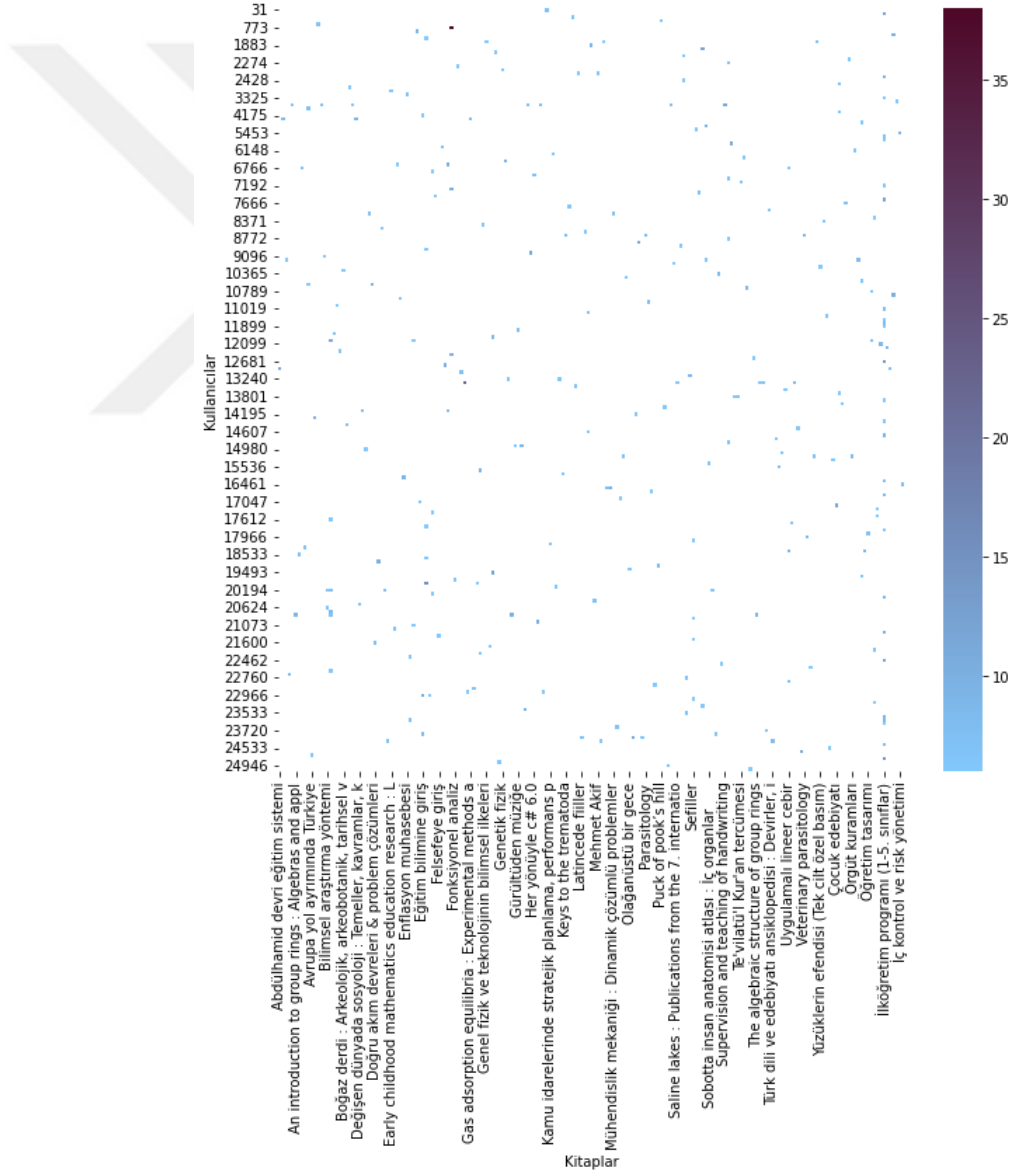
Algoritma	Seyreklik	Zaman (Saniye)	RMSE
SVD	0,9997	3306,03	0,01881
SVDS		61,73	0,01881

Tablo incelendiğinde veri setinin seyreklik oranı %99,97, SVD algoritması için işlem süresi 3306 saniye (~55 dk); SVDS algoritması için işlem süresi 61 saniye ve her iki algoritma için de RMSE değeri 0,01881 olarak gerçekleşmiştir. 221592 kayıttan oluşan nispeten büyük bir veri seti ile çalışılıyor olması, seyreklik oranının yüksek olması ve işlem süresinin uzun sürmesi SVDs algoritmasının kullanılması için başlıca sebepler olmuştur. RMSE değerleri ise her iki algoritma için aynı olarak gözlenmiş, algoritma seçiminde RMSE değerinden feragat etme gibi bir durum ortaya çıkmamıştır.

SVDs algoritması için Python dildinde, pandas, numpy, scipy.sparse.linalg, sklearn.metrics, math.sqrt kütüphane ve fonksiyonlar Spyder IDE paket programında çağrılmış ve analiz yürütülmüştür. Tekil değer ayrışımı algoritmasına dayalı kitap öneri sistemi için kütüphane otomasyon sisteminden temin edilen ve veri ön işleme sürecinden geçirilen nihai veri setinden ‘Üye-Id’ Ve ‘Eser-Adı’ değişkenleri analize dahil edilmiştir. Her iki değişken sütununda da eksik değer bulunmamaktadır. xlsx ortamında tutulan veriler Spyder IDE paket programına çağrılmış, veri setine ön bakış yapılmış ve ‘Üye-Id’

Ve ‘Eser-Adı’ sütunlarında bulunan 221592 kayıt ile analizler başlatılmıştır. `pd.pivot_table` fonksiyonu ile ‘Üye-Id’ Ve ‘Eser-Adı’ dan oluşan kullanıcı-kitap etkileşim matrisi oluşturulmuştur. Ayrıca `fill_value` fonksiyonu ile boş değerler 0 ile doldurulmuştur.

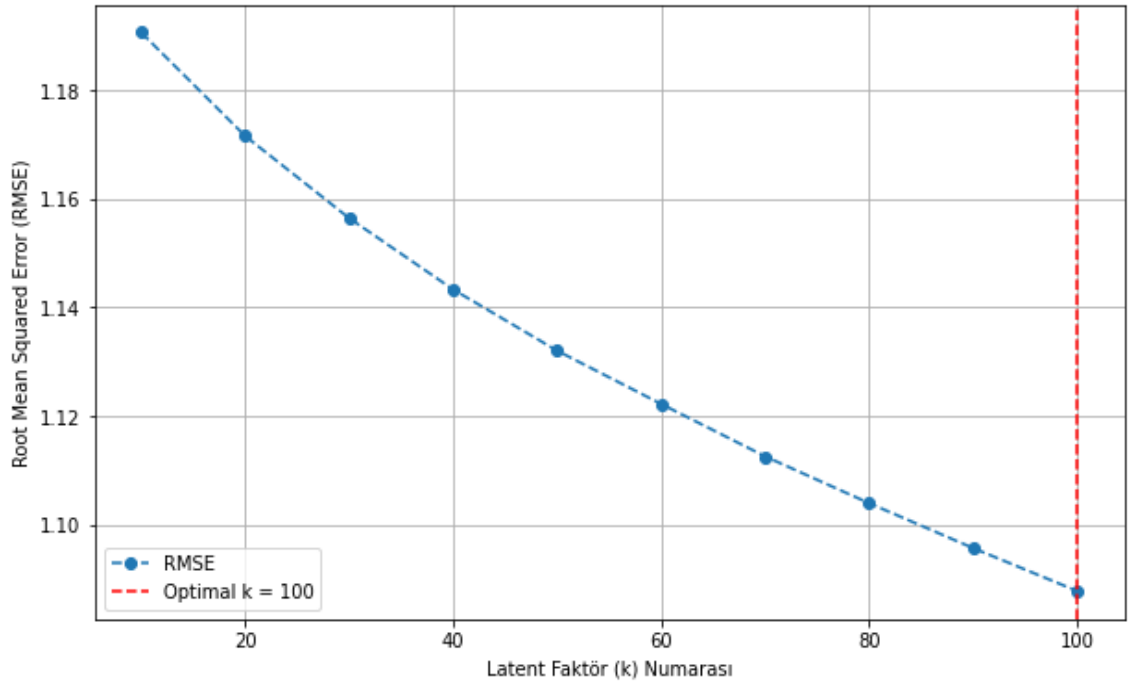
`pd.pivot_table` fonksiyonu ile oluşturulan kullanıcı-kitap pandas dataframe’i veri setindeki kayıt sayısı dikkate alınarak performans ve hesaplama verimliliği adına numpy dizisine çevrilmiştir. Ardından satır bazında her kullanıcının kitap alıp almadığını belirlemek için kullanıcı-kitap matrisine (Şekil 18) kitap ödünç alınmışsa 1 alınmamışsa 0 kodlanır. Böylelikle kullanıcıların kitaplarla olan etkileşimi sayısal olarak belirlenir.



Şekil 18: Kullanıcı-Kitap Etkileşim Matrisi Isı Haritası

Şekil 18’de 5’ten fazla ödünç alınan kitaplar için kullanıcı-kitap etkileşim matrisi verilmiştir. Analizlerde ödünç verilen tüm kitaplar ve kullanıcılar için etkileşim matrisi oluşturulmuş fakat grafikte 25093 x 28650 satır ve sütundan oluşan matrisin gösterilmesi görsel açıdan anlaşılabilir olmadığı için filtreleme işlemi yapılmıştır. Buna göre 5’ten fazla ödünç alınan basılı kitaplar ile kullanıcılar arasındaki etkileşim (ödünç) matrisi verilmiştir. Isı haritasının yoğunlukla açık mavi renkten müteşekkil oluşu kitapların genel itibarıyla düşük ödünç alınma sayılarını ifade etmektedir.

Bu aşamada SVDs algoritması ile analize başlamadan önce algoritmanın optimum k latens (gizli) faktör sayısının belirlenmesi gerekmektedir. Optimum k latent sayısını belirlemek üzere 10 ile 100 arasındaki k sayılarının RMSE değerleri hesaplanmış ve $k=10$ için $RMSE=1,19061$; $k=20$ için $RMSE=1,17154$; $k=30$ için $RMSE=1,15641$; $k=40$ için $RMSE=1,14326$; $k=50$ için $RMSE=1,13199$; $k=60$ için $RMSE=1,12219$; $k=70$ için $RMSE=1,11252$; $k=80$ için $RMSE=1,10390$; $k=90$ için $RMSE=1,09563$; $k=100$ için $RMSE=1,08773$ değerleri elde edilmiştir (Şekil 19). RMSE değerine göre en düşük değer k faktör sayısının 100 olduğu noktada sağlanmış ve kitap öneri sistemi için $k=100$ latent faktör sayısı seçilmiştir.



Şekil 19: k Latent Sayısı Belirleme

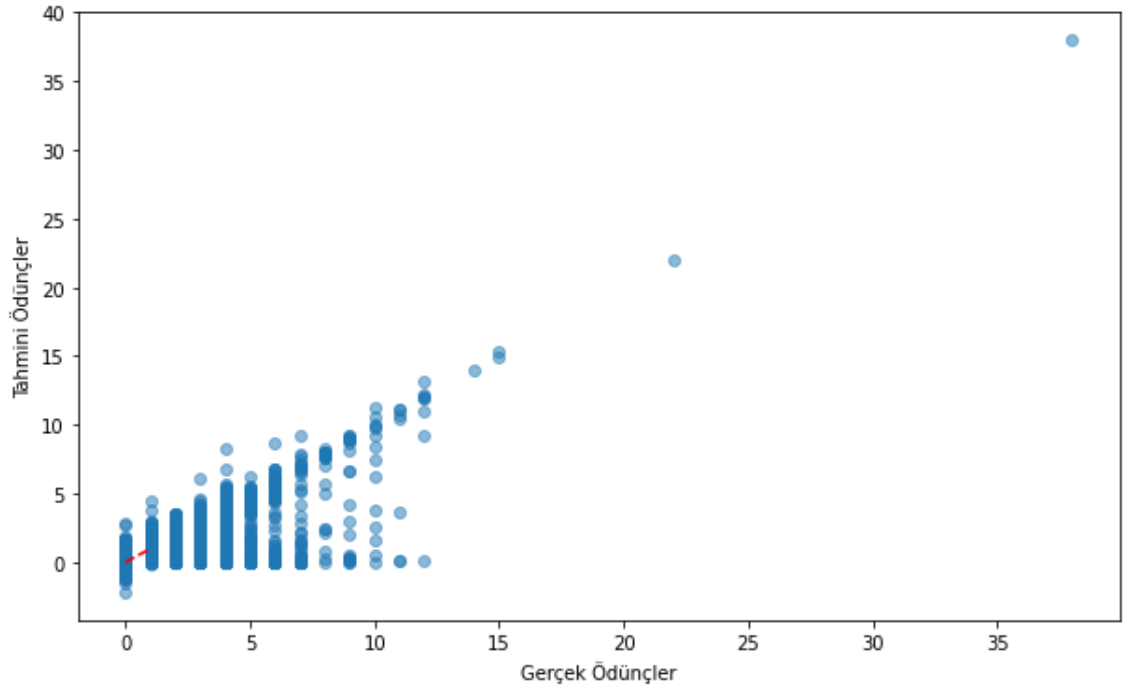
$K=100$ latent faktör sayısına göre SVDS algoritması uygulanmış ve yeniden yapılandırılan matris ile tahmini derecelendirmeler uygulanmış; her kullanıcı için 5 kitap

önerisi hesaplanmıştır. Ardından oluşturulan kitap önerileri dataframe formatına dönüştürülmüş, her bir kitap adı bir sütunda ve her kullanıcı numarası satır başında olacak şekilde düzenlenmiştir. Analizin son adımı olarak RMSE performans değerlendirmesi yapılmıştır. Aşağıda ilk 10 kullanıcı için yapılan kitap önerileri verilmiştir (Tablo 18). Analiz 115 saniyede tamamlanmış ve RMSE=0,018333 değeri elde etmiştir.

Tablo 18: Kullanıcı Kitap Tavsiyeleri (SVDS)

ÜyeID	Tavsiye 1	Tavsiye 2	Tavsiye 3	Tavsiye 4	Tavsiye 5
0	On medicine : Books 1-4	Ateşten gömlek	İncir kuşları	İntibah	İki kişilik yalnızlık
1	Genetik kavramlar	Anatomi	Anlamdan anlatıma Türkçemiz	Sosyoloji	İnsan anatomisine giriş
2	Dünya ve ülkeler coğrafyası	Anlamdan anlatıma Türkçemiz	Canlılar bilimi	Kuramdan uygulamaya fen ve teknoloji öğretimi	Beyaz zambaklar ülkesinde
3	Uçurtma avcısı	Genel fizik ve teknolojinin bilimsel ilkeleri	Dönüşümler ve geometriler	Yaşamın sihirli yılları: Okul öncesi dönem	Fonksiyonel analiz
4	Sis ve gece	Şeytan ayrıntıda gizlidir	Bab-ı esrar	Aşk köpekliktir	Da Vinci şifresi
5	Siyah süt: Yeni başlayanlar için postpartum depresyon	Küçük arı	Aşk köpekliktir	Simyacı	Bit palas
6	Yeni programa uygun Türkçe öğretim yöntemleri	Dünya ve ülkeler coğrafyası	Psikolojik danışmanlık ve rehberlik	Mutluluk	Edebiyat bilgi ve kuramları
7	Fen ve mühendislik için fizik 1 : Mekanik-mekanik dalgalar-termodinamikler	Türk eğitim sistemi ve okul yönetimi	Psikolojik danışma ve rehberlik	Anna Karenina	Fareler ve insanlar
8	Sefiller	Şeker portakalı	Matematik analiz	Çözümlü analitik geometri alıştırmaları	Biyokimya
9	Biyokimya	Şah ve Sultan	Mikro ekonomi	Uluslararası iktisadi örgütler ve az gelişmiş ülkeler	Piruze : Şam'da bir Türk gelini

Kitap önerileri incelendiğinde üye numarasına göre sıralanan kullanıcılar için her biri ayrı sütunda 5 kitap önerisi verilmiştir. 0 numaralı kullanıcı için On medicine: Books 1-4, Ateşten gömlek, İncir kuşları, İntibah, İki kişilik yalnızlık kitaplarının önerildiği görülmektedir. 1 numaralı kullanıcı için Genetik kavramlar, Anatomi, Anlamdan anlatıma Türkçemiz, Sosyoloji, İnsan anatomisine giriş kitapları sistem tarafından önerilmiştir. Kitap öneri listesi incelendiğinde her bir kullanıcı için kişiselleştirilmiş ve birbirini tekrar etmeyen kitap önerileri yapıldığı görülmektedir.



Şekil 20: Tahmin-Gerçek Karşılaştırması

Son olarak kitap öneri sisteminin yapmış olduğu kullanıcılara yönelik kitap tahminlerinin gerçek kitap ödünçleri ile karşılaştırılması Şekil 20’de verilmiştir. Buna göre ortaya koyulan modelin genel olarak doğru tahmin yaptığı ancak 20 ve üzeri değerlerde bazı aykırı tahminler olduğu saptanmıştır. Ortaya koyulan tahminler kullanıcı-kitap etkileşim matrisinin yeniden yapılandırılması ile tahmini derecelendirmeleri hesaplamaktadır. Daha önce verildiği üzere $RMSE=0,018333$ değeri, grafiği doğrular nitelikte 0’a yakın çıkmış ve modelin yüksek oranda başarılı olduğunu göstermiştir.

SONUÇ

Gelişen bilgi teknolojisi ile insanlık ihtiyaç duyduğu bilgiye her an her ortamda erişebilir düzeye gelmiş, yakın döneme kadar büyük bir keşmekeş olan bilgi kaynaklarına fiziksel erişim engelleri büyük oranda yıkılmıştır. İnsanlar her ne kadar dijital ortamda ihtiyaç duyduğu bilgiye zahmetsizce erişebildiğini düşünse de bilgi teknolojilerinin dezavantajları neticesinde de karşı karşıya kaldığı bilgi bombardımanı ile doğru ve nitelikli bilgiye erişim konusunda halen kütüphane (bilgi merkezi) kurumuna -fiziksel bir bina olmasa dahi- ihtiyaç duymaktadır. Zira basılı veya dijital ortamlarda üretilen içerik insanoğlunun ömrü boyunca tüketebileceğinin çok ötesindedir. Tüm platformlar müzik, film, resim, oyun ve kitaplar ile dolup taşmakta, kullanıcılar üretilen bu içerikler arasından ihtiyaç veya ilgi duyduğu içeriklere erişimde çaresiz kalabilmektedir. Dijital içerik sağlayıcılar ve e-ticaret platformları ürünlerini kullanıcıların ihtiyaçlarına yönelik olarak pazarlama noktasında kişiselleştirilmiş ürün önerileri ile etkin bir pazarlama tekniği gütmeye gayretindedir. Her ne kadar kütüphaneler binlerce yıllık kadim usuller üzerine bina edilmiş bir hizmet politikasına sahip olsa da kütüphanelerin her geçen gün biraz daha “etüt merkezi” olma problemi bilgi kaynaklarını kullanıcılarına etkin şekilde ulaştırılması ile aşılabilecektir. Veri madenciliği ise diğer alanlarda olduğu gibi kütüphanelere hem bilgi kaynaklarının kullanım etkinliği hem de kullanıcıların eğilimlerini belirleme noktasında fayda sağlamakta; kullanıcılara ihtiyaç ve eğilimlerine uygun bilgi kaynaklarının sunulması adına çözümler üretmektedir.

Bu çalışma kapsamında Burdur Mehmet Akif Ersoy Üniversitesi Kütüphane ve Dokümantasyon Daire Başkanlığı tarafından kullanılan kütüphane otomasyon sisteminden temin edilen ödünç (dolaşım), üye ve katalog verileri ile RapidMiner programı vasıtasıyla kütüphane kullanıcılarının yaş, cinsiyet, grup, akademik birim, gibi demografik özellikleri verilmiş; 2009-2023 yılları arasında en çok ödünç alınan kitaplar belirlenmiş, ödünç verilen kitapların kümülatif konu dağılımları tespit edilmiş, yıllık verilen ödünç kitap sayıları ortaya koyulmuş; cinsiyet ve yaş grubuna göre kitap tercihleri açıklanmıştır. Ayrıca Spyder IDE programı vasıtasıyla kullanıcılar *k-means* kümeleme analizine tabi tutulmuş, Silhouette ve Elbow değerleri baz alınarak belirlenen küme sayısına göre kümelere ayrılmış, kullanıcı kümelerinin kitap ödünç alma davranışları, küme yoğunlukları, kümelerin cinsiyet, yaş, konu başlığı ve grup dağılımları irdelenmiş, son olarak kümelerin kitap ödünç alma davranışlarına yönelik her kullanıcı için kitap

tavsiyeleri verilmiştir. Kümeleme analizi ile oluşturulan kitap tavsiye modelinin Precision değeri 0,0208, Recall değeri 0,0162, F1 Score 0,1530 ve RMSE değeri 0,9895 olarak ölçülmüştür. Kümeleme analizinden sonra işbirlikçi filtreleme yöntemi ile kullanıcıların kitap ödünç alma eğilimlerine uygun kişiselleştirilmiş kitap öneri sistemi sunulmuştur. İşbirlikçi filtreleme yöntemlerinden SVDs algoritması ile kullanıcı-kitap etkileşim matrisi oluşturulmuş, optimum k latent sayısına göre SVDs algoritması uygulanarak kullanıcılar için tahmini derecelendirmelere göre en uygun 5'er kitap önerisi verilmiştir. Geliştirilen kişiselleştirilmiş kitap öneri sistemi 0,18333 RMSE değeri alarak başarılı bir performans ortaya koymuştur.

Tanımlayıcı istatistiklere bakıldığında kütüphaneden ödünç kitap alan kullanıcıların %65,07'si kadın, %34,93'ü de erkeklerden oluşmaktadır. Bu oran Burdur Mehmet Akif Ersoy Üniversitesi Prof. Dr. İlhan Varank Kütüphanesi'nin kadınlar tarafından daha çok kullanıldığına işaret etmektedir. Bu sonuç 2020 yılında Trakya Üniversitesi Merkez Kütüphanesi'ne yönelik yapılan araştırma ile benzerlik göstermektedir (Taşkın, 2020: 52). Kütüphaneden ödünç kitap alan kullanıcıların yaş aralığına bakıldığında % 85,6 16-25 yaş aralığında, %10,54 26-35 yaş aralığına, %3,06 36-45 yaş aralığında, %0,7 46-55 yaş aralığında ve %0,1 56 yaş ve üzerindeki kullanıcılardan oluşmaktadır. Yaş hesaplaması yapılırken kullanıcıların doğum yılı ile ödünç işleminin yapıldığı yıl hesaba katılarak yaş verisi oluşturulmuştur. Çünkü geçmişe dönük olarak kullanıcıların bugünkü yaşının verilmesi işlemin yapıldığı ana herhangi bir ilişki kuramayacağı çıkarımından hareket edilmiştir. Ancak kullanıcının kitabı ödünç aldığı anda kaç yaşında olduğu ve bu yaş grubunun kitap alma davranışlarına ne gibi bir etkisinin olabileceği bu araştırmanın irdelediği konulardan biridir. Yaş aralıklarına bakıldığında kütüphane kullanıcıların ekseriyetle 16-25 yaş aralığında olduğu ortaya çıkmaktadır. Burdur Mehmet Akif Ersoy Üniversitesi Prof. Dr. İlhan Varank Kütüphanesi'nin potansiyel kullanıcı kitlesi düşünüldüğünde üniversite mensuplarının büyük bir kısmının öğrenciler tarafından oluşuyor olması kütüphane kullanıcılarının da yaş aralığını yoğunlukla 16-25 yaş aralığına çekmektedir. Öte yandan özellikle 46 yaş üzerindeki kullanıcıların -ki bu yaş grubundaki kullanıcılar genellikle akademisyen ve idari personellerden oluşmakta- kütüphane kullanımının sayısal ve oransal olarak düşük olması bir diğer dikkat edilmesi gereken konudur. Kütüphaneden ödünç kitap alan üniversite mensuplarının oranını daha iyi anlayabilmek için 2024 yılı itibariyle

üniversitenin öğrenci ve personel sayısına bakıldığında fakülte, yüksekokul ve meslek yüksekokullarda toplam 32530, yüksek lisansta 2620, doktora 370 öğrenci bulunurken, 1120 akademisyen, 408 idari personel ve 485 sözleşmeli personel bulunmaktadır (Burdur Mehmet Akif Ersoy Üniversitesi, 2024). Buna karşın kütüphaneden 2008 ila 2024 yılları arasında toplam 22431 öğrenci (%89,39), 1463 yüksek lisans öğrencisi (%5,83), 701 akademisyen (%2,79), 271 idari personel (%1,08), 116 sözleşmeli personel (%0,46) ve 111 doktora öğrencisi (%0,44) ödünç kitap almıştır. Ortaya çıkan sayılar pek çok üniversite öğrenci, akademisyen ve personellerinin kütüphaneden hiç kitap ödünç almadığı sonucunu göstermektedir. Kütüphaneden ödünç kitap alan kullanıcıların akademik birimleri irdelendiğinde %40 oranında Eğitim Fakültesi, %12,65 oranında İktisadi ve İdari Bilimler Fakültesi, %11,33 oranında Fen-Edebiyat Fakültesi, %5,93 oranında Mühendislik Mimarlık Fakültesi ve %5,71 oranında Sağlık Bilimleri Fakültesi ilk beşi oluşturmaktadır. Eğitim Fakültesi (1965), Veteriner Fakültesi (1992), Sağlık Bilimleri Fakültesi (1996) ve Fen Edebiyat Fakültesi (2006) Burdur Mehmet Akif Ersoy Üniversitesi'nin en eski fakülteleri olarak kütüphaneden en fazla ödünç kitap alan akademik birimler arasında yer alması şaşırtıcı değildir. Ancak İktisadi ve İdari Bilimler Fakültesi (2011) nispeten geç kurulmasına karşın ikinci en çok kitap alan fakülte olması dikkat çekicidir.

2009 ila 2023 yılları arasında en çok okunan ilk iki kitaba bakıldığında 2009 yılında “Aşk-ı memnu” ve “Sevda linki”, 2010-2011 ve 2012 yıllarında üç yıl üst üste “İlköğretim programı (1-5. Sınıflar)” adlı kitap en fazla ödünç alınmış, 2010 ve 2011 yılında ikinci olarak en fazla “Öğretim ilke ve yöntemleri” ve 2012 yılında “Öğretim teknolojileri ve materyal tasarımı” en fazla ödünç alınan ikinci kitap olmuştur. 2013 yılında “Öğretim teknolojileri ve materyal tasarımı” ve “Bilimsel araştırma yöntemleri”; 2014 yılında “Eğitim bilimine giriş” ve “Öğretim ilke ve yöntemleri”; 2015 yılında “Öğretim ilke ve yöntemleri” ve “Bilimsel araştırma yöntemleri”; 2016 yılında “Eğitim bilimine giriş” ve “Bilimsel araştırma yöntemleri”; 2017 yılında “Bilimsel araştırma yöntemleri” ve “Öğretim ilke ve yöntemleri”; 2018 yılında “Eğitim psikolojisi” ve “Bilimsel araştırma yöntemleri”; 2019 yılında “Öğretim ilke ve yöntemleri” ve “Eğitim psikolojisi”; 2020 yılında “Bilimsel araştırma yöntemleri” ve Çalığı; 2021 yılında “Olağanüstü bir gece” ve “Şeker portakalı”; 2022 yılında “Eğitim sosyolojisi” ve “Eğitim psikolojisi”; 2023 yılında “Anna Karenina” ve “Eğitim sosyolojisi” en fazla ödünç alınan

kitaplar olmuştur. Kronolojik olarak en fazla ödünç alınan kitapların dağılımı özellikle eğitim bilimleri alanındaki kullanım oranlarını doğrular niteliktedir. Öte yandan ödünç verilen kitapların kümülatif konu dağılımlarına bakılmış ve en fazla %41,06 oranında dil ve edebiyat, % 11,69 oranında bilim, % 10,74 oranında eğitim, % 8,94 oranında sosyal bilimler, %7,44 oranında felsefe-psikoloji-din konulu kitaplar ödünç alınmıştır. Dil ve edebiyat, bilim ve eğitim konulu kitapların çokça tercih edilmesi daha önce yapılan araştırmalar (Taşkın, 2020: 57; Uçan, 2010: 67) ile benzerlik göstermektedir. Ayrıca 2009-2023 yılları arasında toplam ödünç verilen kitap sayıları irdelenmiş ve 2009-2011 yılları arasında istikrarlı bir yükselişin yaşandığı, 2014 yılına kadar ciddi bir düşüş yaşandığı, 2014 yılı itibariyle 2019 yılına kadar dalgalı da olsa bir yükselişin olduğu 2019 yılında pik noktayı gördükten sonra 2020 itibariyle Covid-19 pandemisinin etkisiyle düşüş ve sonra 2021, 2022 yıllarında toparlanmaların yaşandığı ve en nihayetinde 2023 yılında tekrar düşüş ile sonlandığı tespit edilmiştir. Yıllık ödünç verilen kitap sayısını etkileyen bir husus olarak kütüphane binasının üç kez yer değiştirmiş olması gösterilebilir. Zira kütüphane 2012 yılına kadar üniversitenin eski Eğitim Fakültesi'nde hizmet vermiş, 2013 yılında İstiklal Yerleşkesi'nde Sağlık Kültür ve Spor Daire Başkanlığı'nda bulunan mekâna ardından da 2018 yılında da mevcut müstakil binasına taşınmıştır. Yıllık ödünç sayıları da bu yer değişiklikleri ile ilişkili olabileceğini göstermektedir. Düşüş eğilimindeki sayılar 2012 yılında İstiklal Yerleşkesine taşınma sonrası toparlanmış, 2018 yılında ise mevcut modern kütüphane binasına taşındıktan sonra zirveye ulaşmıştır.

Kullanıcıların demografik özelliklerinin kitap tercihlerine etkisine bakıldığında en çok tercih edilen kitap konuları erkeklerde %16,97 oranında Türk dili ve edebiyatı, % 7,56 oranında eğitim teorisi ve pratiği, % 5,33 oranında matematik olurken kadınlarda % 24,63 oranında Türk dili ve edebiyatı, % 9,75 eğitim teorisi ve pratiği, % 8,5 oranında Amerikan edebiyatı konulu kitaplar olduğu görülmektedir. Analizin devamında cinsiyete göre bazı farklılıklar olmakla birlikte cinsiyet değişkeninin konu başlığı bakımından kitap tercihlerine ciddi bir etkisinin olmadığını söylemek mümkündür. Öte yandan yaş gruplarına göre kitap tercihleri irdelenmiş ve tüm yaş gruplarında Türk dili ve edebiyatı konulu kitapların en çok tercih edildiği sonucuna varılmıştır. Yine 16-25, 26-35- ve 36-45 yaş aralığındaki kullanıcıların en çok tercih ettiği ikinci konu başlığı eğitim teorisi ve pratiği iken 46 yaş üzerindeki kullanıcılarda bu eğilim yerini psikoloji ve Türkiye hukuku

konulu kitaplara bırakmaktadır. Amerikan edebiyatı konulu kitapların 26-35 yaş aralığındaki kullanıcılar hariç tüm gruplarda tercih ediliyor olması bir diğer dikkat çeken konu başlığıdır.

K-means kümeleme analizi için kullanıcıların cinsiyet, yaş, fakülte, bölüm, grup gibi demografik niteliklerle birlikte daha önce ödünç almış olduğu kitap konusu seçilmiş ve model bu değişkenlerle kurulmuştur. K-means algoritması kategorik değişkenlerle çalışmadığı için fakülte, bölüm, konu değişkenleri sayısallaştırılmış, yaş değişkeni için eksik değerler medyan ile doldurulmuş ve yaş aralıklarına bölünerek sayısallaştırılmıştır. Kümeleme analizi yapılmadan önce veri setinin bölüneceği küme sayısını belirlemek için Silhouette index ve Elbow yöntemi kullanılmış ve her iki yöntemin işaret ettiği 4 kümeye veri seti bölünmüştür. Oluşan küme sayısı alanda yapılan diğer çalışmalarla benzerlik göstermektedir (Tempelman-Kluit ve Pearce, 2014; Uçan, 2010). Öte yandan 2 ve 3 küme oluşan bazı çalışmalarla da ayrılmıştır (Fırat, 2019; Taşkın, 2020).

Bölme işleminde en yoğun küme Küme 1 (%37,7), en seyrek küme ise Küme 0 (%13,22) olmuştur. Küme 2 (%21,11), Küme 3 (%17,97) diğer iki kümedir. Tüm kümelerde kadınlar erkeklere oranla daha baskın taraftır. Yaş aralıkları açısından ise 16-25 yaş aralığındaki kullanıcılar diğer yaş aralıklarına göre tüm kümelerde baskın olan yaş aralığıdır. Kümelerin en fazla tercih ettiği konu başlıklarına bakıldığında Küme 0 en fazla Amerikan edebiyatı (%18,59); Küme 1 Türk dili ve edebiyatı (%43,45); Küme 2 Eğitim teorisi ve pratiği (%23,05); Küme 3 ise Türk dili ve edebiyatı (%31,78) olarak şekillenmiştir. Kullanıcı gruplarının kümelere dağılımında ise öğrenci grubu diğer gruplara oranla en baskın grup olmuştur. Kümelerin en fazla ödünç aldığı kitaplar incelenmiş ve Küme 0 ile Küme 2'nin; Küme 1 ile de Küme 3'ün kitap tercihleri bakımından birbirine benzediği gözlemlenmiştir. Küme 0 ve Küme 2 daha çok akademik ders kitapları ödünç alırken Küme 1 ve Küme 3 yoğunlukla edebi eserler ödünç almaktadır. Tüm kümeleme analizi sonucunda kullanıcıların en fazla yer aldığı kümeler tespit edilmiş ve her bir kullanıcı bir kümede olacak şekilde kümelerin kitap ödünç alma eğilimlerine göre kullanıcılara 5'er kitap tavsiyesi verilmiştir. Verilen kitap tavsiyelerinin performansı değerlendirilmiş ve Precision değeri 0,0208, Recall değeri 0,0162, F1 Score 0,1530 ve RMSE değeri 0,9895 olarak ölçülmüştür. Ortaya çıkan sonuçlar kitap tavsiye sisteminin k-means kümeleme analizi ile geliştirilmesinin düşük performans gösterdiğini ortaya koymuştur.

Bu araştırma, literatürde kütüphane kullanıcılarına yönelik kümeleme analizi yapan diğer çalışmalar ile (Tempelman-Kluit, Pearce, 2014; Glind, 2019; Fırat, 2019; Taşkın, 2020; Uçan, 2010) hem kullanılan değişkenler hem de kümeleme analizi sonunda uygulanan kitap öneri sistemi ile ayrılmaktadır. Diğer çalışmalar, kullanıcıları veri setindeki değişkenler dahilinde kümelere ayırmakta ve kümelerin ilişkin bilgiler paylaşılmaktadır. Kümeleme analizi sonrası herhangi bir kitap önerisi düşünülmemiştir. Ancak iş bu araştırma ile *k*-means kümeleme analizi sonrasında kullanıcıların en çok yer aldığı kümeye göre kitap önerileri verecek bir kitap öneri sistemi sunulmuştur. Bu bakımdan literatüre katkıda bulunmaktadır.

İkincil bir analiz olarak işbirlikçi filtreleme yöntemlerinden matris faktörizasyon algoritması olan SVD kullanılmıştır. Veri seti üzerinde yapılan değerlendirme sonucu SVD algoritmasının uzun bir analiz süreci gerektirmesi ve veri setinin seyreklik değerinin yüksek olması sebebiyle SVD algoritmasının uyarlanmış bir sürümü olan SVDs algoritması kullanılmıştır. Analiz boyunca kullanıcı-kitap etkileşim matrisi ısı haritası verilmiş, *k* latent faktör sayısı belirlenmiş ve kullanıcılar için kişiselleştirilmiş kitap tavsiye sistemi inşa edilmiştir. Ortaya çıkan modelde her bir kullanıcı için kendi geçmiş ödünç kitap tercihlerine göre 5'er kitap tavsiyesi verilmiştir. Sistemin performansı ölçmek için RMSE değeri hesaplanmış ve 0,018333 ile modelin yüksek performans gösterdiği belirlenmiştir. SVDs algoritmasının gösterdiği performans Satyal'ın (2019) çalışması ile benzerlik göstermektedir. Literatürde SVD algoritması ile yapılan çalışmalar (Satyal, 2019; Ahmed, Letta, 2023) kullanıcıların kitaplar için vermiş olduğu puanlar üzerinden bir kullanıcı-kitap matrisi oluşturmakta ve bu çalışma ile bu bakımdan ayrılmaktadır. İş bu çalışma kütüphane otomasyon sistemi üzerinde tutulan ödünç kayıtları ile kullanıcı-kitap matrisi oluşturmakta ve bu anlamda literatüre katkıda bulunmaktadır. Hali hazırda kütüphane otomasyon sistemlerinin çalışma mantığına bağlı olarak kullanıcılardan kitaplar için puanlama beklemeden kitap önerisi oluşturmaktadır.

Bu çalışma, gelişen bilgi teknolojileri ve dijitalleşme çağında kütüphanelerin rolünü yeniden tanımlamaya yönelik önemli bir katkı sunmaktadır. Bilgiye erişimin hızla dijital platformlara kaydığı günümüzde, kütüphaneler sadece fiziksel mekanlar olmaktan çıkarak kullanıcılarına doğru ve nitelikli bilgiyi sunmada kilit bir rol üstlenmeye devam etmektedir. Çalışma, dijital ortamda yaşanan bilgi bombardımanına karşı, kütüphanelerin

veri madenciliği ve kişiselleştirilmiş öneri sistemleri aracılığıyla kullanıcılarına daha etkili hizmet sunabileceğini göstermektedir.

Bu araştırma, Burdur Mehmet Akif Ersoy Üniversitesi Kütüphanesi'ndeki kullanıcıların ödünç alma verilerini analiz ederek, kullanıcıların demografik özelliklerine göre kitap tercihlerini ve okuma eğilimlerini ortaya koymuştur. Bu bağlamda, veri madenciliği tekniklerinin kütüphane hizmetlerinde uygulanabilirliği üzerine önemli bulgular sunmaktadır. K-means kümeleme analizi ve işbirlikçi filtreleme yöntemleri kullanılarak geliştirilen kitap öneri sistemleri, kütüphane kullanıcılarının ihtiyaçlarına ve ilgi alanlarına göre kişiselleştirilmiş öneriler sunma potansiyelini göstermiştir.

Araştırma soruları ile elde edilen bulguları toplamak gerekirse:

- S1: Kütüphane kullanıcı kümeleri nelerdir?
- C1: Veri setinden 'Bölüm', 'Cinsiyet', 'Grup', 'Fakülte', 'Konu' ve 'Yaş' değişkenleri dahil edilerek yapılan kümeleme analizi sonucunda Silhouette index ve Elbow yöntemi dikkate alınarak kullanıcılar 4 kümeye ayrılmıştır.
- S2: Kullanıcı kümelerinin özellikleri nedir?
- C2: 4 kümenin yoğunluk, cinsiyet, yaş, grup, konu dağılımları verilmiştir. Buna göre tüm kümelerde kadınlar, gençler ve öğrenciler baskın taraf olmuştur.
- S3: Kullanıcı kümelerinin en çok ödünç aldığı kitaplar hangileridir?
- C3: Küme 0'da Öğretim ilke ve yöntemleri, Eğitim bilimine giriş, Öğretim teknolojileri ve materyal tasarımı, Olasılıksız, Eğitim psikolojisi; Küme 1'de Suç ve ceza, Genel kimya, Aşk, Piraye, Çalıkuşu; Küme 2'de Öğretim ilke ve yöntemleri, Bilimsel araştırma yöntemleri, Eğitim psikolojisi, Eğitim bilimine giriş, İlköğretim programı (1-5. sınıflar); Küme 3'te Suç ve ceza, Genel kimya, Doğum ve kadın sağlığı hemşireliği, Piraye, Anatomi kitapları en fazla ödünç alınan kitaplar olmuştur.
- S4: Kullanıcıların demografik özellikleri kitap okuma davranışını etkiler mi?

- C4: Kullanıcıların cinsiyet ve yaş değişkenlerine göre en çok ödünç aldığı kitaplar incelendiğinde demografik değişkenlerin kitap tercihlerine ciddi bir etkisinin olmadığı sonucuna varılmıştır.
- S5: Kişiselleştirilmiş bir kitap öneri sistemi oluşturmak mümkün müdür?
- C5: İki farklı algoritma ile kullanıcı eğilimlerine yönelik kitap öneri sistemleri oluşturulmuştur.
- S6: Oluşturulan kitap öneri sistemlerinin performans değerleri nedir?
- C6: *k*-means kümeleme analizi ile oluşturulan kitap tavsiye modelinin Precision değeri 0,0208, Recall değeri 0,0162, F1 Score 0,1530 ve RMSE değeri 0,9895 olarak tespit edilmiştir. SVDs işbirlikçi filtreleme analizi ile oluşturulan kitap tavsiye sisteminin performans değeri RMSE=0,018333 olduğu tespit edilmiştir.

Sonuç olarak bu çalışma, kütüphanelerin veri madenciliği kullanarak nasıl daha etkin ve kullanıcı odaklı hizmetler sunabileceğine dair önemli bir model sunmakta ve literatürde bu alandaki boşluğu doldurmaktadır. Bu hem akademik kütüphaneler hem de genel kütüphane hizmetleri için yol gösterici nitelikte olup, gelecekteki araştırmalara ve uygulamalara ışık tutmaktadır. Gelecek çalışmalarda, *k*-means ve SVDs algoritmalarının performans değerleri göz önünde bulundurularak, eklerde sunulan Python kodları üzerinde geliştirilmiş ve kütüphane otomasyon sistemlerinin OPAC arayüzünde kullanılmak üzere bir modül geliştirilmesi önerilmektedir.

KAYNAKÇA

- AAAI. (t.y.). The first international conference on knowledge discovery and data mining (KDD 95). *Association for the Advancement of Artificial Intelligence*. 1 Kasım 2024 tarihinde <https://aaai.org/conference/kdd/kdd95/> adresinden erişildi.
- Aabø, S. (2005). The role and value of public libraries in the age of digital technologies. *Journal of Librarianship and Information Science*, 37(4), 205–211.
- Ackoff, R. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis*, 16, 3–9.
- Adams, S. (1964). MEDLARS and the Library Community. *Bulletin of the Medical Library Association*, 52(1), 171–177.
- Adomavicius, G. ve Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749.
- Agee, J. (2005). Collection evaluation: a foundation for collection development. *Collection Building*, 24(3), 92–95.
- Agrawal, R., Imieliński, T. ve Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2), 207–216.
- Agrawal, R. ve Srikant, R. (1994). Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases* içinde , VLDB '94 (ss. 487–499). San Francisco: Morgan Kaufmann Publishers Inc.
- Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M. ve Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10(1), 311.
- Ahmed, E. ve Letta, A. (2023). Book recommendation using collaborative filtering algorithm. *Applied Computational Intelligence and Soft Computing*, 2023, 1–12.
- Akalın, Ş. H. (Ed.). (2019). *Türkçe Sözlük*. Ankara: Türk Dil Kurumu.
- Akerkar, R. (2019). *Artificial intelligence for business*. SpringerBriefs in Business. Cham: Springer International Publishing.
- Akkaya, M. A. ve Odabaş, H. (2019). Önsöz. *Bilgi Merkezleri: Kütüphaneler - Arşivler - Müzeler* içinde . İstanbul: Hiperyayın.
- Akpınar, H. (2014). *Data Veri Madenciliği Veri Analizi*. İstanbul: Papatya Yayıncılık Eğitim.
- Alpaydın, E. (2017). *Yapay öğrenme*. İstanbul: Boğaziçi Üniversitesi Yayınevi.
- Amazon. (t.y.). Otomatik akıl yürütme nedir? *Amazon*. 1 Mayıs 2024 tarihinde <https://aws.amazon.com/tr/what-is/automated-reasoning/> adresinden erişildi.
- Anderson, D. R., Sweeney, D. J. ve Williams, T. A. (1998). Statistics. *Britannica Academic*. 1 Mayıs 2023 tarihinde <https://academic.eb.com/levels/collegiate/article/statistics/108592> adresinden erişildi.

- ANSI. (2015). *Information retrieval (Z39.50): Application service definition and protocol specification*. Maryland: National Information Standards Organization.
- ANSI. (2016). *Information Interchange Format*. Maryland: National Information Standards Organization. doi:10.3789/ansi.niso.z39.2-1994R2016
- Ardolino, M., Rapaccini, M., Saccani, N., Gaiardelli, P., Crespi, G. ve Ruggeri, C. (2018). The role of digital technologies for the service transformation of industrial companies. *International Journal of Production Research*, 56(6), 2116–2132.
- Asimov, I. (1986). *Kurtarıcı*. İstanbul: Altın Kitaplar.
- Asimov, I. (2016). *Ben, robot*. İstanbul: İthaki Yayınları.
- Aslan, H. (2022). LIBRA kütüphane yönetim bilgi sistemi. 7 Mayıs 2023 tarihinde http://www.huryasa.com/brosur_libra.pdf adresinden erişildi.
- Association American Library. (2016). Academic libraries. 2 Ağustos 2023 tarihinde <http://www.ala.org/educationcareers/libcareers/type/academic> adresinden erişildi.
- Austrian, G. D. (1982). *Herman Hollerith: Forgotten giant of information processing*. New York: Columbia University Press.
- Aven, T. (2013). A conceptual framework for linking risk and the elements of the data–information–knowledge–wisdom (DIKW) hierarchy. *Reliability Engineering & System Safety*, 111, 30–36.
- Avram, H. D. (1975). *MARC: Its history and implications*. Washington: Library of Congress.
- Banerjee, K. (1998). Is data mining right for your library? *Computers in Libraries*, 18(10), 28–31.
- Barnes, R. (2014). Musaların tavuk kümesinde dünyadan uzak yaşayan kitap kurtları: Antik İskenderiye Kütüphanesi. R. Macleod (Ed.), *İskenderiye Kütüphanesi Antik Dünya'nın Öğrenim Merkezi* içinde (ss. 83–103). Ankara: Dost Kitabevi.
- Bartlett, J. A. (2021). *Knowledge Management: A Practical Guide for Librarians*. Practical Guides for Librarians. Lanham: Rowman & Littlefield Publishers.
- Başkarada, S. ve Koronios, A. (2013). Data, information, knowledge, wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and its Quality Dimension. *Australasian Journal of Information Systems*, 18(1), 5–24.
- Basu, C., Hirsh, H. ve Cohen, W. (1998). Recommendation as classification: Using social and content-based information in recommendation. *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence* içinde , AAAI '98/IAAI '98 (ss. 714–720). USA: American Association for Artificial Intelligence.
- Bawden, D. ve Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180–191.
- Bawden, D. ve Robinson, L. (2020). Information overload: An introduction. *The Oxford Encyclopedia of Political Decision Making* içinde . Oxford: Oxford University Press.
- Bayes, M. ve Price, M. (1763). An essay towards solving a problem in the doctrine of

- chances. *Philosophical Transactions (1683-1775)*, 53, 370–418.
- Baysal, J. (1987). *Kütüphanecilik alanında yeni kavramlar araçlar yöntemler* (2. bs.). İstanbul: İstanbul Üniversitesi Edebiyat Fakültesi Yayınları.
- Becker, J. (1964). Circulation and the Computer. *ALA Bulletin*, 58(11), 1007–1010.
- Bell, S. (2019). Academic librarians' c-word problem : From the bell tower. *Library Journal*.
- Bellinger, G., Castro, D. ve Mills, A. (2004). Data, Information, Knowledge, and Wisdom. 16 Şubat 2023 tarihinde <http://www.systems-thinking.org/dikw/dikw.htm> adresinden erişildi.
- Bhatia, P. (2019). *Data mining and data warehousing*. Cambridge: Cambridge University Press.
- Bhavana, P. ve Padmanabhan, V. (2021). Matrix factorization of large scale data using multistage matrix factorization. *Applied Intelligence*, 51(6), 4016–4028.
- Bigus, J. P. (1996). *Data mining with neural networks: Solving business problems from application development to decision support*. Indiana: The McGraw-Hill Companies.
- Bimbrow, K. (2015). Autonomous cars: Past, present and future a review of the developments in the last century, the present scenario and the expected future of autonomous vehicle technology. *2015 12th International Conference on Informatics in Control, Automation and Robotics (ICINCO)* içinde (C. 01, ss. 191–198).
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Singapore: Springer Science+Business Media, LLC.
- Boser, B. E., Guyon, I. M. ve Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory* içinde (ss. 144–152). New York, NY, USA: ACM.
- Boter, J. ve Wedel, M. (2005). User categorization of public library collections. *Library & Information Science Research*, 27(2), 190–202.
- Bozkurt, A. (2014). Ağ Toplumu ve Bilgi. *Türk Kütüphaneciliği*, 28(4), 510–525.
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E. ve Yergeau, F. (2008). Extensible markup language (XML) 1.0 (fifth edition). 12 Mart 2023 tarihinde <https://www.w3.org/TR/REC-xml/#sec-terminology> adresinden erişildi.
- Breeding, M. (2009). Next generation library automation: Its impact on the serials community. *The Serials Librarian*, 56(1–4), 55–64.
- Breeding, M. (Ed.). (2016). *Library technology buying strategies*. Chicago: American Library Association.
- Breeding, M. (2023). A directory of libraries throughout the world. 08 Mayıs 2023 tarihinde <https://librarytechnology.org/libraries/> adresinden erişildi.
- Breiman, L., Friedman, J. H., Olshen, R. A. ve Stone, C. J. (1984). *Classification and regression trees*. Florida: CRC Press.
- Brin, S. ve Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine.

- Computer Networks and ISDN Systems*, 30(1–7), 107–117.
- Buckland, M. K. (2017). Library technology in the next 20 years. *Library Hi Tech*, 35(1), 5–10.
- Burdur Mehmet Akif Ersoy Üniversitesi. (2024). Kalite Bilgi Yönetim Sistemi. 1 Mayıs 2024 tarihinde <https://kbys.mehmetakif.edu.tr/> adresinden erişildi.
- Burke, P. (2017). *Bilgi Tarihi Nedir?* İstanbul: Islık Yayınları.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2015). *Bilimsel araştırma yöntemleri* (19. bs.). Ankara: Pegem Akademi.
- Byrne, D. J. (1998). *MARC Manual: Understanding and using MARC records* (2. bs.). London: Libraries Unlimited.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. ve Zanasi, A. (1998). *Discovering data mining: From concept to implementation*. New Jersey: Prentice Hall PTR.
- Çağıltay, N. E. (2010). *İş zekâsı ve veri ambarı sistemleri*. ODTÜ Yayıncılık.
- Carvalho, J. de ve Cordeiro, M. I. (2002). XML and bibliographic data: The TVS (transport, validation and services) model. *68th IFLA Council and General Conference* içinde .
- Casson, L. (2001). *Libraries in the ancient world*. New Haven: Yale University Press.
- Çelik, H. (1972). Otomasyon-dokümantasyon ve biz. *Türk Kütüphaneciler Derneği Bülteni*, 21(4), 229–242.
- Çelik, M. (2004). *Bilgi ve hikmet Enformasyon Toplumunun Belleği*. İstanbul: Kaknüs Yayınları.
- CERN. (t.y.). A short history of the web. CERN. 15 Şubat 2024 tarihinde <https://press.cern/science/computing/birth-web/short-history-web> adresinden erişildi.
- Çetin, N. M. (2011). *Eski Arap şiiri*. İstanbul: Alfa Basım Yayım Dağıtım Ltd. Şti.
- Chen, C. C. ve Chen, A. P. (2007). Using data mining technology to provide a recommendation service in the digital library. *Electronic Library*, 25(6), 711–724.
- Chen, C. L. P. ve Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314–347.
- Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S. ve Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of Computer Science*, 7(2), 157–164.
- Chengwang, L. (2010). Singular value decomposition in active monitoring data analysis. *Handbook of Geophysical Exploration: Seismic Exploration* içinde (C. 40, ss. 421–430). Oxford: Elsevier.
- Clifton, C. (2009). Data mining. *Encyclopedia Britannica*. 1 Aralık 2023 tarihinde <https://academic.eb.com/levels/collegiate/article/data-mining/437561> adresinden erişildi.
- Coenen, F. (2011). Data mining: past, present and future. *The Knowledge Engineering Review*, 26(1), 25–29.

- Committee on Technical Standards for Library Automation. (1974). Standards for Library Automation and ISAD's Committee on Technical Standards for Library Automation (TESLA). *Information Technology and Libraries*, 7(2 SE-Articles), 126–138.
- Copeland, B. J. (1998). Artificial intelligence (AI). *Britannica Academic*. 12 Şubat 2024 tarihinde <https://academic.eb.com/levels/collegiate/article/artificial-intelligence/9711#article-contributors> adresinden erişildi.
- Council on Library Resources. (1973). *Council on Library Resources Inc: Seventeenth annual report*. Washington.
- Cox, M. ve Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. *Proceedings. Visualization '97 (Cat. No. 97CB36155)* içinde (ss. 235–244).
- Crossley, J. N. ve Henry, A. S. (1990). Thus spake al-Khwārizmī: A translation of the text of Cambridge University Library Ms. Ii.vi.5. *Historia Mathematica*, 17(2), 103–131.
- Davenport, T. (2018). *Big data @ work*. İstanbul: Türk Hava Yolları.
- Davenport, T. H. (2006). Competing on analytics. *Harvard business review*, 84(1), 98–107,134.
- Davenport, T. H. ve Bean, R. (2023). *Data and analytics leadership annual executive survey 2023*. 1 Aralık 2023 tarihinde https://www.newvantage.com/_files/ugd/e5361a_247885043758499ba090f7a5f510cf7c.pdf adresinden erişildi.
- Davenport, T. H. ve Prusak, L. (2000). *İş Dünyasında Bilgi Yönetimi: Kuruluşlar Elleriindeki Bilgiyi Nasıl Yönetirler*. İstanbul: Rota Yayın Yapım.
- Davies, D. L. ve Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224–227.
- Decker, S., Melnik, S., Harmelen, F. van, Fensel, D., Klein, M., Broekstra, J., ... Horrocks, I. (2000). The Semantic Web: the roles of XML and RDF. *IEEE Internet Computing*, 4(5), 63–73.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. ve Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407.
- Dempster, A. P., Laird, N. M. ve Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.
- Dennett, D. C. (2004). Can Machines Think? C. Teuscher (Ed.), *Alan Turing: Life and legacy of a great thinker* içinde (ss. 295–316). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Dewey, H. (1959). Punched card catalogs- Theory and technique: I. history. *American Documentation*, 10(1), 36.
- Diebold, F. X. (2012). *On the origin(s) and development of "big data": The phenomenon, the term, and the discipline*. PIER Working Paper No. 12-037.

- Dikeç, M. (1988). *Bilgisayara dayalı kütüphane ve bilgi sistemleri: Bilkent Üniversitesi kütüphanesi için alternatif bir sistem seçimi ve değerlendirilmesi*. Hacettepe Üniversitesi.
- Djuraskovic, O. (2023). Big data statistics 2023: How much data is in the world? *First Site Guide*.
- Dunham, M. H. (2003). *Data mining: Introductory and advanced topics*. New Jersey: Prentice Hall/Pearson.
- Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1), 95–104.
- EATCS. (2003). Gödel Prize - 2003. *European Association for Theoretical Computer Science*. 10 Aralık 2023 tarihinde <https://eatcs.org/index.php/component/content/article/505> adresinden erişildi.
- Eliot, T. S. (1934). *The Rock*. Faber & Faber.
- Encyclopædia Britannica. (t.y.). Natural language processing (NLP). *Britannica Academic*. 1 Nisan 2024 tarihinde <https://academic.eb.com/levels/collegiate/article/natural-language-processing/636044> adresinden erişildi.
- Encyclopædia Britannica. (2020). Automation. *Britannica Academic*. 15 Aralık 2023 tarihinde <https://academic.eb.com/levels/collegiate/article/automation/109393> adresinden erişildi.
- Encyclopædia Britannica. (2023). Henry Ford on mass production. *Britannica Academic*. 14 Aralık 2023 tarihinde <https://academic.eb.com/levels/collegiate/article/Henry-Ford-on-mass-production/635570> adresinden erişildi.
- Erdoğan, P. L. (2011). *Bir kütüphanenin hikayesi : Bilkent Üniversitesi Kütüphanesi'nin yirmi yılı (1986-2005), kurumsal ve kişisel anılar*. İstanbul: Hiperlink Yayınları.
- Erünsal, İ. E. (2018). *Orta çağ İslam Dünyasında kitap ve kütüphane*. İstanbul: TİMAŞ Yayınları.
- Fayyad, U. M., Piatetsky-Shapiro, G. ve Smyth, P. (1996). From data mining to knowledge discovery: An overview. *AI Magazine*, 17(3), 37–54.
- Fazlıoğlu, İ. (1997). Hârizmî, Muhammed b. Mûsâ. *TDV İslâm Ansiklopedisi*. <https://islamansiklopedisi.org.tr/harizmi-muhammed-b-musa> adresinden erişildi.
- Fırat, İ. (2019). *Veri madenciliği teknikleri uygulayarak kütüphane otomasyonundan bilgi keşfi elde edilmesi*. Erzincan Binali Yıldırım Üniversitesi.
- Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21(3), 768–769.
- Forsythe, G. E. ve Malcolm, M. A. (1977). *Computer methods for mathematical computations*. New Jersey: Prentice-Hall, Inc.
- Freund, Y. ve Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. P. Vitányi (Ed.), *EuroCOLT 1995 içinde* (ss. 23–37). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:https://doi.org/10.1007/3-540-59119-2_166

- Freund, Y. ve Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference* içinde , ICML'96 (ss. 148–156). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Freund, Y. ve Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139.
- Frické, M. (2008). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science*, 35(2), 131–142.
- Friedman, H. P. ve Rubin, J. (1967). On some invariant criteria for grouping data. *Journal of the American Statistical Association*, 62(320), 1159.
- Frontz, S. J. ve Alvey, C. (1982). Academic tol column: The integrated automated library system. *Art Documentation: Journal of the Art Libraries Society of North America*, 1(6), 196–200.
- Gaillard, M. (2017). CERN data centre passes the 200-petabyte milestone. *CERN*. 1 Aralık 2023 tarihinde <https://home.cern/news/news/computing/cern-data-centre-passes-200-petabyte-milestone> adresinden erişildi.
- Galabova, S., Trencheva, T. ve Trenchev, I. (2009). Standardized structure of electronic records for information exchange. *Third International Scientific Conference* içinde (ss. 230–233). Blagoevgrad: South-West University.
- Gandomi, A. ve Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
- Garrett, J. (1999). Redefining order in the German library, 1775-1825. *Eighteenth-Century Studies*, 33(1), 103–123.
- Gelfand, M. A. (1968). *University libraries for developing countries*. Switzerland: UNESCO.
- Geyer-Schulz, A., Neumann, A. ve Thede, A. (2003). An architecture for behavior-based library recommender systems. *Information Technology and Libraries*, 22(4), 165–174.
- Giannella, C., Han, J., Pei, J., Yan, X. ve Yu, P. S. (2004). Mining frequent patterns in data streams at multiple time granularities. H. Kargupta, A. Joshi, K. Sivakumar ve Y. Yesha (Ed.), *Data mining: Next generation challenges and future directions* içinde (ss. 105–124). California: AAAI Press/ The MIT Press.
- Glind, Y. M. van de. (2019). *Library user segmentation by transactional log data using k-means*. Tilburg University.
- Goes, P. B. (2014). Editor's comments: Big data and is research. *MIS Quarterly*, 38(3), iii–viii.
- Goldberg, D., Nichols, D., Oki, B. M. ve Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70.
- Golub, G. H. ve Loan, C. F. Van. (2013). *Matrix computations* (Fourth Edi.). Baltimore: The John Hopkins University Press.

- Gomez-Urbe, C. A. ve Hunt, N. (2016). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 6(4), 1–19.
- Goodfellow, I., Bengio, Y. ve Courville, A. (2016). *Deep Learning*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts: The MIT Press.
- Görz, G. ve Nebel, B. (2002). *Yapay zekâ*. İstanbul: İnkilap Kitabevi.
- Greider, E. M. (1956). Kütüphane üniversitenin kalbidir. *Türk Kütüphaneciler Derneği Bülteni*, 5(2), 93–100.
- Gürsoy, E. (1975). *Hacettepe Üniversitesi kütüphanelerinde bulunan süreli yayınlar kataloğunun bilgisayar katkısıyla geliştirilmesi*. Hacettepe Üniversitesi.
- Gürsoy, E. (1978). *Gerçek zamanlı bilgisayarla kütüphane dolaşım izleme sistemi*. Hacettepe Üniversitesi.
- Haider, S., Francis, F. C., Foscett, D. J. ve Estabrook, L. S. (2022). Library. *Encyclopedia Britannica*. 17 Şubat 2023 tarihinde <https://www.britannica.com/topic/library> adresinden erişildi.
- Halpin, T. ve Morgan, T. (2008). *Information modeling and relational databases*. Burlington: Elsevier.
- Han, J., Kamber, M. ve Pei, J. (2012). *Data mining: Concepts and techniques* (Third Edit.). Massachusetts: Elsevier.
- Han, J., Pei, J. ve Tong, H. (2022). *Data mining concepts and techniques* (Fourth Edi.). Cambridge: Morgan Kaufmann.
- Hand, D. J., Mannila, H. ve Smyth, P. (2001). *Principles of data mining*. Cambridge, Massachusetts: Massachusetts Institute of Technology.
- Harari, Y. N. (2018). Our nonconscious future. *Britannica Academic*. 1 Aralık 2023 tarihinde <https://academic.eb.com/levels/collegiate/article/Our-Nonconscious-Future/630820> adresinden erişildi.
- Haravu, L. J. (2006). Standards in library automation and networking. 8 Ekim 2023 tarihinde <https://drtc.isibang.ac.in/ldl/handle/1849/291> adresinden erişildi.
- Hastie, T., Tibshirani, R. ve Friedman, J. (2009). *The elements of statistical learning*. Springer Series in Statistics. New York: Springer New York. doi:10.1007/978-0-387-84858-7
- Hey, J. (2004). The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link. 1 Mayıs 2023 tarihinde <https://www.jonohey.com/files/DIKW-chain-Hey-2004.pdf> adresinden erişildi.
- Himmelfarb, G. (1997). Revolution in the Library. *The American Scholar*, 66(2), 197–204.
- Hopkins, J. (1973). The Ohio college library center. *Library Resources and Technical Services*, 17(3), 308–319.
- Hopkinson, A. (1984). International access to bibliographic data: Marc and marc-related activities. *Journal of Documentation*, 40(1), 13–24.

- Hosch, W. L. (2009). Robotics. *Britannica Academic*. 1 Ağustos 2023 tarihinde <https://academic.eb.com/levels/collegiate/article/robotics/473204> adresinden erişildi.
- IBM. (2011). *IBM SPSS Modeler CRISP-DM guide*. IBM Corporation. https://inseaddataanalytics.github.io/INSEADAnalytics/CRISP_DM.pdf adresinden erişildi.
- IFLA. (t.y.-a). Library types. 23 Ocak 2023 tarihinde <https://librarymap.ifla.org/data-glossary/library> adresinden erişildi.
- IFLA. (t.y.-b). ISO 2709. *IFLA*. 23 Ocak 2023 tarihinde <https://www.ifla.org/references/best-practice-for-national-bibliographic-agencies-in-a-digital-age/resource-description-and-standards/metadata-transmission-standards/specialist-bibliographic-transmission-standards/iso-2709/> adresinden erişildi.
- IFLA. (2022). Library map of the world. *IFLA*. 27 Ocak 2023 tarihinde <https://librarymap.ifla.org/map> adresinden erişildi.
- International Organization for Standardization. (1986). Information processing - Text and office systems - Standard Generalized Markup Language (SGML). (*ISO Standard No. 8879*).
- International Organization for Standardization. (1998). Information and documentation - Information retrieval (Z39.50) - Application service definition and protocol specification. *ISO Standart No. 23950*.
- International Organization for Standardization. (2008). Information and documentation - Format for information exchange. (*ISO Standard No. 2709*).
- Işık, M. (2013). *Google pagerank & trustrank*. İstanbul: İkinci Adam Yayınları.
- Jayaraman, S. (2023). How big is Big? 85+ big data statistics you should know in 2023. G2. 23 Ocak 2023 tarihinde <https://www.g2.com/articles/big-data-statistics> adresinden erişildi.
- Jin, R. ve Agrawal, G. (2007). Frequent pattern mining in data streams. C. C. Aggarwal (Ed.), *Data Streams içinde*, Advances in Database Systems (ss. 61–84). Boston, MA: Springer US.
- Johnson, B. C. (2001). XML and MARC: Which is “Right”? *Cataloging & Classification Quarterly*, 32(1), 81–90.
- Johnson, D. G. ve Verdicchio, M. (2017). AI anxiety. *Journal of the Association for Information Science and Technology*, 68(9), 2267–2270. doi:10.1002/asi.23867
- Kantardzic, M. (2019). *Data mining: Concepts, models, methods, and algorithms*. New Jersey: Wiley.
- Karahan Adalı, G. (2016). Birliktelik kuralları ile veri madenciliğinde bir pazarlama uygulaması gerçekleştirilmesi. M. E. Balaban ve E. Kartal (Ed.), *R ile veri madenciliği uygulamaları içinde* (ss. 221–242). İstanbul: Çağlayan Kitabevi.
- Kargupta, H., Joshi, A., Sivakumar, K. ve Yesha, Y. (Ed.). (2004). *Data mining: Next generation challenges and future directions*. AAAI Press / The MIT Press.

- Kendrick, K. M. (1998). Intelligent perception. *Applied Animal Behaviour Science*, 57(3), 213–231.
- Keseroğlu, H. S. (2019). Kütüphane Üstüne Bir Deneme. M. A. Akkaya ve H. Odabaş (Ed.), *Bilgi Merkezleri: Kütüphaneler - Arşivler - Müzeler*2 içinde (ss. 11–19). İstanbul: Hiperyayın.
- Khare, R. ve Rifkin, A. (1997). XML: a door to automated Web applications. *IEEE Internet Computing*, 1(4), 78–87.
- Kilgour, F. G., Long, P. L., Landgraf, A. L. ve Wyckoff, J. A. (1972). The shared cataloging system of The Ohio College Library Center. *Journal of Library Automation*, 5(3).
- King, G. (2016). Preface: Big data is not about the data! R. M. Alvarez (Ed.), *Computational Social Science Discovery and Prediction* içinde (ss. vii–x). New York: Cambridge University Press.
- King, G. W., Edmundson, H. P., Flood, M. M., Kochen, M., Libby, richard L., Swanson, D. R. ve Wylly, A. (1963). *Automation and the Library of Congress*. Washington: Library of Congress.
- Kitchin, R. ve McArdle, G. (2016). What makes big data, big data? exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1). doi:10.1177/2053951716631130
- Konya, Ü. (2019). Milli kütüphane. M. A. Akkaya ve H. Odabaş (Ed.), *Bilgi Merkezleri: Kütüphaneler - Arşivler - Müzeler* içinde (ss. 20–40). İstanbul: Hiperyayın.
- Koren, Y., Bell, R. ve Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37.
- Krass, U., Allen, M., White, E., Cybelle Ferrari, A., Brigant, A., Prucková, L., ... Ezenel, E. (2022). IFLA-UNESCO halk kütüphanesi bildirgesi 2022. *IFLA*. IFLA. 13 Şubat 2023 tarihinde <https://repository.ifla.org/handle/123456789/2159> adresinden erişildi.
- Kuroiwa, T., Sarcon, A., Ibara, T., Yamada, E., Yamamoto, A., Tsukamoto, K. ve Fujita, K. (2023). The Potential of chatGPT as a self-diagnostic tool in common orthopedic diseases: Exploratory study. *Journal of Medical Internet Research*, 25, e47621.
- Külcü, Ö. (2019). *Bilgi kuramı ve bilgi yönetimi : Kuramsal bilginin oluşumu ve toplumsal bilgiye dönüşümü*. İstanbul: Hiperyayın.
- Kütüphaneler ve Yayınlar Genel Müdürlüğü. (2021). 2021 Yılı İstatistikleri. *Kültür ve Turizm Bakanlığı*. 17 Şubat 2023 tarihinde <https://kygm.ktb.gov.tr/TR-310238/2021-yili-istatistikleri.html> adresinden erişildi.
- Lambert, F. (2023). Elon Musk says Tesla is now working on ‘final piece of the FSD AI puzzle’. *Electrek*. 1 Şubat 2024 tarihinde <https://electrek.co/2023/08/02/elon-musk-tesla-working-final-piece-fsd-ai-puzzle/> adresinden erişildi.
- Laney, D. (2001). *3D data management: controlling data volume, velocity, and variety*. 1 Şubat 2023 tarihinde <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...> adresinden erişildi.
- Langville, A. N. ve Meyer, C. D. (2006). *Google’s PageRank and beyond*. Princeton

University Press.

- Larose, D. T. (2005). *Discovering knowledge in data: an introduction to data mining*. New Jersey: John Wiley & Sons.
- Larose, D. T. ve Larose, C. D. (2015). *Data mining and predictive analytics* (Second Edi.). Newark, UNITED STATES: John Wiley & Sons, Incorporated.
- Larsen, R. M. (2012). Propack. <http://sun.stanford.edu/~rmunk/PROPACK/> adresinden erişildi.
- Lee, K.-F. (2018). *Yapay zekâ ve yeni dünya düzeni : Çin ve Silikon Vadisi*. İstanbul: Optimist Yayın Grubu.
- Lee, K.-F. ve Qiufan, C. (2022). *Yapay zekâ 2041: Geleceğimiz için on vizyon*. İstanbul: HYV Havacılık Yönetim Çözümleri.
- Leupp, H. L. (1924). The library the heart of the university. *Bulletin of the American Library Association*, 18(4A), 193–197.
- Levine-Clark, M. ve Carter, T. M. (Ed.). (2013). *ALA glossary of library and information science* (4. bs.). Chicago: American Library Association.
- LHCb. (2023). LHCb begins using unique approach to process collision data in real-time. CERN. 10 Ocak 2023 tarihinde <https://home.cern/news/news/experiments/lhcb-begins-using-unique-approach-process-collision-data-real-time> adresinden erişildi.
- Li, X., Ng, S.-K. ve Wang, J. T. L. (2014). *Biological data mining and its applications in healthcare*. Science, Engineering, and Biology Informatics. Singapore: World Scientific Publishing Co. Pte. Ltd.
- LIBRA. (t.y.). Yazılım Hakkında. 7 Mayıs 2023 tarihinde <http://libra.huryasa.com/> adresinden erişildi.
- Library of Congress. (1969). *MARC manuals used by the Library of Congress*. Chicago: American Library Association.
- Library of Congress. (1996). The MARC 21 formats: Background and principles. 8 Ocak 2023 tarihinde <https://www.loc.gov/marc/96principl.html> adresinden erişildi.
- Library of Congress. (2000). MARC 21 specifications for record structure, character sets, and exchange media. 8 Ocak 2023 tarihinde <https://www.loc.gov/marc/specifications/specrecstruc.html> adresinden erişildi.
- Library of Congress. (2008). MARC in XML. 12 Mayıs 2023 tarihinde <https://www.loc.gov/marc/marcxml.html> adresinden erişildi.
- Library of Congress. (2015). Z39.50. 8 Ekim 2023 tarihinde <https://www.loc.gov/z3950/agency/> adresinden erişildi.
- Library of Congress. (2022a). Library of Congress Launches Effort to Transform Collections Management and Access. 12 Mayıs 2023 tarihinde <https://newsroom.loc.gov/news/library-of-congress-launches-effort-to-transform-collections-management-and-access/s/c432d3c2-780b-4bfe-9123-bbb6c25631bc> adresinden erişildi.
- Library of Congress. (2022b). MARC XML design considerations. 12 Haziran 2023 tarihinde <https://www.loc.gov/standards/marcxml/marcxml-design.html> adresinden

erişildi.

- Liu, J., Wang, Z., Peng, Z., Cui, J.-H. ve Fiondella, L. (2014). Suave: Swarm underwater autonomous vehicle localization. *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications* içinde (ss. 64–72).
- Liu, Z., Zhang, Y., Yu, X. ve Yuan, C. (2016). Unmanned surface vehicles: An overview of developments and challenges. *Annual Reviews in Control*, 41, 71–93.
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137.
- Locke, J. (2018). *An Essay Concerning Human Understanding*. Charles River Editors.
- Lor, P. J. (1997). *Guidelines for legislation for national library services*. Paris: UNESCO.
- Luccioni, A. S., Jernite, Y. ve Strubell, E. (2023). Power hungry processing: Watts driving the cost of AI deployment? *ArXiv*, abs/2311.1(16863).
- Lynch, C. A. (1997). The Z39.50 information retrieval standard: Part I: A strategic view of its past, present and future. *D-Lib Magazine*, 3(4), 16–27.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. L. M. Le Cam ve J. Neyman (Ed.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* içinde (C. 1, ss. 281–297). Oakland: University of California.
- Martin, L. A. (1968). The changes ahead. *Library Journal*, 93(January-December), 711–716.
- Mashey, J. (1998). Big data and the next wave of infrastress. 15 Haziran 2023 tarihinde <https://web.stanford.edu/class/ee380/9798win/lect08.html> adresinden erişildi.
- McCallum, S. H. (2002). MARC: keystone for library automation. *IEEE Annals of the History of Computing*, 24(2), 34–49.
- McCarthy, J. (2006). The Dartmouth Workshop--as planned and as it happened. 12 Haziran 2023 tarihinde <https://www-formal.stanford.edu/jmc/slides/dartmouth/dartmouth/node1.html> adresinden erişildi.
- McCarthy, J., Minsky, M., Rochester, N. ve Shannon, C. (1955). A proposal for the Dartmouth summer research project on artificial intelligence.
- McIntosh, T. R., Liu, T., Susnjak, T., Watters, P., Ng, A. ve Halgamuge, M. N. (2023). A culturally sensitive test to evaluate nuanced GPT hallucination. *IEEE Transactions on Artificial Intelligence*, 1–13.
- McKee, G. T. (2006). The maturing discipline of robotics. *International Journal of Engineering Education*, 22(4), 692–701.
- Mehar, A. M., Matawie, K. ve Maeder, A. (2013). Determining an optimal value of K in K-means clustering. *2013 IEEE International Conference on Bioinformatics and Biomedicine* içinde (ss. 51–55). IEEE.
- Merliana, N. P. E., Ernawati ve Santoso, A. J. (2015). Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering. *Seminar Nasional Multi Disiplin Ilmu & Call For Papers* içinde . Universitas Stikubank.

- Miller, D. R. (2000). XML and MARC: A choice or replacement? 12 Mayıs 2023 tarihinde <https://lane.stanford.edu/elane/public/L210158/ALACHicago2000.html> adresinden erişildi.
- Miller, E. W. ve Hodges, B. J. (1971). Shawnee mission's on-line cataloging system. *Journal of Library Automation*, 4(1), 13–26.
- Miller, W. L. (1971). A probabilistic search strategy for medlars. *Journal of Documentation*, 27(4), 254–266.
- Milli Kütüphane. (t.y.). Tarihçe. *Milli Kütüphane*. 8 Şubat 2023 tarihinde <http://www.millikutuphane.gov.tr/page/Tarihce> adresinden erişildi.
- Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
- Mittler, E. (2016). The Library as History: Library History Research after the Cultural Turn. *Quaerendo*, 46(2–3), 222–240. d
- Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, 38(8).
- Morton, K. D. (1986). The marc formats: An overview. *The American Archivist*, 49(1), 21–30.
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. London: The MIT Press.
- Naisbitt, J. (1984). *Megatrends: Ten new directions transforming our lives*. New York: Warner Book.
- Nalinipriya, G., Geetha, M., Cristin, R. ve Maram, B. (2021). Biomedical data mining for improved clinical diagnosis. *Artificial Intelligence in Data Mining* içinde (ss. 155–176). London: Academic Press.
- National Information Standards Organization. (2002). *Z39.50 — A primer on the protocol*. Maryland: NISO Press.
- Naughton, J. (2023). Why AI is a disaster for the climate. *The Guardian*. 1 Eylül 2024 tarihinde <https://www.theguardian.com/commentisfree/2023/dec/23/ai-chat-gpt-environmental-impact-energy-carbon-intensive-technology> adresinden erişildi.
- Neal, R. M. ve Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. *Learning in Graphical Models* içinde (ss. 355–368). Dordrecht: Springer Netherlands.
- Nicholson, S. (2003). The bibliomining process: Data warehousing and data mining for library decision making. *Information Technology and Libraries*, 22(4), 146–151.
- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164–182.
- Nicholson, S., Hwang, S., Keezer, P. ve O'Neill, E. T. (2003). The bibliomining process: Data warehousing and data mining for libraries. *Proceedings of the American Society for Information Science and Technology*, 40(1), 478–479.
- Nicholson, S. ve Stanton, J. (2004). Gaining strategic advantage through bibliomining : Data mining for management decisions in corporate, special, digital and, traditional

- libraries. H. R. Nemati ve C. D. Barko (Ed.), *Organizational data mining : Leveraging enterprise data resources for optimal performance* içinde (ss. 247–262). Hershey: Idea Group Publishing.
- Nicholson, S. ve Stanton, J. (2009). Bibliomining for library decision-making. J. Wang (Ed.), *Encyclopedia of Data Warehousing and Mining* içinde (Second Edi., ss. 153–159). Hershey: IGI Global.
- Oakleaf, M. (2010). *Value of academic libraries : A comprehensive research review and report*. Chicago: Association of College and Research Libraries.
- OCLC. (2023). Inside WorldCat. OCLC. 5 Ekim 2023 tarihinde <https://www.oclc.org/en/worldcat/inside-worldcat.html> adresinden erişildi.
- Odabaş, H. (2007). Bilgi Yönetimi. H. Odabaş ve H. Anameriç (Ed.), *Bilgi...* içinde (ss. 99–108). Ankara: Referans Yayınevi.
- Ohlhorst, F. J. (2012). *Big data analytics: Turning big data into big money*. Newark, UNITED STATES: John Wiley & Sons, Incorporated.
- Ortaylı, İ. (2006). İskenderiye Kütüphanesi. *2Türk Kütüphaneciliği*, 20(1), 85–88.
- Ötüken, A. (1955). Milli Kütüphane nasıl kuruldu? *Türk Kütüphaneciler Derneği Bülteni*, 4(1), 1–40.
- Öztemel, E. (2020). *Yapay sinir ağları*. İstanbul: Papatya Yayıncılık Eğitim.
- Page, L., Brin, S., Motwani, R. ve Winograd, T. (1998). The PageRank citation ranking : Bringing order to the web. 12 Aralık 2023 tarihinde <https://www.cis.upenn.edu/~mkearns/teaching/NetworkedLife/pagerank.pdf> adresinden erişildi.
- Papadakis, P. (2013). Terrain traversability analysis methods for unmanned ground vehicles: A survey. *Engineering Applications of Artificial Intelligence*, 26(4), 1373–1385.
- Parthasarathy, S., Ghoting, A. ve Otey, M. E. (2007). A survey of distributed mining of data streams. C. C. Aggarwal (Ed.), *Data streams* içinde , *Advances in Database Systems* (ss. 189–307). Boston, MA: Springer US.
- Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., ... Dean, J. (2021). Carbon emissions and large neural network training. *arXiv, abs/2104.10350*, 1–22.
- Peek, N., Combi, C. ve Tucker, A. (2009). Biomedical data mining. *Methods of Information in Medicine*, 48(03), 225–228.
- Piatetsky-Shapiro, G. (1990). Knowledge discovery in real databases: A report on the IJCAI-89 workshop. *AI Magazine*, 11(5), 68–70.
- Piatetsky-Shapiro, G., Brachman, R., Khabaza, T., Kloesgen, W. ve Simoudis, E. (1996). An overview of issues in developing industrial data mining and knowledge discovery applications. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* içinde (ss. 89–95). Oregon: The Association for the Advancement of Artificial Intelligence.
- Polat, C. (2019). Üniversite kütüphaneleri. M. A. Akkaya ve H. Odabaş (Ed.), *Bilgi*

Merkezleri: Kütüphaneler - Arşivler - Müzeler içinde . İstanbul: Hiperyayın.

- Potts, D. T. (2014). İskenderiye'den Önce: Antik Yakındoğu'daki Kütüphaneler. R. Macleod (Ed.), *İskenderiye Kütüphanesi Antik Dünya'nın Öğrenim Merkezi içinde* (ss. 35–52). Ankara: Dost Kitabevi.
- Pyne, L. (2019). *Kitaplık*. (Ü. Gurbanov, Ed.). İstanbul: İthaki.
- Quin, L. (2016). Extensible markup language (XML). <https://www.w3.org/XML/> adresinden erişildi.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. London: Morgan Kaufmann Publishers.
- Radebaugh, J. (2007). MARC 21 / MARCXML. *Computers in Libraries*, 27(4), 15. 12 Haziran 2023 tarihinde <https://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=24590341&lang=tr&site=ehost-live> adresinden erişildi.
- Raj, P., Surianarayanan, C., Seerangan, K. ve Ghinea, G. (2022). *Streaming analytics: Concepts, architectures, platforms, use cases and applications*. *Streaming Analytics: Concepts, architectures, platforms, use cases and applications*, IET Computing Series. Stevenage: The Institution of Engineering and Technology.
- Ramos de Carvalho, J., Inês Cordeiro, M., Lopes, A. ve Vieira, M. (2004). Meta-information about MARC: an XML framework for validation, explanation and help systems. *Library Hi Tech*, 22(2), 131–137.
- Ranganathan, S. R. (2006). *The five laws of library science*. New Delhi: Ess Ess Publications.
- Reid, E. (2023). Then and now: 5 ways we're continuing to improve Search. *Google*. 12 Haziran 2023 tarihinde <https://blog.google/products/search/google-search-breakthroughs-over-25-years/> adresinden erişildi.
- Reitz, J. M. (2004). *Dictionary for library and information science*. London: Libraries Unlimited.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. ve Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM conference on Computer supported cooperative work - CSCW '94 içinde* (ss. 175–186). New York, New York, USA: ACM Press.
- Rogers, F. B. (1964). The Development of MEDLARS. *Bulletin of the Medical Library Association*, 52(1), 150–151.
- Rogers, R. D. ve Weber, D. C. (1971). *University library administration*. New York: The H.W. Wilson Company.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180.
- Rowley, Jennifer. (2006). Where is the wisdom that we have lost in knowledge? *Journal of Documentation*, 62(2), 251–270.

- Russell, S. J. ve Norvig, P. (2010). *Artificial intelligence a modern approach* (Third Edit.). New Jersey: Pearson Education.
- SAE International. (2021). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. 22 Haziran 2023 tarihinde https://www.sae.org/standards/content/j3016_202104/ adresinden erişildi.
- Sağdıç, L. L. (1987). *Kütüphane otomasyonu ve T.B.M.M. uygulaması*. Ankara Üniversitesi.
- Salihoğlu, R. (2012). *Açık kaynak kodlu kütüphane otomasyon sistemlerinin akademik kütüphanelerde kullanımı*. Ankara Üniversitesi.
- Salmon, S. R. (1975). *Library automation systems*. New York: Marcel Dekker INC.
- Sarul, S. (2016). Veri madenciliği bankacılık uygulaması. *R ile veri madenciliği uygulamaları* içinde (ss. 243–253). İstanbul: Çağlayan Kitabevi.
- SAS. (2006). *Getting started with SAS® Enterprise Miner™ 5.2*. Cary, North Carolina: SAS Institute Inc.
- Satyral, U. R. (2019). *A study on recommendation system with library data*. 12 Mayıs 2023 tarihinde <https://www.doria.fi/handle/10024/170787> adresinden erişildi.
- Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, 5(2), 197–227.
- Schapire, R. E., Freund, Y., Bartlett, P. ve Lee, W. S. (1998). Boosting the margin : A new explanation for the effectiveness of voting methods. *The Annals of Statistics*, 26(5), 1651–1686.
- Schilling, F., Lecoeur, J., Schiano, F. ve Floreano, D. (2019). Learning vision-based flight in drone swarms by imitation. *IEEE Robotics and Automation Letters*, 4(4), 4523–4530. v
- Schoenung, J. G. (1974). *The Ohio college library center: An overview*.
- Scime, A. (Ed.). (2005). *Web mining : Applications and techniques*. Hershey: Idea Group Publishing.
- Şeker, Ş. E. (2015). Veri ambarı (Data warehouse). *YBS Ansiklopedi*, 2(4). 12 Şubat 2023 tarihinde https://ybsansiklopedi.com/wp-content/uploads/2015/09/veri_ambari.pdf adresinden erişildi.
- Şentürk, A. (2006). *Veri madenciliği: Kavram ve teknikler*. Bursa: Ekin Basım Yayın dağıtım.
- Sharma, N. (2008). The Origin of Data Information Knowledge Wisdom (DIKW) Hierarchy.
- Sherine, A., Jasmine, M., Peter, G. ve Alexander, S. A. (2023). *Algorithm and design complexity*. Boca Raton: CRC Press.
- Shetkar, A. A. ve Fernandes, S. (2016). Text categorization of documents using k-means and k-means++ clustering algorithm. *International Journal on Recent and Innovation Trends in Computing and Communication*, 4(6), 485–489.
- Shu, K. ve Liu, H. (2019). *Detecting fake news on social media*. Synthesis Lectures on

- Data Mining and Knowledge Discovery. Cham: Springer International Publishing.
- Siau, K. ve Wang, W. (2020). Artificial Intelligence (AI) Ethics. *Journal of Database Management*, 31(2), 74–87.
- Siciliano, B., Sciavicco, L., Villani, L. ve Oriolo, G. (2009). *Robotics*. Advanced Textbooks in Control and Signal Processing. London: Springer London.
- Silahtaroğlu, G. (2013). *Veri madenciliği kavram ve algoritmaları* (2. bs.). İstanbul: Papatya Yayıncılık Eğitim.
- Silberschatz, A., Korth, H. F. ve Sudarshan, S. (2020). *Database system concepts* (Seventh Ed.). McGraw-Hill Education.
- Simpson, J. (2016). The Heart of the University: Library Link Location on Doctoral Granting Institutions Webpages and Correlation with Research Output. *The Journal of Academic Librarianship*, 42(5), 503–508.
- Smithsonian Institution. (1851). *Fifth annual report of the board of regents of the Smithsonian institution*. Washington.
- Soysal, Ö. (1965). İlmi ve teknik literatürün kontrolü meselesi otomasyon ve kütüphaneler. *Türk Kütüphaneciler Derneği Bülteni*, 14(3–4), 19–23.
- Srivastava, J., Cooley, R., Deshpande, M. ve Tan, P.-N. (2000). Web usage mining. *ACM SIGKDD Explorations Newsletter*, 1(2), 12–23.
- Srivastava, J., Desikan, P. ve Kumar, V. (2004). Web mining - concepts, applications, and research directions. H. Kargupta, A. Joshi, K. Sivakumar ve Y. Yesha (Ed.), *Data mining: Next generation challenges and future directions* içinde (ss. 405–423). California: AAAI Press/ The MIT Press.
- Steinberg, D. ve Colla, P. (1995). *CART: Tree-structured nonParametric data analysis*. San Diego: Salford Systems.
- Strubell, E., Ganesh, A. ve McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* içinde (ss. 3645–3650). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Tabachnick, B. G. ve Fidell, L. S. (2015). *Çok değişkenli istatistiklerin kullanımı*. (M. Baloğlu, Ed.) (6. bs.). Ankara: Nobel Akademik.
- Tambe, N. ve Jain, A. (2023). Top website statistics for 2024. *Forbes Advisor*. 02 Haziran 2023 tarihinde <https://www.forbes.com/advisor/in/business/software/website-statistics/> adresinden erişildi.
- Tan, P.-N., Steinbach, M., Karpatne, A. ve Kumar, V. (2019). *Introduction to data mining* (Second Edi.). Harlow: Pearson Education.
- Taşkın, E. (2020). *Nicel birikimin nitel değişime etkisi: Kütüphane yönetiminde veri madenciliği uygulaması*. Trakya Üniversitesi.
- Taylor, P. (2023). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. *Statista*. 12 Kasım 2023 tarihinde <https://www.statista.com/statistics/871513/worldwide-data-created/> adresinden erişildi.

- Tekin, E. C. (2019). *Asurbanipal kütüphanesi*. Hiperyayın.
- Tempelman-Kluit, N. ve Pearce, A. (2014). Invoking the user from data to design. *College & Research Libraries*, 75(5), 616–640. doi:10.5860/crl.75.5.616
- Tennant, R. (2002). MARC must die. *Library Journal*, 15(October), 26–28.
- Terzi, R., Sağiroğlu, Ş. ve Demirezen, U. (2017). Büyük veri ve açık veri: Temel kavramlar. Ş. Sağiroğlu ve O. Koç (Ed.), *Büyük veri ve açık veri analitiği: Yöntemler ve uygulamalar* içinde (ss. 14–29). Ankara: Grafiker Yayınları.
- Tesendic, D., Krsticev, D. B. ve Surla, B. D. (2011). XML in the BISIS library management system. *Novi Sad Journal of Mathematics*, 41(2), 131–148.
- Thorpe, C., Herbert, M., Kanade, T. ve Shafer, S. (1991). Toward autonomous driving: the CMU Navlab. I. Perception. *IEEE Expert*, 6(4), 31–42.
- Tian, Y., Chen, X., Xiong, H., Li, H., Dai, L., Chen, J., ... Gao, W. (2017). Towards human-like and transhuman perception in AI 2.0: A review. *Frontiers of Information Technology & Electronic Engineering*, 18(1), 58–67.
- Tomic, T., Schmid, K., Lutz, P., Domel, A., Kassecker, M., Mair, E., ... Burschka, D. (2012). Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue. *IEEE Robotics & Automation Magazine*, 19(3), 46–56.
- Tonta, Y. (1999). Bilgi toplumu ve bilgi teknolojisi. *Türk Kütüphaneciliği*, 13(4), 363–375.
- Trustees of Dartmouth College. (2023). Artificial intelligence coined at Dartmouth. *Trustees of Dartmouth College*. 12 Ekim 2023 tarihinde <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth> adresinden erişildi.
- Tsuji, K., Takizawa, N., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F. ve Itsumura, H. (2014). Book recommendation based on library loan records and bibliographic information. *Procedia - Social and Behavioral Sciences*, 147, 478–486.
- Tunay, M. İ. (1970). Dünyanın En Eski Kütüphanelerinden İskenderiye Kütüphanesi. *Türk Kütüphaneciliği*, 19(1), 25–26.
- Tunçkanat, H., Karakaş, Ü., Yaşın, C., Tüzün, F. ve Çomuoğlu, D. (1988). Hacettepe Üniversitesi tıp merkezi kütüphane otomasyon projesi-Bilgieriş. *Türk Kütüphaneciliği*, 2(3), 109–116.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX(236), 433–460.
- Uçan, Ö. (2010). *Dijital kütüphanelerde veri madenciliği uygulamaları: Akdeniz Üniversitesi Merkez Kütüphanesi örneği*. Akdeniz Üniversitesi.
- Ünal Orhan, İ. (2006). *Türkçe tabanlı kütüphane otomasyon programları ve üniversite kütüphanelerinde uygulanması*. Ankara Üniversitesi.
- UNESCO. (1971). Records of the General Conference. Paris: UNESCO.
- Ürün, A. K. (2015). *Klasik Arap edebiyatı*. Konya: Çizgi Kitabevi.
- Üstün, A. (2017). *Bilgi Merkezleri: Üniversite ve Bilimin Anahtarı*. Anka Matbaa.
- Vapnik, V. N. ve Chervonenkis, A. Y. (1964). On a class of perceptrons. *Automation and*

- Remote, 25(1), 103–109.
- Vardi, M. Y. (2012). Artificial intelligence: Past and future. *Communications of the ACM*, 55(1), 5–5.
- Vercellis, C. (2009). *Business intelligence: Data mining and pptimization for decision making*. Cornwall: John Wiley & Sons Ltd.
- Vincze, M., Wachsmuth, S. ve Sagerer, G. (2014). Perception and computer vision. *The Cambridge Handbook of Artificial Intelligence* içinde (ss. 168–190). Cambridge University Press.
- Wang, B., Tao, J. ve Wu, H. P. (2011). The application of knowledge representation on the design of loading device of garbage truck. *Applied Mechanics and Materials*, 121–126, 3568.
- Watson, H. J. ve Wixom, B. H. (2007). The current state of business intelligence. *Computer*, 40(9), 96–99.
- Wayner, P. (2022). What is machine perception? How artificial intelligence (AI) perceives the world. *VentureBeat*. 1 Şubat 2024 tarihinde <https://venturebeat.com/programming-development/what-is-machine-perception/> adresinden erişildi.
- Williams, S. ve Williams, N. (2007). *The profit impact of business intelligence*. San Francisco: Morgan Kaufmann Publishers.
- Witten, I. H., Frank, E., Hall, M. A. ve Pal, C. J. (2017). *Data mining: Practical machine mearning tools and techniques* (Fourth Edi.). Boston: Elsevier.
- Wu, J. (2012). *Advances in K-means clustering*. Springer Theses. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-29807-3
- Wu, X. ve Kumar, V. (2009). *The top ten algorithms in data mining*. Boca Raton: Taylor & Francis Group.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37.
- Yang, D., Ma, Z. ve Buja, A. (2014). A Sparse singular value decomposition method for high-dimensional data. *Journal of Computational and Graphical Statistics*, 23(4), 923–942.
- Yeşilyurt, T. (2004). Kur'an' da Bilgi. *İlahiyat Fakültesi Dergisi*, 9(1), 1–12.
- Yılmaz, A. ve Aslan, H. (1992). Veri tabanına dayalı Türkçe kütüphane otomasyonu yazılımı: KYBELE. *Türk Kütüphaneciliği*, 6(1), 10–17.
- Yılmaz, M. (2006). T.C. Anadolu Üniversitesi Kütüphane ve Dokümantasyon Merkezi. *Bilgi Dnyasi*, 7(1), 147–151.
- Yohannes, Y. ve Webb, P. (1999). *Classification and regression trees, CART: a user manual for identifying indicators of vulnerability to famine and chronic food insecurity*. Washington: International Food Policy Research Institute.
- Yontar, A. (1995). *Kütüphane ve Belge-bilgi Merkezlerinde bilimsel yönetimin Önemi*. İstanbul: Türk Kütüphaneciler Derneği İstanbul Şubesi.

- Yükseköğretim Kurulu. (t.y.). Üniversitelerimiz. 7 Ekim 2023 tarihinde Yükseköğretim adresinden erişildi.
- Zafarani, R., Abbasi, M. A. ve Liu, H. (2014). *Social media mining: An introduction*. New York: Cambridge University Press.
- Zeleny, M. (1987). Management support systems: Towards integrated knowledge management. *Human Systems Management*, 7(1), 59–70.
- Zgurovsky, M. Z., Kasyanov, P. O. ve Levenchuk, L. B. (2023). Formalization of methods for the development of autonomous artificial intelligence systems. *Cybernetics and Systems Analysis*, 59(5), 763–771.
- Zhang, H. ve Thurber, C. H. (2007). Estimating the model resolution matrix for large seismic tomography problems based on Lanczos bidiagonalization with partial reorthogonalization. *Geophysical Journal International*, 170(1), 337–345.
- Zhihui, X. (Ed.). (2008). *Computer Vision*. Croatia: InTech. doi:10.5772/72
- Zikopoulos, P. C., DeRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D. ve Giles, J. (2013). *Harness the power of big data: The IBM big data platform*. New York: The McGraw-Hill Companies.

EKLER



EK 1. k-means Kümeleme Analizi Python Kodu

```
1. import pandas as pd
2. import matplotlib.pyplot as plt
3. from sklearn.cluster import KMeans
4. from sklearn.metrics import silhouette_score, precision_score, recall_score,
   f1_score, mean_squared_error
5. from sklearn.preprocessing import LabelEncoder
6. from math import sqrt
7. from sklearn.impute import SimpleImputer
8.
9. data_path = r"/Users/bayraktar/Desktop/ODUNC-kullanilan.xlsx"
10. data = pd.read_excel(data_path)
11. data.head(), data.info(), data.describe()
12.
13. missing_data = data.isnull().sum()
14. basic_stats = data.describe(include='all')
15. missing_data, basic_stats
16.
17. imputer = SimpleImputer(strategy="median")
18. data['YAS'] = imputer.fit_transform(data[['YAS']])
19. bins = [15, 25, 35, 45, 55, float('inf')] # 56 ve üzeri için float('inf')
   kullanılır
20. labels = ['16-25', '26-35', '36-45', '46-55', '56+']
21. data['YAS_GRUBU'] = pd.cut(data['YAS'], bins=bins, labels=labels, right=False)
22.
23. label_encoders = {}
24. for col in ['BOLUM', 'FAKULTE', 'KONU', 'YAS_GRUBU']:
25.     le = LabelEncoder()
26.     data[col] = le.fit_transform(data[col])
27.     label_encoders[col] = le
28.
29. selected_features = data[['BOLUM', 'CINSIYET', 'GRUP', 'FAKULTE', 'KONU',
   'YAS_GRUBU']]
30. selected_features.head()
31.
32.
33. #Silhouette Skoru
34. cluster_range = range(2, 11)
35. silhouette_scores = []
36. for n_clusters in cluster_range:
37.     kmeans = KMeans(n_clusters=n_clusters, random_state=42)
38.     cluster = kmeans.fit_predict(selected_features)
39.     silhouette_avg = silhouette_score(selected_features, cluster)
40.     silhouette_scores.append(silhouette_avg)
41. silhouette_scores
42.
43. plt.figure(figsize=(10, 6))
44. plt.plot(cluster_range, silhouette_scores, marker='o')
45. plt.title('')
46. plt.xlabel('Küme Numaraları')
47. plt.ylabel('Silhouette Skoru')
48. plt.grid(True)
49. plt.show()
50.
51.
52. #Elbow Grafiği
53. sse = []
54. k_range = range(1, 11) # 1'den 10'a kadar küme sayısı denenecek
55. for k in k_range:
56.     kmeans = KMeans(n_clusters=k, random_state=42)
57.     kmeans.fit(selected_features)
```

```

58.     sse.append(kmeans.inertia_)
59.
60. plt.figure(figsize=(10, 6))
61. plt.plot(k_range, sse, marker='o')
62. plt.title('')
63. plt.xlabel('Küme Numaraları')
64. plt.ylabel('Sum of Squared Errors (SSE)')
65. plt.xticks(k_range)
66. plt.grid(True)
67. plt.show()
68.
69. #K-MEANS
70. kmeans = KMeans(n_clusters=4, random_state=42)
71. clusters = kmeans.fit_predict(selected_features)
72.
73. data['Cluster'] = clusters
74. cluster_centers = pd.DataFrame(kmeans.cluster_centers_,
    columns=selected_features.columns)
75. cluster_sizes = data['Cluster'].value_counts()
76. print(cluster_centers, cluster_sizes)
77.
78. selected_features['Cluster'] = kmeans.labels_
79. selected_features.head()
80. cluster_summary = selected_features.groupby('Cluster').mean()
81.
82. cluster_summary[['BOLUM', 'CINSIYET', 'GRUP', 'FAKULTE', 'KONU', 'YAS_GRUBU']]
83. top_books_per_cluster = data.groupby('Cluster')['ESER-ADI'].apply(lambda x:
    x.value_counts().nlargest(5))
84.
85. def safe_mode(x):
86.     try:
87.         return x.mode()[0]
88.     except IndexError:
89.         return None # Mod bulunamazsa None döner
90.
91. most_common_cluster = data.groupby('UYE-
    ID')['Cluster'].agg(safe_mode).reset_index(name='Küme')
92. user_recommendations = most_common_cluster.copy()
93.
94. def get_recommendations(cluster):
95.     return list(top_books_per_cluster[cluster].index) if cluster in
        top_books_per_cluster else []
96. user_recommendations['Recommended Books'] =
    user_recommendations['Küme'].apply(get_recommendations)
97.
98. recommendation_columns = ['Tavsiye Kitap 1', 'Tavsiye Kitap 2', 'Tavsiye Kitap
    3',
99.                             'Tavsiye Kitap 4', 'Tavsiye Kitap 5']
100.
101.     user_recommendations[recommendation_columns] =
        pd.DataFrame(user_recommendations['Recommended Books'].tolist(),
            index=user_recommendations.index)
102.     user_recommendations.drop('Recommended Books', axis=1, inplace=True)
103.     print(user_recommendations.head())
104.
105.     output_path = '/Users/bayraktar/Desktop/tavsiye-kitaplar-k-means.xlsx'
106.     user_recommendations.to_excel(output_path)

107. # Tavsiye sisteminin performansını değerlendirme
108. true_books = data.groupby('UYE-ID')['ESER-ADI'].apply(list)

```

```

109. pred_books = user_recommendations.set_index('UYE-
    ID')[recommendation_columns].apply(lambda x: x.dropna().tolist(), axis=1)
110. precision = []
111. recall = []
112. f1 = []
113. for user_id in true_books.index:
114.     true_set = set(true_books[user_id])
115.     pred_set = set(pred_books.get(user_id, []))
116.     tp = len(true_set & pred_set)
117.     fp = len(pred_set - true_set)
118.     fn = len(true_set - pred_set)
119.     if len(pred_set) > 0:
120.         precision.append(tp / (tp + fp))
121.     if len(true_set) > 0:
122.         recall.append(tp / (tp + fn))
123.     if tp > 0:
124.         f1.append(2 * tp / (2 * tp + fp + fn))
125.
126. print("Precision: ", sum(precision) / len(precision))
127. print("Recall: ", sum(recall) / len(recall))
128. print("F1 Score: ", sum(f1) / len(f1))
129.
130. #RMSE
131. true_books = data.groupby('UYE-ID')['ESER-ADI'].apply(list)
132. def calculate_rmse(true_books, recommended_books):
133.     true_list = []
134.     pred_list = []
135.     for user in recommended_books['UYE-ID']:
136.         true_set = set(true_books.get(user, []))
137.         pred_set = set(recommended_books[recommended_books['UYE-ID'] ==
            user][recommendation_columns].values[0])
138.         for book in pred_set:
139.             true_list.append(1 if book in true_set else 0)
140.             pred_list.append(1)
141.     return sqrt(mean_squared_error(true_list, pred_list))
142.
143. rmse = calculate_rmse(true_books, user_recommendations)
144. print(f"RMSE: {rmse:.4f}")

```

EK 2. SVDs İşbirlikçi Filtreleme Analizi Python Kodu

```
1. import pandas as pd
2. import numpy as np
3. from scipy.sparse.linalg import svds
4. from sklearn.metrics import mean_squared_error
5. from math import sqrt
6. from tqdm import tqdm
7. import time
8. import matplotlib.pyplot as plt
9. import seaborn as sns
10. from matplotlib.colors import LinearSegmentedColormap
11.
12. data_path = r"/Users/bayraktar/Desktop/ODUNC-kullanilan.xlsx"
13. data = pd.read_excel(data_path)
14. print(data.head(), data.info(), data.describe())
15.
16. interaction_matrix = pd.pivot_table(data, index='UYE-ID', columns='ESER-ADI-KISA', aggfunc='size', fill_value=0)
17. print("Kullanıcı-Kitap Etkileşim Matrisi:")
18. print(interaction_matrix)
19. R = interaction_matrix.values
20. user_ratings_mean = np.mean(R, axis=1)
21. R_demeaned = R - user_ratings_mean.reshape(-1, 1)
22. # SVD
23. start_time = time.time()
24. U, sigma, Vt = svds(R_demeaned, k=100) # k, latent faktörlerin sayısını temsil eder
25. svd_time = time.time() - start_time
26. sigma = np.diag(sigma)
27. all_user_predicted_ratings = np.dot(np.dot(U, sigma), Vt) + user_ratings_mean.reshape(-1, 1)
28.
29. num_recommendations = 5 # Her kullanıcı için 5 kitap
30. recommendations_dict = {}
31. for idx, user_ratings in enumerate(all_user_predicted_ratings):
32.     user_id = interaction_matrix.index[idx]
33.     user_data = interaction_matrix.loc[user_id]
34.     user_full = (user_data - user_data.mean()) * 0 + user_ratings
35.     recommendations = user_full[user_data == 0].sort_values(ascending=False)[:num_recommendations]
36.     recommendations_dict[user_id] = recommendations.index.tolist()
37.
38. # Performans ölçümleri
39. results = {}
40. results['svd'] = {'time': svd_time}
41. num_users = R.shape[0]
42. recommendations_df = pd.DataFrame.from_dict(recommendations_dict, orient='index')
43. recommendations_df.columns = [f'Tavsiye Kitap {i+1}' for i in range(num_recommendations)]
44. # RMSE
45. mse = mean_squared_error(R, all_user_predicted_ratings)
46. rmse = sqrt(mse)
47.
48. output_path = r"/Users/bayraktar/Desktop/tavsiye-kitaplar-SVD.xlsx"
49. recommendations_df.to_excel(output_path)
50.
51. print(f'RMSE: {rmse}')
52. print(recommendations_df.head(100))
```


EK 3. Veri Seti Kullanım İzni



T.C.
BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
REKTÖRLÜĞÜ
Kütüphane ve Dokümantasyon Daire Başkanlığı



Sayı : E-27032400-903.99-390282
Konu : Kütüphane Verilerine Ulaşım İzni.

07.05.2024

SAYIN SEFA BAYRAKTAR

İlgi : 22.04.2024 tarihli, sayılı ve "Tezli Yüksek Lisans Programı İçin Veri Kullanım İzni H.k" konulu yazı

İlgi tarihli dilekçeniz incelenmiş olup, bilimsel çalışmalarda (Yüksek Lisans tez çalışmasında) kullanılmak üzere talep ettiğiniz bilgiler KVKK gereğince kişisel bilgilerin gizlenerek Pdf yada Excel ortamında verilebilecektir.
Bilgilerini ve gereğini rica ederim.

Prof. Dr. Oğuz GÜRSOY
Rektör Yardımcısı

Belge Doğrulama Kodu :BSCLNUZR27

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Adresi :

https://ebys.mehmetakif.edu.tr/enVision/Validate_Doc.aspx

İstiklal Yerleşkesi 15030 BURDUR

Telefon: +90 248 213 12 31 Faks: +90 248 213 12 33

e-Posta: kddeb@mehmetakif.edu.tr Elektronik Ağ: <http://kutuphane.mehmetakif.edu.tr>

Kep Adresi : maku@hs01.kep.tr

Ayrıntılı bilgi için irtibat: İsmail Mamiş
Evrak Pin Kodu: 57772



Bu belge, güvenli elektronik imza ile imzalanmıştır.

ÖZGEÇMİŞ

Kişisel Bilgiler:

Adı ve Soyadı : Sefa Bayraktar

Eğitim Bilgileri:

Lisans: Atatürk Üniversitesi, Edebiyat Fakültesi, Bilgi ve Belge Yönetimi Bölümü, 2009-2013

Yüksek Lisans: Burdur Mehmet Akif Ersoy Üniversitesi, Sosyal Bilimler Enstitüsü, Yönetim Bilişim Sistemleri Anabilim Dalı, 2016- 2024

Yabancı Dil/Düzeyi:

İngilizce: Başlangıç

İş Deneyimleri:

- Burdur Mehmet Akif Ersoy Üniversitesi Merkez Kütüphanesi, Kütüphaneci, 2013-2019
- Cumhurbaşkanlığı Millet Kütüphanesi, Kütüphaneci, 2019-

Bilimsel Yayınlar ve Çalışmalar:

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında basılan bildiriler:

- Bilgi Merkezlerinde Kullanıcı Odaklı Web Siteleri: MAKÜ Merkez Kütüphanesi İncelemesi- II. Uluslararası Sosyal Bilimler Sempozyumu 18-19-20 Mayıs 2017, Alanya
- Bilgi Merkezlerinde Bulut Bilişim: MAKÜ Merkez Kütüphane Örneği- II. Uluslararası Sosyal Bilimler Sempozyumu 18-19-20 Mayıs 2017, Alanya
- SVDs Algoritmasını Kullanan Kişiselleştirilmiş Kitap Öneri Sistemi- III. BİLSEL Uluslararası Truva Bilim Araştırmaları ve İnovasyon Kongresi 25-26 Mayıs 2024, Çanakkale
- K-Means Algoritmasını Kullanarak Kütüphanede Kullanıcı Segmentasyonu- III. BİLSEL Uluslararası Truva Bilim Araştırmaları ve İnovasyon Kongresi 25-26 Mayıs 2024, Çanakkale

Ulusal bilimsel toplantılarda sunulan ve bildiri kitabında basılan bildiriler:

- Lisans Öğrencileri İçin Web Tabanlı Bilgi Okuryazarlığı Programı
Geliştirme- 19. Akademik Bilişim Konferansı 8-10 Şubat 2017, Aksaray



MAKÜ

Burdur Mehmet Akif Ersoy Üniversitesi, Sosyal Bilimler Enstitüsü

İstiklal Yerleşkesi, 15030 BURDUR

0 248 213 31 65

0 248 213 25 04

sbe@mehmetakif.edu.tr