T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

BİLGİ MERKEZİ HİZMETLERİNİN KULLANICILAR TARAFINDAN KULLANIM TEPKİ DURUMLARININ VERİ MADENCİLİĞİ YAKLAŞIMI İLE İNCELENMESİ

DOKTORA TEZİ

Engin DAYAN
Danışman: Prof. Dr. Mahmut KARA

VAN – 2025

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

BİLGİ MERKEZİ HİZMETLERİNİN KULLANICILAR TARAFINDAN KULLANIM TEPKİ DURUMLARININ VERİ MADENCİLİĞİ YAKLAŞIMI İLE İNCELENMESİ

DOKTORA TEZİ

Engin DAYAN

Tez Savunma Sınavı Jüri Üyeleri

Prof. Dr. Mehmet Akif ARVAS (Başkan)

Prof. Dr. Mahmut KARA (Danışman)

Doç. Dr. Alkan ÖZKAN (Üye)

Doç. Dr. Cem TIRINK (Üye)

Dr. Öğr. Üyesi Burak UYAR (Üye)

VAN – 2025

KABUL VE ONAY SAYFASI

Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı'nda Prof. Dr. Mahmut KARA danışmanlığında, Engin DAYAN tarafından sunulan "Bilgi Merkezi Hizmetlerinin Kullanıcılar Tarafından Kullanım Tepki Durumlarının Veri Madenciliği Yaklaşımı ile İncelenmesi" başlıklı bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince 12/02/2025 tarihinde aşağıdaki jüri tarafından oy birliği ile başarılı bulunmuş ve Doktora tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. Mehmet Akif ARVAS İmza:

Üye: Prof. Dr. Mahmut KARA İmza:

Üye: Doç. Dr. Alkan ÖZKAN İmza:

Üye: Doç. Dr. Cem TIRINK İmza:

Üye: Dr. Öğr. Üyesi Burak UYAR İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 04/02/2025 tarih ve 2025/5-9 sayılı kararı ile onaylanmıştır.

İmza

……

Enstitü Müdürü

ETİK BEYAN

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

İmza
Engin DAYAN

ÖZET

BİLGİ MERKEZİ HİZMETLERİNİN KULLANICILAR TARAFINDAN KULLANIM TEPKİ DURUMLARININ VERİ MADENCİLİĞİ YAKLAŞIMI İLE İNCELENMESİ

DAYAN, Engin
Doktora Tezi, İstatistik Anabilim Dalı
Danışman: Prof. Dr. Mahmut KARA
Şubat 2025, 109 sayfa

Geleneksel kütüphane hizmet anlayışından farklı olarak günümüzde çok yönlü ve büyük miktarlarda üretilen bilgi gerçek kullanıcısı için erişilebilir olmuştur. Bu paradigma değişimi, kullanıcıların bilgiye erişimini kolaylaştırma görevini üstlenen kütüphaneler için çeşitli zorlukları beraberinde getirmiştir. Bu zorlukların yanı sıra, kullanıcıların bilgiyi elde etme davranışında karşılaştığı olumsuzluklar çevrimiçi bilgiye erişim sürecini erken sonlandırmalarına veya tamamen terk etmelerine neden olabilmektedir. Araştırma sürecinden ayrılma riski taşıyan kullanıcıların davranışlarının ve bilgi erişimindeki olası başarısızlıklarının erken aşamada tahmin edilmesi, kütüphane kullanım sürecinin etkinliğini artırmak açısından kritik bir rol oynamaktadır. Bu çalışma kapsamında, kütüphane kullanıcılarının bilgi erişim süreçleri ve kütüphane kaynaklarının kullanımına ilişkin veriler titizlikle analiz edilerek anlamlı bilgilere dönüştürülmüştür. Elde edilen bulgular doğrultusunda, kullanıcıların bilgiye erişim sürecinde karşılaştıkları güçlükleri ortadan kaldırmaya veya en aza indirmeye yönelik uygulanabilir bir model önerisi geliştirilmiştir.

Bu çalışma kapsamında kullanıcıların araştırma süreçleri ile ilgili verilerin toplandığı üniversite kütüphanesi otomasyonu araştırma amacı çerçevesinde kullanılmıştır. Çalışmada kullanılan veriler bu otomasyonda biriken büyük web günlük dosyalarından elde edilmiştir. Kullanıcıların araştırma başarım performansının, çevrimiçi ortamdan elde edilen değişkenlere dayalı olarak modellenmesi, araştırma sürecini terk etme eğilimi gösteren bireylerin erken aşamada belirlenmesi ve araştırma sürecinin sonunda ortaya çıkabilecek olası başarısızlıkların öngörülmesi açısından büyük önem taşımaktadır. Elde edilen modeller, uyarlanabilir çevrimiçi araştırma ortamlarında kullanıcıların faaliyet düzeylerini değerlendirerek otomatik olarak sınıflandırılmasına veya otomatik uyarlamaların gerçekleştirilmesine yönelik araçlar olarak işlev görebilir.

Çalışma Ocak 2022'den Haziran 2023 tarihleri arasındaki 18 takvim ayını kapsayan kullanıcı verilerinden oluşmaktadır. Veriler, **Van Yüzüncü Yıl Üniversitesi Ferit Melen Merkez Kütüphanesi otomasyonu** üzerinden elde edilmiştir. Otomasyon sağlayıcı firmadan günlük olarak tutulmuş log dosyası şeklindeki veriler alınmıştır. Veriler günlük işlemleri kapsayan toplam **459 adet txt.** formatlı dosyadan oluşmaktaydı. Log dosyasını excel dosyası dönüşümü sonrasında toplam **328 bin 100 satırdan** oluşmaktadır. Her bir eylem, kullanıcılara özgü benzersiz bir kimlik kodu (kullanıcı IP'si) ile temsil edilmiştir. Bu şekilde bireysel düzeyde izlemi anonim bir şekilde sağlamıştır. Arama sürecinden ayrılma riski taşıyan kullanıcıları belirlemek amacıyla toplam yapılmış kullanım verileri üzerinden veri analizi gerçekleştirilmiştir. Kullanıcıların etkinlik değişkenleri hesaplanmıştır. Bununla birlikte kullanıcı eylemleri, veri kümesinde bulunan diğer kullanıcıların eylemleriyle karşılaştırılarak çeşitli ölçümler oluşturulmuş ve arama davranışını yansıtan **18 adet değişken** belirlenmiştir. Kullanıcıların araştırma

performansları ise hipotetik hesaplama kullanılarak **"geçti" – "kaldı"** biçiminde kodlanmıştır.

Bu araştırmanın parametreleri dâhilinde, kullanıcıların araştırma faaliyetlerine ilişkin verilerin toplanmasını kolaylaştıran üniversite kütüphanesi otomasyon sistemi, bilimsel inceleme amacıyla kullanılmıştır. Bu analizde kullanılan veriler, söz konusu otomasyon sistemi içerisinde biriken kapsamlı web günlük dosyalarından elde edilmiştir. Çevrimiçi ortamdan çıkarılan değişkenlere dayalı olarak kullanıcıların araştırma başarı performanslarının modellenmesi, araştırma sürecinden ayrılma eğilimi gösteren bireylerin öngörülmesi ve olası başarısızlıkların erken aşamada belirlenmesi açısından büyük önem taşımaktadır. Elde edilen modeller, uyarlanabilir çevrimiçi araştırma bağlamlarında kullanıcıların otomatik sınıflandırılmasına olanak tanıdığı gibi, ortamda gözlemlenen etkinlik yoğunluğuna bağlı olarak otomatik düzenlemeler yapılmasını da sağlayabilir.

İlk araştırma sorusuna yanıt aramak amacıyla, kullanıcıların araştırma başarı performanslarını en iyi şekilde tahmin edebilecek algoritma ve değişkenleri belirlemek için çeşitli sınıflandırma algoritmaları ve ön işleme teknikleri sistematik bir şekilde karşılaştırılmıştır. Araştırma başarısıyla ilgili bir diğer problem kapsamında, seçilen algoritma ve değişkenler kullanılarak, kullanıcıların araştırma performanslarının önceki haftalarda öngörülüp öngörülemeyeceği araştırılmıştır. Son araştırma sorusu bağlamında ise, çevrimiçi öğrenme ortamında benzer davranış örüntüleri sergileyen kullanıcı grupları belirlenmiş ve bu grupların araştırma başarı performansı ile olan ilişkisi incelenmiştir. Veritabanlarında gizli kalmış içgörüleri ve örüntüleri ortaya çıkarmak için kabul gören yöntemlerden biri olan veri madenciliğinden yararlanılmıştır. Öngörüsel analizler için sınıflandırma algoritmaları kullanılırken, benzer kullanıcı gruplarını tespit etmek amacıyla kümeleme algoritmalarından faydalanılmıştır. Öngörü analizlerinde elde edilen sonuçların genelleştirilmesi ve kümeleme analizinde en uygun küme sayısının belirlenmesi için çapraz geçerlilik yöntemi uygulanmıştır.

Çalışma sonucunda, araştırma süresi boyunca kullanıcı aktivitelerindeki değişimlerin belirlenmesinin, risk altındaki kullanıcıların henüz araştırma sürecini bırakmadan önce gerçek zamanlı olarak tespit edilmesine yardımcı olabileceği bulunmuştur. Süreç boyunca kullanıcı verilerinin aylık olarak incelenmesi, genel ortalamalardan sapma gösteren, olağan dışı davranış sergileyen kullanıcıları belirleyerek erken uyarı sistemleri geliştirmesine olanak tanıyabilir. Ayrıca, sorumlu mentörlerin ya da öğretim üyelerinin sorumlu oldukları kullanıcıları web sitesi kullanım verileri sonucundaki analizler yardımıyla gözlemlemelerine ve beklentilerine uygun hareket etmeyen kullanıcıları belirlemelerine olanak sağlamaktadır.

Araştırmanın sonuçları, kullanıcıların çevrimiçi arama etkileşim verilerinin araştırma başarı performanslarını tahmin etmek için etkin bir şekilde kullanılabileceğini ortaya koymuştur. En yüksek doğruluk oranı, verilerin eşit genişlik yöntemiyle kesikli hâle getirildiği ve en iyi on değişkenin ReliefF skorlama sistemine göre seçildiği durumda elde edilmiştir. Bu durumda, Destek Vektör Makinesi (SVM) algoritması, kullanıcıların %99'unu doğru bir şekilde sınıflandırmayı başarmıştır.

Kullanıcıların nihai araştırma başarı performanslarının önceki haftalardan tahmin edilip edilemeyeceğine ilişkin analizler incelendiğinde, dördüncü hafta itibarıyla %97 doğruluk oranıyla tahmin edilebileceği belirlenmiştir.

Kümeleme analizlerinin sonuçları değerlendirildiğinde, kullanıcıların çevrimiçi araştırma ortamındaki etkinlik seviyelerine göre ideal olarak üç farklı kümede gruplandığı görülmüştür: *aktif olmayan, aktif* ve *çok aktif.* Daha sonra bu kümelerin araştırma başarı

performanslarıyla olan ilişkisi incelenmiştir. Bulgular, aşağıdaki sonuçları ortaya koymuştur: Süreç boyunca minimum düzeyde etkileşim gösteren kullanıcılar, düşük araştırma performansı sergilemiştir (düşük düzeyde arama katılımı), orta düzeyde etkileşim gösteren kullanıcılar, orta düzeyde araştırma başarısı göstermiştir (orta düzeyde arama katılımı), yüksek etkileşim seviyesine sahip kullanıcılar ise, yüksek araştırma performansı sergilemiştir (yüksek düzeyde arama katılımı).

Bu sonuçlar, veri madenciliği ve öğrenme analitiklerinin çevrimiçi araştırma ortamlarında kullanıcı etkileşimini artırmak ve akademik başarıyı iyileştirmek için etkin bir şekilde kullanılabileceğini göstermektedir.

Anahtar kelimeler: Bilgi arama davranışı, Karar ağacı, Kişiselleştirme, Kullanıcı davranışları, Kümeleme, Kütüphaneler ve bilgi merkezleri, Kütüphane kullanıcıları, Öğrenme analitiği, Sınıflandırma, Veri madenciliği, Yapay sinir ağı

ABSTRACT

ANALYSIS OF THE USAGE OF THE USERS' RESPONSES TO THE INFORMATION CENTER SERVICES WITH IN THE CONTEXT OF DATA MINING APPROACH

DAYAN, Engin
Ph.D. Thesis, Department of Statistics
Supervisor: Prof. Dr. Mahmut KARA
February 2025, 109 pages

In contrast to the conventional library service paradigm, the contemporary landscape has witnessed the emergence of multifaceted and extensive volumes of information that are readily accessible to actual users. This paradigm shift has engendered a myriad of challenges for libraries, whose imperative is to facilitate user access to information. Furthermore, the complexities faced by users in their quest for information often lead to premature abandonment or cessation of the online information access process. The capacity to forecast, at an incipient stage, the behaviors of users likely to disengage from the research process and their potential failures in information retrieval is vital for optimizing library utilization. Through the present study, data pertaining to the usage of library resources and the processes of information access among library users were meticulously analyzed and subsequently transformed into actionable insights. Drawing upon the insights gleaned from this analysis, a pragmatic model proposal was formulated with the objective of mitigating or alleviating the challenges users encounter in their pursuit of information access.

Within the scope of this study, the university library automation system, where data related to users' research processes were collected, was used for research purposes. The data used in the study were obtained from large web log files accumulated in this automation system. Modeling the performance of users in terms of research success, predicated upon variables derived from the online environment, is of paramount importance for the early identification of individuals who exhibit a propensity to abandon their research endeavors and may experience potential failures at the culmination of the research process. The resultant models can serve as instrumental tools for the automatic categorization of users by assessing their levels of activity within adaptive online research environments or for implementing automated modifications.

The study consists of user data covering an 18-calendar-month period between January 2022 and June 2023. The data were obtained from the **Ferit Melen Central Library Automation System at Van Yüzüncü Yıl University**. The log files, recorded daily by the automation provider, were collected in a total of **459 text files** containing daily operations. After converting the log files into an Excel file, the data consisted of **328,100 rows** in total. Each action was represented by a unique identification code specific to users (user IP), ensuring anonymous tracking at an individual level. A data analysis was conducted based on the total usage data to identify users at risk of leaving the search process. The activity variables of users were calculated. Additionally, various measurements were created by comparing user actions with those of other users in the dataset, and **18 variables reflecting search behavior** were identified. Users' research performance was hypothetically coded as **"passed" – "failed."**

Within the parameters of this investigation, the automation system of the university library, which facilitated the collection of data pertinent to users' research activities, was employed for scholarly inquiry. The data utilized in this analysis were derived from extensive web log files amassed within this automation framework. The modeling of users' research success performance, predicated upon variables extracted from the online domain, is of paramount importance in forecasting individuals who are likely to disengage from the research endeavor and identifying potential failures at an incipient stage. The resultant models are amenable to automatic classification of users within adaptive online research contexts or to facilitate automated modifications predicated on the intensity of activities present in the environment.

In addressing the initial research inquiry, various classification algorithms and preprocessing techniques were systematically compared to ascertain the most efficacious algorithm and variables for predicting users' research success performance. In relation to another research challenge concerning success, an investigation was conducted to ascertain whether users' research performance could be anticipated in the earlier weeks leveraging these variables and the selected algorithm. For the concluding research inquiry, user cohorts exhibiting analogous behavioral patterns within the online learning milieu were delineated, and their correlation with research success performance was scrutinized.

To unearth latent patterns and insights embedded within databases, widely recognized data mining methodologies were employed. Classification algorithms were deployed for predictive analytics, while clustering algorithms were utilized to identify akin user groups. In order to generalize the predictive outcomes and ascertain the optimal number of clusters within the clustering analysis, the cross-validation technique was instituted.

The outcomes of the research elucidated that the online search interaction data of users could be adeptly leveraged to predict research success performance. The highest rate of accurate classification was accomplished when the data were discretized utilizing the equal-width binning methodology, and the top ten variables were selected based on the ReliefF scoring system. In this instance, the Support Vector Machine (SVM) algorithm successfully classified 99% of users.

Upon examining the analyses pertaining to the predictability of users' final research success performance from earlier weeks, it was determined that such performance could be forecasted with 97% accuracy by the fourth week.

Upon review of the clustering analyses results, it was discerned that users were optimally categorized into three distinct clusters based on the levels of activity they manifested within the online research setting: inactive, active, highly active. Subsequently, the interrelation between these clusters and research success performance was investigated. The findings revealed that: Users exhibiting minimal engagement throughout the process demonstrated low research performance (low search engagement), users displaying moderate engagement manifested moderate research performance (intermediate search Engagement), users with high engagement levels showcased high research performance (high search engagement).

These results signify that data mining and learning analytics can be effectively harnessed to enhance user engagement and academic achievement within online research environments.

Keywords: Artificial eural network, Classification, Clustering, Data mining, Decision trees, İnformation-seeking behavior, Learning Analytics, Library and İnformation Center, Library user, Personalization, User behavior

TEŞEKKÜR

Bu tez çalışmasında, her türlü ilgi ve yardımlarını esirgemeyen danışmanım Sayın Prof. Dr. Mahmut Kara'ya, ayrıca bilgi ve deneyimlerini benimle paylaşarak yol gösterdikleri için Sayın Prof. Dr. M. Akif ARVAS'a, Dr. Öğr. Üyesi Burak UYAR'a, Doç. Dr. Cem TIRINK'a ve Doç. Dr. Alkan ÖZKAN'a teşekkür ederim.

Bu süreçte bana moral veren, destek olan tüm arkadaşlarıma ve tezime katkı sunan ve bana ilham veren tüm kişi ve kurumlara teşekkürlerimi sunuyorum.

Annem, sevgili eşim Reyhan, canım kızım Zeynep Liya ve diğer tüm değerli aile üyelerim, birlikte olduğumuz her an, bu süreci daha kolay ve anlamlı kılarak sağlamış olduğunuz destekleriniz ve sevginizden dolayı sizlere minnettarım.

Bu çalışmayı, O'nun sevgisi, öğütleri ve her daim yanımda hissettiğim desteği ile bir adım daha ileriye taşımama olanak sağlayan, henüz doktoraya başladığımda kaybetmiş olduğum canım babamın anısına itham ediyorum.

2025

Engin DAYAN

İÇİNDEKİLER

Sayfa

ÇİZELGELER LİSTESİ

Sayfa

ŞEKİLLER LİSTESİ

Sayfa

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamalarıyla aşağıda sunulmuştur.

Simgeler	**Açıklama**
AUC	ROC Eğrisi Altındaki Alan
CART	Sınıflandırma ve Regresyon Ağacı
ID3	Yinelemeli Dikotomizer 3

Kısaltmalar	**Açıklama**
BIT	Bilgi ve İletişim Teknolojileri
BES	Bilgi Erişim Sistemi
CHAID	Ki-Kare Otomatik Etkileşim Dedektörü
DBSCAN	Gürültülü Uygulamaların Yoğunluk Tabanlı Mekansal Kümelenmesi
DSO	Doğru Sınıflama Oranı
DVM	Destek Vektör Makinaları (Support Vector Machine)
EAKA	ROC Eğrisi Altında Kalan Alan
EVM	Eğitsel Veri Madenciliği
IIR	Etkileşimli Bilgi Erişim Sistemleri
IR	Bilgi Erişim
IoT	Nesnelerin İnterneti
KDD	Veri Tabanlarından Bilgi Keşfi (Knowledge Discovery in Databases)
Knn	k En Yakın Komşuluk (k-Nearest Neighbors)
MARC	Makine Okunabilir Kataloglama
	Doğal Dil İşleme

NLP	Optimizasyon Algoritmaları
OA **OPAC**	Online Toplu Erişim Kataloğu
ÖA	Öğrenme Analitiği
ROC	Alıcı İşletim Karakteristiği (Receiver Operating Characteristic)
RSS	Zengin Site Özeti
SAS	Statistical Analysis Software
VA	Veri Analitiği
VM	Veri Madenciliği
WEKA	Waikato Bilgi Analizi Ortamı
YZ	Yapay Zeka
YÖK	Yükseköğretim Kurulu

1. GİRİŞ

Geleneksel kütüphane hizmet anlayışından farklı olarak günümüzde çok yönlü ve büyük miktarlarda üretilen bilginin gerçek kullanıcısı için erişilebilir olma durumu bilgiyi kullanıcısına eriştirme görevine sahip kütüphaneleri ve bilgi merkezlerini farklı açılardan zorluklarla karşı karşıya bırakmıştır (Dayan, 2023). Dijital çağın etkilerinin yoğun olarak yaşandığı günümüzde, kütüphaneler ve bilgi merkezlerinin sadece zengin kaynak koleksiyonlarına ve geliştirilmiş nitelikli hizmetlere sahip olmaları yeterli olmamakta, aynı zamanda kütüphane deneyimlerini kullanıcıların bireysel ihtiyaçlarına daha uygun hale getirebilmek için sunulan hizmetlerin kullanıcı beklentilerine göre özelleştirilmesi önemli olmuştur (Liu ve Hu, 2020). Hem bilgi çağının yaşanan etkileri sonucu olarak elektronik ortamlara aktarılan, elektronik ortamlarda üretilen veri miktarındaki artış hem bu verilere ulaşmaya çalışan farklı özelliklere sahip bilgi kullanıcıları ve hem de bu kullanıcıların öğrenme sürecinde yer alan mentorlerinin izleme, yönlendirme gereklilikleri doğru kullanıcıların ihtiyaç duyduğu bilgiye erişmelerini zorlaştırabilmektedir. Ayrıca kullanıcıların bilgiye erişim sürecindeki davranışlarında da (ödünç alma, arama yapma ve hizmete dâhil olma davranışları) büyük değişimleri beraberinde getirmiştir. Bu dönemde kütüphaneler, kişiselleştirilmiş hizmet sunumu için kullanıcı verilerinden daha fazla yararlanma ihtiyaçları giderek artmıştır. Öte yandan sistemde toplanan veriler çok büyük miktarlara ulaştığında geleneksek araçlar ve uygulamalarla hızlı yanıt verilmesi zorlaşmakta ve belli bir yerden sonra ise sisteme çeşitli zararlarla sonuçlanabilmektedir. Benzer şekilde süreçteki önemli bir diğer zorluk, yönetsel kararların kalitesini arttırmak için büyük miktardaki veri içerisinde mevcut bilgiye ulaşma çabasıdır (Uppal ve Chindwani, 2013). Veri analizi, veri tabanı çalışmaları ve dosyalama sistematiğindeki yeniliklerle bu problemlere çözüm aranmıştır. Sistemlerdeki veri miktarı yükünün ortaya çıkardığı bu problemlerin çözümüne yönelik "Veri Madenciliği" (Data Mining) teknikleri; istatistiksel yöntemler, makine öğrenimi ve küme eğitimi kullanılmaktadır. Bu tekniklerin sistemlerde kullanılmasıyla kullanıcıların bilgiye erişme süreçleri daha verimli ve etkili gerçekleşebildiği gibi erişilmek istenen bilgi doğru kullanıcısına kanalize olabilmektedir.

Veri Madenciliği, büyük ölçekli veri kümeleri içerisindeki önemli, anlamlı ve değerli olan örüntüleri otomatik veya yarı-otomatik tekniklerle elde etmede kullanılan bir

süreçtir. Büyük ölçekli veri yığınlarında ki bilinmeyen örüntüler veri madenciliği ve çeşitli istatistiksel analizler kullanılarak uygun formatta saklanan yapılandırılmamış verilerden çıkarılır. Elde edilen bu değerli örüntüler, gelecekte daha etkili stratejiler geliştirmek ve veri odaklı kararlar almak için kullanılabilir. Veri madenciliği kavramı için Han vd. (2011), "Bilgi keşfinin önemli adımı olarak büyük miktardaki verilerden bilgi çıkarma, amacının ise keşfedici veri analizi, örüntüler, kurallar ve algoritmalar keşfetmeyi, öngörücü modellemeyi ve sapmaları bulmayı amaçlamakta" şeklinde açıklamaktadırlar. Genel olarak, kullanıcıların bilgi gereksinimlerini yönetmek ve anlamak için veri madenciliği kullanmak, veri ilişkilerinin keşfedilmesine ve yararlı bilgilerin sağlanmasına yol açabilir (Suresh vd. 2018b). Bilgisayar sistemlerinin hem giderek ucuzlaması hem de daha güçlü hale gelmesi, daha büyük miktarlarda verinin saklanmasını mümkün kılmaktadır. Bu nedenle, büyük veri yığınlarını işleyebilen tekniklerin kullanılması ve veri madenciliği modern teknolojisinin bilgi merkezleri ve kütüphane yönetim sistemlerinde kullanılması oldukça önemlidir (Zhang, 2018: Yi vd. 2014). Bilgi merkezleri ve kütüphanelerin görevlerini kolaylaştırmada potansiyel işlevselliğe sahip olan bu yeni ve güvenilir teknoloji, bilgiye dayalı kararlar için kullanıcıların gelecekteki davranışlarını tahmin etmek için uygun bir araçtır (Mishra ve Mishra, 2013). Bu şekilde, veri madenciliği ödünç verme işlemlerini analiz ederek kullanıcılara ilgi alanlarına uygun kitaplar önerebilir. Bu da kütüphane kaynaklarının kalitesinin artırılmasına ve yönetiminin kolaylaşmasına katkı sağlar (Chen ve Chen, 2007a; Jomsri 2014; Liu ve Wang, 2018; Morawski, 2017; Tsuji vd. 2012b; Yi vd. 2018; Yi vd. 2014; Zhang, 2018). Kütüphaneler üzerinde tavsiye edici ve çeşitli akıllı sistemleri uygulamak için veri madenciliği tekniklerini kullanmışlardır. Veri madenciliği, özellikle müşteri profillerinin oluşturulması, hizmet pazarlaması alanı ve hedef kitlenin taleplerinin belirlenmesi gibi uygulamalarda giderek artan bir popülerlik kazanmıştır.

Geleneksel bilgi erişim paradigmaları, kullanıcı sorgularıyla karşılaştırma ve eşleştirme yoluyla, veri tabanlarından ilgili bilgi ve belgelerin elde edilmesine odaklanmaktadır (Saracevic, 1997). Bununla birlikte, günümüz bağlamında bilgi erişim sürecinin etkili ve verimli bir şekilde gerçekleştirilmesi, yalnızca kullanıcıya ilgili bilgi ve belgelerin sunulmasını aşmaktadır. Bu süreç, kullanıcıların davranışları ve etkileşimlerine odaklanarak erişim deneyimine ilişkin somut verilerin elde edilmesini gerektirmektedir. Bu yöntem, kullanıcıların bilişsel, duygusal ve fiziksel etkileşimlerini

bilgi erişim sistemiyle kurdukları bağ üzerinden vurgulamaktadır (Aydemir Şenay, 2021). Bu bağlamda, bilgi merkezlerinin hizmet verdiği kitlenin koleksiyon geliştirme, hizmet planlama, tasarım, uygulama ve değerlendirme süreçleri açısından kapsamlı bir şekilde anlaşılması gereklidir. Kullanıcıların bilgi davranışlarının ve sunulan hizmetlerden memnuniyet düzeylerinin belirlenmesi, hizmet kalitesini artırmada ve kullanıcı merkezli hizmet ve sistemlerin oluşturulmasında temel bir rol oynamaktadır. Ayrıca, bilgi davranışlarına yönelik yapılan araştırmalar, hem ulusal hem de yerel düzeyde bilgi politikalarının oluşturulması ve bilgi erişim yöntemlerinin ve tekniklerinin geliştirilmesi için bir temel sağlayabilir (Najjari, 2010; Uçak, 1997).

Bilgi erişim süreci, hem bilgi erişim sistemlerinin ilgili bilgi ve dokümanlara eriştirip eriştirmediğiyle hem de kullanıcıların sistemi nasıl kullandığıyla ilgilenmelidir. Bireylerin, bilgi arama davranışlarının belirlenerek sistem ile etkileşimlerinin üst düzeyde olmasının sağlanması, tarama sonuçlarının anlamlılığını artırdığı gibi kullanıcı memnuniyetini de bilgi erişim süreci açısından artırmaktadır. Söz konusu bu memnuniyetin artırılması mutlaka kullanıcıların sistemle etkileşiminin ele alındığı yöntemlerle belirlenip ve değerlendirilip bu duruma uygun bilgi erişim süreci bileşenlerinin hazırlanmasıyla mümkündür.

Gelinen noktada bilgi merkezleri ve kullanıcıları arasında interaktif bir çevrimiçi bilgi edinme-kullanma ortamının olduğunu görmekteyiz. Bilgi merkezleri, genel verimliliği artırmak, değişen kullanıcı taleplerini karşılamak ve hizmetleri optimize etmek için dijital çağda giderek daha fazla ileri teknoloji kullanıyor. Dijital bilgi merkezi/kütüphane, çok çeşitli kitap koleksiyonlarıyla bilgi üretimi, paylaşımı ve korunması hizmetlerini teşvik etmeye kendini adamış bilgili personellere sahip olan, sürekli gelişen bir platforma dönüşmüştür. Çevrimiçi ortamda kullanıcı aktiviteleri sonucunda ortaya çıkan çok çeşitli (gezinim, kullanıcı vs.) veri kümeleri kayıt edilmektedir. Bu verilerin Veri Madenciliği (VM), Veri Analitiği (VA), Optimizasyon Algoritmaları (OA) ve Öğrenme Analitiği (ÖA) gibi yaklaşımlarla analiz edilmesi ile kullanıcı davranışları, benzer kullanıcı profillerinin oluşturtulması, uyarlanabilir ve kişiselleştirilebilir çevrimiçi bilgi paylaşım ortamlarının oluşturulması mümkün hale gelebilir.

1.1 Problem Durumu

Araştırmanın asıl problemi "kütüphane otomasyonundan bilgi erişim ve kütüphane kullanım süreçlerinden elde edilen verilerin veri madenciliği yöntemleri ile analiz edilerek kullanıcıların bilgi arama davranışlarının modellenmesinde kullanılıp kullanılmayacağının araştırılmasıdır" biçimindedir.

1.2 Araştırmanın Amacı ve Hipotezi

Araştırmanın temel amacı, bir bilgi erişim merkezi olarak üniversite kütüphanesi kullanıcılarının otomasyon kullanarak bilgiyi arama ve bilgiye erişme süreçlerindeki verileri kullanılarak araştırma performanslarının modellenmesidir. Bu doğrultuda;

- Kullanıcının bilgi edinme sürecinde süreci tamamlama, tamamlayamama/ terk etme,
- Sürecin devam durumunun tahmin edilip edilemeyeceğinin araştırılması,
- Süreci tamamlamadan terk etme durumunu süreç içerisinde erken tahmin edilip edilememesinin araştırılması,
- Benzer davranış örüntüsüne sahip süreci tamamlamadan terk eden kullanıcı gruplarının kümelenmesi ve karşılaştırılması amaçlanmıştır.

Bu amaçlar doğrultusunda kullanıcıların araştırma süreçleri ile ilgili verilerin toplandığı üniversite kütüphanesi otomasyonu araştırma için bir araç olarak kullanılmıştır. Çalışmada kullanılan veriler bu otomasyonda biriken büyük web günlük dosyalarından elde edilmiştir. Veri madenciliği ve veri analitiği gibi konular kütüphane kullanıcı hizmetleri ilişkisi kapsamında ele alınmıştır. Çevrimiçi araştırma ortamındaki kullanıcı verileri istatistik ve veri madenciliği yöntemleri ile analiz edilerek kullanıcıların davranışları incelenmiştir. Ayrıca araştırma süreci verileri kapsamında araştırma süreci takibini gösteren uyarı verebilecek aylık değişkenler ve ölçümler geliştirilmiştir.

1.3 Araştırmanın Problemleri ve Alt Problemler

Araştırmanın temel problemi, kütüphane kullanıcılarının bilgi erişim süreçlerini kapsayan ve otomasyon ssistemi aracılığıyla elde edilen web günlük verilerinin veri

madenciliği yöntemleri ile analiz edilerek kullanıcıların bilgi arama davranışlarının modellenmesi ile kullanıcıların erişim süreçleri iyileştirilebilir mi? Ayrıca, belirli bir araştırma sürecinden ayrılan kullanıcıların erken tespit edilmesine yardımcı olup olmayacağını incelemek amaçlanmıştır. Bu temel problem çerçevesinde belirlenen araştırma problemleri aşağıdaki gibidir.

1. Bir bilgi arama sürecinin başlama ve bitme süreleri arasında bilgiye erişimin gerçekleşip, gerçekleşmeme ve sürecin tahmini yapılabilir mi?
2. Oluşturulan farklı sınıflama modellerinin süreçlere ilişkin araştırma performansını tahmin etme başarısı nasıldır?
3. Farklı veri dönüştürme tekniklerinin ön işleme sürecinde sınıflandırma performansı üzerindeki etkileri nasıldır?
4. Farklı özellik seçme tekniklerinin ön işleme sürecinde sınıflama performansı üzerindeki etkileri nasıldır?
5. Kullanıcının araştırma performansının tahmininde hangi değişkenler önemlidir?
6. En iyi performans sergileyen sınıflama algoritması ve ön işleme yöntemleri kullanılarak, dersten başarısız olacak kullanıcıların önceden tahmin edilmesi mümkün müdür?
7. Bilgi arama sürecini sonlandıran araştırmacıların verileri kullanılarak benzer kullanım profillerine sahip kullanıcılar gruplanabilir mi?
8. Farklı kümeleme algoritmalarına göre kullanıcılar kaç farklı gruba ayrılmaktadır ve küme gruplarına göre bu kullanıcılar nasıl tanımlanabilir?
9. Farklı kümeleme algoritmalarına göre araştırma sürecini tamamlayanlar, tamamlamayan kullanıcıların kümelerine göre dağılımları incelendiğinde hangi algoritma araştırma sürecini tamamlayanlar, tamamlamayanları ayırt etmede daha iyi performans gösterir?

1.4 Çalışmanın Sınırlılıkları

Bilgi merkezlerinde bilgi teknolojilerinin getirmiş olduğu yeniliklerle kullanılan yazılımlar, otomasyon sistemleri vb. araçlar sonucunda bilgi merkezi kullanıcılarına ve kaynaklara yönelik çok büyük miktarlarda çeşitli yeni veriyi meydana getirmiştir. Diğer

birçok alanda olduğu gibi bilgi merkezleri de sunmuş oldukları hizmet kapsamında ortaya çıkan bu devasa boyuttaki veriden faydalı bilgi dönüşümü için yeni yöntem/ler arayışında olmuşlardır. Veri madenciliğinin getirmiş olduğu uygulamalar bu süreçte bilgi merkezlerinin de dikkatini çekmiş ve kullanılmıştır. Literatürde özellikle Üniversite kütüphaneleri tarafından kullanımı ile ilgili Altay, bu kütüphanelerin veri madenciliği uygulamalarını kullanarak ham veriden faydalı bilgi çıkarılıp kullandığının önemi ve yaygınlaştığını dile getirmiştir (Altay, 2018).

Üniversite kütüphanelerindeki veri madenciliği süreci aşağıdaki adımları takip edilmesi ile gerçekleştirilebilir (Uçan, 2010):

• **Odaklanacak alanların tespit edilmesi**: Üniversite kütüphanelerinde veri madenciliği alanındaki ilk aşama, madencilik yöntemlerinin uygulanacağı belirli alanların, madencilik sürecinin genel hedefleri doğrultusunda belirlenmesidir. Madencilik çabası, belirli bir sorunla sınırlanabilir veya alternatif olarak, karar verme süreçlerini ve ilgili bilgilerin edinilmesini sağlamak amacıyla kapsamlı bir tarama yapılabilir. Dolayısıyla, veri madenciliği faaliyetlerinin gerçekleştirilmesinden önce, madenciliğin denetimli mi yoksa denetimsiz mi olacağı belirlenmelidir. Madenciliğin denetimli olduğu, yani bir sorun odaklı bir yaklaşım izlediği durumlarda, önceden bir stratejik yol haritası hazırlanmalıdır. Örneğin, mali kısıtlamalarla karşılaşan bir kütüphanede, yönetim veri madenciliği tekniklerini kullanarak hangi alanlarda kesinti yapacağına veya tasarruf sağlayacağına karar verebilir. Belirli bir sorun olmadan da, kütüphanenin mevcut durumu hakkında bilgi toplamak amacıyla genel bir veri madenciliği değerlendirmesi yapılabilir. Ancak, böyle bir veri madenciliği çabası birçok zorluk ve komplikasyonla karşılaşabilir. Kütüphanenin tüm verilerinin veri madenciliği işlemlerine tabi tutulması, madencilik işlemlerinden önce yapılması gereken veri temizleme ve uyarlama süreçlerinin önemli ölçüde zaman alacağı anlamına gelir. Özellikle olasılık tabanlı veri madenciliği modellerinin kullanıldığı durumlarda, güçlü hesaplama kaynaklarının kullanılması gereklidir.

• **Veri kaynaklarının belirlenmesi:** Veri madenciliği ile ilgili problem veya hedefin belirlenmesinin ardından, bir sonraki aşama uygun veri kaynaklarının tespit edilmesidir. Üniversite kütüphaneleri bağlamında, veri madenciliği işlemleri genellikle işlemsel, entegre edilmemiş ve düşük seviyeli verilere dayanmaktadır. Bu durum, büyük veri hacimlerinin saklanması, yüksek depolama maliyetleri ve kütüphane kullanıcılarının

gizliliğinin korunması gibi faktörlere bağlı olarak, kütüphane yöneticilerinin bu tür verileri saklama konusunda tereddüt etmelerine neden olmaktadır. Bununla birlikte, veri madenciliği açısından işlemsel veriler, kütüphane varlıkları kadar önemlidir. Bu nedenle, üniversite kütüphanelerinde veri depolama sistemlerinin kurulması kritik bir gereklilik olarak ortaya çıkmaktadır. Üniversite kütüphanelerindeki dijital veri kaynakları çerçevesinde iki veri kategorisi incelenebilir. İlk kategori, kütüphane sisteminde yer alan iç verilerdir. Bu veriler bibliyografik veriler, yazar bilgileri, süreli yayınlar veya tezler, işlemsel metrikler ve web arama günlüklerini içermektedir. Yukarıda belirtilen veriler, karmaşık ve zorlu süreçler aracılığıyla veri madenciliği için uygun formatlara dönüştürülebilir. İkinci kategori ise dışsal verilerdir. Bu veriler, dış kaynaklardan temin edilerek oluşturulan, coğrafi konum, demografik istatistikler ve benzer veriler gibi daha geniş bilgileri kapsamaktadır.

• **Veri ambarının oluşturulması:** Akademik kütüphanelerde veri madenciliği operasyonları bağlamında, çeşitli kaynaklardan gerekli bilgilerin toplanması gerekliliği, veri ambarlarının kurulmasını kritik bir görev haline getirmektedir. Geliştirilen veri ambarı, yalnızca birincil veri kaynağını değil, aynı zamanda yardımcı kaynaklardan elde edilen yaygın ilişkisel alanları da dikkate almaktadır. Veri ambarında bir araya getirilen veriler, titizlikle temizlenmiş ve dönüştürülmüş operasyonel verilerdir. Bir veri ambarının oluşturulmasını kolaylaştırmak için, verilerin, kütüphane yöneticisi veya atanmış yetkililer gözetiminde titizlikle seçilmesi büyük önem taşımaktadır; bu, sürecin sorunsuz bir şekilde yürütülmesini sağlamak için gereklidir. Veri temizleme, düzenleme ve yeniden yapılandırma aşamalarının ardından, işlem otomatik olarak gerçekleştirilir.

• **Veri ambarının yapılandırılması:** Veri madenciliği süreçlerinde en fazla zaman ve çaba gerektiren aşamalar, veri ambarının oluşturulması ve yapılandırılmasıdır. Bu aşamaların gerektiğinde tekrar edilmesi gerekebilir. Ancak, bir kez oluşturulup yapılandırılan veri ambarı, ilerleyen veri madenciliği uygulamaları için bir temel teşkil eder. Bu durum, gelecekte yapılacak çalışmalar için önemli bir zaman ve iş gücü tasarrufu sağlar. Başka bir ifadeyle, veri madenciliği süreçlerinde kullanılan algoritmalar, farklı uygulamalarda da yeniden kullanılabilir, böylece süreçlerin verimliliği artırılabilir.

• **Uygun veri madenciliği araçlarının belirlenmesi:** Veri ambarı oluşturulduktan ve yapılandırıldıktan sonra analiz aşamasına geçilir. Bu aşamada, yalnızca geleneksel istatistiksel raporlar oluşturmakla yetinilmeyip, verilerdeki gizli ve faydalı örüntüler de

keşfedilebilir. Bu örüntüler, özellikle üniversite kütüphaneleri gibi alanlarda anlamlı verilerin elde edilmesi açısından büyük bir öneme sahiptir. Örneğin, kütüphane kullanıcılarının kullanım yoğunlukları belirlenerek personelin verimli çalışma saatleri tespit edilebilir. Benzer şekilde, kitapların sirkülasyon bilgileri incelenerek kütüphane koleksiyonlarının düzenlenmesi veya kitap yerleşim sisteminin yeniden yapılandırılması sağlanabilir.

• **Analiz ve uygulama:** Analizlerin tamamlanmasının ardından karar verme modelleri geliştirilir ve bu modellerin doğruluğu onaylanmalıdır. Veri madenciliği sonuçlarının onaylanması, genellikle kütüphane yöneticileri tarafından gerçekleştirilir. Elde edilen sonuçların değerlendirilmesi, bu sonuçlara dayalı adımların ve uygulamaların planlanması, ayrıca bu planların uygulanabilirliğinin test edilmesi model geliştirme süreci açısından kritik bir aşamadır. Eğer elde edilen örüntüler, kütüphane yöneticileri veya ilgili personel tarafından gerçekçi bulunmazsa, uygun olmayan örüntülerin neden oluştuğu analiz edilmeli ve modeller yeniden test edilmelidir. Üniversite kütüphanelerindeki veri madenciliği süreçlerinin son aşaması, geliştirilen model ve örüntülerin tüm kütüphane sistemi genelinde uygulanmasıdır. Sistemin sorunsuz bir şekilde çalışması için değişkenlerin bir süre izlenmesi ve modellerin sürekli olarak iyileştirilmesi faydalı olacaktır. Sistemin dinamik bir yapıya sahip olması ve değişkenliğin korunması, sürdürülebilir bir yapı için gereklidir.

1.5 Araştırmanın Yöntemi

Yukarıdaki problemler doğrultusunda kullanıcıların araştırma süreçleri ile ilgili verilerin toplandığı Yüzüncü Yıl Üniversitesi Ferit Melen Merkez Kütüphanesi otomasyonu araştırma amacı çerçevesinde incelenip kullanılmıştır. Veri Madenciliği ve Veri Analitiği konularının kütüphane kullanıcı hizmetleri ilişkisi ele alınmış, elde edilen araştırma ortamındaki veriler veri madenciliği yöntemleri ile analiz edilmiş, kullanıcıların bilgi arama-erişme-kullanma davranışları incelenmiştir.

1.6 Araştırmanın Evreni ve Örneklemi

Çalışmada araştırma problemleri doğrultusunda, araştırmanın evrenini bütün kütüphanelerdeki kullanıcıların kullanım verileri; örneklemini ise YYÜ kütüphanesinden elde edilen veriler oluturmaktadır. Verilerimiz YYÜ Kütüphanesinin mevcut otomasyon sisteminden temin edilmiş ve elde edilen veriler işlenmeye uygun bir bilgi haline getirilmiştir.

1.7 Araştırmanın Düzeni

Araştırma yedi bölümden oluşmaktadır.

Giriş bölümü kısmında araştırmanın konusu, amacı, kapsamı, yöntemi, önemi, kavramsal çerçevesi, kapsamı, evreni ve düzeni ve araştırma esnasında yararlanılmış olan kaynaklar ele alınmıştır.

Tezin geri kalanı aşağıdaki şekilde düzenlenmiştir. Bölüm 2'de araştırmamızla ilgili önceki çalışmalar açıklanmaktadır. Bölüm 3 ve 4'de ilişkili konular, ön materyaller ve notasyonlar açıklanmaktadır. Bölüm 5'te önerdiğimiz yaklaşım ayrıntılı olarak açıklanmaktadır. Bölüm 6 deneysel sonuçları, deneylerimizde kullanılan veri kümesini ve performans sonuçlarının değerlendirmesini sunmaktadır. Son olarak, Bölüm 7'de sonuçlar ve bu çalışmanın gelecekteki çalışmalar için potansiyel yönleri işaret edilmiştir.

2. KAYNAK BİLDİRİŞLERİ

Bu bölümde, tez çalışmasının altyapısıyla örtüşen, farklı veri madenciliği yaklaşımlarını kullanan kütüphane ve bilgi merkezlerinde yürütülmüş olan ilgili araştırmalar sunulmuştur.

Bilgi ve belge yönetimi alanında kullanıcıların bilgi arama, Bilgi Erişim Sistemi (BES) kullanmasıyla ilgili çalışmaların 1900'lü yılların başlarından günümüze kadar araştırılmaya devam edilen önemli konulardan biridir. Literatürde araştırmacılar kullanıcıların bilgi arama süreçlerini veya bilgi arama davranışlarını konu alan çeşitli araştırmalar yapmıştır. Günümüz bilgi merkezlerinin temeli ileri teknolojilerin yoğun ve etkin kullanımına bağlı durumdadır. Son teknolojiler arasında bulunan giyilebilir teknoloji araçları, mobil internet, akıllı işleme teknolojisi, bulut bilişim ile birlikte yeni nesil bilgi merkezlerinin gelişiminin özünde veri madenciliği, yapay zekâ ve nesnelerin interneti (IoT) bulunmaktadır (Kassim ve Zakaria, 2006).

Veri madenciliği uzun zamandır bilgi ve belge yönetimi çalışmaları içerisinde tartışılmaktadır. Bilgi merkezleri, yönetim kararlarına yardımcı olmak için veri madenciliğini kullanmaktadır (Cullen, 2005). Cox ve Jantti (2012), Wollongong Üniversitesi kütüphanesinde, kütüphane kullanımı ile daha yüksek öğrenci notları arasındaki güçlü korelasyonu göstermek için kütüphane verilerini üniversite demografik ve akademik performans verileriyle birleştirerek bir çalışma gerçekleştirilmiştir. Bazı bilgi bilim araştırmacıları, bilgi merkezlerinde veri madenciliğinin, bilgi merkezi materyalleri için öneri hizmetleri sağlamak amacıyla kullanmanın önemine dikkat çekmişlerdir (Chen ve Chen, 2007a; Kovacevic vd. 2010; Tsai ve Chen, 2008). Veri madenciliği yöntemleri kullanılarak, kullanıcılar için kişiselleştirilmiş birçok hizmet üretebilir (örneğin, kullanıcılara anlık bilgi göndererek) ve karar alma süreçlerine yardımcı olabilirler. Bununla birlikte, farklı kaynaklar arasında (örneğin, sağlıklı beslenme kitapları ile spor alanı kitapları arasında) bağlantılar kurabilirler ve bu şekilde kullanıcılar için zengin ve kişiselleştirilmiş kaynak önerilerinde bulunabilir (Renaud vd. 2016).

Long ve Wu (2012b), eğitim ve araştırma faaliyetlerinin merkezinde yer alan akademik bilgi merkezlerine yönelik, hizmetlerinin daha kolay erişimi için veri tabanı teknolojilerini kullandıklarına dikkat çekmektedirler. Akademik bilgi merkezleri sürekli

olarak çok yönlü verilerle ilgilenen bilgi merkezleridir. Yaşanan dijital bilgi çağında veri madenciliği teknolojisi, veri araştırmalarında bilgi merkezlerinin üzerindeki işlevsel etkisi nedeniyle, veri ve özelliklerinin verimli bir şekilde nasıl çıkarılacağı ve kullanıcılara kaliteli kişisel hizmetlerin nasıl sağlanacağı, akademik kütüphanelerde önemli konular olmuştur (Duan ve Wang, 2021). Bilgi merkezleri de tıpkı diğer birçok alan gibi, yeni bilgi teknolojilerindeki her yeni gelişimle hizmetlerini istatistiksel ve niteliksel olarak artırmaktadır. Öte yandan, devasa miktarda bilgiye erişim söz konusu olduğunda yerleşik araç ve teknikler kullanıcılara hızlı ve yeterli yanıt veremeyebilmektedir. Veri madenciliği gibi araçlar ve yaklaşımlar, verilere erişmek ve değerli bilgileri verimli bir şekilde çıkarmak için giderek daha önemli hale geliyor. Bunlardan en önemlisi, yönetim tercihlerinin kalitesini artırmak için büyük miktarda veriden bilgi elde etmektir (Uppal ve Chandwani, 2013). Bu denli büyük veri hacminde farklı parametreler arasındaki ilişkiler, desenler ve bağlantılar açık şekilde görülmemektedir. Bu nedenle modern veri madenciliği teknolojisi bilgi merkezleri ve kütüphane yönetim sistemleri için gereklidir (Zhang, 2018).

Kullanıcıların bilgi ihtiyaçlarını yönetmek ve anlamak için veri madenciliği kullanıldığında yararlı bilgiler ortaya çıkarılabilir ve sağlanabilir (Suresh vd. 2018b). Bilgiye dayalı kararlarda önceki kullanıcıların davranışlarına uygun bir araç olarak son teknolojiler ve veri bazlı sonuç üreten teknolojiler bilgi merkezlerinin işlerini kolaylaştırabilir (Mishra ve Mishra, 2013). Silwattananusarn ve Kulkanjanapiban (2020), akademik kütüphanenin kullanıcıların kullanım davranışlarını, yani kullanıcıların kitap ödünç alma davranışlarını inceledi. Veri madenciliği birliktelik kuralları ve kümeleme için kullanılan araç WEKA'dır. Etkin yönetim ve kütüphane hizmetleri için örnek bir model sağlarlar. Veri madenciliği, ödünç verme sürecini analiz ederek kullanıcılara ilgili kitapları sağlayabilir, bu da bilgi merkezlerinde kaynak kalitesinin iyileştirilmesine ve yönetiminin kolaylaştırılmasına yol açabilir (Yi vd. 2018). Bu tür pek çok çalışma, bilgi merkezlerindeki önerici ve akıllı sistemlerde veri madenciliği tekniklerinin kullanıldığını göstermektedir (Chen ve Chen, 2007a; Jomsri, 2014; Liu, 2018; Tsuji vd. 2012b; Yi, vd. 2018). Prehanto vd. (2020), kütüphane kitaplarının modellenmesinde, birliktelik kuralları (Apriori algoritması) kullanılarak kitap yerleşiminin belirlenmesinde ve kitap ödünç verme analizinde, ortalama destek (%6) ve güven (%67-100) değerlerine sahip birliktelik kurallarının kullanıcıların ödünç alma işlemlerini tahmin edebildiğini ortaya koymuştur.

Zhou (2020), bir kitap öneri sisteminin tasarlanması ve uygulanmasında birliktelik kurallarını (Apriori algoritması) kullanmış ve algoritmayı geliştirerek iyi performansa sahip bir sistemin kullanıcılara kişiselleştirilmiş bilgileri etkili bir şekilde önerebileceğini göstermiştir.

Büyük bilgi koleksiyonlarından kullanıcı merkezli bilginin elde edilmesi amacıyla Bhopale ve Tiwari (2020) tarafından gerçekleştirilen bir çalışmada, sürü zekâsı destekli optimize edilmiş küme tabanlı bir model önerisi yapılmıştır. Bu model, aynı zamanda veri madenciliğinde çoğunlukla market sepet analizlerinde kullanılan örüntü madenciliği (pattern mining) tekniğini de içermektedir. Bilgi kaynaklarının kümelenmesi işleminin, yazarlar tarafından önerilen ve denetimsiz öğrenme altyapılı bir "K-Flock" isimli bir algoritmayla gerçekleştirildiği bu çalışmada, önerilen model sayesinde, geleneksel bilgi erişim yaklaşımlarından daha etkili ve verimli performans gösterildiği belirtilmiştir. Wang vd. (2020), verilerin etkin işlenmesi ve kullanılması sorununun çözülmesi gerektiğini belirtmiştir. Veri segmentasyonu, birliktelik analizi ve sapma analizi, veri madenciliği için kullanılan analiz yöntemleridir.

Başka bir çalışmada Rakhmanov (2019), eğitim sırasında %71.9 ve test verileri üzerinde %72 doğruluk oranına sahip karar ağacı tekniğini kullanarak, veri madenciliğinin kütüphane kullanıcılarının ödünç alma işlemlerine dayalı davranışlarını tahmin edebildiğini göstermiştir. Li vd. (2019), kullanıcının ihtiyacını belirlemek ve veri madenciliği yoluyla mevcut hizmetlerin iyileştirilmesine yönelik etkili ve verimli fikirler geliştirmek için kütüphanelerdeki dijital kaynaklar ve kullanıcı kullanım davranışı gibi büyük verilerin ve dış bilimsel verilerin analiz edildiğini belirtmiştir. Ansari vd. (2020), kütüphanelerin ve bilgi sistemlerinin verilerinin doğru bir şekilde analiz edilerek kullanıcı davranışlarının öğrenilebileceğini ve her kütüphane için bir öneri sistemi geliştirilebileceğini belirtmiştir.

Yi vd. (2018), kullanıcıların kredi işlemlerini temel alan bir çalışmada, birliktelik kuralları ve yapay arı kolonisi (ABC) algoritmasına dayalı bir öneri sistemi tasarlamaya çalışmıştır. Başka bir çalışmalarında ise, birliktelik kurallarını optimize etmek için yapay arı kolonisi (ABC) algoritmasını kullanmış ve buna dayanarak bir öneri sistemi oluşturmuşlardır. Huang vd. (2015), başka bir çalışmada FP-Growth algoritmasını kullanarak kütüphane ödünç dolaşım kayıtlarına ilişkin verileri incelemiş ve ödünç dolaşım kayıtlarının sayısının yürütme süresiyle pozitif, destek derecesi ile ise olumsuz

ilişkili olduğunu göstermiştir. Ayrıca minimum eşik sınırı azaltılarak kural sayısı artırılmış, yürütme süresi artırılarak kural sayısı azaltılmıştır. Sonuçları ayrıca birliktelik kuralları yaklaşımının büyük veri setlerinin analizinde verimlilik kaybı olmadan iyi çalıştığını göstermiştir (Huang vd. 2015). Long ve Wu (2012b), başka bir çalışmada birliktelik kuralları yaklaşımı ve Maksimum Sürü Desenleri (MFP)-Miner algoritmasını kullanarak işlem verilerini analiz etmiş ve destek ile güven derecesine sahip güçlü kurallar sağlamıştır. Kütüphanelerde en etkili kaynak önerme sistemlerinin belirlenmesine yönelik bir çalışmada, Tsuji vd. (2012b) işbirlikçi filtrelemeye, birliktelik kurallarına ve Amazon'a dayalı üç yöntem kullanılmıştır. Öğrencilerin görüşlerinin araştırılmasının sonuçları, birliktelik kuralları yaklaşımının işbirlikçi filtreleme algoritmasından daha uygun maliyetli olmasına rağmen, kütüphane kullanıcılarına kaynak önermek için en iyi yöntem olduğunu gösterilmiştir (Tsuji vd. 2012b). Yapmış oldukları bir diğer çalışmada, kullanıcıların kredi işlemlerini analiz etmek için kümeleme, zaman serisi ve birliktelik kuralları teknikleri kullanmışlardır. Kümeleme sonuçları, her kümedeki öğrenciler tarafından en çok kullanılan kitapları belirlemiş ve akademik derece düzeyi arttıkça kullanıcıların ödünç verme sürecinin mesleki alanlarına yöneldiğini göstermiştir. Ayrıca birliktelik kuralları tekniği, kullanıcıların kredi işlem sürecini belirli bir güven derecesiyle tanımlamaktadır (Yu, 2011). Chen ve Chen (2007a), Bayes ağı ve birliktelik kuralları kavramlarını kullanarak bir kitap tavsiye sistemi geliştirmiştir. Daha sonra, önerilen kitapların uygunluğunu ve kullanıcıların memnuniyetini değerlendirmek için bir anket kullanmışlardır.

Kütüphane kullanıcılarının bilgi erişim kayıtlarını inceleyen bir çalışmada, birlikte kitap alma örüntüleri araştırılmıştır. Dwivedi ve Bajpai (2007), birliktelik kuralları madenciliği kullanarak bilgi erişimin bir boyutu olan benzer yayınları belirlemiş ve bilgi erişim performansı açısından önemli bir çalışma gerçekleştirmiştir. Ma ve Lund (2020), büyük dergilerdeki makalelerin içeriğini analiz ederek, kütüphane ve bilgi biliminde veri madenciliği stratejilerinin rolünü incelemişlerdir. Kelime frekansı dikkate alınarak küme frekansı analizi yapılmış ve üç ana tema belirlenmiştir: Yayın/alıntı, bilgi/medya kullanımı ve tüketici davranışı. Yayınlardan/alıntılardan tüketici davranışlarına doğru bir geçiş olduğu anlaşılmaktadır. Papadopoulos vd. (2020), büyük verilerden mantıksal ve anlaşılır anlamlar çıkarmak, indekslemek, analiz etmek,

değerlendirmek ve yorumlamak için metin ve veri madenciliği tekniklerinin kullanıldığını belirtmiştir.

Başka bir çalışmada Li (2017), geri yayılımlı (BP) yapay sinir ağını temel alarak elektronik kütüphane kaynaklarının kalitesini değerlendirerek, bu tür bir yaklaşımın kütüphanenin elektronik kaynaklarının kalitesini yüksek doğruluk ve hassasiyetle değerlendirdiğini ve tahmin ettiğini göstermiştir. Zhang ve Wang (2016), yaptığı bir çalışmada BP yapay sinir ağını temel alarak, kütüphane okuyucularının bilgi kaynakları, hizmet içeriği, personel, hizmetler ve çevresel olanaklarla ilgili memnuniyetlerini incelemiş ve kullanıcı memnuniyeti çalışmalarında sinir ağlarının kullanılabileceğini göstermiştir. Kümeleme yaklaşımı ve k-ortalamalar algoritmasını kullanan başka bir çalışmada, Jin ve Sheng (2017), kullanıcıları dört kategoriye ayırmış ve hangi kullanıcıların kütüphane kaynaklarını kolaylıkla, hangilerinin zor kullandığını ve hangi kümelerin daha fazla ödünç verme süresine sahip olduğunu belirlemişlerdir. Roy vd. (2018) bir çalışmalarında, kütüphanenin elektronik kaynaklarındaki çeşitli öğeleri kümeleyerek, ARM tabanlı kümelemenin diğer algoritmalara göre daha doğru sonuçlar verdiğini göstermiştir. Bu nedenle, bir kümedeki öğeler, kullanıcı tavsiyesi için iyi bir kaynak olarak değerlendirilebilir. Bir diğer çalışmada, Ruixiang (2018), kitap satın almak için makul bütçe tahsisinin tahmininde BP yapay sinir ağı tekniğini kullanmış ve sinir ağı hesaplamalarının tahmin edilen sonuçlarının gerçek duruma çok yakın olduğunu, ayrıca BP yapay sinir ağının akademik kitapların oluşturulmasında önemli bir rol oynayabileceğini göstermiştir.

Kullanıcı işlemlerini tahmin etme yeteneği ve kullanıcıların erişim kalıplarını tanımlayıp işlemlerini doğru bir şekilde tahmin edebilme yetenekleri, bir dizi çalışmada gösterilmiştir (Yu, 2011; Long ve Wu, 2012b; Tsuji vd. 2012a; Chen vd. 2014; Huang vd. 2014; Krishnamurthy ve Balasubramani, 2014; Liu ve Wang, 2018; Renaud vd. 2015; Yi vd., 2018a). Ayrıca, kütüphane hizmetlerinin kalitesi (Li, 2017), kullanıcı memnuniyeti (Zhang ve Wang, 2016) ve kullanıcıların kümelenmesi (Chen ve Chen, 2006; Chen ve Chen, 2007b; Kovacevic vd. 2010; Yu, 2011; Jin ve Sheng, 2017) konularında veri madenciliği tekniklerinin kütüphane verilerinin analizinde oldukça uygun ve etkili olduğu gösterilmiştir. Bu nedenle, yukarıdaki çalışmaların incelenmesi, yöneticilerin ve politika yapıcıların, kullanıcıların ihtiyaçlarını anlaması, veri madenciliği tekniklerini kullanarak analiz yapması ve buna göre kısa ve uzun vadeli planlamalar

yapması gerektiğine dair acil bir ihtiyaç olduğunu ortaya koymaktadır. Kütüphaneler ve bilgi merkezleri, kullanıcıların ödünç verme işlemlerini, bilgi ihtiyaçlarını uzmanlaşmış bir yaklaşımla ve en az hatayla karşılamak için bu yeni teknolojinin önemine işaret etmektedir.

Yurtiçi mesleki yazın incelendiğinde, bilgi ve belge yönetimi alanında veri madenciliği kavramının ele alındığı birçok önemli çalışmanın bulunduğu görülmektedir. Bu çalışmaların veri madenciliği analizlerinin ve uygulamalarının örneklerini oluşturduğu, konu ile ilişkili çeşitli literatür incelemelerinden veya kavramsal düzeyde konunun değerlendirildiği araştımalardan oluştuklarını söyleyebiliriz. YÖK Ulusal Tez Merkezi'ni taradığımızda ise veri madenciliğiyle ilgili sekiz doktora ve yirmi lisansüstü tez çalışmasına ulaşılmıştır. Bazı çalışmaların veri madenciliği konusunu kavramsal bir çerçevede ele aldığı ve bu yaklaşımlarını anket çalışmalarıyla desteklediği görülmektedir. Diğer çalışmaların ise metin madenciliği teknik ve yöntemlerini uygulamalı bir şekilde gerçekleştiren uygulama çalışmaları olduğu gözlemlenmiştir. Yapılan çalışmalar incelendiğinde, kütüphane bilimleri ve bilgi merkezleri alanındaki çeşitli veri madenciliği tekniklerinin yönetim işlerinde kullanılmasının ve kütüphane hizmetlerinin kalitesinin artırılmasının etkinliğini göstermektedir. Bu alandaki çalışmaların çoğu birliktelik kuralları tekniğini kullanmış ve sonuçları bu tekniğin yüksek bir performansa sahip olduğunu göstermiştir.

Bu çalışmada alandaki diğer çalışmalardan farklı olarak veri madenciliği yöntemlerinin kullanılmasıyla çevrim içi ortamda kullanıcı arama davranışlarının günlük ayak izlerine karşılık gelen veriler sınıflandırılmış ve analiz edilmiştir. Önceki çalışmalardan farklı olarak, kullanıcı etkileşim verileri büyük ölçekli log dosyalarından çıkarılmış ve bu veriler, kullanıcıların bilgi erişim süreçlerindeki eğilimlerini belirlemek amacıyla çeşitli veri madenciliği teknikleri ile incelenmiştir. Özellikle, kullanıcıların bilgi arama süreçlerinde sergiledikleri davranış örüntülerinin belirlenmesi ve bu örüntüler üzerinden gelecekteki bilgi erişim başarısının tahmin edilmesi hedeflenmiştir.

Çalışmada, çevrim içi ortamda bilgi arama sürecini etkileyen faktörler belirlenmiş, kullanıcıların araştırma süreçlerindeki etkileşim yoğunlukları analiz edilerek farklı davranış grupları oluşturulmuştur. Bu kapsamda, sınıflama ve kümeleme algoritmaları kullanılarak kullanıcıların bilgiye erişim başarıları modellenmiş, bilgi erişim sürecinde olası sorunlarla karşılaşan kullanıcıların erken tespit edilmesine olanak

sağlayacak analizler gerçekleştirilmiştir. Böylece, kütüphane ve bilgi merkezleri bağlamında kullanıcı deneyiminin iyileştirilmesi ve bilgi erişim hizmetlerinin daha verimli hale getirilmesine yönelik uygulanabilir bir model önerisi sunulmuştur.

Bu çalışmanın, veri madenciliği ve öğrenme analitiği tekniklerinin bilgi ve belge yönetimi alanına entegrasyonu açısından önemli bir katkı sağlaması beklenmektedir. Özellikle, büyük veri analizi ve yapay zekâ tabanlı tahminleme modellerinin kütüphane hizmetleriyle bütünleştirilmesi, kullanıcıların araştırma süreçlerinde daha verimli bilgiye erişmelerine yardımcı olabilecek yeni yöntemlerin geliştirilmesine olanak tanıyacaktır.

3. ARAŞTIRMANIN KAVRAMSAL TEMELİ

3.1 Temel Kavram ve Tanımlar

Bu bölümde çalışmanın konusuna uygun olarak kavramsal tanımlamalar yapılmış, teknik detaylar ve terimlere açıklık getirilmiştir.

3.1.1 Bilgi

Hemen her çağda sosyal bir varlık olan insanların yaşamında vazgeçilmez öneme sahip olan bilgi, çok farklı şekillerde tanımlanmaktadır. Sözlük anlamıyla bilgi: "Öğrenme, araştırma ve gözlem aracılığıyla varılan gerçekler, kavrayışlar, malumat, vukuf" şeklindedir (Türk Dil Kurumu, 2024). Bir başka tanımı ise "belirli bir düzen içerisindeki tecrübelerin, değerlerin, amaca yönelik enformasyonun ve uzmanlık görüşünün yeni deneyimlerin ve enformasyonun birleştirilerek değerlendirilmesi amacı ile çerçeve oluşturan bir bilişimdir" (Davenport ve Prusak, 1998). Bilgi, bir alandaki anlaşılır ve iletilebilir verilerden oluşan enformasyonun ürünüdür (Gürdal, 1991). Farklı bakış açıları, yaklaşımlar veya alanlara göre bilginin tanımları değişebilmektedir.

Bireyler yaşamları süresince hemen her dönemde karşı karşıya kaldıkları olaylara ait çeşitli verileri kullanma ve anlamlandırma çabası içerisinde olmuşlardır. Farklı yaş ve uğraş grubunda ki insanlar, yaşadıkları çevreyi daha iyi tanımak, çevreyle uyumlu bir biçimde yaşamak, yeni şeyler üretmek veya var olan şeyleri iyileştirmek için veri (bilgi) kullanımına başvurmuşlardır. Diğer taraftan bilindiği üzere kültürel birikim çabası içerisinde bulunan insan sürekli bir arayış içerisindedir. Bu süreci elde ettikleri bilgileri organize etme, toplama ve farklı amaçla kullanma ve paylaşma adımlarının izlediğini görmekteyiz. Bilgi her geçen gün insan yaşamındaki önemini korumuş ve günümüzde bilgi teknolojileriyle birlikte kullanılarak bilgi merkezleri hizmetleri alanı da dâhil olmak üzere insan yaşamının çeşitli yönleri üzerinde önemli bir etkiye sahip olmuştur (Mathar vd. 2022). Bilgi teknolojisinin gelişimi, bilgiye erişme, onu yönetme ve yayma biçimimizde devrim niteliğinde değişikliklere götürmüştür. Geçmişte, bilgi merkezleri genelde düzgün bir şekilde düzenlenmiş fiziksel kitap koleksiyonlarıyla özdeşleştirilirdi. Günümüze gelindikçe bilgi teknolojisinin gelişmesiyle yalnızca kitap depolama alanı

olmaktan çıkmış, çeşitli dijital kaynaklara hızlı ve verimli bir şekilde erişilebilen dinamik bilgi merkezleri haline gelmiştir. Bilgi teknolojisinin kullanımıyla bilgiye erişilebilirlik önemli ölçüde artmıştır. İnternet ve elektronik veri tabanları aracılığıyla, araştırmacılar kitapları, bilimsel dergileri, makaleleri ve diğer multimedya kaynakları da dâhil olmak üzere çeşitli türdeki materyalleri kolayca arayabilir ve bunlara erişebilirler. Bu, kütüphane kullanıcılarının ilgili bilgileri daha etkili ve verimli bir şekilde elde etmelerini sağlar (Buana ve Linarti, 2021).

Bunların sonucunda, Kütüphane ve Bilgi Sistemlerinin (LIS) geliştirilmesinde, örneğin yapılandırılmamış verilerin toplanması, düzenlenmesi, depolanması, işlenmesi ve kullanımı konusunda büyük zorluklar ortaya çıkarmaktadır (Bowker, 2018). Buna bağlı olarak bir taraftan mevcut verilerin hacmi ve karmaşıklığını arttırmış diğer taraftan bilgi miktarında yaşanan hızlı artış ve bilişim teknolojilerinde yaşanan ilerlemeler ve gelişmeler bilgi erişim sürecinde kullanıcıların bilgi arama davranışlarını da değiştirmiştir.

3.1.2 Bilgi Davranışı

Bireylerin bilgi kaynakları ve bilgi erişim kanallarıyla olan ilişkileri (bilgi arama, kullanma, dağıtma) bilgi davranışı (information behavior) olarak nitelendirilmektedir (Wilson, 2000).

1960'lı yıllarda alanda yaygınlaşan kullanıcı araştırmaları ile birlikte ortaya çıkmış olan bilgi davranışı kavramı günümüzde bilgibilim alan çalışmaları için kapsayıcı bir terim olarak karşımıza çıkmaktadır (Wilson, 1994). Geleneksel yapılarda sistemi merkeze alan anlayışın yerini 1980'li yıllardan sonra kullanıcı yönlü yaklaşım ile tasarlanmış anlayışa bırakmıştır. Çağdaş tartışmalarda, kullanıcılara ve onların karşılaştığı sorunlara ilişkin geleneksel yaklaşımlar büyük bir dönüşüm geçirmiştir. Kullanıcılar artık yalnızca bir sistemin ayrılmaz parçaları olarak değil, davranış bilimleri, psikoloji, sosyoloji ve bilgi yönetimi gibi çeşitli disiplinlerin perspektiflerinden değerlendirilen bir araştırma odağı olarak görülmektedir. Kullanıcıları yalnızca bir sistemin unsurları olarak incelemek yerine, araştırmalar giderek kullanıcıları kendi doğal ortamlarında etkileyen faktörlere odaklanmıştır (Yıldız ve Uçak, 2014). Bu dönüşüm, sistem merkezli çerçevelerden kullanıcıyı önceliklendiren paradigmalarına geçişi

sağlamıştır. (Wilson, 2000). Böylece sistem yapısalı kadar önemli bir diğer etken olarak kullanıcı özellikleri ve beklentileri ağırlık kazanmış ve çalışmalar bu yöne kaymıştır. Bireysel bilgi davranışları alanında önemli modeller ve kuramlar geliştirilmiştir (Yıldız ve Uçak, 2014).

1980'lerden itibaren bilgi davranışı araştırmalarının evriminde belirleyici bir faktör, teknolojinin gelişimi olmuştur. Özellikle bilgisayarların ve İnternet'in yaygınlaşmasıyla birlikte bilgi teknolojilerindeki yenilikler, bilginin üretildiği, saklandığı ve erişildiği bağlamları ve biçimleri köklü bir şekilde değiştirmiştir. Bunun sonucunda kullanıcılar, bilgi arama süreçlerinde kütüphaneleri fiziksel olarak ziyaret etmek yerine elektronik bilgi erişim sistemlerini tercih etmeye başlamıştır (Sonnenwald ve Iivonen, 1999).

Bu bağlamsal değişim, kullanıcıların dijital ortamda bilgi arama yöntemleri, tercih ettikleri bilgi kaynakları ve karşılaştıkları engeller üzerine artan bir incelemeyi beraberinde getirmiştir. Bilgi sistemlerinin tasarımı ve değerlendirilmesinde kullanıcı merkezli bir bakış açısının benimsenmesi ve sorunların bu perspektiften çözülmesi büyük önem kazanmıştır (Uçak ve Al, 2000). Zamanla, bilgi teknolojilerinin insan bilgi davranışlarını nasıl şekillendirdiğini ve etkilediğini anlamaya yönelik ilgi artmıştır. Dolayısıyla, bilgi davranışı araştırmaları, bilgi merkezlerinden bağımsız düşünülemez. Yeni teknolojileri hizmetlerine entegre eden ve daha etkili çözümler geliştiren bilgi merkezleri, bilgi arama faaliyetlerinin şekillenmesinde kritik bir rol üstlenmiştir.

Bilgi merkezlerinin ve sistemlerinin temel amacı, kullanıcıların bilgiye etkin bir şekilde erişimini sağlamaktır. Kullanıcıların bilgi davranışları, hizmetlerden memnuniyet düzeyleri ve bireysel ihtiyaçlarının incelenmesi yoluyla onların daha iyi anlaşılması; koleksiyon geliştirme, hizmet planlama, tasarım, iyileştirme ve değerlendirme gibi çeşitli görevler açısından vazgeçilmez hale gelmiştir. Ayrıca, bilgi davranışı üzerine yapılan araştırmalar, ulusal ve yerel düzeyde bilgi politikalarının oluşturulmasına ve bilgi erişim yöntemlerinin optimize edilmesine yönelik önemli bulgular sağlamaktadır. Bu araştırmalardan elde edilen sonuçlar, kullanıcı ihtiyaçlarına göre uyarlanmış, kullanıcı merkezli hizmet ve sistemlerin tasarlanmasını mümkün kılmaktadır (Najjari, 2010; Uçak, 1997).

Kullanıcıları anlamayı amaçlayan çalışmalarla yakından ilişkili olan bilgi davranışı araştırmaları, 1950'li yıllarda başlamıştır (Uçak ve Al, 2000). İlk başlarda

sistem merkezli bakış açısını benimseyen yaklaşımlar, 1980'lerde kullanıcının önemine vurgu yapan bir anlayışa evirilmiştir. Bu paradigma değişimi, kullanıcı özellikleri ve beklentilerinin analizine odaklanarak bireysel bilgi davranışlarını ele alan temel teori ve modellerin geliştirilmesiyle sonuçlanmıştır (Yıldız ve Uçak, 2014).

3.1.3 Bilgi Tarama Davranışı

Bilgi davranışının önemli bir boyutu olan bilgi arama davranışı, genel olarak bir kullanıcı ile bir bilgi sistemi arasındaki etkileşim unsurlarını ifade eder (Spink ve Cole, 2004). Bilgi arama davranışı yalnızca insan-bilgisayar etkileşimlerini değil, aynı zamanda bilişsel süreçleri de kapsar. Örneğin, bağlantılara tıklamak, Boole arama stratejileri kullanmak, erişilen bilgi kaynaklarının uygunluğunu değerlendirmek ya da iki bilgi kaynağından hangisinin daha faydalı olduğuna karar vermek gibi eylemler bilgi arama davranışlarına örnek teşkil eder. (Wilson, 2000). Bu etkinlikler, kullanıcının bilgi edinmeye yönelik eylemlerini yansıtmaktadır.

Teknolojik gelişmeler, özellikle bilgisayarların ve internetin yaygınlaşması, bilginin üretilmesi, depolanması ve sunulma alanlarını ve formatlarını önemli ölçüde yeniden şekillendirmiştir. Bunun sonucunda, kullanıcılar bilgi aramak için geleneksel kütüphaneleri ziyaret etmek yerine, giderek daha fazla elektronik bilgi erişim sistemlerini kullanmaya başlamışlardır (Sonnenwald ve Iivonen, 1999).

Bu bağlamda, araştırmacılar kullanıcıların elektronik ortamlarda bilgi arama yöntemlerini, tercih ettikleri bilgi formatlarını ve arama sırasında karşılaştıkları hataları incelemeye başlamışlardır. Bilgi sistemlerinin tasarımı ve değerlendirilmesi, giderek daha fazla kullanıcı perspektifine odaklanmış ve sorunlara bu açıdan yaklaşılarak etkili çözümler üretilmiştir (Uçak ve Al, 2000). Bakıldığında kullanıcılar bilgi davranışı süreci içerisinde sürekli olarak gelişmiş olan yeni teknolojilerin güdümünde kalmaktadırlar. Kullanıcılar bilgi davranışları sürecinde sürekli gelişen yeni teknolojilerle etkileşimde bulunmaktadır. Sonuç olarak, bilgi teknolojisinin bilgi davranışlarını şekillendirme ve etkilemedeki rolüne ilişkin araştırmalar önemli ölçüde artış göstermiştir. Bilgi teknolojisinin bilgi davranışları üzerindeki etkisi, bilgi merkezlerinin rolünden bağımsız düşünülemez. Yeni teknolojileri hizmetlerine uyarlayarak, bilgi merkezleri bilgi davranışlarını şekillendirme ve geliştirme sürecinde kritik bir rol oynamışlardır.

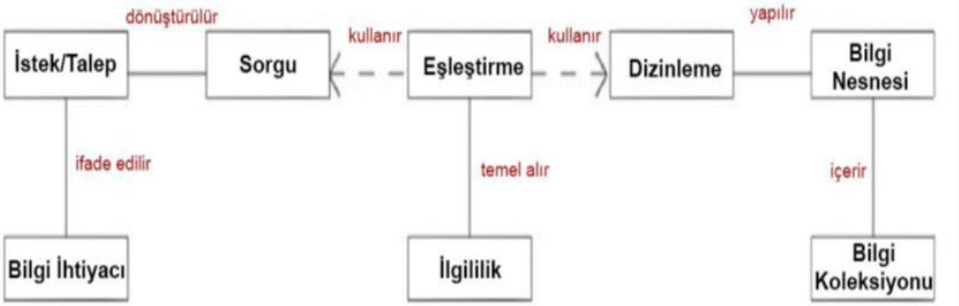
Ülkemizde örneklerin yoğunlukla geleneksel yöntemlerle sunulan bilgi hizmetleri kapsamının dijitale uyarlamasında danışma, eğitim, ödünç verme, kaynak sağlama gibi hizmet yönlü yeni modeller ya da uygulamalar biçiminde olduğunu görmekteyiz. Bilgi merkezlerinin etkili bilgi erişim (Information Retrieval - IR) rolü üstlenen yapısı gereği yeni teknoloji imkânlarını kullanarak kullanıcı davranışları çerçevesinde çevrimiçi bilgiye eriştiren ortamları kullanıcı istatistiklerinin işlenmesi, veri madenciliği tekniklerinin kullanılması, kullanıcı tiplerine göre uyarlanan sistemlerle daha etkileşimi yüksek ve verimli bir öğrenme ortamına dönüştürülebilir.

3.2 Bilgiye Erişim

"Bilgi erişim" (Information Retrieval) terimi ilk olarak 1951'de Mooers tarafından önerilmiştir (Ehsanifar, 2016). Türk Dil Kurumu (2024), "bellekte saklı verilerden belli bir konuda bilgi alma yöntem ve yordamları" şeklinde tanımlanmıştır. Tonta (2001), araştırmasında bilgi erişimine yönelik olarak, ihtiyaç duyulan bilginin toplanması, sınıflandırılması, erişiminin sağlanması, kataloglaması ve depolaması sonrasında biriken büyük veri yığınlarında arama yapılması ve istenilen bilgilerin elde edilmesi veya sunulması süreci şeklinde ortaya koymuştur. Bilgi erişim, bir dizi bilginin kullanıcının bilgi ihtiyaçlarını karşılamak için işlendiği, depolandığı, alındığı ve yayıldığı bir süreci ifade eder. Bilgi erişim işlemi manuel ve elektronik olmayan bir süreç olsa da (örneğin, bir kitaptan bilgi almak için bir dizinden yararlanmak gibi), genellikle elektronik olarak depolanana veya derlenen bilgi alma/arama sürecinin bilgisayar aracılığıyla yapıldığında kullanılmaktadır (Seyyeddokht ve Emami, 2021). Kavramla ilgili önemli bir diğer vurguda ise bilgi ihtiyacını daha belirginleştirecek şekilde "kullanıcıların ilgi duydukları bilgiye hızlı erişimlerine odaklanmış bir alan" olması yönündedir (Baeza-Yates ve Ribeiro-Neto, 2011). Bilgi erişimde amaç, kullanıcının bilgi ihtiyaçlı arayışında ona çok yönlü bir ilgi, ilişki işlemi sonrasında kullanacağı hedef bilginin sunulmasıdır (Ceri vd. 2013). Kesin arama hedefleri belirli olan kullanıcıların hedef bilgilerini hızlı ve doğru bir şekilde bulmalarını sağlamayı amaçlamaktadır.

Yeni bilgi teknolojilerinin kullanılması ve bilgi toplumlarının yaşamış oldukları erişim zorlukları, kütüphanelerin geleneksel rolü değiştirmiş ve bilgisayar becerilerinin etkin kullanımı ile eriştirme, öğretme ve oluşturma gibi işlevler daha belirgin hale

gelmiştir. Bilgi erişim içerisinde bilgi ihtiyacı, arama yapma ve erişim gibi başlıklar yer almaktadır (Göker ve Davies, 2009).

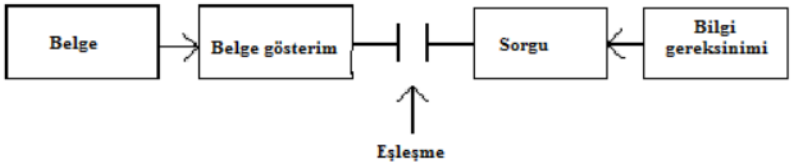


Şekil 3.1 Bilgi erişimde gerçekleşen etkileşim

Şekil 3.1'de görüleceği gibi, bilgi erişim sürecinde iki temel taraf yer almaktadır. Bunlar: bilgiye ihtiyaç duyan kişi ve bu ihtiyacı karşılayabilecek bir bilgi koleksiyonu olmaktadır. Bilgi ihtiyacı, bir talep olarak ortaya konmakta ve araştırma yapmak amacıyla bir sorguya dönüştürülmektedir. Bu sorguya yanıt verebilecek bilgileri içeren bir koleksiyonun var olması gerekmektedir. Gerekli bilginin koleksiyonda bulunabilmesi için, bu bilginin uygun bir şekilde temsil edilmesi ve dizinlenmesi gereklidir. Bilgiye erişimdeki önem ise, kişinin ihtiyaç duyduğu bilgiyle, dizinlenmiş bilgi nesnesi arasında doğru ve beklenen bir eşleştirmenin sağlanmış olmasıdır (Göker ve Davies, 2009).

Geleneksel "sorgu-yanıt" modeli artık kullanıcıların bilgi ihtiyaçlarını karşılamak için yeterli ve uygun değildir (Marchionini, 2006). Kullanıcıların keşif, öğrenme ve araştırmaya yönelik beklentileri giderek artmıştır ve bu da keşifsel arama faaliyetlerini karşımıza çıkartmaktadır (Hoeber ve Yang, 2008). Geleneksel bilgi erişim modelleri, kullanıcıların sorgularına uygun bilgi ve belgeleri sistemden çağırmayı ve bu sorgularla eşleştirmeyi temel almaktadır (Saracevic, 1997). Ancak günümüzde bilgi erişim sürecinin başarılı bir şekilde tamamlanabilmesi, yalnızca sistemin kullanıcı sorgularına uygun içerikleri sağlamasından ziyade, kullanıcıların erişim süreciyle ilgili somut verilerden yararlanılmasını gerektirmektedir. Bu süreç, kullanıcıların davranışları ve deneyimlerine odaklanarak, bilişsel, duygusal ve fiziksel boyutlarda sistemle etkileşim kurmalarını

içermektedir. Bilinen ve çok uzun yıllardır araştırma sürecinde kullanılan geleneksel bilgi erişim modeli Şekil 4.'de gösterilmiştir (Bates, 1989).



Şekil 3.2 Geleneksel bilgi erişim modeli

Geleneksel bilgi erişim modeli, Şekil 3.2'de gösterildiği gibi, kullanıcıların bilgi gereksinimlerine uygun sorgular oluşturarak koleksiyonda yer alan ve bu sorgularla ilişkili olan belgeleri bulmaya odaklanan bir süreci ifade etmektedir. Başka bir deyişle, bu model kullanıcı gereksinimlerini temsil eden sorgular ile koleksiyondaki belgeler arasında bir eşleşme mantığına dayanmaktadır.

Bilgi erişim (geri alma) süreci, bilgi erişim sistemlerinin ilgili bilgi ve belgelerin erişimini sağlama kapasitesinin yanı sıra, kullanıcıların sistemle etkileşim kurma yöntemlerini de kapsamalıdır. Kullanıcıların bilgi arama davranışlarının analiz edilmesi ve sistemle etkileşimlerinin optimize edilmesinin sağlanması, yalnızca arama sonuçlarının önemini artırmakla kalmaz, aynı zamanda bilgi erişim süreciyle ilgili kullanıcı memnuniyetini de yükseltir. Bu memnuniyeti artırmak, yalnızca sistemle kullanıcı etkileşimlerinin değerlendirilmesi ve bilgi geri alma sürecinde uygun bileşenlerin geliştirilmesi yoluyla başarılabilir.

Kullanıcıların bilgi arayışları sırasında sistemle etkileşimli olarak sergiledikleri davranışlar, bilgi arama sürecinin bağlamında gerçekleşen kullanıcı-sistem etkileşimi olarak nitelendirilir ve bu sürecin tüm kritik bileşenlerini kapsamaktadır (Ingwersen, 1992). Bilgi arama sürecinin çerçevesinde kullanıcılar, bilgi ihtiyaçlarını karşılamak için özel ihtiyaçlarına uygun sorgular oluşturmaya çalışırlar. Her kullanıcının bilgi ihtiyaçları önemli ölçüde farklılık gösterdiğinden ve kullanıcılar kişisel ihtiyaçlarına göre uyarlanmış sorgular oluşturduğundan, bu durum her kullanıcı için farklı sorgular ve buna karşılık gelen farklı arama sonuçları ile sonuçlanır.

Bilgi erişim (IR) sistemleri, kullanıcıların bilgiye erişmesi için önemli araçlardır ve arama motorları, soru cevaplama ve öneri sistemleri gibi senaryolarda yaygın olarak

uygulanır. Günümüzün dijital ortamında bilgi erişim sistemleri, çevrimiçi bilginin engin denizinde gezinmek için çok önemlidir. Tamdoğan (2009), yapmış olduğu çalışmasında BES'in, bilgi arayışında bireysel ihtiyaçları özelinde yaklaşan 'kullanıcı' ile veri tabanında sistematik bir biçimde yapılandırılmış 'bilgi' arasında kesintisiz bağlantıyı sağlamak üzere bir arada çalışan, etkileşim içerisindeki parçaların bir entegrasyonu biçiminde tanımlamıştır. Bu sistemler yalnızca güvenilir olmakla kalmaz, aynı zamanda bilgi ve fikirlerin küresel olarak yayılmasında da önemli bir rol oynar. Kullanıcıların ihtiyaç duydukları bilgileri büyük veri tabanı koleksiyonlarından bulmalarına yardımcı olur. Dijital çağda, bilgi alma etkinliği akademik ve profesyonel başarının kritik bir yönü haline gelmiştir. Veri tabanlarında ve çevrimiçi kaynaklarda etkili bir şekilde gezinme yeteneği, araştırmanın kalitesini ve öğrenme sonuçlarını önemli ölçüde etkileyebilir.

Etkileşimli Bilgi Erişim (IIR) Sistemleri: bilgisayar sistemleri ile insan kullanıcıları arasında iletişimsel bir değişim sağlayan etkileşim seanslarını ifade etmektedir (Reitz, 2004). Sonuç olarak, bilgi erişim sistemleri ile bu sistemlerin kullanıcıları arasındaki etkileşimi destekleyen organizasyonel çerçeveler, interaktif bilgi erişim sistemleri olarak adlandırılır. Etkili bilgi alma, dijital çağda akademik ve profesyonel başarıyı etkileyen temel bir beceridir. Veritabanlarında etkin bir şekilde gezinme ve bunları kullanma becerisi, arama davranışını ve sonuçlarını etkileyen çeşitli faktörleri vurgulayan çok sayıda çalışmanın odak noktası olmuştur.

Bilgiye Erişim Sistemi'nin (BES) amacı, bir kullanıcının ihtiyaç duyduğu değerli bilgiye ulaşmasına yönelik olarak bu işlem için harcanması gereken sürenin/emeğin azaltılmasıdır (Kowalski, 2011; Latha, 2018). Büyük bir öğe koleksiyonu içerisinden bilgi ihtiyacı olan kullanıcı için ilgili olması beklenen öğelerin (metin, resim, video, vb.) seçilmesidir (Larson, 2011). Ayrıca, kullanıcının isteğini karşılamaya yönelik gerekli bilginin sağlanması olmakla birlikte, konuya özel tüm bilgiyi bulmak değildir (Kowalski, 2011). Bir BES'in desteklemek zorunda olduğu üç temel işlem bulunmaktadır. Bunlar aşağıdaki şekilde belirtilebilir (Göker ve Davies, 2009):

1. Kaynakların içeriğinin betimlenmesi: Dizinleme işlemi olarak da ifade edilmektedir. Bu işlem, kullanıcının dâhil olmadığı bir biçimde gerçekleştirilir. Çoğunlukla belgeler parçalar halinde (sadece başlık, özet ve asıl belgenin konumu ile ilgili bir bilgi) depolanır, bazen de asıl belgenin depolanmasını içerebilir. Bu işlemin en basit hali, her bir metnin içinde yer alan kelimelerin çıkartılması ve bu kelimelerin hangi

metine ait olduğunu belirten bir gösterge ile birlikte bir dizin değeriyle depolanması şeklinde gerçekleştirilmektedir (Larson, 2011).

2. Kullanıcının bilgi ihtiyacının betimlenmesi: çoğunlukla sorgu formülleştirme olarak ifade edilmektedir. Uygun bir sorgulamaya ve kullanıcı ihtiyaçlarının daha iyi anlaşılmasına önderlik eden sistem ve kullanıcı arasındaki etkileşimli iletişim olarak değerlendirilmektedir. Buradaki "betimleme" kavramı kullanıcının yaptığı sorgulama işlemidir. Yani; bir kullanıcının ihtiyacını tanımlayan bir terimle (kelime ya da kelimeler) BES üzerinde sorgulama yapmasıdır. Kullanıcı, bilgi ihtiyacını betimleyen bir terimle yaptığı sorgulama sonucunda karşısına getirilen kaynakları yeterli görmediği takdirde, betimleme şeklini değiştirebilir. BES ise, değişen bu betimleye yönelik olarak yeni bilgi kaynaklarını listeleyebilir.

3. Her iki betimlemenin eşleştirilmesi: kullanıcının yaptığı sorgulamayla, dizinlenmiş olan bilgi kaynaklarının eşleştirilerek, en uygun kaynakların kullanıcıya sunulması anlamına gelmektedir. Kullanıcı, karşısına getirilen listede yukarıdan aşağıya doğru tarama yaparak ihtiyacına uygun olan kaynağı bulmaya çalışacaktır.

Genel itibariyle özetlemek gerekirse; BES, belirli bir konuya ya da duruma özel bilgi ihtiyacını karşılayabilecek kaynakları organize edip depolayan, bilgi ihtiyacını giderebilmek için kullanıcıyla bir arayüz üzerinden etkileşime giren ve kullanıcıyla ihtiyacı giderebilme potansiyeli olan bilgi kaynaklarını ortak bir çerçevede buluşturabilen sistem olarak tanımlanabilir.

3.2.1 Bilgiye Erişim Sistemleri

Hitap ettiği kullanıcı kitlesinin çeşitli bilgi ihtiyaçlarını karşılamak amacıyla hizmet vermekte olan BES'lere, günümüzde sıklıkla faydalanmakta olduğumuz Google, Yahoo, Bing, vb. arama motorları en bilinen örnekler olarak verilebilir. Bununla birlikte, çoğu üniversitelerin ve kütüphanelerin kitaplara, dergilere ve belgelere erişim sağlamak amacıyla kullandıkları sistemler de BES'lere örnek verilebilir (Latha, 2018). BES olarak değerlendirilmekte olan sistem çeşitlerinden bazıları aşağıda açıklanmıştır:

1. Web Arama Motorları: Kullanıcıların web üzerinden ihtiyaçlarını giderebilmelerine yardımcı olan arama motorları, herkes tarafından görülebilen en önemli bilgi erişim teknolojilerinden birisidir (Latha, 2018). Farklı nitelikteki bilgiye hızlı bir

şekilde erişim imkânını arttıran önemli bir BES çeşididir. Bilgi ihtiyacı bulunan kişi, arama motoru ara yüzlerindeki metin yazma kısmına ihtiyacını tanımlayan terimleri girerek sorgulamalar gerçekleştirebilmekte ve bu sorgulamalar doğrultusunda ihtiyacına cevap olabilecek en uygun sonuçlara ulaşabilmektedir. Bu sistemin kullanıcıları, kısa sorgulamalarına (birkaç kelimeden oluşan terimler) karşılık olarak ihtiyaçlarına doğru ve hızlı cevap beklemektedir (Büttcher vd. 2010). Kullanıcı tarafında bu kadar basit bir işlem yapılmasına rağmen, arka planda bilgi ihtiyacını giderebilecek web sayfalarının sıralamasını yapan çok fazla bilgisayar koordineli bir şekilde çalışmaktadır (Büttcher vd. 2010). Aynı zamanda bu sistemler, kullanıcıların bilgi ihtiyaçlarına cevap verebilmek için "Web Tarayıcısı (Crawler)" adı verilen bir yapı kullanırlar. Bu yapı, ihtiyaca karşılık olarak doğru sonuçların sunulması için web'de yer alan sayfaların belirli zaman aralıklarında kopyasını alıp, kayıt altında tutmaktadır (Büttcher vd. 2010).

2. Masaüstü ve Dosya Sistemi Arama Araçları: Yerel bir hard disk üzerinde ya da belki de yerel bir internet ağı üzerinden ulaşılabilen diskler üzerinde bulunan dosyalara erişmeye yarayan uygulamalardır (Büttcher vd. 2010; Latha, 2018). Web arama motorlarının aksine, dosya format ve üretilme sürelerine yönelik olarak farkındalık bakımından daha hassas çalışan bir yapıya sahiplerdir (Büttcher vd. 2010; Latha, 2018).

3. Dijital Kütüphaneler: Kaliteli ve çoğu zaman özel mülkiyete sahip materyal koleksiyonuna erişim amacıyla kullanılan bir BES çeşididir (Büttcher vd. 2010). Bu koleksiyon, telif hakkı sebebiyle web siteleri üzerinde barındırılmayan makaleler, kitaplar, vb. kaynakları içerir (Latha, 2018). Yayıncı kalitesi ve sınırlı kapsamı göz önüne alındığında, arama isteğini daraltıp, bilgi erişim verimliliğini arttırmak için yapısal özelliklerinin (yazar, başlık, tarih ve diğer yayın verisi) avantajından faydalanmak çoğunlukla mümkündür (Büttcher vd. 2010).

4. Seçme Bilgi Yayımı Sistemleri: Arama sistemleri "çekme (pull)", dağıtım sistemleri ise "itme (push)" sistemleri olarak ifade edilmektedir (Meadow vd., 2007; Kowalski, 2011). Arama sistemlerinde bir kullanıcı, ihtiyaçları giderebilmek için sisteme sorgu yöneltirken; dağıtım sistemlerinde ise kullanıcı, arama yapılan terimlerin bir arada kaydedildiği bir profil tanımlamaktadır (Kowalski, 2011). Büttcher vd. (2010), bu davranışı bilgi erişim sürecinin tersten işletilmesi şeklinde ifade etmiştir. Latha (2018), tarafından dokümantasyon izleme araçları olarak tanımlanmış olan bu sistem, yeni bir bilgi eklendiğinde otomatik olarak kullanıcının profiliyle karşılaştırma yapmaktadır.

Eşleşme bulunduğundaysa, kullanıcıyı bilgilendirmektedir. Bu bilgilendirmeler, e-posta, sesli mesaj, anlık mesajlaşma, cep telefonu mesajı ve Zengin Site Özeti (Rich Site Summary - RSS) beslemeleri gibi çeşitli şekillerde yapılabilmektedir (Latha, 2018). Genel olarak bu sistemler, kullanıcı profiline bağlı kalarak yeni bir bilgi kaynağı ortaya çıktığında kullanıcıyı bilgilendiren otomatik arama sistemleri şeklinde ifade edilebilir. Web arama motorlarında sonuçların gösterilmesi esnasında bir ürünün reklamının sayfada yer alması da buna örnek gösterilebilir (Zhai ve Massung, 2016).

5. Bilgi Filtreleme Sistemleri: Bir kullanıcıya sunulmadan önce otomatik/yarı otomatik veya bilgisayarlı yaklaşımları kullanarak gereksiz bilgileri bilgi akışından kaldıran sistemlerdir (Latha, 2018). Büttcher vd. (2010) tarafından "tersten işletilen bilgi erişim süreci" olarak ifade edilmiş diğer bir BES çeşididir. Bu sistemlere, spam filtreleme sistemleri örnek gösterilebilir.

6. Öneri Sistemleri: Latha (2018), bu sistemleri, kullanıcıların ilgisini çekebilecek bilgi öğelerini kullanıcıya sunan bilgi filtreleme sistemleri olarak tanımlamıştır. Bu sistemler, kullanıcıya iletilen bilgi akışından bilgi öğelerini çıkarmak yerine, bu akışa yeni öğeler ekleyen sistemlerdir (Latha, 2018). Bu sistemlerde yapılan analizlerin genellikle kullanıcılar ve öğeler arasındaki geçmiş etkileşimlere dayanmakta olduğu, çünkü geçmiş eğilim ve ilgilerin gelecek tercihlerinin iyi bir göstergesi olduğu belirtilmiştir (Aggarwal, 2016).

7. Uzman Arama Sistemleri: Kurumlarda belirli bir alanda uzmanlaşmış kişilerin aranabilmesi amacıyla kullanılan sistemlerdir (Büttcher vd. 2010). Bu sistemde, kullanıcıların bir konu üzerinden gerçekleştirdikleri sorgulamalara karşılık olarak bilgi koleksiyonunda sorguyla güçlü ilişkisi olan kişilerin bulunması işlemi gerçekleştirilmektedir (Latha, 2018). Bulunan uzmanlar, sorguyla ilişkilerinin kuvvetlilik durumlarına göre sıralanarak listelenmektedir (Latha, 2018). Laudon ve Laudon (2012) tarafından "Bilgi Ağı Sistemleri (Knowledge Network Systems)" olarak da isimlendirilmiş olan bu BES, çalışanların bir şirkette uygun uzmanı bulmasını kolaylaştırmak için iletişim teknolojilerini kullanan bir sistem olarak tanımlanmıştır.

8. Soru Cevaplama Sistemleri: İnsanların doğal dilleriyle sordukları soruları otomatik olarak cevaplamaya yönelik işlem yapan sistemlerdir (Latha, 2018). Bu işlemi yapabilmek için farklı kaynaklardan bilgi toplamaktadır (Büttcher vd. 2010). Ek olarak, doğal dille oluşturulmuş, yapılandırılmamış belge koleksiyonlarındaki cevapları da

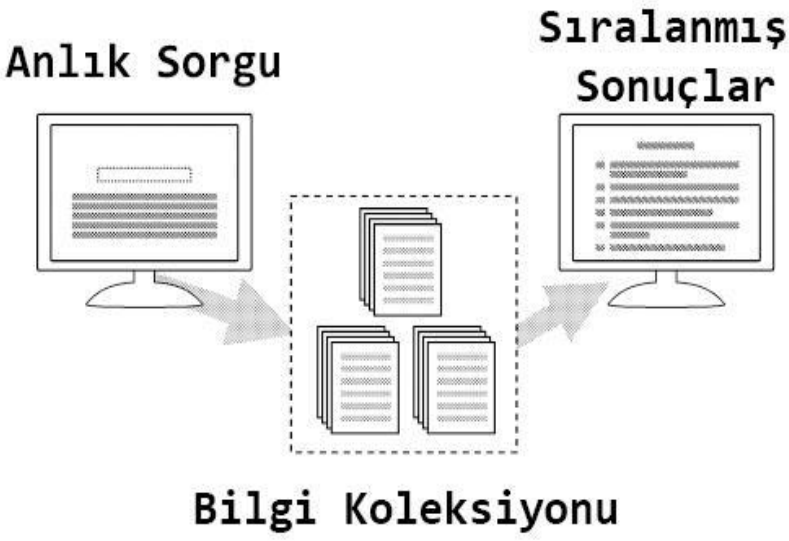
bünyesine alabilmektedir (Latha, 2018). Bilgi kaynaklarını listelemek yerine sorulara karşılık özel cevaplar vermek bu sistemlerin amacıdır (Croft vd. 2009).

9. Multimedya Bilgi Erişim Sistemleri: Resim, müzik, video ve konuşma gibi farklı multimedya kaynaklarından anlamsal bilgiye erişmeyi amaçlayan sistemlerdir (Büttcher vd. 2010).

10. Özetleme Sistemleri: Belgeleri, içeriklerini açıklayan az miktarda ana paragraf, cümle veya kelime öbeklerine küçülten sistemlerdir (Latha, 2018). Web arama motorlarında yapılmış olan bir sorgulama sonucunda, ortaya çıkmış olan sonuç listesindeki metin parçacıkları buna örnek gösterilebilir (Büttcher vd. 2010).

Meadow vd. (2007) ve Kowalski (2011) tarafından "çekme" ve "itme" olmak üzere iki farklı biçimde kategorize edilmiş olan BES'ler, Zhai ve Massung (2016) tarafından da bilgiye erişimde zaman kavramı temel alınarak farklı bir şekilde açıklanmıştır. Zhai ve Massung (2016)'un "kısa vadeli" olarak ifade ettiği çekme biçimindeki BES'lerde, kullanıcının geçici bir bilgi ihtiyacının (ad hoc information need) olması söz konusudur. Bu bilgi ihtiyacı karşılanır karşılanmaz, bilgi erişim isteği ortadan kaybolmaktadır. Web arama motorlarında gerçekleştirilen anlık, ihtiyaca özel bilgi erişim davranışı bu duruma örnek olarak gösterilebilir. Diğer taraftan, "uzun vadeli" olarak ifade edilmiş olan BES'lerde ise, sistem, kullanıcıya birtakım ilgili bilgi öğelerini tavsiye etme işlemini gerçekleştirmektedir. Bu BES'ler, bir kullanıcının uzun süredir devam eden bilgi ihtiyacını karşılamak için genellikle daha kullanışlıdır. Bilgi filtreleme ve öneri sistemleri bu türdeki BES'lere örnek olarak gösterilebilir.

Zhai ve Massung (2016), anlık bilgi erişimin (ad hoc retrieval) daha önemli olduğunu, çünkü anlık bilgi ihtiyacının uzun vadeliden daha sık ortaya çıktığını belirtmiştir. Bu çalışma kapsamında gerçekleştirilmiş işlemler de anlık bilgi erişim çerçevesinde yer alan web arama motorlarına dayanmakta olup, bu BES'lerdeki temel işlemler Şekil 3.3'de aktarılmıştır:



Şekil 3.3 Temel anlık bilgi erişim işlemi.

Şekil 3.3'e göre bir kullanıcı ihtiyaç duyduğu bilgi için arama motoruna anlık olarak sorgu göndermekte ve arama motoru, içinde barındırdığı bilgi koleksiyonunu kullanıcının sorgusuna göre analiz edip, ilgili olan sonuçları sıralayarak kullanıcıya listelemektedir (Grossman ve Frieder, 2004).

3.2.2 Üniversite Kütüphaneleri ve Bilgiye Erişim Çalışmaları

Üniversite kütüphaneleri, akademik büyüme için hayati önem taşıyan dijital veritabanları, e-kitaplar ve çevrimiçi dergiler gibi temel bilgi kaynakları sunma uğraşındadır (Bachynska vd. 2024). Bu kütüphaneler, bağlı oldukları kurumların eğitim, öğretim ve araştırma faaliyetlerini kolaylaştırmak amacıyla; fakülte, öğrenci ve personelin çeşitli bilgi ve belge ihtiyaçlarını karşılamak ve bilgi birikiminin ulusal ve uluslararası platformlarda toplanması, kullanılması ve yayılmasını teşvik etmek üzere stratejik olarak tasarlanmıştır. Bu hedeflere ulaşmak için bilgi kaynaklarının sistematik olarak organize edilmesi ve kullanıcıların ihtiyaçlarına uygun şekilde erişilebilir hale getirilmesi zorunludur.

Günümüz kütüphane kullanıcıları, genellikle araştırma veya inceleme konularıyla ilgili materyalleri hızlı ve kolay bir şekilde bulabilmeyi beklentisindedirler. Teknolojideki ilerlemeler, bu kişilerin kullanıcı demografisi ve bilgi arama davranışlarında önemli bir

dönüşüme neden olmuştur. Bilgi teknolojilerinin hızlı evrimiyle birlikte, yeni nesil kullanıcılar, bilgisayar, internet ve mobil cihazlar gibi dijital medya araçları bakımından zengin bir ortamda yetişmiş bir demografik grubu oluşturmaktadır ve bu da web ekosistemini günlük yaşamlarının önemli bir parçası haline getirmektedir.

Online Toplu Erişim Kataloğu (OPAC) sistemleri, bilgi teknolojilerindeki gelişmelere paralel olarak önemli dönüşümler geçirmiştir. 1970'lerde ortaya çıkan ilk nesil OPAC'lar, kütüphanelerin mevcut materyallerine ait Makine Okunabilir Kataloglama (MARC) kayıtlarını inceleyerek bibliyografik bilgiye erişim sağlamıştır. Daha sonrasında, Boolean operatörlerinin ve anahtar kelime kullanımına olanak tanıyan sistemlerin ortaya çıkması önemli bir gelişme olmuştur. 1990'larda, grafik kullanıcı ara yüzleri, hiperbağlantılar ve gelişmiş arama ve erişim özellikleriyle karakterize edilen ikinci nesil web tabanlı OPAC'ların ortaya çıkışı, alanda önemli bir evrim olarak kabul edilmiştir (Özel ve Çakmak, 2011). Takip eden süreçte Web 2.0 teknolojileriyle uyumlu olarak geliştirilen yeni nesil OPAC'ların uygulanması, kullanıcı erişilebilirliğinde ve çeşitli tarama seçeneklerinde önemli gelişmeler sunmaktadır. Bu sistemler, diğer kataloglar ve veri tabanları dâhil olmak üzere çeşitli kaynaklardan indeksleme bilgilerini birleştirir, ödünç verme verilerine erişimi kolaylaştırır ve sorunsuz veri entegrasyonuna olanak tanımıştır. Bununla birlikte sosyal ağlar, etiketleme, kullanıcı geri bildirimi ve arama sonuçlarını daraltma veya genişletme yeteneği yoluyla kişiselleştirme de mümkündür. Ayrıca konu tarama, hem özlü hem de kapsamlı kayıt görüntüleme, alaka düzeyine göre sıralama ve gelişmiş keşif özellikleri de dâhil edilmiştir. Buna ek olarak, bu OPAC'lar ilgili arama terimleri sağlayabilir, yazım hatalarını düzeltebilir, çeviriler sunabilir ve çeşitli sıralama ölçütleri uygulayabilir (Wilson, 2007).

Web 2.0 teknolojileri tarafından kolaylaştırılan kullanıcı etkileşimli çerçeve, genellikle Kütüphane 2.0 ortamı olarak adlandırılan bir paradigmanın oluşturulmasına olanak sağlamıştır. Bu çerçeve, kullanıcı odaklı bir dönüşüm modelini temel alır. Yaklaşım, kullanıcıların hem fiziksel hem de dijital hizmetlerin tasarımına ve geliştirilmesine katılımının önemini vurgulayarak, hizmetlerin kullanıcı ihtiyaçlarındaki ve tercihlerindeki değişimlere yanıt olarak gelişmesi gerektiğini önermektedir (Casey ve Savastinuk, 2006). Konu ile ilgili yapılmış olan araştırma bulguları gösteriyor ki kullanıcıların çoğunlukla bilimsel bilgi arayışındayken arama motorlarına başvurup ve daha sonra kütüphane kataloglarına yöneliyorlar. Ayrıca, kütüphane kataloglarındaki

kaynakların içeriğine erişim sağlayan bilgileri görüntülemeyi tercih ettiklerini ve aynı zamanda arama kolaylığı sağlayan arayüz tasarımı ve yazım algoritmalarını önceliklendirdiklerini belirtmektedir (Özel ve Çakmak, 2011).

Web keşif araçları üzerine kapsamlı incelemeler yoluyla, kullanıcıların değişen beklentilerinin belirlenmesi ve kütüphane hizmetlerinin buna göre uyarlanması gerektiği öne sürülmektedir. Bu, kütüphanelerin temel rolünün, kullanıcıların kütüphane kaynaklarından en iyi şekilde yararlanmasını sağlayan gerekli altyapıyı ve hizmetleri sağlamak ve bu kaynaklardan elde edilen bilgilerin araştırma çalışmalarına etkin bir şekilde entegre edilmesinin önünü açmak için oldukça önemlidir.

Günümüzde etkin kütüphane hizmetleri kapsamında önemli yere sahip olan Yapay Zekânın (YZ) akademik kütüphanelerdeki Bilgi Erişim (IR) sistemlerine entegrasyonu, kullanıcı deneyimini, arama verimliliğini ve genel erişilebilirliği modernize etme ve geliştirme yönünde önemli bir ilerlemeye işaret etmektedir. Yapay zekânın çağdaş kütüphanelerde kataloglama ve sınıflandırma çerçeveleri üzerindeki etkisi önemlidir. Özellikle verimlilik, doğruluk ve kullanıcı etkileşimi açısından çok sayıda avantaj sunar. Bununla birlikte, veri gizliliği sorunları, algoritmik önyargı ve insan gözetimi gerekliliği gibi engellerin, yapay zekâ teknolojilerinin yeteneklerini tam olarak kullanabilmek için üstesinden gelinmesi gerekmektedir. Yapay zekâ, makine öğrenimi algoritmalarından yararlanarak, kullanıcı davranışına ve geri bildirimlerine dayalı olarak arama sonuçlarını dinamik bir şekilde iyileştirebilir ve zaman içinde uyum sağlayarak giderek daha alakalı öneriler sunabilir. Bu yetenek, kullanıcıların bilgiye daha verimli bir şekilde erişmesine yardımcı olmakla kalmaz, aynı zamanda daha ilgi çekici ve kişiselleştirilmiş bir deneyimi de teşvik eder. Yapay zekâ destekli öneri sistemleri, doğal dil arama işlevleri ve sanal asistanlar, akademik kütüphaneleri bireysel araştırma ihtiyaçlarına ve öğrenme tercihlerine daha uyumlu hale getirerek bilgi erişimini kolaylaştırma potansiyelini ortaya koymaktadır.

Bilginin yayılması ve dağıtımının yöntemleri, her dönemin yeniliklerine uygun olarak sürekli bir dönüşüm geçirmiştir. Bilgi teknolojisi ve iletişim alanlarındaki ilerlemeler, bilgi aktarımını kolaylaştıran araçların geliştirilmesine yol açarak e-öğrenme ile bilgi arama arasındaki ilişkiyi önemli ölçüde güçlendirmiştir; e-öğrenme bu tür bir araçtır. Genellikle e-öğrenme veya sanal öğrenme olarak adlandırılan web tabanlı öğrenme, temel olarak sanal öğrenme ortamlarında yönetilen kurslar aracılığıyla

çevrimiçi eğitimin deneyimini kapsar. Sanal öğrenme ortamının kullanımıyla, eğitim süreçleri elektronik formatta yürütülebilir (Ghalyan ve Zalpour, 2019).

4. VERİ MADENCİLİĞİ

4.1 Veri Madenciliği Tanımı ve Tarihçesi

Veri madenciliği, büyük veri kümelerindeki önemli ve anlamlı örüntüleri bulmak için kullanılan bir süreçtir. Bir diğer tanımı ise, verilerin uygun analiz teknikleri kullanılarak analiz edilmesiyle büyük veri tabanlarında bulunan verilerin sahip oldukları özniteliklerini, aralarındaki veri ilişkilerini ya da sahip oldukları modellerdeki gizli örüntüleri ortaya çıkarmayı ve sonuçlardan faydalı bilgiler özetlemeyi amaçlayan süreçler olarak tanımlamaktadır (Vikipedi, 2024). Teknik yönden veri madenciliğine baktığımızda, bilgi toplama ve kataloglama yöntemini birleştirmesi ve sonrasında ise büyük miktarda veriden kural benzeri bilgi üretebilmesi öne çıkan yönleri olmaktadır. Ticari yönden bu kavram, bir işletmenin bilgi işleme teknolojisine karşılık gelmektedir. Mevcut veri madenciliği algoritmalarındaki en yaygın olanları arasında sınıflandırma, regresyon, kümeleme, ilişki kuralları, kural oluşturma, özetleme, bağımlılık modelleme ve dizi analizi bulunur (Mitra vd. 2002). Veri madenciliğinde süreç: Bilinmeyen örüntüler uygun formatta saklanan yapılandırılmamış verilerden çıkarılır ve bu çıktılar gelecekteki stratejileri planlamak ve geliştirmek amacıyla kullanılır. Veri madenciliği, ham veriyi anlamlı bilgiye dönüstürmek için kullanılır.

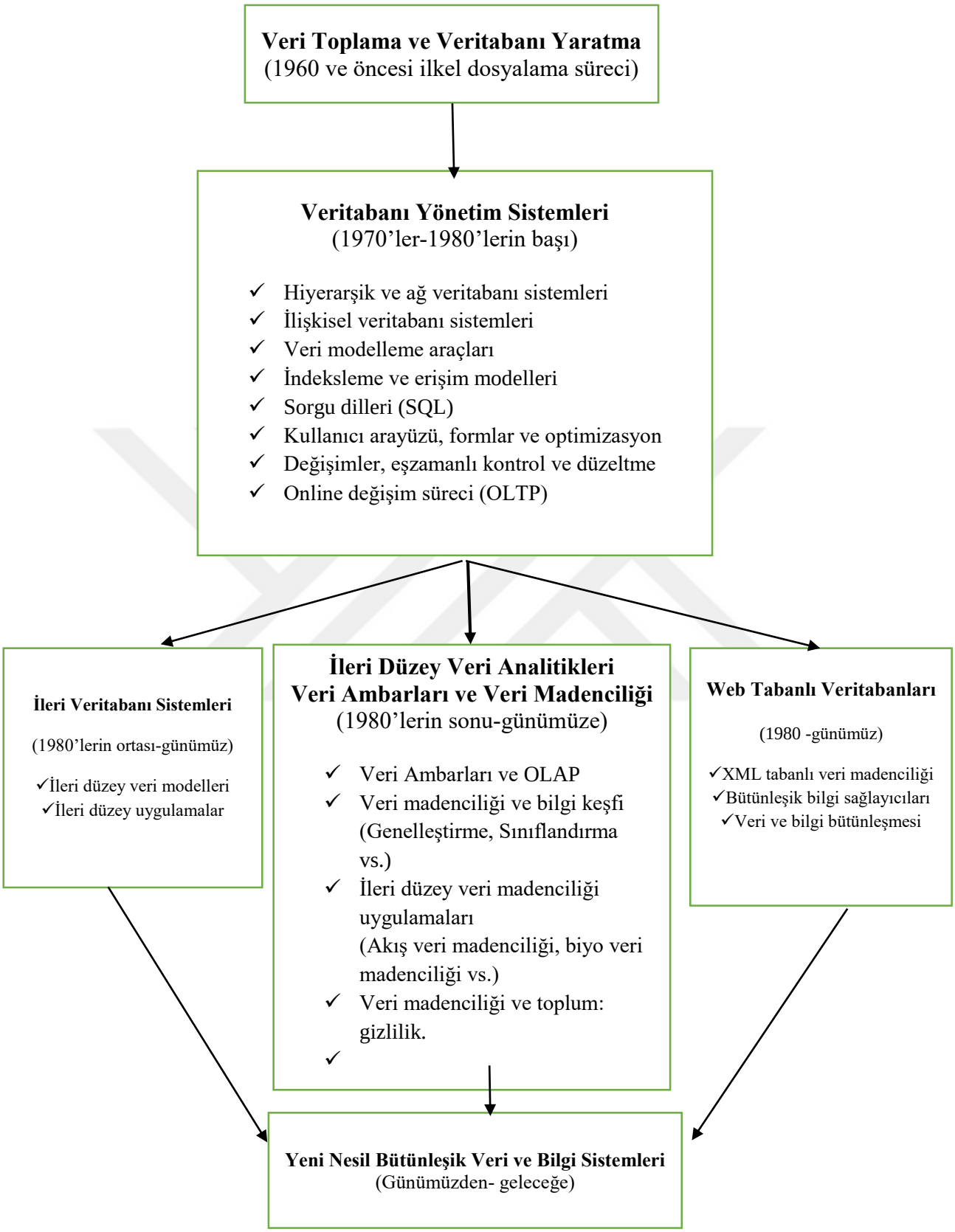
Veri madenciliği, istatistik, makine öğrenimi, veri yönetimi, örüntü tanıma ve yapay zekâ gibi disiplinlerin kesişim noktasında yer alan bir alandır. Bu disiplinin temel amacı, büyük veri kümeleri üzerinde gerçekleştirilen analizlerle, gizli ve faydalı bilgilerin açığa çıkarılmasıdır (Coşkun ve Baykal, 2011). Veri madenciliği, mevcut problemlere çözüm bulmak, stratejik kararlar almak ve geleceğe dair tahminler geliştirmek için gerekli bilgileri sağlayan güçlü bir araçtır. Elde edilmesi hedeflenen bilgiler; açık bir şekilde belirgin olmayan, daha önce fark edilmemiş, bilinmeyen ancak kullanıldığında anlamlı ve kritik değer taşıyan bilgilerdir.

Günümüzde, bilgi toplama, depolama ve işleme teknolojilerindeki ilerlemeler, mevcut veri kaynaklarının daha ayrıntılı incelenmesini ve bu verilerden anlamlı sonuçlar çıkarılmasını mümkün kılmaktadır. Bu durum, veri madenciliğinin karar verme süreçlerindeki önemini giderek artırmaktadır. Bilgi ve iletişim teknolojilerinin (BİT) hızla gelişmesi, insan hayatının her alanında büyük değişimlere yol açmaktadır. Bu değişimlere

uyum sağlamak amacıyla kütüphaneler geliştirilmiş, koleksiyonları dijital ortama aktarılmış ve çevrimiçi ya da uzaktan erişim imkânı sunularak ürün ve hizmetler kağıt formdan elektronik forma dönüştürülmüştür (Rafique vd. 2021). Bilgi merkezleri, sistem günlükleri, dolaşım geçmişi ve kullanıcıların okuma alışkanlıklarını analiz ederek müşteri ihtiyaçlarını belirleyebilir, hizmet ve içeriklerini daha etkili hale getirebilir ve daha doğru pazarlama stratejileri geliştirebilirler.

Veri madenciliği araçları, yapılandırılmış verileri etkili bir şekilde işlerken, yapılandırılmamış veya yarı yapılandırılmış verileri analiz etmek için metin madenciliği önemli bir rol oynamaktadır (Salloum vd. 2018). Günümüzde kütüphane ve bilgi merkezleri kullanıcılarına, doğru ve hızlı yanıt verebilmek için hizmet kalitesini ve kapasitesini artırmaya çalışmaktadır. Ancak, geleneksel araçlar ve teknikler hızla büyüyen bilgi miktarıyla başa çıkmakta yetersiz kalmaktadır. Bu bilgi talebiyle basa çıkmak için kütüphane sistemi ve bilgi merkezleri veri madenciliği tekniklerini kullanmaktadır (Ansari vd. 2020). Veri madenciliği, kütüphane ve bilgi biliminde disiplinler arası bakış açısı gibi yeni fırsatlar da sunmaktadır (Virkus ve Garoufallou, 2020).

Büyük veri hacimleri ve çeşitliliği içinden anlamlı bilgiye ulaşma ihtiyacı veri madenciliğini oldukça önemli hale getirmektedir. Büyük hacimde veri olan her yerde veri madenciliği kullanılabilir. Teknolojik gelişmeler ile birlikte verilerin toplanması, depolanması ve çağrılması hususlarında kolaylık yaşanmaya başlanmıştır. Şekil 4.1'de görüldüğü üzere veri madenciliği belli aşamalardan geçerek günümüze gelmiştir (Jiawei ve Kamber, 2006).



Şekil 4.1 Geçmişten günümüze veri madenciliğinin gelişimi "Data mining: concepts and techniques.", Jiawei ve Kamber (2006), kaynağından dönüştürülmüştür. Telif hakkı San Francisco, CA, itd: Morgan Kaufmann yayıncısına aittir

1950’li yıllarda matematikçiler, veri madenciliği teknikleri üzerinde çalışmalar başlatmış ve bu süreçte mantık ile bilgisayar bilimlerine dayalı olarak yapay zekâ ve makine öğrenimi alanlarını geliştirmişlerdir. 1960’lı yıllarda ise istatistikçiler, teorik bir hipoteze dayanmaksızın veri analizi yapma sürecini tanımlamak amacıyla “veri balıkçılığı” veya “veri taraması” gibi terimler kullanmışlardır. "Veri madenciliği" terimi ise ilk kez 1990’lı yıllarda veri tabanı topluluğu içerisinde ortaya çıkmış ve bu süreçte yaygın bir kullanım kazanmıştır. Diğer bazı eş anlamlı terimler olarak veri arkeolojisi hasadı, bilgi keşfi, bilgi çıkarımı vb. kullanılmışlardır. 1960'lı yıllarda istatistikçiler, veri madenciliğinin temellerini oluşturan yeni algoritmalar üzerinde çalışmalar yapmışlardır. Bu dönemde regresyon analizi, en büyük olabilirlik kestirimleri ve sinir ağları gibi yöntemler ön plana çıkmıştır. Aynı zamanda, veritabanı sistemlerinin gelişmesiyle büyük miktarda metin dokümanının saklanması ve bu verilere erişim mümkün hale gelmiştir. Veri madenciliğinin altyapısında önemli bir yer tutan unsurlar, istatistik çalışmaları ve veritabanı uygulamalarıdır. Özellikle büyük veri ambarlarının ortaya çıkışı, veri madenciliği çalışmalarının hız kazanmasına yol açmıştır.

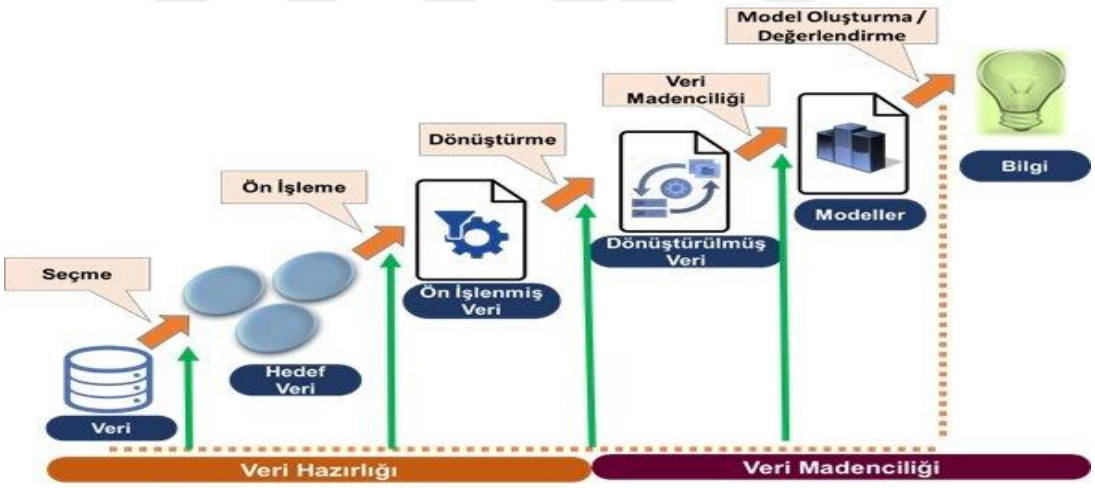
1960'larda veriler elektronik ortamda toplanmaya başlanmış ve geçmiş veriler bilgisayar teknolojileri ile analiz edilmiştir. 1980'li yıllarda ise bağıntılı veritabanlarının (relational databases) ve SQL gibi dillerin geliştirilmesi, verilerin dinamik olarak analiz edilmesine olanak sağlamıştır. Bu süreçte, yeni programlama dillerinin ve bilgisayar tekniklerinin yaygınlaşmasıyla kümeleme yöntemleri, genetik algoritmalar ve karar ağaçları gibi gelişmiş algoritmalar kullanılmaya başlanmıştır. 1990'lı yıllarda ise veri hacmindeki büyük artış nedeniyle veri ambarları, verilerin depolanması ve analiz edilmesi için temel bir çözüm haline gelmiştir. Bu dönemde, istatistik ve yapay zekâ tekniklerinin kullanımıyla veri madenciliği modern anlamıyla şekillenmiş ve bilgi keşfinin ilk adımları atılmıştır. Ayrıca, veri tabanı ambarlarının geliştirilmesi bu sürecin önemli bir dönüm noktası olmuştur.

Günümüzde veri madenciliği, neredeyse tüm sektörlerde geniş bir uygulama alanı bulmuş ve veri tabanı ile bilgi sistemleri alanlarındaki en önemli yeniliklerden biri haline gelmiştir. Bu nedenle, veri madenciliği hem bilgi iletişimi hem de teknoloji alanlarında en umut verici disiplinler arası gelişmelerden biri olarak kabul edilmektedir. Bilgi merkezleri özelinde ise sınıflandırma, edinim, dolaşım, pazarlama, kullanıcı izleme ve

referanslama gibi işleyişlerinin veri madenciliği gibi yeni nesil teknikler kullanılarak yeniden tasarlanması yararlı olacaktır.

4.2 Veri Madenciliği Süreci

Bilgi yoğun kurumsal yapıların ön planda olduğu günümüzde, veri tabanı sistemlerinin yoğun kullanımı kaçınılmaz hale gelmiştir. Zamanla büyük hacimlere ulaşan bu verilerin organizasyonlara nasıl fayda sağlayabileceği önemli bir sorun olarak ortaya çıkmıştır. Geleneksel sorgu ve raporlama araçları, bu büyük veri yığınlarını analiz etmede yetersiz kalırken, veri madenciliği süreci ve Veri Tabanı Bilgi Keşfi gibi yaklaşımlar, verilerin daha etkin kullanılmasını sağlamak için geliştirilmiştir. Bu süreç, gizli kalmış örüntüleri ve anlamlı bilgileri ortaya çıkararak stratejik karar alma süreçlerini desteklemektedir. Veri Tabanı Bilgi Keşfi sürecindeki iş süreci temel olarak beş adımdan oluşmaktadır: Şekil 4.2'de gösterilmiştir:



Şekil 4.2 Veri tabanlarında bilgi keşfi

Veri madenciliği teknolojisinin iş süreci temel olarak yukarıda verilen beş adımdan oluşmaktadır:

1. Veri Seçimi ve Derlenmesi: Öncelikle veri madenciliği görevinin genel amacı belirlenir. Gerekli veri türü, kaynak maliyeti, karşılaşması muhtemel riskler, vb. belirlenip uygun bir uygulama planı seçilir. Bu süreçte tutulan veriler arasında sorguya uygun olanlar seçilir, elde edilir ve derlenip depolanır.

2. Veri Temizleme ve Ön işleme: Veri seçimi sürecinde elde edilen verilerdeki hatalı tutulanların tespit edilip çıkarılması ve eksik olan veri değerlerinin değiştirilip tamamlandığı aşamadır. Bu süreç eksik değerleri belirlemeyi, veri tutarlılığı sağlamayı, veri dönüştürmeyi ve optimizasyon gibi işlemleri içerir.

3. Veri Dönüşümü: Verilerin yeniden birleştirilmesinde veya dönüştürülmesinde yumuşatma, normalleştirme ve istenen formatta toplama yapılır (Gul vd. 2021). Veriler madencilik işlemleri için uygun bir yapıya dönüştürülür (Sharma vd. 2021). Bu aşamada yapılanların amacı verilerin uygun formda birleştirilmesidir (Nalawade ve Joshi, 2021).

4. Veri Madenciliği: Bir önceki işlem sonrasında elde edilen veri kümesine veri madenciliği sorgusunun uygulanması sürecidir. Toplanan ve işlenen verilerdeki örüntüleri bulmak için özel araçlar ve teknikler uygulanır (Gul vd. 2021). İşlenmiş verilerin veri madenciliği fonksiyonları ve algoritmaları ile değerlendirilerek verilerden anlamlı eğilimlerin, desenlerin (pattern) çıkarılması yapılır. Yukarıda belirtilen sorgular sınıflandırma, kümeleme, ilişkilendirme biçimindeki sorgulardır (Nalawade ve Joshi, 2021).

5. Bilgi - Yorumlama: Dördüncü işlem sonrasında ulaşılan sonuçların yorumlanması, anlamlandırılmasıdır. Veri madenciliğindeki son adımdır ve veri madenciliğinde elde edilen veri modellerinin yorumlanmasına ve kapsamlı bir biçimde gösterilmesine yardımcı olan çeşitli tekniklerden oluşmaktadır (Gul vd. 2021). Model sonunda işlemler sonrasında çıkan bilgiler kullanıcısına daha iyi bir formatta sunulur (Nalawade ve Joshi, 2021).

4.3 Veri Madenciliği Modelleri

Veri madenciliği süreçlerinde kullanılan modeller, genellikle **tahminleyici** ve **tanımlayıcı** olmak üzere iki temel kategoride ele alınır. Bu modeller, verilerden bilgi çıkarma ve anlamlı ilişkiler kurma süreçlerinde farklı amaçlara hizmet eder (Han vd. 2012; Witten vd. 2016).

4.3.1 Tahminleyici Modeller

Tahminleyici modeller, sonuçları bilinen bir veri kümesinden hareketle bir model oluşturmayı ve bu modeli, sonuçları bilinmeyen veri kümelerindeki hedef değişkenlerin

tahmin edilmesi amacıyla kullanmayı hedefler. Örneğin, bir bankanın geçmiş kredi ödemelerine ilişkin verileri üzerinden bir model geliştirilerek, gelecekteki müşterilerin kredi ödeme davranışları tahmin edilebilir (Cristobal Romero vd. 2010). Bu tür modeller, genellikle "Ne olacak?" ve "Neden olacak?" gibi sorulara yanıt arar. Tahminleyici modelleme süreçlerinde sınıflandırma, regresyon ve zaman serileri analizi gibi yöntemler öne çıkar. Bu modeller, özellikle gelecekteki durumları öngörmek amacıyla veri madenciliği uygulamalarında sıklıkla tercih edilir (Fayyad vd. 1996). Örneğin, pazarlama alanında Heinzelmann'ın yaptığı çalışmada, tahminleyici analiz yöntemleriyle pazarlama stratejileri değerlendirilmiş ve en popüler yöntemlerden biri olarak işbirlikçi filtreleme yönteminin tavsiye sistemlerinde yaygın şekilde kullanıldığı sonucuna ulaşılmıştır (Heinzelmann, 2002).

4.3.2 Tanımlayıcı Modeller

Tanımlayıcı modeller, verideki yapıları, desenleri veya ilişkileri belirlemek ve veri setinin genel özelliklerini keşfetmek için kullanılan tekniklerdir. Bu modeller, geçmişe yönelik analizler yaparak, "Ne oldu?" sorusuna yanıt arar ve gelecekteki sonuçların anlaşılmasına katkı sağlar (Osmanbegović ve Suljić, 2012). Tahminleyici modeller geleceğe odaklanırken, tanımlayıcı modeller verinin mevcut yapısını ve geçmişteki davranışlarını analiz etmeye odaklanır. Bu tür modeller genellikle denetimsiz öğrenme kapsamında değerlendirilir ve kullanıcı müdahalesi olmadan verideki ilişkileri belirler (Agrawal vd. 1993). Örneğin, farklı gelir seviyelerine sahip ailelerin satın alma alışkanlıkları arasındaki benzerliklerin analiz edilmesi, tanımlayıcı modelleme kapsamında değerlendirilebilir. Birliktelik kuralları ve kümeleme algoritmaları tanımlayıcı modellerin en bilinen uygulamalarıdır (Han vd. 2012).

4.4 Veri Madenciliği Yöntemleri

Veri madenciliği yöntemleri, farklı veri analizi ve bilgi çıkarımı ihtiyaçlarına göre beş ana başlıkta incelenebilir, bunlar; sınıflandırma ve regresyon, kümeleme, birliktelik kuralları, dizi örüntüleri ve korelasyon analizidir. Sınıflandırma ve Regresyon, belirli bir hedef değişkeni tahmin etmeye yönelik yöntemlerdir. Sınıflandırma, verileri önceden

tanımlanmış kategorilere ayırırken, regresyon sürekli bir hedef değişkenin değerlerini tahmin eder (Quinlan, 1986). Kümeleme, verilerin doğal gruplarını belirlemek ve benzer özelliklere sahip veri noktalarını bir araya getirmek için kullanılır. Denetimsiz öğrenme kategorisindedir (Demšar vd. 2013). Birliktelik Kuralları, veriler arasındaki korelasyonları ve bağlantıları belirlemek için kullanılan bir yöntemdir. Örneğin, bir markette birlikte satın alınan ürünlerin analiz edilmesi birliktelik kurallarına bir örnektir (Agrawal vd. 1993).

4.4.1 Veri Sınıflandırma Yöntemi

Veri sınıflandırma yöntemi, veri madenciliğinin temel tekniklerinden birisidir. Temel teknik süreç, mevcut veri setini analiz ederek bir sınıflandırma modeli oluşturmayı, ardından bu modeli kullanarak yeni verileri doğru sınıflara ayırmayı içerir. Veri sınıflandırma tekniklerinin temel amacı, bir veri setindeki örnekleri önceden tanımlanmış kategorilere atamak, bu şekilde verileri örnek özelliklerine göre doğru bir biçimde sınıflandırabilen uygun bir model oluşturmaktır. Sınıflandırma için kullanılan model ya da algoritmaya sınıflandırıcı adı verilir.

Veri sınıflandırma teknolojileri arasında başlıcaları: Karar Ağaçları Algoritması, Destek Vektör Makineleri Algoritması, K-en Yakın Komşu Algoritması, Naive Bayes Algoritması, Genetik Algoritmalar, Sinir Ağları Algoritması, Lojistik Regresyon Algoritması vb. yer almaktadır. Bu algoritmalar bir sonraki bölümde detaylı olarak açıklanmaktadır.

4.4.2 Veri Kümeleme Yöntemi

Kümeleme, aynı gruptaki verilerin birbirine benzer, farklı gruplardaki verilerin ise farklı olması için veri nesnelerini gruplandırma tekniğidir. Bir küme, kendi içindeki elemanların birbiriyle benzer olduğu, ancak diğer kümelerin elemanlarından farklılık gösterdiği bir grup olarak tanımlanabilir (Sun, 2024). İşlemin özünde kümeleme uğraşı vardır. Yani veri kümesi bir dizi ayrı alt kümeye bölünür ve kümeleme analizinin amacı küme içi benzerliği en üst düzeye çıkarmak ve kümeler arası benzerliği en aza

indirmektir. Kümeleme, sınıflandırma yöntemlerinden farklı olarak, hedef niteliğe ihtiyaç duymaz ve tamamen etiketsiz veriler üzerinde çalışır (Li vd. 2024). Bu yöntem; istatistik, biyoloji ve makine öğrenmesi gibi birçok alanda yaygın olarak kullanılır. Örneğin, çevrimiçi alışveriş sistemlerinde, müşterilerin davranışlarına göre gruplandırılması ve bu grupların satın alma alışkanlıklarının analiz edilmesi kümelemeye dayalı bir uygulamadır.

Veri kümeleme teknikleri arasında başlıca K-ortalamalar kümeleme, hiyerarşik kümeleme, Gürültülü Uygulamaların Yoğunluk Tabanlı Mekansal Kümelenmesi (DBSCAN), Kümeleme Yapısını Belirlemek İçin Noktaların Sıralanması (OPTICS), Gaussian karışım modeli ve diğer yaygın algoritmalar yer almaktadır.

K-ortalamalar kümeleme analizi sürecinde, önce küme sayısı K belirlenir ve ardından K örnek veri kümesi rastgele olarak başlangıç küme merkezleri olarak seçilir. Benzerliği ölçmek için standart olarak Öklid mesafesi kullanılır ve kare hata kümeleme kriteri fonksiyonu olarak kullanılır. Ardından, hedef fonksiyonunun değerini en aza indirmek için yinelemeyi tekrarlar. K-ortalamalar algoritması basittir ve yüksek hesaplama verimliliğine sahiptir, ancak başlatmaya duyarlıdır ve yerel en iyiliğe düşmeye eğilimlidir (Zhang vd. 2021).

Hiyerarşik kümeleme, hiyerarşik bir ağaç (dendrogram) oluşturarak veri kümelemesini gerçekleştirir. Farklı yapılandırma yöntemlerine göre, toplayıcı hiyerarşik kümeleme ve bölücü hiyerarşik kümeleme olarak ayrılır. Hiyerarşik kümeleme, önceden belirlenmiş sayıda kategori gerektirmez ve yüksek esnekliğe sahiptir, ancak hesaplama karmaşıklığı yüksektir.

DBSCAN, verilerin kümelenmesini yoğunluğa göre ayıran yoğunluk tabanlı bir mekansal kümeleme yöntemidir. Yöntem, bir noktanın bulunduğu alanın yoğunluğunu, o noktanın etki alanındaki veri noktalarının sayısı açısından ölçer ve kümelere sınıflandırılabilecek kadar yoğun olan verileri sınıflandırmak için sürekli olarak komşu noktaları arar. Gürültülü mekansal verileri keyfi olarak şekillendirilmiş kümelere sınıflandırabilir (Huang ve Yang, 2024).

OPTICS, yoğunluk tabanlı ve parametrik olmayan bir kümeleme algoritmasıdır. Belirli bir uzayda bir dizi nokta bulutu verildiğinde, farklı yoğunluk eşiklerine göre kümeleme bilgileri içeren bir dizi nokta bulutu dizisi çıktısı verir. Bu bilgilere dayanarak, çıktı sonuçları esnek bir şekilde ayarlanabilir (Pan vd. 2022).

Gaussian Karışım Modeli (GMM), bir veri kümesini, birden fazla Gaussian dağılımının karışımını temel alarak analiz eden olasılıksal bir kümeleme yöntemidir. Bu model, verilerin birden fazla Gaussian dağılımından türediğini varsayar. Her bir Gaussian dağılımı, bir küme ile ilişkilendirilir ve her veri noktasının bir kümeye ait olma olasılığının bilindiğini varsayar.

Kümeleme yöntemleri: Bölümleme metotları, hiyeraşik metotlar, yoğunluk tabanlı metotlar, model tabanlı metotlar olarak sınıflandırılır. Bunlar arasında sıklıkla tercih edilen yöntem bölümleme metotları yöntemidir. En yaygın bölümleme yöntemlerinden biri K-Means yöntemidir (Layep, 2023).

4.4.3 Birliktelik Kuralı Madenciliği

Birliktelik analizi, veri öğeleri arasındaki ilişkileri araştırmak için metodolojik bir yaklaşım oluşturur ve sık görülen kalıpları, birliktelik kurallarını veya bağımlılıkları ortaya çıkarmak için kullanılabilir. Bu analitik teknik, belirli bir veri kümesi içinde birlikte ortaya çıkan öğe kümelerinin tanımlanmasını kolaylaştırır. Burada amaç, belirli bir veri seti içerisinde birlikte olan öğelerin belirlenmesidir.

Birliktelik kuralları, öğe kümeleri arasındaki ilişkiyi tanımlayan resmi ifadeler olarak hizmet eder. Birliktelik kurallarını değerlendirmek için üç önemli ölçüt destek, güven ve kaldırmadır. Destek, veri setinin tamamında öğe setlerinin yaygınlığını ölçerken güven, B öğe setinin A öğe setinin varlığında ortaya çıkma olasılığını değerlendirir. Kaldırma, iki öğe seti arasında var olan ilişkinin gücünü ölçer (Veri Bilimi Okulu, 2017).

Birliktelik kuralı madenciliğinde kullanılan önde gelen algoritmalar Apriori, FP-Growth, Carma ve Eclat algoritmalarını kapsamaktadır.

4.4.4 Dizi Örüntüsü Madenciliği

Zaman serisi verileri içindeki sık alt dizileri belirleme süreci, GSP (Genelleştirilmiş Sıralı Desen) ve PrefixSpan bu alanda yaygın olarak tanınan algoritmalar olmak üzere, sıra deseni madenciliği olarak adlandırılır.

GSP algoritması, tanımlanmış bir dizi adımı izleyerek dizi kalıplarının sistematik olarak genişletilmesi yoluyla sık alt dizileri tanımlar. GSP algoritmasının doğasında bulunan prosedürel adımlar şunları içerir: aday dizilerin oluşturulması, veritabanı taraması, budama ve yinelemeli tekrarlamadır.

Buna karşılık, PrefixSpan algoritması, veritabanının özyinelemeli projeksiyonu yoluyla sık alt dizilerin madenciliğini gerçekleştirir. PrefixSpan algoritması için operasyonel adımlar şunlardan oluşur: Veritabanı projeksiyonu ve özyinelemeli madencilik (Hevo data, 2018).

4.4.5 Korelasyon Analizi

Korelasyon analizi, değişkenler arasındaki ilişkilerin gücünü ve yönünü ölçmek için sıklıkla kullanılır. Yaygın olarak kullanılan metodolojiler arasında Pearson korelasyon katsayısı ve Spearman sıra korelasyon katsayısı bulunmaktadır.

Pearson korelasyon katsayısı öncelikle iki sürekli değişken arasındaki doğrusal ilişkiyi değerlendirmek için tasarlanmıştır ve değerleri -1 ile 1 arasında değişir.

Buna karşılık, Spearman sıra korelasyon katsayısı, iki değişken arasındaki monotonik ilişkiyi değerlendirmek için kullanılır ve hem sürekli hem de sıralı kategorik değişkenlere uygulanabilir.

4.5 Veri Madenciliği için Araçlar/Yazılımlar

Veri madenciliği sürecinde, büyük veri setlerinden anlamlı bilgiler çıkarabilmek için çeşitli yazılımlar ve araçlar kullanılmaktadır. Bu araçlar, veri temizleme, desen tanıma, sınıflandırma, kümeleme ve tahmin gibi farklı teknikleri destekleyerek analiz süreçlerini kolaylaştırır. Açık kaynak ve ticari yazılımlar olmak üzere iki ana gruba ayrılan bu araçlar, işletmelerden akademik araştırmalara kadar geniş bir kullanım alanına sahiptir. Yaygın kullanılanları aşağıda sunulmuştur.

Rapid Miner

Rapid Miner, makine öğrenimi ve veri madenciliği için geliştirilmiş bir araçtır. Analitik süreç iş akışlarını tasarlamak için "sürükle ve bırak" arayüzüne sahiptir. Verileri elektronik tablolar, ilişkisel veritabanları veya istatistiksel paket dosya formatlarından

yükleyebilir. WEKA algoritmalarını kullanarak süreci tanımlamak için XML kullanır. Rapid Miner, veri madenciliği için problemlerin çözülmesi ve eğilimlerin tahmini amacıyla kullanılır (Thakur ve Vinit Kumar, 2020).

Weka

Waikato Environment for Knowledge Analysis (WEKA), GNU lisansına sahip açık kaynaklı bir veri madenciliği aracıdır ve JAVA ile yazılmıştır. Çeşitli makine öğrenimi algoritmalarını kullanır ve sonuçları görselleştirir. 1997 yılında Waikato Üniversitesi tarafından tanıtılmıştır. WEKA üç grafik arayüze sahiptir:

- Explorer (ön işleme süreci için),
- Experimenter (makine öğrenimi algoritmalarını test etmek ve değerlendirmek için),
- Knowledge Flow (veritabanlarında bilgi keşfi (KDD) sürecini görselleştirmek için).

Ayrıca komutları yazmak için bir komut satırı gezgini de bulunmaktadır (Bansal ve Srivastava, 2018).

SAS Veri Madenciliği

SAS (Statistical Analysis Software), istatistiksel analiz ve veri işleme için kullanılan bir tescilli yazılımdır. Bu yazılım, metin analizi için bir programlama aracı olarak kullanılır (Thakur ve Vinit Kumar, 2020). Kuzey Karolina Eyalet Üniversitesi'nin tarımsal araştırma projesi olarak 1976 yılında kurulmuştur. Küçük, orta ve büyük veri kümeleri için iyi bir veri işleme kapasitesine sahiptir. Ancak en büyük dezavantajı oldukça pahalı olmasıdır (Bansal ve Srivastava, 2018).

Python

Python, yüksek seviyeli bir programlama dilidir ve veri madenciliği alanındaki kullanımı hızla artmaktadır. Kelime ve cümle ayrıştırma, konuşma etiketleme, sınıflandırma ve parçalara ayırma gibi Doğal Dil İşleme (NLP) işlemleri gerçekleştirebilir. Şirketler ve organizasyonlar Python'u; NLP, konu modelleme, yapay zeka bilimsel hesaplama, pazar analizi gibi alanlarda kullanmaktadır (Thakur ve Vinit Kumar, 2020). Python, veri bilimciler tarafından veri madenciliği ve diğer veri manipülasyon işlemleri için kullanılan bir araçtır. Açık kaynaklıdır, hızlıdır, taşınabilirdir ve diğer teknolojileri destekler (Bansal ve Srivastava, 2018).

R Paketi

R, metin analizi için açık kaynaklı bir yazılımdır. Doğal Dil İşleme (NLP), konu modelleme, kelime sıklığı belirleme, duygu analizi gibi işlemleri gerçekleştirebilecek araçlar içerir. Lemmatizasyon, ayrıştırma, ön işleme, kök sözcük çıkarımı, LDA ve diğer algoritmalara sahiptir (Thakur ve Vinit Kumar, 2020). R, veri madenciliği, veri analitiği, istatistiksel hesaplama ve grafikler için kullanılır. Özellikle araştırma ve akademik alanlarda tercih edilmektedir (Bansal ve Srivastava, 2018).

Diğer Yazılımlar: Orange, KNIME, Sisense, SSDT (SQL Server Data Tools), Apache Mahout, Oracle Data Mining, Rattle, DataMelt, IBM Cognos, IBM SPSS Modeler, SPSS, Teradata, Board, Dundas BI, Spark ve H2O'dur.

4.6 Veri Madenciliği Teknolojisinin Uygulama Alanları

Veri madenciliğinin; istatistik, matematik, makine öğrenimi, bilgi bilimi, yapay zekâ ve bilgisayar-iletişim gibi çeşitli alanların birleştiği bir alanı temsil ettiği düşünüldüğünde, uygulama alanları oldukça geniş ve farklıdır. Sıklıkla kullanıldığı alanlar ve uygulamaların örnekleri aşağıda verilmiştir:

Üretim: Ürün kalitesi takip programı, hata belirleme ve takibi, arz planlama, yeni ürünlerin geliştirilmesi,

Sağlık: Hastalıkların belirlenmesi, hasta bakım kalitesi belirleme ve takip, kişisel tedavi planlaması,

Eğitim: Öğrenci öğrenme süreç takip programı, davranışsal sınıflandırma, öğretmen performans değerlendirme,

Taşımacılık: Akıllı ulaşım sistemi, trafik kontrol, rota planlaması, yakıt takip,

Bilim-araştırma: Veriden bilgi üreten çalışmalar, veri yönetimi programı ve veri yönetim modelleri öneri sistemleri,

Telekominikasyon: Dolandırıcılık tespit sistemi, müşteri kayıp önleme programları, ağ güvenlik sistemleri,

Sosyal Medya: Kötüye kullanım tespit sistemi, tıklama akış analiz programı, reklam hedefleme, hedef kitle tespit ve takip sistemi.

Bankacılık: Veri madenciliğinin en yaygın uygulamalarından biri bankacılık sektöründe görülmektedir. Veri madenciliği teknolojisi finans kuruluşlarının iş

süreçlerini optimize etmesine, riskleri azaltmasına, müşteri deneyimini ve kârlılığı iyileştirmesine yardımcı olmaktadır. Veri madenciliği teknolojisinin finansal alandaki uygulamalarından biri de kredi riski yönetimidir. Veri madenciliği teknolojisi, müşterilerin kredi kayıtlarını, finansal işlemlerini ve demografik bilgilerini analiz ederek kredi riskini değerlendirmek için kredi puanlama modelleri oluşturur ve bu değerlendirme modelleri finansal kuruluşların kredi onaylarını, kredi limitlerini ve faiz oranlarını belirlemesine yardımcı olabilir.

Tavsiye sistemleri ağırlıklı olarak insan-bilgisayar etkileşimi veya bilgi erişimi (IR) dâhil olmak üzere komşuluk temelli metodolojiler ve teknikler kullanmaktadır. Bununla birlikte, bu sistemlerin çoğu temelde bir veri madenciliği tekniği olarak kavramsallaştırılabilecek bir algoritma tarafından desteklenmektedir. Gerçekten de, veri madenciliğinde karşılaşılan çok sayıda zorluk benzer şekilde tavsiye sistemleri bağlamında da mevcuttur (Tafralı, 2022).

5. MATERYAL VE METOT

Bu bölümde; araştırma yöntemi, çalışma grubu, veri toplama araçları, veri toplama süreci ve veri değerlendirme süreçleri ayrıntılı olarak açıklanmıştır.

5.1 Yöntem

Bu araştırmada izlenen metodoloji, veri keşfi ile başlayarak ihtiyaçları belirleme ve tanımlama sürecini içermektedir. İlk olarak, veri madenciliğinden anlamlı sonuçlar elde edebilmek için veri hazırlama aşaması gerçekleştirilmiştir. Özellikle büyük veri kümeleri ve aykırı değerlerin varlığı nedeniyle, ön işleme aşaması kritik bir adım olarak ele alınmıştır. Ön işleme yapılmadan elde edilen sonuçların yanıltıcı olabileceği göz önünde bulundurularak, veri temizleme ve düzenleme işlemleri uygulanmıştır.

Ardından, kullanıcıların segmente edilmesi amacıyla gruplandırma işlemi gerçekleştirilmiştir. Elde edilen segmentasyon sonuçları, kullanıcıları sınıflandırma ve tahmin modellerini uygulama sürecinde kullanılmıştır. Böylece, karar verme süreçlerine destek sağlayacak bir karar destek aracı oluşturulması hedeflenmiştir.

Son aşamada, kullanılan yöntem ve algoritmaların değerlendirilmesi yapılmış, bulgular detaylı bir şekilde raporlanmış ve sonuçlar doğrultusunda karar destek mekanizması oluşturulmuştur. İşlem akış şeması Şekil 5.1'de sunulmuş olup, süreçte yer alan her bir bileşen ayrıntılı şekilde açıklanmıştır.

Şekil 5.1 Araştırma süreci modeli

5.1.1 Veri Keşfi Aşaması

Veri keşfi adımı, çalışmanın amacına bağlı olan ihtiyaçların belirlenmesiyle başlar ve bu ihtiyaçlar daha fazla fayda ile eşleştirilmesi hedeflenir. Veri keşfi, bir veri setinin boyutlarını, değişkenlerin dağılımlarını, değişkenler arasındaki ilişkileri ve genel kalıpları anlamak amacıyla istatistiksel özetleme yöntemleri ve veri görselleştirme araçlarının kullanıldığı bir süreçtir. Bu süreç, modelleme veya daha derin analizlere geçmeden önce veri kalitesini değerlendirmenin yanı sıra, verilerle ilgili hipotezlerin geliştirilmesine olanak tanır (Tukey, 1977). Bir faaliyet içerisinde olan yapılar, pazarlama stratejileri, kaynakların dağıtımı, ürünlerin dağıtımı ve diğer işletme davranışları hakkında doğru ve uygun kararlar veremezse başarısızlığa uğrar. Doğru ve uygun iş kararları almak, en önemli işletme ihtiyaçlarından biridir. Veri keşfi, karar verme sürecinde kullanılabilecek basit bilgi kaynaklarını sunabilir.

5.1.2 Veri Hazırlama Aşaması

Veriyle çalışma başladığında bunu temelde veri madenciliği olarak kabul edebiliriz, bu nedenle verilerin neyi temsil ettiğini, neden ve nasıl dönüştürüldüğünü ve hangi çıkarımların yapılabileceğini daha ayrıntılı incelemek önemli olmaktadır. Sorun belirlenmeden veri madenciliği çalışmasıyla bir değer elde etmenin yollarını belirlemek zordur. Çözümün neye benzediği ya da benzeyeceği tahmin edildikten ve uygun veri bulunduktan sonra, veri madenciliği süreci devreye girmeye hazırdır (Doğan ve Arslantekin, 2023).

5.1.2.1 Veri Temizleme

Han vd. (2012), veri temizlemeyi "veri kalitesini artırmak ve hatalı sonuçların önüne geçmek için eksik, tutarsız ve hatalı verilerin analizden önce işlenmesi" olarak tanımlamışlardır. Eğer veri kümesinde talebe uygun olmayan veriler varsa, öncelikle veri temizleme süreci uygulanmalıdır. Örneğin, eksik, uygunsuz veya aykırı veriler olabilir. Bu tür veriler "gürültülü veri" olarak adlandırılır. Bu durumda, veri analizcinin kararına göre temizlenmeli ve yenileriyle değiştirilmelidir. Kullanılabilecek yöntemlerden bazıları şunlardır:

- Eksik, uygunsuz veya aykırı değerler veri kümesinden çıkarılabilir.
- Tüm eksik, uygunsuz veya aykırı değerler için genel bir sabit değer kullanılabilir (örneğin, "bilinmiyor"). Ancak tüm değişkenler için aynı sabit değerin kullanılması sorun yaratabilir.
- Değişkenin ortalaması hesaplanarak bu ortalama, eksik veya uygunsuz değer yerine kullanılabilir.
- Sadece bir sınıfa ait değişkenin ortalaması hesaplanarak eksik veya uygunsuz değer yerine bu kullanılabilir.
- Regresyon veya karar ağaçları gibi yöntemlerle eksik veya uygunsuz değerler tahmin edilebilir ve bu tahminler yerine kullanılabilir.

Veri kümesinde pozitif veya negatif aykırı değerler bulunabilir. Bu aykırı değerler yukarıdaki yöntemlerle temizlenmelidir. Ancak bazı durumlarda pozitif aykırı değerler

talep edilebilirken, sadece negatif aykırı değerler temizlenmelidir. Bu durumda, negatif aykırı değerler karar vericinin seçimiyle değiştirilmelidir.

Veri temizleme sürecinin temel amacı, veri setinin doğruluğunu, tutarlılığını ve analiz edilebilirliğini artırmaktır.

5.1.2.2 Veri Azaltma

Rahm ve Do (2000), veri azaltmayı "işlenebilir veri boyutlarını oluşturmak için orijinal veri kümesinden önemli bilgiyi koruyarak veri hacmini veya boyutunu azaltma süreci" olarak tanımlar. Analiz sürecinin beklenenden uzun sürebileceği düşünüldüğünde veri azaltma uygulanmalıdır. Yani yüksek boyutlu veri kümelerinde (büyük veri seti gibi) performansın artırılması için veri azaltma gereklidir. Analiz sonuçlarının değişmeyeceği veya iyileşmeyeceği düşünülüyorsa veri veya değişken sayısı azaltılabilir. Veri azaltma şu şekilde yapılabilir: Veri birleştirme, boyut azaltma, veri sıkıştırma, örnekleme, genelleştirmedir.

5.1.2.3 Veri Dönüşümü

Bazı durumlarda veriyi doğrudan kullanmak doğru olmayabilir. Değişkenlerin ortalamaları ve varyasyonları arasında büyük farklar varsa, bu farklar diğer değişkenlerin rolünü ciddi şekilde azaltabilir. Bu nedenle, veri dönüşümü için veri ayıklama işlemi uygulanmalıdır.

5.1.3 Kümeleme Aşaması

Kümeleme analizi, veri madenciliği alanında önemli bir tekniktir ve amacı, benzer özelliklere sahip verileri gruplamak, yani kümeler halinde organize etmektir. Kaufman ve Rousseeuw (1990), kümelemeyi "veri setindeki nesneleri benzerlik ölçütlerine göre gruplandırarak, verilerin yapısal özelliklerini anlamayı ve özetlemeyi sağlayan bir süreç" olarak tanımlamışlardır. Bu analiz, özellikle etiketlenmemiş verilerle çalışan veri madenciliği tekniklerinden biri olarak, veri setindeki örnekleri benzer özelliklere sahip gruplara ayırmayı amaçlar. Kümeleme analizi, veri içerisindeki yapıları keşfetmeye

yönelik bir yaklaşım sunar ve veri kümelerini, içerdikleri benzerliklere göre anlamlı şekilde gruplar. Örneğin, bir şirketin müşteri kayıtlarını ve niteliklerini gruplandırmak, aynı gruptaki müşterilerin birbirine diğer gruplardakilere göre daha çok benzediği birkaç grup ortaya çıkarabilir (Jain vd. 2000). Kümeleme, özellikle keşifsel veri analizi ve desene dayalı veri analizi için kullanılır ve farklı alanlarda, örneğin müşteri segmentasyonu, biyolojik veri analizi, pazar araştırması gibi birçok uygulama alanına sahiptir (Han vd. 2011; Xu ve Wunsch, 2005). Kümeleme, özellikle benzerlik veya mesafe ölçütleri kullanarak gruplama yapmayı sağlar, bu da veriler arasındaki örtük ilişkileri ortaya çıkarmada faydalıdır (Kaur vd. 2021). Bu süreç, ayrıca verilerin daha iyi anlaşılmasını, karar verme süreçlerinin iyileştirilmesini ve daha etkili modelleme yapılmasını sağlayabilir (Yun vd. 2020). Kümeleme işlemine geçmeden önce veri hazırlanmalı, kümeleme algoritmalarının uygulanması, kümeleme sonuçlarının çıkarılması ve en iyi kümeleme yönteminin belirlenmesi adımları yürütülmelidir.

5.1.3.1 Kümeleme Algoritmalarının Uygulanması

Birden fazla kümeleme yöntemi uygulanmalı ve yöntemler birbirleriyle karşılaştırılarak çalışmaya en uygun olanı seçilmelidir. Hangi algoritmaların kullanılacağı tanımlanmalı ve veri kümesini daha iyi kümeleyebilecek algoritmalar belirlenerek karar verilmelidir.

5.1.3.2 Tüm Kümeleme Sonuçlarının Çıkarılması

Kümeleme algoritmaları uygulandıktan sonra, tüm kümelerin özellikleri anlaşılmalıdır. Her bir küme, kümeleme sonuçlarından çıkarımlar yapılarak tanımlanmalıdır.

5.1.3.3 Çalışma için En İyi Kümeleme Yönteminin Seçilmesi

Her bir kümeleme yöntemi için kümeler anlaşıldıktan sonra, çalışmanın amacı doğrultusunda hangi kümeleme yönteminin sonuçlarının daha faydalı olduğuna bakılarak en iyi yöntem seçilmelidir.

5.1.4 Sınıflandırma Aşaması

Jain vd. (2000), sınıflandırmayı "veri örneklerini bir veya daha fazla sınıfa ayırma ve bu sınıfların sınırlarını belirleme işlemi" olarak tanımlar. Sınıflandırma, geçmiş verilerden öğrenilen bilgiyi kullanarak yeni verilere uygulanabilir bir model geliştirmeyi amaçlar. Veri madenciliğinde sınıflandırma teknikleri, büyük miktarda veriyi işleme kapasitesine sahiptir. Bu teknikler, kategorik sınıf etiketlerini tahmin etmek ve veriyi eğitim seti ve sınıf etiketlerine dayalı olarak sınıflandırmak için kullanılabilir. Ayrıca, yeni veri setlerini sınıflandırmak için de kullanılabilir. Sınıflandırma terimi, mevcut bilgiler temelinde bir kararın veya tahminin yapıldığı herhangi bir bağlamı kapsar (Hestie vd. 2009).

Önerilen metodoloji, sınıflandırma tekniklerinin bu özelliklerinden faydalanmaktadır. Ancak, sınıflandırma süreci doğru bir şekilde uygulanmalıdır; aksi takdirde yanlış ve yanıltıcı tahminler ortaya çıkabilir. Sınıflandırma aşamasının temel adımları aşağıda sunulmuştur.

5.1.4.1 Karar Sütununun Seçilmesi

Veri kümesindeki sınıflandırmayı temsil eden hedef değişken belirlenir. Bu değişken, modelin öğrenmesi için temel oluşturur. Kümeleme yöntemi seçildikten sonra, seçilen kümeleme yönteminin sonucu, sınıflandırma sürecinin ön işleme aşaması olarak kullanılabilir. Örneğin, müşteri memnuniyetini analiz eden bir veri setinde hedef değişken "Memnuniyet Düzeyi" olabilir. Bu yaklaşım, daha etkili ve güvenilir bir karar destek aracı hazırlamak açısından fayda sağlar. Çünkü bir kümenin veri seti, sınıflandırma süreci için eğitim veri setini daha iyi temsil edebilir. Bu durumda, uygulama sırasında belirli bir sütuna odaklanılarak sınıflandırma sürecinde karar sütunu olarak kullanılması sağlanabilir. Başlangıç veri kümesi karar sütununu içermediği için bu sütun, kümelerin özelliklerinden türetilebilir.

Eğitim veri setini hazırlamak için öncelikle bir sütun, karar sütunu olarak belirlenmelidir.

5.1.4.2 Veri Kümesini Eğitim ve Test Veri Seti Olarak Bölme

Veri kümesi, eğitim ve test veri seti olarak ikiye bölünmelidir. Eğitim seti, sınıflandırma modelini öğrenmek için kullanılırken, test seti modelin doğruluğunu ve genel performansını değerlendirmek için kullanılır. Bu sayede, eğitim veri seti sınıflandırma algoritması tarafından eğitilebilir ve ardından test verisi ile değerlendirilebilir.

5.1.4.3 Sınıflandırma Algoritmalarının Uygulanması

Eğitim veri seti, sınıflandırma algoritmasıyla eğitilir. Daha sonra, test veri seti, eğitim sonuçları kullanılarak sınıflandırma algoritması ile tahmin edilir. Çalışma kapsamında, birden fazla sınıflandırma algoritması uygulanmalı ve bu algoritmaların birbirleriyle karşılaştırılmasıyla en iyi olan seçilmelidir.

5.1.5 Karar Destek Aracının Hazırlanması

Kümeleme aşaması ve tahmin aşamasından elde edilen sonuçlara göre, bu modelle uyumlu algoritmalar değerlendirilmelidir. Bu değerlendirme önemlidir, çünkü model için en iyi algoritmalar, gelecekte daha gerçekçi sonuçlar ve karar önerileri sunacaktır.

Sınıflandırma algoritmasının seçiminde yalnızca eğitim sürecindeki değerlendirme sonuçları yeterli değildir. Çünkü tahmin aşamasında da bir sınıflandırma algoritması uygulanır ve bu sürecin doğruluğu büyük önem taşır. Bu nedenle, gelecekte daha gerçekçi sonuçlar ve etkili karar önerileri elde edebilmek için tahmin başarısı da dikkate alınmalıdır.

Bu değerlendirme sonucunda, en başarılı sınıflandırma algoritması seçilmeli ve ardından karar destek aracı bu sonuçlara göre hazırlanmalıdır.

5.2 Araştırma Alanı

Çalışma, Ocak 2022 ile Haziran 2023 tarihleri arasında toplanan ve 18 aylık bir süreyi kapsayan kullanıcı verilerinden oluşmaktadır. Bu dönemdeki veriler, kullanıcı

davranışlarını analiz etmek, eğilimleri belirlemek ve çeşitli parametreler doğrultusunda istatistiksel çıkarımlar yapmak amacıyla kullanılmıştır. Veriler, Van Yüzüncü Yıl Üniversitesi Ferit Melen Merkez Kütüphanesi otomasyonu üzerinden elde edilmiştir. Otomasyon sağlayıcı firmadan günlük olarak tutulmuş log dosyası şeklindeki veriler alınmıştır. Veriler günlük işlemleri kapsayan toplam 459 adet txt. formatlı dosyadan oluşmaktadır. Log dosyasını excel dosyası dönüşümü sonrasında toplam 328 bin 101 satırdan oluşmaktadır. Her bir satır IP tanımlı bir kullanıcının zaman ve tarih damgası, anahtar kelime, arama süresi, gösterim süresi ve bulunan satır sayısını gösteren değişkenleri kapsamaktadır.

Date	Duration	Term:	Facet:	Hits:	Start:	Profile:	Sources:	Rows:
1.01.2022 00:44	680ms	Van YÃ¼zÃ¼ncÃ¼ YÄ±l Ãœni	(SUBJECT_facet:"KÃ¼rt Dili	6	0	DEFAULT	[SD_ILS]	6
1.01.2022 01:09	225ms	Sosyal deÄŸiÅŸme_TÃ¼rkiye		133	48	DEFAULT	[SD_ILS]	12
1.01.2022 01:10	151ms	TÃ¼rk ÅŸiiri_20. yÃ¼zyÄ±	(SUBJECT_facet:"TÃ¼rk ede	6	0	DEFAULT	[SD_ILS]	6
1.01.2022 02:17	21ms		(SUBJECT_facet:"TÃ¼rk ede	75	12	DEFAULT	[SD_ILS]	12
1.01.2022 03:08	204ms	TÄ±p_TÃ¼rkiye_Tarih		29	0	DEFAULT	[SD_ILS]	12
1.01.2022 03:21	211ms	EÄŸitim_TÃ¼rkiye_Kongreler		111	0	DEFAULT	[SD_ILS]	12
1.01.2022 04:38	110ms	afet sosyolojisi		3	0	DEFAULT	[SD_ILS]	3
1.01.2022 04:43	27ms	af		11	0	DEFAULT	[SD_ILS]	11
1.01.2022 04:45	39ms			73199	0	DEFAULT	[SD_ILS]	12
1.01.2022 04:46	407ms	Sosyal sektÃ¶rler_GeliÅŸme_TÃ¼rkiye		66	0	DEFAULT	[SD_ILS]	66
1.01.2022 05:10	209ms	TÃ¼rk edebiyatÄ±_Åiir_20. yÃ¼zyÄ±l		1145	12	DEFAULT	[SD_ILS]	12
1.01.2022 07:13	83ms	TÃ¼rkiye_DÄ±ÅŸ iliÅŸkiler_NATO		7	0	DEFAULT	[SD_ILS]	7
1.01.2022 07:44	238ms	Avrupa BirliÄŸi_DÄ±ÅŸ iliÅŸkiler_TÃ¼rkiye		54	48	DEFAULT	[SD_ILS]	6
1.01.2022 07:58	232ms	TÃ¼rkiye_DÄ±ÅŸ iliÅŸkiler_A	(SUBJECT_facet:"DÄ±ÅŸ ekc	3	0	DEFAULT	[SD_ILS]	3
1.01.2022 08:22	1159ms	TÃ¼rk edebiyatÄ±_Denemeler_20. yÃ¼zyÄ±l		498	0	DEFAULT	[SD_ILS]	300
1.01.2022 08:56	174ms	TÄ±p_TÃ¼rkiye_Tarih		29	12	DEFAULT	[SD_ILS]	12

Şekil 5.2 Kullanıcı Log veri örneği

Örnek Log txt ve csv formatlı veri dosyası görüntüleri Şekil 5.2 ve 5.3'te gösterilmiştir.

Şekil 5.3 Kullanıcı Log veri örneği (txt. formatı)

5.3 Veri Kaynakları

Araştırma problemleri doğrultusunda yapılan analizlerde, kullanıcıların günlük log verileri incelenmiş ve bu veriler kullanılarak, kullanıcıların araştırmayı yarıda bırakma veya terk etme eğilimleri (beklenmedik kullanıcı etkinliği) hakkında uyarı verebilecek çeşitli değişkenler hesaplanıp uygulanmıştır. Ayrıca, mevcut otomasyon sistemlerinde yer almadığı için elde edilemeyen bir başka önemli değişken olarak araştırma başarı puanı değerlendirilmiştir. Bu başarı puanı, kullanıcıların arama performanslarına dayalı olarak oluşturulmuş hipotetik bir veri seti üzerinden hesaplanmış ve kullanıma sunulmuştur.

5.4 Araştırmada Kullanılan Değişkenler ve Araştırma Süreci

Çalışma iki aşamada gerçekleştirilmiştir:

Aşama 1: Veri Toplama ve Veri Ön İşleme

Araştırmanın ilk aşamasında, Yüzüncü Yıl Üniversitesi Ferit Melen Merkez Kütüphanesi'nden talep edilen kullanıcı etkinliğiyle ilgili büyük miktardaki veri (örneğin, eylem sayısı, zamanlaması ve sıklığı gibi), Otomasyon sağlayıcı firmadan log dosyası şeklinde toplanmıştır. Araştırma verilerinin kapsadığı dönemde toplam 328.101 kullanıcı etkinlik satırına ilişkin veri elde edilmiştir. Kullanıcıların gizliliği ve veri koruma

ilkelerine özen gösterilerek, kullanıcı etkinliklerine ilişkin veriler sunucu günlük dosyalarından alınmış ve bir veri dosyasında düzenlenmiştir.

Her bir eylem, kullanıcılara özgü benzersiz bir kimlik kodu (kullanıcı IP'si) ile temsil edilmiştir. Bu şekilde bireysel düzeyde izlemi anonim bir şekilde sağlamıştır. Ayrıca, her eylemin tarih ve saati ile türü (örneğin anahtar kelimeleri, arama sayısı, gösterim süresi) veri dosyasına dâhil edilmiştir.

Veri madenciliğinde analize başlamadan öncesinde yapılan bir diğer aşama ise veri ön işlemedir. Çalışmada farklı araştırma problemlerine yönelik farklı ön işleme teknikleri kullanılmıştır. Buna bağlı olarak ön işleme veri analizi başlığı altında ele alınmıştır.

Aşama 2: Veri Analizi

Bu aşamada, arama sürecinden ayrılma riski taşıyan kullanıcıları belirlemek amacıyla toplam yapılmış kullanım verileri üzerinden veri analizi gerçekleştirilmiştir. Kullanıcıların etkinlik değişkenleri hesaplanmıştır. Bununla birlikte kullanıcı eylemleri, veri setinde bulunan diğer kullanıcıların eylemleriyle karşılaştırılarak çeşitli ölçümler oluşturulmuştur. Bu ölçümler, her kullanıcı için olağandışı etkinliklere ilişkin uyarıları temsil eden bir indeks halinde oluşturulmuştur. Yarıda bırakma/sonlandırmayı tanımlamak için iki tür veri kullanılmıştır:

- *Aylık Etkinlik Yoğunluğu Verileri:* Her kullanıcının etkinlik verilerinin yoğunluğunu aylık olarak gösteren ilk veri türünden oluşmaktadır.

- *Bağıl Ölçümler:* Her kullanıcının günlük verilerinden oluşturulan ve gözlenen aktivitenin beklenip beklenmediğini anlamak için araştırma etkinlikleri ile diğer araştırmacıların benzer araştırma etkinlikleri arasında karşılaştırma yapılmasına olanak sağlayan göreceli ölçümlerden oluşmaktadır. Bu ölçümler, gözlemlenen etkinliğin beklenen bir etkinlik olup olmadığını anlamaya olanak tanımaktadır.

Değişken Grupları: Analiz için 18 değişken tanımlanmış ve üç gruba ayrılmıştır:

- **Grup 1:** Tüm süreç boyunca kullanıcıların toplam kullanımını ölçer: (V1–V3).
- **Grup 2:** Kullanıcı etkinlik yoğunluğunu ölçer ve beklenmeyen değişiklikleri (örneğin etkinlik azalması ya da etkinliği sonlandırma) tespit eder (V4–V8).
- **Grup 3:** Kullanıcıların yılsonu notlarını değerlendirir (V9-V14).

Bu analizlerde kullanılmak amacıyla on sekiz değişken tanımlandı ve üç grupta düzenlenmiştir (Çizelge 5.1). Tüm süreç boyunca yapılan araştırmalar toplam kullanıcı kullanımını üç farklı yönden incelemek üzere yapılandırılmıştır.

Çizelge 5.1 Kullanıcı eylemlerini ölçen değişkenler

Kullanıcıların Arama Eylemlerini Ölçen Genel Düzeydeki Değişkenler	
Değişken İsmi	**Değişken Açıklaması**
1. Tüm Kullanıcı Eylemlerinin Sayısı V1	Günlük dosyasındaki tüm kullanıcılar için tıklama akışı kayıtlarının toplam sayısı.
2. Bir aylık dönemde tüm kullanıcıların eylemlerinin sayısı V2	Her takvim ayına ait günlük dosyasındaki tüm kullanıcılara ait tıklama akışı kayıtlarının toplam sayısı.
3. Belirli bir takvim ayında tüm kullanıcı eylemlerinin ortalaması V3	Her takvim ayı için günlük dosyasındaki tüm kullanıcıların ortalama tıklama akışı kayıtları.
Kullanıcıların Arama Eylemlerinin Yoğunluğunu Kullanıcı Düzeyinde Ölçen Değişkenler	
4. Tüm süreçte bir kullanıcının yaptığı toplam eylem sayısı_V4	Günlük dosyasındaki belirli bir kullanıcıya ait tıklama akışı kayıtlarının toplam sayısı.
5. Tüm süreçte diğer kullanıcıların eylemlerinin ortalamasına göre eylemlerin bağıl ortalaması V5	Belirli bir kullanıcının ortalama etkinlik sayısı ile tüm dönem boyunca tüm kullanıcıların etkinliklerinin ortalaması arasındaki oran
6. Belirli bir dönemde (Bir Aylık) eylemlerinin sayısı V6	Her takvim ayına ait günlük dosyasındaki belirli bir kullanıcıya ait tıklama akışı kayıtlarının toplam sayısı
7. Belirli bir dönemde (örn. Bir Aylık) diğer kullanıcılarla ilgili göreceli eylem sayısı V7	Belirli bir kullanıcıya yönelik etkinlik sayısı ile belirli bir ay boyunca tüm kullanıcıların etkinliklerinin ortalaması arasındaki oran. Dönem boyunca her ay, her kullanıcı için göreceli bir puan hesaplandı.
8. Faaliyet günlerinin sayısı V8	Eylemlerin gerçekleştirildiği toplam gün sayısı.
Kullanıcı Durumunu tanımlayan Değişkenler (Kullanıcı Düzeyi)	
9. Kullanıcı performans notu V9	Arama performansına göre her kullanıcıya atanmış olan puan değeri.
10. Yılsonu performansı V10	Her kullanıcı için arama performansı değişkenine bağlı yılsonunda performans değeri. Bu değer "geçti", "kaldı" biçiminde oluşturuldu.
11. Arama süresi V11	Kullanıcının bir aramayı başlattığı süre ile bitirdiği süre arasında geçen toplam süre.
12. İşlem süresi V12	Kullanıcının sistemde kaldığı toplam süre.
13. Bulunan satır sayısı V13	Bir arama sonrasında bulunan sonuçlara ait toplam satır sayısı.
14. Gösterim süresi V14	Kullanıcının eriştiği kaynağı gösterdiği süre.

Ek Uyarı Değişkenleri:

Araştırma belirlenen sürelerde (bu çalışmada kullanıcı aktivitenin kontrol edilmesi için makul süre bir ay) olarak belirlenmiştir. Birinci gruptaki değişkenler genel araştırma düzeyiyle (V1–V3) ilgilidir. İkinci grup, etkinlikteki beklenmedik değişiklikleri (etkinlik azalması gibi) belirlemeyi amaçlayan kullanıcı düzeyinde değişkenlerle (V4–V8) ilgilidir. Üçüncü değişken grubu, kullanıcıların araştırma performans puanlarına (V9–V14) yöneliktir. Bu değişken aynı zamanda kullanıcının aramayı bırakıp bırakmadığını göstermektedir.

İlk üç gruptaki değişkenlerden yararlanılarak, dördüncü bir değişken grubu hesaplanmıştır. Bu dördüncü grup, uyarı indeksinde temsil edilen, beklenmeyen kullanıcı etkinliklerini yansıtan uyarı değişkenlerini içermektedir (Çizelge 5.2). Bu grupta, kullanıcıların etkinliksizlik durumlarını veya zaman içinde etkinlik yoğunluğundaki beklenmeyen değişiklikleri (azalmaları) yansıtan bir dizi ölçüm (14–18) tanımlanmıştır. Örneğin, bir ay boyunca herhangi bir etkinlik göstermeme veya düşük düzeyde etkinlik sergileme (bu durumda, takvim ayındaki genel kullanıcı etkinlik ortalamasının %50 oranında altında bir düzey) gibi durumlar bu ölçümlerle ifade edilmiştir.

Çizelge 5.2 Kullanıcı etkin olmama veya beklenmeyen uyarı değişkenleri

Etkin Olmama veya Beklenmeyen Uyarı Değişkenleri	
Değişken Adı	**Değişken Açıklaması**
15. Aylık hareketsizlik uyarılarının sayısı V15	Her ay için, etkinlik gözlemlenmediğinde bir uyarı bayrağı açıldı. Her Kullanıcı için bu uyarı bayrağının açıldığı ayların sayısı hesaplandı.
16. Genel Kullanıcı ortalamasına kıyasla aktivitedeki aylık düşüşlere ilişkin uyarı sayısı (düşük bağıl aktivite) V16	Kullanıcı eylemlerinin sayısı, o ay genel kullanıcı ortalamasının %50'inden daha düşük olduğunda bir uyarı bayrağı açıldı. Her kullanıcı için, bu bayrağın açıldığı ayların sayısı hesaplandı.
17. Etkin olmama uyarısının göründüğü takvim ayı V17	Arama süreci sırasında ilk kez hareketsizlik uyarı bayrağının belirdiği takvim ayı.
18. Düşük etkinlik uyarısının göründüğü takvim ayı V18	Arama süreci sırasında düşük etkinlik uyarı bayrağının ilk kez belirdiği takvim ayı.

Bu aşamanın sonunda, araştırmayı yarıda bırakma ya da sonlandırma tahmini için analiz gerçekleştirilmiştir. Üç ana bileşen analizde yer almıştır:

1. **Girdi Dosyası:** Otomasyondaki günlük dosyalardan alınan kullanıcı etkinlik değişkenlerini içerir.

2. **Uyarı Mekanizması:** Kullanıcının belirli bir zaman birimi için minimum etkinlik düzeyini tanımlayan ölçülebilir değişkenleri içeren dosya.

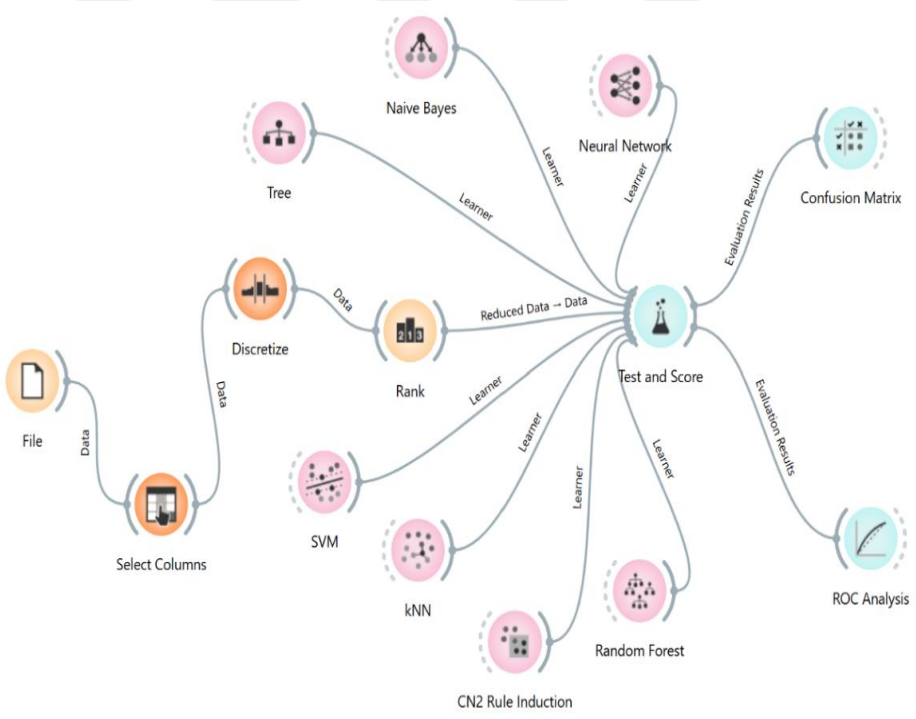
3. **Çıktı Dosyası:** Belirli kullanıcılar için olağandışı etkinliklerin tespit edilmesine dayalı olarak, potansiyel araştırmayı yarıda bırakma riski taşıyan kullanıcıları işaretleyen dosya.

5.5 Araştırma Problemlerine Göre Ön Analizler

Birincil araştırma konusu, öngörücü analiz üzerine odaklanmaktadır. Amaç, dijital ortamdaki kullanıcıların etkileşim verilerini kullanarak, belirli bir zaman dilimi içinde kullanıcıların araştırma performansını tahmin edebilen ve yönetebilen bir model geliştirmektir. Kullanıcıların nihai performanslarının "Geçti" veya "Kaldı" olarak kategorize edilmesi, bir sınıflandırma problemi ortaya çıkarmaktadır. Veri madenciliği alanında sınıflandırma görevleri için birçok algoritma bulunmaktadır. Bu algoritmaların etkinliği, veri kümesinin içsel özelliklerine bağlı olarak önemli ölçüde değişkenlik gösterebilir. Bu nedenle, veri kümesinin belirli özelliklerine iyi uyum sağlayan bir sınıflandırma algoritmasının seçilmiş olması gereklidir (Romero vd. 2013). Yaygın olarak kullanılan bir strateji, aynı analitik süreci farklı algoritmalarla uygulamak ve ardından belirlenen performans ölçütleri dâhilinde veri kümesinde en yüksek performansı sergileyen algoritmayı seçmektir (Akçapınar vd. 2013; Osmanbegović ve Suljić, 2012; Romero vd. 2010).

Sınıflandırma analizi sonuçlarını kullanmanın amacı, uygun bir algoritma belirlerken dikkate alınması gereken önemli bir değişkendir. Örneğin, yaygın olarak "kara kutu" olarak adlandırılan Naïve Bayes algoritması, sınıflandırma görevlerinde övgüye değer bir performans sergiler. Bu tür algoritmalar, sınıflandırmanın etkinliğinin model yorumlanabilirliği gerekliliğinden daha öncelikli olduğu senaryolarda tercih edilmelidir (Dreiseitl ve Ohno-Machado, 2002). Tersine, kural tabanlı veya ağaç tabanlı modeller de dâhil olmak üzere "beyaz kutu" algoritmaları, uzmanlıktan yoksun bireyler tarafından dahi kolayca anlaşılabilen sonuçlar üretir (Suljić ve Osmanbegović, 2012).

Mevcut araştırmada, literatürde sınıflandırma için yaygın olarak kullanılan yedi farklı algoritma seçilerek benzer bir metodoloji kullanılmıştır. Bunlar kural tabanlı, fonksiyon tabanlı, ağaç tabanlı ve çekirdek tabanlı algoritmaları kapsar: Naïve Bayes Algoritması, Rastgele Orman Algoritması, Destek Vektör Makineleri Algoritması, Karar Ağacı Algoritması, Yapay Sinir Ağları Algoritması, CN-2 Kuralları Algoritması ve k-En Yakın Komşu Algoritmalarıdır. Sınıflandırma algoritmalarının etkinliği, Doğru Sınıflandırma Oranı (CCR), Doğruluk, Hassasiyet, F-Ölçümü ve ROC Eğrisi Altındaki Alan (AUC) gibi performans ölçütleri kullanılarak değerlendirilmiştir. Sınıflandırma analizleri, Şekil 5.4'te açıklanan prosedüre uygun olarak, Orange Veri Madenciliği yazılımı (Demšar vd. 2013) kullanılarak gerçekleştirilmiştir. Bu çalışmada kullanılan sınıflandırma algoritmalarının ve performans ölçütlerinin ayrıntılı bir açıklaması aşağıda verilmiştir



Şekil 5.4 Sınıflandırma Performansıyla İlişkili Analiz Süreci

Naive Bayes Algoritması: Naive Bayes algoritması, olasılıksal bir sınıflayıcı olarak işlev görmektedir. Sınıflandırmayı etkileyen özelliklerin birbirinden bağımsız olduğunu varsayan Bayes Teoremine dayalı basit bir olasılıksal sınıflandırıcıdır. Gerçekte, özelliklerin tamamen bağımsız olma olasılığı düşük olsa da Naive Bayes

algoritması, Bayes formülüne dayanır ve bir dizi özellik verildiğinde bir kategorinin arka olasılığını hesaplayabilir. Naive Bayes algoritması, her kategorinin olasılığını hesaplamak için özellik verilerini kullanır ve ardından en yüksek olasılığa sahip kategoriyi nihai sonuç olarak seçer (Li vd. 2024). Eğitim verisi kullanılarak bağımlı değişkenlerle ilişkili koşullu olasılıklar hesaplanır ve ardından bu olasılıklar, yeni veri örneklerini sınıflandırmak için kullanılır. Bayes ağları, diğer yöntemlere kıyasla daha kapsamlıdır, çünkü tek bir değişken yerine herhangi bir olasılık dağılımını temsil edebilirler.

Karar Ağacı Algoritması: Karar ağaçları, veri madenciliği araştırmalarında tahminsel modeller oluşturmak için yaygın biçimde kullanılan bir tekniktir. Karar ağacı yöntemlerinden yararlanan birçok algoritma bulunmaktadır. Bunlar arasında; CveART, CHAID, C4.5 ve ID3 gibi algoritmalar yer almaktadır. Yinelemeli dikotomizer 3 (ID3), C4.5, sınıflandırma ve regresyon ağacı (CART), ki-kare otomatik etkileşim dedektörü (CHAID) ve hızlı tarafsız verimli istatistiksel ağaç (QUEST) yaygın karar ağacı modeli algoritmalarıdır (Shen ve Wang, 2024). Karar ağaçları bağlamında, veriyi en etkili şekilde ayıran değişken, seçilen bir bilgi ölçütüne (örneğin, Gini indeksi, kazanım oranı vb.) göre belirlenir ve bu, ilk düğümün oluşturulmasına yol açar. Bu işlem, yeni bölümlenen veri için tekrarlanır ve sonuçta ek alt düğümlere bölünme gerçekleşir. Bu süreç, her bir düğümde belirli bir sayıda eleman kalana kadar veya belirli bir düğüm derinliği elde edilene kadar devam eder. Karar ağaçları, çıktılarının netliği ve yorumlanabilirliği nedeniyle tercih edilmektedir.

Rastgele Orman Algoritması: Breiman tarafından geliştirilen Random Forest (RF) algoritması, ağaç tabanlı bir algoritmadır. RF algoritması, Karar ağaçları algoritmalarında yapılan tek bir ağaç üretmek yerine, bootstrap yöntemi kullanarak binlerce sayıda karar ağacını üretebilir. Karar ağacındaki her bir yapı için veri kümesinden rastgele bir biçimde örneklem seçimi yaplılır ve değişkenler de rastgele belirlenir. Seçilen örneklemin üçte ikisi, her bir karar ağacını oluşturmak için kullanılırken, her bir ağacın sınıflama başarısı veri kümesi üzerinde test edilir. Her bir karar ağacına bu süreç boyunca oy verilir. Karar ağaçları arasında en yüksek oyu alan ağaç (yani en düşük olan hata oranı olan) hangisi ise işlemin sonunda seçilir ve bu ağaç sınıflandırmanın temelini oluşturur.

Destek Vektör Makinaları Algoritması: Destek Vektör Makinaları algoritması, çekirdek (kernel) tabanlı bir algoritmadır. istatistiksel öğrenme teorisine dayanan yaygın olarak kullanılan bir desen tanıma ve ikili sınıflandırma algoritmasıdır. Giren ham verileri yüksek boyutlu uzayda belirli bir noktaya eşleyebilir. Farklı tipteki giriş örnek verileri yüksek boyutlu uzayda farklı konumlarda kümelendiğinden, uygun hiper düzlemi bularak farklı tipteki giriş verilerinin sınıflandırılmasını ve tanınmasını gerçekleştirebilir. Böylece, yeni veriler aynı yüksek boyutlu uzaya eşlendiğinde, eşlendikleri noktaların konumuna göre hangi kategoriye ait oldukları tahmin edilebilir (Zeng vd. 2024). Bu algoritmanın amacı, veri kümelerinde doğrusal olarak ayrılabilen sınıfları ayıran marjini en üst düzeye çıkarmak için doğrusal bir fonksiyonu bulmaktır. Diğer yandan, doğrusal olmayan veri kümelerinde veriler daha yüksek boyutlu bir uzaya aktarılır ve sınıflandırma, en büyük marjine sahip olan hiper-düzleme kullanılarak yapılır (Chang ve Lin, 2011).

Yapay Sinir Ağları Algoritması: Yapay sinir ağları algoritması, insan beynindeki nöronların nasıl çalıştığını öğrenerek geliştirilmiştir (Nisbet vd. 2009). Temelde matematiksel bir modeldir. İnsan beyni sinir ağına benzer şekilde, yapay sinir ağı yapay nöronlardan ve nöronlar arasındaki bağlantılardan oluşur. Bu nöronlar arasında, biri harici bilgileri almak için diğeri bilgi çıkışı için olmak üzere iki tür özel nöron vardır. Bu nedenle, sinir ağı, girdiden çıktıya bilgi için bir işleme sistemi olarak kabul edilebilir. Sinir ağına bilgi girerek, sınıflandırma sonucu çıktı katmanı aracılığıyla elde edilebilir. Sınıflama ve regresyon analizlerinde lojistik veya doğrusal fonksiyonlar kullanılarak hesaplamalar yapılmaktadır. Elde edilen sonuçlar, alanda uzman olmayanlar için anlaşılması ve yorumlanması zor olabilir (Hämäläinen ve Vinni, 2010). Yapısal olarak, analiz sürecindeki önemli bir bölüm gizli (genellikle "kara kutu" şeklinde isimlendirilmektedir) olduğundan, görüntü işleme, örüntü tanıma ve ses tanıma gibi alanlar için tercih edilebilmektedir (Stolzer, 2009).

Tahmin edici modellemelerin en yaygın kullanılan yöntemleri sınıflandırma ve regresyondur. İki model arasındaki temel fark sınıflandırma için hedef değişkenin kategorik, regresyon için ise hedef değişkenin süreklilik gösteren bir değere sahip olmasıdır.

CN-2 Kuralları Algoritması: CN-2 algoritması, kural tabanlı bir algoritma olarak sınıflandırılmaktadır. ID3 ve AQ sınıflandırma algoritmalarının güçlü olduğu

yönleri esas alınarak hazırlanmıştır (Clark ve Niblett, 1989). ID3 algoritmasında olduğu gibi gürültülü (veri girişinden veya veri toplanmasından kaynaklanan dışsal hatalar) verilerle de çalışabilme yeteneğine sahiptir. Ayrıca, AQ algoritmasındakine benzer şekilde, sınıflandırma sonucu "EĞER-İSE" kuralları üretmeyi de yapmaktadır.

K-en Yakın Komşuluk Algoritması: Algoritmanın temel prensibi, yeni bir veri noktasının sınıfını, en yakın k adet örneğin sınıf bilgilerine dayanarak çoğunluğun yer aldığı sınıfa atama yapmaktadır (Hand vd. 2001). Bu algoritmada sınıflandırma görevinde ki başarıyı etkileyen kritik bir belirleyici, k-değerinin uygun seçimidir. k'nin değeri aşırı büyükse, farklı sınıflardan örnekler tek bir sınıfta birleştirilebilir. Tersine, k'nin değeri çok düşük ayarlanırsa, aynı sınıfa ait olması gereken numuneler farklı bir sınıfa yanlış sınıflandırılabilir (Mitchell, 1997).

Aşağıda sınıflama algoritmalarının performansını karşılaştırmak için kullanılan performans metrikleri gösterilmektedir. Sınıflama modelinin doğru ve hatalı olarak sınıfladığı örneklerin bir araya getirildiği bir çapraz tablo, bu metrikleri hesaplamak için kullanılır. Sonuç olarak, Çizelge 5.3, kullanıcıların ders başarısı ile ilgili olarak "Geçti" ve "Kaldı" biçiminde iki değişkene sahip bir sınıflandırma problemine ait örnek bir çapraz tablo kullanılmıştır.

Doğru sınıflama oranı (DSO): Bir modelin doğru bir şekilde sınıflandırdığı örneklerin, toplam örnek sayısına oranı, Doğru Sınıflama Oranı (Accuracy) olarak adlandırılır ve aşağıda verilen formül kullanılarak hesaplanır. Sınıf değişkeninin dağılmadığı durumlarda ulaşılan metrikler yanıltıcı yorumlara yol açabilir, bu nedenle Accuracy, modelin performansını değerlendirmek için tek başına yetersiz bir ölçü olabilir. Accuracy, sınıflama görevlerinde modelin doğruluğunu ölçmek için kullanılan temel bir metriktir. Eşitlik 5.1'deki gibi ifade edilir:

$$\textit{Doğru Sınıflandırma Oranı (Accuracy)} = \textit{Doğru Tahmin Sayısı} / \textit{Toplam Örnek Sayısı} \tag{5.1}$$

Daha detaylı olarak Eşitlik 5.2'de ifade edilmiştir:

$$Accuracy = \frac{TP + TN}{TN + TP + FP + FN} \tag{5.2}$$

Burada:

TP (True Positive): Doğru bir şekilde pozitif sınıflandırılmış örnekler. Model tarafından "Kaldı" olarak sınıflandırılan başarısız kullanıcı sayısı,

TN (True Negative): Doğru bir şekilde negatif sınıflandırılmış örnekler. Model tarafından "Geçti" olarak sınıflandırılan başarılı kullanıcı sayısı,

FP (False Positive): Yanlış bir şekilde pozitif sınıflandırılmış örnekler (Hatalı alarm). Model tarafından "Geçti" olarak sınıflandırılan başarısız kullanıcı sayısı,

FN (False Negative): Yanlış bir şekilde negatif sınıflandırılmış örnekler. Model tarafından "Kaldı" olarak sınıflandırılan başarılı kullanıcı sayısıdır.

Çizelge 5.3 Örnek bir çapraz çizelge

		Gerçek Durum	
		Geçti	**Kaldı**
Tahmin	**Geçti**	TP (a)	FP (b)
	Kaldı	FN (c)	TP (d)

Duyarlılık (Sensitivity): Aşağıdaki formül kullanılarak yapılan hesaplama sonucunda, modelin başarısız kullanıcıları (Kaldı) doğru olarak sınıflandırma olasılığı ortaya çıkmaktadır (Eşitlik 5.3). Anma (Recall) metrik olarak da bilinir.

$$Duyarlılık = \frac{a}{a+b} \tag{5.3}$$

Seçicilik (Specificity): Yapılan hesaplama sonucunda, modelin başarılı kullanıcıları (Geçti) doğru olarak sınıflandırma olasılığı ortaya çıkmaktadır (Eşitlik 5.4).

$$Seçicilik = \frac{c}{c+d} \tag{5.4}$$

Kesinlik (Precision): Doğru sınıflandırılmış başarısız kullanıcı sayısının toplam başarısız kullanıcı sayısına oranıdır (Eşitlik 5.5).

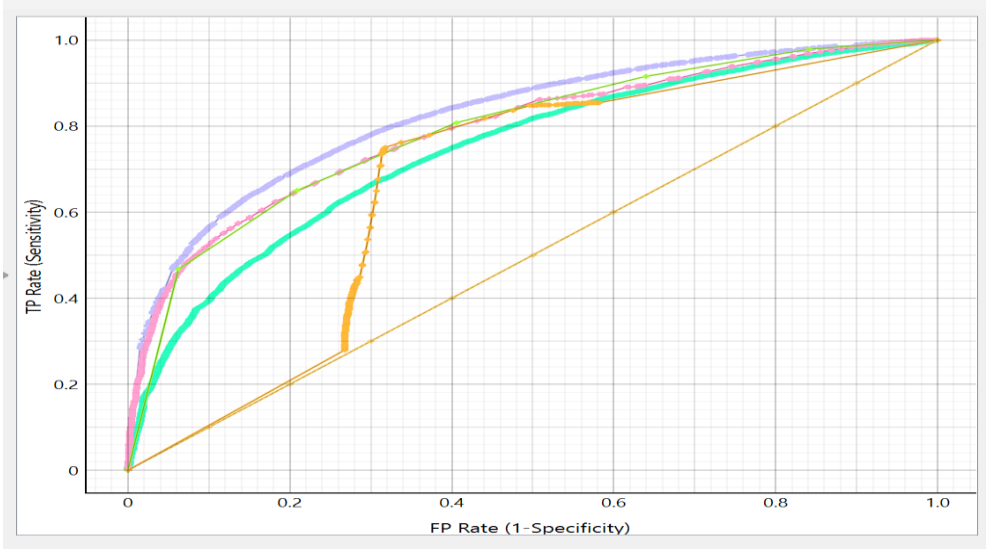
$$Kesinlik = \frac{a}{a+c} \tag{5.5}$$

F-Ölçütü (F-measure): Kesinlik ve duyarlılık ölçütlerinin tek başına kullanılması modellerin değerlendirilmesinde yanlı yorumlar yapılmasına neden olabilir, bu nedenle iki ölçütün beraber kullanıldığı F-Ölçütü üretilmiştir. F-Ölçütü, kesinlik ve duyarlılığın harmonik ortalamasıdır (Eşitlik 5.6).

$$F - \text{Ölçütü} = \frac{2a}{2a + b + c} \tag{5.6}$$

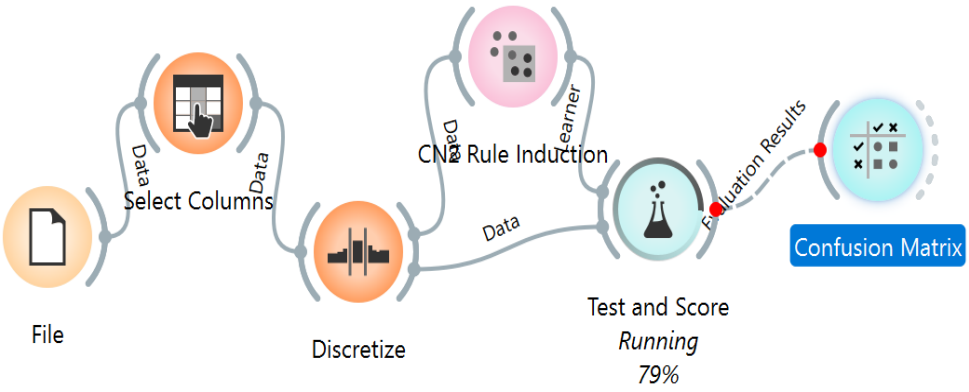
ROC (receiver operating characteristic) altında kalan alan (EAKA): İşaret işlemeyi alanında, gürültülü bir iletişim kanalında bir dedektörün doğru algılama oranının yanlış alarm oranına karşı çizilen grafik, ROC eğrisi olarak adlandırılır. Şekil 5.5'te örnek olarak verilen ve farklı sınıflayıcıların karşılaştırılmasını sağlayan ROC eğrilerinde, y ekseninde Doğru Pozitif Oranı (TPR), $a + (a + b)$ olarak hesaplanırken, x ekseninde Yanlış Pozitif Oranı (FPR), $c + (c + b)$ olarak hesaplanır. Farklı renklerde gösterilen eğriler, kullanılan yapay sinir ağları, CN2, k-en yakın komşu (k-NN), random forest, destek vektör makineleri (SVM), Naïve Bayes ve karar ağacı gibi algoritmaların performanslarını temsil etmektedir. ROC eğrisindeki her bir nokta, belirli bir sınıflayıcı tarafından oluşturulan bir modele karşılık gelmektedir. Grafikte kullanılan mor renk ile Destek Vektör Makineleri (SVM), yeşil ile Random Forest, pembe ile Yapay Sinir Ağları (ANN), sarı ile Naïve Bayes, açık mavi ile K-En Yakın Komşu (KNN), turuncu ile Karar Ağacı ve koyu yeşil ile CN2 Algoritması gösterilmektedir. Bu eğrinin altındaki alan ise sınıflayıcının etkinliğini gösteren bir metrik olarak kabul edilir.

Şekil 5.6'da gösterildiği gibi, sınıflama modelleri, ikinci araştırma problemi kapsamında dönem sonunda başarısız olma ihtimali yüksek olan kullanıcıların önceden tahmin edilmesi için kullanılmıştır. Bu, kullanıcıların etkileşim verilerinin yer aldığı analiz veri tabanından 4, 8, 12 ve 16 haftaya ait veriler kullanılarak oluşturulmuştur. Birinci araştırma problemi için en iyi sonuç veren algoritma ve ön işleme yöntemleri, sınıflama modelleri oluşturulurken dikkate alınmıştır.



Şekil 5. 5 Örnek ROC Eğrisi

18. haftadan alınan veriler kullanılarak oluşturulan sınıflama modeli ile elde edilen modellerin performansı karşılaştırılmıştır. Birinci araştırma problemi, buradan da elde edilen sınıflama modellerinin performanslarının karşılaştırılmasında ve sonuçların genelleştirilmesinde kullanılmıştır.

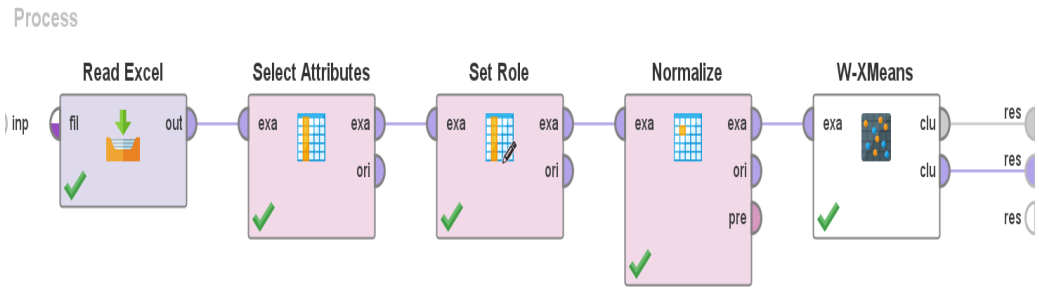


Şekil 5.6 Haftalık oluşturulan sınıflama problemine ilişkin analiz süreci

Üçüncü araştırma problemi çerçevesinde, kümeleme analizi, çevrimiçi öğrenme ortamında benzer kullanım davranışları gösteren çeşitli kullanıcı profillerini belirlemek için kullanılmıştır.

Şekil 5.7, Rapidminer VM yazılımının Weka eklentisinde bulunan W-XMeans algoritmasını kullanarak kümeleme analizlerini gerçekleştirme süreçlerini göstermektedir. Bu kümeleme teknikleriyle ilgili ayrıntılar aşağıda açıklanmıştır.

X-Ortalamalar: X-Ortalamalar kümeleme algoritması, K-means algoritmasının gelişmiş bir versiyonudur. K-means, veri setini sabit bir sayıda kümeye ayıran ve bu kümeleri belirli bir optimizasyon kriterine göre güncelleyen bir algoritma olarak yaygın kullanıma sahiptir. Ancak K-means algoritması, kümelerin sayısının önceden belirlenmesini gerektirir, bu da belirli veri setleri için sınırlayıcı olabilir. X-Ortalamalar Kümeleme algoritması, bu sorunu çözmek için tasarlanmış bir yöntemdir ve K-means algoritmasının başlangıçtaki küme sayısının adaptif bir şekilde belirlenmesini sağlar (Pelleg ve Moore, 2000). X-Ortalamalar Kümeleme algoritması, veri setinde en uygun küme sayısını dinamik olarak belirler. Bu işlem, her adımda mevcut küme sayısını değerlendirip gerektiğinde daha fazla küme oluşturur ya da mevcut kümeleri birleştirir. Algoritma, küme sayısının her bir aşamada nasıl değişmesi gerektiğini belirlemek için bir istatistiksel yaklaşım kullanır. X-Ortalamalar, özellikle büyük ve karmaşık veri setlerinde, daha doğru ve verimli sonuçlar elde etmek için kullanılır. Yöntem, veri kümesinin iç yapısını keşfetmede daha esnek ve otomatik bir yaklaşım sunarak, modelin genel performansını artırır (Cicek ve Gok, 2020; Xu, 2010). Bu algoritmanın avantajlarından biri, verinin doğasında bulunan gizli yapıları keşfetme yeteneğidir. Ayrıca, küme sayısının önceden belirlenmesine gerek olmadığı için kullanıcıdan daha az bilgi gerektirir ve veri kümesi büyüdükçe dinamik olarak uyum sağlar (Xu, 2010). Ayrıca, Bayes Bilgi Kriteri (BIC)'nin kullanılmasıyla algoritma optimal küme sayısının belirlenmesini sağlayabilmektedir.



Şekil 5.6 X Ortalamalar kümeleme analiz süreci

5.6 Araştırmanın İç ve Dış Geçerliği

Araştırmanın her bir sınıflama analizi sonucunda elde edilmiş sonuçların güvenilir bir şekilde genellenebilmesi için 5-katlı Çapraz Geçerlilik (5-fold Cross-Validation) yöntemi tercih edilmiştir. Bu yöntemde veri, 5 eşit alt gruba bölünür. İlk dört grup, sınıflama modelinin oluşturulması için eğitim verisi olarak kullanılırken, geri kalan bir grup test verisi olarak değerlendirilir. Bu süreç, her bir alt grubun bir kez test verisi olarak kullanılması sağlanacak şekilde tekrarlanır. Bu şekilde, her biri farklı bir test grubuna ait olmak üzere 5 farklı sınıflama modeli oluşturulur ve her modelin performansı ayrı ayrı hesaplanır. Son olarak, bu performans ölçümlerinin ortalaması alınarak, elde edilen modelin genel performansı değerlendirilir (Kohavi, 1995; Stone, 1974). Araştırmada sunulan bulgular, bu ortalama performans ölçümlerine dayalı olarak oluşturulmuştur.

Kümeleme analizlerinde ise, optimal küme sayısının belirlenmesi amacıyla çapraz geçerlilik yöntemi uygulanmıştır. Bu yöntem, verideki en uygun küme sayısının güvenilir bir şekilde tespit edilmesine olanak sağlamıştır (Han vd. 2012). Bu tür bir yaklaşım, hem sınıflama hem de kümeleme analizlerinin iç geçerliğini artırarak, elde edilen sonuçların daha tutarlı ve güvenilir olmasını sağlamaktadır.

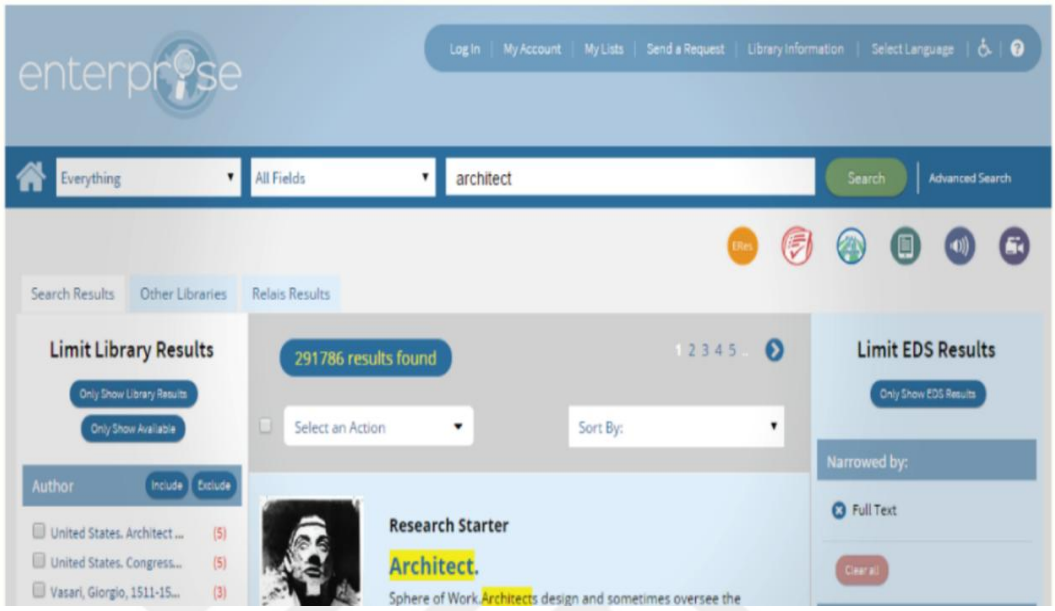
Veri madenciliği çalışmalarında dış geçerlik, oluşturulan modellerin performansının yeni ve farklı bir veri seti üzerinde test edilmesi ile değerlendirilir (Hastie vd. 2009). Ancak bu çalışmada geliştirilen öğrenme ortamı ilk kez uygulandığından, oluşturulan tahmin ve kümeleme modellerinin dış geçerliliği henüz doğrulanamamıştır.

5.7 Çevrimiçi Arama Ortamı

SirsiDynix, kütüphanelerin kataloglama, ödünç verme, kullanıcı yönetimi ve araştırma işlemlerini dijitalleştiren ve kolaylaştıran kapsamlı bir kütüphane yönetim sistem sağlayıcısıdır. Çeşitli özellikleriyle özellikle büyük kütüphane sistemlerinde tercih edilir. SirsiDynix'in teknik özellikleri, sayfa görüntüsü, Şekil 5.8 ve araştırma sayfası bileşenleri sunulmuştur:

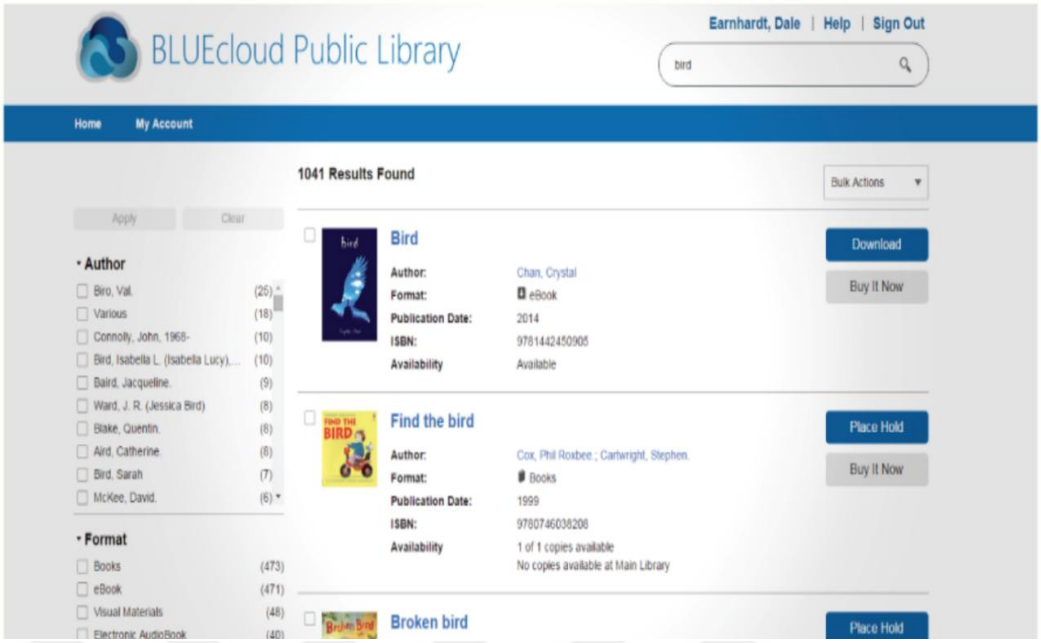
- **Arama Kutusu (Search Box):** Sayfanın üst kısmında, kullanıcıların basit veya gelişmiş arama yapabileceği bir alan olarak konumlandırılmıştır. Kullanıcıların anahtar kelime, yazar adı, başlık veya konu gibi kriterlerle

arama yapmalarına olanak tanır. Gelişmiş arama seçenekleri de sunarak, kullanıcıların aramalarını daha spesifik hale getirmelerini sağlar.



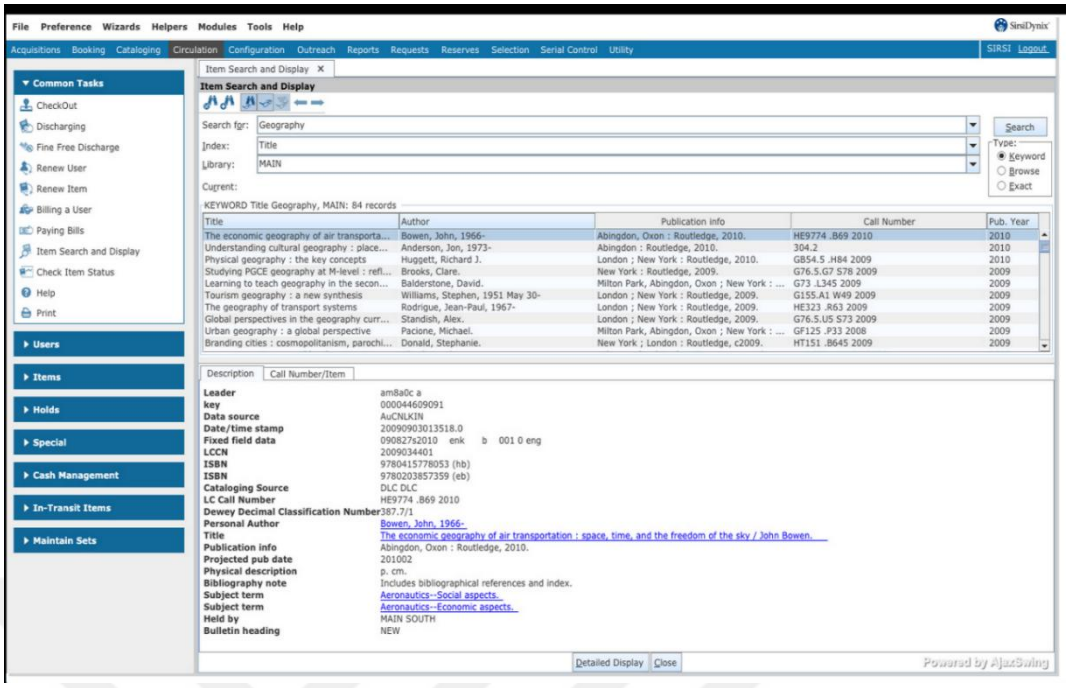
Şekil 5. 7 SirsiDynix'in araştırma (arama) sayfası

- **Filtreler (Filters):** Sol panelde, format, yayın tarihi ve dil gibi filtre seçeneklerini içerir. Arama sonuçlarını daraltmak için kullanılan bölümdür. Kullanıcılar, materyal türü (kitap, dergi, e-kitap vb.), yayın tarihi, dil ve diğer özelliklere göre filtreleme kullanabilirler.
- **Arama Sonuçları Listesi (Search Results List):** Şekil 5.9'da görüldüğü üzere, sayfanın merkezinde görüntülenir ve kullanıcıların arama kriterlerine uygun materyalleri listenmektedir. Her bir sonuç, başlık, yazar, yayın tarihi ve mevcutluk durumu gibi temel bilgileri içerir.



Şekil 5. 8 SirsiDynix'in arama sonuç sayfası

- **Kullanıcı Seçenekleri (User Options):** Sağ panelde, favoriler, ödünç alınan materyaller ve rezervasyonlar gibi kullanıcıya özel araçlar bulunur.
- **Materyal Detayları (Item Details):** Kullanıcılar, arama sonuçlarındaki bir materyale tıkladıklarında, detaylı bilgilere erişebilirler. Bu bölümde özet, konu başlıkları, ISBN ve raf yeri gibi bilgiler bulunur.
- **Kullanıcı Hesabı Bağlantıları (User Account Links):** Sayfanın üst kısmında veya sağ üst köşede yer alır. Kullanıcılar, hesaplarına giriş yaparak ödünç aldıkları materyalleri, rezervasyonlarını ve favori listelerini görüntüleyebilirler.
- **Erişim ve Rezervasyon Seçenekleri (Access and Reservation Options):** Kullanıcılar, arama sonuçları üzerinden materyalleri rezerve edebilir, dijital içeriklere erişebilir veya ödünç alma işlemleri yapabilirler.
- **Kütüphane İçi Yardım:** Araştırma sırasında kütüphane personeliyle iletişim amaçlı canlı sohbet veya destek paneli ve arama ve erişimle ilgili rehberlik amaçlı kullanılmak üzere sık sorulan sorular (SSS) alanları yer almaktadır. Ayrıca Şekil 5.10'da dolaşımdaki materyallerin anlık olarak izlenmesi amaçlı kullanılan sayfa bulunmaktadır.



Şekil 5. 9 SirsiDynix'in dolaşımdaki materyalleri gösteren sayfası

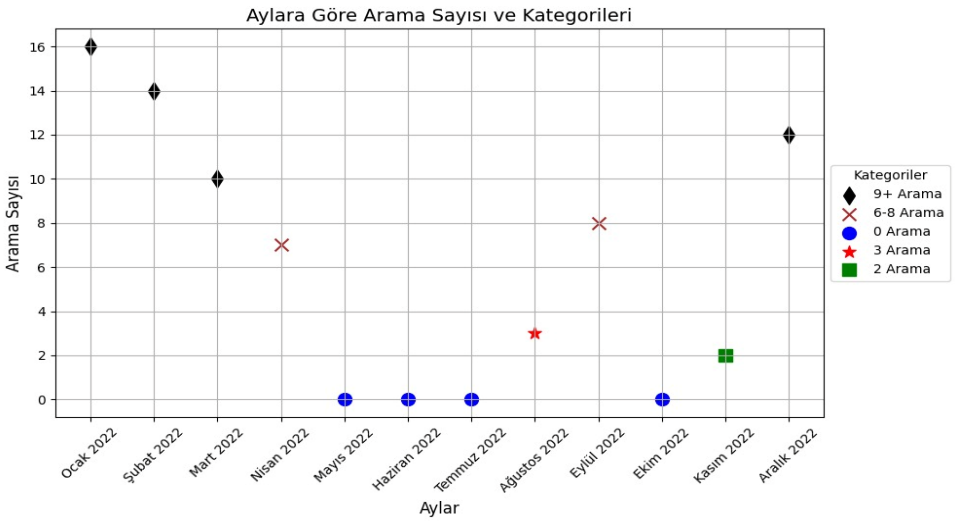
Bu bileşenler, kullanıcıların kütüphane kaynaklarına kolay ve hızlı bir şekilde erişmelerini sağlamak amacıyla tasarlanmıştır.

6. BULGULAR VE TARTIŞMA

Bu bölümde, analiz sonucundan elde edilmiş olan bulgular araştırmanın amacına göre ele alınmıştır. Araştırmanın asıl problemi "kütüphane otomasyonundan bilgi erişim ve kütüphane kullanım süreçlerinden elde edilen verilerin veri madenciliği yöntemleri ile analiz edilerek kullanıcıların bilgi arama davranışlarının modellenmesinde kullanılıp kullanılmayacağının araştırılmasıdır" biçimindedir. Belirlenen araştırma problemleriyle ilgili elde edilen bulgular şu şekildedir.

6.1 Kullanıcı Etkinlikleri ve Aramayı Yarıda Bırakma/Terk Etme Riskine Yönelik Uyarılar

Otomasyon veri setindeki kullanıcı etkinlik yoğunluğuna dayalı olarak aramayı yarıda bırakma/terk etme riskiyle ilgili uyarılar sağlayabilecek değişkenler ve ölçümler tanımlanmıştır. Kullanıcı etkinlik yoğunluğu, tutulabilecek farklı veriler aracılığıyla kullanıcılara sunulan farklı özelliklerle ilişkilendirilerek analiz edilebilir. Örneğin, Şekil 6.1'de bir kullanıcının belirlenen aylara göre arama etkinlik sayılarını göstermektedir.



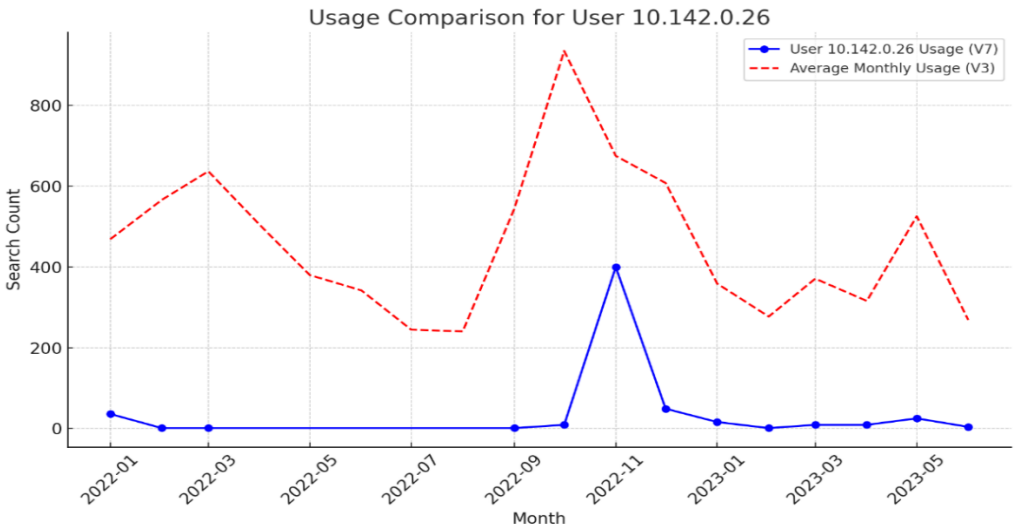
Şekil 6.1 Bir kullanıcının aylara göre arama etkinlik sayılarına bir örnek

Ancak, yalnızca bir kullanıcının etkinliği incelenerek yeterli bilgiye ulaşılamaz. Bu nedenle, bir kullanıcının etkinliği, diğer kullanıcıların etkinlikleriyle

karşılaştırılmalıdır. Bu karşılaştırma, gözlemlenen etkinliğin genel araştırma yeterliliğine göre beklenen bir davranış olup olmadığını anlamak için gereklidir.

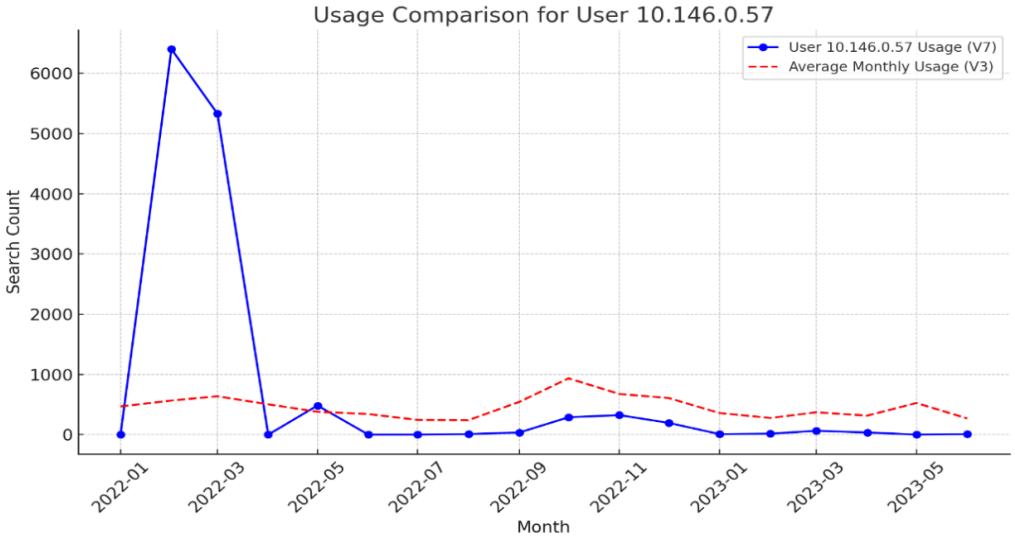
Kullanıcı etkinlikleri incelendiğinde farklı davranış kalıpları tespit edilmiştir. Şekil 6.2, 6.3, 6.4, 6.5 ve 6.6 aramayı yarıda bırakma/terk etme riski taşıyan kullanıcıları işaret edebilecek farklı davranışlara dair örnekler sunmaktadır. Bu beş şekil ile her biri farklı kullanıcı için aylık bazda toplam arama başlatma kayıtlarının (V6) diğer kullanıcıların aylık ortalama arama başlatma kayıtlarının yoğunluğuyla (V3) karşılaştırıldığındaki sonuçları göstermektedir.

Kullanıcı 1: Diğer kullanıcı ortalamasına benzer bir arama sürec kalıbına sahiptir, ancak bazı dönemlerde arama etkinlikleri genelden daha az olduğu görülmektedir (Şekil 6.2).



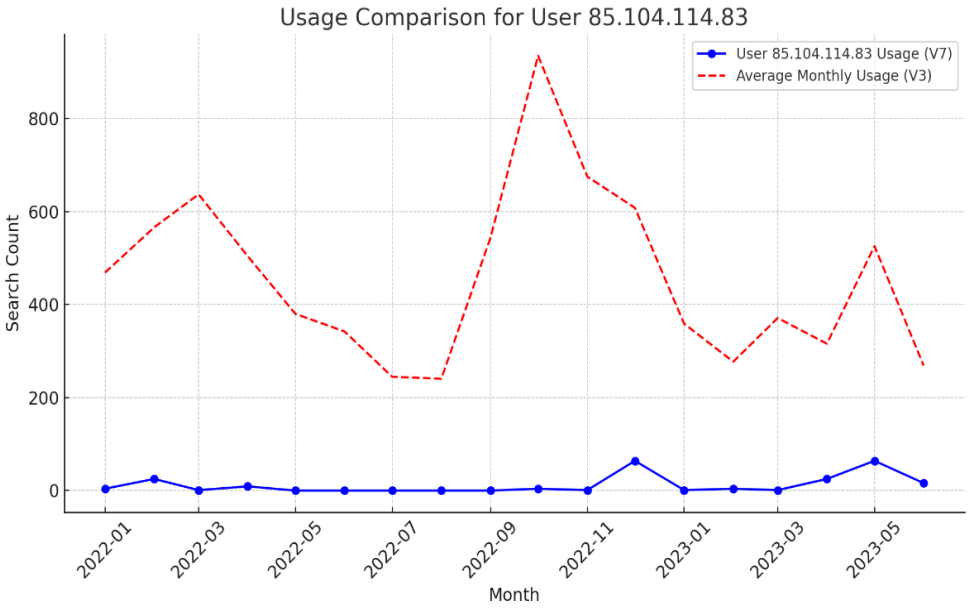
Şekil 6.2 Kullanıcı 1'in aktivite yoğunluğunun her takvim ayında diğer tüm kullanıcıların ortalama yoğunluğuna kıyasla durumu

Kullanıcı 2: Ocak ve Şubat aylarında özellikle çok yüksek sayıda arama etkinliğine sahip olduğu Şubat ayından sonra ise genel ortalamaya benzer etkinlik gösterdiği izlenmektedir (Şekil 6.3).



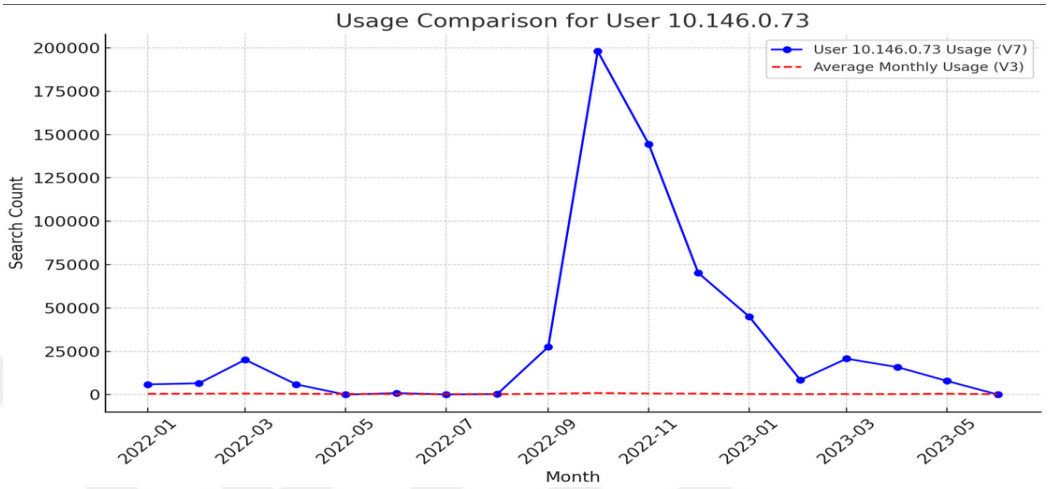
Şekil 6.3 Kullanıcı 2'nin aktivite yoğunluğunun her takvim ayında diğer tüm kullanıcıların ortalama yoğunluğuna kıyasla durumu

Kullanıcı 3: İlk aydan itibaren arama etkinliği çok düşük sadece belirli birkaç arama etkinliğine sahip bir kullanımı görülmektedir (Şekil 6.4).



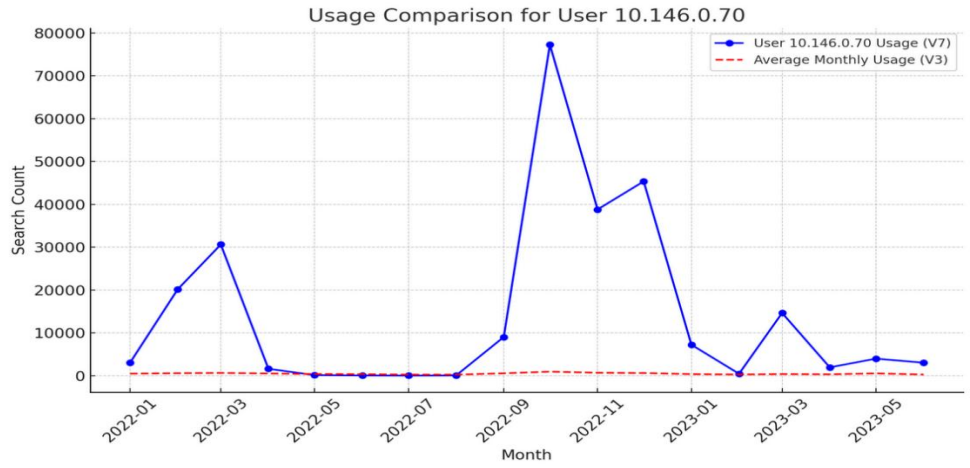
Şekil 6.4 Kullanıcı 3'ün aktivite yoğunluğunun her takvim ayında diğer tüm kullanıcıların ortalama yoğunluğuna kıyasla durumu

Kullanıcı 4: Temmuz Eylül ayları arasında özellikle çok yüksek sayıda arama etkinliğine sahip olduğu diğer aylarda ise genel ortalamaya benzer etkinlik gösterdiği izlenmektedir (Şekil 6.5).



Şekil 6.5 Kullanıcı 4'ün aktivite yoğunluğunun her takvim ayında diğer tüm kullanıcıların ortalama yoğunluğuna kıyasla durumu

Kullanıcı 5: Hem Temmuz Eylül ayları arasında çok yüksek sayıda arama etkinliğine sahip olduğu hem de diğer aylarda da genel ortalama üzerinde bir etkinliğe sahip olduğu izlenmektedir (Şekil 6.6).

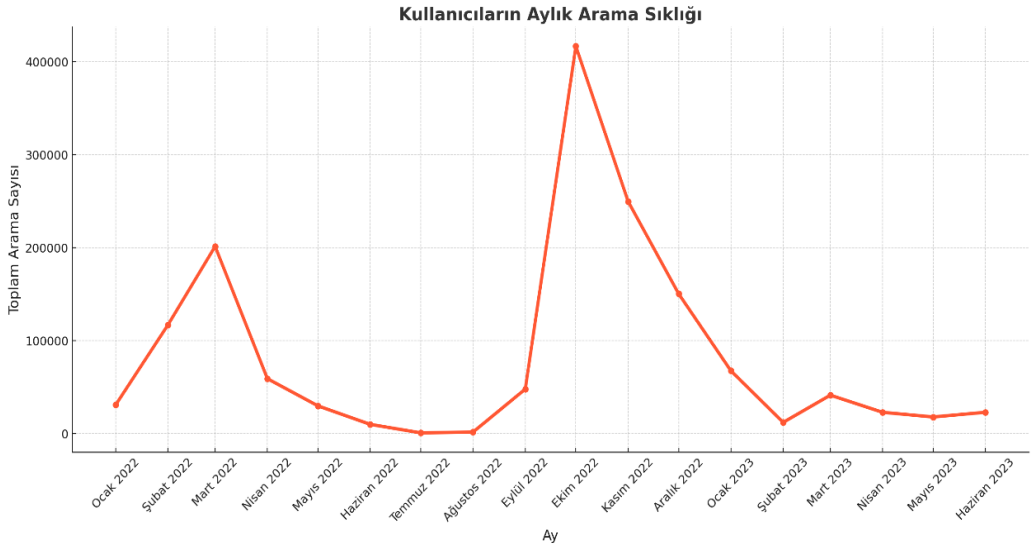


Şekil 6.6 Kullanıcı 5'in aktivite yoğunluğunun her takvim ayında diğer tüm kullanıcıların ortalama yoğunluğuna kıyasla durumu

Bu beş farklı durum, kullanıcıların kütüphane otomasyonunu beklenildiği gibi kullanıp kullanmadıklarını belirlemek için basit bir etkinlik izleme işlemi sunmaktadır. Bu kullanıcılar hakkında otomasyon yöneticisine gerçek zamanlı bir uyarı gönderilmesi sağlanmıştır.

6.2 Arama Yarıda Bırakma/ Terk Etme Uyarılarına Yönelik Tanımlanan Değişken ve Ölçümlerin Test Edilmesi

Kullanıcılar, genel olarak araştıma süreci boyunca otomasyon kullanımında belli düzeyde aktif görünmektedirler. Arama etkinliklerinde genel olarak bir sürekliliğin olması yanında kullanıcıların belirli tarih aralığında oldukça yoğun kullandıkları gözlemlenmektedir. Araştırma alanımızın bir üniversite kütüphanesi olmasından kaynaklı olarak dönemin ortasında ve final sınavı gibi üniversite sınav dönemi öncesinde yoğun bir etkinlik gözlemlemektedir (Şekil 6.7).



Şekil 6.7 Kullanıcıların otomasyon kullanım etkinliklerinin sıklığı ($N_{etkinlik}$ = 141.082)

6.3 Araştırma Problemlerine Göre Analizler

İlk araştırma problemine yönelik olarak altı farklı analiz gerçekleştirilmiş ve analizlere ait ayrıntılar Çizelge 6.1'de verilen altı farklı analiz gerçekleştirilmiştir. Çalışmada, bir veri kümesi üzerinde çeşitli veri ön işleme tekniklerinin ve özellik seçimi

yöntemlerinin sınıflandırma performansı üzerine etkileri incelenmiştir. İlk olarak, ham veri üzerinde herhangi bir işlem yapılmadan temel bir sınıflandırma modeli oluşturulmuş ve performansı değerlendirilmiştir. Ardından, veri dönüştürme tekniklerinin (örneğin, normalizasyon, ölçeklendirme) sınıflandırma başarısına etkileri araştırılmıştır. Son olarak, farklı özellik seçimi yöntemleri (örneğin, korelasyon analizi, bilgi kazancı) uygulanarak, modelin performansındaki değişimler gözlemlenmiştir. Tüm analizlerde, tutarlılık sağlamak amacıyla aynı sınıflandırma algoritmaları ve performans metrikleri kullanılmıştır.

Çizelge 6.1 Araştırma problemi kapsamında yapılmış olan analizler

No	**Ön İşleme Yöntemi**		**Sınıflama Yöntemleri**	**Performans Metrikleri**	**Sonuçlar**
	Veri Dönüştürme	**Özellik Seçme**			
1	Eşit Frekans	Kazanç Oranı	Yapay Sinir Ağları	Duyarlılık Seçicilik DSO EAKA F-Ölçütü	Çizelge 6.6
2	Eşit Frekans	Gini İndeksi	CN2 Kuralları		Çizelge 6.5
3	Eşit Frekans	DVM Ağırlıkları	k- En Yakın Komşuluk		Çizelge 6.4
4	Eşit Frekans	Yok	Random Forest		Çizelge 6.3
5	Eşit Frekans	Yok	Destek Vektör M.		Çizelge 6.2
6	Yok	Yok	Naive Bayes		Çizelge 6.1

Birinci analizde değişkenlere herhangi bir müdahale yapılmamıştır ve tüm değişkenler herhangi bir veri dönüştürme işlemi yapılmadan analize katılmıştır. Çizelge 6.2'de verilen analiz sonuçları incelendiğinde en yüksek doğru sınıflama oranına %91'le birden çok algoritma (Random Forest ve Naive Bayes) performansı sonuçları ile ulaşılmıştır. Diğer algoritmalardan yapay sinir ağları %55 ve destek vektör makineleri algoritmaları % 49 oranında sınıflama performansı ile en düşük iki algoritma olarak görülmüş, CN2 kuralları, k-en yakın komşuluk ve karar ağacı algoritmalarının performansı ise %78 ve üzerinde sergiledikleri görülmüştür.

Veri Madenciliği araştırmalarında ön işleme aşaması önemli bir yere sahiptir. Bu aşamada uygulanan müdahaleler, modellerin öngörü performansını artırma potansiyeline sahiptir. Bu bağlamda, başlangıç analizlerinden elde edilen bulgularla karşılaştırıldığında daha iyi sonuçlar elde edilip edilemeyeceğini belirlemek amacıyla iki farklı ön işleme yöntemi değerlendirilmiştir.

Çizelge 6.2 Verilerin sürekli olduğu durumda yapılan analizler

Yöntem	DSO	Duyarlılık	Seçicilik	EAKA	F-Ölçütü
Yapay Sinir Ağları	0.56	0.56	0.43	0.50	0.55
CN2 Kuralları	0.86	0.85	0.85	0.86	0.85
k- En Yakın Komşuluk	0.88	0.87	0.82	0.85	**0.87**
Random Forest	0.91	0.91	0.90	0.91	**0.91**
Destek Vektör M.	0.58	0.58	0.31	0.47	0.49
Naive Bayes	0.92	0.91	0.91	0.89	**0.91**
Karar Ağacı	0.78	0.78	0.72	0.70	0.78

İlk alt problem ile ilgili olarak, ön işleme aşamasında kullanılan çeşitli veri dönüştürme tekniklerinin sınıflandırma performans metrikleri üzerinde etkili olup olmadığını incelemek üzere bir araştırma gerçekleştirilmiştir. Bu doğrultuda, mevcut literatürde yaygın olarak kullanılan iki veri dönüştürme tekniğinin uygulandığı sınıflandırma modellerinin etkinlikleri karşılaştırılmıştır. İlk analizde, sürekli değişkenler eşit frekans yöntemi kullanılarak üç eşit frekans aralığına bölünmüş ve bu şekilde incelenmiştir. İkinci analizde ise değişkenler, eşit genişlik yöntemi ile üç eşit genişlik aralığına ayrılmış ve aynı analitik prosedür bu değişkenlerle tekrar gerçekleştirilmiştir. Bu iki analize ait sınıflandırma sonuçları sırasıyla Çizelge 6.3 ve Çizelge 6.4'te sunulmaktadır.

Çizelge 6.3 Verilerin eşit frekans yöntemi ile kesikli hale dönüştürüldüğü durumda yapılan analiz sonuçları

Yöntem	DSO	Duyarlılık	Seçicilik	EAKA	F-Ölçütü
Yapay Sinir Ağları	0.93	0.93	0.93	0.98	0.93
CN2 Kuralları	0.93	0.93	0.93	0.97	0.93
k- En Yakın Komşuluk	0.92	0.92	0.92	0.96	0.92
Random Forest	0.93	0.93	0.93	0.97	0.93
Destek Vektör M.	0.45	0.45	0.57	0.53	0.44
Naive Bayes	0.90	0.90	0.94	0.96	0.90
Karar Ağacı	0.72	0.72	0.68	0.66	0.72

Verilere uygulanan eşit frekans ve eşit genişlik hale dönüştürülme adımları sonrasında özellikle ile Yapay sinir ağları, karar ağacı ve CN2 kuralları algoritmalarının performans sonuçlarının doğru sınıflama oranlarında %38 ile %21 arasında artırmışlardır. Eşit frekans ve eşit genişlik yöntemlerine ilişkin sonuçları karşılaştırdığımızda ise eşit

genişlik yönteminin özellikle araştırma sürecinde başarısız olan kullanıcıları tahmin edilmesinde tüm algoritmalarda en iyi performansı sergilediğini görmekteyiz.

Çizelge 6.4 Verilerin eşit genişlik yöntemi ile kesikli hale dönüştürüldüğü durumda yapılan analiz sonuçları

Yöntem	DSO	Duyarlılık	Seçicilik	EAKA	F-Ölçütü
Yapay Sinir Ağları	0.93	0.93	0.93	0.98	0.93
CN2 Kuralları	0.93	0.93	0.93	0.97	0.93
k- En Yakın Komşuluk	0.92	0.92	0.92	0.96	0.92
Random Forest	0.93	0.93	0.93	0.97	0.93
Destek Vektör M.	0.45	0.45	0.57	0.52	0.44
Naive Bayes	0.90	0.90	0.94	0.96	0.90
Karar Ağacı	0.93	0.93	0.93	0.97	0.93

Çizelge 6.2'de sonuçları verilmiş olan taban modelle (herhangi bir müdahale yok) veri dönüştürme yöntemlerinin performanslarını karşılaştırıldığımızda ise her iki veri dönüştürme işlemi sonuçlarında da algoritmalardan çoğunun taban modele kıyasla sınıflama performanslarının artırdığını gözlemlemekteyiz. Yapay sinir ağları algoritmasında bu değişim %55'ten %93 gibi çok büyük bir sınıflama performansına dönüşmüştür. Sadece destek vektör makinaları algoritması düşük performans göstermiştir.

Bu araştırma problemine ilişkin ele alınan bir diğer sorgulama, veri ön işleme aşamasında kullanılan farklı özellik seçimi yöntemlerinin sınıflandırma performansı üzerindeki etkisidir. Özellik seçimi işleminin temel amacı, tahmin analizi çerçevesinde hedef değişkeni öngörmede daha büyük öneme sahip olan değişkenleri belirlemektir. Bu sayede, tüm değişkenlerin dâhil edilmesi yerine, daha az sayıda değişken kullanılarak tahmin işlemi gerçekleştirilebilmektedir.

Önemli değişkenlerin belirlenmesi amacıyla birçok algoritma kullanılmakta olup, bu algoritmalar her bir değişkene kullanılan yönteme bağlı olarak bir skor atamaktadır. Daha sonra değişkenler, bu skor doğrultusunda sıralanmakta ve belirli bir sayıda seçilen değişken kullanılarak analizler yapılmaktadır.

Bu çalışmada, özellik seçimi süreci, Orange veri madenciliği yazılımı kullanılarak üç farklı özellik seçimi algoritması (kazanım oranı, Gini indeksi, reliefF oranı) çerçevesinde gerçekleştirilmiştir.

Özellik seçme analizi sonucu her bir değişkenin farklı algoritmalardan aldıkları puanlar Çizelge 6.5.'de sunulmuştur.

Çizelge 6.5 Farklı özellikler seçme yöntemine göre değişkenlerin önem puanları

No	Değişkenler	Kazanç Oranı	Gini İndeksi	ReliefF Oranı
1	Tüm Kullanıcı Eylemlerinin Sayısı_V1	0	0	0
2	Belirli bir dönemde (Bir Aylık) tüm kullanıcıların eylemlerinin sayısı_V2	0.069	0.043	**0.019**
3	Belirli bir takvim ayında tüm kullanıcı eylemlerinin ortalaması_V3	0.057	0.036	**0.016**
4	Tüm süreçte bir kullanıcının yaptığı toplam eylem sayısı_V4	0.014	0.009	**0.031**
5	Diğer kullanıcıların eylemlerinin ortalamasına göre eylemlerin bağıl ortalaması_V5	0.694	0.337	**0.298**
6	Belirli bir dönemde (Bir Aylık) eylemlerinin sayısı_V6	0.038	0.024	**0.077**
7	Belirli bir dönemde (örn. Bir Aylık) diğer kullanıcılarla ilgili göreceli eylem sayısı_V7	0.011	0.007	**0.079**
8	Faaliyet günlerinin sayısı_V8	0.349	0.185	**0.176**
9	Kullanıcı performans notu search count_V9	0.694	0.337	**0.298**
10	Arama süresi_V10	0.001	0	0
11	İşlem süresi_V11	0.0.17	0.010	**0.001**
12	Bulunan satır sayısı_V12	0.003	0.002	0
13	Gösterim süresi_V13	0.009	0.005	**0.001**

Verilen veri setinde hangi özellik seçimi yönteminin üstün sonuçlar sağlayacağını belirlemek amacıyla üç ayrı analiz gerçekleştirilmiştir. İlk analizde, Kazanç oranı kriterine göre en yüksek skora sahip ilk 10 değişken belirlenmiş ve tüm algoritmalar için performans metrikleri hesaplanılmıştır. İkinci analizde, Gini indeksine göre en yüksek skoru alan ilk 10 değişken seçilmiş ve sınıflandırma analizleri tekrar edilmiştir. Son olarak, ReliefF yöntemiyle en yüksek skora sahip ilk 10 değişken seçilmiş ve benzer analizler bir kez daha gerçekleştirilmiştir.

Bu üç araştırma kapsamında yapılan sınıflama analizlerinin sonuçları sırasıyla Çizelge 6.6, Çizelge 6.7 ve Çizelge 6.8'de sunulmuştur. Her bir analizde, değişkenler daha önceki deneylerde en iyi sonuçları veren eşit genişlik yöntemini kullanarak kesikli hale dönüştürülmüş ve analizler bu yaklaşım doğrultusunda gerçekleştirilmiştir.

Çizelge 6.6 Kazanç oranı yöntemine göre en önemli 10 değişken seçilerek yapılan analiz sonucu

Yöntem	DSO	Duyarlılık	Seçicilik	EAKA	F-Ölçütü
Yapay Sinir Ağları	0.53	0.53	0.46	0.50	0.53
CN2 Kuralları	0.72	0.72	0.64	0.79	**0.75**
k- En Yakın Komşuluk	0.73	0.73	0.66	0.78	0.73
Random Forest	**0.75**	**0.75**	**0.69**	**0.82**	**0.75**
Destek Vektör M.	0.65	0.65	0.34	0.50	0.51
Naive Bayes	0.70	0.69	0.60	0.74	0.70
Karar Ağacı	0.72	0.72	0.68	0.66	0.72

Çıkan analiz sonuçları önceki araştırma sorusundan bağımsız değerlendirildiğinde (Çizelge 6.6, 6.7 ve 6.8 verilerinin karşılaştırılmasıyla) en yüksek doğru sınıflama oranına Destek Vektör Makinaları (%99) algoritmaları ile ReliefF skoru ile özellik seçme ve bu değere çok yakın bir diğer değere (CNN2 kuralları %98) ise Gini indeksi ile yapılmış olan özellik seçme işlemiyle ulaşıldığını görmekteyiz (Çizelge 6.7 ve Çizelge 6.8).

Çizelge 6.7 Gini indeksi yöntemine göre en önemli 10 değişken seçilerek yapılan analiz sonucu

Yöntem	DSO	Duyarlılık	Seçicilik	EAKA	F-Ölçütü
Yapay Sinir Ağları	**0.99**	0.95	0.95	0.99	0.95
CN2 Kuralları	**0.98**	**0.98**	**0.97**	**0.99**	**0.98**
k- En Yakın Komşuluk	0.95	0.95	0.95	0.95	0.95
Random Forest	0.95	0.95	0.95	0.95	0.95
Destek Vektör M.	0.95	0.95	0.91	0.96	0.95
Naive Bayes	0.88	0.88	0.86	0.97	0.88
Karar Ağacı	0.96	0.96	0.98	0.97	0.96

Farklı bir alt problemde, özellik seçimi aşamasında daha iyi performans gösteren ve daha az sayıda değişken kullanılarak oluşturulan sınıflama modellerinin performans sergileyen ve az sayıda değişkenle ile oluşturulan sınıflama modellerinin performansı, tüm değişkenler kullanılarak oluşturulan modellerle karşılaştırılmıştır. Bu karşılaştırmayı kolaylaştırmak amacıyla, önceki analizde elde edilen en yüksek sınıflama oranlarını içeren Çizelge 6.4 ile seçilen önemli değişkenlerle elde edilen en iyi sınıflama oranlarını sunan Çizelge 6.8 karşılaştırıldığında, ReliefF Skora kullanılarak seçilen 10 değişkenle

elde edilen sınıflama sonuçlarının, tüm değişkenlerle elde edilen sonuçlardan daha başarılı olduğu gözlemlenmiştir.

Çizelge 6.8 ReliefF skoru yöntemine göre en önemli 10 değişken seçilerek

Yöntem	DSO	Duyarlılık	Seçicilik	EAKA	F-Ölçütü
Yapay Sinir Ağları	0.96	0.96	0.97	0.99	0.96
CN2 Kuralları	0.98	0.98	0.97	0.99	0.98
k- En Yakın Komşuluk	0.95	0.95	0.95	0.95	0.95
Random Forest	0.95	0.95	0.95	0.95	0.95
Destek Vektör M.	**0.99**	**0.99**	**0.99**	**1.00**	**0.99**
Naive Bayes	0.91	0.91	0.89	0.97	0.91
Karar Ağacı	0.96	0.96	0.98	0.97	0.96

Doğru sınıflama oranları açısından sonuçlar değerlendirildiğinde, DVM algoritmasındaki performansın %55 oranında arttığı gözlenmektedir. Ayrıca, CN2 kural tabanlı algoritması doğru sınıflama oranında %5'lik bir artış gösterirken, diğer algoritmalar doğru sınıflama oranlarını %1 ile %3 arasında arttırmıştır. Algoritmaların başarılı ve başarısız kullanıcıları ayırt etme kapasitesinin ölçüsünün bir göstergesi olan EAKA metriği açısından incelendiğinde ise, daha az değişken kullanıldığında tüm algoritmaların ayırt ediciliklerinde bir artış gösterdiği anlaşılmaktadır.

Bir sonraki araştırma sorusunda, ilk araştırma probleminin en iyi performans gösteren sınıflama algoritması ve ön işleme teknikleri, araştıma süreci sırasında başarısız olma ihtimali yüksek kullanıcıların tahmin edilmesinde kullanılmıştır. Bu bağlamda, daha erken haftalarda başarısız olma olasılığı yüksek kullanıcıları tahmin etme olasılığı, 4, 8, 12 ve 16. Haftalardan elde edilen verilerle oluşturulan sınıflama modellerinin, tüm veri kullanılarak (18. haftada) oluşturulan modelle karşılaştırılması yoluyla araştırılmıştır.

Modeller, ilk araştırma probleminde elde edilen bulgulardan yararlanılarak ve o bağlamda belirlenen en iyi performansa sahip algoritma ile ön işleme teknikleri kullanılarak geliştirilmiştir. Oluşturulan tüm modellerde, ilk araştırma probleminde seçilen on değişken (Çizelge 6.5) ve yüksek sınıflama performansına sahip CN-2 kural tabanlı algoritma kullanılmıştır. Veriler eşit genişlik yöntemi kullanılarak kesikli hale dönüştürülmüştür. Araştırma performansını temsil eden değişkeni de ilk problemde olduğu gibi "Kaldı" (n = 23.053, başarı puanı <= 40) ve "Geçti" (n = 12.124, başarı puanı > 40) şeklinde kodlanmıştır.

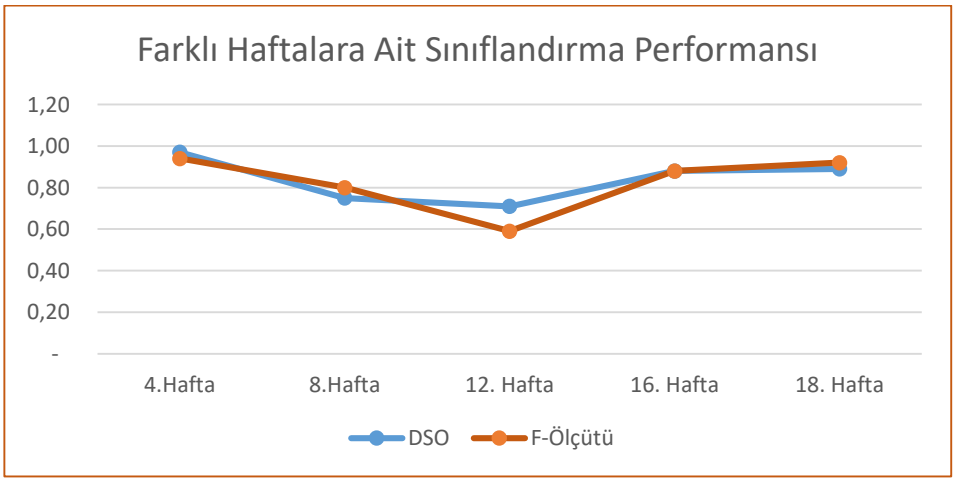
4, 8, 12, 16 ve temel model (18. Hafta) açısından performans metriklerine ilişkin sonuçlar Çizelge 6.9'da sunulmaktadır. Bu sonuçlar, 5-k Çapraz geçerlilik yöntemi kullanılarak ulaşılmış olunan genelleştirilmiş sonuçlardır.

Çizelge 6.9 CN-2 Kural Algoritması'na göre farklı haftalardaki veriler ile oluşturulan sınıflama modeline ilişkin analiz sonuçları

Haftalar	DSO	Duyarlılık	Seçicilik	EAKA	F-Ölçütü
4.Hafta	0.97	0.97	0.82	0.92	0.94
8.Hafta	0.75	0.91	0.57	0.81	0.80
12. Hafta	0.71	0.52	0.83	0.83	0.59
16. Hafta	0.88	0.88	0.78	0.93	0.88
18. Hafta	0.89	0.96	0.78	0.94	0.92

Bu araştırma sorusunda, başarısız olacak kullanıcıların tahmin edilmesi amaçlandığından, sınıflama modellerinin çeşitli haftalardaki çapraz geçerlilik sonuçlarından elde edilen "F-Ölçütü" değerleri ile modellerin genel verimliliğini değerlendirmek için kullanılan "DSO" değerleri Çizelge 6.9'da verilmiştir.

12. Haftadan sonra oluşturulan sınıflama modellerinin doğru sınıflama oranının, Şekil 6.8'de görsel olarak sunulan bu iki metriğe ilişkin grafikte artış eğilimi gösterdiği görülmektedir.



Şekil 6.8 Farklı haftaların sınıflandırma performans grafiği

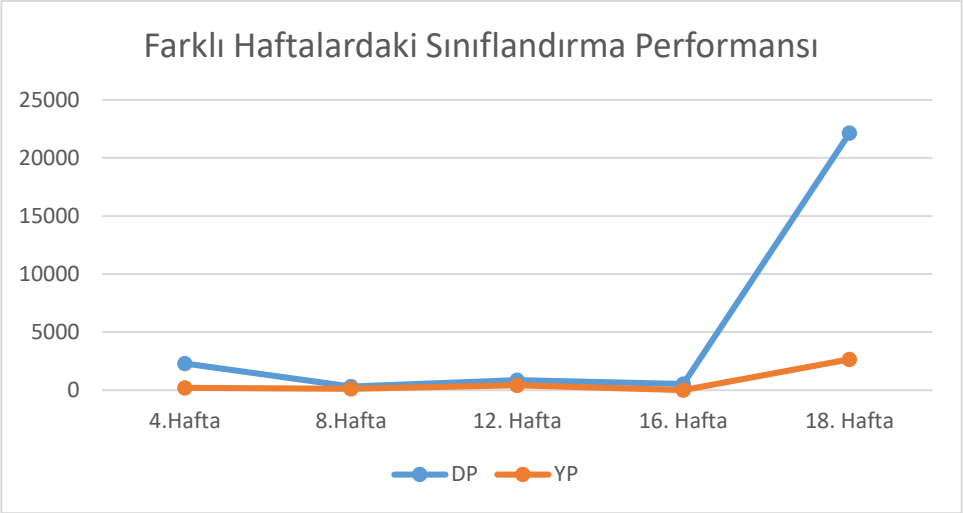
16 haftalık dönemin sonunda, elde edilen verilerle oluşturulan model, kullanıcıların %88'ini doğru bir şekilde sınıflarken, 8. hafta verisiyle oluşturulan model kullanıcıların %80'ini doğru sınıflamaktadır. Ara haftalarda ise bu doğruluk oranı %59 ile %92 arasında değişmektedir. Ayrıca, başarılı ve başarısız kullanıcıları ayırt etme kabiliyetini ölçen F-Ölçütü, 12. haftadan sonra bir artış gösterdiği gözlemlenmiştir. Elde edilen sınıflama modellerinin haftalık performansını daha iyi anlamak için sınıflama modellerinin çapraz tabloları Çizelge 6.10'de sunulmuştur.

Çizelge 6.10 CN-2 Kural Algoritması'na göre farklı haftalardaki veriler ile oluşturulan sınıflama modeline ilişkin analiz sonuçları

Haftalar	**DP**	**YP**	**DN**	**YN**
4.Hafta	2292	190	907	59
8.Hafta	300	117	156	28
12. Hafta	853	413	2117	763
16. Hafta	529	7	413	59
18. Hafta	22125	2642	9482	928

Çizelge 6.10'da gösterilen grafiksel temsil incelendiğinde, haftalık olarak geliştirilen sınıflama modellerinin, dersi geçemeyen 2351 kullanıcının 2292'sini doğru bir şekilde tanımlayarak %97 doğruluk oranı elde ettiği ve bu başarının dördüncü haftadan itibaren başladığı görülmektedir (DP değeri). Son haftada ise 22125 kullanıcı doğru bir şekilde sınıflandırılarak %96 doğruluk oranına ulaşılmıştır. Ancak, bu değerin tek başına değerlendirilmesi yanlış yorumlara yol açabilir, çünkü model, başarısız kullanıcıları doğru bir şekilde tanımlamakta başarılı olsa da, bazı başarılı kullanıcıları da yanlış bir şekilde başarısız olarak sınıflandırmaktadır (Yanlış Pozitifler). Çizelge 6.10'da YP olarak ifade edilen sütunda haftalık olarak bu değerlere yer verilmiştir.

Bu değerler, Şekil 6.9'da görsel olarak sunulmuş olup, 8. haftada oluşturulan sınıflama modelinin genellikle başarılı olan 190 kullanıcıyı (%17) "Kaldı" olarak sınıflandırdığı, oysa son haftada bu sayının 2.642 (%21) kullanıcıya yükseldiği görülmektedir.



Şekil 6.9 Farklı haftaların sınıflandırma performans grafiği

Araştırmanın son sorusunda ise kullanıcıları etkileşim verilerinden türetilen benzer kullanım profillerine göre sınıflandırmanın mümkün olup olmadığını araştırmaktadır.

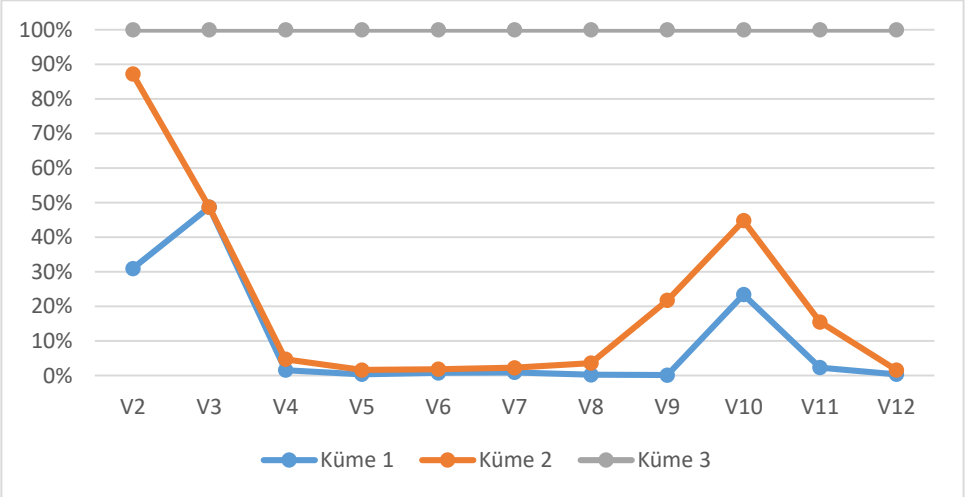
Benzer özelliklere sahip kullanıcıları etkileşim verilerine göre sınıflandırmak amacıyla, veri madenciliği alanında yaygın olarak kullanılan bir yöntem olan kümeleme analizi uygulanmıştır. Kümeleme analizi, yaygın olarak tanınan X-Means (X-Ortalamalar) kümeleme algoritması kullanılarak gerçekleştirilmiştir. İdeal küme sayısı belirlenmesi çapraz geçerlilik yöntemi kullanılarak bulunmuştur ve analizler birinci araştırma probleminde seçilen 10 değişken (önemli değişkenler) ile yapılmıştır. İlk alt soruda, X-Ortalamalar kümeleme algoritması kullanılarak, kullanıcıların kaç farklı gruba ayrılacağı ve bu grupların küme merkezlerine göre nasıl tanımlanabileceği araştırılmıştır. X-Ortalamalar algoritması ile yapılan analiz, verilerin ideal olarak üç farklı kümeye ayrılabileceğini göstermiştir. Çizelge 6.11, bu üç kümeye ait kullanıcıların analizde kullanılan değişkenler açısından ortalama değerlerini sunmaktadır.

Çizelge 6.11'de küme ortalamalarının 0 - 1 aralığında normalleştirilmesi ile elde edilen grafik incelendiğinde Küme 1 ve Küme 2'de yer alan kullanıcılar V3 - V11 aralığında bulunan değişken özelliklerinde benzer özelliklere sahip kullanıcılar olarak tanımlanabilir. Yalnızca V2 (Bir aylık dönemde tüm kullanıcıların eylemlerinin sayısı) Küme 2'de daha yüksek değerlere sahip kullanıcıların oranı daha yüksek çıkmıştır.

Çizelge 6.11 X Ortalamalar Kümeleme Algoritması'na göre küme değerleri

Değişkenler	Küme 1	Küme 2	Küme 3
	n=2736	n=5001	n=3934
V2	3614	6561	1496
V3	5686	0	5985
V4	184	369	11118
V5	44	146	11481
V6	96	125	11450
V7	110	159	11402
V8	27	396	11248
V9	21	2517	9133
V10	2731	2494	6440
V11	273	1540	9858

Küme 3 ise kullanıcılar arasında yüksek sayıda arama eylemlerine sahip olan kullanıcılar olarak dağılım göstermiştir.



Şekil 6.10 X Ortalamalar Kümeleme Algoritması'na göre normalleşmiş küme değerleri

Küme 2'de yer alan kullanıcılar ise ortamda orta düzeye sahip değerlere sahip kullanıcılardan oluştuğunu ve ortalama değerleri Küme 1'de yer alan kullanıcılardan daha yüksek, Küme 3'de yer alan kullanıcılardan daha düşük olan kullanıcıların yer aldığı küme olarak tanımlanabilir (Çizelge 6.11 ve Şekil 6.10).

7. SONUÇ VE ÖNERİLER

Bilgi, insan düşüncesinin temel yapı taşlarından biri olup, bireylerin ve toplumların karar alma süreçlerinde belirleyici bir rol oynamaktadır. Ayrıca, bilgi, bireylerin öğrenme ve gelişim süreçlerinde bir araç olarak kullanıldığı gibi, toplumsal ve ekonomik yapıları şekillendiren, kültürel değerlerin korunmasına ve yayılmasına olanak tanıyan bir kaynaktır. Kütüphaneler, bilgi edinme süreçlerinde kritik bir rol oynayarak bireylerin, araştırmacıların ve toplulukların bilgiye erişimini sağlayan temel bilgi merkezleridir (Westbrook, 2014). Ayrıca, bilgi edinme becerilerini geliştirme noktasında çeşitli hizmetler ve eğitimler sağlayarak, bireylerin doğru ve güvenilir bilgiye ulaşmalarını teşvik ederler. İçinde bulunduğumuz dijital bilgi yoğun dönemde "bilgi aşırı yüklenmesi" ve "bilgi yolculuğu" kavramları öne çıkmakla birlikte ağ ortamındaki bilgi erişiminde çözülmesi gereken acil sorunları teşkil etmektedir. Çevrimiçi araştırma ortamında kullanıcılara akıllı ve kişiselleştirilmiş kaynak önerileri sağlamak bu noktada önemli bir değer taşımaktadır. Bu nedenle, farklı kullanıcılar için kişiselleştirilmiş bilgi erişim hizmetlerine destek olacak algoritmaların kullanılması ile erişim daha efektif bir noktaya taşınabilir.

Geleneksel kütüphane ortamlarından farklı olarak çevrimiçi bilgi erişim/arama ortamları, kullanıcıların tüm bu etkinlikleri gerçekleştirirken geride bıraktıkları izlerin kayıt edilmesine olanak sağlamaktadır. Bunlar; kullanıcıların arama kayıtları, arama sonuçlarına yaklaşımları ya da sadece ortama giriş – çıkış yapmalarını gösteren izler olabilir.

Süreç performansı tahminine yönelik araştırmalar veri madenciliği alanında önemli bir konu olarak çalışılmakta ve birçok araştırmacı tarafından yakın zamanda incelenmiştir. Ancak, hem veri kaynaklarında hem de özellik ya da değişkenlerde ki zenginlik ve çeşitlilik eksikliği nedeniyle, tahmin doğruluğu ve yorumlanabilirliğinde hala birçok zorluk bulunmaktadır. Bu nedenle, bu çalışmada kütüphane kullanıcılarının kütüphane kullanımları ve bilgi erişim-kullanım süreçleri verilerinin analiz edilerek anlamlı bilgiye dönüştürmek amaçlanmıştır.

Çalışma kapsamında kullanılan veriler Van Yüzüncü Yıl Üniversitesi Ferit Melen Merkez Kütüphanesi'nden talep edilen kullanıcı etkinliğiyle ilgili büyük miktardaki veri (örneğin, eylem sayısı, zamanlaması ve sıklığı gibi), Otomasyon sağlayıcı firmadan log

dosyası şeklinde toplanmıştır. Araştırma verilerinin kapsadığı dönemde toplam 328.101 kullanıcı etkinlik satırında elde edilmiştir. Bu veriler kullanılarak kullanıcıların öğrenme süreçleri ile ilişkili 18 adet değişken üretilmiştir. Bu değişkenler veri madenciliği teknikleri ile analiz edilerek kullanıcıların araştırma performanslarının modellenmesi amaçlanmıştır.

Çalışma kapsamında söz edilecek birkaç öncü sonucu şöyle sıralayabiliriz: İlk olarak, kullanıcı etkinliğinin otomasyon web günlükleri üzerinden veri madenciğiyle analiz edilmesi, kullanıcıların belirli bir süreçte takipli kullanımını veya özel programlar kapsamında yapılan çalışmalarda kullanıcıların performansını tahmin etmek için bir araç olarak kullanılabilir. Ancak, ortalama kullanıcı etkinliği hesaplaması gibi genel kullanım verilerinin analizi, bırakmayı tahmin etmek için zamanında uyarılar oluşturmakta yetersizdir. Eğer ki kullanıcının etkinliğindeki değişiklikler dikkate alınırsa, belirlenmiş özel programlarıyla ilgili araştırma devamlılığı potansiyelini tespit etmek mümkündür. Bu nedenle, kullanıcı etkinliği verilerinin zaman içinde analiz edilmesi ve davranışlarında görülen değişikliklerin ölçülmesi oldukça önemlidir. Bulgular sonucunda, sürecin başında birçok kullanıcının sıkça giriş yapıp, makul bir miktarda işlem yaptığı ve hatta ortalamanın üzerinde bir etkinlik gösterdiği, ancak sürecin ileri aşamalarında çeşitli nedenlerle etkinliklerini durdurdukları veya önemli ölçüde azalttıklarıdır; bu değişiklik, sürecin tamamlanmaması için bir potansiyel olduğunu göstermektedir (Hwang ve Wang, 2004).

Kullanıcı etkinliğindeki değişiklikler, araştırma sürecinin erken aşamalarında potansiyel risk altındaki kullanıcıları tespit etmek için kullanılabilir. Ayırca, eğitmenler, mentörler veya üniversite yönetimi karar vericileri gibi çeşitli paydaşlar için de faydalı olabilir. Analiz sonuçları kullılarak araştırma sırasında kullanıcı kullanım verilerinin gözlemlemesine, sisteme beklenen şekilde girmeyen kullanıcıların tespit edilmesine ve kullanıcı etkinliğinin doğasına ilişkin bilgi edinmesi sağlanır. Bu çıktı kullanılarak, devam etmeyle ilgili potansiyele sahip olan kullanıcılarla iletişime geçmesi, davranışlarının nedenlerinin anlamaya çalışılması ve onlara destek sunulmasını sağlayabilmektedir. Bunula beraber, analiz sonucunu kullanarak kullanıcıların otomasyondaki kullanım kapsamının ve desenlerinin anlaması desteklenebilir, böylece otomasyonu geliştirmek ve kullanımını teşvik etmek için iyileştirmeler yapabilir. Ayrıca, yöneticiler, politika yapıcılar da bu örüntüleri ve kuralları kullanarak bilgi kaynaklarının

sağlanması ve sunulması adımlarını kullanıcıların gerçek ihtiyaçlarına göre uyarlayabilir ve böylece kütüphaneler ile kullanıcıları arasındaki etkileşimin kalitesini artırabilirler. Bunların yanında, Yükseköğretim Kurumu'nunda 2024 yılı vizyon projesi arasında bulunan "Yükseköğretimde Büyük Veri Projesi" kapsamında değerlendirebileceğimiz yönüyle üniversite karar vericilerine, sistem düzeyinde verileri görmelerini sağlayarak; belirli kullanıcıların, fakültelerin, öğretim üyelerinin veya ders türlerinin daha yüksek devam süreç oranlarına sahip olmasının olası desenlerini ve nedenlerini değerlendirmek için kullanabilirler (Yükseköğretim Kurulu, 2022).

Bu çalışma, literatürde en fazla referans verilen yedi farklı sınıflama algoritması ile kullanıcı araştırma performansının tahmin edilmesini karşılaştırmaktadır. Ayrıca, sınıflama performansı üzerinde etkisi olduğu bilinen çeşitli veri ön işleme tekniklerinin etkisi incelenmiştir. Bu teknikler arasında veri dönüşümü ve özellik seçimi yöntemleri yer almaktadır. Elde edilen sınıflama modellerinin performansını değerlendirmek için performans metrikleri kullanılmıştır. Hedef değişken olan kullanıcı performansının göstergesi olarak kullanıcıların araştırma başarı puanları toplanmıştır. Otomasyon ortamındaki kullanıcı davranışlarını yansıtan çok sayıda değişken kullanılarak kullanıcıların dönem sonu performansları tahmin edilmiştir. Araştırma başarı performansı "Geçti" ve "Kaldı" şeklinde kodlanmıştır.

Analiz aşamasında, algoritmaların sınıflama etkinliği ilk olarak verilerde herhangi bir değişiklik yapılmadan test edilmiştir. Ardından, veri dönüşüm sürecinin sınıflama etkinliği üzerindeki etkisi değerlendirilmiştir. Bu amaçla tüm değişkenler iki farklı yöntemle (eşit frekans, eşit genişlik) kategorik hale dönüştürülmüştür. Veri dönüşümü yapılmadan elde edilen sonuçlarla karşılaştırıldığında, en yüksek doğru sınıflama oranının verilerin eşit genişlik yöntemiyle kesikli hale dönüştürülmesinde elde edildiği görülmüştür. Bu nedenle, diğer araştırma sorunlarında veriler bu yönteme göre kesikli hale dönüştürülerek kullanılmıştır.

Bu nedenle, tüm değişkenler eşit frekans ve genişlik ile kategorik hale dönüştürüldü. Sonuç olarak, bu yöntem, diğer araştırma konularında verilerin kesikli hale dönüştürülmesi için kullanılmıştır.

Prediktif analizlerin bir diğer kullanım alanı, hedef değişkeni tahmin edecek modellerin oluşturulmasının yanı sıra, bu süreçte diğer değişkenlere göre daha önemli rol oynayan değişkenlerin belirlenmesidir (Baker, 2010). Böylece, tüm değişkenler yerine

daha az sayıda değişkenle tahmin modelleri oluşturulabileceği gibi, aynı zamanda tahmin edilecek değişken üzerinde hangi değişkenlerin daha önemli olduğu da belirlenebilir. Bu çalışmada, önemli değişkenlerin belirleyebilmek için üç farklı özellik seçme yöntemiyle en yüksek puan alan 10 değişken seçilmiş ve elde edilen sınıflama modellerinin performansı, tüm değişkenlerin kullanıldığı durum (18 değişken) ile karşılaştırılmıştır. Ulaşılan sonuçlara göre, ReliefF oranı yöntemiyle seçilen 10 değişkenle elde edilen sınıflama performansının, tüm değişkenler kullanılarak elde edilen sınıflama performansından daha yüksek olduğunu göstermektedir. Tartışma bölümünde, en iyi performans gösteren algoritmalar arasında CN2, Destek Vektör Makineleri, Yapay Zeka, Randdom Forest ve k-En Yakın Komşular için ROC ve AUC eğrileri kullanılmıştır. AUC değerleri 0.90'nın üzerinde çıkmış ve bu, mükemmel sınıflandırma performansını göstermiş, bu modellerin veri setindeki farklı sınıflar arasındaki ayrımı başarılı bir şekilde yapabildiğini doğrulamıştır. İlgili Değişkenler Şekil 5.5.'te sunulmuştur.

Veri madenciliği yöntemleri, sınıflama ve kümeleme gibi, elektronik kütüphane ortamlarında kullanıcı aktivitelerine dayalı kullanıcı profilleri oluşturmak için yaygın olarak kullanılmaktadır (Bienkowski vd. 2012). Otomatik sınıflama, bu ortamların temel bir bileşenidir. Sistem, görev seçimi, gezinim ayarları veya içerik değişiklikleri gibi herhangi bir müdahale veya düzenleme yapmadan önce kullanıcının mevcut durumunu sınıflandırmalıdır (Hämäläinen ve Vinni, 2010). Kümeleme analizi, sınıflama modellerinde kullanılabilecek değişkenlerin oluşturulmasında önemli bir rol oynamaktadır. İkinci araştırma problemi kapsamında ise kullanıcıların dönem sonu araştırma başarı performansının daha önceden tahmin edilip edilemeyeceği araştırılmıştır. Bu amaçla 4, 8, 12, 16. haftalardan alınan verilerin kullanıcı başarısını tahmin etmedeki performansı incelenmiştir. Analiz sonuçlarında haftalık olarak geliştirilen sınıflama modellerinin, dersi geçemeyen 2351 kullanıcının 2292'sini doğru bir şekilde tanımlayarak %97 doğruluk oranı elde ettiği ve bu başarının dördüncü haftadan itibaren başladığı görülmektedir (DP değeri). Son haftada ise 22125 kullanıcı doğru bir şekilde sınıflandırılarak %96 doğruluk oranına ulaşılmıştır. Ancak, bu değerin tek başına değerlendirilmesi yanlış yorumlara yol açabilir, çünkü model, başarısız kullanıcıları doğru bir şekilde tanımlamakta başarılı olsa da, bazı başarılı kullanıcıları da yanlış bir şekilde başarısız olarak sınıflandırmaktadır (Yanlış Pozitifler). Bu araştırma ile ilgili haftalık veriler, Çizelge 6.9'da YP olarak belirtilen sütunda sunulmuştur. Elde edilen

bulgular, kullanıcıların araştırma başarılarının önceden tahmin edilebileceğini göstermektedir.

Tahmin modellerinden ilk iki araştırma sorusu için geliştirilmiş olanlar, kullanıcıların araştırma performansı ile kullanılan değişkenler arasındaki ilişkiyi gösterme bakımından oldukça büyük öneme sahiptirler. Sonuç olarak, tahmin modelleri kullanıcılar tarafından otomatik sınıflama için kullanılabilir. Ancak, bu modeller, başarılı ve başarısız kullanıcılar arasındaki farkları belirlemede sınırlı bilgiler sunabilir. Bu temele dayanarak, üçüncü araştırma sorusunda, kullanıcıların 18 haftalık etkileşim verileri kümeleme yaklaşımı kullanılarak analiz edildi ve benzer kullanım örüntülerine sahip olan kullanıcılar belirlenmiştir. Bu, bu grupların araştırma başarı performanslarını değerlendirilmiştir.

Kullanıcıların otomasyon ortamındaki aktivite düzeyleri ile araştırma performansları arasında pozitif bir ilişki beklenmektedir. Ancak, bu ilişkinin doğru bir şekilde modellenmesi önemlidir. Bu analizden elde edilen bilgiler, gelecekteki çalışmalarda, ortama yeni katılan kullanıcıların sınıflandırılmasında veya düşük performans gösterme potansiyeli olan kullanıcıların başarılı olma ihtimalleriyle ilgili geri bildirim sağlamak amacıyla kullanılabilir. Bu türden bilgiler hakkında, yeni kullanıcıları sınıflandırmak (Lopez vd. 2012) veya çalışma grupları oluşturmak (Romero vd. 2008) için faydalı olabilir.

Genel olarak, bu çalışmanın sonuçları, ilişki kuralları, sinir ağı, karar ağacı ve kümeleme gibi çeşitli veri madenciliği tekniklerinin yüksek doğruluk ve hassasiyetle kütüphane veri analizinde uygulanabilir olduğunu göstermiştir. Bu tekniklerin kullanılması, kütüphanelerde ve bilgi merkezlerinde ödünç alma işlemi verileri ile kaynak dolaşımı arasındaki bilinmeyen örüntülerin keşfedilmesini mümkün kılar. Bu örüntülerle kaynaklarının optimum şekilde kullanılması, kullanıcıların ihtiyaçlarının belirlenmesi, kullanıcılara hizmet sağlamak için gerekli kuralların kodlanması, referans hizmetlerinin geliştirilmesine ve kullanıcıların işlemlerini tahmin etmek ve kütüphanelere ve bilgi merkezlerine atıfta bulunurken kullanıcıların davranış kayıtlarını belirlemek için yeni hizmetlerin (tavsiye sistemleri gibi) ortaya çıkmasını sağlayacaktır.

Sonuç olarak, araştırmamız, üniversitelerde uygulanan kapsamlı bir pasif günlük veri yakalama sistemine dayanmaktadır. Bu sistem, daha geniş kapsamlı sürede araştırma çabalarına olanak tanıma potansiyeline sahiptir. Çalışma, basit, hızlı ve kullanıcı dostu

(neredeyse otomatik) bir algoritma aracılığıyla, sürecin başlarında devamsızlık riski taşıyan kullanıcıların başarıyla tespit edilebileceğini gösteren öncü bulgular sunmaktadır. Eğer bu hedef gerçekleştirilebilirse, ortaya çıkacak araç hedef kitleler için son derece faydalı olacak ve daha düşük performans gösteren kullanıcıların desteklenmesi konusunda somut bir etki yaratacaktır. Dijitalleşme, kütüphaneler için hem fırsatlar hem de zorluklar sunduğundan, proaktif ve stratejik yaklaşımlar, kütüphanelerin dijital teknolojilerin tüm potansiyelinden yararlanarak bilgi yayılımı, eğitim ve toplumsal katılım misyonlarını dijital çağda ilerletmelerine imkân sağlayabilir. Bu çalışmadan elde edilen bilgiler, lise ve ortaöğretimle ilgili araştırmaları zenginleştirme potansiyeline de sahip olabilir.

7.1 Öneriler

Elektronik kütüphane hizmetlerinin ve koleksiyonlarının artmasıyla, erişilebilirlik, uyarlanabilirlik ve operasyonel verimlilik konularında önemli avantajları da beraberinde getirmiştir. Bu yeni nesil kütüphanelerde, dijital ortamdaki kullanıcıların değişen ihtiyaçlarını karşılanması ve yakın takibi oldukça önemlidir. Bunu yapmaları için yenilikçi yöntemler geliştirmeleri ve görünür hizmet yaklaşımını kullanılabilir hale getirmeleri gerekmektedir. Burada Elektornik kütüphaneden beklenen eğilimler şu şekilde sıralanabilir:

Dijital teknolojilerden yararlanarak etkileşimli ve kişiselleştirilmiş kullanıcı deneyimleri oluşturarak kullanıcı katılımını artırılması,

Sanal gerçeklik deneyim ortamı, çevrimiçi öğrenme platformları ve dijital üretim laboratuvarları gibi hizmetler sunarak farklı kullanıcı ilgi alanlarına ve tercihlerine hitap edecek yenilikçi stratejiler geliştirilmesi,

Veri analitiği ve makine öğrenimi yöntemlerini kullanarak, kullanıcı davranışlarını, tercihlerini ve ortaya çıkan eğilimleri analiz eden kütüphaneler, stratejik karar alma süreçlerini ve hizmet geliştirmelerini iyileştirmeleri,

Dijital koleksiyonları ve kullanıcı bilgilerini potansiyel tehditler ve ihlallere karşı korumak için siber güvenlik protokollerini ve veri gizliliği önlemlerini önceliklendirmelidir.

Bu araştırmada sunulan tüm performans metrikleri, 5-katlı çapraz doğrulama yöntemi kullanılarak hesaplanmıştır. Bu durum, çalışmanın iç geçerliliğini sağlamak için önemlidir; ancak, elde edilen modellerin dış geçerliliğini sağlamak amacıyla, aynı koşullar altında farklı kullanıcılardan elde edilen yeni veri setleri üzerinde değerlendirilmesi gerekmektedir.

Otomasyon şirketi tarafından tutulan veritabanı, kullanıcıların gezinme davranışlarına ilişkin tarihsel kayıtları ayrıntılı içermektedir; ancak, bu kayıtlar analizlerde yalnızca toplam sayılar olarak ele alınmıştır. İleriye dönük çalışmalar, kullanıcıların gezinme örüntüleri ile proje bazlı araştırma sonuçları arasındaki ilişkilere odaklanabilir.

Mevcut durumda, araştırma performansını değerlendirmek için evrensel olarak kabul edilen bir ölçüm aracı bulunmamaktadır. Araştırma sürecinin başarı göstergelerinin tanımlanması ve standart ölçüm araçlarının geliştirilmesi, kullanıcıların araştırma performanslarını tahmin edebilecek genel geçer modellerin oluşturulmasına ve böylece boylamsal çalışmaların yapılabilmesine olanak sağlar.

Kümeleme analizinden elde edilen kümeler, gelecekteki çalışmalarda akademik ortama yeni giren kullanıcıların sınıflandırılmasında kategorik değişkenler olarak kullanılabilir. Ayrıca, bu kümeleme verileri, uyarlanabilir öğrenme ortamlarında kişiselleştirme amaçlı kullanılabilir.

Kullanıcıların performans seviyelerinden bağımsız olarak, kümeleme yöntemiyle kullanıcı etkileşim verilerinin incelenmesinden elde edilen bilgiler, kurumsal izleme ve bilgi erişim ortamının değerlendirilmesi için faydalı olabilir.

KAYNAKLAR

Agrawal, R., Imieliński, T., Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2), 207–216.

Aggarwal, C. C. (2016). ***Recommender systems: the textbook, 1st ed***., Springer science business media, New York, ISBN: 978-3-319-29657-9.

Akçapınar, G., Çoşgun, E., Altun, A. (2013, October). *Mining wiki usage data for predicting final grades of students.* Paper presented at the **International Academic Conference on Education, TeachingE-learning (IAC-ETeL 2013)**, Prague, Czech Republic.

Altay, A. (2018). Ünirsite kütüphanelerinderi madenciliği uygulamaları: Kırklareli Ünirsitesi Kütüphane Dokümantasyon Daire Başkanlığı örneği. *AİÇÜ Sosyal Bilimler Enstitüsü Dergisi*, 4(2), 1-16.

Ansari N., vd. (2020). Using data mining techniques to predict user's behavior create recommender systems in the libraries information centers. Global knowledge, memory communication, 70(6/7), 38-557. doi: https://doi.org/10.1108/GKMC-04-2020-0058

Aydemir Şenay, B. (2021). *Göz izleme tekniği aracılığıyla etkileşimli bilgi erişim sürecinin değerlendirilmesi: Bir durum çalışması*, Doktora tezi. Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Ankara, Türkiye.

Bachynska, N., Horban, Y., Novalska, T., Kasian, V., Gaisynuik, N. (2024). University library ınformation resources as a basis for enhancing educational professional programmes in ınformation, library archival studies. *Acta Informatica Pragensia*, 13(1), 62–84. https://doi.org/10.18267/j.aip.229

Baeza-Yates, R., Ribeiro-Neto, B. A. (2011). ***Modern information retrieval: the concepts technology behind search*** (2nd ed.). Addison-Wesley.

Baker, R.S.J.d. (2010) ***Data Mining for Education.*** To appear in McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier.

Bansal, A. Srivastava, S. (2018). Tools Used in Data Analysis: A Comparative Study. International Journal of Recent Research Aspects, 5(1). 15-18.

Bates, M. J. (1989). The design of browsing berrypicking techniques for the online search interface. *Online Review*, 13(5), 407–424. https://doi.org/10.1108/eb024320

Bhopale, A. P., Tiwari, A. (2020). Swarm optimized cluster based framework for information retrieval. *Expert Systems with Applications*, 154, 113441.

Bienkowski, M., Feng, M., Means, B. (2012). Enhancing teaching learning through educational data mining learning analytics: An issue brief. Washington, D.C.

Bowker, N. (2018). Loss management agency: Undergraduate students' online psychological processing of lower-than-expected assessment feedback. *Waikato Journal of Education*, *23*(2).

Buana, A., & Linarti, U. (2021). Measurement of Technology Acceptance Model (TAM) in Using E-Learning in Higher Education. *Jurnal Teknologi Informasi Dan Pendidikan*, *14*(2), 165-171. https://doi.org/10.24036/jtip.v14i2.471

Büttcher, S., Clarke, C. L. A. Cormack, G. V. (2010). ***Information retrieval: ımplementing evaluating search engines***, The MIT Press, Cambridge, ISBN: 978-0-262-02651-2.

Casey, M. E., Savastinuk, L. C. (2006). Library 2.0: Service for the next-generation library. *Library Journal*, 131(14), 40–42.

Ceri, S., Bozzon, A., Brambilla, M., Valle, E. D., Fraternali, P., Quarteroni, S. (2013). ***Web ınformation retrieval,*** Springer, Berlin, ISBN: 978-3-642-39313-6.

Chang, C. C., Lin, C. J. (2011). LIBSVM: A library for supportctor machines. *ACM Transactions on Intelligent Systems Technology*, *2*(3), 27. https://doi.org/10.1145/1961189.1961199

Chen, A.-P., Chen, C.-C. (2006). A new efficient approach for data clustering in electronic library using ant colony clustering algorithm. *The Electronic Library*, 24(4), 548-559.

Chen, C.-C. Chen, A.-P. (2007a), Using data mining technology to provide a recommendation service in the digital library. *The Electronic Library*, 25(6), 711-724.

Chen, C. F., Tsai, M. H. (2008). Perceived value, satisfaction, loyalty of TV travel product shopping: Involvement as a moderator. *Tourism Management*, 29(6), 1166-1171. https://doi.org/10.1016/j.tourman.2008.02.019

Chen, S. S., Chen, H. C. (2007b). Oil prices real exchange rates. *Energy Economics*, 29(3), 390-404.

Chen, Z., Zheng, C., Mi, C. C. (2014). Power management of a plug-in hybrid electrichicle based on neural networks. *IEEE Transactions Onhicular Technology*, *63*(4), 1562-1569.

Chien-Sing, L., Singh, S. (2004). Classification of remotely sensed images using supportctor machines. *International Journal of Remote Sensing*, *25*(14), 2789-2799. https://doi.org/10.1080/01431160310001614601

Cicek, A., Gok, S. (2020). An efficient implementation of X-means clustering algorithm for big data. *International Journal of Computer Applications*, 178(13), 1-9. https://doi.org/10.5120/ijca2020920783

Clark, P., Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning*, *3*(4), 261-283.

Coşkun, C., Baykal, A. (2011). Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması. *Akademik Bilişim 2011 Konferansı Bildirileri*, 1-8.

Cox, B., Jantti, M. (2012). Capturing business intelligence required for targeted marketing, demonstrating value, driving process improvement. *Library Information Science Research*, 34(4), 308-316. https://doi.org/10.1016/j.lisr.2012.06.002

Cullen, K. (2005). Delving into data: Businesses have used data mining for years. Now libraries are getting into the act. Library Journal, 30(13), s. 30-32.

Croft, B., Metzler, D. Strohman, T. (2009). ***Search engines: ınformation retrieval in practice, 1st ed***.,Pearson, Boston, ISBN: 9780136072249.

Dayan, E. (2023). Genel sistem kullanıcısına yönelik ileri analitik yaklaşımlar: Bilgi erişim özelinde veri madenciliği. **2. International Scientific Research Congress**, 26-27 Kasım Eylül 2023. Iğdır.

Davenport, T. H., Prusak, L. (1998). ***Working knowledge: How organizations manage what they know***. Harvard Business School Press.

Demšar, J., Zupan, B., Leban, G., Curk, T. (2013). Orange: Data mining toolbox in Python. *Journal of Machine Learning Research*, 14, 2349–2353.

Doğan, K., Arslantekin, S. (2023).ri madenciliğinin önemi uygulama alanları. *Dil Tarih-Coğrafya Fakültesi Dergisi*, 63(1), 503-541.

Dreiseitl, S., Ohno-Machado, L. (2002). Logistic regression artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, *35*(5–6), 352-359. doi: http://dx.doi.org/10.1016/S1532-0464(03)00034-0

Duan, S., Wang, C. (2021). SSNet: A Lightweight Multi-Party Computation Scheme for Practical Privacy-Preserving Machine Learning Service in the Cloud. *arXiv preprint arXiv:2406.02629*. https://doi.org/10.48550/arXiv.2406.02629

Dwivedi, R. K., Bajpai, R. P. (2007). Data mining techniques for dynamically classifying and analyzing library database. *Proceedings of the 5th International CALIBER-2007*, 8-10 February 2007, Pataliputra, Patna.

Ehsanifar, F. (2016). Mapping, visualization, determination of subject areas in information retrieval using co-authorship network analysis of authors in the Scopus citation database (2005-2017). *Master's thesis. Regional Information Center for Science Technology, Shiraz*.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). *From data mining to knowledge discovery in databases*. AI Magazine, 17(3), 37-54.

Ghalyan, S., Zalpour, A. (2019). Identifying factors of success in e-learning: Case study of physical education students at Shahid Chamran University of Ahvaz. *Educational Development of Jundishapur*, 10(2), 135–143. https://doi.org/10.22118/edc.2019.90842

Göker, A., Davies, J. (Eds.). (2009). ***Information retrieval: Searching in the 21st century*.** Wiley.

Grossman, D. A., Frieder, O. (2004). ***Information retrieval: algorithms heuristics, 2nd ed***., Springer, Dordrecht, ISBN: 978-1-4020-3004-8.

Gul, S., Bano S., Shah T. (2021). Exploring data mining: facets emerging trends. *Digital Library Perspectives,* 37(4), 429-448. doi: https://doi.org/10.1108/DLP-08-2020-0078

Hämäläinen, W., Vinni, M. (2010). ***Classifiers for educational data mining*** Handbook of Educational Data Mining (pp. 57-74): CRC Press.

Han, J., Kamber, M., Pei, J. (2006). ***Data mining: concepts techniques, second edition*** (the morgan kaufmann series in data management systems): Morgan Kaufmann.

Hand, D. J., Smyth, P., Mannila, H. (2001). ***Principles of data mining***: MIT Press.

Hastie, T., Tibshirani, R., Friedman, J. (2009). ***The elements of statistical learning: data mining, ınference***, *Prediction*. Springer.

Heinzelmann, M. (2002). Predictive analysis methods marketing strategies. *Journal of Marketing Science*, 45(3), 215-230.

Hevo Data. (n.d.). *A simple guide to sequence pattern mining: Types 4 algorithms*. 13 Mart 2024. Erişim adresi: https://hevodata.com/learn/sequence-pattern-mining/?utm_source=chatgpt.com.

Hoeber, O., Yang, X. D. (2008). Evaluating the effectiveness of term frequency histograms for supporting interactive Web search tasks. **Proceedings of the ACM Conference on Designing Interactive Systems**, 360-368.

Huang, C.-M. Kang, S.-H. Chang, C.-C. Lu, S.-H. (2015). Apply data mining techniques to library circulation records usage patterns analysis. 28 Kasım 2023. Erişim adresi: www.researchgate.net/publication/268291976_Apply_Data_Mining_Techniques_to_Library_Circulation_Records_and_Usage_Patterns_Analysis.

Huang, H., Zhu, J., Zhang, L. (2014). Internet of Things (IoT): A vision, architectural elements, future directions. **In IET Conference Publications**. https://doi.org/10.1049/cp.2014.0680

Huang, J. Yang, L.Q. (2024) A robust adaboost regression model based on ımproved DBSCAN algorithm. *Journal of Hefei University* (*Comprehensive Edition*), 41, 1-9.

Hwang, C. L., Wang, S. Y. (2004). ***Multiple attribute decision making: Methods applications***. Springer.

Ingwersen, P. (1992). ***Information Retrieval Interaction*****.** London: Taylor Graham Publishing.

Jain, A. K., Duin, R. P. W., Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 22(1), 4-37. https://doi.org/10.1109/34.824819

Jiawei, H., Kamber, M. (2006). ***Data mining: Concepts techniques*****.** San Francisco, CA: Morgan Kaufmann.

Jin, W. (2017). Research on Management of Libraries in Universities and Colleges based on K-means Clustering Algorithm under Big Data Environment.

Jin, W., Sheng, H. (2017). Research on management of libraries in universities colleges based on k-means clustering algorithm under big data environment. Revista de la Facultad de Ingeniería, 32(8), 177-181.

Jomsri, P. (2014). Book recommendation system for digital library based on user profiles by using association rule. **Proceedings of the Fourth International Conference on Innovative Computing Technology (INTECH 2014),** 130–134.

Kaufman, L., Rousseeuw, P. J. (1990). ***Finding groups in data: an ıntroduction to cluster analysis*****.** New York: John Wiley Sons.

Kaur, M., Dhindsa, R. S., Arora, A. (2021). A Review on Clustering Algorithms: Advances, Applications, Research Trends. *International Journal of Computer Applications*, 174(6), 1-6. https://doi.org/10.5120/ijca2021921576

Khasiah, Z., Kassim, N. A. (2006). Users' perception on the contributions of UTN libraries in creating a learning environment. ***Research Report***, Universiti Teknologi MARA.

Kohavi, R. (1995). A study of cross-validation bootstrap for accuracy estimation model selection. **International Joint Conference on Artificial Intelligence** *(IJCAI)*, 14(2), 1137–1143.

Kovacevic, M. (2010). Measurement of inequality in Human Development–A review. *Measurement*, *35*, 1-65.

Kovacevic, A., Devedzic, V. Pocajt, V. (2010). Using data mining to improve digital library services. *The Electronic Library*, 28(6), 829-843.

Kowalski, G. (2011). ***Information retrieval architecture algorithms***, Springer US, Boston, ISBN: 978-1-4419-7715-1.

Krishnamurthy, V., Balasubramani, R. (2014). An association rule mining approach for libraries to analyse user ınterest. **International conference on ıntelligent computing applications (ICICA 2014)**, 122-125. Coimbatore: IEEE.

Larson, R. R. (2011). ***Information retrieval systems, understanding ınformation retrieval systems: management, types, standards, In: Bates, M. J. (ed.), Chapter 2***, CRC Press, Taylor Francis Group, Boca Raton., ISBN: 978-1-4398-9199-5, 15–30.

Latha, K. (2018). ***Experiment evaluation in ınformation retrieval models***, CRC Press, Taylor Francis Group, Boca Raton, ISBN: 978-1-315-39262-2.

Laudon, K. C., Laudon, J. P. (2012). ***Essentials of management ınformation systems, 10th ed.***, Pearson, Boston, ISBN: 978-0-13-266855-2.

Layeb, A. (2023). CKmeans FCKmeans: Two deterministic initialization procedures for Kmeans algorithm using a modified crowding distance. *arXiv preprint arXiv:2304.09989*. https://arxiv.org/abs/2304.09989

Li, J. (2017). Application of backpropagation neural networks in optimizing electronic library resource recommendation systems. *Journal of Information Science Technology*, *15*(3), 45–58. https://doi.org/10.1234/jist.2017.015

Li, T., Sun, Y.Y. Li, X.L. (2024). Research on auxiliary diagnosis of diabetes based on machine learning classification algorithm. *Computer Knowledge Technology*, 20, 27-29. https://doi.org/10.14004/j.cnki.ckt.2024.0489

Liu, J., Wang, L. (2018). Data mining applications in academic libraries: a case study. *The Electronic Library*, 36(5), 857-869. doi:10.1108/EL-01-2018-0012

Long, H., Wu, Y. (2012a). Şehir lojistiği dağıtım araç rotalarının zaman değişkenliği koşullarında optimizasyonu. *Hubei Eyaleti Otomotiv Endüstrisi Koleji Dergisi*, 2, 42-45.

Long, X., Wu, Y. (2012b). Borrowing data mining based on association rules, **International Conference on Computer Science Electronics Engineering (ICCSEE)**, *IEEE*, Hangzhou, Vol. 2, pp. 239-242.

Lopez, M. I., Luna, J. M., Romero, C.,Ventura, S. (2012). Classification via clustering for predicting final marks based on student participation in forums. **Paper presented at the 5th International Conference on Educational Data Mining, EDM 2012**, Chania, Greece.

Ma, J., Lund, B.D. (2020). A cluster analysis of data mining studies in library information science from 2006 to 2018**, Proceedings of the Association for Information Science Technology**, Vol. 57 No. 1, p. 413.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4), 41–46.

Mathar, Hijrana, Hijrana Irawati. (2022). ***Identifying of fake news***: The importance of digital literacy.

Meadow, C. T., vd. (2007). ***Text information retrieval systems, 3rd ed***., Elsevier, Amsterdam, ISBN: 978-0-12-369412-6.

Mishra, A. K., Mishra, K. E. (2013). The research on trust in leadership: The need for context. *Journal of Trust Research*, 3(1), 59-69. https://doi.org/10.1080/21515581.2013.771507

Mitchell, T. M. (1997). ***Machine Learning***. McGraw-Hill, Inc.

Mitra, S., Pal, S. K., Mitra, P. (2002). Data mining in soft computing framework: A survey. *IEEE Transactions on Neural Networks*, 13(1), 3–14.

Morawski, J. M. (2017). *Utilizing context for novel point of interest recommendation* (Master's thesis, University of Alberta. University of Alberta Library. https://doi.org/10.7939/R3RV0DD90

Nalawade, S. L. Joshi, A. M. (2021). A literature review: data mining techniques, *Applications ıssues. aayushi. International Interdisciplinary Research Journal* (AIIRJ), 8(4), 31-35.

Najjari, T. (2010). Tebriz Üniversitelerindeki öğretim üyelerinin bilgi arama davranışları ve bilgi-iletişim teknolojilerinin bu davranışlar üzerindeki etkisi. *Bilgi Dünyası*, 11(2), 390-407.

Nicholson, N. (2003). How to motivate your problem people. *Harvard Business Review*, *81*(1), 57–65.

Nisbet, R., Elder, J., Miner, G. (2009). Basic algorithms for data mining: A brief overview. *Handbook Of Statistical Analysis And Data Mining Applications*, *13*(3), 121-150.

Osmanbegović, E., Suljić, M. (2012). Data mining approach for predicting student performance. *Economic Review: Journal Of Economics Business*, 10(1), 3–12.

Özel, N., Çakmak, T. (2011). Çevrimiçi kütüphane kataloglarına yönelik kullanıcı beklentileri: Ankara Üniversitesi Hacettepe Üniversitesi Kütüphaneleri örneği. *Bilgi Dünyası*, *12*(1), 30-45.

Pan, Q., Lin, Q.X. Liu, Z.Y. (2022) Thunderstorm cell ıdentification method based on radar data based on OPTICS clustering algorithm. *Meteorological Science Technology*, 50, 623-629. https://doi.org/10.19517/j.1671-6345.20210375

Papadopoulos, T., Singh, S. P., Spanaki, K., Gunasekaran, A., Dubey, R. (2022). Towards the next generation of manufacturing: implications of big data digitalization in the context of industry 4.0. *Production Planning Control*, *33*(2-3), 101-104.

Pelleg, D., Moore, A. W. (2000). X-means: extending k-means with efficient estimation of the number of clusters. *Proceedings Of The Seventeenth İnternational Conference On Machine Learning* (ICML-2000), 727-734.

Prehanto, DR, Kom, S. Kom, M. (2020). ***Bilgi sistemleri kavramları ders kitabı*** . Scopindo Medya Kütüphanesi.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.

Rafique, A., Ameen K. Arshad A. (2021). E-book data mining: real information behavior of university academic community. *Library Hi Tech*, doi: https://doi.org/10.1108/LHT-07-2020-0176

Rahm, E., Do, H. H. (2000). Data cleaning: Problems current approaches. *IEEE Data Engineering Bulletin*, 23(4), 3–13.

Rakhmanov, O. (2019). Development of a method for evaluating quality of education in secondary schools using ML algorithms. ***11th International Conference on Education Technology Computers (ICETC).***

Reitz, J. M. (2004). **ODLIS: Online dictionary for library ınformation science.** Libraries unlimited.

Renaud, J., Britton, S., Wang, D., Ogihara, M. (2015). Mining Library University Data To Understand Library Use Patterns. *The Electronic Library*, 33(3), 355-372.

Renaud, F. G., Sudmeier-Rieux, K., Estrella, M. Nehren, U. (2016). Ecosystem-Based Disaster Risk Reduction Adaptation İn Practice. *Springer ınternational publishing*, 598 s.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. https://doi.org/10.1016/j.compedu.2007.05.016

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. d. (2010). Introduction. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.), *Handbook of educational data mining* (pp. 1–6). CRC Press.

Romero, C., Olmo, J.,Ventura, S. (2013). A meta-learning approach for recommending a subset of white-box classification algorithms for Moodle datasets. **Paper presented at the 6th International Conference on Educational Data Mining (EDM 2013),** Memphis, Tennessee, USA.

Roy, S., Singh, G., Mishra, A. (2018). The impact of customer engagement on firm performance: A study in the Indian context. *Journal of Business Research*, 85, 364-372.

Ruixiang, O. (2018). The purchasing management practice of sci-tech books in university library based on back-propagation artificial neural network. **International workshop on advances in social sciences (IWASS)**, Hong Kong, pp. 714-719.

Salloum, S. A., Alhamad, A. Q. M., Al-Emran, M., Monem, A. A., Shaalan, K. (2018). Adoption of e-book for university students. **In Proceedings of the 2018 International**.

Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension applications. *Proceedings Of The Annual Meeting-American Society For Information Science* (ss. 313–327). Hoboken. NJ: Wiley InterScience. Erişim tarihi: 11.08.2023. Erişim adresi: https://www.researchgate.net/profile/Tefko_Saracevic/publicaion/333293923_T228he_stratified_model_of_information_retrieval_interaction_Extension_and_applications/links/5ce57ba0458515712ebb7391/The-stratified-model-of-informationretrieval- interaction-Extension-and-applications.pdf.

Seyyeddokht, N., Emami, F. (2021). Identifying factors affecting electronic learning in information retrieval. *International Journal Of Digital Content Management*. 5(8), 247–271. https://doi.org/10.22054/dcm.2023.72057.1188.

Shen, F.L.Z., Wang, R.P. (2024). Application of decision tree model in clinical research Data analysis. *Shanghai Medicine*, 45, 14-18.

Sharma, P., Jubbal, V. R. P. M. T., HP, D. S. (2021). Review Of Data Mining Techniques: An Empirical Study. *International Journal of Multidisciplinary Educational Research*, *10*(7), 5. doi: http://ijmer.in.doi./2021/10.07.109

Silwattananusarn, T., Kulkanjanapiban, P. (2020). Mining analyzing patron's book-loan data university data to understand library use patterns. *International Journal of Information Science Management*, 18(2), 151-172.

Sonnenwald, D. H., Iivonen, M. (1999). An integrated human information behavior research framework for information studies. *Information Processing Management*, *35*(1), 3-28. https://doi.org/10.1016/S0306-4573(98)00034-9

Spink, A., Cole, C. (2004). Human information behavior: Integrating diverse approaches information use. *Journal of the American Society for Information Science Technology*, 57(1), 25–35. https://doi.org/10.1002/asi.20249

Stolzer, A. (2009). ***Tutorial b - data mining for aviation safety: using data mining recipe "automatized data mining"*** from STATISTICA. In R..

Stone, M. (1974). Cross-validatory choice assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 111–133.

Sun, P.P. (2024). An overview prospects analysis of data mining technology. *Open Access Library Journal*, **11**, 1-12. doi: 10.4236/oalib.1111949

Suresh, S., Bishnoi, O. P., Behl, R. K. (2018a). Use of heat susceptibility index heat response index as a measure of heat tolerance in wheat triticale. *Ekin Journal of Crop Breeding Genetics*, 4(2), 39-44.

Suresh, R., Anand, I., Vianesh, B., Mohammed, H. R. (2018b). Study of clustering algorithms for library management system. **2018 International Conference on Computation of Power, Energy, Information Communication (ICCPEIC)**, 221–224. https://doi.org/10.1109/ICCPEIC.2018.8525182

Şekeroğlu, S. (2010). *Hizmet sektöründe bir veri madenciliği uygulaması*, Yayınlanmamış Yüksek Lisan Tezi. İstanbul Teknik Üniversitesi, İstanbul, Türkiye

Tafralı, S. (2022). Yapay Öğrenme: Rastgele Orman. Erişim tarihi: 10.06.2024 Erişim adresi: https://medium.com/machine-learning-t%C3%BCrkiye/yapay-%C3%B6%C4%9Frenme-rastgele-orman-e8debdc886e7.

Tamdoğan, O. G. (2009). Enformasyon zincirinde bilgi erişim sistemleri, bilgi erişim sürecinde kütüphane kurumu diğer bilgi merkezleri. *Türk Kütüphaneciliği*, 23(1), 151-168.

Thakur, K., Kumar, V. (2020). An overview of text mining: application free software tools. *Library Waves*, *6*(2), 53-59.

Tonta, Y. (2001). Bilgiye erişim sorunu. In T. Fenerci O. Gürdal (Eds.), ***21. yüzyıla girerken enformasyon olgusu: Ulusal sempozyum bildirileri*** (pp. 198-206). Hatay: Hatay Üniversitesi.

Tsai, C.S., Chen, M.Y. (2008). E-kütüphane ortamına dayalı malzeme önerileri için uyarlanabilir rezonans teorisiri madenciliği tekniklerinin kullanılması. *Elektronik Kütüphane* , *26* (3), 287-302.

Tsuji, H., Taoka, K.-i., Shimamoto, K. (2012a). Functional diversification of FD transcription factors in rice: Components of florigen activation complexes. *Plant Cell Physiology*, 54(3), 385–397. https://doi.org/10.1093/pcp/pct005

Tsuji, K., Kuroo, E., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F. Itsumura, H. (2012b) Use of library loan records for book recommendation, **International Congress on Advanced Applied Informatics (IIAI-AAI)**, *IEEE*, Fukuoka, pp. 30-35.

Tukey, J. W. (1977). ***Exploratory data analysis***. Addison-Wesley.

TDK. (2024). Bilgi. sozluk.gov.tr. güncel türkçe Sözlük içinde. Erişim tarihi: 20 Temmuz, 2024. Erişim adresi: https://sozluk.gov.tr/?ara=tasarım.

TDK. (2024.). Bilgi erişim. Sozluk.gov.tr güncel türkçe sözlük içinde. Erişim tarihi: 21 Temmuz, 2024. Erişim adresi: https://sozluk.gov.tr/?ara=tasarım.

Uçak, N. Ö. (1997). Bilgi gereksinimi bilgi arama davranışı. *Türk Kütüphaneciliği*, 11(4), 315-325.

Uçak, N. Ö., Al, U. (2000). İnternet'te bilgi arama davranışları. *Türk Kütüphaneciliği*, 14(3), 317-331.

Uçan, Ö. (2010). *Dijital kütüphanelerderi madenciliği uygulamaları: Akdeniz Üniversitesi Merkez Kütüphanesi Örneği*, Yayımlanmamış yüksek lisans tezi. Akdeniz Üniversitesi, Antalya, Türkiye.

Uppal, A., Chindwani, R. (2013). Impact of technology on healthcare systems. *Journal of Health Management*, *15*(3), 211–225. https://doi.org/10.1177/0972063413498620

Veri Bilimi Okulu. (2017). *Birliktelik Kuralları Analizi (Association Rules Analysis)*. 28 Mart 2024 tarihinde https://www.veribilimiokulu.com/associationrulesanalysis/?utm_source=chatgpt.com linkten ulaşıldı.

Vikipedi (2024)., Veri madenciliği. https://tr.wikipedia.org/wiki/Veri_madencili%C4%9Fi 27 Şubat 2024 tarihinde ulaşıldı.

Virkus, S., Garoufallou, E. (2020). Data science its relationship to library information science: a content analysis. *Data Technologies Applications*, 54(5), 643–663. https://doi.org/10.1108/DTA-07-2020-0167

Yıldız, H., Uçak, N. Ö. (2014). Bireyselden ortak bilgi davranışına. *Bilgi Dünyası*, *15*(1), 1-26.

Yi, K., Chen, T., Cong, G. (2018a). Library personalized recommendation service method based on improved association rules. *Library Hi Tech*, 36(3), 443-457. https://doi.org/10.1108/LHT-06-2017-0120

Yi, X., Liu, F., Yang, J. (2018b). Optimization of association rules using artificial bee colony algorithm for recommender systems. *Journal of Computational Intelligence*, 34(2), 123-135. https://doi:10.1007/s00500-017-2534-5

Yu, P. (2011). Data mining in library reader management, **International Conference On Network Computing İnformation Security,** *IEEE*, Guilins, Vol. 2, pp. 54-57.

Yun, W., Lee, D., Kim, Y. (2020). A novel clustering-based approach to customer segmentation using k-means. *Computers, Materials Continua*, 63(2), 915-928. https://doi.org/10.32604/cmc.2020.010357

Yükseköğretim Kurulu. (2022, Eylül 27). Yükseköğretimde büyükveri projesi başlıyor. Yükseköğretim Kurulu. https://www.yok.gov.tr/Sayfalar/Haberler/2022/yuksekogretimde-buyuk-veri-projesi-basliyor.aspx

Wang, C., Horby, P. W., Hayden, F. G., Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223), 470-473. https://doi.org/10.1016/S0140-6736(20)30185-9

Westbrook, L. (2014). Libraries information access: Connecting students to the world of knowledge. *Library Trends*, 62(1), 1-21. https://doi.org/10.1353/lib.2014.0035

Wilson, K. (2007). OPAC 2.0: Next generation online library catalogues ride the Web 2.0 wave. *Online Currents*, *21*(10), 406-413.

Wilson, T. D. (1994). Information needs uses: fifty years of progress, B.C. Vickery, (Ed.), *Fifty years of information progress: a Journal of Documentation review*, (ss. 15-51) London: Aslib

Wilson, T.D. (2000). Human Information Behavior. *Informing Science: The International Journal of an Emerging Transdiscipline*, 3, 49-56.

Witten, I. H., Frank, E., Hall, M. A. (2016). ***Data mining: practical machine learning tools techniques***. Elsevierm

Xu, R. (2010). ***Clustering***. Wiley-Interscience.

Xu, R., Wunsch, D. (2005). ***Clustering***. Wiley-Interscience.

Zhai, C. Massung, S., (2016). ***Text data management analysis: a practical ıntroduction to ınformation retrieval text mining***, ACM Books, New York, ISBN: 978-1-970001-16-7.

Zhang, D. (2018). *Big Data Security Privacy Protection*. **8th International Conference on Management Computer Science (ICMCS 2018)**, 275-278. Atlantis Press.

Zhang, Q., Wang, P. (2016). A Study on Readers' Satisfaction of University Library Based on BP Neural Network. *Journal of Computational Theoretical Nanoscience*, 13(9), 6005-6010.

Zhang, Y., Xu, Y.M. Zhang, Y. (2021). A multivariate linear regression prediction model for substation line loss rate based on a new k-means clustering algorithm. *Journal of Electric Power Science Technology*, 36, 179-186. https://doi.org/10.19781/j.issn.1673-9140.2021.05.022

Zeng, Q.T., Chen, G.H. Li, W.X. (2024). Rapid detection classification of steel by laser ınduced breakdown spectroscopy based on particle swarm-supportctor machine algorithm. *Spectroscopy Spectral Analysis*, 44, 1559-1565.

Zhou, Y.W. (2020). Design ımplementation of book recommendation management system based on ımproved apriori algorithm. *Intelligent Information Management*, 12, 1-10.

ÖZ GEÇMİŞ

Kişisel Bilgiler
Adı Soyadı :Engin DAYAN

Eğitim Bilgileri
Lisans
Üniversite :HacettepeÜniversitesi
Fakülte :Edebiyat Fakültesi
Bölüm :Bilgi ve Belge Yön.
Mezuniyet Yılı : 2011

Yüksek Lisans
Üniversite :ÇankırıKaratekinÜniversitesi
Enstitü :Edebiyat Fakültesi
Anabilim Dalı :Bilgi ve Belge Yön. Bölümü
Mezuniyet Yılı : 2015

Akademik Yayınlar

Dayan, E. (2018). Üniversite kütüphanelerinde elektronik bilgi kaynaklarının duyurulması ve elektronik danışma hizmetleri ile sürdürülmesi. **1. Uluslararası Iğdır Multi Disipliner Çalışmalar Kongresi**, 6-7 Kasım 2018.

Dayan, E. (2018). Teknik Dergisi'nin bibliyometrik analizi (2004-2013). **1. Uluslararası Iğdır Multi Disipliner Çalışmalar Kongresi**, 6-7 Kasım 2018.

Dayan, E. (2018). *Ticari bilgi ve lojistik sektörü rekabet gücü: Iğdır örneği* (Yüksek lisans tezi). Çankırı Karatekin Üniversitesi, Sosyal Bilimler Enstitüsü, Bilgi ve Belge Yönetimi Anabilim Dalı.

Dayan, E., & Polat, C. (2023). Ticari bilginin lojistik sektöründe rekabet gücüne etkisi üzerine bir değerlendirme. *ETÜ Sentez İktisadi ve İdari Bilimler Dergisi*, 12, 41-68.

Dayan, E. (2023). Genel sistem kullanıcısına yönelik ileri analitik yaklaşımlar: Bilgi erişim özelinde veri madenciliği. **2. International Scientific Research Congress**, 26-27 Kasım Eylül 2023. Iğdır.

Dayan, E. (2023). Günümüz teknolojilerinin geldiği nokta ve sağlık alanında kullanılan bilginin geleceği. **2. International Scientific Research Congress**, 26-27 Kasım Eylül 2023. Iğdır.

Dayan, E. (2023). Yakın gelecekte sağlık verilerinin güvenlik ve gizlilik gereklilikleri - Blockchain. Aydın N., (Ed.), ***İşletmelerde Yönetim Bilişim Sistemleri*** (s. 39-53) içinde. Bidge Yayınları. Ankara, Türkiye.

Dayan, E. (2023). Tıbbi bilgi grafikleri üzerine bir araştırma. Aydın N., (Ed.), ***İşletmelerde Yönetim Bilişim Sistemleri*** (s. 54-64) içinde. Bidge Yayınları.

ETİK KURUL KESİN SONUÇ ONAY BELGESİ

	T.C. **VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ** **FEN VE MÜHENDİSLİK BİLİMLERİ YAYIN ETİK KURULU BAŞKANLIĞI** **ETİK KURUL KARARLARI**
TOPLANTI TARİHİ: 20/12/2022 **OTURUM SAYISI: 2022/10** **TOPLANTIDA ALINAN KARAR SAYISI: 03**	**Sayfa: 02/03**

Van Yüzüncü Yıl Üniversitesi Fen ve Mühendislik Bilimleri Yayın Etik Kurulu 20/12/2022 tarihinde saat 11.00'de Prof. Dr. İsmail ÇELİK başkanlığında online olarak yapmış olduğu toplantıda aşağıdaki karar/kararları almıştır:

KARAR NO 2022/10-02. Danışmanlığını, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı öğretim üyesi Prof. Dr. Mahmut KARA'nın yapmış olduğu doktora öğrencisi Engin DAYAN'a ait, "Bilgi Merkezi Hizmetlerinin Kullanıcılar Tarafından Kullanım Tepki Durumlarının Veri Madenciliği Yaklaşımı ile İncelenmesi" adlı tez çalışmasında kullanılacak olan araçlar incelenmiş olup, söz konusu araçların ilgili kişilere uygulanmasında Fen ve Mühendislik Etik Kuralları ve İlkeleri çerçevesinde herhangi bir sakınca olmadığına toplantıya katılan üyelerin oy birliğiyle karar verilmiştir.

	BAŞKAN **Prof. Dr. İsmail ÇELİK** Fen Fakültesi	
ÜYE **Prof. Dr. Rüveyde TUNÇTÜRK** Ziraat Fakültesi	**ÜYE** **Prof. Dr. Seval ANDİÇ** Ziraat Fakültesi	**ÜYE** **Prof. Dr. Cabir TEMİRCİ** Fen Fakültesi
ÜYE **Prof. Dr. İbrahim Hakkı YÖRÜK** Fen Fakültesi	**ÜYE** **Prof. Dr. Musa ÇAKIR** Fen Fakültesi	**ÜYE** **Prof. Dr. Suat ŞENSOY** Ziraat Fakültesi

VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
LİSANSÜSTÜ TEZ ORİJİNALLİK RAPORU

Tarih 20/01/2025

Tez Başlığı: Bilgi Merkezi Hizmetlerinin Kullanıcılar Tarafından Kullanım Tepki Durumlarının Veri Madenciliği Yaklaşımı ile İncelenmesi

Yukarıda başlığı belirtilen tez çalışmamın, kapak sayfası, giriş, ana bölümler ve sonuç bölümlerinden oluşan toplam 81 (seksen bir) sayfalık kısmına ilişkin, 12/02/2025 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre tezimin benzerlik oranı %12 (on iki) dır.

Uygulanan filtreler aşağıda verilmiştir:

- Kabul ve onay sayfası hariç,
- Teşekkür hariç,
- İçindekiler hariç,
- Simge ve kısaltmalar hariç,
- Gereç ve yöntemler hariç,
- Kaynakça hariç,
- Alıntılar hariç,
- Tezden çıkan yayınlar hariç,
- 7 kelimeden daha az örtüşme içeren metin kısımları hariç (Limit match size to 7 words)

Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim.

Tarih ve İmza

Adı Soyadı: Engin DAYAN
Öğrenci No: 20910001176
Anabilim Dalı: İstatistik A.B.D.
Programı: İstatistik Doktora
Statüsü: ( ) Yüksek lisans (X) Doktora

DANIŞMAN
UYGUNDUR

ENSTİTÜ ONAYI
UYGUNDUR