

**T.C.**

**BAHÇEŞEHİR ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**

**ELEKTRONİK VERİTABANLARININ VERİ  
MADENCİLİĞİ YÖNTEMİ İLE ANALİZ EDİLMESİ**

**Yüksek Lisans Tezi**

**Erdem KARAASLAN**

**İSTANBUL, 2019**



**T.C.**

**BAHÇEŞEHİR ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİ TEKNOLOJİLERİ**

**ELEKTRONİK VERİTABANLARININ VERİ  
MADENCİLİĞİ YÖNTEMİ İLE ANALİZ EDİLMESİ**

**Yüksek Lisans Tezi**

**Erdem KARAASLAN**

**Tez Danışmanı: PROF. DR. TUFAN ADIGÜZEL**

**İSTANBUL, 2019**

**T.C.**  
**BAHÇEŞEHİR ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**  
**BİLGİ TEKNOLOJİLERİ YÜKSEK LİSANS PROGRAMI**

Tezin Adı: Elektronik Veritabanlarının Veri Madenciliği Yöntemi ile Analiz Edilmesi

Adı Soyadı: Erdem KARAASLAN

Tez Savunma Tarihi: 16.08.2019

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğu Fen Bilimleri Enstitüsü tarafından onaylanmıştır.

Dr. Öğr. Üyesi Yücel Batu SALMAN  
Enstitü Müdürü

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğunu onaylarım.

Prof. Dr. Mehmet Alper TUNGA  
Program Koordinatörü

Bu Tez tarafımızca okunmuş, nitelik ve içerik açısından bir Yüksek Lisans tezi olarak yeterli görülmüş ve kabul edilmiştir.

Jüri Üyeleri

İmzalar

Tez Danışmanı

Prof. Dr. Tufan ADIGÜZEL

-----

Üye

Dr. Öğr. Üyesi Serkan AYVAZ

-----

Üye

Doç. Dr. Tuncay SEVİNDİK

-----

## TEŐEKKÜR

Arařtırmanın hazırlanma sürecinde bana devamlı yol gösteren ve desteęini hiçbir zaman esirgemeyen deęerli hocam Prof. Dr. Tufan ADIGÜZEL'e, veri madencilięi konusunda arařtırmaya katkı saęlayan saygıdeęer hocam Dr. Öğretim Üyesi Serkan AYVAZ'a ve tez jürisi olarak davetimizi kabul eden saygıdeęer hocam Doç. Dr. Tuncay SEVİNDİK'e sonsuz teşekkürlerimi sunuyorum.

Arařtırma sürecinde bana desteęini hiçbir zaman esirgemeyen sevgili eřim Semanur DEMİREL'e, hayatımın her zamanında yanımda olan annem Naciye KARAASLAN'a, aramızda bulunmayan ve her zaman eksiklięini hissettięim babam Nurettin KARAASLAN'a, ablam Özlem HERDEM ve kardeřim Çiędem Karaaslan'a sonsuz teşekkürlerimi sunuyorum.

İstanbul, 2019

Erdem KARAASLAN

## ÖZET

### ELEKTRONİK VERİTABANLARININ VERİ MADENCİLİĞİ YÖNTEMİ İLE ANALİZ EDİLMESİ

Erdem KARAASLAN

Bilgi Teknolojileri Yüksek Lisans Programı

Tez Danışmanı: Prof. Dr. Tufan ADIGÜZEL

Ağustos 2019, 42 sayfa

Araştırma kütüphanelerinde elektronik kaynakların koleksiyon geliştirme politikalarının kütüphane yönetiminin ve kullanıcıların beklentileri doğrultusunda yapılması çok önemlidir. Bu çalışmada, İstanbul'da özel bir üniversite kütüphanesinde bulunan elektronik veritabanlarının kapsam, içerik, işlevsellik, maliyet gibi özelliklerinden elde edilen veriler kullanılarak performans durumları ortaya çıkartılmaya çalışılmıştır. Bu durumları araştırmak amacıyla veri madenciliği yöntemlerinden K-Means algoritması kullanılmıştır.

Araştırmadan elde edilen bulgular incelendiğinde, en fazla ve en az kullanıma sahip olan veritabanlarının kullanıcı taleplerinde yer aldığı bulgusu ortaya çıkmıştır. Elektronik veritabanlarını işlevsel özelliklerinden olan Web site tarama arayüzü, filtreleme özelliği ve düzenli güncelleme gibi durumlarının elektronik veritabanı kullanımında etkili olduğu bulgusu elde edilmiştir. Yıllık artış oranı yüksek olan veritabanlarının kullanıcı taleplerinde yer alan konuları içerdiği gözlemlenmiştir. Arşive ek ücret istenilen veritabanlarının da kullanımını en az olan elektronik veritabanlarında yer aldığı gözlemlenmiştir.

Araştırmanın sonucunda, en az kullanılan veritabanlarının kullanıcı taleplerinde yer alması bulgusu, bu elektronik veritabanlarının tanıtım faaliyetlerinin artırılması gerektiğini, artış oranı yüksek olan veritabanlarının kullanıcı taleplerindeki konu ve disiplinleri içermesi bulgusu ise bütçelerin çok etkin kullanılması gerektiğini ortaya çıkarmıştır. Web site arayüzlerinin kullanışlı olması, filtreleme özelliklerine sahip olması ve düzenli güncellenmesi elektronik veritabanları kullanımında etkili olduğu saptanmıştır. Bu bilgiler kapsamında elektronik veritabanları seçim ve değerlendirme çalışmalarında yol gösterici sonuçlar elde edildiğine inanılmaktadır.

**Anahtar Kelimeler:** Koleksiyon Geliştirme, Elektronik Veritabanları, Data Mining, K-Ortalamlar

## ABSTRACT

### ANALYSIS OF ELECTRONIC DATABASES WITH DATA MINING METHOD

Erdem KARAASLAN

Information Technology

Thesis Supervisor: Prof. Dr. Tufan ADIGÜZEL

August 2019, 42 pages

It is very important that the collection development policies of electronic resources in research libraries are made in line with the expectations of library management and users. In our research, data mining method was used to perform a detailed analysis by considering user demands and selection criteria in order to reveal the current status of electronic databases. In this study, a detailed analysis was conducted to discover the performance of electronic databases in a private university library in Istanbul. The data obtained from the features such as scope, content, functionality and cost were tried to be extracted from the performance status of the electronic databases.

In this study, the following questions were asked to investigate the user requests and the effective budget usage in order to reveal the performance status of the electronic databases: (1) Are the topics and disciplines covered by the widely used electronic databases included in user requests? (2) Are the topics and disciplines covered by the less used electronic databases included in user requests? (3) Are features such as scanning interfaces, filtering, and regular updates effective in the use of electronic databases ? (4) What is the status of using electronic databases with a high annual increase rate? (5) Do electronic databases with a high annual increase meet user demands ? (6) Is it sufficient to use electronic databases with archives and additional fees? In order to investigate these questions, K-Means algorithm which is one of the data mining methods is used. The data used in the analysis and the variables obtained under the scope, content, functionality and cost headings of the electronic databases were used.

When the findings obtained from the research were examined, it was found that the databases with the most and least usage were included in the user demands. One of the functional features of electronic databases, such as website browsing interface, filtering feature and regular updates, was found to be effective in the use of electronic databases. It is observed that the databases, which have a high annual increase rate, contains the subjects included in the user demands. It has been observed that the databases that require additional fees to the archive are one of least usage of the electronic databases.

As a result of the research, the current state of electronic databases was revealed. The fact that the least used databases are included in the usage demands indicates that the promotion activities of these electronic databases should be increased. The fact that

databases with high growth rate include the subjects and disciplines in user demands revealed that budgets should be used very effectively. The usefulness of the website interfaces, the fact that it has filtering features and its regular updating has proved to be effective in the use of electronic databases. Within the scope of this information, it is believed that guiding results are obtained in the selection and evaluation of electronic databases.

**Keywords:** Collection Development, Electronic Database, Data Mining, K-Means





## İÇİNDEKİLER

TABLolar .....	ix
ŞEKİLLER.....	x
1. GİRİŞ .....	1
1.1 PROBLEM DURUMU.....	1
1.2 ARAŞTIRMANIN AMACI.....	2
1.3 ARAŞTIRMANIN ÖNEMİ .....	2
2. ALANYAZIN TARAMASI.....	4
2.1 ELEKTRONİK KAYNAKLAR .....	4
2.1.1 Elektronik Kaynakların Türleri .....	6
2.1.2 Elektronik Kaynaklar Koleksiyon Geliştirme Politikası.....	11
2.2 VERİ MADENCİLİĞİ.....	12
2.2.1 Veri Madenciliğinin Uygulama Alanları.....	13
2.2.2 Veri Madenciliği Süreci .....	15
2.2.3 Veri Madenciliği Teknikleri ve Yöntemleri .....	17
2.2.4 Yapılan Çalışmalar .....	21
3. YÖNTEM.....	25
3.1 ARAŞTIRMANIN MODELİ .....	25
3.2 ÖRNEKLEM .....	25

<b>3.3 VERİLERİN DÜZENLENMESİ VE TOPLANMASI .....</b>	<b>26</b>
<b>3.3.1 Kullanıcı İstatistikleri .....</b>	<b>27</b>
<b>3.3.2 İçerik .....</b>	<b>28</b>
<b>3.3.3 İşlevsellik.....</b>	<b>28</b>
<b>3.3.4 Maliyet.....</b>	<b>28</b>
<b>3.4 VERİLERİN ANALİZİ .....</b>	<b>30</b>
<b>4. BULGULAR .....</b>	<b>32</b>
<b>4.1 BİRİNCİ ARAŞTIRMA SORUSUNA GÖRE BULGULAR.....</b>	<b>34</b>
<b>4.2 İKİNCİ ARAŞTIRMA SORUSUNA GÖRE BULGULAR .....</b>	<b>35</b>
<b>4.3 ÜÇÜNCÜ ARAŞTIRMA SORUSUNA GÖRE BULGULAR.....</b>	<b>35</b>
<b>4.4 DÖRDÜNCÜ ARAŞTIRMA SORUSUNA GÖRE BULGULAR.....</b>	<b>36</b>
<b>4.5 BEŞİNCİ ARAŞTIRMA SORUSUNA GÖRE BULGULAR .....</b>	<b>36</b>
<b>4.6 ALTINCI ARAŞTIRMA SORUSUNA GÖRE BULGULAR.....</b>	<b>36</b>
<b>5. TARTIŞMA VE ÖNERİLER .....</b>	<b>37</b>
<b>5.1 BULGULARININ TARTIŞILMASI.....</b>	<b>37</b>
<b>5.2 ÖNERİLER.....</b>	<b>39</b>
<b>6. SONUÇ.....</b>	<b>41</b>
<b>KAYNAKÇA .....</b>	<b>43</b>

## TABLÖLAR

Tablo 2.1: Elektronik Kaynakların Tarihsel Gelişimi .....	5
Tablo 3.1: 2017-2018 Konularına Göre Yayın Talep Sayıları .....	27
Tablo 3.2: Birim Başlı Maliyetlerin Düzeyleri .....	28
Tablo 3.3: Yıllık Fiyat Artış Oranlarının Düzeyleri .....	29
Tablo 3.4: Değişkenlerin Düzenlenmeden Önceki Durumu .....	29
Tablo 3.5: Değişkenlerin Düzenlendikten Sonraki Durumu .....	30
Tablo 3.6: Küme Sayısına Göre Küme İçi Hata Kareler Toplamı .....	31
Tablo 4.1: Küme Üyeleri .....	33
Tablo 4.2: Küme Merkezlerin Birbirine Uzaklığı .....	33



## ŞEKİLLER

Şekil 2.1: Veri Madenciliği Süreci .....	16
Şekil 4.1: Kümelerin Dağılımı .....	32
Şekil 4.2: Değişkenlerin Kümelere Dağılımı .....	34



# 1. GİRİŞ

## 1.1 PROBLEM DURUMU

Gelişen teknoloji ile birlikte yayınların elektronik ortama geçmesiyle bilginin paylaşılması ve erişimi daha kolay olmuştur. Bilgiye olan talep artmakla birlikte bilgi üretimi de sürekli artış göstermektedir. Bilgi toplumunda bulunduğumuz günümüzde bilginin üretilmesi akıl almaz boyutlarda artmıştır. Bilim insanlarının güncel kalabilmesi için büyük çabalar gerekmektedir. Güncel bilginin toplum tarafından kolay erişilmesi, kullanıcıların öğrendiklerini yaşamlarında uygulamasını da hızlandırmıştır (Kıpçak, 2018). Bu ihtiyacı karşılamak için devreye giren üniversite kütüphaneleri, gerekli bilgiyi bütün kullanıcılarına ekonomik bir biçimde ulaştırmak zorundadır (Çakın, 2000). En çok ihtiyaç duyulan alanlar ile ilgili veritabanlarının kütüphane koleksiyonlarında bulunması kullanıcılar için büyük önem arz etmektedir. Kütüphanelerin koleksiyon geliştirme politikasındaki amacı, bütün kaynakların değil, kullanıcıların ihtiyaç duydukları kaynakları koleksiyonlarında bulundurmasıdır (Agee, 2005 ve Ünal, 2008). Bu doğrultuda, elektronik kaynakların güncel bilgiye hızlı ve kolay erişim sağladıkları için kullanıcılar için önemi büyüktür.

Diğer yandan, akademik kütüphane bütçelerinin yaklaşık %95'i elektronik kaynaklara ayrılmaktadır (Gürdal ve Can, 2017). Elektronik kaynakların sayısında görülen artıştan dolayı kütüphane bütçelerindeki payının artmasıyla, seçim ve değerlendirilme konusu büyük önem kazanmıştır (Atılğan ve Yalçın, 2009). Elektronik kaynakların seçim kriterleri doğası gereği basılı kaynaklardan daha karmaşık değerlendirme süreçlerini zorunlu kılmıştır. Bu bağlamda, sınırlı olan bütçeyi en verimli şekilde kullanmak için kullanıcıların ihtiyaçlarına göre hareket etmeleri ve kapsamlı bir değerlendirme yapmaları gerekmektedir. Üniversite kütüphanelerinde kullanıcıların elektronik veritabanlarını etkin kullanabilmesi için mevcut durumun ortaya çıkarılması ve kullanıcıların istedikleri yayınları sağlaması için kullanıcıların ne istediğinin ayrıntılı olarak analiz edilmesi gerekmektedir. Bununla birlikte elektronik kaynaklar arasında doğru kaynakları sağlamaya yardımcı olacak seçim ölçütlerinin de kullanılması gerekmektedir (Al ve Al, 2003, s. 4; Davis, 1997, s. 392; Guenther, 2000, s. 58). Bu

çalışmada gelişmiş analiz yöntemlerinden olan veri madenciliği kullanılarak elektronik veritabanlarının detaylı analizi amaçlanmış ve mevcut durumu ortaya çıkartılmaya çalışılmıştır.

## **1.2 ARAŞTIRMANIN AMACI**

Bu tez çalışmasında, sınırlı olan bütçeyi kullanıcının ihtiyaçlarını gözeterek verimli bir şekilde kullanmak için elektronik veritabanlarının performans özelliklerinin veri madenciliği yöntemleriyle ortaya çıkartılması amaçlanmıştır. Bu amaca yönelik araştırmanın soruları aşağıdaki biçimde belirlenmiştir:

- a. Kullanımı çok olan elektronik veritabanlarının kapsadığı konu ve disiplinler kullanıcı taleplerinde yer almakta mıdır?
- b. Kullanımı az olan elektronik veritabanları kullanıcı taleplerindeki konu ve disiplinlerde yer almakta mıdır?
- c. Tarama arayüzleri, filtreleme ve düzenli güncelleme gibi özellikler elektronik veritabanlarının kullanılmasında etkili midir?
- d. Yıllık artış oranı yüksek olan elektronik veritabanlarının kullanılma durumları nasıldır?
- e. Yıllık artış oranı yüksek olan elektronik veritabanları kullanıcı taleplerini karşılamakta mıdır?
- f. Arşivi olan ve ek ücret istenilen elektronik veritabanlarının kullanımı yeterli midir?

## **1.3 ARAŞTIRMANIN ÖNEMİ**

Akademik kütüphanelerin bütçelerinin verimli kullanılabilmesi, bütçe planlamalarının doğru yapılabilmesi için elektronik kaynakların seçimi, değerlendirilmesi ve yönetilmesi süreçleri büyük önem taşımaktadır (Gürdal ve Can, 2017). Bu çalışmada, elektronik veritabanlarının seçim ve değerlendirilmesinde bazı ölçütler kullanılarak veri madenciliği kümeleme yöntemiyle detaylı bir analiz gerçekleştirilmek istenmiştir. Elde edilen sonuçlar, sonraki dönemler için elektronik kaynakların seçim ve değerlendirilmesinde karar destek sağlanması açısından önemli olduğu düşünülmektedir. Alanyazı

incelendiğinde ülkemizde kütüphane uygulamalarında veri madenciliđi yöntemlerinin fazla kullanılmadıđı görölmektedir. İşletme, sađlık, pazarlama gibi alanlarda oldukça fazla kullanılan veri madenciliđi yöntemlerinin kütüphanecilik alanında kullanılarak elektronik veritabanlarının seçim ve deđerlendirilmesinde detaylı bir analiz yapılması ve bu dođrultuda bütçe planlamasının yapılmasının önemli olduđu düşünölmektedir.



## 2. ALANYAZIN TARAMASI

### 2.1 ELEKTRONİK KAYNAKLAR

Bilimin daha hızlı gelişmesi bilginin etkin kullanılması ile sağlanmaktadır. Bilgi, teknolojinin gelişimi ile birlikte herkes tarafından kolayca erişilmektedir. İnternet'in yaygın kullanılmasıyla bilginin üretiminde, depolanmasında ve erişiminde önemli değişimler meydana gelmiştir (Kaur, Kaur, Valia ve Valia, 2016). Bu değişimler aynı zamanda yayıncılık sektörünün de gelişmesine yol açmıştır.

Elektronik yayıncılık belgelerin sanal ortamda sunucular tarafından dağıtımı, depolanması ve erişimi olarak tanımlanmaktadır (Tonta, 2000). Elektronik yayıncılığın kullanıldığı alanlar genel olarak kitap, dergi, gazete yayıncılığı, bülten panoları, haber grupları, posta listeleri ve web siteleri gibi dijital ortamlarda üretilen materyalleri kapsamaktadır (Özarslan, Balaban-Sali ve Demiray, 2013). Dünyada internet ağının yaygınlaşmasıyla elektronik yayıncılık, akademik alanların yanı sıra devlet kuruluşları, uluslararası örgütler, ticari ve askeri kuruluşlar tarafından kullanılmaya başlanmıştır (Çağıltay, 2015). Bu gelişmelerle birlikte araştırma kütüphaneleri de büyük ölçüde etkilenmiş ve derme geliştirme politikalarının büyük çoğunluğunu bilgisayar, mobil cihazlar ve elektronik kitap okuyucu cihazları tarafından erişilebilen elektronik kaynakların sağlanması ve yönetilmesi üzerine kurmuşlardır. Diğer yandan, elektronik kaynakların uzaktan erişiminin olması, güncel bilgiye hızlı erişim sağlaması ve de maliyetlerinin düşük olması kütüphane bütçelerinin büyük çoğunluğunun elektronik kaynaklarda kullanılmasına yol açmıştır.

Elektronik kaynakların tarihsel süreçte, çevrimiçi, uzaktan erişim ve web tabanlı olarak gelişmiştir (Tablo 2.1). İlk elektronik kaynak kullanımı; 1964 yılında tıp ve sağlık bilimleri alanlarındaki literatüre erişim için kullanılan MEDLARS (Medical Literature Analysis and Retrieval System) ve çevrimiçi olarak kullanılan ürünü MEDLINE'dır. 1971 yılında ERIC (Education Resource Information Center) bibliyografik veri tabanına uzaktan erişim sağlanmıştır (Johnson, 2013). Elektronik kaynakların kütüphaneler tarafından çevrimdışı olarak sağlanma şekli ise CD-ROM üzerinden olmakla birlikte



(Polat, 2017), CD-ROM veri tabanları basılı süreli yayınları iptal etmek yerine daha çok koruma amaçlı kullanılmıştır (Johnson, 2013). Elektronik kaynakların 1990 yılların başında web tabanlı olarak ticareti yapılmaya başlanmıştır. LexisNexis (1997) yılında hukuk konusunda hizmet vermiş ilk web tabanlı veri tabanı olmuştur (Cotter ve Koehler, 2002). 2000’li yılların başında elektronik kaynak sağlayıcıları bibliyografik veri tabanları ile birlikte tam metin içerik de sunmaya başlamışlardır (Johnson, 2013). Teknolojinin gelişmesi ile paralel olarak elektronik kaynaklar kütüphaneler tarafından çevrimdışı ve çevrimiçi olarak kullanılmıştır. Basılı olan kitaplar ve süreli yayınlar zamanla elektronik ortamda üretilmeye başlanılmıştır. Elektronik ortamda, e-kitap, e-süreli yayınlar olduğu gibi elektronik ortamın doğası gereği çoklu ortam materyalleri olarak da kütüphaneler tarafından sunulmaktadır.

**Tablo 2.1: Elektronik kaynakların tarihsel gelişimi**

Adı	Üreticisi	Konusu	Türü	Tarihi
MEDLARS VE MEDLINE	US National Library of Medicine	Tıp ve Sağlık Bilimleri	Çevrimiçi	1964
ERIC	Institute of Education Sciences (IES) of the U.S. Department of Education	Eğitim Bilimleri	Uzaktan Erişim	1971
LexisNexis	LexisNexis	Hukuk	Web tabanlı	1997

Elektronik kaynakların kullanımının kolay olması ile birlikte birçok avantajı da bulunmaktadır (Chandra, 2002):

- a. Küreselleşmiş erişime imkan sağlar.
- b. Kullanıcılar aynı kaynakları, aynı zamanda, herhangi bir yerde kullanabilirler,
- c. Kolayca kopyalanır, saklanır ve dağıtılır.
- d. Kolay güncellenir, yönetilir ve entegre edilir.
- e. Kağıttan daha az hantaldır.
- f. Hızlı teslim edilir.
- g. Hizmetlere değer katar.
- h. Elektronik kaynaklar kullanıcılara zaman kazandırır.

- i. Kullanıcılarda memnuniyet yaratır.
- j. Elektronik kaynakların satımı WWW (World Wide Web) üzerinden yapılır.
- k. Uzaktaki kullanıcılar ile etkileşimi destekler.
- l. Çevrimiçi kullanıcıların takibi mümkündür.
- m. Elektronik kaynaklar Ar-Ge'yi etkin ve verimli bir şekilde aktarır. (s.2)

Waghmare ve Salunkhe'ye (2013,) göre elektronik kaynakların sağladığı avantajlar ise şöyledir;

- a. Çoklu Erişim - Birden çok eşzamanlı kullanıcı için 7/24 erişim imkanı sağlar.
- b. Hız - Bilgiyi arama ve göz atmada elektronik cihazlar çok hızlıdır, çeşitli yayınlar arasında çapraz referans arama yapılabilir.
- c. İşlevsellik - Kullanıcılara yayınların içerik analizini yapmasına olanak sağlar.
- d. Tür - İçerik - Farklı bilgi kaynakları ve karma ortamları içerir (s. 262-264).

Basılı yayınların fiyatlandırılmasında basım, dağıtım ve editoryal giderler etkili olmakla birlikte elektronik yayınların fiyatlandırılması yapılırken birçok değişkene bakılmaktadır: ağ ve sunucu maliyetleri, kullanıcı sayıları ve kullandığı teknolojik özellikler.

### **2.1.1 Elektronik Kaynakların Türleri**

Elektronik kaynaklar çevrimiçi ve çevrimdışı olarak ayrılrsa da e-kitap, süreli yayınlar ve çoklu ortam materyalleri olarak kütüphaneler tarafından dermelerine eklenilmektedir.

#### **2.1.1.1 E-Kitap**

E-Kitap, Rao (2003)'e göre bilgisayar dosya formatı ile sayısal biçime dönüştürülen metin, kitap, okuma materyali ya da görüntülerin elektronik biçimlerde bilgisayar ve onlar için üretilmiş e-kitap okuyucular ile okunabilen materyaller olarak tanımlarken; Konya (2004) ise e-kitapları, sayısal ortamlarda kullanılmak için sayısal formata çevrilmiş referans kitabı, ders kitabı, ticari kitap gibi her türlü yayınlanmış materyal ile sayısal ortamda üretilmiş metin ve resmin yanında ses ve hareketli görüntüleri de içeren,

onlar için üretilmiş cihazlar ile bilgisayarlarda okunabilen materyaller olarak tanımlamıştır.

“Alan Key “Dynabook” adlı eserini 1960 yılında yayınlayarak e-kitaplardan bahsetmiştir” (Ruknacı ve Anameriç, 2003, s.149). İlk e-kitapların ortaya çıkışı, 1971 yılında İllinois Üniversitesin’de Micheal Hurt tarafından bilgisayarların sunucuya çevrilerek ABD Özgürlük Bildirgesinin yayınlanması olarak gösterilmiş ve bunun sonucunda 10.000 kitaplık elektronik halk kütüphanesinin yapılabilirliğini göstermek amacıyla Gutenberg Projesi yapılmıştır. (Grimes, 2011; Project Gutenberg, 2017). Kütüphanelerde ilk e-kitapların kullanımı 1990’lardan itibaren ABD’deki bazı kütüphanelere çevrimiçi kataloglar üzerinden link verilerek gerçekleştirilmiştir (Small, 2000). Elektronik Kaynak sağlayıcıları 1980’lerde e-kitapları CD-ROM’larda pazarlamaya başlamış olsalar da 1990’ların ortalarında Ebrary, NetLibrary ve Questia kullanıcılarına ve kütüphanelere bu ürünlerini çevrimiçi olarak sunmuşlardır (Johnson, 2013).

E-kitap hizmetleri yaygınlaşarak ücretsiz okuma ve indirme hizmeti sunan platformların oluşmasını sağlamıştır. Project Gutenberg bu konuda öncü girişim olarak 53.000’den fazla e-kitabın html, e-pub ya da Kindle formatında indirilerek okunmasını sağlamıştır (Project Gutenberg, 2017). 2004 yılında ise Google tarafından Bodleian Kütüphanesi’nin 19. Yüzyıla ait telif hakkı olmayan yaklaşık bir milyondan fazla kitabı dijital ortama aktarılmıştır. ([https://www.bodleian.ox.ac.uk/news/2009/2009\\_mar\\_26](https://www.bodleian.ox.ac.uk/news/2009/2009_mar_26), 2019).

E-kitaplar, kullanımının yaygınlaşması ile kütüphanelerde dermelerinin büyük çoğunluğunu oluşturmuşlardır. 2015 yılında Library Journal tarafından yapılan araştırma ile ABD’deki halk kütüphanelerinin %94’ü e-kitap hizmeti sunduklarını ve e-kitap dermesinin de %300 arttığını ortaya çıkarmıştır (Ebook Usage in U.S. Public Libraries, 2015).

E-kitaplar için kullanılan formatlar Adobe firması tarafından “.pdf” olarak, Microsoft Windows için ise diskler gibi çeşitli biçimlerde geliştirilmiştir. Uluslararası Dijital Yayıncılık Forumu (IDPF) tarafından geliştirilen EPUB, XML tabanlı bir format olup

yayıncılara e-kitap oluşturmak ve dağıtmak için imkan sunmaktadır. Daha sonra EPUB 2 olarak geliştirilen bu format şifrelenmiş olarak tüketicilerle birlikte çalışabilirlik sunmaktadır. Son versiyonu olan EPUB 3 formatı, 2010 yılında IDPF'nin üyeliği ile onaylanan bir tüzükle geliştirilmiştir. E-kitaplar için günümüze kadar geliştirilen diğer formatlara HTML5, .exe, .app, .mobi ve tablet uygulamaları için geliştirilen versiyonları örnek verilebilir (Millidere, 2013). Kullanılabilirlik ve erişim özellikleri dikkate alınarak elektronik kitaplar Crawford tarafından dokuz türde gruplandırılmıştır (Rao, 2005):

- a. Tescilli elektronik kitaplar: Metin tek bir kullanıcıya ait olmak üzere tescilli olarak indirilip kullanılır. Belli kitap okuyucu aygıtlar için özel üretilen türden kitaplardır.
- b. Açık elektronik kitaplar: Her türlü donanımla okunabilecek belli bir sınırlamanın olmadığı genelde XML ortamında üretilmiş kitaplardır. Bu tür kitaplar açık elektronik kitap formunca hazırlanmıştır.
- c. Ücretsiz veya halka açık elektronik kitaplar: Bu tür kitaplar ücretsiz olarak kopyalanabilir ve basılabilir. Bu tür kitapların en meşhur sağlayıcıları Project Gutenberg, Inter Public Library'dir.
- d. Nüsha/kopya kitaplar: Kütüphane tarafından okuyucuya sunulmak için alınan bazı kitaplar kullanıcı tarafından bilgisayarlarına indirilir. Kütüphanenin indirilen elektronik kitap sayısına göre ödeme yaptığı elektronik kitap türüdür.
- e. Instabooks: Basılı kitapların taranarak oluşturulduğu ya da dijital verilerin bir araya getirilip oluşturulduğu kitap türüdür.
- f. Tam bir kitap sayılmayan elektronik kitaplar: Kısa roman ve hikâyelerin PDF formatında dönüştürülüp internet sitelerinde yayınlanması amacıyla oluşturulan kitap türüdür.
- g. E-vanity ve kendi kendine yayıncılık: Yazarın kendi çalışmalarını masrafını kendisi karşılamak şartıyla internette ya da kendi sitesinde yayınlamasıdır.
- h. Web kullanımından önceki kitaplar: İnternetin kullanımından önceki yıllarda CD-Rom ve disket ortamına kayıt edilip yayınlanan kitaplardır.
- i. Genişletilmiş elektronik kitaplar: İnternet ortamında veya CD-Rom ortamında hazırlanan kitapların çoklu ortam özellikleri ile zenginleştirilip oluşturulan kitap türüdür ( s. 121).

### 2.1.1.2 E-Sürelı Yayınlar

Kütüphane dermelerinin en önemli parçaları olan E-sürelı yayınlar Cole (2004) tarafından bilgisayar ağıları üzerinden elde edilen her türlü sürelı yayın olarak tanımlanırken; Ashcroft ve Langdon (1999, s. 105) tarafından “CD-ROM ve elektronik ortamlarda kullanılan sürelı yayın” olarak tanımlanmıştır. İnternetin yaygınlaşması ile günümüzde sürelı yayınlara artık Web üzerinden erişilebilmektedir. Ann C. Weller’ in 2002 yılında yayınlanan makalesinde elektronik dergi türleri üç başlık altında değerlendirmiştir: (i) Dergiler sadece elektronik ortamdadır, basılı versiyonu yoktur, (ii) Yazılı ve elektronik olanların başlıkları aynıdır. Her iki yayının kendine özgü bir içeriği vardır ve her ikisinin de ayrı ISSN’leri vardır, ve (iii) Hem elektronik hem de basılı olarak vardır ve aynı başlıkları içerir. Tek bir ISSN’si vardır.

Sürelı yayınların kütüphanelerin en önemli parçalarından biri olmasının en önemli sebebi kullanıcıların güncel bilgiye olan ihtiyaçlarıdır. Tennessee Üniversitesi’nde yapılan bir araştırmada makalelerin en çok kullanılma sebebi %75 oranında araştırma yapmak, %58 oranında eğitime destek, %32 oranında resmi yayınlardan haberdar olmak ve %20 oranında resmi sunumlardan haberdar olmak olarak ortaya çıkmıştır (Tenopir ve King, 1998). Crane (1972) araştırma çalışmalarının daha fazla kullanıcıya ulaşabilmesi için hakemli dergilerde yayınlanmasına “Resmi İletişim Kanalı” adını vermiştir. Bilimsel iletişim olabilmesi için profesyonel meslektaşlar ve öğrenciler arasında resmi ya da resmi olmadan değerlendirilme ve paylaşılması gerekmektedir (Rao, 2001). Rowland (1997), bilimsel dergilerin bilimsel iletişimde dört ana işlevi olduğunu belirtmiştir:

- a. Bilginin dağıtımı (dissamination of information)
- b. Kalite kontrolü (quality control)
- c. Doğal, standart arşiv (canonical archive)
- d. Yazarların tanınması (recognition of authors) (s. 170).

Elektronik sürelı yayınlar; kullanıcılara ilk olarak e-posta, CD-ROM ve FTP sunucular ile ulaştırılırken (Curtis, 2000), “internet üzerinden 1993 yılında dergiler tam metin olarak

Amerikan Kimya Kurumu tarafından sunulmuştur” (Tonta, 1997, s. 306). Bilginin güncel olarak erişilmesinde süreli yayınların akademik dünya açısından önemi büyüktür. Yayınevleri 90’lı yılların başında bu alana daha fazla girmesiyle elektronik yayıncılık artmıştır. 1991-1995 yılları arasında Elsevier Science, TULIP (The University Licensing Program)’ı devreye sokarak dokuz üniversite için 43 dergiyi elektronik ortama aktarmıştır. 1993 yılında Andrew W. Mellon Foundation tarafından fazla talep gören çeşitli disiplinlerdeki 10 derginin eski sayılarını elektronik ortama aktarımı yer sıkıntısına çözüm sağlamak için beş kütüphanede başlamıştır (Curtis, 2000).

İnternetteki gelişmelere paralel olarak e-süreli yayınlar alanyazında daha fazla kullanılmaya başlanılmıştır. 2000 -2009 yılları arasında bilimsel dergiler elektronik ortamda %560 artış gösterirken (Johnson, 2013), 2004 yılında e-süreli yayınların 30.000’nin üzerinde olduğu belirtilmiştir (Cole, 2004). 2016 yılına gelindiğinde ise bilimsel süreli yayınlar olan hakemli dergilerin elektronik ortamda sayıları 300.000’i geçmiştir (Ulrichsweb, 2017). Bilginin güncel erişilmesinde vazgeçilmez olan ve kütüphanelerin bütçelerinde fazla yer kaplayan e-süreli yayınlar bilimsel çalışmaların her yıl artmasıyla önemi de artmaktadır. Erişimi ve paylaşılması kolay olan e-süreli yayınların basılı formatları azaldığı gibi sadece elektronik ortamda üretilmeye devam etmektedir.

Elektronik süreli yayınların kendine özgü bazı avantajları ve dezavantajları bulunmaktadır. Elektronik süreli yayınlara, zaman ve mekan sınırlaması olmaksızın erişilebilmesi, süreli yayınların içerisinde tam metin arama yapılabilmesi, yeni dokümanlar yayınlandığında e- posta yolu ile kullanıcılara bildirilmesi, bazı süreli yayınlara basılışından haftalar öncesinden elektronik olarak erişilebilmesi, basılı süreli yayına göre daha ucuz maliyetlerinin olması, bir ekranda birden çok sekme açılarak aynı anda birden çok makale okunabilmesi, video, interaktif vb. seçenekli sunum zenginliği sunması, basılı süreli yayına oranla daha çok atıf alması elektronik süreli yayınların başlıca avantajlarından. “Bilgisayar ekranından okuma, arşivleme, bağlantı ömrü ve telif hakkı sorunu elektronik süreli yayınların başlıca dezavantajlarından” (Atılğan ve Yalçın, 2009; s. 774 – 776)

### 2.1.2 Elektronik Kaynaklar Koleksiyon Geliştirme Politikası

Kütüphaneler elektronik kaynaklara erişimde farklı yöntemler kullanmaktadır. Bu yöntemler; kurumsal abonelikler, aracı şirketler, ya da konsorsiyumlar aracılığı ile abonelik işlemleri sağlamakta veya satın almaktadırlar. Kurumsal aboneliklerde kütüphanenin ihtiyacına göre elektronik dergilere doğrudan abone olunmaktadır. Aracı şirketlerle yapılan abonelikler maliyet açısından daha avantajlı seçenekler sunmaktadır. Ülkemizde Ankos (Anadolu Üniversite Kütüphaneleri Konsorsiyumu) ile daha uyguna kütüphanelerin elektronik kaynaklara toplu abonelikleri yapılmaktadır.

Kütüphane bütçelerinin büyük çoğunluğunun elektronik kaynaklara ayrıldığı ve elektronik kaynakların kapsam, erişim ve depolanmasının farklı özelliklerde olması, seçim ve değerlendirmede değişik metotları zorunlu kılmıştır. Buna göre, doğru kaynakların sağlanmasına yardımcı olacak seçim ölçütlerinin değerlendirilmesi durumu ortaya çıkmaktadır. Bu ölçütler; kullanıcı talepleri, kullanıcı istatistikleri, kapsam, içerik, işlevsellik, maliyet, lisans ve teknik özelliklerdir (Al ve Al, 2003, s. 4; Davis, 1997, s. 392; Guenther, 2000, s. 58).

Kullanıcı talepleri, kullanıcıların talepte bulunduğu elektronik kaynaklardır. Kütüphane koleksiyon geliştirme politikalarında kullanıcılarının gereksinimlerini tam olarak karşılayıp karşılamadığının gözden geçirilmesi gerekmektedir (Al ve Al, 2003; Evans, 1995; Kluegel, 1997). Taleplerini mail yoluyla iletebildikleri gibi sistem üzerinden de yapabilmektedirler. Kullanıcı istatistikleri, elektronik kaynakların ne kadar kullanıldığını göstermektedir. Kullanıcı istatistiklerinin COUNTER üzerinden yapılıyor olması istatistiklerin güvenilir olduğunu göstermektedir. Bazı veritabanlarında COUNTER özelliği mevcut değildir. Kapsam dahilinde sahip olunan kaynakların hangi disiplin ve konuları kapsadığı ve elektronik süreli yayınların hakemli olup olmadığı gibi özelliklere bakılmaktadır. Elektronik kaynakların içeriğinin kütüphane yönetiminin beklentilerini karşılaması için şu sorulara cevap vermelidir: (1) Geriye dönük erişim olanağı nedir? (2) Düzenli güncelleme yapılıyor mudur? (3) Açıklayıcı bilgiler var mıdır? (Yardım sayfaları, Türkçe kullanım kılavuzları), (4) Dergilere erişimde kısıtlamalar var mıdır?

(Üçüncü parti olarak adlandırılan veritabanlarında belirli zaman aralıklarında ambargolara rastlanmaktadır.) (Akçal ve Kaya, 2006; Al ve Al, 2003; Polat, 2017).

İşlevsellik göz önüne alındığında, elektronik kaynakların sahip olduğu teknolojik özellikler farklılık göstermektedir. Teknolojik özellikler kolay kullanım olması açısından şu sorulara cevap vermelidir: (1) Uzaktan erişime imkan var mıdır? (2) Tarama arayüzleri kullanışlı mıdır (Kovacs ve Elkordy, 2000; Guenther, 2000; Gürdal ve Can, 2017) ? (3) Filtreleme özelliklerine sahip midir (Atılğan ve Yalçın, 2009) ? (4) Açıklayıcı bilgiler var mıdır? (Yardım sayfaları, Türkçe kullanım kılavuzları) (Allison, 2000). Elektronik kaynakların maliyeti, erişim yöntemlerine ve aynı anda erişilen kullanıcı sayısına göre belirlenmektedir (Salman, 2012). Elektronik kaynakların maliyetine ilişkin şu unsurlara bakılmaktadır: (1) Geriye dönük erişim hangi yıla kadar inmektedir? (2) Arşive ulaşımında ek bir ücret talep edilmekte midir? (3) Kullanıcı sınırlaması var mıdır? (4) Veritabanı ücretlendirilmesi neye göre yapılmaktadır? (5) Yıllık fiyat artış oranları nedir? (Atılğan ve Yalçın, 2009; Gürdal ve Can, 2017). Son özellik olarak, kütüphane yönetimi ve konsorsiyum, veritabanlarını almak istediğinde üretici veya satıcı ile elektronik kaynakların kullanımına yönelik sınırlamaların ve izinlerin yer aldığı lisans anlaşması imzalamaktadırlar (Atılğan ve Yalçın, 2009).

## **2.2 VERİ MADENCİLİĞİ**

Veri madenciliği, büyük verilerin içindeki gizli kalmış bilgilerin ortaya çıkarılması veya gelecekte olabilecek olayların tahmin edilmesi için bilgisayar uygulamaları ile analiz edilme sürecidir (Doğan ve Türkoğlu, 2007). Kalikov (2006) ise, çok miktardaki verilerden gizli kalmış bilgileri ve bu veriler arasındaki ilişkileri ortaya çıkarmak için kullanılan veri analizi yöntemi olarak tanımlamıştır. Hand (1998, s.112), “Veri madenciliğini istatistik, veritabanı teknolojisi, örüntü tanıma makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veri tabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi”, Kitler ve Wang (1998, S.46), “Veri madenciliğini oldukça tahminci anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesini sağlama



yeteneđi”, Jacobs (1999, s.8) ise, “Veri madenciliđini, ham datanın tek bařına sunamadıđı bilgiyi ıkararak veri analizi sreci ” olarak tanımlamıřlardır.

### 2.2.1 Veri Madenciliđinin Uygulama Alanları

Veri madenciliđi raporlamanın ve veri tabanının olduđu her yerde kullanılabilir (am, 2014). Veri madenciliđi sektrel olarak dřnldđnde yaygın olarak kullanılan uygulama alanlarının bařlıca olanları ařađıda sıralanmıřtır (Akpınar, 2000; Silahtaroglu, 2013; Han ve Kamber, 2006). Klasik hesaplama yntemlerinin yeterli olmadıđı durumlarda gizli bilgilerin ortaya ıkarılması ve gelecekte veriler arasındaki iliřkilerin tahmin edilmesi amacıyla verilerin analizi sađlanmaktadır.

#### Pazarlama

- a. Mřterilerin satın alma rntlerinin belirlenmesi
- b. Mřterilerin demografik zellikleri arasındaki bađlantıların bulunması
- c. Posta kampanyalarında cevap verme oranının artırılması
- d. Pazar sepeti analizi (Market Basket Analysis)
- e. Mřteri iliřkileri ynetimi (CRM: Customer Relations Management)
- f. Mřteri deđerlendirme
- g. Satıř tahmini
- h. Mřteri dađılımında
- i. eřitli pazarlama kampanyalarında
- j. Mevcut mřterilerin elde tutulması iin geliřtirilecek pazarlama stratejilerinin oluřturulmasında
- k. apraz satıř analizleri
- l. eřitli mřteri analizlerinde

#### Banka ve Sigortacılık

- a. Farklı finansal gstergeler arasında gizli korelasyonların bulunması
- b. Kredi kartı ve sigorta dolandırıcılıklarının tespiti
- c. Kredi taleplerinin deđerlendirilmesi.
- d. Mřteri dađılımında

- e. Usulsüzlük tespiti
- f. Yeni poliçe talep edecek müşterilerin tahmininde
- g. Riskli müşterilerin örüntülerinin belirlenmesinde

#### Perakendecilik

- a. Satış noktası veri analizleri
- b. Alış-veriş sepeti analizleri
- c. Tedarik ve mağaza yerleşim optimizasyonu
- d. Hisse senedi fiyat tahmini
- e. Genel piyasa analizleri
- f. Alım-satım stratejilerinin optimizasyonu

#### Telekomünikasyon

- a. Kalite ve iyileştirme analizleri
- b. Hatların yoğunluk tahminleri

#### Biyoloji, Tıp, Genetik ve Kimya

- a. Bitki türlerinin ıslahı
- b. Gen haritasının analizi ve genetik hastalıkların tespiti
- c. Kanserli hücrelerin tespiti
- d. Yeni virüs türlerinin keşfi ve sınıflandırılması
- e. Yeni kimyasal moleküllerin keşfi ve sınıflandırılması
- f. Yeni ilaç türlerinin keşfinde

#### Endüstri

- a. Kalite kontrol analizleri
- b. Lojistik
- c. Üretim süreçlerinin optimizasyonu

#### Yüzey Analizi ve Coğrafi Bilgi Sistemleri

- a. Bölgelerin coğrafi özelliklerine göre sınıflandırılması
- b. Kentlerde yerleşim yerlerinin belirlenmesi

- c. Kentlerde suç oranı, zenginlik-yoksulluk, köken belirlemede
- d. Kentlere yerleştirilecek posta kutusu, otomatik para makineleri, otobüs durakları gibi hizmetlerin konumlarının tespitinde

#### Navigasyon Uygulamaları

- a. Yaya navigasyonu uygulamalarında kullanıcı profilinin belirlenmesinde

#### Görüntü Tanıma ve Robot Görüş Sistemleri

- a. Çeşitli algılayıcılar aracılığı ile tespit edilen görüntülerden yola çıkarak engel tanıma, yol tanıma, yüz tanıma, parmak izi tanıma gibi tekniklerde

#### Uzay Bilimleri ve Teknolojisi

- a. Gezegen yüzey şekillerinin ve gezegen yerleşimleri, yeni galaksiler keşfinde
- b. Yıldızların konumlarına göre gruplandırılmasında

#### Meteoroloji ve Atmosfer Bilimleri

- a. Bölgesel iklim, yağış haritalarını oluşturmada
- b. Hava tahminleri, ozon tabakası deliklerinin tespitinde
- c. Çeşitli okyanus hareketlerinin belirlenmesinde

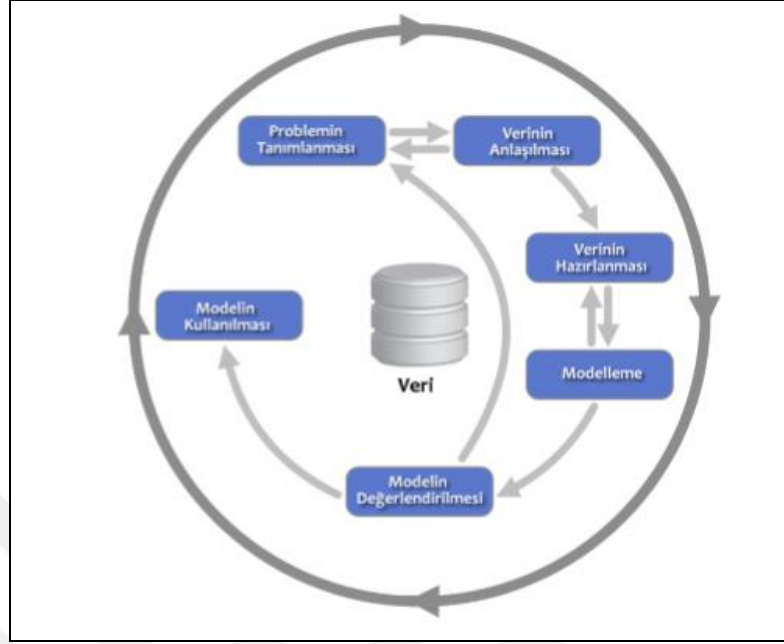
#### Metin Madenciliği

- a. Çok büyük ve anlamsız metin yığınları arasından anlamlı ilişkiler elde etmede

### 2.2.2 Veri Madenciliği Süreci

Veri madenciliği organize yürütülmesi gereken bir süreçtir. Bu süreçlerin sınırlarının ve veri boyutlarının net olmaması, karmaşıklığa yol açarak geri dönülemez hatalara sebep olabilmektedir (Gökmen, 2014). Bu sıkıntıların önüne geçebilmek amacıyla The Cross-Industry Standard Process for Data Mining (CRISP-DM) konsorsiyumu tarafından veri madenciliği süreci belirlenmiştir (<http://crisp-dm.eu/home/about-crisp-dm/>).

Şekil 2.1: Veri madenciliği süreci



Kaynak: (<http://crisp-dm.eu/home/about-crisp-dm/>)

Veri madenciliğinde izlenen adımları (Şekil 2.1) Shearer (2000) şöyle tanımlamıştır:

- Problemin tanımlanması:** Problemin hangi işletme alanında yapılacağını ve çıkan sonuçlar doğrultusundaki analizlerin nasıl değerlendirileceğini tanımlanmasında, veri madenciliği çalışmalarında başarıya götüren en önemli kurallarından biridir.
- Verilerin hazırlanması** süreci, verilerin toplanması, verilere değer biçme, verilerin temizlenmesi, verilerin birleştirilmesi, verilerin bütünden seçimi ve verilerin dönüştürülmesi basamaklarından meydana gelmektedir.
- Modelin kurulması ve değerlendirilmesi:** Belirlenen sorun için uygun modelin seçilebilmesi ve diğer modeller kullanılarak çıkan analizler sonucunda en uygun modelin değerlendirilmesidir. Bundan dolayı verinin hazırlanması ve modeli kurma basamaklarında en iyi analiz sonucunun alındığı düşünülen model bulunana kadar bu işlem devam eden süreçtir.

d. **Modelin kullanılması:** Kullanılan ve geçerli olduđu düşünölen model bir başka uygulamanın alt yapısı olabileceđi gibi bir uygulama olarak da kullanılabilir.

**Modelin izlenmesi:** Kurulan modellerin ürettikleri verilerde çıkabilecek deđişikliklerden dolayı, verilerin düzenlenmesini gerektiren durumlar için kurulan modellerin izlenmesi gerekmektedir (s. 13-23).

### 2.2.3 Veri Madenciliđi Teknikleri ve Yöntemleri

Veri madenciliđi, tanımlayıcı ve tahminleyici olmak üzere iki modele ayrılmaktadır. Tanımlayıcı modellerde mevcut durumun analiz edilmesi ve yorumlanması üzerine odaklanılırken tahmin edici modellerde gelecekteki sonuçların eğilimine odaklanılmaktadır. Veri madenciliđinin temelini, veriler üzerinde yapılan çalışmalarda kullanılan yöntemler oluşturmaktadır. Veriler üzerindeki çalışmalara göre veri madenciliđi üç grupta incelenmektedir: (i) Sınıflama ve Regresyon, (ii) Kümeleme ve (iii) Birliktelik Kuralları. En az iki deđişken arasındaki ilişkileri ölçmek için sınıflama modelleri kullanılırken, tanımlayıcı modeller için de kümeleme ve birliktelik kuralları kullanılmaktadır (Özekes, 2003).

#### 2.2.3.1 Sınıflama

Veri madenciliđinde en çok kullanılan yöntemlerden biridir. Sınıflama kuralları kullanılırken eldeki eğitim kümesine göre ortaya çıkan durumlar için karar verilmesi sağlanır. Genel olarak resim, ses ve örüntü tanıma, hastalık tanıları, dolandırıcılık tespiti, kalite kontrol çalışmaları ve pazarlama konularında bu yöntem kullanılmaktadır (Dursun, 2016). Sınıflamanın şirketler açısından kullanıldıđı alanlar şöyle sıralanmaktadır (Alagöz, 2014, s.17):

- a. Kredi, sigorta işlemlerinde risklerin belirlenmesi
- b. Kredi verilecek kişilerin çeşitli özellikleri ve mali durumlarına göre risk sınıfının belirlenmesi sonucu müşteriye verilecek kredi tutarına karar verilmesi

- c. Yaş, meslek, eğitim, araba markası gibi bilgilere göre farklı kasko prim oranlarının uygulanmasının karar ağaçlarının kurulduğu risk faktörleriyle saptanması
- d. Ürünlerin geleceğe yönelik satış miktar ve fiyatlarının tahmin edilmesi
- e. Yeni bir ürünün satış imkanının incelenmesi (tahmin edilmesi)
- f. Ürünlerin satış miktarı, ciro, mevsimsel değişiklikler, şikâyet, arıza, memnuniyet oranı vb. faktörlerin değerlendirilmesi sonucu en iyi satan ürünün saptanması
- g. Bir ürünün bir bölgede satılma olasılığının incelenmesi
- h. Müşterinin risk sınıfının belirlenmesi ve risk sınıflarına göre müşterilere risk limiti atanması

Ekibin eğitim seviyesi, tecrübesi, yaptığı projeler gibi özelliklerine göre sınıflandırılarak ücretlendirilmesi gibi örnek analizler sınıflama algoritmaları veya karar modelleri yoluyla gerçekleştirilebilir. Sınıflama yöntemlerinde kullanılan algoritmalarından bazıları Bayes, Regresyon, En Yakın Komşu, CART, ID3 algoritmalarıdır.

### **2.2.3.2 Birliktelik Kuralları**

Veri madenciliğinde kullanılan verilerde hangi durumun aynı anda gerçekleştiğini gösteren modeldir. Literatürde sepet analizi olarak geçen bu yöntemi (Dursun, 2016) işletmeler oldukça fazla kullanmaktadır. Örneğin bir markette X ürünü alan bir müşteri Y ürününü de alıyorsa raflarda yan yana sıralama için kullanılabilir. Birliktelik Kuralları yöntemlerinde kullanılan algoritmalarından bazıları APRIORI, OCD, AIS, PDM ve DMA gibi algoritmalarıdır. Birliktelik analizinin şirketler açısından kullanıldığı alanlar şöyle sıralanmaktadır (Alagöz, 2014, s.16):

- a. Müşteri satın alma eğilimlerinin bölgesel, kesim esasında, aylık, günlük ve hatta saatlik periyotlarda tespit edilmesi
- b. Farklı semtlerde bulunan birden fazla satış şubesi için farklı eğilimler tespit edilmesi ve mağaza esasında doğru stok politikaları geliştirilmesi
- c. Ürün kataloglarında birlikte satılan ürünlerin aynı sayfaya yerleştirilerek dikkat çekici hale getirilmesi

- d. Müşteri memnuniyet anketlerinde ve servis kayıtlarında ne tür müşterilerin hangi sıklıkla nelerden şikayet ettikleri ya da memnun olduklarının ortaya çıkarılması
- e. Birbirini izleyen ödeme ve gecikmelerdeki benzerliklerin belirlenebilmesi
- f. Üretimdeki insan kaynakları ile ürün kalitesi arasındaki bağıntıların sağlanması

### 2.2.3.3 Kümeleme

Verilerin içindeki ortak benzerliklerine göre oluşturulan yöntemdir. Kümelenen verilerin arasındaki uzaklık ve yakınlıkların ortaya çıkartılmasında kullanılır. Kümeleme modelinde, hiyerarşik kümelemede en uzak komşu algoritmasıyla, en yakın komşu algoritması durumudur. Kümeleme analizini şirketler açısından kullanıldığı alanlar şöyle sıralanmaktadır (Alagöz, 2014, s.16):

- a. Bayii büyüklüklerine, müşteri sayılarına, cirosuna, rakip ürün pazar derinliği veya nüfuz durumuna göre kümeleme yapılabilmektedir.
- b. Müşteriler, ciroları, sıklıkla aldıkları ürün grubu, ödeme karakteristikleri, yaş ve gelir gruplarına göre bölümlenebilmektedir.
- c. Birden fazla satış noktası, satılan ürün çeşidi, ciro, müşteri profili vb. özelliklere göre tasnif edilebilmektedir.
- d. İnsan kaynakları eğitim seviyesi, bir zaman biriminde yaptıkları iş miktarı, yapılan hata miktarı, oturdukları bölge gibi özellikleri kullanılarak belli gruplarda toplanarak analiz edilebilmektedir.
- e. Ürün ya da hammaddeler için satış/sarf miktarı, ciro, mevsimsel dağılımları gibi özellikleri üzerinden tüketim oranlarına göre tasnifleyerek ABC analizi yapılabilmektedir.

Kümeleme analizi sonuçları ile şu konular araştırılabilmektedir (Alagöz, 2014, s.16.): (i) En iyi müşterilerin kimler olduğu, (ii) Farklı müşteri gruplarının neler olduğu ve bu grupların alışveriş örüntüleri hakkında bilgi, (iii) Müşterilerin satın alma davranışlarının neler olduğu, (iv) Hangi gelir grubunun hangi markayı satın almayı tercih ettiği bilgisi, (v) Müşteri sadakat derecesi ve bunun öngörülebilirliği. Kümeleme analizi yöntemlerinde birçok algoritma geliştirilmiştir. Algoritmalar, yapılacak çalışmanın amacına göre

sınıflandırılmaktadır. Veri setine ve kümeleme oluşturma biçimine göre dört başlık altında incelenmektedir (i) Hiyerarşik yöntemler, (ii) Yoğunluk tabanlı yöntemler, (iii) Grid temelli yöntemler ve (iv) Bölümlemeli yöntemler

Hiyerarşik kümeleme yöntemi, veri nesnelерinin birbirlerine uzaklık ve yakınlık ilişkilerine göre ağaç yapısı içinde gruplamaya çalışılmasıdır. Bu kümeleme metodunda hiyerarşinin aşağıdan yukarıya (birleştirme) ya da yukarıdan aşağıya (ayırma) formuna bağlı olarak birleştirici ya da ayırıcı olmak üzere sınıflandırılırlar (Han ve Kamber, 2006). Hiyerarşik Yöntemlerde en bilinen algoritmalar AGNES ve DIANA algoritmalarıdır. Hiyerarşik yapının içinde bulunması gereken özellikleri şu şekilde sıralamak mümkündür (Levent, 2016): (i) Hiyerarşik bir yöntem kullanıyorsak; birkaç kümeye ayrıştırabilen bir veritabanına sahip olduğumuz anlaşılabilir, (ii) Veri tabanını birkaç kümeye ayrıştırmamızda bize yardımcı olan dendogram adı verilen ağaçsı bir yapı bulunmaktadır, (iii) Dendogram yapısını yapraklardan gövdeye doğru oluşturabildiğimiz gibi gövdeden yapraklara doğru da oluşturabiliriz bununla beraber oluşturduğumuz bu yapıyı istenilen yerden bölerek kümeleri elde etmek mümkündür.

Yoğunluk tabanlı tekniklerde uzaklığa dayalı küme seçiminden ziyade verilerin yoğunluğuna göre bir kümeleme işlemi yapılmaktadır (Bircan ve Çam; 2016, S.89). Yoğunluğa dayalı kümeleme tekniklerinde, kümeler veri tabanındaki daha yüksek yoğunluklu alanlar olarak tanımlanmaktadır (Ester vd., 1996). Küme yoğunluklarının seyrek olduğu alanlarda ise ya gürültülü veriler ya da küme sınırını oluşturan veriler bulunmaktadır (Jiang vd., 2003). Bu algoritmayı büyük veri tabanları ve gürültülü verisi çok olan yapılarda kullanmak oldukça uygundur (Ester vd., 1996; Kriegel vd., 2011). Yoğunluk Tabanlı Yöntemlerde en yaygın kullanılan algoritmalar, DBSCAN ve OPTICS algoritmalarıdır.

Grid temelli yöntemlerde veriler, ızgara yapısı gibi bölümlere ayrılarak kümeleme işlemi her bir ızgara yapısı için ayrı ayrı yapılır (Aydın, 2017). Bu yöntemin en büyük avantajı ise veri boyutundan bağımsız ve hızlı işlem süresinin olmasıdır. Bu yöntemle büyük veri setlerinin işlenmesinde kolaylık olması açısından grid kümelemesinin veritabanını bir kez



taraması yeterli olmaktadır (Levent, 2016). Grid Temelli Yöntemlerde en yaygın kullanılan algoritmalar, STING, Dalga Kümeleme ve CLIQUE algoritmalarıdır.

Bölümlemeli yöntemler, küme yapılmak istenen  $n$  sayıdaki nesneyi önceden belirlenen  $k$  sayıda kümelere bölünme işlemidir. Kümeleme işleminde en çok kullanılan yöntemdir. Bölümlemeli yöntemler, hiyerarşik yöntemlerde olduğu gibi benzerlik matrisi kullanılmadığı için daha büyük verilerde çalışmak için uygundur (Levent, 2016). Küme sayısının önceden belirlenmesi sonuçlardaki güvenilirliği azaltmaktadır. Bu sebeple önceden belirlenen küme sayısının kesin olması dışında en iyi sonuca ulaşmak için algoritma birden fazla tekrarlanır (Dunham, 2003). Bölümlemeli Yöntemlerde, K-Means, PAM, CLARA ve CLARANS algoritmalar kullanılmaktadır. K-Means, en çok kullanılan ve en popüler olan algoritmadır. Bu algoritmada kümelerin en faydalı sonucu vermesi için en uygun çözüme ulaşana kadar kümelerin sürekli yenilediği döngüsel bir algoritmadır (Silahtaroglu, 2013). K-Means algoritmasıyla  $n$  adet verilerden oluşan veri kaynakları ile  $c$  merkez noktası belirlenerek  $k$  adet küme oluşturulmasıdır (Vattani, 2011). Bu algoritma şu şekilde gösterilebilmektedir (Lin ve Wu, 2009): (i)  $k$  adet küme keyfi olarak atanır ve küme merkezleri belirlenir, (ii) Her bir veri belirlenen küme merkezlerine göre yakınlıkları belirlenerek ilgili kümeye dahil edilir, (iii) Her bir küme merkezi tekrardan hesaplanır ve (iv) ii.ve iii. Adımlar küme merkezleri değişmeyecek duruma gelene kadar tekrarlanır. Hata, kareler toplamının minimum olmasıyla merkezin değişmeyecek olmasıyla anlaşılır.

#### **2.2.4 Yapılan Çalışmalar**

Birçok alanda etkin bir şekilde kullanılan veri madenciliği, her geçen yıl kullanım alanları da yaygınlaşmaktadır (Savaş, Topaloğlu ve Yılmaz, 2012). Veri madenciliği, kurumların karar vermelerini kolaylaştıran ve hızlı karar almalarını sağlayan bir araçtır. Alanyazın incelendiğinde farklı disiplinlerde kullanıldığı görülmektedir. Bu çalışmanın kapsamında alanyazında K-Means yönteminin kullanıldığı veri madenciliği çalışmalarından bazıları taranmış ve aşağıda özetlenmiştir.

Çam (2014) çalışmasında Cumhuriyet Üniversitesi Hastanesi'ne 2011 yılında başvurmuş olan hastaların 2006-2011 arasındaki kayıtlar, hasta başvuru davranışlarının belirlenmesi amacıyla incelenmiştir. Bu bağlamda, iş gücü potansiyeli olan 18-65 yaş arasındaki hastaların verilerine veri madenciliği kümeleme analizi yöntemlerinden olan K-Means ve Yoğunluk Tabanlı Kümeleme Algoritması yöntemleri uygulanmıştır. Bu analiz ile hastaların yaşları ile kümelere göre farklılık gösterdiği görülmüş, ayrıca genç hastaların daha farklı teşhisler ile ileri yaşlardaki hastaların ise daha az çeşitli hastalıklarla hastaneye başvurdukları ortaya çıkartılmıştır. Çalışma esas itibariyle hastaneye başvuru yapan bir hasta için tedavi prosedürleri, personel tahsisi, ilaç ve tıbbi malzeme ihtiyaçlarının düzenlenmesi gibi hususlarda yardımcı bir nitelik taşımaktadır.

Kaçmaz Dağ (2015) çalışmasında banka müşteri verilerinden yararlanılarak kredi verilen müşterilerin analizi yapılmıştır, ikinci kısımda ise farklı müşteri grupları için uygun kredi verme politikaları geliştirilmiştir. Uygulamanın ilk kısmında elde edilen sonuçlara göre müşterilerin kredi ödemelerindeki gecikmelere veya aksamalara en çok çocuk sayısı, gelir, medeni hal ve yaş nitelikleri etki etmektedir. Çocuk sayısının artması gelir aralığının azalması doğal olarak kredi ödemelerine açık bir şekilde etkilemektedir. Uygulamanın ikinci kısmında elde edilen 5 müşteri grubu içerisinde orta yaşlı diyebileceğimiz gelir aralığı düşük erkek müşterilerin ödemelerinde gecikme olduğu belirlenmiştir. Aynı zamanda ileri yaşlı, ortalama gelir aralığına sahip kadın müşterilerin ödemelerinde gecikme olduğu belirlenmiştir.

Çerçevik (2016) çalışmasında bir banka için değeri yüksek müşterilerin veri madenciliği tekniklerinden kümeleme analizi yöntemini kullanarak müşteri bölümlenmesi yapılmış ve bu müşteri bölümleri için müşteri sadakatini ve cüzdan payını artırıcı tavsiyeler sunulmuştur. Çalışmanın kümeleri incelendiğinde kaybedilen müşteri grupları, gelişime açık olanlar ve en değerli müşterilerden oluşan kümeler bulunmaktadır. Analizin sonucunda pazarlama personeli için yararlı bilgiler bulunmaktadır. Hangi müşteri bölümleri hangi amaçla iletişime geçilmesi gerektiği konusunda önemli bilgiler ortaya çıkarılmıştır.

Öztürk (2015) çalışmasında 2014-2015 öğretim yılında Anadolu Üniversitesi Açıköğretim Sistemine kayıtlı öğrenenlerin özelliklerini analiz ederek kümelere ayırmak, öğrenen türleri ve özelliklerine uygun stratejilerin belirlenmesine yön göstermektir. Bu çalışmadan elde edilen sonuçlar ile öğrenen odaklı akademik, yönetsel, kurumsal ve destek hizmetlerine yönelik kararların alınması kolaylaşabileceği, bu grupların özelliklerine uygun olarak hizmetlerin farklılaştırılmasına ve tasarlanmasına destek olunabileceği sonucu çıkmıştır.

Levent (2016) çalışmasında havacılık sektöründe karşılaşılan fazla ikram problemi, veri madenciliği yöntemlerinden olan kümeleme analizi ile incelenerek analiz edilmiştir. Kümeleme Analizi Yöntemlerinden KMeans Algoritması ile bilgiler analiz edilerek bu sorunun ortaya çıkarılmasına yönelik açıklayıcı ve önleyici öneriler sunulmuştur.

Üstünel (2018) çalışmasında, K-Means algoritmasından faydalanılarak müşteri satın alma davranışları analizi gerçekleştirmiş ve kümelenen veriler ile hangi müşteri profilinin hangi markayı, hangi ürünü ne zaman ve ne miktarda tercih ettiği belirlemeye çalışmıştır. Yapılan çalışmada amaç, müşteri tercihleri dikkate alınarak firma için hem talep yaratma hem de doğru talebi doğru zamanda karşılama ve sunma gibi avantajların sağlanacağı bir sistem oluşturmak ve bu sistemden yararlanarak firma için veri analizi gerçekleştirmesini sağlamaya çalışılmıştır. Kümeleme sonuçları analiz edildiğinde; toplam satış, ağırlık, zaman ve kategori öz nitelikleri dikkate alınarak gerçekleştirilen kümeleme işleminde, kış sezonunda şarküteri pazarlama kategorisinde yer alan farklı iki markaya ait sade sütlerin en fazla tercih edilen ürünler olduğu görülmüştür. Diğer ürünler ise sırasıyla; kuru gıda kategorisinde bulunan tablet çikolata, meyve suları, sular ve Ayçiçek yağı olarak belirlenmiştir. Diğer yandan, yaz sezonunda en çok tercih edilen ürünler, kuru gıda kategorisinde bulunan kolalı içeceklerdir ve meyve sularıdır. Kış sezonunda en sık tercih edilen ürünler gıda dışı kategorisinde yer alan Türkçe gazeteler ve şarküteri pazarlama kategorisinde yer alan sade sütler iken; diğer yandan, yaz sezonunda en sık tercih edilen ürünlerin gıda dışı kategorisinde yer alan Türkçe gazeteler ve tütün alkollü içecek kategorisinde yer alan sigaralar olduğu görülmüştür. Yaz sezonundaki satışların kış sezonuna göre yaklaşık %8 oranında daha fazla olduğu da tespit edilmiştir.

Aydın (2015) çalışmasında, Elektronik Harp ile elde edilen verilerin, DBSCAN, K-Means ve PAM algoritmaları ile kümeleme analizini yaparak hedef unsurların yoğunlaştığı bölgelerin belirlenmeye çalışılmasını ve buna uygun olarak zaman serisi analiz teknikleri ile hedef unsurların iletişim zaman sıklıklarıyla paralel olarak buldukları koordinatların ortaya çıkarılmasını amaçlamıştır. K-Means ve PAM algoritmalarında düşük küme sayılarında kümelenemeyen veri miktarı fazla olduğu ve küme sayısını yüksek olduğu durumlarda ise veri setinin %93,7 sini kümelenebilmiştir.

Çağlar (2018) çalışmasında, Türkiye'deki 2011, 2012, 2013 yıllarına ait Trafik Kaza istatistik veri setleri üzerinde K-Means yöntemi, K-Medoids yöntemi ve Birleştirici Hiyerarşik Kümeleme (AGNES) yöntemlerini kullanılarak kümeleme analizi yapmış ve bu kümeleme analizlerinin sonuçları kullanılarak çok değişkenli haritalar üretmiştir. AGNES kümeleme analizi sonuçlarıyla hazırlanan çok değişkenli haritaların birbirleriyle oldukça uyumlu olduğu belirtilmiştir. K-Means ve K-Medoids kümeleme analizleri sonuçlarıyla üretilen çok değişkenli haritalarda farklı küme sayıları için kümeleme sonuçları gözlemlenmiş ve her iki algoritmanın da kümeleme performansları benzerlik gösterse de K-Medoids algoritmasında kümelerin birbirinden daha iyi ayrıldığı ortaya çıkmıştır.

### 3. YÖNTEM

Bu bölümde, araştırmanın çözümlerine yer verilmiş olup araştırmanın modeli, örnekleme, verilerin toplanması ve verilerin analizinin nasıl gerçekleştiği açıklanmıştır.

#### 3.1 ARAŞTIRMANIN MODELİ

Bu araştırma kütüphanedeki elektronik veritabanlarının performans durumlarını ortaya çıkarmak amacıyla genel tarama modellerinden ilişkisel tarama türünde bir araştırmadır. İlişkisel tarama modeli, iki ya da daha fazla sayıdaki değişken arasında, birlikte değişim varlığı ve/veya derecesini belirlemeyi amaçlayan bir araştırma modelidir (Frankel ve Wallen, 2009; Karasar, 2005). Model gereği elektronik kaynakların seçimi için birçok değişken kullanılmaktadır. Bu değişkinlerin veritabanı seçiminde etkin bir şekilde kullanılabilmesi için kümeleme yöntemlerinden K-Means algoritması kullanılmıştır. K-Means algoritmasının kullanılmasının sebebi her veritabanının bir kümeye ait olmasını sağlamaktır. Oluşturulan kümeleme yöntemiyle kümelerin kendi aralarında heterojen, elemanları arasında homojen özellik gösterilmesi çalışılmıştır. Veri setinde, elektronik kaynakların seçim ve değerlendirilmesinde kullanılan kriterler kullanılmıştır: kullanıcı istatistikleri, kapsam, içerik, işlevsellik ve maliyet.

#### 3.2 ÖRNEKLEM

Çalışma kapsamında, 2017-2018 eğitim-öğretim döneminde İstanbul'da "Üniversite Kütüphanedir" anlayışını benimsemiş bir Vakıf Üniversitesinin Kütüphanesi'nde kullanılan elektronik veritabanları kullanılmıştır. Kütüphanenin kuruluş çalışmaları Üniversite ile birlikte yapılmıştır. Kütüphanenin bütçesi yıllar içinde artarak elektronik kaynaklar açısından zengin bir koleksiyon geliştirilmeye çalışılmıştır (Erzurum, 2010). Bu çalışmada, kütüphane koleksiyonunda bulunan tüm elektronik veritabanları (54 adet) kullanılmıştır. Elektronik veritabanlarının sağlanması konsorsiyumlar aracılığıyla, doğrudan ve Tübitak Ekual kapsamında aboneliklerle gerçekleştirilmektedir. Elektronik veritabanlarının hepsinin uzaktan ve çoklu erişime imkanı bulunmaktadır. Arşive erişimi bulunan 12 elektronik veritabanı bulunmakla birlikte 42'sinde arşive erişim

bulunmamaktadır. Arşive erişimi bulunan elektronik veritabanlarının 5'inde ek ücret istenilmektedir. Düzenli güncellenen 51 elektronik veritabanı bulunurken 3'ü düzenli olarak güncellenmemektedir. Web sitelerinin modern arayüzlere sahip olması ve kullanıcı dostu olması açısından incelendiğinde 46'sı kullanışlıyken, 8'inin kullanışlılığı iyi değildir. Web site filtreleme özellikleri 46'sında bulunurken 8'inde bu özellik bulunmamaktadır. Elektronik veritabanlarının kapsadığı konular bakımından tek bir konu ve disiplin içermekle birlikte birden fazla konu ve disiplini içeren elektronik veritabanları da bulunmaktadır. Çalışmaya bütün elektronik veritabanlarının dahil edilmesinin sebebi performans durumlarının ortaya çıkartılmasının amaçlanmasıdır. Bu amaç doğrultusunda her bir elektronik veritabanının hangi kümede yer aldığı belirlenmesine çalışılmıştır.

### **3.3 VERİLERİN DÜZENLENMESİ VE TOPLANMASI**

Uygulamada kullanılan veriler, Kütüphanenin 2017-2018 eğitim öğretim döneminde tutulan kullanıcı istatistiklerinden ve kütüphane yönetiminin elektronik veritabanlarının seçiminde kullandığı kriterlerden belirlenmiştir. Bununla birlikte elektronik veritabanlarının kullanıcıların beklentilerinin karşılayıp karşılamadığının ortaya çıkartılması amacıyla kullanıcıların istekte bulunduğu yayınların konu ve disipline göre sayıları kullanılmıştır. Kullanıcıların taleplerinin belirlenmesi için istekte buldukları yayınların konularına göre sayıları Tablo 3.1'de gösterilmektedir. 2017-2018 eğitim döneminde kütüphaneden sözlü olarak, sistem üzerinden eser istekte bulunarak veya mail yoluyla yayın istekleri yapılmıştır. İstenilen yayınların genelde kütüphane yönetimi tarafından elektronik versiyonu tercih edilmekle birlikte en fazla basılı yayınlar sağlanmıştır. Bir yayın istenildiğinde elektronik versiyonu varsa genelde ilk olarak o tercih edilmektedir. Bu sayede birden fazla kullanıcının erişimine imkan tanınmaktadır.

Tablo 3.1'deki verilere göre en fazla yayın talepleri Dil ve Edebiyat, Hukuk, Tarih, Sosyal Bilimler, Tıp, Teknoloji ve Mühendislik alanlarından gelmiştir. Dil ve Edebiyat alanında ise romanların büyük bir payı bulunmaktadır. Romanların tümü basılı kaynak olarak sağlanmıştır. Kullanıcıların en fazla talepte bulunduğu bu alanlardan her konu için tek bir veritabanı olmakla birlikte birden fazla alanları içeren veritabanları da bulunmaktadır.

**Tablo 3.1: 2017-2018 Konularına göre yayın talep sayıları**

Konu	İstek Sayısı
Dil ve Edebiyat	524
Hukuk	380
Tarih	279
Sosyal Bilimler (İşletme – Finans- Sosyoloji-Spor-Tarım-Genel Konular)	252
Tıp	208
Teknoloji ve Mühendislik	186
Mimarlık	89
Felsefe – Psikoloji - Din	75
Siyaset Bilimi	25
Eğitim	15
Müzik	15
Coğrafya	6

### 3.3.1 Kullanıcı İstatistikleri

Elektronik kaynakların istatistikleri o kaynakların ne kadar kullanıldığını bize gösterse de değerlendirme yapılırken kullanıcı istatistiklerinden ziyade abone olunan veritabanının birim başı maliyeti göz önüne alınmaktadır. Mesela abone olunan X firmasına ait bir veritabanına \$10.000 ödenmiş olsun, bu veritabanındaki kullanım sayısı ne kadar artarsa birim başına maliyet o kadar düşmektedir. 2017-2018 eğitim döneminde birim başına maliyet \$0.11 ile \$55.27 arası değişmektedir. Birim başı maliyetler “çok kullanılan veritabanları”, “orta kullanılan veritabanları” ve “az kullanılan veritabanları” olmak üzere üç düzeyde belirlenmiştir. Düzeyler Tablo 3’de gösterilmiştir. Bununla birlikte, veri setinde kullanıcı istatistiklerinin COUNTER uyumlu olup olmama durumuna göre 1 ve 0 olarak değer verilmiştir. Eğer COUNTER uyumlu ise 1, değilse 0 olarak değer verilmiştir.

**Tablo 3.2: Birim Başı Maliyetlerin Düzeyleri**

Etiket	Aralık Değeri	Düzy
Çok kullanılan veritabanları	\$0 ile \$5	1
Orta kullanılan veritabanları	\$5,1 ile \$12	2
Az kullanılan veritabanları	12,1 ve fazlası	3

### 3.3.2 İçerik

Elektronik kaynakların içeriği kütüphane yönetiminin beklentilerini karşılması için şu sorulara cevap vermelidir: (i) Geriye dönük erişim olanağı nedir? (ii) Düzenli güncelleme yapılıyor mudur? (iii) Dergilere erişimde kısıtlamalar var mıdır? Kısaca ambargo var mıdır? (Üçüncü parti olarak adlandırılan veritabanlarında belirli zaman aralıklarında ambargolara rastlanmaktadır.) (Akçal ve Kaya, 2006). Bu başlık altında bulunan nitelikler “Evet” ve “Hayır” durumuna göre belirlenmiştir. Eğer “Evet” ise 1, “Hayır” ise 0 değerini almaktadırlar.

### 3.3.3 İşlevsellik

Elektronik kaynakların sahip olduğu teknolojik özellikler farklılık göstermektedir. Teknolojik özellikler kolay kullanım olması açısından şu sorulara cevap vermelidir: (i) Uzaktan erişime imkan var mıdır? (ii) Tarama arayüzleri kullanışlı mıdır? (iii) Filtreleme özelliklerine sahip midir? (Allison, 2000). Bu başlık altında bulunan nitelikler “Evet” ve “Hayır” durumuna göre belirlenmiştir. Eğer “Evet” ise 1, “Hayır” ise 0 değerini almaktadırlar.

### 3.3.4 Maliyet

Elektronik kaynakların maliyeti erişim yöntemlerine ve aynı anda erişilen kullanıcı sayısına göre belirlenmektedir (Salman, 2012). Elektronik kaynakların maliyetine ilişkin



şu unsurlara bakılmaktadır: (i) Arşive ulaşımında ek bir ücret talep edilmekte midir? (ii) Kullanıcı sınırlaması var mıdır? (iii) Yıllık fiyat artış oranları nedir? (Atılğan ve Yalçın, 2009). Bu başlık altında oluşan niteliklerden; “Geriye dönük erişimin olup olmaması”, “Arşive ulaşımında ek ücret talep edilmesi” ve “Kullanıcı sınırlaması” verilerinin değeri “Evet” ve “Hayır” durumuna göre belirlenmiştir. Eğer “Evet” ise 1, “Hayır” ise 0 değerini almaktadırlar. Yıllık fiyat artış oranları firmalarla yapılan anlaşmada belirtilen artış oranlarını içermektedir (Tablo 3.3). Yıllık fiyat artış oranları “Artış oranı az”, “Artış oranı orta” ve “Artış oranı fazla” olmak üzere üç düzeyde belirlenmiştir. Yıllık fiyat artış oranlarının düzeyleri Tablo 3.3’de verilmiştir.

**Tablo 3.3: Yıllık Fiyat Artış Oranlarının Düzeyleri**

Etiket	Değer Aralığı	Düzye
Artış oranı az	0 ile 3,99	1
Artış oranı orta	4 ile 5,99	2
Artış oranı fazla	6 ve fazlası	3

**Tablo 3.4: Değişkenlerin Düzenlenmeden Önceki Durumu**

Kullanıcı İstatistiği	İçerik	İşlevsellik	Maliyet
Birim Başlı Maliyet	Arşive Ulaşım İmkânı Var mı?	Uzaktan Erişime İmkân Var mı?	Arşive Ulaşımında Ek Ücret Var mıdır?
	Düzenli Güncelleme Yapılıyor mu?	Tarama Arayüzleri Kullanışlı mı?	Kullanıcı Sınırlaması Var mıdır?
Counter Uyumluluğu Var mı?	Ambargo Var mı?	Filtreleme Özelliklerine Sahip Midir?	Yıllık Artış Oranı

Çalışma verilerinin kullanıldığı Kütüphane’de veritabanları seçiminde kullanılan kriterlerden “uzaktan erişim”, yayınlarda “ambargo olması” ve “kullanıcı sınırlaması” nitelikleri analizden çıkartılmıştır. Elektronik veritabanlarının anlaşması yapılırken bu üç özelliğe büyük önem verilmektedir. O yüzden bütün veritabanlarının uzaktan erişime imkanı vardır ve kullanıcı sınırlaması yoktur. Ambargolu yayınlara ise çok az rastlanılmaktadır. Bu yüzden bu çalışmadaki seçim kriterlerinden çıkartılmıştır (Tablo 3.4). Bu bağlamda, değişkenlerin düzenlenmiş son hali Tablo 3.5’de verilmiştir.

**Tablo 3.5: Değişkenlerin Düzenlendikten Sonraki Durumu**

Değişken Adı	Değişken Tipi	Değişken Aralığı	Değişken Kodu
Birim Maliyeti	Sayısal Değer	0,11 - 55,27	Çok Kullanılan Veritabanı = 1 Orta Kullanılan Veritabanı = 2 Az Kullanılan Veritabanı = 3
Counter	Metinsel Değer	-	Evet = 1 Hayır = 0
Arşiv Erişim	Metinsel Değer	-	Evet = 1 Hayır = 0
Düzenli Güncelleme	Metinsel Değer	-	Evet = 1 Hayır = 0
Tarama Arayüzü	Metinsel Değer	-	Evet = 1 Hayır = 0
Filtreleme	Metinsel Değer	-	Evet = 1 Hayır = 0
Arşiv Ek Ücret	Metinsel Değer	-	Evet = 1 Hayır = 0
Artış Oranı	Sayısal Değer	0 - 10,55	Az Artış Oranı = 1 Orta Artış Oranı = 2 Fazla Artış Oranı = 3

### 3.4 VERİLERİN ANALİZİ

Uygulamada kümeleme yöntemi seçilerek K-Means algoritması seçilmiştir. Veri noktaları uzaklık ve benzerlik durumuna göre kümelere ayrılmaktadır. Nesnelere arası uzaklıkları doğru bir şekilde ölçmek algoritmaların özünü oluşturur (Wang ve Sun 2015). Kümeleme algoritması veri tabanını belirli altkümelere ayırırken birimler arası uzaklıklar için Euclidyen, Standardize Euclidyen, Manhattan Mahalanobis, Kareli Euclidyen, Minkowski veya Canberra gibi alternatif ölçü ve yöntemler kullanılmaktadır. (Sarıman, 2011, s.193). Öklid uzaklığının kullanılmasının nedeni aşırı değerlerden kaynaklanabilecek olan olumsuzluklardan etkilenmemesidir (Pasin, 2015). Uzaklık ölçüsü olarak uygulamada öklid uzaklığı seçilmiştir. Uygun k değerlerinin bulunması için k = 2, 3, 4, 5, 6, 7, 8, 9 rastgele değerler verilerek hata değerlerinin belirlenmesi

sağlanmıştır. K değeri seçilirken kümelerin birbiriyle arasında heterojen, kümeler içinde homojen olmasına dikkat edilmiştir (Han, Kamber, Pei 2012). Bundan dolayı, k değeri seçilirken hata kareler toplamının kırılma noktası olan  $k = 4$  değeri kabul edilmiştir (Tablo 3.6). Algoritmanın başarısı uygun k değerinin seçilmesiyle ölçülmektedir (Han, Kamber & Pei 2012); k değeri 5 alındığı takdirde hata kareler toplamı çok değişmeyecektir ve fazladan bir küme daha oluşacaktır. Bu durum, birbirine yakın elemanların farklı kümelerde yer almasına neden olabilir ve araştırmanın amacını olumsuz yönde etkileyebilir. Bu bakımdan optimal küme sayısı 4 seçilerek birbirine en yakın elemanların aynı kümede yer alması sağlanmıştır. Kütüphanede 2017-2018 döneminde kullanılan elektronik veritabanlarının performans durumlarını ortaya çıkartmak amacıyla yapılan bu çalışma, IBM SPSS 25.0 programında gerçekleştirilmiştir.

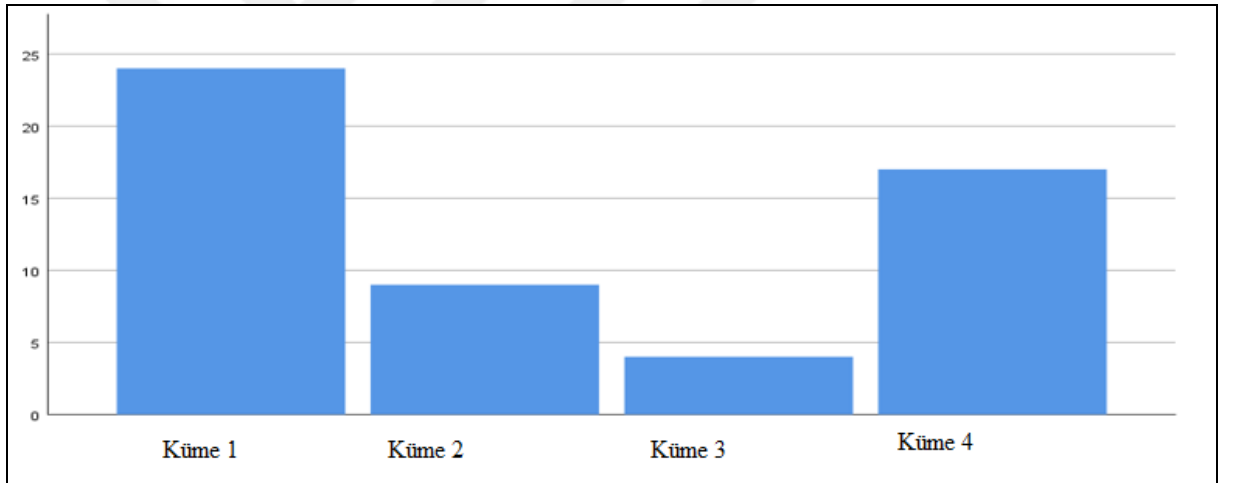
**Tablo 3.6: Küme sayısına göre küme içi hata kareler toplamı**

Küme Sayısı (k)	Küme İçi Hata Kareler Toplamı (SSE)
2	39.363
3	24.698
4	14.482
5	17.802
6	11.062
7	9.6279
8	9.6279
9	8.7727

#### 4. BULGULAR

Bu bölümde, elektronik veritabanlarının seçim kriterleri değişkenlerinden toplanan veriler araştırma sorularına göre incelenmiştir. 2017-2018 döneminde kütüphanede kullanılan elektronik veritabanları, seçim kriterleri değişkenleri kullanılarak K-Means algoritması ile analiz edilmesi sonucunda Şekil 4.1’ de verilen dört küme ortaya çıkmıştır. Bu kümelerin oransal olarak dağılımları birbirinden uzaktır. Elektronik veritabanlarının kümelere göre dağılımları birinci kümede 24, ikinci kümede 9, üçüncü kümede 4 ve dördüncü kümede 17 tane olarak yer almıştır. Tablo 4.1’de elektronik veritabanlarının hangi kümelere olduğunu gösterilmekle birlikte küme merkezlerine olan uzaklığı da gösterilmektedir.

**Şekil 4.1: Kümelerin dağılımı**



**Tablo 4.1: Küme üyeleri**

Elektronik Veritabanları	Küme	Küme Merkezine Uzaklık	Elektronik Veritabanları	Küme	Küme Merkezine Uzaklık	Elektronik Veritabanları	Küme	Küme Merkezine Uzaklık
1	1	,331	19	1	,331	37	4	1,322
2	2	,754	20	4	1,077	38	4	1,693
3	3	,935	21	1	,331	39	1	,331
4	1	,331	22	4	1,049	40	3	,935
5	2	1,006	23	2	,824	41	1	,331
6	1	,881	24	4	1,794	42	1	,781
7	2	,676	25	1	,331	43	1	1,235
8	2	1,160	26	2	,754	44	1	,781
9	2	1,006	27	4	1,794	45	1	1,423
10	4	1,077	28	2	,824	46	4	,716
11	1	1,235	29	4	1,408	47	1	,331
12	4	,716	30	4	,716	48	3	,935
13	1	,781	31	4	,716	49	4	,716
14	1	,331	32	4	,716	50	1	,331
15	2	,754	33	1	,331	51	4	1,622
16	1	,331	34	1	,781	52	1	1,013
17	1	,331	35	1	1,235	53	4	1,622
18	1	,331	36	4	,716	54	3	1,369

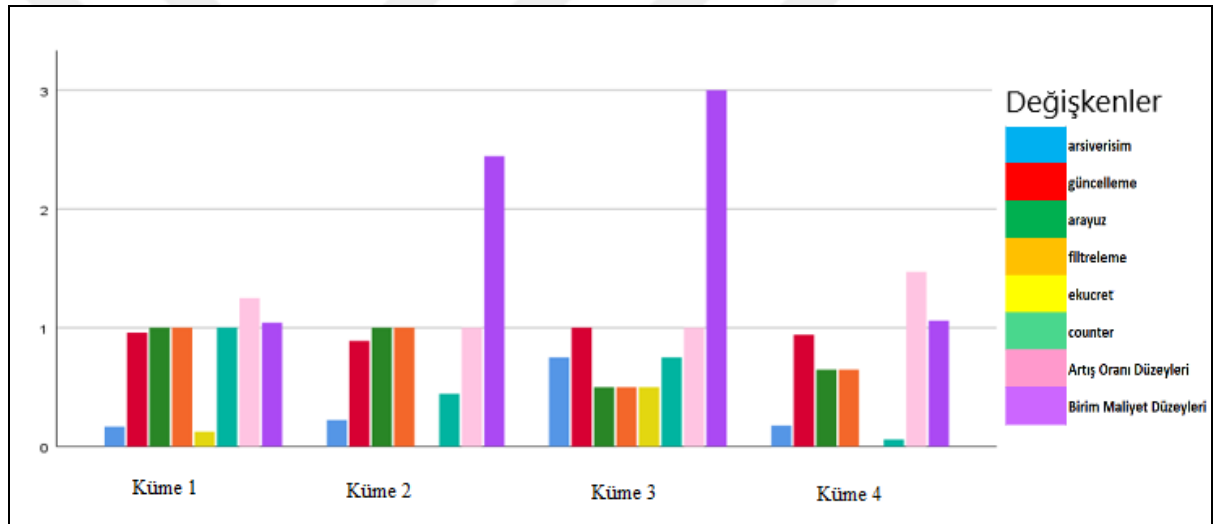
**Tablo 4.2: Küme Merkezlerin Birbirine Uzaklığı**

Küme	1	2	3	4
1	-	1,537	2,223	1,095
2	1,537	-	1,201	1,595
3	2,223	1,201	-	2,257
4	1,095	1,595	2,257	-

Tablo 4.2’de küme merkezlerinin birbirlerine olan uzaklıkları verilmiştir. Uzaklık değeri büyüdükçe benzerlikler azalmaktadır. Üçüncü ve dördüncü kümeler birbirine en uzak iki kümedir. Birbirine en yakın iki küme ise birinci ve dördüncü kümedir.

Kümeleme analizinde önem derecesine göre değişkenler incelendiğinde birim maliyet düzeyleri, artış oranı düzeyleri, arayüz tasarımı, filtreleme özellikleri, counter uyumluluğu, arşive erişim ve arşive erişimde ek ücret talep edilmesi en ayırt edici girdiler olarak ortaya çıkmıştır. Güncelleme durumları ise en düşük ayırt edici özelliğe sahiptir (Şekil 4.2).

**Şekil 4.2: Değişkenlerin Kümelere Dağılımı**



#### 4.1 BİRİNCİ ARAŞTIRMA SORUSUNA GÖRE BULGULAR

Bu kısımda, elektronik veritabanlarının işlevsellik özelliklerinin elektronik veritabanlarının kullanımında etkili olup olmadığının bulguları incelenmiştir. İşlevsellik özellikleri incelenirken web sitelerinin modern arayüzlere ve kullanıcı dostu olup olmamasına dikkat edilmekle birlikte filtreleme özelliklerine ve düzenli güncelleme gibi özellikleri dikkate alınmıştır.

Birinci kümede 24, ikinci kümede ise 17 elektronik veritabanı bulunmaktadır (Şekil 4.2). Birim maliyeti en düşük olan elektronik veritabanları bu kümelere yer almaktadır. Birim

maliyeti \$0 - \$5 arasında bulunduđu için en çok kullanılan elektronik veritabanlarıdır. Burada bulunan elektronik veritabanlarının kapsadıđı konu ve disiplinler kullanıcıların en fazla talep ettiđi konu ve disiplinleri içerdiđi saptanmıştır.

#### **4.2 İKİNCİ ARAŞTIRMA SORUSUNA GÖRE BULGULAR**

Bu kısımda kullanımı az olan elektronik veritabanlarının kullanıcı taleplerindeki yayınların konu ve disiplinlerinde yer alıp almadıđının bulguları incelenmiştir. Kullanımı en az olan elektronik veritabanları ikinci ve üçüncü kümelerde toplandıđı gözlenmiştir (Şekil 4.2). Bu kümelerdeki elektronik veritabanlarının birim maliyet düzeyleri yüksek çıkmıştır. Küme üyelerinin içerdiđi konu ve disiplinler kullanıcılar tarafından en çok talepte bulunan konu ve disiplinleri kapsadıđı ortaya çıkmıştır.

#### **4.3 ÜÇÜNCÜ ARAŞTIRMA SORUSUNA GÖRE BULGULAR**

Bu kısımda, elektronik veritabanlarının tarama arayüzlerinin kullanışlılık, filtreleme özelliklerine sahip olma ve düzenli güncellenme durumları gibi işlevsellik özelliklerinin elektronik veritabanlarının kullanılmasında etkili olup olmadıđının bulguları incelenmiştir.

Web site arayüz kullanışlılığı en fazla olan elektronik veritabanları birinci ve ikinci kümelerde yer almaktadır (Şekil 4.2). En az kullanıma sahip elektronik veritabanlarının oluşturduđu üçüncü kümenin üyelerinin web site arayüz tasarımının kullanışlılığı en az olarak gözlenmiştir. Kullanımı en fazla olan elektronik veritabanlarının toplandıđı birinci kümede web site arayüzlerinin kullanışlı olduđu gözlenmiştir. İkinci kümede yer alan 9 elektronik veritabanının kullanım sayıları az olmasına rağmen web sitesi arayüzlerinin kullanışlı olduđu gözlenmiştir.

En fazla filtreleme özelliđine sahip elektronik veritabanları birinci kümede yer almaktadır. İkinci kümede yer alan elektronik veritabanlarının büyük kısmının filtreleme özelliđi olduđu gözlenmiştir. En az filtreleme özelliđine sahip elektronik veritabanları ise üçüncü kümede olduđu gözlenmiştir (Şekil 4.2).

Elektronik veritabanlarının düzenli güncelleme durumları kümelere orantılı bir şekilde dağılmıştır (Şekil 4.2). Kümelerde yer alan elektronik kaynaklarının çoğunda güncellenme durumlarının olduğu gözlenmiştir.

#### **4.4 DÖRDÜNCÜ ARAŞTIRMA SORUSUNA GÖRE BULGULAR**

Bu kısımda yıllık artış oranı yüksek olan elektronik veritabanlarının kullanım durumlarının bulguları incelenmiştir. Yıllık artış oranları düzeylerinin kümelere dağılımı orantılı bir şekilde dağılmıştır (Şekil 4.2). Artış oranı düzeylerinin ortalamalarına bakıldığında en az artış ikinci ve üçüncü kümede yer aldığı gözlenmiştir. Artış oranı düzeyleri ortalaması birinci kümede 1,25, dördüncü kümede 1,47 olmakla birlikte çok yüksek olmadığı gözlenmiştir.

#### **4.5 BEŞİNCİ ARAŞTIRMA SORUSUNA GÖRE BULGULAR**

Bu kısımda yıllık artış oranı yüksek olan elektronik veritabanlarının kullanıcı taleplerini karşılama göre incelenmiştir. Yıllık artış oranı kümelere dağılımı düşük düzeyde gözlenmiştir (Şekil 4.2). Artış oranı yüksek olan elektronik veritabanları konu ve disiplinleri kullanıcıların talepte bulunduğu yayınların konu ve disiplinlerinde olduğu gözlenmiştir.

#### **4.6 ALTINCI ARAŞTIRMA SORUSUNA GÖRE BULGULAR**

Bu kısımda arşive erişimde ek ücret talep eden elektronik veritabanlarının kullanılma durumları incelenmiştir. Üçüncü kümedeki elektronik veritabanlarının en fazla arşive erişime sahip elektronik veritabanları olduğu gözlenmiştir (Şekil 4.2). Birinci, ikinci ve dördüncü kümelerde orantılı bir dağılım gözlemlenmekle birlikte bu kümelerde yer alan elektronik veritabanlarının çoğunda arşive erişimin olmadığı bulgusu ortaya çıkmıştır. Arşive erişim için ek ücret talebinin en fazla üçüncü kümede olduğu gözlenmiştir. Üçüncü kümedeki elektronik veritabanlarının en az kullanıma sahiptir. Birinci kümede yer alan elektronik veritabanlarının bir kısmında ek ücret talep edilmektedir.



## 5. TARTIŞMA VE ÖNERİLER

### 5.1 BULGULARININ TARTIŞILMASI

Bu bölümde, elektronik veritabanlarının performans durumlarının değerlendirilmesi amacıyla veri madenciliği yöntemi kullanılması sonucunda elde edilen bulgular araştırma soruları ile birlikte alanyazındaki çalışmalarla kıyaslanarak tartışılmıştır.

Araştırmada, birinci ve dördüncü kümelerde en fazla kullanılan elektronik veritabanları bir araya gelmiştir. Bu kümelerdeki elektronik veritabanlarının kapsadığı konu ve disiplinlerin, kullanıcıların yayın taleplerinde bulunan yayınların konu ve disiplinlerini içerdiği gözlemlenmekle birlikte kullanıcı sayıları az olan elektronik veritabanlarının da kullanıcı taleplerinde yer aldığı saptanmıştır. Bu bulgu da alanyazında kütüphanelerin koleksiyon geliştirme politikalarında kullanıcı beklentilerini göz önünde bulundurması gerekliliği (Al ve Al, 2003; Evans, 1995; Kluegel, 1997) durumu ile benzerdir.

Araştırmada, elektronik veritabanlarının web site arayüzlerin kullanılabilirliği en fazla olan kümeler birinci ve ikinci kümelerdir. Kullanıcı sayısı en fazla olan elektronik veritabanları bu kümelerde yer almaktadır. Web site arayüzü kullanılabilirliği en az olan üçüncü kümede ise kullanıcı sayıları en az olan elektronik veritabanları bulunmaktadır. Alanyazında elektronik kaynakların seçim kriterlerinde kullanılabilir arayüzü tasarımlarına sahip olması gerekliliği ile araştırmanın bulguları örtüşmektedir (Kovacs ve Elkordy, 2000; Guenther, 2000; Gürdal ve Can, 2017). Ekranda yer alacak bilgilerin kullanıcıların gereksinimleri ve beklentileriyle uyumlu olmasına dikkat edilmelidir (Kovacs ve Elkordy, 2000). Elektronik kaynağın tasarımıyla ilgili eksiklikler kullanımı olumsuz yönde etkileyecektir. Kullanıcının bakış açısıyla düzenlenmiş bir kaynağın, kullanımı sezgisel olarak anlaşılabilir. Kullanıcı arayüzü kaynağın içeriğinin rahat bir şekilde gözden geçirilmesini sağlayacak biçimde iyi planlanmış olmalıdır. Ekranda görünen yazıların okunaklı olması gerekmektedir (Guenther, 2000). Gürdal ve Can (2017), elektronik veritabanlarının arayüz tasarımlarının kullanıcı dostu olması gerektiği, kolay bir şekilde gezinilmesi ve sezgisel olması (eğitsel dokümanlar, tanıtım ekranları, bağlam duyarlı yardım, kişiselleştirme seçenekleri, arama geçmişinin kaydedilebilmesi vb.) gerektiğini belirtmişlerdir.

Araştırmada, elektronik veritabanları web sitelerinin filtreleme özelliklerine sahip olma durumuna göre incelendiğinde en fazla filtreleme özelliği olan elektronik veritabanlarının 1. kümede toplandığı görülmüştür. Birinci kümenin birim başı maliyetleri düşük olduğu için en fazla kullanılan elektronik veri tabanlarının bulunduğu kümedir. Alanyazında elektronik veritabanlarının web site arayüzlerinin işlevsel olmasının önemine dair bulgular araştırmanın bulguları ile örtüşmektedir (Atılğan ve Yalçın, 2009; Allison, 2000; Al ve Al, 2003). Web sitelerinin tarama sonuçlarında filtreleme gibi özelliklerinin bulunması elektronik veritabanlarının avantajına olmaktadır (Atılğan ve Yalçın, 2009). Elektronik veritabanlarının seçimlerinde kullanılan önemli ölçütlerden birisi de web sitelerinin işlevsel olması bakımından filtreleme özelliklerine sahip olması gerekliliğidir (Allison, 2000; Al ve Al, 2003).

Araştırmada, elektronik veritabanları düzenli güncelleme durumlarına göre incelendiğinde bütün kümelere orantılı bir şekilde dağılmıştır. Düzenli güncelleme neredeyse bütün elektronik veritabanlarında kullanılmaktadır. Alanyazında elektronik veritabanlarının düzenli güncellenmesi gerekliliği ile ilgili çalışmalar bulunmaktadır (Akçal ve Kaya, 2006; Al ve Al, 2003; Polat, 2017). Kütüphanelerin koleksiyon geliştirme politikalarında, kullanıcıların bilgi gereksinimlerini karşılayacak güncel elektronik kaynakların seçimi önemli bir konudur (Akçal ve Kaya, 2006; Al ve Al, 2003; Polat, 2017).

Araştırmada, yıllık artış oranları yüksek olan elektronik veritabanları kullanıcı durumlarına göre incelendiğinde artış oranları kümelere orantılı bir şekilde dağılmıştır. Kümelerdeki yıllık artış oranının ortalamaları düşük düzeyde saptanmıştır. En fazla kullanılan elektronik veri tabanlarının bulunduğu 1. kümede yıllık artış oran ortalaması 1.25, dördüncü kümede 1.47'dir. Artış oranı yüksek olan elektronik veritabanlarının içerdiği konu ve disiplinler kullanıcıların yayın taleplerinde de bulunmaktadır. Benzer şekilde, Gürdal ve Can (2017), elektronik veritabanlarının seçiminde yıllık artış oranlarının kontrolü ve satın alma/abonelik fiyatlama modelindeki değişikliklerin değerlendirilmesi gerektiğini belirtmiştir.

Arşive erişimi bulunan elektronik veritabanlarının ek ücret talep etmesi durumuna göre incelendiğinde üçüncü kümede en çok ek ücret talep edilen elektronik veritabanlarının bulunduğu saptanmıştır. Elektronik veritabanlarını seçim kriterlerinde kütüphaneler, arşive erişim için ek ücret talep edilmesini önemli bir kriter olarak değerlendirmektedir (Atılğan ve Yalçın, 2009).

## 5.2 ÖNERİLER

Bu kısımda, araştırma sonuçlarına yönelik ve ileride yapılabilecek araştırmalara yönelik önerilerde bulunulmuştur.

Araştırma sonuçlarına yönelik öneriler;

- a. Kullanıcı sayıları az olan veritabanlarının kapsadığı konular kullanıcı taleplerinde yer aldığı için bu veritabanlarının tanıtım çalışmaları artırılabilir.
- b. Elektronik veritabanı seçimlerinde kullanıcıların taleplerindeki yayınları konu ve kapsamını içeren elektronik kaynaklar için ayrıntılı çalışma yapılabilir.
- c. Kullanıcı sayıları az olup arşive erişimde ek ücret isteyen elektronik veritabanlarının kullanıcılar için çok önemi yoksa yenilenmesi durdurulabilir.
- d. Yıllık artış oranı fazla olan elektronik veritabanlarının kullanıcı sayıları düşükse yenilenmesi durdurulabilir.
- e. Arayüz tasarımı kullanıcı dostu olmayan elektronik veritabanlarının yenilenmesi durdurulabilir ve yeni elektronik veritabanı seçiminde bu kriter göz önünde tutulabilir.

İlerideki araştırmalara yönelik öneriler;

- a. Elektronik veritabanlarının seçim kriterleri çalışmalarında veri madenciliği algoritmalarından sınıflayıcı yöntemler kullanılarak mevcut elektronik veritabanlarının durumlarına göre yeni elektronik veritabanlarının seçiminin tahmin edilmesine yönelik çalışmalar yapılabilir.

- b. Öğrencilerin akademik başarıları ile kütüphaneyi kullanma ilişkilerini analiz eden çalışmalar yapılabilir.
- c. Elektronik veritabanlarının arayüz tasarımlarında kullanıcıların beklentilerinin neler olduğunu araştıran çalışmalar yapılabilir.
- d. Kütüphane web sitelerinde kullanıcıların gezinme durumları ve web aramaları verilerinden yararlanılarak kullanıcıların beklentilerini içeren elektronik veritabanları seçimi yapılabilir.



## 6. SONUÇ

Kütüphanelerin koleksiyon geliştirme stratejilerinde elektronik kaynakların değerlendirilmesi, seçimi ve yönetilmesi büyük önem taşımaktadır. Elektronik veritabanlarının çoğunun yurtdışı kaynaklı olması, döviz ile alınmasını gerektirmektedir. Kısıtlı olan bütçenin etkin ve verimli kullanılması gerekmektedir. Elektronik veritabanlarının mevcut durumunun analizini gerçekleştirmek amacıyla İstanbul'da bir vakıf üniversitesinin kütüphanesinde yer alan elektronik veritabanlarının performans durumları veri madenciliği kullanılarak analiz edilmiştir.

En çok kullanılan veritabanlarının kapsadığı konu ve disiplinlerin kullanıcıların yayın taleplerindeki konularla benzerlik göstermesi sonucunda kütüphane kapsamında kullanıcıların ilgisine yönelik elektronik veritabanlarının bulunduğu söylenebilir. Bununla birlikte en az kullanılan veritabanlarının kapsadığı konu ve disiplinler, kullanıcı taleplerindeki konuları kapsamaması bu veritabanlarının kullanıcılar tarafından bilinmemesi sonucunu ortaya çıkartabilir. Ayrıca, en az kullanılan veritabanlarının tanıtım faaliyetlerinin yeterli olmadığı düşünülebilir.

Yapılan analiz sonucunda yıllık artış oranlarının bütün kümelere eşit dağılması sonucunda elektronik veritabanlarının her yıl belli bir oranda artacağı durumu ortaya çıkmaktadır. Kullanıcıların yayın taleplerinin kapsadığı konu ve disiplinler, yıllık artış oranlarının yüksek olan elektronik veritabanlarında bulunması bütçenin artırılmasını ve en çok ihtiyaç gerektiren elektronik veritabanlarının alınmasını gerektirmektedir. Elektronik veritabanlarının bazılarında arşive erişim bulunmaktadır. Arşive erişimi bulunan elektronik kaynakların erişiminde ek ücret istenmesi bu elektronik veritabanlarının değerlendirilmesini zorunlu kılmaktadır. Arşive erişimde ek ücret isteyen elektronik veritabanlarının üçüncü kümede toplandığı görülmektedir. Üçüncü kümedeki elektronik kaynakların kullanıcı sayılarının çok az olması yeni dönem seçim değerlendirmesinde önemli bir bilgi olabilir. Elektronik veritabanlarının işlevsel olması bakımından web sitelerinin kullanıcı dostu olması son derece önem teşkil etmektedir. Çalışmanın sonuçlarında kullanıcı dostu olan; arayüz tasarımı kullanışlı ve filtreleme özelliklerine sahip olan elektronik veritabanlarının en çok kullanılan elektronik

veritabanları arasında yer aldığı sonucu ortaya çıkmıştır. En az kullanılan elektronik veritabanlarının da web site arayüz kullanışlılığı ve filtreleme özelliklerine sahip olmaması yeni seçilecek elektronik veritabanların seçim kriterlerinde dikkat edilmesi gereken konular arasında yer alabilir.



## KAYNAKÇA

### *Kitaplar*

Crane, D. (1972). *Invisible colleges; diffusion of knowledge in scientific communities*. Chicago, IL: University of Chicago Press.

Curtis, D., Scheschy, M. V., ve Tarango A. R. (2000). *Developing and managing electronic journal collection*. New York: Neal-Schuman Publishers.

Dunham, M. (2003). *Data Mining Introductory and Advanced Topics*, Prentice Hall, *Pearson Education Inc.*, New Jersey, 2003.

Evans, G.E. (1995). *Developing Library and Information Center Collection*. Englewood: Libraries Unlimited.

Frankel, J. R., & Wallen, N. E. (2009). *Single-subject research. How to design and evaluate research in education. (7th ed.)*. New York, NY: McGraw-Hill.

Han, J. ve Kamber, M., (2006) *Data Mining: Concepts and Techniques*, *San Francisco*.

Han, J., Kamber, M. ve Pei, J. (2012). *Data Mining Concepts and Techniques (3. Baskı)*. Boston, USA, *Elsevier*.

Karasar, N. (2005). *Bilimsel araştırma yöntemi (17. Baskı)*. *Ankara: Nobel yayın dağıtım*, 81- 83.

Konya, Ü. (2004). *Elektronik Kitaplar. Aysel Yontar Armağanı, Yay. Haz.: Bekir*.

Silahtaroglu, G. (2008). *Kavram ve Algoritmalarıyla Temel Veri Madenciliği*. İstanbul, Türkiye, *Papatya Yayıncılık*.

Silahtaroglu, G. (2013).Veri Madenciliđi Kavram ve Algoritmaları. 2.Baskı. *İstanbul: Papatya Yayıncılık Eđitim A.Ş.*

Tsiptsis, K. (2009). Data Mining Techniques in CRM. A John Wiley and Sons, Ltd., Publication.





### ***Sürekli Yayınlar***

Agee, J. (2005). Collection evaluation: A foundation for collection development.

*Collection Building*, 24 (3), 92-95.

Akçal, İ. ve Kaya, S. (2006). Bilgi merkezleri için veritabanı seçim kriterleri ve kullanıcı eğitimi, *Bilgi Dünyası*, 6 (1), 78-88.

Akpınar, H., (2000), “Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği”. İ.Ü. *İşletme Fakültesi Dergisi*, Cilt 29, S 1, 1-22

Umut, A. L., & Pınar, A. L. (2003). Elektronik bilgi kaynaklarının seçimi. *Bilgi Dünyası*, 4(1), 1-14.

Alagöz, A., S. Öge ve M. Ortakarpuz (2014). Bir Kurumsal Zeka Teknolojisi Olarak Veri Madenciliği ile Muhasebe Bilgi Sistemi İlişkisi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*. 2014. Vol.1. No.21, s.16.

Allison, S., Beth M. ve Swanson S. (2000). Database selection: One size does not fit all, *College & Research Libraries*, 61 (1), 56-63

Ashcroft, L. ve Langdon C. (1997). Electronic journals and university library collections. *Collection Building*, 18 (3), 105-114

Atılğan, D., & Yalçın, Y. (2009). Elektronik kaynakların seçimi ve değerlendirilmesi. *Türk Kütüphaneciliği*, 23(4), 769-802.

Bircan, H., & Çam, S. (2016). Veri Madenciliğinde Kümeleme Analizi ve Sağlık Sektöründe Bir Uygulaması. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 17(2), 85-96.

Chandra, H. (2002). E-Resources management with specific reference to E-reference sources: initiatives and issues. Proceedings 21st Annual Convention and

Conference of Society for Information Science (SIS), Roorkee, (India). 2002; 180-190.

Cole, L. (2004). Back to basics: what is the e-journal?. *The Serials Librarian*, 47(1-2), 77-87.

Cotter, G. ve Koehler, W. (2002). Electronic collection management and electronic information services. Paper presented at the RTO IMC Lecture Series on “Electronic Information Management for PFP Nations”, *Vilnius, Lithuania*, 24-26 September 2002

Çakın, İ. (2000). Üniversitelerimizin bilgiye erişim sorunları ile Ulusal Akademik Ağ ve Bilgi Merkezi. *Bilgi Dünyası*, 1(1), 26-4

ÇALIK, T., & Sezgin, F. (2005). Küreselleşme, bilgi toplumu ve eğitim. *Kastamonu Eğitim Dergisi*, 13(1), 55-66.

Davis, T.L. (1997). The evaluation of selection activities for electronic resources. *Library Trends*, 45(3): 391-403.

Doğan, Ş., ve Türkoğlu, Ğ., 2007. Hypothyroidi and Hyperthyroidi Detection from Thyroid Hormone Parameters by Using Decision Trees. *Doğu Anadolu Bölgesi Araştırmaları Dergisi*. 5 (2), ss. 163-16.

Erzurum, K. (2010). Bir duygu yaratmak:“üniversite kütüphanedir”. *Türk Kütüphaneciliği*, 24(3), 533-539.

Guenther, K. (2000). Making smart licensing decision. *Computer in Libraries*, 20(6), 58-60.

Gürdal, G., ve Kafalı Can, G. (2018). E-Kaynakların Seçimi ve Yönetimi: *İYTE Kütüphanesi*.

Hand, D. J. (1998). Data mining: statistics and more?. *The American Statistician*, 52(2), 112- 118.

Jacobs, P., 1999. Data Mining. *Harvard Management Update*. 4 (10), s. 8.

Jiang, D., Pei, J. ve Zhang A. (2003). *DHC: A Density-Based Hierarchical Clustering Method for Time Series Gene Expression Data*. Third IEEE Symposium on BioInformatics and BioEngineering, Washington, DC, USA, 12 March.

Kaur, M., & Walia, P. K. (2016). Collection development of electronic resources in management libraries of India. *Collection Building*, 35(3), 73-83.

Kitler, R. ve Wang, W., 1998. The Emerging Role of Data Mining. *Solid State Technology*. 42 (11), s. 45.

Kovacs, D. K., & Elkordy, A. (2000). Collection development in cyberspace: building an electronic library collection. *Library Hi Tech*, 18(4), 335-361.

Kluegel, K. (1997). Redesigning our future. *Reference Quarterly*, 36 (3), 330-334.

Özarslan, Y., Balaban-Sali, J., & Demiray, U. (2013). TOJDE: Electronic Publishing and a Review of Ten Years' Experience in Turkey. *Turkish Online Journal of Distance Education*, 13(3), 316-346.

Özekes, S., (2003), "Data Mining Models and Application Areas", *İstanbul Commerce University Journal of Science*, No.3, 65-82.

Rao, M. K. (2001). Scholarly communication and electronic journals: issues and prospect for the academic and research libraries. *Library Review*, 50 (4): 169-175.

Rao, S. S. (2003), "Electronic books: A review and evaluation", *Library Hi Tech*, 21(1), 85-93.

Rao, S. S. (2005). Electronic books: their integration into library and information centers. *The Electronic Library*, 23(1), 116-140

Rukancı, F. ve Anameriç, H. (2003). E-Kitap teknolojisi ve kullanımı. *Türk Kütüphaneciliği*, 17(2), 147-166.

Savaş, S., Topaloğlu, N. & Yılmaz, M., 2012. Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen bilimleri Dergisi*. 11 (21), ss. 1-23.

Shearer, C., 2000. The New Blueprint for Data Mining. *Journal of Data Warehousing*. 5 (4), ss. 13-23.

Small, D. (2000). E-books in libraries: Some early experiences and reactions. *Searcher* 8(9), 63-65.

Sarıman, G.(2011). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma KMeans ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 15-3(2011),192-202, s.193

Tenopir, C., & King, D. W. (1998). Designing the future of electronic journals with lessons learned from the past: economic and use patterns of scientific journals. In *Proceedings. Socioeconomic Dimensions of Electronic Publishing Workshop*(pp. 11-17). IEEE.

Tonta, Y. (1997). Elektronik Yayıncılık, Bilimsel İletişim ve Kütüphaneler. *Türk Kütüphaneciliği*, 11 (4), 305-314

Tonta, Y. (2000). Elektronik yayıncılıkta son gelişmeler. *Bilgi Dünyası*, 1(1), 89-132.

Vattani, A. (2011). K-Means Requires Exponential Many Iterations Even in the Plane.  
*Discrete Computer Genom.* 45, 596-616

Waghmare, B. ve Salunkhe, S. (2013). Usage of e-resource in academic institution.  
*International Journal of Innovative Research and Development*, 2(12), (Special Issue), 262-264.

Wang, F., ve Sun, J., 2015. Survey on distance metric learning and dimensionality reduction in data mining. *Data Mining & Knowledge Discovery*. 29 (2), pp. 534564.

Weller, Ann C., “Qualitative and Quantitative Measures of Indexed Health Sciences Electronic Journals”, *The Journal of American Medical Association*, 287(21), 2002: 2875-2876.

## ***Diğer Yayınlar***

Aydın, Ö. (2017) *Elektronik Harp ile Toplanan Verilerin Veri Madenciliği Yöntemleri ile Analiz Edilmesi*, Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2017.

Besimoğlu, C. (2007). *Akademisyenlerin elektronik dergi kullanımında disiplinler arasındaki fark* (Doctoral dissertation, Hacettepe University).

Çağlar, Burak. *Mekansal Verilerin Kümeleme Analizi ile Değerlendirilmesi*, Necmettin Erbakan Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2018.

Çağiltay, K. (2015). “Türkiye’de internet: dünü, bugünü, yarını – 1996 (bölüm-4)”. [Çevrimiçi] <http://blog.metu.edu.tr/kursat/2015/02/02/turkiyede-internetdunu-bugunu-yarini-1996-bolum-4/> adresinden 07.07.2019 tarihinde erişildi.

Çam, Selim. *Veri Madenciliğinde Kümeleme Analizi ve Sağlık Sektöründe Bir Uygulama*, Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 2014.

Çerçevik, Fikriye Çiğdem. *Müşteri İlişkileri Yönetiminde İş Zekası ve Bankacılıkta Bir Uygulama*, Eskişehir Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 2016.

Dağ, Burcu Kaçmaz. *Veri Madenciliği ve Hizmet Sektöründe Bir Uygulama*, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2015.

Dursun, Gökçe Ezer. *İç Kontrol Sisteminde Bir Araç Olarak Veri Madenciliğinin Kullanılmasının İşletmenin Bilançosuna Etkisi*, Marmara Üniversitesi, Sosyal Bilimler Üniversitesi, Yüksek Lisans Tezi, 2016.

Ebook usage in U.S. academic libraries, sixth annual survey (2016). New York: Library Journal, 27.07.2019 tarihinde

[https://s3.amazonaws.com/WebVault/research/LJ\\_2016\\_EbookUsage\\_AcademicLibraries.pdf](https://s3.amazonaws.com/WebVault/research/LJ_2016_EbookUsage_AcademicLibraries.pdf) adresinden erişildi.

Ester, M., Kriegel, H. P., Sander, J. ve Xu, X. (1996). *A Density-Based Algorithm for Discovering Cluster in Large Spatial Databases with Noise*. 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, 2-4 August.

Gökmen, Şenol. *Müşteri İlişkileri Yönetiminde Bir Araç Olarak Veri Madenciliği ve Parekende Sektöründe Bir Uygulama*, Yıldız Teknik Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 2014.

Grimes, W. (2011). Michael Hart, a Pioneer of E-Books, Dies at 64. *The New York Times*, 8, 2011. Viittu 10.08.2019  
<http://www.nytimes.com/2011/09/09/business/michael-hart-a-pioneer-of-ebooks-dies-at-64.html>.

Johnson, P. (2013). *Developing and managing electronic collections*. Chicago: American Library Association. 07.07.2019 tarihinde  
<https://ebookcentral.proquest.com/lib/karatekin-ebooks/detail.action?docID=1636018> adresinden erişildi.

Kalikov, A., (2006). *Veri madenciliği ve bir e-ticaret uygulaması*, Ankara Gazi Üniversitesi FBE, Yüksek Lisans Tezi.

Kıpçak, H. (2018). *Türkiye’de bilginin toplumsallaşmasında elektronik yayınların rolü*, Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 2018.

Levent, Eyyüp Burak. *Veri Madenciliği ve Hava Sektöründe Bir Uygulama*, Yıldız Teknik Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 2016.

Lin, D.,ve Wu, X. (2009). Phrase clustering for discriminative learning. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2* (pp. 1030-1038). Association for Computational Linguistics.

Millidere, D. (2013). *Sosyal Bilimler Dallarında 2004-2009 yılları arasında etki faktörü değeri en yüksek olan süreli yayınların Ankara'daki üniversite kütüphanelerinde bulunabilirliği/erişebilirliği*, Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 2013.

Öztürk, Aylin. *Açık ve Uzaktan Öğrenme Sistemlerinde Kümeleme Analizi Yöntemiyle Öğrenen Grupların Belirlenmesi*, Eskişehir Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 2015.

Pasin, Ö. (2015). *Sağlık alanında yapılan araştırmalarda kümeleme algoritmalarının kullanımı: bir uygulama*, Düzce Üniversitesi, Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2015.

Polat, C. (2017). Elektronik Kaynak Yönetimi. In *Bilişim Teknolojilerinin Bilgi Merkezlerine ve Hizmetlerine Etkileri* (pp. 227-252). HiperLink.

Project Gutenberg. (2017). About. 27.07.19 tarihinde <https://www.gutenberg.org/wiki/Gutenberg>About> adresinden erişildi.

Rowland, F. (1997). Print journals: fit for the future?. *Ariadne*, (7).

Salman, P. (2002). *Elektronik bilgi kaynaklarının seçimi ve değerlendirilmesi: Hacettepe Üniversitesi kütüphaneleri örneği*. Yayımlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi, 2002.



Anameriç, H., & Rukancı, F. (2004). Bilgi toplumu ve toplumun bilgilenmesinde kütüphanelerin rolü. S. Arslantekin ve F. Özdemirci (Yay. Haz.). *Kütüphaneciliğin Destanı Uluslararası Sempozyumu Bildiriler* 21-24 Ekim 2004, 330-338.

Ulrichwebs Global Serials Directory. (2017). 27.07.19  
<http://www.ulrichsweb.com/ulrichsweb/faqs.asp> adresinden erişildi.

Ünal, Y. (2008). *Dergi kullanım verilerinin bibliyometrik analizi ve koleksiyon yönetiminde kullanımı*, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Doktora Tezi, 2008.

Üstünel, Merve. *K – Ortalamalar Algoritmasına Dayalı Kümeleme Analizi Sistemi ve Perakendecilik Sektöründe Uygulanması*, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2018.