

T.C.
ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

VERİ MADENCİLİĞİ TEKNİKLERİ UYGULAYARAK
KÜTÜPHANE OTOMASYONUNDAN BİLGİ KEŞFİ ELDE
EDİLMESİ

İshak FIRAT

Danışman: Dr. Öğr. Üyesi Fulya ASLAY

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI

ERZİNCAN

2019

Her Hakkı Saklıdır.

Kabul ve Onay Sayfası

Dr. Öğr. Üyesi Fulya ASLAY danışmanlığında, İshak FIRAT tarafından hazırlanan bu çalışma 19/08/2019 tarihinde aşağıdaki jüri tarafından Elektrik Elektronik Mühendisliği Anabilim Dalı'nda Yüksek Lisans/Doktora Tezi olarak oybirliği/oy çokluğu (.../...) ile kabul edilmiştir.

Başkan : Dr. Öğr. Üyesi Fulya ASLAY

İmza: 

Üye : Dr. Öğr. Üyesi Kamil ORMAN

İmza: 

Üye : Doç. Dr. Handan ÇAM

İmza: 

Yukarıdaki sonuç Enstitü Yönetim Kurulunun 11 / 09 / 2019 tarih ve 32/16..... sayılı kararı ile onaylanmıştır.



Prof. Dr. Mustafa Fatih ERTUGAY
Enstitü Müdürü

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildirişlerin, şekil ve tabloların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

Bilimsel Etięe Uygunluk Sayfası

“Veri Madencilięi Teknikleri Uygulayarak Kütüphane Otomasyonundan Bilgi Keşfi Elde Edilmesi” isimli “Yüksek Lisans” tezim tarafımda intihal tespit programı ile incelenmiştir. Buna göre tezimde bilimsel etik ihlali ve intihal olarak nitelendirilebilecek herhangi bir durum olmadığını taahhüt ederim.

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir biçimde elde edildiğini; aynı zamanda bu kural ve davranışların gerektirdiğı gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi beyan ederim. 19/08/2019



İshak FIRAT

ÖZET

Yüksek Lisans Tezi

VERİ MADENCİLİĞİ TEKNİKLERİ UYGULAYARAK KÜTÜPHANE OTOMASYONUNDAN BİLGİ KEŞFİ ELDE EDİLMESİ

İshak FIRAT

Erzincan Binali Yıldırım Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik Elektronik Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Fulya ASLAY

Bilişim sistemlerinin oldukça geliştiği ve dijital çağ olarak adlandırılan günümüzde, kayıt altına alınan veri miktarının sürekli artması, büyük veri kavramını ortaya çıkarmıştır. Özellikle yönetsel anlamda karar vericiler için oldukça önemli olan stratejik bilgilerin elde edilmesi ise ancak bu büyük verilerin doğru tekniklerle işlenmesi ile mümkün olmaktadır. Veri yığınları içerisinde anlamlı ve yararlı bilgiye ulaşmak için son yıllarda veri madenciliği tekniklerinden sıklıkla yararlanılmaktadır. Bu çalışmada, veri odaklı olarak hizmet sunan üniversite kütüphane bilgi sistemlerinin yöneticilerine, kullanıcı profillerini oluşturabilmeleri ve bu doğrultuda gelecekte alacakları yatırım kararlarını da etkin bir şekilde verebilmeleri amaçlanmıştır. Çalışma kapsamında Erzincan Binali Yıldırım Üniversitesi Merkez Kütüphanesi'ne ait bilgi sistemi üzerinde bir uygulama gerçekleştirilmiştir. Veri madenciliğinin tanımlayıcı modellerinden olan kümeleme analizi ve birliktelik kuralları aracılığıyla; kütüphane envanteri, üye profili, yayın ödünç alma alışkanlıkları ile birlikte ödünç alınan yayınlar gibi bilgiler elde edilerek, çalışma sonunda bu bilgiler doğrultusunda değerlendirmelerde bulunulmuştur.

2019, 70 Sayfa

Anahtar Kelimeler: Bibliomining, Birliktelik Analizi, Kümeleme Analizi, Kütüphane Bilgi Sistemi, Veri madenciliği

ABSTRACT

Master Thesis

KNOWLEDGE DISCOVERY IN LIBRARY DATABASE USING DATA MINING TECHNIQUES

İshak FIRAT

Erzincan Binali Yıldırım University
Institute of Natural and Applied Sciences
Department of Electrical and Electronics Engineering

Supervisor: Asist. Prof. Dr. Fulya ASLAY

A continuous increase in the amount of recorded data has introduced a new thing “Big Data” nowadays, which is named as “the IT Era” when information systems are quite developed. Obtaining highly important strategic informations for especially managerial decision makers is only possible when big data is processed correctly. Data mining techniques are often used to reach meaningful and useful information from the data stacks. The purpose of this study is that, information systems managers of university libraries providing data-oriented services can create user profiles and accordingly, they will be able to effectively make future investment decisions. In this study, an application has been made of the information system of Erzincan Binali Yıldırım University central library. Through clustering and association rule analyses which are among descriptive models, data such as library inventory member profile, publication borrowing habits and borrowed publications were obtained and an evaluation was made according to these data at the end of the study.

2019, 70 Page

Anahtar Kelimeler: Bibliomining, Association Analyze, Clustering Analyze, Library Information System, Data Mining

TEŐEKKÜR

Yüksek Lisans Tezi olarak sunduđum bu alıőma Erzincan Üniversitesi Fen Bilimleri Enstitüsü Mühendislik Fakültesi Elektrik Elektronik Mühendisliđi Bölümünde yapılmıőtır.

alıőmanın gerçekleştirilmesinde, deđerli bilgileri ile bana yol gösteren, kullandıđı her kelimenin önemini asla unutmayacađım, kıymetli danıőman hocam Sayın Dr. Öğr. Üyesi Fulya ASLAY' a teşekkürlerimi sunarım.

Beni bu günlere nice emek ve zorlukla yetiőtirerek getiren ve benden hiçbir zaman desteđini esirgemeyen annem ve babama, hayattaki en büyük őansım olan eőime ve kızlarım Elif ile Melike'ye sonsuz teşekkürler.

İőhak FIRAT

Ađustos 2019

İÇİNDEKİLER

	Sayfa
ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
İÇİNDEKİLER	iv
ŞEKİLLER LİSTESİ	vi
TABLolar LİSTESİ.....	vii
SİMGELER ve KISALTMALAR	viii
1. GİRİŞ	1
2. KURAMSAL TEMELLER.....	3
2.1. Literatürdeki Kütüphane Veri Madenciliği Çalışmaları	3
2.2. Veri Ve Veri Madenciliği Kavramı.....	5
2.3. Verilerin Gizliliği Ve Kişisel Verilerin Korunması Kanunu	9
2.4. Veri Tabanı Ve Veri Tabanı Yönetim Sistemleri	10
2.4.1. Nesne yönlü veri tabanları	12
2.4.2. Nesne ilişkili veri tabanları	13
2.4.3. Uzamsal veri tabanları.....	13
2.4.4. Zaman serisi veri tabanları	13
2.4.5. Metin ve multimedya veri tabanları	14
2.4.6. Heterojen veri tabanları.....	14
2.4.7. World wide web (Dünya Çapında Ağ).....	14
2.5. Veri Ambarı	15
2.6. Verilerin Hazırlanma Süreci	16
2.6.1. Veri temizleme	16
2.6.2. Veri uyarlama.....	17
2.6.3. Veri dönüşümü	17
2.6.4. Veri daraltma.....	18
2.7. Üniversite Kütüphanelerinde Veri Madenciliği (Bibliomining).....	18
2.7.1. Kütüphanelerde veri madenciliği aşamaları.....	19
2.8. Veri Madenciliği Metotları Ve Algoritmalar	21
2.8.1. Regresyon ve sınıflandırma modelleri	22
2.8.2. Kümeleme	24

2.8.3. Birliktelik kuralları ve sıralı örüntü analizi	27
2.8.3.1. Apriori algoritması	29
3. MATERYAL VE YÖNTEM.....	32
3.1. Araştırmanın Amacı Ve Verinin Elde Edilmesi.....	32
3.2. Verinin Temizlenmesi Ve Dönüştürülmesi.....	32
3.3. SPSS Modeller Uygulaması İle Modelleme	43
3.3.1. Kümeleme analizi ile modelleme.....	44
3.3.2. Birliktelik analizi ile modelleme.....	48
4. ARAŞTIRMA BULGULARI.....	56
4.1. Kümeleme Analizi Sonuçlarının Yorumlanması	56
4.2. Birliktelik Analizi Sonuçlarının Yorumlanması	60
5. SONUÇ VE ÖNERİLER.....	64
KAYNAKLAR	67
ÖZGEÇMİŞ	71

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 2.1. Bilgi Keşfi Sürecinde Veri Hazırlık Aşamaları.....	8
Şekil 2.2. Veri Madenciliği Metotları.....	22
Şekil 2.3. Kümeleme	25
Şekil 2.4. Kümeleme Yöntemleri	26
Şekil 2.5. Hiyerarşik Kümeleme.....	26
Şekil 2.6. Alışveriş Sepeti Analizi.....	28
Şekil 2.7. Apriori Algoritması	30
Şekil 3.1. Kütüphanede bulunan yayınların konularına göre dağılımları	38
Şekil 3.2. Kütüphanede bulunan yayınların yayım tarihlerine göre dağılımları	39
Şekil 3.3. Kütüphanede bulunan yayınların ödünç alınma sayıları.....	40
Şekil 3.4. Üyelerin unvanlarına göre ödünç alma grafiği	41
Şekil 3.5. Üyelerin cinsiyetlerine göre ödünç alma grafiği.....	42
Şekil 3.6. Ödünç alınan gün sayısı	43
Şekil 3.7. Kümeleme analizi için akış şeması	44
Şekil 3.8. Üyelere ait verilerin genel görünümü	45
Şekil 3.9. Ödünç alma sayılarına göre üyelerin kümelere dağılımı	46
Şekil 3.10. Ödünç alma sayılarının histogram grafikleri	47
Şekil 3.11. Ödünç alma işlemlerine ait verilerin genel görünümü.....	47
Şekil 3.12. Ödünç alma sürelerinin kümelere dağılımı.....	48
Şekil 3.13. Birliktelik analizi için oluşturulan akış şeması	49
Şekil 3.14. Üyelerin fakülteleri ile ödünç alınan yayın kodları arasındaki ilişki.....	50
Şekil 3.15. Üyelerin unvanları ile ödünç alınan yayın kodları arasındaki ilişki	51
Şekil 3.16. Üyelerin cinsiyetleri ile ödünç alınan yayın kodları arasındaki ilişki.....	52
Şekil 3.17. Üyelerin unvanları ile ödünç alma süreleri arasındaki ilişki	52
Şekil 3.18. Akademisyen üyelerin fakülteleri ile cinsiyetleri arasındaki ilişki.....	53
Şekil 3.19. Akademisyenlerin cinsiyetleri ile ödünç alma süreleri arasındaki ilişki	54
Şekil 3.20. Güven değerine göre sıralanmış birliktelik kuralları	55

TABLolar LİSTESİ

	Sayfa
Tablo 3.1. Kütüphaneye ait kayıtları içeren dosyalar	33
Tablo 3.2. Dewey onlu sınıflama sistemi.....	35
Tablo 3.3. Üyelerin kayıtlı oldukları birimler.....	36
Tablo 3.4. Önişlemden geçtikten sonra dosyaların durumu.....	37
Tablo 3.5. Üyelerin ünvanlarına göre ödünç alma sayıları	41
Tablo 3.6. Üyelerin cinsiyetlerine göre ödünç alma sayıları	42
Tablo 4.1. Üyelerin kümelere dağılımı.....	57
Tablo 4.2. Ödünç alma işlemlerinin kümelere dağılımı.....	59
Tablo 4.3. Birliktelik kurallarının farklı konu başlıklarında sıralanması.....	61

SİMGELER ve KISALTMALAR

Simgeler

χ^2	Ki-kare
<	Küçüktür
>	Büyüktür
∞	Sonsuzluk

Kısaltmalar

BC	Bliss Bibliographic Classification
CC	Colon Classification
CLC	Chinese Library Classification
DDC	Dewey Decimal Classification
DNA	Deoxyribonucleic Acid
e-CRM	Electronic Customer Relationship Management
IBM	International Business Machines
KVKK	Kişisel Verilerin Korunması Kanunu
LCC	Library of Congress Classification
MS	Microsoft
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
VTYS	Veri Tabanı Yönetim Sistemleri
WEB	World Wide Web

1. GİRİŞ

Endüstri 4.0 ile birlikte üretimden yönetime birçok işleyişin değiştiği ve geliştiği günümüz işletmelerinde, tüm süreçlerde kayıt altına alınan verilerin işlenmesi için yapay zeka, makine öğrenmesi ve veri madenciliği gibi tekniklerden yararlanılmaktadır. Bu teknikler içerisinde adı son yıllarda oldukça fazla duyulan ve disiplinler arası bir disiplin olarak ortaya çıkan veri madenciliğinin geçmişi, 1700'lü yıllardaki Bayes Teoremi ile 1800'lü yıllardaki Regresyon analizi gibi istatistiksel yöntemlere dayanmaktadır. Veri madenciliği; istatistik başta olmak üzere veri bilimi, veri tabanı teorisi ve makine öğrenimi gibi pek çok teknik ile de ilişkilidir. Özellikle 1990'lı yılların sonunda, işletmelerin karlılığına katkı sağlaması ile birlikte veri madenciliği, birçok işletme tarafından kullanılan teknikler bütününe dönüşmüştür. İnternet çağı ile birlikte her geçen gün daha da artan büyük veri kümelerinin farklı bakış açıları ile ele alınarak analiz edilmesi, elde edilen bilgilerin amaca uygun olarak yararlı hale getirilip özetler halinde sunulması ve böylece veri içerisindeki gizli örüntülerin ortaya çıkarılması, veri madenciliği ile mümkün olabilmektedir. Günümüzde veri madenciliği uygulamaları iş dünyasında, kamu kuruluşlarında, üniversitelerde ve bilimin her alanında kullanılabilmektedir. Tüm bilgi teknolojileri servislerinden elde edilen ve artık nesnelerin interneti ile bilginin, nesnelerin ve insanların aynı ortamda birbirine bağlanabildiği teknoloji ile birlikte büyük veri yığınlarının depolanması ve bu yığınlar içerisinde gerçek zamanlı karar verme stratejilerinin oluşturulması gerekmektedir.

Stratejik hedefler doğrultusunda rekabet üstünlüğü elde etmek isteyen birçok örgüt gibi, Üniversiteler de rekabet üstünlüğü sağlamak ve kaliteli bir eğitim öğretim sağlamak için öğrencileri ile çalışanlarına daha iyi bir hizmet vermeyi hedeflemektedirler. Bu sebeple mevcut bilgi sistemlerinde çok boyutlu olarak depolanan veriler içerisinde kendilerine yararlı olan bilgileri üretebilmeyi istemektedirler. Bu çalışmada Erzincan Binali Yıldırım Üniversitesi Merkez Kütüphanesi'ne ait bilgi sistemi içerisindeki verilerin, uygun veri madenciliği teknikleriyle işlenmesi ile birlikte, karar vericiler için anlamlı ve faydalı olacak örüntülerin tanımlanması amaçlanmıştır. Bunun için tanımlayıcı veri madenciliği tekniklerinden olan kümeleme analizinin two-step algoritması ile birliktelik kuralları analizi için geliştirilen apriori algoritmalarından faydalanılmıştır. Analizler sonucu elde edilen bilgiler ile kütüphane karar vericilerine yol göstermek amaçlanmıştır. Ayrıca,

ödünç alma alışkanlıkları, kütüphane envanter bilgisi, kütüphane üye profili, birlikte ödünç alınan yayınlar gibi yararlı olabilecek sonuçlar da üretilmiştir. Söz konusu bilgilerin analizi ile kütüphanenin fiziksel olarak daha verimli planlanması, kaynakların ihtiyaçları karşılayacak şekilde yenilenmesi ve üyelere yeni kitap tavsiyelerinde bulunarak kullanım oranının artırılması hedeflenmektedir.

Bu kapsamda çalışmanın ikinci bölümünde veri, veri tabanı ve veri madenciliği kavramları genişçe ele alınarak; veri madenciliğinin süreci ile birlikte yöntemlerine ve bu alanda yapılan çeşitli çalışmalara yer verilmiştir. Ayrıca bu bölümde veri madenciliği yöntemlerinin kütüphanelere ait veriler üzerinde uygulandığı bibliomining kavramı tüm yönleriyle ve aşamalarıyla birlikte açıklanmıştır. Üçüncü bölümde veri madenciliği için kullanılan verinin elde edilmesi, ön işlenmesi ve temizlenmesi süreci ile birlikte yapılan kümeleme ve birliktelik kuralları analizinin olduğu uygulama bulunmaktadır. Uygulama sonucunda elde edilen tüm bulgular dördüncü bölümde paylaşılmış olup, çalışma elde edilen bulguların yorumlandığı ve çalışmanın benzer çalışmalarla kıyaslandığı beşinci bölüm olan sonuç ve öneriler bölümüyle sonlanmaktadır.

2. KURAMSAL TEMELLER

Bu bölümde, *Bibliomining* olarak da bilinen Kütüphane Veri Madenciliği bağlamında daha önce yapılmış olan çalışmalardan bazıları sunulmuştur. Diğer veri madenciliği alanlarına nazaran Kütüphane Veri Madenciliği daha yeni bir alan olarak kabul edilebilir. Bununla birlikte çalışma alanı hızla gelişmekte ve özellikle üniversite kütüphanelerindeki uygulamalarının artacağı düşünülmektedir.

2.1. Literatürdeki Kütüphane Veri Madenciliği Çalışmaları

Takçı ve Soğukpınar (2002) tarafından yapılan bir çalışmada bir kütüphanenin internet sitesine ait web günlükleri veri madenciliği teknikleri ile analize tabii tutulmuş ve kütüphane kullanıcılarının erişim örüntüleri bulunmaya çalışılmıştır. Bu çalışma yapılırken istatistiksel yöntemler kullanılmıştır.

Pu ve Yang (2003) tarafından yapılan çalışmada ise benzer kullanıcıların kaynak tercih örüntülerinin sirkülasyonu temel alınarak dağınık konu hiyerarşileri içindeki farklı birliktelik sınıflarının keşfedilebileceği değerlendirilmiştir. Bu sebeple, oluşturulacak örüntü şemalarından farklı kütüphanelerin kullanımları ve kullanıcıları için yararlanılabileceği düşünülmüştür. Çalışma sürdürülürken, kullanıcıların sınavlarına hazırlanmak için belirli kaynakları belirli dönemlerde tercih etme davranışları da göz önünde bulundurulmuştur. Öte yandan, çalışma için Tayvan da bulunan bir üniversitenin kütüphanesinin 2 yıllık kitap ödünç alma kayıtları kullanılmıştır. Söz konusu üniversite kütüphanesinde Dewey Onlu Sınıflandırma yöntemine benzeyen Çin Onlu Sınıflandırma yöntemi kullanılmaktadır.

Decker ve Höppner, (2006) Almanya da bulunan Bielefeld Üniversitesi kütüphanesinin sunduğu hizmetin etkinlik ve verimliliği ile akademik kütüphanelerdeki kullanıcı zekâsı uygulamalarının incelenmesi üzerine yoğunlaştıklarını bildirmişlerdir. Bu şekilde kullanıcıların kaynak arayışları içerisindeki örüntülerin tespit edilmesi ve bu sayede hizmet kalitesinin yükseltilmesi hedeflenmiştir. Öte yandan, stratejik planlama sürecini destekleyecek iç ve dış verilerle bağlantılı veri ambarı temelli kütüphane yönetim bilgi sistemleri için kavramsal bir çerçevenin tanıtımı yapılmıştır.

Uçan (2010) Akdeniz Üniversitesi Kütüphanesinde gerçekleştirdiği çalışmada, Birliklilik ve Kümeleme Analizi modellerinden Apriori ve TwoStep algoritmalarını kullanmak suretiyle, üniversite kütüphane veri tabanı çerçevesinde, kaynak ödünç alma bilgileri üzerinde odaklanmış ve kütüphane kullanıcılarının davranışlarını incelemiştir. Çalışma kapsamında, yayınlar arasında ilişkilendirme yapılmış ayrıca bölüm, cinsiyet grup, toplam ödünç başlıkları altında 4 grup oluşturulmuştur. Söz konusu gruplar grafikler ve sayısal istatistikî veriler yardımıyla da belirtilmiştir. Çalışmada kütüphane eser bilgileri göz önüne alınarak oluşturulan yığın verilerin karar mekanizmasına nasıl katılacağı tespit edilmeye çalışılmıştır.

Hajek ve Stejskal (2012) tarafından gerçekleştirilen bir başka çalışmada ise Çek Cumhuriyetinde bulunan büyük bir halk kütüphanesinde, kütüphaneye aktarılan kamu kaynaklarının verimli kullanılıp kullanılmadığını ve kullanıcıların en verimli şekilde kitap ve dokümanlara ulaşım ulaşımadığını incelemek amacıyla veri madenciliği uygulaması gerçekleştirilmiştir. Çalışmada kamu kurumlarının, halk kütüphaneciliğini de içine alan kamu hizmetlerini, merkezi otorite idaresiyle ya da komisyonlar aracılığıyla ve ayrılan fonlar çerçevesinde gerçekleştirdiği belirtilmiştir. Bununla birlikte gerek mali kaynakların yerinde kullanılması gerek kütüphanecilik faaliyetlerinin daha verimli icra edilmesi için geliştirilmiş bir yöntem bulunmadığı ifade edilmiştir. Bahse konu sebepler nedeniyle kütüphane kullanıcılarının bu kamu hizmetinden faydalanma noktasındaki davranış formlarının belirlenmesi hedeflenmiştir. Uygulama sürecinde, gerekli veriler kütüphane kaynaklarından temin edilmiş ve kümeleme analizi K-Means algoritması ile kullanılmıştır. Ayrıca hangi grup kullanıcıların hangi grup kaynaklardan faydalandığının tespit edilmesi ile kitap bağışlayan kişilerinde daha uygun ve ihtiyaç duyulan kitapları bağışlamasının sağlanması amaçlanmıştır.

Kütüphane sistemlerinde veri madenciliği uygulamasına yönelik deneysel bir çalışmada Uppal ve Chandwani (2013) tarafından kütüphaneden ödünç alınan kitapların sıklığının ve dizinlerinin ortaya çıkarılması amacıyla birliklilik analizi, kümeleme ve sıralı örüntü analizlerini uygulamışlardır. Araştırma sonuçlarından bir tanesi göstermiştir ki biyografi ve Çin edebiyatı kitapları alan okuyucular aynı zamanda ortak yabancı diller kitaplarından da almışlardır. Bu örüntünün, kütüphane yetkilileri tarafından öğrencilere sunulan hizmetin geliştirilmesi amacıyla kullanılabilmesi düşünülmüştür. Söz konusu

hizmet kapsamında, kütüphanenin, okuyucular ellerindeki kitapları bitirdikten sonra almaları için başka kitap tavsiyesinde de bulunabileceği değerlendirilmektedir. Öte yandan, kitapların kullanım sıklığı temeline göre diziliminde kullanıcıların almaları muhtemel kitapları bulmaları daha kolay olacaktır. Çalışma neticesinde, kitap tavsiye modelinin kitap mağazaları, bilgi alma sistemleri ve ağ referanslı veri tabanları gibi alanlarda da kullanılabilmesi düşünülmektedir.

Kütüphane kullanıcılarının üniversite bölümü, yaş aralığı, cinsiyet gibi kategorik bilgilerinin işlenmesinin ve kütüphane hizmetlerinin geliştirilmesi amacıyla kullanıldığı bir çalışma Zhang ve Wang (2013) tarafından gerçekleştirilmiştir. Çalışmada, Birliktelik analizi uygulaması olan Apriori algoritması kullanılarak, kütüphane kullanıcılarının bilgilerinden ve yaptıkları işlemlerden (ödünç aldıkları yayınlar, ödünç alma süresi vs.) oluşan veri tabanında örüntülerin tespit edilmesi amaçlanmıştır. Örüntüler neticesinde elde edilen modeller aracılığıyla kütüphane kullanıcılarına daha iyi hizmet sunulması düşünülmektedir.

Tempelman-Kluit ve Pearce (2014) tarafından, gerçek kişilerin amaçlarını, faydalarını ve değerlerini temsil eden kurgusal karakterler üzerinden akademik kütüphanelerde profil oluşturulmasını sağlamak amacıyla, 2014 yılında, bir çalışma yapılmıştır. Kümeleme analizi yöntemlerinden K-ortalama uygulamasının tercih edildiği çalışmada, kütüphanelerde üretilen hizmet verilerinden faydalanmak için sanal ortamda gerçekleştirilen sohbet üzerinden basit bir sohbet referans dökümü kodlanmıştır. Söz konusu döküm X/Y grafiği üzerine çizilmiş olup X doğrusu grafikte kullanıcının bilgi ihtiyacının niteliğini, Y ise kullanıcının motivasyonunun niteliğini temsil etmektedir.

2.2. Veri ve Veri Madenciliği Kavramı

Genel anlamda veri, bağımsız ve soyut gerçeklerden, ölçümlerden, sembol ve sayısal karakterlerden oluşan bir dizidir. Veri işlenmeden tek başına bir anlam ifade etmez. Bir verinin anlamlandırılması için analiz edilerek yorumlanması gerekir (Yılmaz, 2013). Veri madenciliğinin ana ögesi olan verinin depolanması ve analiz edilebilmesi için ise bilişim teknolojileri vasıtasıyla kayıt altına alınması gerekmektedir.

Günümüzde pek çok şekilde ifade edilebilen veri madenciliği genel anlamda, düzensiz biçimde ve büyük yığınlar halinde depolanan ham verinin analiz edilmesi, bu veriden kıymetli biçimler ve anlamlı ilişkisel bilgilerin ortaya çıkartılması şeklinde açıklanabilir

Bununla birlikte, veri madenciliğinin bazı tanımları aşağıdaki biçimlerde ifade edilmiştir.

Öte yandan, Argüden ve Erşahin (2008) ise veri madenciliğini; doğru kararlar alabilmek için büyük miktardaki karmaşık veri yığınları arasından faydalı olabilecek, anlamlı ve uygulanabilir bilgileri bulmak olarak tanımlamışlardır.

Zahedi ve Zare-Mirakabad (2014) göre veri madenciliği, veri tabanlarında örtülü halde var olan bilgilerin ortaya çıkarılması veya veri toplulukları içerisinde bulunan ilişki ve bağlantıları keşfetmek için verileri inceleme ve analiz etme sürecidir.

Veri madenciliği; büyük veri yığınları içerisinde anlamlı ve kullanılabilir bilgileri açığa çıkartmak amacıyla, modern teknolojinin, istatistiksel ve matematiksel tekniklerini kullanma işlemidir (Tozak vd., 2017).

Veri toplama ve depolama teknolojilerindeki gelişmeler, muhtelif kurum ve işletmelerde büyük miktarda veri depolanabilmesine olanak sağlamıştır. Depolanan söz konusu veriler çok büyük miktarlarda olduğu için anlamlı bilgilerin elde edilmesi zor bir hal almıştır. Geleneksel veri analiz teknikleri büyük miktardaki veriyi çözümlenmekte yetersiz kalmış olup veri madenciliği bu soruna çözüm üreten bir tekniktir. Veri madenciliği geleneksel veri analiz teknikleri ile geliştirilmiş algoritmaları birleştiren bir teknolojidir (Ceylan vd., 2017).

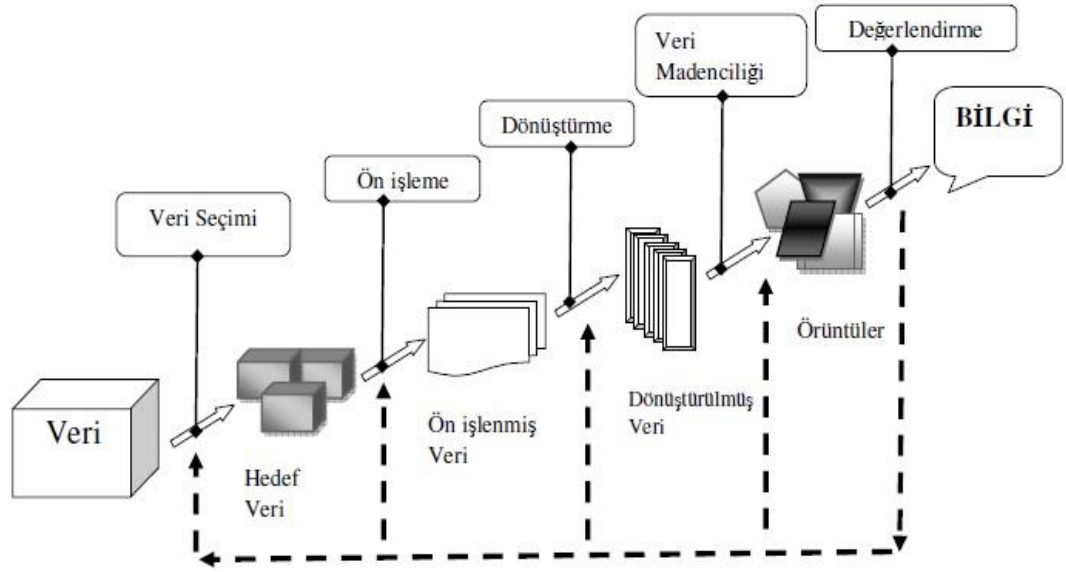
Veri madenciliğinin amacı büyük miktardaki ve büyük çoğunluğu denetlenmemiş veriyi tanım kümesi içerisinde anlamlandırmaktır. Veri madenciliğinde modellemeler yapılırken sadece veri yığınları arasında değil, tek bir veri ile diziler arasındaki ilişkiler de göz önüne alınır. Bu nedenle veri madenciliğinde modelleme çeşitleri belirlenerek kurallar bütünü haline getirilmeli, veriler gruplandırılarak anlamlı ilişkiler araştırılmalıdır. Bu ilişkisel çıkarımlar veri madenciliği uzmanları tarafından yapılmakla beraber yöneticiler, veri madenciliğini yorumlaması gereken karar vericiler için anlaşılabilir düzeyde olması yine veri madenciliği uzmanlarının sorumluluğundadır.

Karar vericiler için gerekli bilgiyi yorumlayacağı şekilde sunmak, araştırmanın güvenilirliğini sağlamak, veri madenciliği uzmanlarının görevidir (Cios vd., 2007).

Veriler birçok yapıda sınıflandırılmış ya da sınıflandırılmamış şekilde depolanabilmektedir. Veri madenciliği verilerin kalitesine ve miktarına bağlı olarak değerlendirilir. Tüm bu verilerin bir arada bulunmasıyla oluşan veri setleri temel tanımıyla çevresel ya da işlemsel süreçlerden elde edilen değerlerdir. Veri madenciliği çalışmasının yapılabilmesi amacıyla yeterli boyutta veri yığına ihtiyaç duyulmaktadır. Tek başlarına ya da yetersiz sayıda olduğunda anlamlandırılmaya müsait olmayan veriler, ancak veri setleri haline geldiklerinde bir anlam ifade eder ve çıkarım yapılabilir durumda olurlar. Öte yandan, bu noktaya gelmeden önce verilerin birçok aşamadan geçmesi ve işlenmesi gerekmektedir. Bununla birlikte günümüzde veri akışının kontrol edilebilirlikten uzak olması, veri madenciliği için verileri yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış olarak gruplandırma zorunluluğu getirmiştir (Kantardzic, 2001).

Yapılandırılmamış veriler arasında en sık, çoklu ortam (Multimedya) dosyaları ile karşılaşmaktadır. Günümüz koşullarında kapasite probleminin ortadan kalkmasıyla görsel ve işitsel verilerin varlığı ve depolanabilirliği artmaktadır. Ancak bu verilerin içerdiği bilgiler depolama aşamasında kategorize edilmediğinden ilerleyen süreçte sağlıklı ve anlamlı bir hale dönüştürülmesi gerekmektedir. Yarı yapılandırılmış verilere örnek ise yazılı dokümanlar gösterilebilir. İşletmelerde kullanılan dokümanlar, raporlar ve yazışmalar yeni nesil veri tabanları sayesinde kayıt altına alınmakta ve işlenebilir veri olarak kullanılabilir. Öncelikle içerik olarak sınıflandırılmalı ve sonrasında, yapılandırılmamış veriler de olduğu gibi, ham veri setinden, kullanılabilir veri setine dönüştürülmesi gerekmektedir (Kantardzic, 2001).

Veri madenciliği için kullanılmaya en uygun veriler ise yapılandırılmış verilerdir. Bu veri setlerinde önemli olan husus hangi bilginin bizim için işe yarar durumda olduğunu belirleyebilmektir. Ancak dikkat edilmesi gereken nokta alanların niceliksel ya da kategorik olup olmadığıdır. Yaş ve gelir gibi değerler herhangi bir değer almaları gerekirken, cinsiyet medeni durum gibi ölçütler daha önceden belirlenmiş bazı kategoriler altında toplanmalıdırlar. Bu ayırım analitik teknikler açısından önemlidir. Niceliksel ya da kategorik verilere göre analitik teknikler uygulanmalıdır (Hand vd., 2001).



Şekil 2.1. Bilgi Keşfi Sürecinde Veri Hazırlık Aşamaları (Han ve Kamber, 2001)

Veri madenciliğinin günümüzde yaygın bir kullanım alanı bulunmaktadır. Bu alanlar pazarlama, bankacılık, sigortacılık, elektronik ticaret, telekomünikasyon, tıbbi araştırmalar ve astronomi gibi bilimsel araştırmalar olarak sıralanabilir. Bunlar kullanım yerlerine göre aşağıdaki gibi sınıflandırılmıştır (Özkan, 2008):

- Pazarlama alanında; Müşterilerinin satın alma alışkanlıklarının belirlenmesi, Müşterilerin demografik özellikleri arasındaki bağlantıların ortaya konması, Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması, Pazar sepeti analizi, Müşteri ilişkileri yönetimi, Müşteri değerlendirme, Satış tahmini
- Bankacılık alanında; Farklı finansal göstergeler arasındaki gizli korelasyonların tespiti, Kredi kartı dolandırıcılıklarının tespiti, Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi, Kredi taleplerinin değerlendirilmesi
- Sigortacılık alanında; Yeni poliçe talep edecek müşterilerin tahmin edilmesi, Sigorta dolandırıcılıklarının tespit edilmesi, Riskli müşteri gruplarının belirlenmesi
- Elektronik Ticaret alanında; Saldırıların çözümlenmesi, e-CRM uygulamalarının yönetimi, WEB sayfalarına yapılan ziyaretlerinin çözümlenmesi, Kullanıcı davranışlarına göre web sitesinin yenilenmesi,

- Telekomünikasyon alanında; İletişim ağlarında sorunlu bölgelerin tespiti, Kaçak hat kullanımlarının belirlenmesi, Kullanıcı davranışlarının belirlenmesi, Müşteri davranışlarına göre yeni hizmetlerin sunulması
- Tıbbi Araştırmalar alanında; DNA içerisindeki genlerin sıralarının belirlenmesi, Protein analizlerinin yapılması, Hastalık haritalarının hazırlanması, Hastalık tanıları, Sağlık politikalarına yön verilmesi

Bunların dışında da veri madenciliğinin faydalı olabileceği ve kullanılabileceği alanlardan bazıları; Taşımacılık ve Ulaşım, Turizm ve Otelcilik, Belediyeler, Eğitim, Bilim ve mühendislik konularıdır.

2.3. Verilerin Gizliliği ve Kişisel Verilerin Korunması Kanunu

Verilerin işlenebilmesi amacıyla, gerektiğinde ve kolayca ulaşım sağlayabilmenin önemi büyüktür. Ancak bahse konu verilerin, bireylerin özel yaşamlarına ilişkin bilgileri ihtiva ediyor olması nedeniyle gizliliği ve güvenliği durumları söz konusu olmaktadır. Bu aşamada, kişilerin kimlik bilgilerinden tüketim alışkanlıklarını gösteren bilgilere kadar tüm verileri korumak, ilgili kurum ve kuruluşların görevleri arasındadır.

Günlük yaşamda, teknolojinin büyük oranda insan hayatının içine nüfuz etmesiyle birlikte bireyin kimlik, iletişim, sağlık ve mali bilgileri, dini inancı, siyasi görüşü gibi kişisel verilerinin mahremiyetini korunması büyük önem arz etmektedir. Kişisel veriler, gerek özel sektör gerek kamu sektörü tarafından bilişim sistemleri üzerinden otomatik yollarla sıkça işlenmektedir. Bu verilerin işlenmesi bireyler ile mal ve hizmet sunanlar bakımından bazı kolaylıklar ve avantajlar sağlasa da, söz konusu verilerin istismar edilme riskini de beraberinde getirmektedir. Dolayısıyla bu iki menfaat arasında meşru ve makul dengenin kurulması gerekmektedir (KVKK, 2016).

Ülkemizde 1981 yılından bu yana kişisel verilerin korunması konusunda mevzuat oluşturma çalışmaları yapılmaktadır. Söz konusu çalışmalar neticesinde 6698 sayılı Kişisel Verilerin Korunması Kanunu 7 Nisan 2016 tarihinde yürürlüğe girmiştir.

6698 sayılı Kanun kapsamında koruma altına alınan kişisel veriler sadece gerçek kişilere ait veriler olup, tüzel kişilere ait olanlar koruma altında bulunmamaktadır. Fakat tüzel

kişiyeye ait verinin elde edilmesi, bir veya birden fazla gerçek kişinin kimliğinin belirlenmesine neden oluyor ise, bu tür verilerin de Kanunun korumasından yararlanması mümkün olabilir. Bu durumda da esasen korunan, gerçek kişiyeye ilişkin kişisel veriler olmaktadır.

2.4. Veri Tabanı ve Veri Tabanı Yönetim Sistemleri

Veri tabanı, veri düzlemi olarak da bilinmekte olup en genel tanımıyla kullanım amacı gözetilerek depolanmış veri topluluğudur. Birbiri ile ilişkili verinin tutulduğu, kullanım amacına uygun biçimde düzenlenmiş veri topluluğunun fiziksel ve mantıksal olarak tanımının olduğu depolardır. Bilişim ortamlarındaki formlardan gelen veri kalıcı değildir. Mevcut sayfa kapatıldığında veride kayıp yaşanacağı için verinin ileride kullanabilmesi amacıyla kayıt altına alınması gerekir. Bir dosyaya veya belgeye yapılan kayıtlara okuma, arama, güncelleme, silme vb. işlemlerinin yapılması oldukça zor ve kullanışsızdır. Veri tabanları verilerin belli bir düzende kayıt edilmesi ile oluşan depolardır. Sık ve yoğun işlemler yapılacağı bilinciyle oluşturulan veri tabanında kayıtların yönetilmesi (sorgulama, silme, güncelleme vb.) dosyalama sistemine göre daha kolaydır (Vural ve Sağırođlu, 2010).

Bir veri tabanına veri eklenip veri silinebilmesi gibi işlemlerin yapılması için VTYS'ye ihtiyaç vardır. VTYS'ler, veri tabanını düzenlemek, tanımlamak, kullanmak, değiştirmek vb. işlemleri yapabilmek amacıyla hazırlanmış yazılımlardır. Yeni bir veri tabanı oluşturmak, geliştirmek, düzenlemek, bakımını yapmak gibi karmaşık işlemlerin yapıldığı yazılımlar VTYS olarak adlandırılmaktadır (Vural ve Sağırođlu, 2010).

MS (Microsoft) SQL Server, My SQL, Oracle, Acces, en çok bilinen ve tercih edilen VTYS yazılımları arasında sıralanabilir (Kaplan vd., 2017; Abrar ve Wahyudi, 2016; Yağcı vd., 2015; Deperliođlu ve Sarpkaya, 2009).

Veri madenciliğinin gerçekleştirilebilmesi için verinin kayıt altına alınması gerekir. Bu noktada VTYS veri tabanlarını oluşturma ve yönetme sistem yazılımıdır. Veri tabanı yönetim sistemi kullanıcıların ve programcıların veriyi, sistematik bir şekilde, oluşturmalarını, geri almasını, güncellemesini ve yönetmesini sağlar.

Gelişmiş veri tabanları gün geçtikçe daha çok nesne odaklı ve nesne ilişkili hale gelmektedir. Veri Madenciliğinde çoğunlukla büyük veri yığınları ile çalışılmaktadır. Büyük veri yığınları ise gelişmiş ve özelleştirilmiş yönetim sistemlerine ihtiyaç duymaktadır. Bunun dört sebebi vardır (Cios, vd., 2007).

- Veri madenciliği için kullanılan veri büyüklüğünün güçlü donanımsal özelliklere ihtiyaç duyması
- Kullanılan veri madenciliği metotlarının farklı veri alt kümelerine gereksinim duyması Buna bağlı olarak gerekli verilerin doğru şekilde kümelendirilmesi zorunluluğu
- Gerekli verilerin eş zamanlı eklenmesine ya da güncellenmesine farklı zamanlarda farklı kişilerce gereksinim olması
- İhtiyaç duyulan verilere ulaşmak için veri yığınları arasından belirli bölümlere ihtiyaç duyulması

Yukarıda belirtilen sebepler veri tabanı yönetim sistemlerinin gerekliliğini ortaya koymaktadır. Veri tabanı yönetim sistemi, birbiriyle ilişkili veri tabanlarından meydana gelen verileri yönetmek ve verilere erişim sağlamak için kullanılan programlardır (Han ve Kamber, 2001). 1960'ların sonlarından itibaren ortaya çıkmaya başlayan veri tabanı yönetim sistemleri, dosyalama sistemleri olarak oluşturulmasına rağmen günümüzde gelişmiş veri tabanı sorgulama dilleri ile yönetsel alanda kullanımı kolaylaştırılmıştır. İyi bir veri tabanı yönetim sistemi aşağıdaki özelliklere sahip olmalıdır (Garcia-Molina vd., 2008):

- Yeni veri tabanları ve şema oluşturabilmesini sağlaması
- Kullanıcılarına veri sorgulama şansı vermesi
- Büyük veri yığınlarını depolama imkânı sağlaması
- Verilere çoklu erişimi sağlaması

Günümüzde gelişmiş veri tabanı sistemleri; uzamsal veriler (haritalar vb.), mühendislik tasarım verileri, hipermetin ve internet ortamında oluşturulan verileri de içermektedir. Oluşturulan sistemler yapılandırılmamış veya yarı yapılandırılmış olsa dahi verileri

işleyebilme ve bu verileri farklı metotlarla yeniden dinamik ortamda şekillendirmeye yardımcı olurlar (Han ve Kamber, 2001).

Han ve Kamber (2001)'e göre gelişmiş veri tabanı sistemlerini yedi kategori altında toplayabilmektedir. Bunlar sırasıyla; Nesne-yönlü veri tabanı sistemleri, Nesne-ilişkili veri tabanı sistemleri, Uzamsal veri tabanı sistemleri, Zaman serileri veri tabanı sistemleri, Metin ve çoklu ortam veri tabanı sistemleri, Heterojen veri tabanı sistemleri, World Wide Web' dir.

2.4.1. Nesne yönlü veri tabanları

Son yıllarda bilgisayar uygulamalarının dosya bazlı veri saklama işlevleri yeterliliğini kaybetmeye başlamasıyla birlikte veri işleme dünyasında hızla yükselen veri tabanları, uygulama yazılımları için de ihtiyaç haline gelmiş ve nesne yönlü veri tabanlarını ortaya çıkmasını sağlamıştır. Yazılımlar için geliştirilen programlama dilleri ile sadece verileri değil ses, görüntü ve çeşitli dokümanlarını yönetmek ihtiyaç haline gelmiştir. C++, Visual Basic, .Net, Python gibi nesne yönlü programlama dillerinin doğmasını sağlamıştır. Web tabanlı uygulamaların gelişmesi ile artık Nesne yönlü veri tabanları geniş bir kullanım alanına erişmiştir (Doğaç vd., 1994).

Nesne yönlü veri tabanları nesnelerin toplanması ile ilgilidir. Her nesnenin bir durum ve davranışı mevcuttur. Nesne davranışı nesnenin durumunu günceller. Veri ve diğer veriler arasındaki ilişkiler (davranışlar) basit bir yapı olan nesnelerle tanımlanmıştır. Nesneler tek başlarına bir anlam ifade etmezler. Nesneyi açıklayan değişkenler, diğer nesnelerle iletişim kurulabilmesi için iletişim setleri ve bu iletişimi sağlayacak metotları barındırmalıdır. Nesne yönlü veri tabanları, veri tabanı yönetim sistemi ve nesne-yönlü sistemlerin özelliklerini taşımalıdır. Veri tabanı yönetim sistemi olmasından dolayı devamlılık, ikincil depolama yönetimi, uyumluluk, geri döndürülebilirlik ve özel amaçlı sorgu araçları barındırmalıdır. Nesne yönlü veri tabanı olmasından dolayı karmaşık nesneler, nesne kişilikleri, kapsülleme, sınıflandırmalar barındırmalıdır (Atkinson, vd., 1989).

2.4.2. Nesne ilişkili veri tabanları

Nesne ilişkili veri tabanları ilişkisel modellerin bulunduğu veri tabanlarında karmaşık nesne ve nesne yönelimlerine zengin veri kaynağı sağlamak için kullanılır. İlişkisel veri tabanlarında bulunan karmaşık veri çeşitleri, sınıf hiyerarşisini ve kalıtımı geniş bir alanda kullanımını sağlar. Bu kullanım veri tabanı şemaları ve sorgulama dilleri ile desteklenir (Han ve Kamber, 2001).

İlişkisel veri tabanları ile Nesne yönlü veri tabanlarının arasında yer alan nesne ilişkili veri tabanları, ilişkisel veri tabanlarında olduğu gibi sınırlı bir veri yığına kapsamaz, geliştiricilerin kendi modelleri ve metotlarını ilişkisel veri tabanlarına adapte etmesini sağlar (Han ve Kamber, 2001).

2.4.3. Uzamsal veri tabanları

Uzamsal veri tabanları bir alanı kapsayan geometrik, coğrafik ya da uzamsal verileri barındırır. Bu, dünyanın iki boyutlu yüzeysel şekilleri olabileceği gibi, tıp araştırmalarında insan beyninin yapısı ya da kimyasal araştırmalarda üç boyutlu protein molekülleri de olabilir. Uzamsal veri tabanları yukarıda belirtilen verileri barındırır da alfa-sayısal verilere her zaman ihtiyaç duymaktadır. Diğer veri tabanlarından ayıran nokta ise uzamsal verilerle ilgili yapabildiği ek işlemlerdir. Uzamsal veriler nokta, çizgi ve bölge şeklinde tutulmaktadır. Uzaysal veri tabanları ise bu bilgileri iki veya üç boyutlu hale getirerek birbirleri arasındaki ilişkiyi ortaya koyabilmektedir (Güting ve Hartmut, 1994).

2.4.4. Zaman serisi veri tabanları

Zaman serisi veri tabanlarında veriler, zaman aralıklarından çıkarım vasıtasıyla elde edilen ilişkili sayılardan oluşur. Bu veriler örneğin hisse senedi borsalarında olduğu gibi zaman içerisinde değişen ardışık değerler olabilirler. Zaman serisi veri tabanları ile kurulan sistemler zaman serilerini belirlenen bir düzen içerisinde oluşturur, değiştirir ve günceller. Bu düzen genelde hiyerarşik bir düzen içerisinde gerçekleşir (Han ve Kamber, 2001).

Zaman serisi veri tabanı sistemlerinde, veri madenciliği özellikle strateji planlamada önemli bir araç olarak kullanılmaktadır. Özellikle yatırım araçlarında teknoloji ile birlikte gelişen anlık verilerin izlenimi zaman serilerine dayanarak tahmin etme yöntemlerini önemli kılmaktadır (Han ve Kamber, 2001).

2.4.5. Metin ve multimedya veri tabanları

Metin veri tabanları anlamlı bir yapısının olduğunu bildiğimiz ancak metinden yapısı hakkında bilgi alamadığımız, bu bilgiyi elde etmek için diğer ilişkili verilere ihtiyacımız olan yapılandırılmamış verilerden oluşur. Yapılandırılmamış veriler salt metinlerden oluşmaktadır. İnternet ortamında kaydedilen veriler, masaüstü yazılımları tarafından saklanan veriler buna örnektir. Ancak metin veri tabanları her zaman yapılandırılmamış metinlerden oluşmayabilir. E-posta içerikleri, programlama dil kodları gibi yarı-yapılandırılmış metinler, ya da dijital kütüphanelerde kullanılan yarı yapılandırılmış içerikler de metin veri tabanlarında kullanılabilirler (Doedens, 1994).

Multimedya veri tabanları ses, görüntü resim gibi verilerin kaydedilmesi için geliştirilmiştir. Multimedya veri tabanlarının ortaya çıkış nedeni verilerin boyutlarının çok büyük olması, bu yüzden de özelleştirilmiş depolama ve arama tekniklerine ihtiyaç duyulmasından dolayıdır (Cios, vd., 2007).

2.4.6. Heterojen veri tabanları

Heterojen veri tabanları birbirinden bağımsız veri tabanlarının kullanıcı sorgularında ilişki sonuçları gösterebilmek için iletişim halinde olması neticesinde oluşur. İlişkisel veri tabanlarında nesnelere tamamen farklı olabilir. Bu yüzden ilişkileri belirlemek oldukça zordur (Han ve Kamber, 2001).

2.4.7. World wide web (Dünya Çapında Ağ)

Günümüzde internet ile sık sık aynı anlamda kullanılan World Wide Web, interneti de içine alan daha geniş bir kavramı yansıtmaktadır. World Wide Web İnternet veya ağ ortamında bulunan dokümanların bağlantılar vasıtasıyla birbirine bağlı bir veri deposu

oluşturmasıyla ortaya çıkmıştır. Tüm verilerin hiperlinkler ile birbirine bağlı olması sayesinde kullanıcıların bilgiye ulaşmasını sağlar (Han ve Kamber, 2001).

Büyük miktarda verilerin birbirine bağlı olması veri madenciliği araştırmaları için de oldukça iyi fırsatlar sunmaktadır. Kullanıcıların bilgiye erişim alışkanlıkları ya da sıklıkları, örneğin bir e-ticaret şirketi için daha iyi pazarlama kararları verebilmek açısından, oldukça elverişli bir bilgidir. Ancak kullanıcılar için internet ortamındaki bilgiler anlaşılabilir düzeyde olsa da veriler yapılandırılmamış halde bulunmaktadır. Yapılandırılmamış verileri kullanıcı sorgularına göre yapılandırarak kullanıcının anlayabileceği bilgiye çevirmek için oldukça gelişmiş algoritmalar kullanılmaktadır. Google, Yahoo, Bing gibi arama motorları internet üzerinde veri madenciliğinin en gelişmiş örneklerindedir. Google'ın Pagerank algoritması oldukça gelişmiş mantıksal sorgularla internet üzerindeki yapılandırılmamış verileri yapılandırılmış ya da yarı yapılandırılmış veriler halinde sunucularında saklayarak kullanıcılarına hızlı bir şekilde bilgi akışı sağlar (Han ve Kamber, 2001).

2.5. Veri Ambarı

Veri Ambarı, yöneticilerin karar verme süreçlerinde konu-odaklı, entegre edilebilir, zaman-değişkenli ve kalıcı verilerin toplanması işlemidir. Bu özellikleri veri ambarlarını diğer veri depolama uygulamalarından ayıran ana özelliklerdir. Veri ambarları geçmiş veriler üzerinde analizler yapılabilmesini ve istenilen bilginin daha kolay elde edilebilmesini sağlar (Inmon, 2002).

Başka bir deyişle Veri Ambarı; farklı bilgisayarlara dağıtılmış, çeşitli formatlarda olabilen bilgilere hızlı erişilebilmesine imkân tanıyan ve karar verme aşamasında stratejik verilerin elde edilmesini sağlayabilecek şekilde düzenlenmiş merkezi veri yapılarıdır (Inmon, 2002).

Bu doğrultuda veri ambarının karakteristik özellikleri arasında (Inmon, 2002);

- İşletmenin ana konuları ve temel faaliyetleri çerçevesinde organize edilmiş olması gerektiği, karar destek için kullanılmayacak olan veriyi barındırmadığı, kullanıcının kendi verilerini analiz etmesine yardımcı olmak üzere tasarlandığı,

- İlişkisel veri tabanı sistemleri ve metin dosyaları gibi entegre edilebilir çoklu heterojen kaynaklardan beslenmekte olduğu, farklı dış kaynaklardan da veri alabildiği ve bu verilerle entegre halde çalışabildiği,
- Tüm verilerin zaman boyutu olduğu, yani bir zaman içerisinde belli bir anı temsil ettikleri,
- Verilerin veri ambarına girdikten sonra değiştirilemediği ve ambarda yer alan verilerin sadece okunabilirlik özelliğinin olduğu hususları yer almaktadır.

2.6. Verilerin Hazırlanma Süreci

2.6.1. Veri temizleme

Veri temizleme süreci veri tabanlarında eksik ve anlamsız değerleri sistemli olarak veri tabanından çıkarma aşamasıdır. Bu süreç planlı bir süreç olmalı, kirli verilerin tespitinin sağlıklı yapılması gerekmektedir. Veri hazırlama süreçleri için ne yazık ki işlemi otomatik sürdürebilecek akıllı yazılımlar bulunmamaktadır. Bu yüzden kullanılacak veri madenciliği araçlarının iyi analiz edilmesi ve dikkatli bir şekilde veri tabanlarına uygulanması gerekmektedir (Pyle, 1999).

Eksik Değerlerin Tespiti, veri tabanında bulunan alanlarda eksikliklerin tespiti halinde; satırın iptal edilmesi, eksik verinin elle girilmesi, eksik değerler için genel bir sabit değer atanması, eksik değerler için ortalama değer atanması, eksik değerler için diğer ilgili alanların ortalama değerinin atanması, eksik alan için en uygun değerlerin hesaplanarak atanması durumları tercih edilebilir (Özkan, 2008).

Eksik değerler için genel bir sabit değer atanması ve eksik alan için en uygun değerlerin hesaplanarak atanması çok doğru sonuçlar vermeyebilir. Bu sebeple eksik alan için en uygun değerlerin hesaplanarak atanması yöntemi ve verilerin çıkartılması en çok kullanılan yöntemler arasındadır (Han ve Kamber, 2001).

Kirli verilerin temizlenmesi: Gürültülü veriler olarak da adlandırılan kirli veriler, bütüne bakıldığında ilgisiz ve anlamsız kalan verileri tanımlamaktadır. Veri temizleme metotlarında yeterince başarılı olamayan veri toplama işlemleri sonucunda oluşan kirli

verilerin bulunması ve temizlenmesi ön plandadır. İlgisiz ve ilgisi oldukça düşük olan veriler veri analizlerine engel teşkil ettiğinden kirli veri olarak ele alınması gerekmektedir (Xiong vd., 2006).

2.6.2. Veri uyarlama

Veri Uyarlaması (Entegrasyonu) değişik kaynaklardan sağlanan verilerin elde edilmesi ve sonrasında birleştirilmesi problemidir (Halevy, 2001; Hull, 1997). Genelde çoklu bilgi sistemlerinin amacı belirlenen sistemlerden alınan bilgileri birleştirerek tek bir bilgi sistem ağında toplamaktır. Verilerin uyarlanması gerekliliği iki sebebe dayanmaktadır. Bunlardan birincisi, var olan bilgi sistemlerini bilgi erişimi oluşumuna ve tekrar kullanıma uygun hale getirmektir. İkinci sebep ise daha kapsamlı tatmin sağlanabilmesi amacıyla farklı tamamlayıcı bilgi sistemlerinden ihtiyacı gidermektir (Ziegler ve Dittrich, 2004).

Uyarlama sırasında karşılaşılan problemlerden en önemlileri “şema uyarlaması” ve “artık bilgilerdir”. Kayıt tanım problemleri de denilen şema uyarlaması esnasında çıkan hatalar birden fazla veri kaynağından gelen verilerin birbiriyle tutarsız olmasından kaynaklanmaktadır. Artık bilgiler ise kaynaktan alınan verilerin başka tablolardan elde edilen verilerle tutarsız olmasından kaynaklanır. Örneğin toplam satış tutar verisinin satışlar tablosundan elde edilen verilerin toplam değerinden farklı olması gibi. Bu hatalar genelde korelasyon analizleri ile bertaraf edilmektedir. Belirlenen iki değer arasında ne kadar güçlü bir ilişki olduğu kestirilebilir olmaktadır (Ziegler ve Dittrich, 2004).

2.6.3. Veri dönüşümü

Veri dönüşüm süreci anlık-ilişkili (instance-related) dönüştürme ve şemalama işlemlerini barındıran birçok adımdan oluşmaktadır. Veri dönüşümü sürecinde dönüşüm kodları oluşturmak ve otomatik bir şekilde gerçekleştirmek için dönüşüm işlemi belirli ve uygun yöntemlerle gerçekleştirilmelidir. Bu yöntemlerden en genel ve esnek olanı veri tabanı sorgulama dillerinden olan SQL'dir. SQL ile veri dönüşümünün yanı sıra kullanıcı-tanımlı fonksiyonların oluşturulmasına ve kullanıcı uygulamaları için özelleştirilmesine olanak sağlamaktadır. Kullanıcı tanımlı fonksiyonlar veri dönüşümünü daha verimli, hızlı

bir şekilde gerekleřmesinde en buyk yardımcılardandır. Maliyet ve zaman konusunda tasarruf saęlar (Rahm ve Hai, 2000).

Veri donuřmnde kullanılan bir dięer yntem ise veri haritalama yntemidir. Veri haritalama veri modelleri arasında eřleřtirme yapılmasıdır. rneęin bir řirketin satıř iřlemleri iin oluřturduęu faturaları standardize etmek iin dięer řirketlerle oluřturulan ortak standartlarla karřılařtırarak fatura verilerini belirli bir sistem iinde tutmasıdır. Kullanım alanı kısıtlı olduęu iin kullanıcı tanımlı fonksiyonlar kadar popler deęildir (Han ve Kamber, 2001).

2.6.4. Veri daraltma

Veri daraltma veri seti ierisinden tm veri setini yansıtacak řekilde daha kk bir veri seti haline donuřtrme iřlemidir. Byk veri tabanlarından hızlı ve uygun cevaplar bulmak her zaman mmkn olmamaktadır. Veri madencilięi iřlemlerini tm veri zerinde uygulamaktansa daha kk ve veri btnlęn yansıtabilecek bir veri setinde alıřmak daha etkili olacaktır (Han ve Kamber, 2001).

Veri daraltma teknikleri iki ana sınıfa ayrılmaktadır. Bunlardan birincisi veri seti iin bir model belirlenip bu modelin parametrelerinin hesaplandıęı parametrik tekniklerdir. Lineer regresyon modelleri, log-lineer modelleri parametrik tekniklere rnektir. Dięer bir veri daraltma teknięi ise modelleme yapmadan veri daraltma iřlemini gerekleřtiren parametrik olmayan tekniklerdir. En yaygın kullanılan parametrik olmayan veri daraltma tekniklerinden birisi de Histogram indeks aęacı teknięidir (Han ve Kamber, 2001).

2.7. niversite Ktphanelerinde Veri Madencilięi (Bibliomining)

Scott Nicholson 2003 yılında ktphane veri madencilięi zerine yaptıęı alıřmaların ardından alıřmalarını “BIBLIOMINING” olarak adlandırmıř ve literatre yeni bir terim kazandırmıřtır. Bibliomining, “bibliyometri” ve “datamining” terimlerinden tretilmiřtir (Nicholson, 2005).

Ktphanelerde oluřan verilerin, veri madencilięi ile incelenmesi ktphanecilik olarak da nitelendirilebilmektedir. Ktphane alıřanları ile bilgi bilimciler bibliyografik bilgi kalıplarını keřfetmek iin bilgisayar kataloglarının yaygın olarak kullanıldıęı

bibliyometrik metotları kullanmışlardır. Ayrıca arařtırmacılar, veri madencilięi tekniklerini geliřtirerek test etmişlerdir. Bibliomining kütüphanelerin ürettięi büyük boyutlardaki verilere ulaşmak için bu teknikleri uygulamaktadır (Nicholson ve Stanton, 2003).

Kütüphane veri madencilięi kütüphanelerdeki kaynakların, hangi kategorideki kullanıcılar tarafından ne sıklıkta ve nasıl bir sıralama ile kullanıldığı yönündeki parametreleri incelemektedir. Söz konusu parametrelerin incelenmesi ve aradaki örüntülerin tespit edilmesi ile karar vericilerin, ilgili kullanıcılara yönelik olan, daha yüksek verim arz eden ve sonradan talep edilmesi muhtemel (birbirinin ardışıęı olacak) kaynakları temin etmesi hedeflenmektedir. Bu şekilde, özellikle, üniversite kütüphanelerinde yapılacak veri madencilięi çalışması ile öğrencilerin kaynak kullanım durumlarının belirlenmesi ve bu durumun başarı düzeyine etkisi de incelenebilecektir (Nicholson, 2003).

2.7.1. Kütüphanelerde veri madencilięi aşamaları

Kütüphanelerde veri madencilięi (bibliomining) süreci aşağıdaki adımlar takip edilerek gerçekleştirilebilir (Nicholson, 2003) :

- Odaklanacak Alanların belirlenmesi: Dijital kütüphanelerde veri madencilięinin ilk aşaması madencilięin gerçekleştirileceęi amacın ve bu amaca baęlı olarak ilgili alanların tespit edilmesidir. Bu odaklanma belli bir problem üzerine olabileceęi gibi karar verme ve bilgi keřif için genel bir tarama da olabilir. Bu nedenle veri madencilięi işleme karar vermeden önce veri madencilięinin önceden yönlendirilmiş ya da yönlendirilmemiş olması belirtilmelidir. Eęer problem odaklı bir tarama söz konusu ise önceden yol haritası belirlenmelidir. Veri madencilięinde belirli bir soruna yönlendirilmeden de genel bir veri madencilięi taraması gerçekleştirerek kütüphanenin genel durumu hakkında bilgi sahibi olunabilmektedir. Ancak böyle bir yol izlenmesi durumunda çeřitli zorluklarla karşılaşıması da ihtimal dâhilindedir. Tüm kütüphane verilerinin veri madencilięi işleminden geçirilmesi, veri madencilięi süreci öncesinde gerçekleştirilmesi gereken veri temizleme ve uyarlama işlemlerinin oldukça vakit almasına neden

olmaktadır. Özellikle olasılık üzerine kurulu veri madenciliği modellemelerinin kullanılması güçleşecek ve güçlü bilgisayarlara gereksinim duyulacaktır.

- Veri Kaynaklarının tespit edilmesi: Problem ya da veri madenciliğinin amacı tespit edildikten sonraki adım uygun veri kaynaklarının belirlenmesidir. Kütüphanelerde veri madenciliği süreci işlemsel, birleştirilmemiş ve düşük-seviye veriler üzerinden gerçekleştirilmektedir. Bu durum büyük veri yığınının depolanması ve depolama maliyetlerinin yüksek bir düzeyde olması, kullanıcı gizliliğinin korunması gibi nedenlerle kütüphane karar vericilerinin bu verileri saklamakta isteksiz olmalarına sebep olmaktadır. Oysaki kütüphane kaynakları kadar işlemsel veriler de veri madenciliği için önemli bir araçtır. Bununla birlikte kütüphanelerde veri depolama sistemleri mutlak bir gereksinimdir.
- Dijital kütüphane veri kaynaklarında iki çeşit veri üzerinde durulmalıdır. Bunlardan ilki kütüphane sisteminde mevcut olan iç verilerdir. Bu veriler yayın bilgileri, yazar bilgileri, süreli yayınlar ya da tezler, işlemsel veriler, web arama kayıtlarından oluşabilir. Bu veriler oldukça karmaşık ve zorlayıcı bir süreç sonucunda veri madenciliğine kullanılabilir şekle getirilebilmektedir. Diğer veri türü ise dışsal verilerdir. Bu veriler demografik bilgiler, yerleşim yeri gibi daha genel bilgiler olup dış bir kaynaktan elde edilebilmektedir.
- Veri ambarının oluşturulması: Belirtildiği üzere veri madenciliği için gerekli olan birçok verinin birden fazla kaynaktan sağlanması veri ambarlarının oluşturulmasını zorunlu kılmaktadır. Veri ambarı ana kaynaktan farklı olarak diğer kaynaklarla ortak ilişkisel alanları dikkate alır. Veri ambarı içerisindeki veriler temizlenmiş ve dönüştürülmüş operasyonel verilerden oluşmaktadır. Veri ambarı oluşturmak için kütüphane karar vericilerinin gözleminde verilerin seçilmesi gerekmektedir.
- Temizlenme, düzenleme ve yeniden oluşturmadan sonra prosedür otomatik bir şekilde devam ettirilmelidir.
- Veri ambarının yapılandırılması: Veri ambarının oluşturulması ve yapılandırılması veri madenciliği süreci içerisinde en fazla zaman ayrılması gereken bölümlerdir. Bu sürecin defalarca tekrarlanması gerekebilir. Ancak bir kez oluşturulduktan sonra,

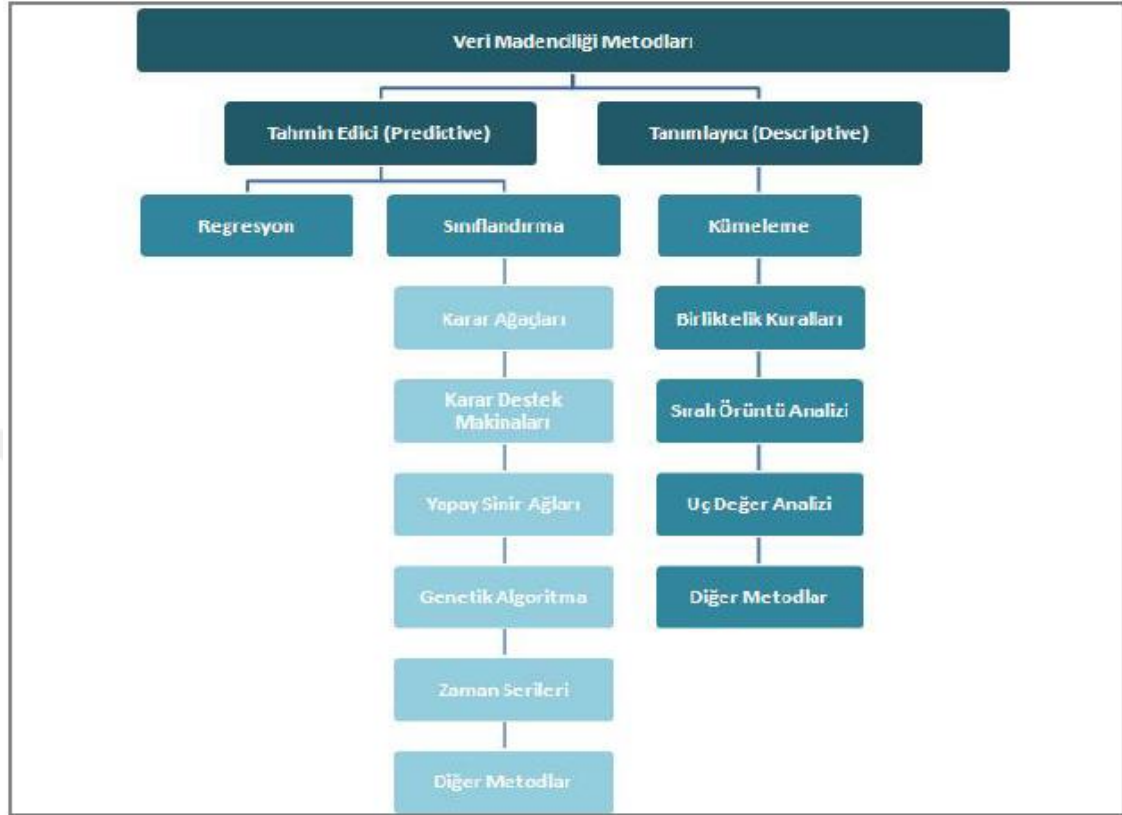
ilerideki veri madenciliği çalışmaları için bir model oluşturulabilir ve sonraki çalışmalarda zamandan tasarruf edilmesini sağlayabilmektedir. Oluşturulan algoritmalar birçok defa kullanılabilir.

- Uygun veri madenciliği araçlarının belirlenmesi; veri ambarı oluşturulduktan sonra analiz aşamasına geçilebilmektedir. Veri ambarı üzerinden geleneksel istatistiksel raporlar oluşturulabileceği gibi, bu istatistiksel verilerin ardında gizlenen ilginç ve faydalı örüntüler de tespit edilebilmektedir. Bu örüntüler kütüphane karar vericileri için anlamlı veriler elde etmesinde önemli bir kaynaktır. Örneğin kullanıcıların kütüphane kullanım yoğunlukları tespit edilerek kütüphane çalışanları için verimli çalışma saatleri belirlenebilir ya da kitap sirkülasyonu göz önüne alınarak kitap yerleşim düzeni tekrar oluşturulabilir.
- Analiz ve Uygulama: Analizler gerçekleştirilerek raporlar ve karar verme modelleri ortaya konulduktan sonra onaylanmalıdır. Veri madenciliği sonuçlarının onaylanması kütüphane karar vericileri tarafından gerçekleştirilmelidir. Sonuçların ortaya konulması ve uygulanabilirliğinin test edilmesi modellerin geliştirilmesi için önemli bir veridir. Oluşturulan örüntüler karar verici için bir örneklem içerisinde gerçekçi bulunmadığı takdirde, uygunsuz örüntülerin neden oluştuğu derinlemesine incelenmeli ve modellemeler birçok defa test edilmelidir. Son adım olarak model ve örüntüler tüm kütüphane sisteminde uygulanmalıdır.
- Değişkenlerin bir süre daha izlenmesi, modellerin güçlendirmesi sistemin kusursuz işleme için yararlı bir hareket olacaktır. Model değişkenliğini korumalıdır.

2.8. Veri Madenciliği Metotları ve Algoritmalar

Veri madenciliği metotları, tanımlayıcı ve tahmin edici olacak şekilde iki ana kategoriye ayrılabilir. Tahmin edici olan modellerde, sonuçları bilinen veriler ele alınarak matematiksel bir model geliştirilmesi ve elde edilen bu model yardımıyla sonuçları bilinmeyen yani araştırılan veri setleri için sonuçların tahmin edilmesi amaçlanmaktadır. (Kaya ve Köymen, 2008).

Tanımlayıcı modellerde ise farklı olarak amaçlanan veri setinde gizli olarak bulunan örüntülerin ortaya çıkarılması yani keşfedilmesidir.



Şekil 2.2. Veri Madenciliği Metodları (Kaya ve Köymen, 2008).

2.8.1. Regresyon ve sınıflandırma modelleri

Regresyon analizi, bir veya daha fazla bağımsız değişken ile bağımlı değişken arasındaki ilişkiyi modellemek için kullanılır. Veri madenciliğinde bağımsız değişkenler zaten bilinen niteliklerdir, istenen ise bağımlı değişkenleri tahmin etmektir. (Baradwaj ve Pal, 2011). Diğer bir deyişle söz konusu analizde, bir bağımlı değişkenin diğer bağımsız değişkenlere olan bağımlılığı, bağımlı değişkenin ana kütle ortalama değeri, bağımsız değişkenin tekrarlanan örneklerdeki bilinen ya da sabit olan değerleri türünden tahmin etmeye veya kestirmeye çalışılır (Damodar, 2001).

İki ya da daha fazla değişken arasında bulunan lineer ilişkiyi açıklamakta doğrusal regresyon modeli kullanılmaktadır. Açıklanan değişken bağımlı değişken olarak adlandırılırken, açıklayıcı değişkenlere ise bağımsız değişken adı verilir. Örneğin, bir

firmanın satış rakamları ve reklam giderleri arasındaki ilişki, öğrencilerin kütüphane kullanım oranı ile başarıları arasındaki ilişki gibi aralarında ilişki bulunduğu düşünülen durumları açıklamak için regresyon modelleri kullanılabilir (Damodar, 2001).

Bağımlı değişken olarak sürekli değişkenleri temel alan doğrusal regresyon modeli yanında, kategorik değişkenlerin tahmini için farklı metotlar üretilmiştir. Lojistik regresyon, bağımlı değişkenin birden fazla sayıda kategorik durum içerdiği, bağımsız değişkenlerin ise yapısal olarak kategorik veya sürekli olduğu durumlarda bağımlı ve bağımsız değişkenler arasındaki ilişkiyi incelemektedir. Lojistik regresyon analizi veya incelemesi iki farklı türde olmaktadır. Hangi türde olduğunu ise bağımlı değişkenin ait olduğu kategori belirlemektedir. Bağımlı değişkenin iki kategoriye sahip olduğu durumda ikili lojistik regresyon, ikiden daha fazla kategoriye sahip olduğu durumda ise çoklu lojistik regresyon adı altında incelenebilmektedir (Damodar, 2001).

Karar ağaçları çok sayıda kayıt içeren bir veri kümesini, bir dizi karar kuralları uygulayarak daha küçük kümelere bölmek için kullanılan bir yapıdır. Karar ağaçları, tek bağımlı değişken ve çok sayıda bağımsız değişkene sahiptirler. Bununla birlikte, eldeki veri setinden regresyon modellerine göre daha farklı ve kullanışlı olabilecek örüntüler ortaya çıkarırlar. Bazı durumlarda bağımlı değişken kategorik halde bulunmaktadır. Böyle durumlarda karar ağaçları, lojistik regresyona alternatif oluşturabilecek bir yöntemdir (Cox, 2002).

CHAID, C&RT en popüler karar ağacı algoritmalarından olup her iki algoritma da sürekli ve kategorik bağımlı ve bağımsız değişkenler ile çalışabilmektedir. CHAID'in çoklu, C&RT'nin ise ikili dallanma yapıları ve dallanma için kullandıkları değişken belirleme şekli söz konusu iki algoritma arasındaki en büyük farktır. CHAID, dallanma yaparken χ^2 ve F testi gibi istatistiksel ölçüler kullanılır; C&RT, χ^2 dışında, safsızlık ölçütü olan Gini İndeksini de kullanabilir. Bununla birlikte, her iki ağaç türünde de ulaşılan sonuçlar birbirine yakın olabilir (Özkan, 2008).

Bir diğer sınıflandırma yöntemi olan karar destek makineleri, lineer ve lineer olmayan verilerin sınıflandırılması ile ilgilenmektedir. Bu yöntem, lineer olmayan bir haritalama yöntemi kullanarak orijinal veriyi sahip olduğundan daha yüksek boyutlara taşımaktadır. Taşıdığı bu yeni ve yüksek boyutta, verileri sınıflandırmak için lineer ayırıcı düzlemler

araştırır ve en iyi orandaki düzlemi yakalamayı amaçlar. Yeterli derecede yüksek boyutta bulunan farklı sınıflara ait veriler uygun bir haritalama yöntemi ile daima ayırıcı bir düzlem tarafından ayrılırlar. Buna göre algoritma, belirtilen düzlemi bulabilmek için, destek vektörleri olarak da adlandırılan eğitim kümesi verilerini ve bu vektörler tarafından tanımlanmış mesafeleri kullanır. (Han ve Kamber, 2006).

Zaman Serileri ise zaman değişkeniyle ilişkili bir değişken hakkında, hafta, ay, mevsim veya yıl bazında elde edilen gözlem değerlerini zamana göre sıralanmış şekilde gösteren seriler olarak tanımlanmaktadır. Buradan hareketle, serilerin geçmiş ve bugünkü değerlerinin kullanılmak suretiyle gelecek dönem hakkında tahminler yapılabilmesinin zaman serileri analizinin konusunu oluşturduğu ifade edilebilir (Han ve Kamber, 2006).

Yapay sinir ağları yöntemi, insan beyninin yapısından yola çıkılarak tasarlanmış olup, örüntü tanıma ve hatanın en aza indirgenmesi üzerine kurulmuş bir yöntemdir. Bilgiyi içeri almak suretiyle hafızasında saklayan, farklı deneyimlerde yeni şeyler öğrenen ve veriler arasındaki ilişkiyi ortaya çıkaran bir yapıyı temsil etmektedir. Sinir sisteminin çalışma şekliyle hareketle oluşturulan yapay sinir ağları yönteminde, nöronları içeren sinir hücreleri bir araya gelerek sinir ağını oluşturmuştur (Rud, 2001).

Genetik algoritmalar yöntemi, bir fonksiyonun optimize edilmesi veya ardışık değerlerin tespit edilmesini içine alan problemlerin çözümü için çalışmaktadır. Genetik algoritmalar, doğal seçim ve en iyinin korunumu ilkelerine dayanırlar. Benzetim yoluyla bilgisayarlara uygulanan ve bilgisayar üzerinde oluşan bir evrim şeklidir. Genetik algoritmaların amacı, problemleri çözerken aynı zamanda evrimsel sistemleri de modellemektir (Gülten ve Doğan, 2008).

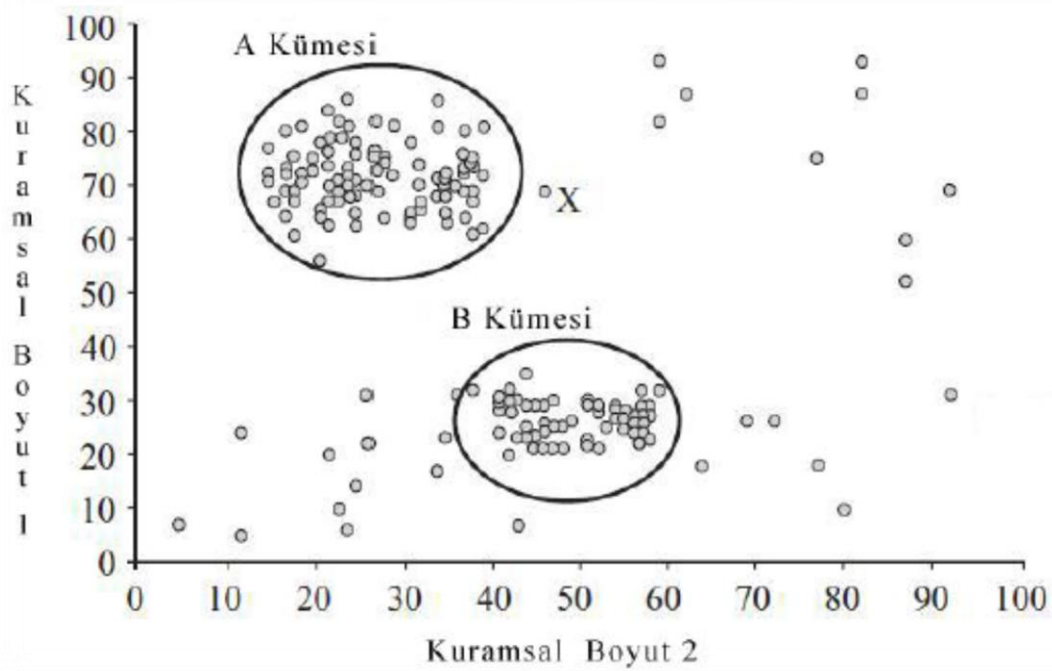
2.8.2. Kümeleme

Kümeleme, denetimsiz bir öğrenme yöntemi olup birbirine benzeyen verileri sınıflara ayırmak suretiyle verileri fiziksel olarak gruplara ayırma veya özetleme sürecidir. Kümeleme yöntemi, büyük veri kümelerinin anlamlı şekilde parçalanıp alt kümelerine ayrıldığı ve birbirine benzer grupların bir arada ele alınmasını ifade eden bir süreçtir. Kümeleme yönteminde herhangi bir çıktı değişkeni bulunmamaktadır. Bu nedenle

kümeleme yöntemi denetimsiz öğrenme metodu olarak da bilinmektedir (Ohsuga ve Hu, 2005).

Öte yandan, kümeleme analizinde küme sayısı bilinmemekte ve yapılan analiz sonucunda veri kümesinden elde edilmektedir. Ayırma işlemi tamamen bahse konu verilerin özellikleri kullanılarak yapılır. Bundan dolayı kümeleme yönteminde herhangi bir fonksiyon elde edilerek daha sonrasında bu fonksiyonun diğer veriler için kullanılma durumu bulunmaz (Han ve Kamber, 2001).

Çok boyutlu uzay içerisinde birbirlerine yakın olan bireylerin oluşturdukları birlik olarak ifade edilebilecek olan küme, uzaklık ve benzerlik kavramlarını çağrıştırmaktadır (Doğan, 2000). Kümeleme yönteminde, verilerin normal dağılım şeklinde dağılması gerektiği varsayımı bulunmaktadır. Ancak bu varsayım prensipte kalmakta, uzaklık değerlerinin normallığı yeterli görülmektedir (Tatlıdil, 2002).



Şekil 2.3. Kümeleme (Myatt, 2007)

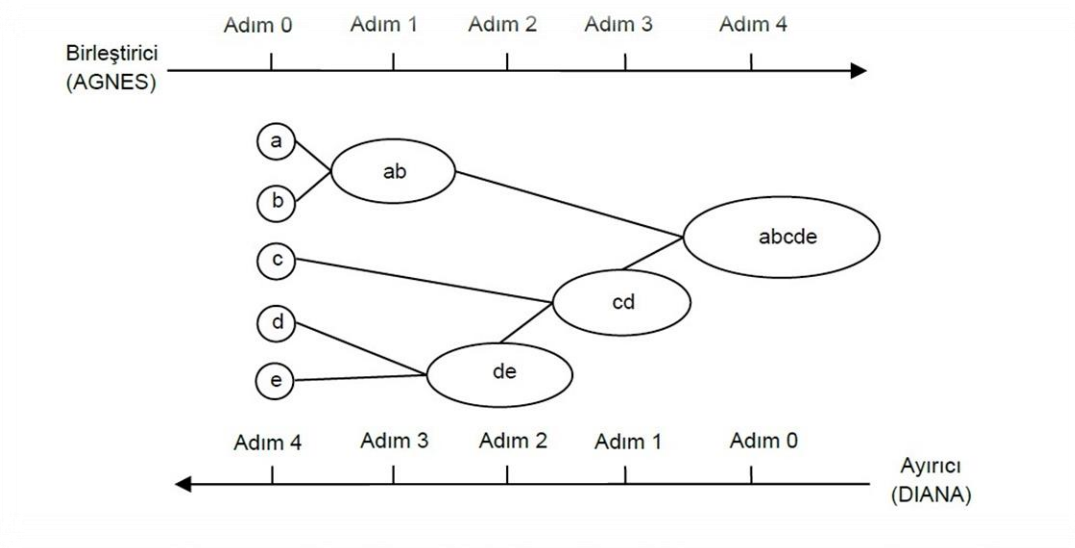
Kümeleme işleminde süreç, her bir verinin farklı özelliklerini belirten değerlerinin, diğer bir veriden farklı ya da o veri ile aynılığına göre çalışmaktadır. Farklılık ya da aynılık durumları kümeleme algoritmaları için uzaklıkları ifade etmektedir. Kümeleme

algoritmaları farklı uzaklık türlerini kullanabilirler. Bu uzaklık türleri öklid uzaklığı, minkowski uzaklığı, manhattan uzaklığı, canberra uzaklığı olarak ifade edilebilir (Özkan, 2008). Bahsedilen bu uzaklık türleri ile çalışan bilinen kümeleme algoritmaları Şekil 2.4.’deki gibidir (Han ve Kamber, 2001).



Şekil 2.4. Kümeleme Yöntemleri

- Aşamalı (Hiyerarşik) Yöntemler: Önce “gruplara ayırma ve sonra kümeleme” veya “tüm veri kümesini aynı kümeye atayıp sonrasında ayırma” gibi çeşitleri vardır. BIRCH algoritması en bilinen algoritmalar birisidir. Yüksek kapasitedeki verilerin kümelmesi için tasarlanmış olan BIRCH algoritmasında, artan halde bulunan ve sıradüzensel bir teknik kullanılmaktadır.



Şekil 2.5. Hiyerarşik Kümeleme (Han ve Kamber, 2001)

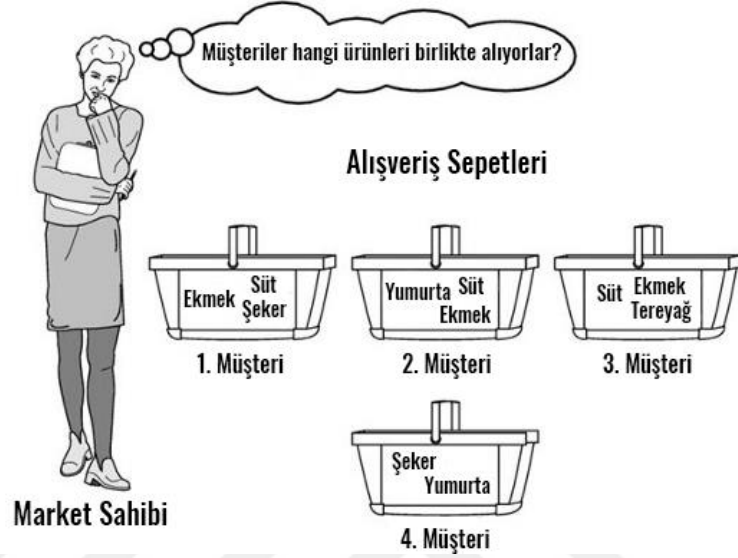
- Model (Bazlı) Yöntemler: Model tabanlı yöntemler her bir küme için bir model varsayımında bulunurlar ve veri kümesine en uygun olan modeli bulmaya çalışırlar. Bu yöntemlere beklenti maksimizasyonu olarak da bilinen EM ve COBWEB örnek olarak gösterilebilecek algoritmalarıdır.
- Yoğunluk Bazlı Yöntemler: yoğunluk tabanlı modeller, genellikle küresel şekilleri bulmada iyi olan ve uzaklıklar ile çalışan yukarıdaki yöntemlerin aksine düzensiz şekilleri tespit etme özelliğine sahiptir. DBSCAN ve OPTICS tipik yoğunluk tabanlı algoritmalarıdır.
- Grid (Izgara) Bazlı Yöntemler: Nesne uzayını, sonlu sayıda hücrenin bulunduğu bir uzaya indirgeyerek işlem yapmakta olan bu yöntemlere STING algoritması örnek olarak verilebilir.
- Bölümlemeli (Paylaşırma) Yöntemleri: Veri kümesini k sayıda kümeye ayırır; her bir kümede en azından bir veri bulunması gerekir ve her bir veri sadece bir kümede olmalıdır. Küçük ve orta büyüklükteki veri setlerinde bulunan küre şekilli kümeleri bulmada önemli bir yer sahibidirler. K-ortalamlar ve k-medoidler algoritmaları bu yöntemlerdeki en bilinen algoritmalarıdır.

Farklı sektörlerde birçok amaçla kullanılmakta olan kümeleme yöntemlerinin en yoğun kullanımlarından biri pazarlama sektöründe olup müşterilerin farklı gruplara ayrılarak oluşturulan her farklı grup için farklı pazarlama stratejilerinin geliştirildiği durumlardır (Özkan, 2008). Kütüphane kullanıcılarının gruplanarak farklı kitap tavsiyeleri oluşturulması işlemi de bir kümeleme uygulaması olarak düşünülebilir.

2.8.3. Birliktelik kuralları ve sıralı örüntü analizi

Veri madenciliğinin tanımlayıcı modellerinden biri olan birliktelik kuralları büyük miktarlardaki veriler arasından ilginç birliktelik örüntülerini keşfederek pazarlama, karar verme ve iş yönetimine fayda sağlamayı amaçlamaktadır. Bu yüzden veri tabanlarında bilgi keşfinde yapılan araştırmalarda birliktelik kuralları konusu genel olarak merkezi noktada bulunmaktadır (Han ve Fu, 1995). Birliktelik kuralları kullanışlı ve anlaşılması kolay olduğundan finans, telekomünikasyon, pazarlama, perakendecilik ve çevrimiçi

(online) ticaret gibi endüstriyel alanlarda geniş bir alana yayılmıştır. Bununla birlikte bir olayın gerçekleşmesinden sonra bir diğerinin gerçekleşiyor olması gibi durumlardaki ardışıklıkların tespit edilmesi durumu ise sıralı örüntü analizini ifade eder (Chen vd., 2006).



Şekil 2.6. Alışveriş Sepeti Analizi (Han ve Kamber, 2001)

Bir diğer adı pazar sepet analizi olan birliktelik kuralları analizi başlangıçta bilgisayar bilimleri alanında geliştirilmiş olup, çoğunlukla alışveriş işlemlerinde müşteriler tarafından alınan ürünlerin birlikteliklerini ölçümlemek veya internet sitelerine erişim sağlayan kişilerin sıklıkla ve sırayla ziyaret ettikleri sayfaları bulmak gibi değişik analizlerde kullanılabilirler. Amaç, işlemsel kayıtların bulunduğu bir veri kümesinde tipik olarak birlikte olan kalemlerin tespitini gerçekleştirmektir (Guidici, 2003).

Veriler arasındaki bu birliktelikleri kurallar şeklinde açıklamak için eğer-sonra ifadeleri kullanılır ve;

“Eğer < bazı şartlar sağlanırsa > sonra < bazı niteliklerin değerlerini tahmin et >”

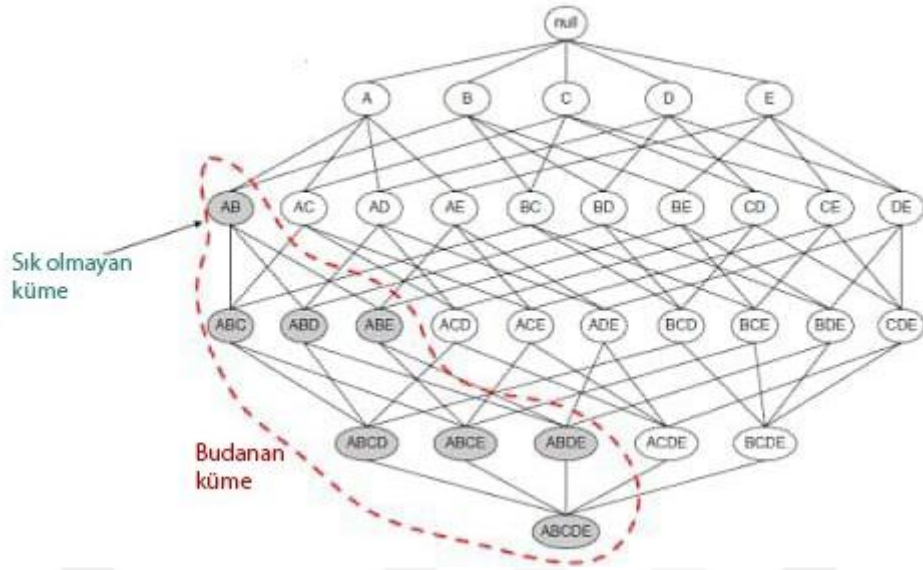
şeklinde belirtilebilir. İfadenin ilk yani eğer bölümündeki durumlar öncül, ikinci yani sonra bölümündeki durumlar ise sonuç olarak adlandırılır. Öncül ve sonuç durumları X ve Y olarak ele alınırsa buradaki ilişki yani birliktelik kuralı $X \rightarrow Y$ şeklinde ifade edilebilir (Emel vd., 2005).

Destek ve güven kavramları, bu analizde önemli olan iki kavramdır. Destek, yukarıda belirtilen kural cümlesini sağlayan olayların birlikte gerçekleştiği durumların tüm durumlara oranı iken, güven, kural cümlesini sağlayan olayların birlikte gerçekleştiği durumların öncül bölümde gerçekleşen olaya oranıdır (Emel vd., 2005).

Sıralı örüntü analizinin çalışma sürecinde ardışık olma durumu bulunmaktadır. Bahse konu şekilde, günümüz internet ortamında arka arkaya ziyaret edilen web sitelerinin analiz edilmesi sonucu ortaya çıkabilecek farklı kurallar bütününe sıralı örüntü analizi olduğu söylenebilir. Bu analiz türü yaygın şekilde de kullanılmaktadır. Öte yandan, bir değişkenin iki yönlü olabilmesi, bazı durumlar için girdi, bazıları için çıktı değişken gibi tanımlanabilmesi, birliktelik kuralları ve sıralı örüntü analizinin sınıflandırma algoritmalarından farkı olarak ifade edilebilir. Birliktelik analizinde kullanılan bazı algoritmaları Apriori, Predictive Apriori, FPGrowth, Filtered Associator, Generalized Sequential Patterns, Tertius biçiminde sıralamak mümkündür (Doğan, 2015).

2.8.3.1. Apriori algoritması

Birçok farklı yöntem kullanılarak, sık geçen nesne-kümeler madenciliğinin hesap karmaşıklığı azaltmak mümkündür. Ancak bu alanda en bilinen yöntem Apriori algoritmasıdır. Apriori algoritması bazı aday nesne-kümelerin destek değerlerini saymadan bu adayların elenmesi ile etkili bir çözüm sunar. Algoritmanın ismi, sık geçen nesne-kümelerin madenlenmesinde önsel (prior) bilgiyi kullanmasını temel almıştır (Agrawal ve Srikant, 1994).



Şekil 2.7. Apriori Algoritması (Han ve Kamber, 2001)

Eğer bir nesne-küme sık geçen ise bu kümenin bütün altkümeleri de sık geçen nesne-küme olmalıdır. Apriori algoritması seviye mantığı arama olarak bilinen yinelemeli bir yaklaşım kullanır. Bu yaklaşımda k ögeli nesne-kümeler $(k-1)$ ögeli nesne-kümelerin birleştirilmesiyle oluşturulur. İlk olarak 1 ögeli nesne-kümeleri oluşturmak için nesnelerin destek değerleri hesaplanır, $(k-1)$ 'inci geçişte bulunan sık öğeler (L_{k-1}), k 'nıncı geçişte sık geçen nesne aday kümesini (C_k) oluşturmak için kullanılır. Veri tabanı taranır ve C_k 'ların destek değerleri hesaplanır ve L_k 'lar oluşturulur. Bu şekilde algoritma sık geçen nesne-küme bulamayınca kadar devam eder (Agrawal ve Srikant, 1994). L_{k-1} 'in L_k 'nın elde edilmesinde nasıl kullanıldığını anlamak için algoritmanın Birleştirme ve Budama işlemlerini kapsayan adımlar aşağıda anlatıldığı gibidir (Han ve Kamber, 2006).

Birleştirme adımı sürecinde, L_k 'yı bulmak için k ögeli aday nesne-kümeler (C_k), L_{k-1} 'in kendi arasında birleştirilmesiyle oluşturulur. I_1 ve I_2 , L_{k-1} 'de bulunan nesne-kümeler olsun. $I_i[j]$ gösterimi, I_i 'deki j . nesneyi temsil eder (örneğin, $I_1[k-2]$, I_1 'deki sondan 2. nesneyi temsil eder). Apriori, bir işlemdeki nesnelerin veya nesne-kümelerin alfabetik sıraya göre sıralandığını varsayar. $k-1$ ögeli nesne-kümeler için nesneler $I_i[1] < I_i[2] < \dots < I_i[k-1]$ şeklinde sıralanmıştır. L_{k-1} 'in ilk $(k-2)$ nesneleri ortak ise $L_{k-1} \bowtie L_{k-1}$ birleşmesi gerçekleştirilebilir (Agrawal ve Srikant, 1994).

Budama Adımı sürecinde C_k 'nın elemanları sık geçen olabilir veya olmayabilir fakat bütün sık geçen nesne-kümeleri içerir ve L_k 'nın üst kümesidir. C_k 'daki her bir adayın destek değerinin hesaplanması için yapılan taramanın sonuçları L_k 'yı belirler (örneğin, minimum destek değerinden büyük ve eşit değere sahip bütün adaylar sık geçen nesne-küme olarak tanımlanır ve L_k 'da yer alır). C_k çok büyük boyutta olabilir ve çok ağır hesaplamalar gerektirebilir. C_k 'nın boyutunu azaltmak için Apriori özelliği kullanılır. Herhangi bir sık olmayan $(k-1)$ -nesne-küme sık geçen k -nesne-kümenin altkümesi olamaz. Bundan dolayı, aday k -nesne-kümenin bütün $(k-1)$ ögeli altkümeleri L_{k-1} 'de yoksa bu aday sık geçen nesne-küme olamaz ve C_k 'dan çıkarılır (Agrawal ve Srikant, 1994).



3. MATERYAL ve YÖNTEM

3.1. Araştırmanın Amacı ve Verinin Elde Edilmesi

Bu araştırmanın amacı Erzincan Binali Yıldırım Üniversitesi Merkez Kütüphanesine ait verilerin veri madenciliği yöntemleri ile incelenerek, kütüphane karar vericileri için yol gösterici olabilecek anlamlı örüntüler üretmeye çalışmaktır.

Bu amaç için Erzincan Binali Yıldırım Üniversitesi Merkez Kütüphanesine ait 2010 ile 2017 yılları arasındaki veriler Etik Onayı alınarak temin edilmiştir. Elde edilen veriler ön işleminden geçirilerek öncelikle temizleme ve dönüştürme işlemlerine tabii tutulmuş, sonrasında ise veriler üzerinde birliktelik analizi ve kümeleme analizi algoritmaları çalıştırılarak ulaşılan sonuçlar yorumlanmıştır.

Bu çalışma esnasında Microsoft firmasına ait Excel ve Access programları ile IBM firmasına ait SPSS Modeler programı kullanılmıştır. Excel dosyası üzerinden alınan veriler Access veri tabanına aktarılarak, verinin temizlenmesi ve dönüştürülmesi aşaması veri tabanı üzerinden gerçekleştirilmiştir. Ayrıca SPSS Modeler programı ikinci bölümde anlatılan veri madenciliği tekniklerinden kümeleme algoritması ile birliktelik kuralları ve sıralı örüntü analizi için kullanılmıştır.

3.2. Verinin Temizlenmesi ve Dönüştürülmesi

Erzincan Binali Yıldırım Üniversitesi Merkez Kütüphanesine ait veriler 3 ayrı Excel dosyası şeklinde alınmıştır. Tablo 3.1' de alınan dosyaların genel yapısı verilmiştir.

Tablo 3.1. Kütüphaneye ait kayıtları içeren dosyalar

Dosya Adı	İçerik	Sütun Sayısı	Kayıt Sayısı
Katalog.xlsx	Kütüphanede mevcut olan yayınların yayın tarihleri, yazarları, vb. gibi temel bilgiler	7	109.840
Ödünç.xlsx	Kütüphaneden yapılan ödünç alma işlemlerinin yayın bilgileri, üye bilgileri, ödünç tarihleri vb. gibi bilgiler	19	59.953
ÜyeTablosu.xlsx	Kütüphaneye üye olan kullanıcılara ait bölüm, unvan, adres vb. gibi bilgiler	11	30.529

Excel dosyası formatında alınan veriler incelendiğinde amaçlanan veri madenciliği çalışması için tüm sütun bilgileri değerlendirilmiş ve veri ön işleme aşamasında çalışma kapsamında oluşturulan modellere uygun olan alanlar belirlenmiştir. Bu belirleme aşamasında, tablonun ilgili alanında bulunan verinin veri madenciliği algoritmaları için uygunluğu, kayıtların doluluk oranı ve kişiye özgü olup gizlilik ihlali içerip içermeyeceği gibi hususlar göz önünde bulundurulmuştur. Daha sonra veriler Microsoft Access programında oluşturulan veri tabanı içerisindeki 3 ayrı tabloya aktarılmıştır.

Yapılan veri ön işleme çalışmasının diğer bir sonucu olarak katalog tablosunda bulunan kitapların tümünün ödünç alınmadığı, 26.205 tanesinin ödünç alındığı tespit edilmiştir. Bu sebeple veri madenciliği analizleri için, ödünç alınan ve alınmayan olmak üzere tüm kayıtlar incelenmiştir. Üye bilgilerini içeren Üye tablosu incelendiğinde ise yine tüm üyelerin kütüphaneden ödünç alma işlemi yapmadığı, sadece 7.802 üyenin kitap ödünç alma işlemi yaptığı anlaşılmıştır. Veri madenciliği aşamasında ise sadece ödünç alma işlemi yapan üyelere ait veriler kullanılmıştır.

Veri madenciliği çalışmalarında verinin temizlenmesi ve dönüştürülmesi büyük önem arz etmektedir. Elde edilen verilerin temizlenme ve dönüştürülme süreci bu çalışmanın teorik kısımlarında açıklanan yöntemler ile gerçekleştirilmiştir.

Öncelikle kütüphanede mevcut olan yayınlara ait veriler için veri tabanında Katalog adında bir tablo oluşturulmuş ve Katalog.xlsx dosyasında bulunan kayıtlar bu veri tabanı tablosuna aktarılarak temizleme ve dönüştürme işlemleri gerçekleştirilmiştir.

Kayıtlar incelendiğinde her kaydın, yayın için belirlenen benzersiz bir kimlik numarasından, yayının tarihinden, ödünç alınma sayısından ve kütüphane içinde kullanılan sınıflama sistemine göre verilmiş sınıf kodundan oluştuğu görülmüştür. Bu sınıflama sistemindeki amaç kütüphanedeki bilgi kaynaklarına kolayca ulaşabilmek adına aynı konuda bilgi içeren bütün kitapların aynı sınıf altında bulunmasını sağlamaktır. Bu amaçla dünyada kullanılan sınıflama sistemleri BC (Bliss Bibliographic Classification), LCC (Library of Congress Classification), CLC (Chinese Library Classification), CC (Colon Classification) ve DDC (Dewey Decimal Classification) olarak sıralanabilir (Aksoy, 2015). Erzincan Binali Yıldırım Üniversitesi Merkez Kütüphanesinde ise yukarıda sayılan bu sınıflandırmalardan en sonuncusu olan ve kısaca dewey sınıflama olarak adlandırılacak olan DDC (Dewey Decimal Classification) kullanılmaktadır. Bu sınıflandırma ABD’li kütüphaneci Melvil Dewey tarafından geliştirilmiş olup Dewey Onlu Sınıflama Sistemi olarak ifade edilebilir (Aksoy, 2015). Bu sistemde kitaplar konularına göre gruplandırılmakta ve bu konular sayısal ifadelerle tanımlanmaktadır. Sistemde 10 ana kategori bulunmaktadır. Bu ana kategoriler de kendi içerisinde alt kategorilere ve konulara yine sayısal ifadelerle bölünmektedir. Tablo 3.2’de Dewey Onlu Sınıflama Sisteminin içeriği sunulmaktadır.

Örnek verilecek olursa, Dewey Sınıflama Sisteminde 400 sınıflama numarası dil ve dilbilim alanındaki eserleri genel olarak tanımlamaktadır. 4 rakamının yanına başka bir sayının eklenmesiyle birlikte genelden ayrıntıya doğru alt kümeler ortaya çıkmaktadır. Örneğin 410 sayısının temsil ettiği alt küme içinde Türk dili ilgili eserler, 413 sayısının temsil ettiği alt küme içinde ise Türkçe ile ilgili sözlükler bulunur. Numaraların hangi konuları temsil ettiğini bilmek kütüphanede araştırma yapanlar için oldukça yararlıdır.

Tablo 3.2. Dewey onlu sınıflama sistemi

DEWEY Onlu Sınıflama Sistemi

0	Genel Konular
100	Felsefe ve Psikoloji
200	Din
300	Toplum Bilimleri
400	Dil ve Dil Bilim
500	Doğa Bilimleri ve Matematik
600	Teknoloji (Uygulamalı Bilimler)
700	Sanat Güzel Sanatlar)
800	Edebiyat ve Retorik
900	Coğrafya ve Tarih

Veri tabanı içerisindeki mevcut kitapların yayın tarihi bilgilerini içeren alan incelendiğinde bu verilerin elle girildiği ve bu nedenle çok fazla hata yapıldığı gözlenmiştir. Bu sebeple SQL komutları yardımıyla anlamlı bulunmayan yayın tarihi bilgilerini içeren tüm kayıtlar silinmiştir. Eser adı ve yazar gibi ne nümerik ne de kategorik olan veriler ise bu tezin kapsamında bulunmadığından incelenmemiştir. Tez kapsamında özellikle istatistikî sonuçlar sunabilen veri alanları incelenmiştir. Ödünç sayısı alanı incelendiğinde ise veri girişi yapılmayan alanlar olduğu gözlenmiştir. Bu kayıtlar ise yukarıda bahsedildiği gibi kitapların ödünç alınmadığını işaret etmektedir.

Kütüphane Otomasyonunda yayın ödünç alma işlemlerinin tutulduğu ödünç tablosu incelendiğinde ödünç tablosunun tamamen katalog ve üyeler tablolarından oluşturulduğu gözlemlenmiştir. Yine bu tabloda da diğer tablolarda olduğu gibi çok sayıda hatalı ve boş kayıta rastlanmış olup tüm bu kayıtlar SQL sorguları vasıtasıyla temizlenmiştir. Bu tabloda bulunan üye cinsiyeti, fakültesi ve unvanı gibi nominal değerlere sahip alanlar kategorik değerlere dönüştürülerek SPSS Modeler programı için uygun hale getirilmiştir. Bahsedilen bu dönüştürme işlemi ile üyelerin cinsiyet bilgileri için 01 kodu erkek üyeler; 02 kodu kadın üyeler; 03 kodu ise cinsiyet bilgisi belirtilmeyen üyeler için kullanılmıştır. Aynı şekilde üyelerin unvan bilgileri için 10 kodu akademisyenler; 12 kodu öğrenciler ve 20 kodu ise idari personel de dahil olmak üzere diğer üyeler için kullanılmıştır.

Kütüphane üyelerinin kayıtlı oldukları eğitim birimleri ise aşağıdaki Tablo 3.3’de belirtildiği şekilde bir dönüşüme tabi tutularak analize devam edilmiştir.

Tablo 3.3. Üyelerin kayıtlı oldukları birimler

uyeler_kodlama	
uye_birim_kodu	Birim adı
100	Diğer Birimler
200	Enstitüler
300	Eczacılık Fakültesi
310	Eğitim Fakültesi
320	Fen-Edebiyat Fakültesi
330	Hukuk Fakültesi
340	İktisadi ve İdari Bilimler Fakültesi
350	İlahiyat Fakültesi
360	Mühendislik Fakültesi
370	Tıp Fakültesi
400	Meslek Yüksekokulları

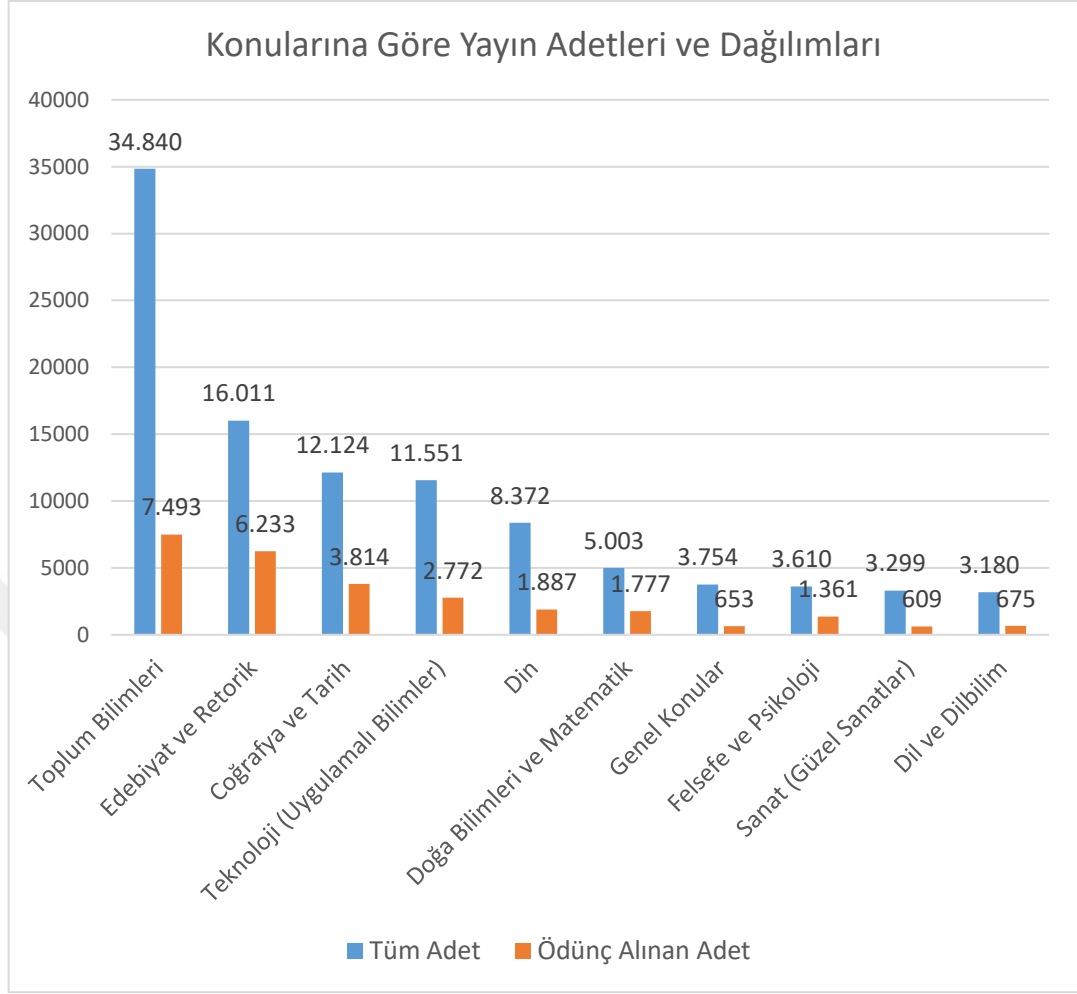
Bununla beraber üyeler tablosunun ilk haliyle kütüphaneden ödünç alma işlemi yapan ve yapmayan tüm üyeleri içerdiğinden ödünç işlemlerinin veri madenciliği ile analiz edilmesinde anlamlı bir sonuç ortaya çıkarmayacağından ötürü ödünç tablosu analizlere uygun hale dönüştürülmüştür. Tüm bu temizleme ve dönüştürme işlemlerinden sonra veri madenciliği sürecinde kullanılacak 3 adet veri tabanı tablosu elde edilmiş olup aşağıdaki çizelgede detaylandırılmıştır. Kütüphane ile ilgili işlemsel kayıtları içeren bu tablolara ek olarak yayın sınıflarını içeren Dewey Tablosu ve üyelere ait nominal bilgilerin kategorik değerlere dönüştürülmüş hallerini içeren üyeler kodlama tablosu da veri madenciliği algoritmalarında kullanılmıştır.

Tablo 3.4. Önişlemeden geçtikten sonra dosyaların durumu

Tablo Adı	İçerik	Alan Sayısı	Alan Adları	Kayıt Sayısı
katalog	Kütüphanede mevcut olan yayınlara ait bilgiler	4	Demirbaş No, Dewey Sınıfı, Yayın Tarihi, Ödünç Alınma Sayısı	101.744
odunc	Kütüphaneden yapılan ödünç alma işlemlerinin yayın bilgileri, üye bilgileri, ödünç tarihleri vb. gibi işlemsel kayıt bilgileri	9	İşlem No ve Ödünç Gün Sayısı alanları ile katalog ve üyeler tablosundaki diğer alanlar	59.156
uyeler	Kütüphaneye üye olan kullanıcılara ait bilgiler	5	Üye No, Üye Fakülte, Üye Unvan, Üye Cinsiyeti, Ödünç İşlemi Sayısı	7.802

Temizleme ve dönüştürme işlemlerinden sonra veri madenciliğine uygun hale getirilen veriler tanımlayıcı bulgular elde etmek adına analiz edilmiştir.

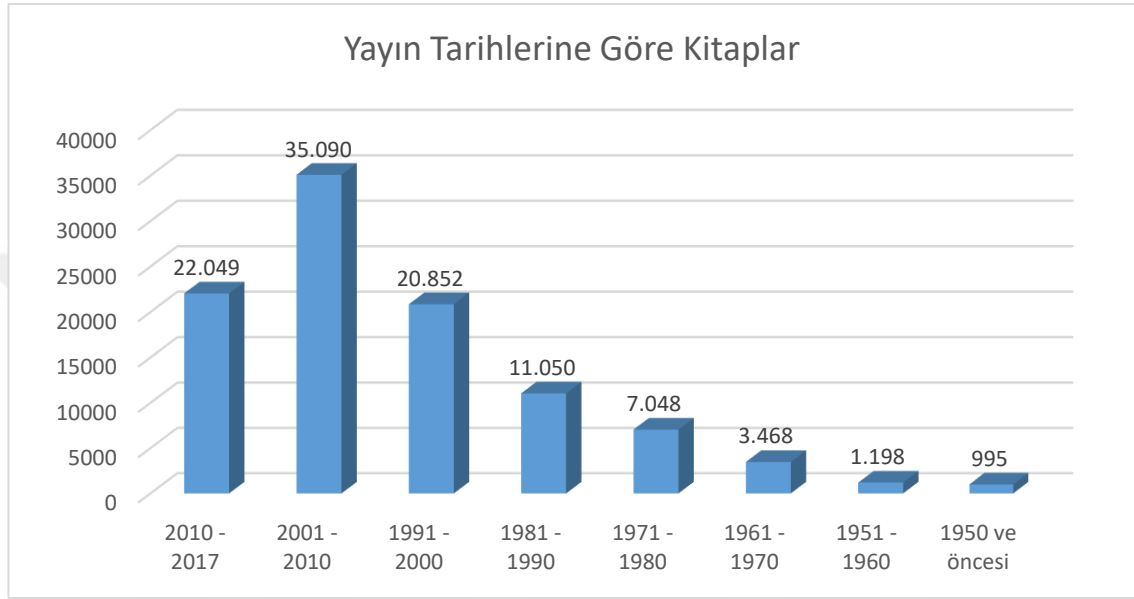
Çok büyük miktarda verinin içinden anlamlı bilgiler çıkarabilmek için yapılan veri madenciliği işleminde madencilik yapılan işin ve verinin temel yapısının kavranması oldukça önemlidir. Temel olarak 3 ayrı tablo şeklinde bulunan verilerden ilk olarak kütüphanede bulunan tüm kaynakların verilerini içeren katalog tablosu incelemeye tabii tutulmuştur. Şekil 3.1’de kütüphanede uygulanan kataloglama sistemi olan Dewey sistemine göre kütüphanede bulunan kaynakların dağılımı görülmektedir.



Şekil 3.1. Kütüphanede bulunan yayınların konularına göre dağılımları

Tablo konu başlıklarına göre kaynakların adetlerini ve o konu başlıklarından incelenen süre içerisinde ne kadar ödünç alındığını beraber göstermektedir. İçerisinde hukuk, ekonomi, eğitim, yönetim bilimleri vb. alt alanlarını bulunduran Toplum Bilimlerine ait kaynakların 34.840 adet ile diğerlerine göre çok daha fazla bulunmaktadır. Bu bilgi ise üniversitenin bu konuları içeren bölümlerinin daha önce faaliyete geçmiş olduğu bilgisi ile uyumludur. Toplum Bilimleri alanını sırası ile edebiyat, coğrafya ve tarih alanları ile uygulamalı bilimler alanları takip etmektedir. 3.180 adet kaynak ile dil ve dilbilim alanları ise tüm kaynakların % 3,12'lik kısmını kapsamakta olup, son sıradadır. Ödünç alınan adetlere ya da kullanım oranlarına bakıldığında % 38,9'luk kullanım oranı ile edebiyat ve retorik alanındaki kitaplar en yüksek ödünç alınma oranına sahip olup, bunu % 35,6'luk kullanım oranı ile Doğa bilimleri ve Matematik alanındaki kitaplar takip etmektedir. Bulunma sayısı açısından en yüksek değere sahip olan toplum bilimleri kaynaklarında ise

kullanım oranı % 21,51 ile yedinci sıradadır. Genel konular ana başlığına sahip kaynaklar ise kullanım oranı bakımından % 17,4 ile son sırada bulunmaktadır. Kütüphanede bulunan fiziki kaynaklar bulunma ve ödünç alınma sayıları açısından değerlendirdikten sonra yayım tarihleri açısından da değerlendirilmiştir ve sonuçlar Şekil 3.2’ de gösterilmektedir.

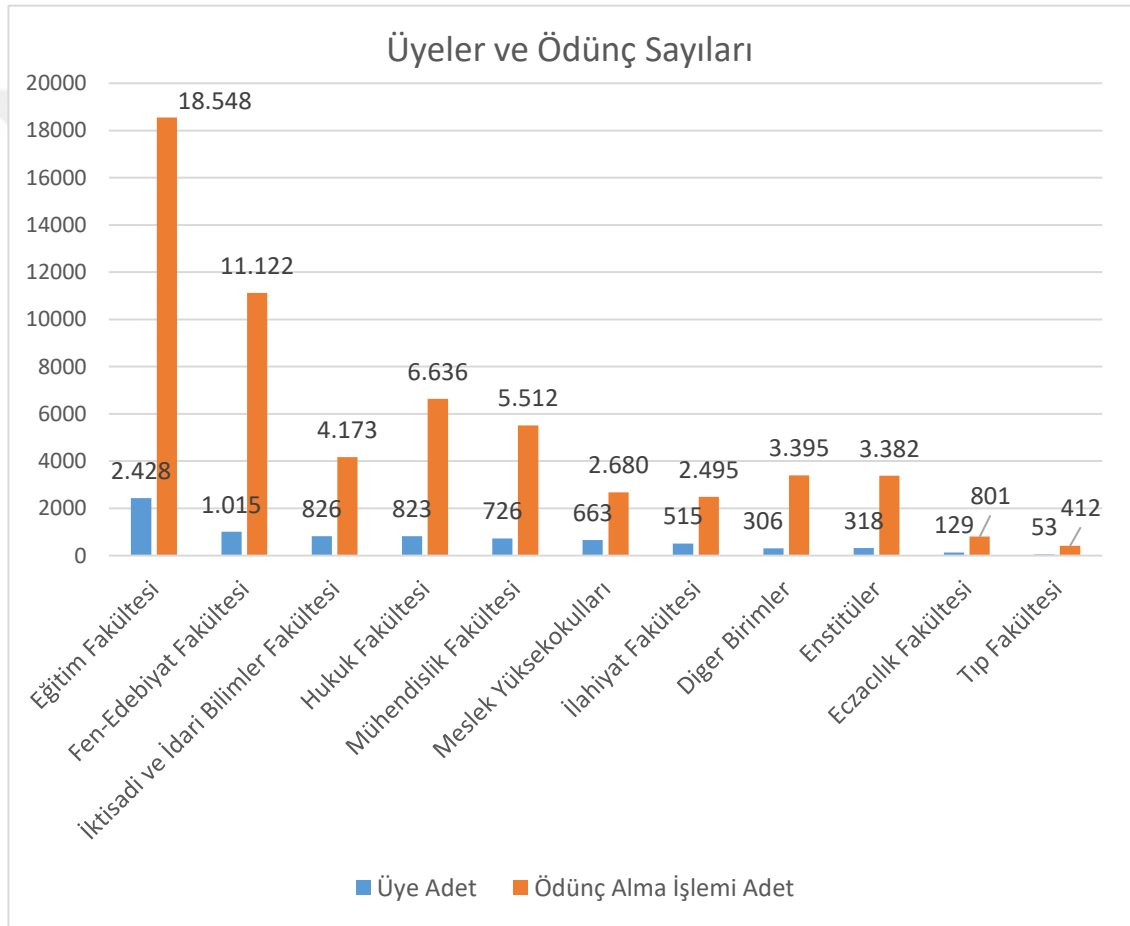


Şekil 3.2. Kütüphanede bulunan yayımların yayım tarihlerine göre dağılımları

Çalışma kapsamında elde edilen veriler 2017 yılına kadarki fiziki kaynaklara ait verileri kapsadığından dolayı yayımlanma tarihi 2017 yılından sonra olan bir kaynak mevcut değildir. Bu şekil incelendiğinde ise %34,49’luk oran ile 2001-2010 tarihleri arasında yayımlanan kaynakların ilk sırada olduğu 1950 ve öncesine ait yayım tarihi bulunan kaynakların ise %0,98’lik oran ile son sırada olduğu görülmektedir. Bu veriler ise kütüphanede bulunan kaynakların sürekli olarak güncel tutulmaya çalışıldığının bir göstergesidir.

Veriler kütüphanede bulunan kaynaklara ait fiziksel bilgileri içermesi ile beraber kütüphaneden yararlanan kullanıcılara ait bilgileri de içermektedir. Bu bilgiler ise yukarıda açıklanan üyeler adlı tabloda bulunmaktadır ve üyelere ait çeşitli bilgiler içermektedir. Aşağıdaki şekilde üyelerin bağlı buldukları eğitim birimlerine göre bulunma ve ödünç alma sayıları gösterilmiştir. Eğitim Fakültesi ve Fen Edebiyat Fakültesine kayıtlı üyeler kullanıcılar arasında en fazla sayıda bulunan üyelere dir. Ödünç

alma sayıları ile değerlendirilecek olduğunda bu iki grubun yine en fazla sayıda ödünç alma işlemi bulunan gruplar olduğu görülmektedir. Kullanım oranına bakıldığında ise idari birimleri içeren diğer birimlerin %11,09'luk oran ile ilk sırada geldiği görülmektedir. Şekilde meslek yüksekokulları ve ilahiyat fakültesi üyelerinde ödünç alma oranlarının %4,04 ve %4,84 ile en az oranlar olduğu görülmektedir. Daha çok lisansüstü eğitim veren enstitülerde ise üye sayısı 318 ile oldukça az olmasına rağmen kullanım oranı %10,64 ile üst sıralarda bulunmaktadır. Bu durum ise lisansüstü eğitim sürecinin çok daha fazla oranda araştırma gerektiren bir süreç olduğu bilgisıyla açıklanabilir.



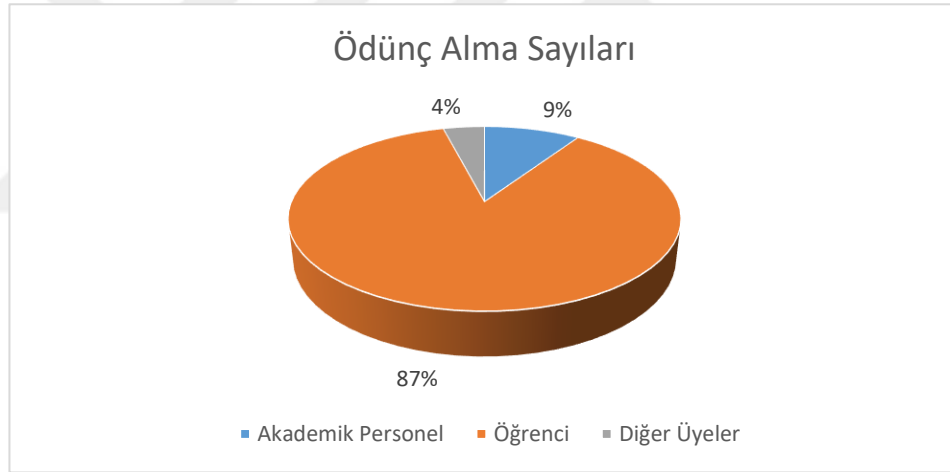
Şekil 3.3. Kütüphanede bulunan yayınların ödünç alınma sayıları

Üyelerin unvanlarına göre incelenmesi ise aşağıdaki tabloda görülmektedir. Tabloya göre kütüphaneye üye olan akademik personel sayısı 361 ile üye olan 7.252 adet öğrenciye göre oldukça az olmasına rağmen kullanım oranlarında durum değişmektedir. Üye başına ortalama ödünç alma sayısı olarak bakılırsa akademik personel grubu 15,43 ile öğrenci grubundan önce gelmektedir.

Tablo 3.5. Üyelerin ünvanlarına göre ödünç alma sayıları

Unvan	Üye Adedi	Ödünç Alma Sayıları	Oran
Akademik Personel	361	5.570	15,43
Öğrenci	7.252	51.189	7,06
Diğer Üyeler	189	2.397	12,68

Ödünç alma sayıları ise aşağıdaki şekilde daha net olarak görülmektedir. Akademik personel unvanlı üyeler tüm ödünç alma işlemlerinin % 9' luk bir bölümünü kapsamaktadır. Unvanı öğrenci olan üyelere ise bu oran % 87'dir. İdari personeli ve unvan bilgisi girilmeyen diğer tüm üyeleri içeren diğer üyeler grubunda ise üye başına ödünç alma oranı kullanıcı başına 12,68 değer ile tüm ödünç alma işlemlerinin %4'ünü kapsamaktadır.



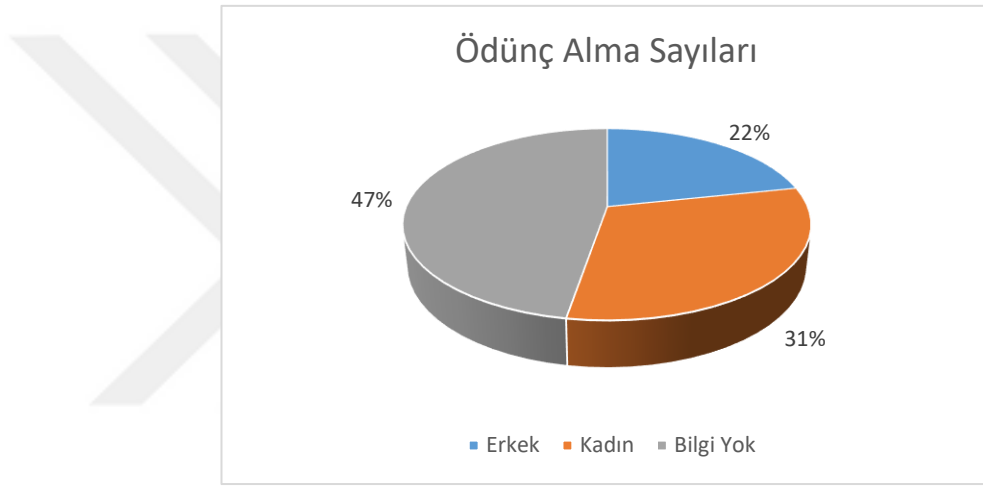
Şekil 3.4. Üyelerin ünvanlarına göre ödünç alma grafiği

Son olarak üyeler cinsiyetleri bakımından incelendiğinde 1.642 erkek ve 2.152 kadın üye olduğu görülmekte olup Tablo 3.6'da ödünç alma sayıları da bulunmaktadır. Dikkat edilmesi gereken bir husus olarak cinsiyet bilgisi içermeyen kayıtların içeren kayıtlardan daha fazla olduğu görülmektedir ve bu durum da kullanılan yazılımın cinsiyet verisini girmeyi zorunlu kılmadığını göstermektedir.

Tablo 3.6. Üyelerin cinsiyetlerine göre ödünç alma sayıları

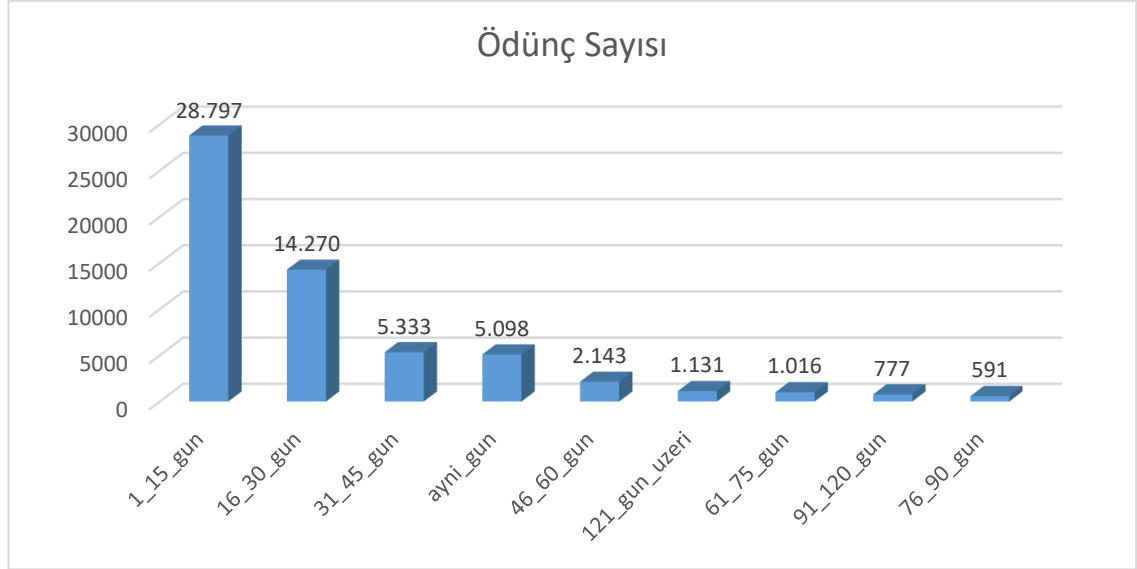
Cinsiyet	Üye Adedi	Ödünç Alma Sayıları	Oran
Erkek	1.642	12.820	7,81
Kadın	2.152	18.339	8,52
Bilgi Yok	4.008	27.997	6,99

Şekil 3.5’ de ise üyelerin ödünç alma sayılarının cinsiyetlerine göre gruplandırılması görülmektedir.



Şekil 3.5. Üyelerin cinsiyetlerine göre ödünç alma grafiği

Buradaki önemli bir nokta ise cinsiyetlerine dair bilgi bulunmayan üyelerin sayılarının fazlalığıdır. Veri madenciliği sürecinde bu çalışmanın teorik kısmında anlatıldığı gibi değer bulunmayan veri alanları temizlemeye tabii tutulur. Bu temizleme alanının tamamen veriden çıkarılması, boş alanların olduğu kayıtların veriden çıkarılması veya alanların çeşitli tekniklerle doldurulması olarak özetlenebilir. Alanların doldurulması durumu veri setimiz için uygun değildir. Tüm verinin veya alanın silinmesi ise veri bütünlüğünü bozacak kadar büyük boyutta olduğu için bu kayıtlara müdahale edilmemiştir.



Şekil 3.6. Ödünç alınan gün sayısı

Araştırmaya konu veri seti içerisindeki en önemli kısımlardan biri olan ödünç alma işlemlerine ait tablo incelendiğinde ise kayıtların ödünç alınan gün sayılarına göre dağılımı Şekil 3.6' da görüntülenmektedir.

Ödünç alınan gün sayısı grafiğinde, aynı gün teslim edenler 5.098 kişi ile % 8,6 oranındadır. Aynı gün teslim edenler grubunda yer alan kişilerin, kaynağın belli bölümlerinin fotokopisini çekirmek ya da kısa bir bölüme çalışmak için ödünç aldığı çıkarımı yapılabilir. Kaynağı ilk on beş gün içerisinde getiren 28.797 kişi ise % 48,6 oranı ile grafikteki en büyük çoğunluğu oluşturmaktadır. Bu sürenin iade yoğunluğu açısından değerlendirildiğinde kaynakların çalışılması için genel olarak ideal süre olduğu yorumu yapılabilir. İade süresi 45 gün ve üzeri olanların sayısı ise 5.658 kişidir. Bu sayı oran olarak ise % 9,5'e tekabül etmektedir ve üyelerin kitap ödünç alma süresini en az iki kez daha uzattıklarını göstermektedir.

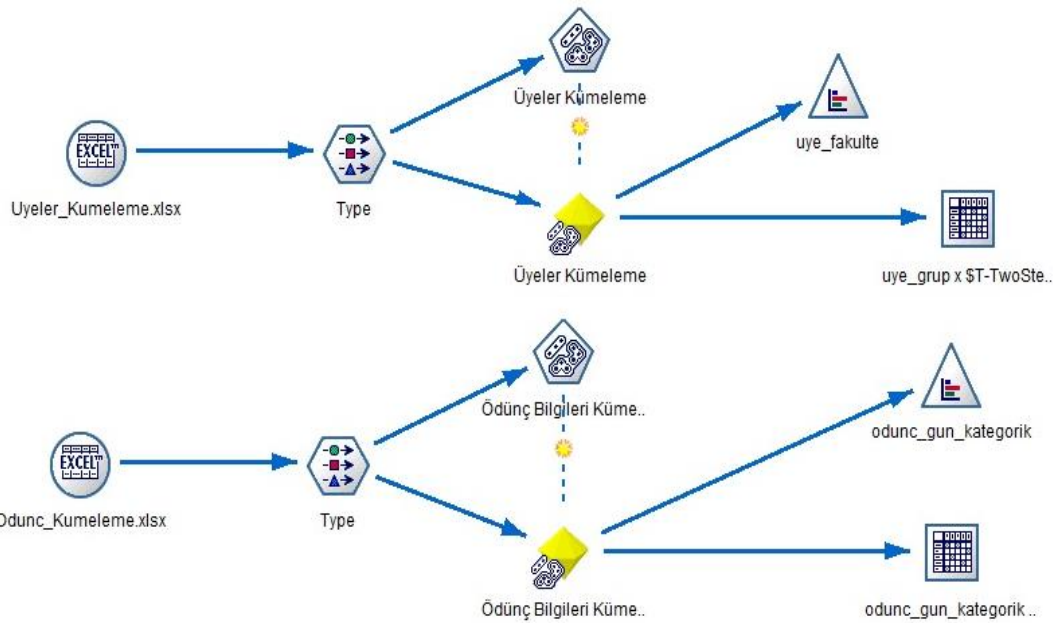
3.3. SPSS Modeller Uygulaması İle Modelleme

Uygulamanın buraya kadar olan kısmında modelleme öncesinde veri seti hakkında genel temel bazı analizler gerçekleştirilmiş ve bu doğrultuda istatistiki bilgiler paylaşılmıştır. İlerleyen bölümlerde ise Veri madenciliği teknikleri kullanılarak daha detaylı bilgi sahibi olmak için veri yığını içerisinde gizli bulunan örüntülerin tespiti gerçekleştirilecektir.

Araştırma sürecinde faydalanılan IBM SPSS Modeler, IBM tarafından geliştirilen klasik istatistik, veri madenciliği, metin analitiği, öngörüselleme, tahmine dayalı analitik gibi birçok analiz için kullanılabilen bir yazılım uygulamasıdır.

3.3.1. Kümeleme analizi ile modelleme

Veri madenciliği yöntemlerinden biri olan kümeleme analizi ile elimizdeki veri içerisinde bulunabilecek gruplar gibi bilgiler elde edilebilmektedir. Bu aşamada kütüphaneye kayıtlı üyeler ve üyelerin yayın ödünç alma alışkanlıkları kümeleme analizine tabi tutulmuştur. Aşağıdaki şekilde görüleceği üzere analiz için üyelere ve ödünç alma alışkanlıklarına ait veriler temizlenmiş ve dönüştürülmüş şekilde iki ayrı Excel dosyası olarak SPSS Modeler programına aktarılmıştır. Daha sonra analize tabii tutulacak alanların belirlenmesi için type adlı düğüm (node) kullanılmış ve çıktısı kümeleme analizi için kullanılan algoritmaya yönlendirilmiştir. Analiz için SPSS Modeler programının segmentation sekmesinde bulunan twostep adlı algoritma kullanılmıştır. Birçok kümeleme analizinde kullanılan k-means algoritmasından farklı olarak küme sayısını kendisi belirlediği için twostep algoritması büyük avantaj sağlamaktadır. Bununla beraber büyük veri setleri üzerinde çalışabilmesi, hem sürekli hem de kategorik değişkenler üzerinde çalışabilmesi yine twostep algoritmasının seçim gerekçelerindedir.



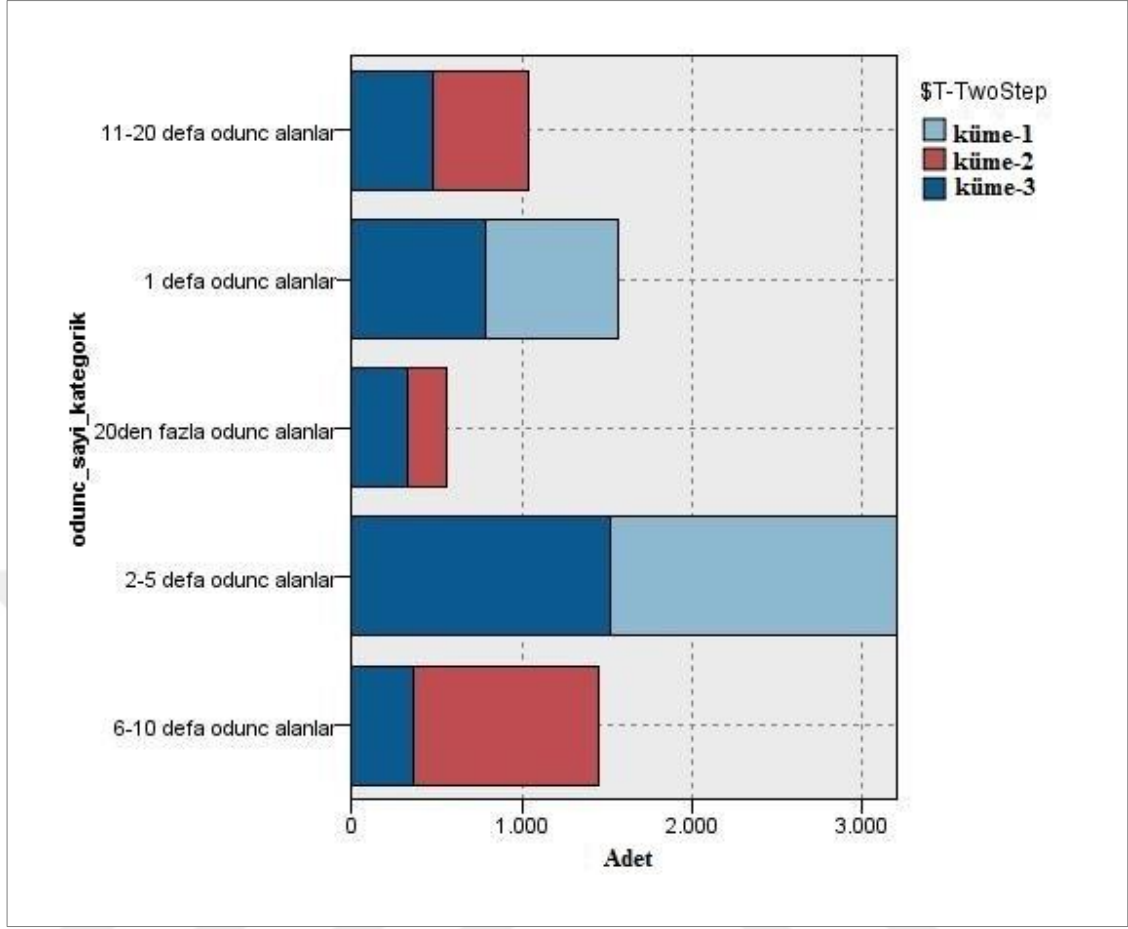
Şekil 3.7. Kümeleme analizi için akış şeması

Analize, üyelerin gruplarının keşfedilmesi yani kümelenmesi ile başlanmıştır. Şekil 3.8 SPSS Modeler programı vasıtasıyla üretilmiş olup analiz için kullanılan veriye ait temel bilgileri görselleştirmektedir. Şekilden anlaşıldığı üzere veri içerisinde beş adet alan bulunmakta olup bunlardan sadece üyelerin kaç adet ödünç alma işlemi yaptıklarını gösteren `odunc_alma_sayisi` isimli değişken süreklidir. Diğer değişkenlerin tümü kategorik değişkenlerdir. Twostep algoritması çalıştırıldıktan sonra üyeleri üç ayrı kümeye ayırdığı gözlemlenmiştir.

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Unique	Valid
uye_fakulte		Nominal	--	--	--	--	11	7802
uye_grup		Nominal	--	--	--	--	3	7802
uye_cinsiyet		Nominal	--	--	--	--	3	7802
odunc_alma_sayisi		Continuous	1.000	274.000	7.582	11.674	--	7802
odunc_sayi_kategorik		Nominal	--	--	--	--	5	7802

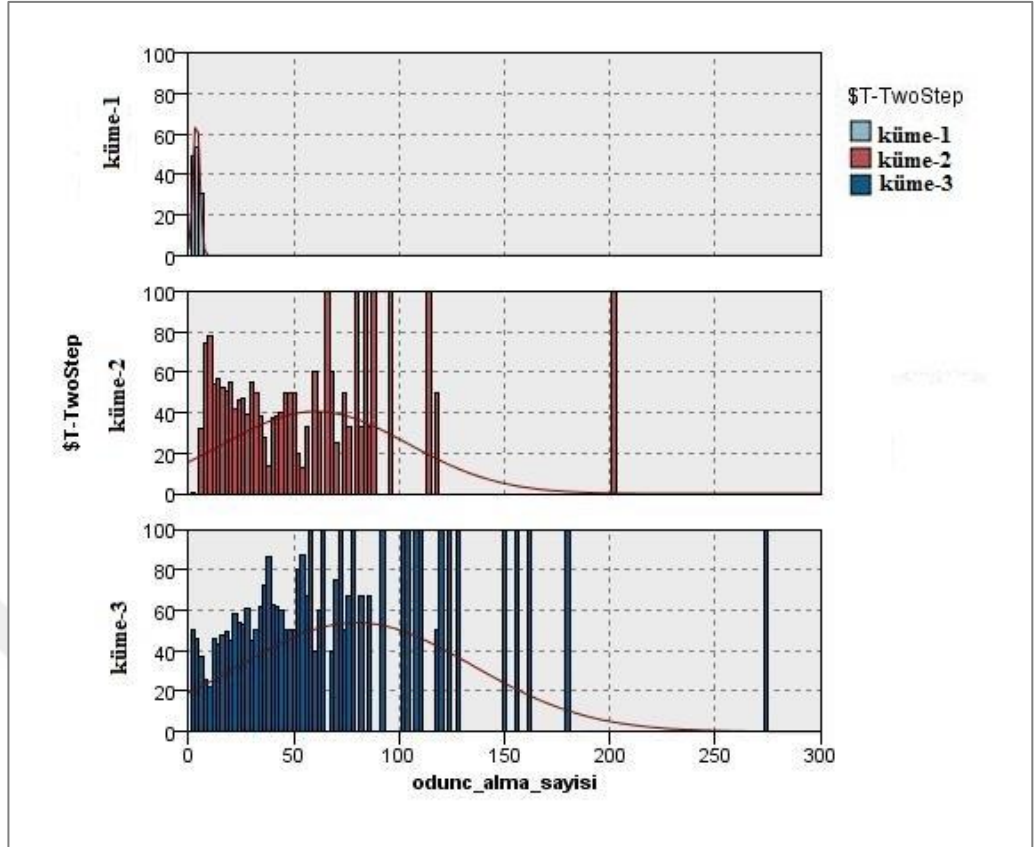
Şekil 3.8. Üyelere ait verilerin genel görünümü

Algoritma tarafından oluşturulan üç ayrı kümenin kullanıcıların ödünç alma sayılarının kategorize edilmesi neticesinde oluşturulan değişkene dağılımı ise aşağıdaki Şekil 3.9’da gösterilmektedir. İnceleme sonucunda iki ile beş defa arasında yayın ödünç alanların sadece 1. ve 3. kümeye dağıtıldığı yine en fazla ödünç alma oranı olan yirmiden fazla ödünç alanların ise 2 ve 3 nolu kümelere dağıtıldığı görülmüştür.



Şekil 3.9. Ödünç alma sayılarına göre üyelerin kümelere dağılımı

Ödünç alma sayısını kategorik olarak değil de, sürekli değişken olarak analiz etmek istediğimizde ise Şekil 3.10’ da görülen 3 ayrı histogram diyagramı elde edilmiştir. Her küme için ayrı bir histogramın elde edildiği şekil incelendiğinde az sayıda ödünç alan üyelerin 1 nolu kümede toplandıkları görülmüş, kalan üyelerin ise 2 ve 3 nolu kümelere kısmen homojen şekilde dağıldığı görülmüştür.



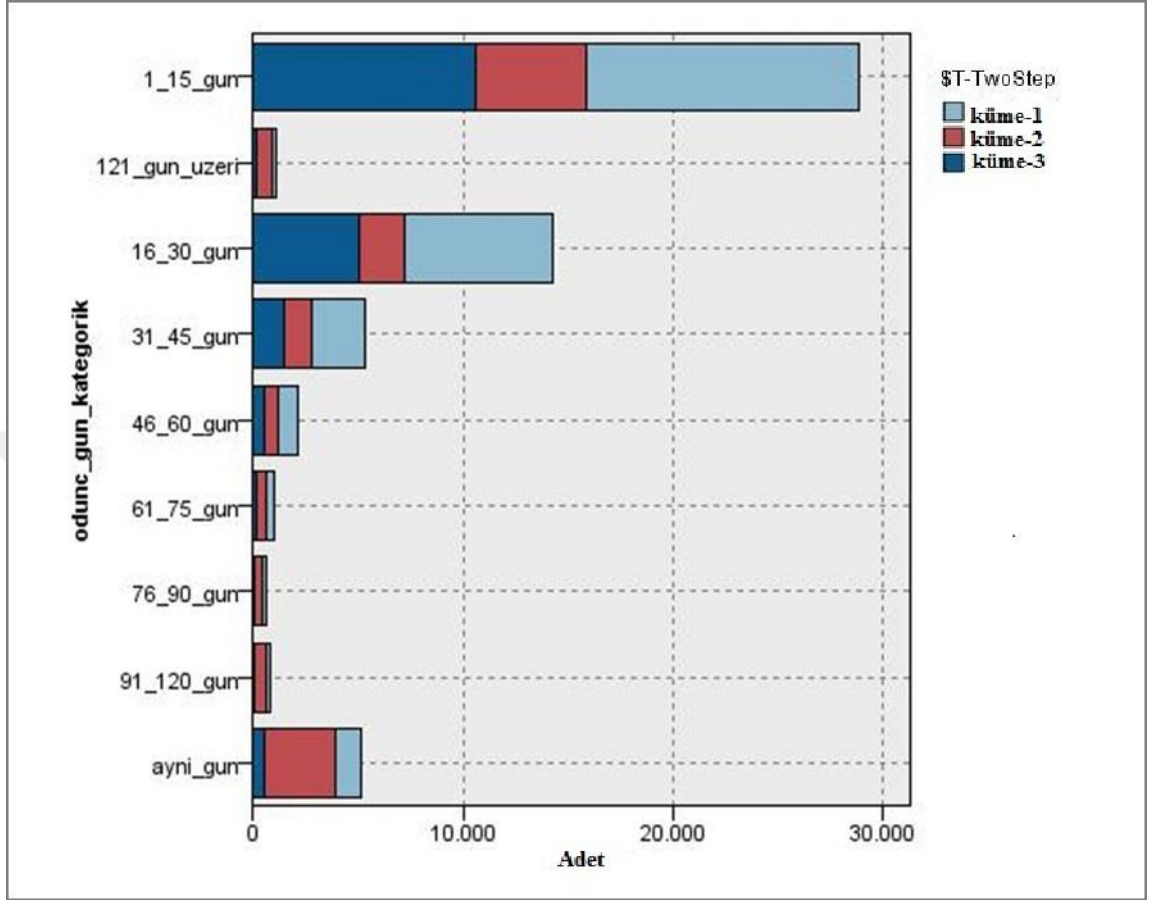
Şekil 3.10.Ödünç alma sayılarının histogram grafikleri

Field	Sample Graph	Measurement	Min	Max	Mean	Unique	Valid
dewey_ana_sinifla...		Categorical	--	--	--	10	59156
odunc_gun_katego...		Categorical	--	--	--	9	59156
uye_cinsiyet		Categorical	--	--	--	3	59156
uye_grup		Categorical	--	--	--	3	59156
uye_fakulte		Categorical	--	--	--	11	59156

Şekil 3.11. Ödünç alma işlemlerine ait verilerin genel görünümü

Kümeleme analizi için kullanılan diğer verilerimiz ise üyelerin kütüphaneden ödünç alma işlemlerini içeren ödünç tablosu olup veriye ait temel bilgiler Şekil 3.11’ de gösterilmiştir. Üyeler tablosundan farklı olarak bu tablomuzdaki tüm verilerin kategorik değişkenler

şeklinde olduğu görülmüş olup 59.156 kayıttan oluşan verimizin analizi için yine twostep algoritmasından faydalanılmıştır.

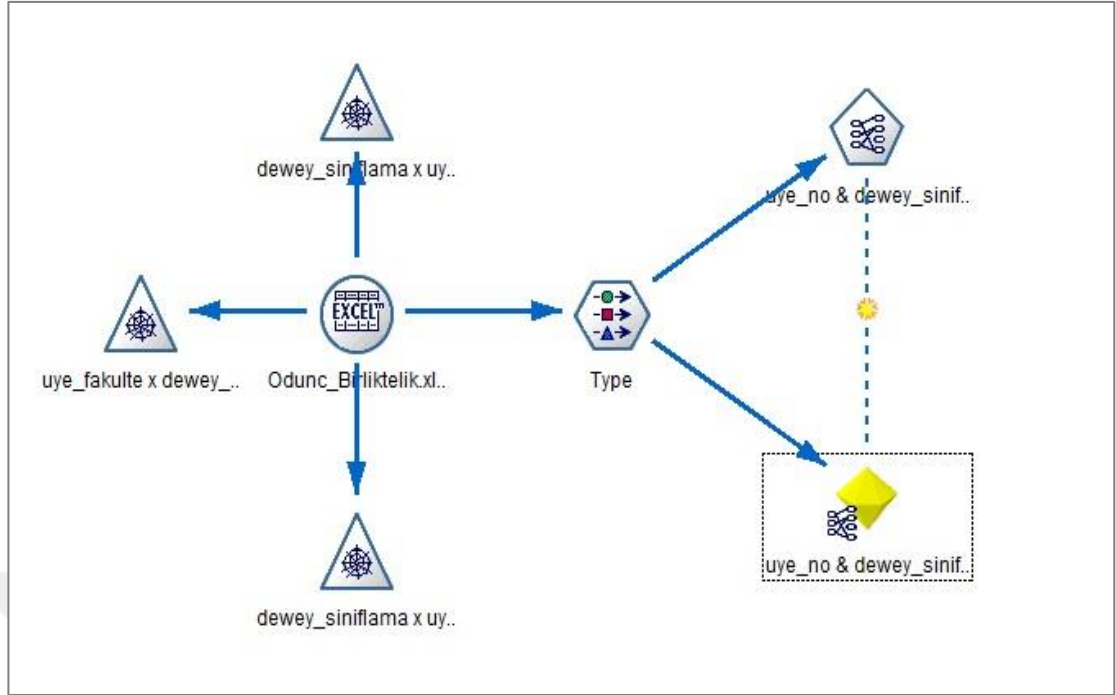


Şekil 3.12.Ödünç alma sürelerinin kümelere dağılımı

Twostep algoritması kullanıldığında ödünç işlemlerinin de üyeler tablosunda olduğu gibi 3 ayrı kümeye bölündüğü gözlemlenmiştir. Şekil 3.12’de ise üyelerin ödünç aldıkları yayımları geri getirme sürelerine göre oluşturulan odunc_gun_kategorik değişkeninin kümelere dağılımı gösterilmiştir. Şekil incelendiğinde en yoğun geri getirme dönemi olan ilk on beş günlük sürenin üç ayrı kümeye eşit düzeyde dağıldığı görülmekle beraber aynı gün geri getirenlerin algoritma tarafından 2 nolu kümeye atandığı görülmektedir.

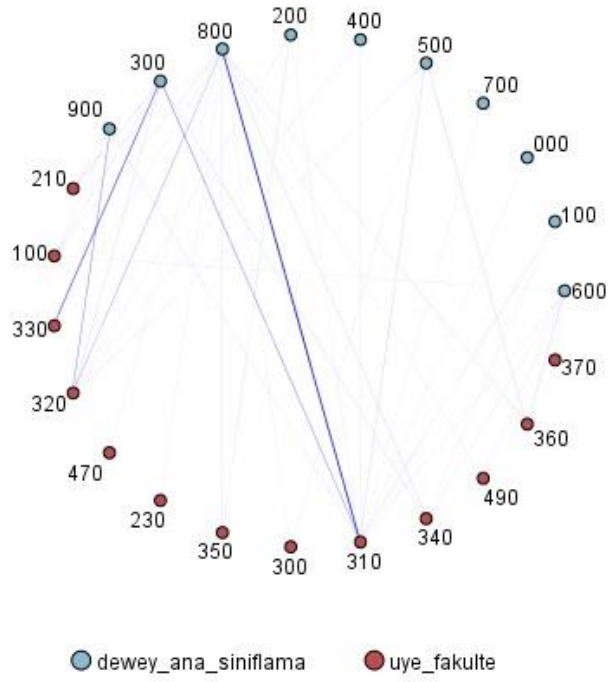
3.3.2. Birliktelik analizi ile modelleme

Öncelikle çalışma alanına madenciliği yapılacak olan dosya Excel formatında aktarılmış ve tip belirlemeleri için type adlı düğüm kullanılmıştır. Birliktelik analizi yapmak için oluşturulan akış şeması (stream) aşağıdaki Şekil 3.13’de gösterilmektedir.



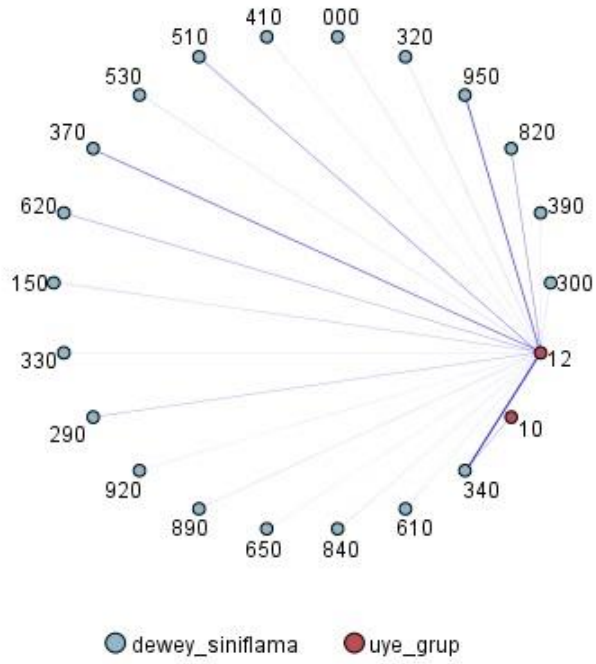
Şekil 3.13.Birliktelik analizi için oluşturulan akış şeması

Elde edilen akış şeması ile birliktelik kuralları oluşturulmadan önce SPSS Modeller programında bulunan ağ grafiği yardımıyla veriler görselleştirilmeye çalışılmıştır. Veri madenciliğinde diğer tüm teknikler kadar önemli olan veri görselleştirme yardımıyla kurallar elde edilmeden önce veri için genel bir görünüm oluşturulabilir. Şekil 3.14’de kütüphane üyelerinin bağlı buldukları fakülteler ile ödünç aldıkları kitapların ana sınıf kodları bahsedilen ağ grafiği yardımıyla görselleştirilmiştir. Veri setimiz içerisinde yayınların ana sınıfları ve üyelerin bağlı buldukları fakültelerin ne sıklıkla beraber bulunduğunu gösteren şekilde koyu olan çizgiler birlikteliklerin yüksek sayıda olduğunu belirtirken daha ince olan çizgiler ise birlikteliklerin düşük sayılarda gerçekleştiğini göstermektedir. SPSS Modeller programında çeşitli ayarlamalar yapılarak belirli sayılar arasında oluşan birliktelikler gözlemlenebilir. Şekil 3.14 incelendiğinde en sık birlikteliğin 310 kodlu Eğitim Fakültesine ait kayıtlı üyelerin 800 dewey kodlu Edebiyat ve Retorik ana yayın grubundan gerçekleştirdikleri ödünç alma işlemlerinden oluştuğu görülmüştür. 320 kodlu Fen Edebiyat Fakültesine kayıtlı üyelerin 900 dewey kodlu coğrafya ve tarih ana yayın grubundan gerçekleştirdikleri ödünç alma işlemlerinin de yine sık tekrarlanan bir birliktelik olduğu gözlemlenmiştir.



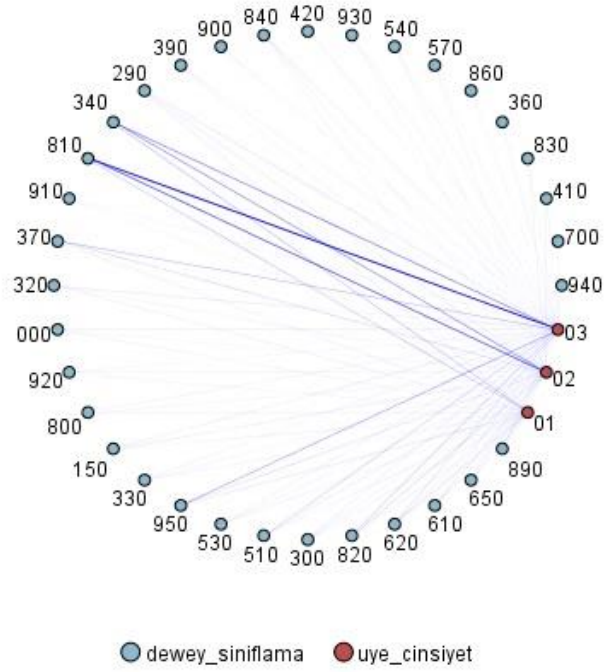
Şekil 3.14. Üyelerin fakülteleri ile ödünç alınan yayın kodları arasındaki ilişki

Şekil 3.15’de ise ödünç alma işlemleri üye grupları ve dewey yayın sınıfları arasındaki birlikteliklerin gerçekleşme sıklığını göstermektedir. Bu şekil incelendiğinde ise 12 kodlu üyeler olan öğrencilerin en sık ödünç aldıkları yayınların 340 kodlu hukuk, 370 kodlu eğitim, 510 kodlu Matematik ve 950 kodlu Türk tarihinin de içinde bulunduğu genel Asya tarihi oldukları gözlemlenmiştir. Üye kodu 10 olan akademisyenlerin ise en fazla aldıkları yayınların 340 kodlu hukuk grubundan olduğu görülmüştür.



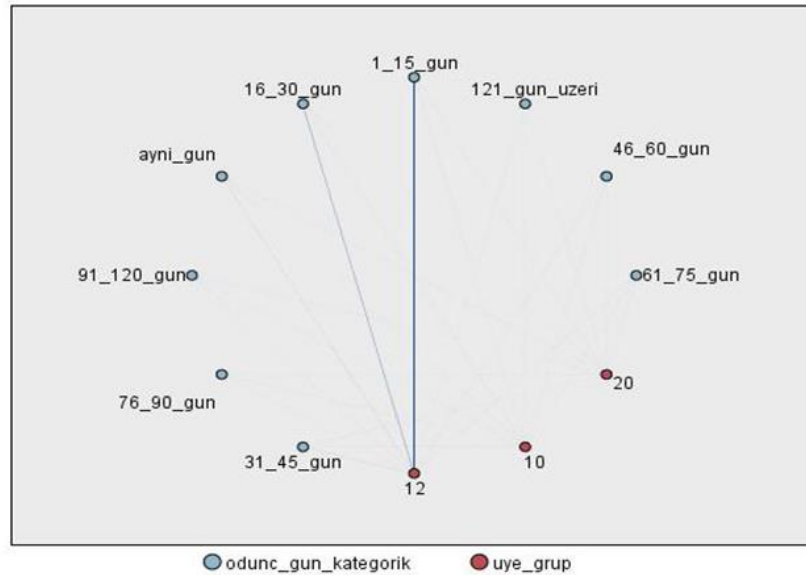
Şekil 3.15. Üyelerin unvanları ile ödünç alınan yayın kodları arasındaki ilişki

Ağ grafiği yardımıyla elde edilen Şekil 3.16’da ise üyelerin cinsiyetlerine göre hangi yayın gruplarını daha fazla tercih ettikleri gösterilmektedir. Şekle göre 02 kodlu kadın üyelerin 810 kodlu Türk edebiyatı yayın grubundan yapmış oldukları ödünç alma işlemlerinin çok sık tekrarlandığı görülmüştür. Burada dikkat edilmesi gereken bir husus da cinsiyet bilgisi içermeyen ödünç alma işlem kayıtlarının da oldukça sık bir şekilde tekrarlandığı durumudur. Veri seti içerisinde boş olan kayıtlar silinmesine rağmen cinsiyet bilgisi içermeyen kayıtların fazlalığı bu silme işlemini riskli hale getirmiştir.



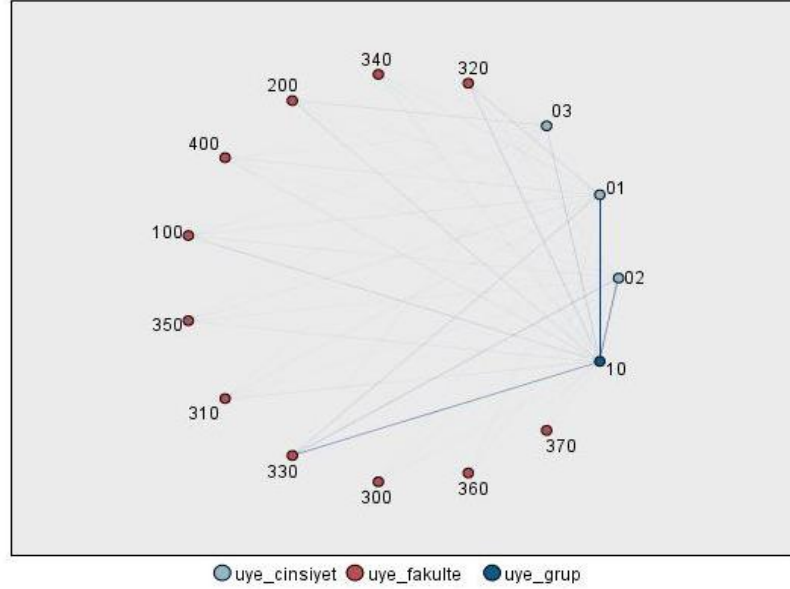
Şekil 3.16. Üyelerin cinsiyetleri ile ödünç alınan yayın kodları arasındaki ilişki

Ağ grafiği yardımıyla elde edilen Şekil 3.17 ise üyelerin bağlı oldukları gruplara göre hangi sürede yayınları geri getirdiklerini göstermektedir. Şekilden ilk on beş günlük süre içerisinde öğrencilerin yayınları geri getirme birlikteliğinin diğer tüm birlikteliklerden çok daha fazla olduğu görülmüştür.



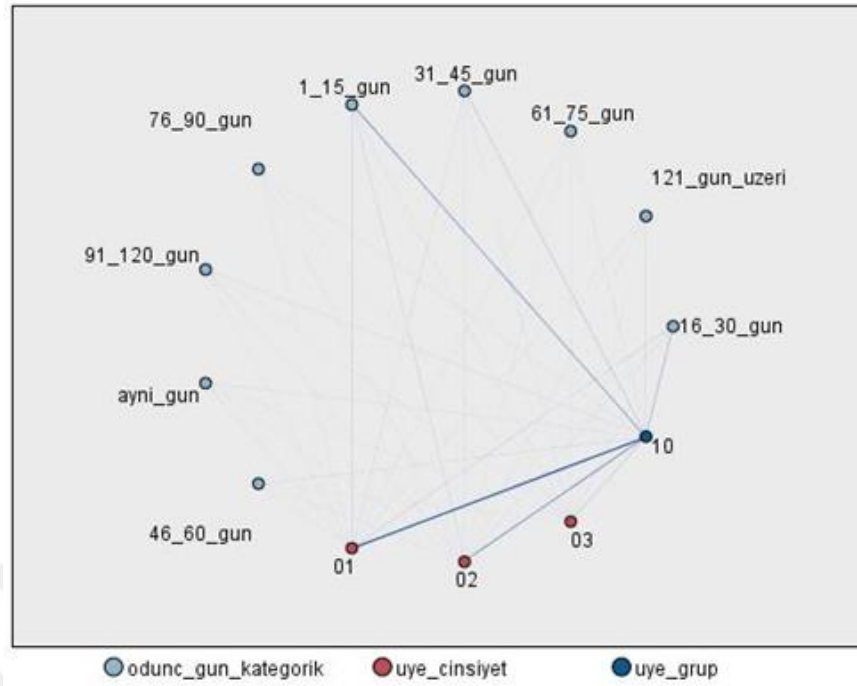
Şekil 3.17. Üyelerin unvanları ile ödünç alma süreleri arasındaki ilişki

Ağ grafikleri yardımıyla yaptığımız görselleştirme işlemlerinde öğrenci grubuna ait üye sayısı çok fazla olduğu için akademisyen grubuna ait verileri baskılamakta ve görselleştirilmesini zorlaştırmaktadır. Bu nedenle sadece akademisyen üye grubuna ait veriler süzülerek Şekil 3.18'deki ve Şekil 3.19'daki iki ağ grafiği elde edilmiştir.



Şekil 3.18. Akademisyen üyelerin fakülteleri ile cinsiyetleri arasındaki ilişki

Elde edilen grafiklerden 10 grup kodlu akademisyenlerin cinsiyetlere, bağlı buldukları eğitim kurumlarına göre ve ödünç aldıkları yayınları geri getirdikleri sürelerle göre en yoğun olan birliktelikler olduğu görülmektedir.



Şekil 3.19. Akademisyenlerin cinsiyetleri ile ödünç alma süreleri arasındaki ilişki

Ağ grafikleri ile görselleştirme aşamasından sonra SPSS Modeler programı ile oluşturulan akış şeması çalıştırılmış ve birliktelik kurallarını listeleyen düğüm (node) oluşturulmuştur. Şekil 3.20’de ise elde edilen çıktı gösterilmektedir. Sisteme girilen destek ve güven eşik değerine göre apriori algoritması 4.470 adet kural üretmiştir. Bu kurallar destek, güven ve kaldıraç değerlerine göre ayrı ayrı sıralanabilmektedir. Destek değeri öncül ve ardıl yayın gruplarının tüm işlemlerde yüzde kaç oranında yer aldığını, güven değeri öncül yayın grubundan alan bir kullanıcının yüzde kaç olasılıkla ardıl yayın grubundan yayın ödünç aldığını ve son olarak kaldıraç değeri ise kuralın ilginçliğini göstermektedir. Eşik değerler olarak ise minimum destek eşik değeri için %1, minimum güven eşik değeri içinse %10 olarak belirlenmiştir. En fazla öncül öge sayısı ise üç olarak belirlenmiştir.

Consequent	Antecedent	Support %	Confidence %	Lift
810 = T	150 = T 890 = T 820 = T	1,243	95,876	2,288
810 = T	840 = T 150 = T 820 = T	1,051	95,122	2,27
950 = T	910 = T 320 = T	1,013	94,937	7,102
810 = T	830 = T 890 = T 820 = T	1,038	93,827	2,239
810 = T	920 = T 820 = T	1,23	93,75	2,237
810 = T	890 = T 370 = T 820 = T	1,307	93,137	2,223
810 = T	830 = T 840 = T 820 = T	1,089	92,941	2,218
810 = T	860 = T 840 = T	1,064	92,771	2,214
810 = T	800 = T 840 = T	1,038	92,593	2,21
810 = T	860 = T 820 = T	1,525	92,437	2,206

Şekil 3.20. Güven değerine göre sıralanmış birliktelik kuralları

4. ARAŞTIRMA BULGULARI

4.1. Kümeleme Analizi Sonuçlarının Yorumlanması

Kümeleme analizi ile üçüncü bölümde elde edilen sonuçlar Tablo 4.1’de bir araya getirilmiştir. Üyelerin kümelenmesi sonucunda oluşan üç adet kümenin üyeler veri tablosundaki her değişkene göre dağılımı tablodan incelenmiş ve aşağıdaki sonuçlar üretilmiştir.

Üyelerin bağlı oldukları birimlere göre küme dağılımları incelendiğinde en fazla üyenin 3 nolu kümede bulunduğu, küme içerisinde en yoğun bulunan birimin %27,71 ile Eğitim Fakültesi olduğu, genel olarak üyelerin tüm kümelere birbirine yakın sayılarda dağıldığı, Tıp Fakültesi üyelerinin 1 ve 3 nolu kümelere bulunduğu görülmüştür.

Üyelerin bağlı oldukları gruplara göre küme dağılımlarına bakıldığında ise en fazla üyenin 3 nolu kümede bulunduğu, küme içerisindeki en yoğun grubun %85,11 ile öğrenciler olduğu ve Akademik personel ile diğer üyelerin yoğunluklarının birbirine yakın olduğu, diğer üyelerin 1. kümeye dağılımının yapılmadığı tespit edilmiştir.

Üyelerin cinsiyet durumlarına göre küme dağılımlarında ise kadın erkek ve değer verilmeyen kısım bulunduğu, yoğunluk dağılımının en fazla 3. Kümeye gerçekleştiği, erkek üyelerin sadece 3. Kümeye dağılımlarının yapıldığı, kadın ve erkek üyelerin 1. Kümeye bulunmadığı ve cinsiyet grubunda değer bilgisi atanmayan üyelerin ise en az 3. Kümeye atandığı tespit edilmiştir.

Kaynak ödünç alma sayılarına göre küme dağılımları incelendiğinde en çok üyenin “2 - 5 arasında ödünç alanlar” grubunda yer aldığı, bu gruptaki dağılımlarında 1. ve 3. kümede birbirlerine yakın yoğunlukta olduğu, “1 defa ödünç alanlar” grubunda da dağılımın aynı şekilde 1. ve 3. kümede birbirlerine yakın yoğunlukta olduğu, 2. Kümeye oldukça az sayıda üyenin bulunduğu gözlenmiştir.

Tablo 4.1. Üyelerin kümelere dağılımı

Özellikler	Değerler	Küme 1		Küme 2		Küme 3	
		Sayı	%	Sayı	%	Sayı	%
Fakülteler	Diğer Birimler	0	0,00	30	1,60	276	7,95
	Enstitüler	121	4,93	121	6,45	76	2,19
	Eczacılık Fakültesi	57	2,32	35	1,86	37	1,07
	Eğitim Fakültesi	706	28,78	760	40,49	962	27,71
	Fen-Edebiyat Fakültesi	215	8,76	346	18,43	454	13,08
	Hukuk Fakültesi	359	14,64	134	7,14	330	9,50
	İktisadi ve İdari Bilimler Fakültesi	335	13,66	182	9,70	309	8,90
	İlahiyat Fakültesi	169	6,89	62	3,30	284	8,18
	Mühendislik Fakültesi	251	10,23	166	8,84	309	8,90
	Tıp Fakültesi	10	0,41	0	0,00	43	1,24
	Meslek Yüksek Okulları	230	9,38	41	2,18	392	11,29
Gruplar	Akademik Personel	7	0,29	24	1,28	330	9,50
	Öğrenci	2446	99,71	1851	98,61	2955	85,11
	Diğer Üyeler	0	0,00	2	0,11	187	5,39
Cinsiyet	Erkek	0	0,00	0	0,00	1642	47,29
	Kadın	0	0,00	337	17,95	1815	52,28
	Değer Yok	2453	100,00	1540	82,05	15	0,43
Ödünç Alma Sayıları	1 defa ödünç alanlar	774	31,55	6	0,32	782	22,52
	2 - 5 arasında ödünç alanlar	1679	68,45	0	0,00	1520	43,78
	6-10 arasında ödünç alanlar	0	0,00	1085	57,81	362	10,43
	11-20 arasında ödünç alanlar	0	0,00	560	29,83	479	13,80
	20'den fazla ödünç alanlar	0	0,00	226	12,04	329	9,48

Üyelerin kümelenmesinden sonra ödünç alma işlemlerinin kümelenmesi ile yine 3 adet küme elde edilmiş ve veri seti içerisindeki değişkenlerin kümelere dağılımı Tablo 4.2’de gösterilmektedir. Tablodaki verilerin incelenmesi ile aşağıdaki sonuçlara ulaşılmıştır.

Üyelerin bağlı buldukları birimlere göre ödünç alma alışkanlıklarının kümelere dağılımı incelendiğinde, 3 nolu kümenin sadece Eğitim Fakültesi ve İktisadi İdari Bilimler Fakültesi üyeleri tarafından yapılan ödünç alma işlemlerinden oluştuğu görülmüştür. Fen Edebiyat Fakültesine kayıtlı üyelerin yaptığı ödünç alma işlemlerinin %39,86 ile büyük oranda 1 nolu kümede buldukları, Hukuk Fakültesine kayıtlı üyelerin yaptığı ödünç alma işlemlerinin ise tümüyle 2 nolu kümede %44,62’lik bir oran ile buldukları görülmüştür.

Üyelerin bağlı buldukları gruplara göre ödünç alma alışkanlıklarının küme dağılımlarına bakıldığında 3 nolu kümenin sadece öğrencilerden oluştuğu, öğrenci olan üyelerin 2 nolu kümede diğer üyelerin toplamına yakın sayıda olduğu tespit edilmiştir.

Üyelerin cinsiyet durumlarına göre ödünç alma alışkanlıklarının küme dağılımlarında ise erkek üyelerin 1 ve 2 nolu kümelerde birbirine yakın sayılarda bulunduğu, kadın üyelerin 1 ve 3 nolu kümelerde birbirine yakın sayılarda bulunduğu, cinsiyet grubunda değer bilgisi atanmayan üyelerin oranlarının 1 ve 3 nolu kümelerde diğer üyelerinden oranlarından yüksek olduğu, cinsiyet bazlı kümeleme dağılımında tüm kümelere atamalar yapıldığı görülmüştür.

Kaynak ödünç alma gün sayılarına göre ödünç alma alışkanlıklarının küme dağılımları incelendiğinde en çok üyenin %48,67 oranı ile “1 – 15 gün arası ödünç alanlar” grubunda yer aldığı, aynı gün geri getirenler grubunda üye dağılımının yarısının 2 nolu kümede toplandığı, “16 – 30 gün arasında ödünç alanların %24,12 ile yoğunluk açısından ikinci sırada yer alan grup olduğu, üyelerin %18,57’sinin ise ödünç aldığı kaynakları 30 günden daha geç sürede getirdikleri gözlenmiştir.

Tablo 4.2. Ödünç alma işlemlerinin kümelere dağılımı

Özellikler	Değerler	Küme 1		Küme 2		Küme 3	
		Sayı	%	Sayı	%	Sayı	Sayı
Fakülteler	Diğer Birimler	281	1,09	3.114	20,94	0	0,00
	Enstitüler	1.521	5,93	1.861	12,51	0	0,00
	Eczacılık Fakültesi	794	3,09	7	0,05	0	0,00
	Eğitim Fakültesi	0	0,00	430	2,89	18.118	97,33
	Fen-Edebiyat Fakültesi	10.232	39,86	890	5,98	0	0,00
	Hukuk Fakültesi	0	0,00	6.636	44,62	0	0,00
	İktisadi ve İdari Bilimler Fakültesi	3.032	11,81	644	4,33	497	2,67
	İlahiyat Fakültesi	2.437	9,49	58	0,39	0	0,00
	Mühendislik Fakültesi	5.451	21,24	61	0,41	0	0,00
	Tıp Fakültesi	0	0,00	412	2,77	0	0,00
	Meslek Yüksek Okulları	1.921	7,48	759	5,10	0	0,00
Gruplar	Akademik Personel	436	1,70	5134	34,52	0	0,00
	Öğrenci	25.180	98,09	7.394	49,72	18.615	100,00
	Diğer Üyeler	53	0,21	2344	15,76	0	0,00
Cinsiyet	Erkek	5.475	21,33	5.257	35,35	2.088	11,22
	Kadın	6.303	24,55	5634	37,88	6402	34,39
	Değer Yok	13.891	54,12	3.981	26,77	10.125	54,39
Ödünç Gün Sayıları	Aynı gün geri getirenler	1.182	4,60	3.399	22,86	517	2,78
	1_15_gun	12.908	50,29	5.314	35,73	10.575	56,81
	16_30_gun	7.083	27,59	2.157	14,50	5.030	27,02
	31_45_gun	2.537	9,88	1.285	8,64	1.511	8,12
	46_60_gun	959	3,74	696	4,68	488	2,62
	61_75_gun	387	1,51	441	2,97	188	1,01
	76_90_gun	172	0,67	333	2,24	86	0,46
	91_120_gun	191	0,74	508	3,42	78	0,42
	121_gun_uzeri	250	0,97	739	4,97	142	0,76
Dewey Kodları	Genel Konular	670	2,61	201	1,35	342	1,84
	Felsefe ve Psikoloji	872	3,40	304	2,04	1.250	6,72
	Din	1.842	7,18	129	0,87	279	1,50
	Toplum Bilimleri	1.866	7,27	10.265	69,02	4.536	24,37

Dil ve Dilbilim	762	2,97	156	1,05	343	1,84
Doğa Bilimleri ve Matematik	2.967	11,56	269	1,81	1.811	9,73
Teknoloji (Uygulamalı Bilimler)	3.391	13,21	992	6,67	298	1,60
Güzel Sanatlar	165	0,64	174	1,17	490	2,63
Edebiyat ve Retorik	7.871	30,66	1.608	10,81	8.384	45,04
Coğrafya ve Tarih	5.263	20,50	774	5,20	882	4,74

Dewey kodlarına göre kaynak ödünç alma alışkanlıklarının küme dağılımları değerlendirildiğinde Edebiyat ve Retorik ile Toplum Bilimleri ders gruplarının sırasıyla %30,20 ve %28,17 oranlarına karşılık gelecek şekilde en yüksek iki grup olduğu, %1,40 oranı ile en düşük yoğunluğun Güzel Sanatlar ders grubu kümelerinde bulunduğu, Teknoloji ve Fen Bilimleri ders gruplarının diğer ders gruplarına göre oldukça düşük yoğunlukta kalmasından hareketle kaynakları ödünç alan üyelerin sözel ve sosyal bilimler ağırlıklı bölümlerde öğrenim görüyor olabilecekleri hususları değerlendirilmektedir.

4.2. Birliktelik Analizi Sonuçlarının Yorumlanması

Birliktelik analizi için apriori algoritmasının uygulanması sonucu elde edilen kurallardan bir kısmı Tablo 4.3'de gösterilmiştir. Hangi tür yayınların birlikte ödünç alındığının görülmesi açısından kuralları güven kriterine göre sıralamak gerekmektedir. Bununla beraber aynı kurallar destek ve kaldıraç değerlerine göre de sıralanarak ortaya çıkan sonuçlar yorumlanabilir.

Tablo 4.3. Birliktelik kurallarının farklı konu başlıklarında sıralanması

Öncül	Ardıl	Destek %	Güven %	Kaldıraç
Türk Edebiyatı	Psikoloji Diğer Diller Edebiyatı İngiliz Edebiyatı	1,243	95,876	2,288
Asya Tarihi	Coğrafya Siyasal Bilimler	1,013	94,937	7,102
İngiliz Edebiyatı	Fransız Edebiyatı Toplum Bilimleri Türk Edebiyatı	1,141	76,404	4,276
Siyasal Bilimler	Biyografi Hıristiyanlık Dışındaki Dinler Asya Tarihi	1,025	65	8,699
Fizik	Mühendislik Matematik Türk Edebiyatı	1,115	62,069	9,666
Psikoloji	Tıp Bilimleri Toplum Bilimleri	1,013	62,025	6,22
Eğitim	Yaşam Bilimleri Türk Edebiyatı	1,295	61,386	3,994
Matematik	Fizik İngiliz Edebiyatı Türk Edebiyatı	1,192	60,215	5,947
Siyasal Bilimler	Hukuk Asya Tarihi	1,871	58,904	7,883
Hıristiyanlık Dışındaki Dinler	Biyografi Siyasal Bilimler Türk Edebiyatı	1,077	57,143	5,184

Güven sıralı olarak ortaya çıkarılan yayın grupları kuralları için sırasıyla aşağıdaki yorumlar yapılabilir.

Türk Edebiyatı → Psikoloji – Diğer Diller Edebiyatı – İngiliz Edebiyatı yayın grupları birliktelik kuralı için;

Türk Edebiyatı ve Psikoloji – Diğer Diller Edebiyatı – İngiliz Edebiyatı yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,243'dür. Ayrıca Türk Edebiyatı yayın grubundan ödünç alan kütüphane üyelerinin % 95,876 olasılıkla Psikoloji

– Diğer Diller Edebiyatı – İngiliz Edebiyatı yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

Asya Tarihi → Coğrafya – Siyasal Bilimler yayın grupları birliktelik kuralı için;

Asya tarihi ve Coğrafya – Siyasal Bilimler yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,013'dür. Ayrıca Asya Tarihi yayın grubundan ödünç alan kütüphane üyelerinin % 94,937 olasılıkla Coğrafya – Siyasal Bilimler yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

İngiliz Edebiyatı → Fransız Edebiyatı – Toplum Bilimleri – Türk Edebiyatı yayın grupları birliktelik kuralı için;

İngiliz Edebiyatı – Fransız Edebiyatı – Toplum Bilimleri – Türk Edebiyatı yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,141'dir. Ayrıca İngiliz Edebiyatı yayın grubundan ödünç alan kütüphane üyelerinin % 76,404 olasılıkla Fransız Edebiyatı – Toplum Bilimleri – Türk Edebiyatı yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

Siyasal Bilimler → Biyografi – Hıristiyanlık Dışındaki Dinler – Asya Tarihi yayın grupları birliktelik kuralı için;

Siyasal Bilimler – Biyografi – Hıristiyanlık Dışındaki Dinler – Asya Tarihi yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,025'tir. Ayrıca Sosyal Bilimler yayın grubundan ödünç alan kütüphane üyelerinin % 65 olasılıkla Biyografi – Hıristiyanlık Dışındaki Dinler – Asya Tarihi yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

Fizik → Mühendislik – Matematik – Türk Edebiyatı yayın grupları birliktelik kuralı için;

Fizik – Mühendislik – Matematik – Türk Edebiyatı yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,115'tir. Ayrıca Fizik yayın grubundan ödünç alan kütüphane üyelerinin % 62,069 olasılıkla Mühendislik – Matematik – Türk Edebiyatı yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

Psikoloji → Tıp Bilimleri – Toplum Bilimleri yayın grupları birliktelik kuralı için;

Psikoloji – Tıp Bilimleri – Toplum Bilimleri yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,013'tir. Ayrıca Psikoloji yayın grubundan ödünç alan kütüphane üyelerinin % 62,025 olasılıkla Tıp Bilimleri – Toplum Bilimleri yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

Eğitim → Yaşam Bilimleri – Türk Edebiyatı yayın grupları birliktelik kuralı için;

Eğitim – Yaşam Bilimleri – Türk Edebiyatı yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,295'tir. Ayrıca Eğitim yayın grubundan ödünç alan kütüphane üyelerinin % 61,386 olasılıkla Yaşam Bilimleri – Türk Edebiyatı yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

Matematik → Fizik – İngiliz Edebiyatı – Türk Edebiyatı yayın grupları birliktelik kuralı için;

Matematik – Fizik – İngiliz Edebiyatı – Türk Edebiyatı yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,1192'tir. Ayrıca Matematik yayın grubundan ödünç alan kütüphane üyelerinin % 60,215 olasılıkla Fizik – İngiliz Edebiyatı – Türk Edebiyatı yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

Siyasal Bilimler → Hukuk – Asya Tarihi yayın grupları birliktelik kuralı için;

Siyasal Bilimler – Hukuk – Asya Tarihi yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,871'tir. Ayrıca Siyasal Bilimler yayın grubundan ödünç alan kütüphane üyelerinin % 58,904 olasılıkla Hukuk – Asya Tarihi yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

Hıristiyanlık Dışındaki Dinler → Biyografi – Siyasal Bilimler – Türk Edebiyatı yayın grupları birliktelik kuralı için;

Hıristiyanlık Dışındaki Dinler – Biyografi – Siyasal Bilimler – Türk Edebiyatı yayın gruplarının toplam ödünç alma hareketlerinde birlikte görülme olasılıkları % 1,077'tir. Ayrıca Hıristiyanlık Dışındaki Dinler yayın grubundan ödünç alan kütüphane üyelerinin % 57,143 olasılıkla Biyografi – Siyasal Bilimler – Türk Edebiyatı yayın gruplarından da ödünç alma işlemi yaptıkları söylenebilir.

5. SONUÇ ve ÖNERİLER

Teknolojinin gelişmesi ile birlikte bilgisayarların ve bilgi sistemlerinin artık her iş alanında kullanılmakta ve depolanan veri miktarı da her geçen gün artmaktadır. Büyük veri kütleleri içerisinde anlamlı ve işe yarar bilgiler elde etmek için istatistik bilimi ve bilgisayar bilimini bir araya getiren teknikler bütünü olarak veri madenciliği sıklıkla tercih edilmektedir.

Bu çalışmada Erzincan Binali Yıldırım Üniversitesi Merkez Kütüphanesi'ne ait 2010-2017 yılları arasındaki verilere, Scott Nicholson tarafından literatüre 2003 yılında kazandırılan bibliomining olarak adlandırılan kütüphanelerdeki veri tabanı bilgi keşfi süreci uygulanmıştır. Uygun veri madenciliği teknikleri uygulayarak kütüphanede bulunan elektronik olmayan tüm kitaplarda ödünç alma alışkanlıkları, envanter bilgisi, kütüphane üye profili ve birlikte ödünç alınan yayınlar gibi yararlı bilgiler elde edilmiştir.

Çalışmada öncelikle veri elde edilerek, veri ön işleme tabii tutulmuş olup üzerinde temizleme ve dönüştürme işlemleri uygulanmıştır. Bu temizleme ve dönüştürme işlemleri için veri tabanına aktarılan veriler üzerinde SQL sorguları kullanılmıştır. Temizleme ve dönüştürme işlemlerine tabii tutulan verilere IBM SPSS Modeler yazılımı yardımıyla, veri madenciliği tekniklerinden olan kümeleme analizi ve birliktelik analizi yöntemleri uygulanmıştır. Veri madenciliğinin tanımlayıcı modellerinden olan kümeleme analizi için two-step algoritması seçilmiştir. Oluşturulacak kümelerin sayısının algoritma tarafından otomatik olarak belirlenmesi bu algoritmanın en önemli seçim gerekçesidir. Analiz neticesinde kütüphane üyelerinin üç ayrı küme oluşturduğu görülerek, kümeler içeriklerine göre incelenmiştir. İnceleme neticesinde kütüphane üyelerinin çoğunlukla beş kitaptan daha az ödünç yayın alan eğitim fakültesi öğrencilerinden oluştuğu görülmektedir. Akademik personelin daha fazla sayıda bulunduğu bir diğer kümede ise ödünç yayın kullanımının ortalama olarak beş ve üzeri olduğu görülmektedir. Mühendislik Fakültesi gibi teknik bölümlerde ise üye sayısının oldukça az olduğu görülmüştür. Bu durumun Üniversite içerisindeki birimlerin ve bölümlerin aktif olma yıllarıyla da ilişkili olabileceği, bu sebeple Mühendislik Fakültesi gibi daha geç kurulan birimlerde öğrenci ve personel sayısının az olmasıyla paralel olarak ödünç alma işlemlerinin de daha az olduğu düşünülmektedir. Literatürde bulunan diğer çalışmalardan

farklı olarak, ödünç alma hareketlerine de kümeleme analizi uygulanmış ve yine üç adet küme oluştuğu görülmüştür. Oluşan kümeler incelendiğinde Eğitim Fakültesi ile İktisadi ve İdari Bilimler Fakültesi öğrencilerinin gerçekleştirdiği ödünç alma işlemlerinin ayrı bir kümede toplandığı ve ödünç alınan yayın gruplarının Toplum Bilimleri ile Edebiyat alanlarında yoğunlaştığı belirlenmiştir. Hukuk Fakültesinde bulunan akademisyenlerin yaptığı ödünç alma işlemlerinin ise bir diğer kümede yoğunlaştığı ve özellikle Toplum Bilimleri yayın grubunun tercih edildiği görülmektedir. Tüm kümelerde ödünç alınan yayınların sıklıkla ilk on beş günlük süre içerisinde geri getirildiği, aynı gün ödünç alınan yayını iade etme durumunun ise akademisyenlerde öğrencilerden daha fazla olduğu tespit edilmiştir. Kümeleme analizi ile elde edilen üye profilleri ve ödünç alma alışkanlıkları, kütüphanenin fiziksel kaynak kullanımının ve üyelerin memnuniyetinin artırılması amaçlı olarak kullanılabilir. Ödünç alma işlemlerinde sıklıkla tercih edilen kitapların stok miktarları gözden geçirilebilir. Kullanılan yöntem ve bulunan sonuçlar itibariyle çalışma Uçan (2010), Decker ve Höppner (2006), Tempelman ve Pearce (2014) çalışmalarını desteklemektedir.

Çalışmada verilere bir diğer tanımlayıcı model olan ve Pazar sepet analizi olarak da bilinen birliktelik kuralları analizi uygulandığında ise 4.470 adet birliktelik kuralı üretilmiştir ve böylelikle üyelerin hangi tür yayınları birlikte aldıklarına dair bazı bilgilere ulaşılmıştır. Birliktelik analizinin gerçekleştirilmesi için en yaygın kullanılan algoritmalarından biri olan Apriori algoritmasından yararlanılmıştır. Algoritmanın ihtiyaç duyduğu destek ve güven değerleri verinin boyutuna göre şekillendirilmiştir. Elde edilen kurallar incelendiğinde, Siyasal Bilimler-Hukuk-Türkiye Tarihi, Eğitim-Yaşam Bilimleri-Türk Edebiyatı ve Türk Edebiyatı-Psikoloji-İngiliz Edebiyatı yayın grubu birlikteliklerinin diğer birlikteliklere göre daha fazla sayıda tekrar ettiği görülmektedir. Birliktelik kuralının kaldıraç değerine göre inceleme yapıldığında ise Siyasal Bilimler-Hristiyanlık Dışındaki Dinler ve Fizik-Türk Edebiyatı birliktelikleri göze çarpmaktadır. Yine kullanılan yöntem ve bulunan sonuçlar itibariyle çalışmanın Pu ve Yang (2003) ile Zhang ve Wang (2013) çalışmalarını desteklediği söylenebilir.

Elde edilen tüm bu sonuçlar kütüphane karar vericilerine yol göstermek amacıyla yapılmış olup, kütüphane envanter bilgisi, kütüphane üye profili, ödünç alma alışkanlıkları, birlikte ödünç alınan yayınlar gibi yararlı olabilecek sonuçlar üretilmiştir.

Bu bilgiler ışığında, kitap sirkülasyonun da göz önüne alınarak, yerleşim düzenlerinin tekrar oluşturulması, kütüphanenin fiziksel planlamasının daha verimli olması, kaynakların ihtiyaçları karşılayacak şekilde yenilenmesi, üyelere yeni kitap tavsiyelerinde bulunarak kullanım oranının artırılması hedeflenmektedir.

Kullanılan kütüphane otomasyon yazılımında, özellikle bazı alanlarda kullanıcının belli bir formatta veri girişine zorlanmamasından dolayı, veri düzensiz ve eksik olarak elde edilmiştir. Bu durum çalışmada bir kısıt olarak, ön işleme aşamasının çok uzun sürmesine, modelleme aşamasına geçişin yavaşlamasına ayrıca veri temizleme aşamasında bir kısım verinin kaybına yol açmıştır. Bu bağlamda yazılımın veri girişlerinde kullanıcı hata payını en aza indirecek şekilde güncellenmesi halinde, ileride yapılabilecek veri madenciliği çalışmaları için daha hızlı ve daha doğru sonuçlar üretilebileceği düşünülmektedir. Ayrıca otomasyondaki veriler, yeni açılan birimlerle birlikte artmakta ayrıca yeni gelen ve mezun olan öğrencilerle, işe yeni başlayan ve ayrılan personelle birlikte sürekli değişkenlik göstermektedir. Veri madenciliği dinamik bir süreç olduğundan yapılan çalışmanın, belirli aralıklarla daha geniş veri setleri ile tekrarlanarak sonuçların kıyaslanması modelin güvenilirliğini artıracaktır. Ayrıca günümüzde elektronik kaynakların kullanımının da giderek yaygınlaştığı düşünüldüğünde, kütüphaneye ait elektronik kaynakların da sürece dahil edildiği yeni çalışmaların yapılması önerilebilir.

KAYNAKLAR

- Abrar, V. A. and Wahyudi, M. D. R. (2016). "Implementasi Heterogenous Distributed Database System Oracle Xe 10g dan MySQL Rekam Medis Poliklinik UIN Sunan Kalijaga". *Creative Information Technology Journal*, 4(1), 9-17.
- Agrawal, R. and Srikant, R., (1994). Fast Algorithms For Mining Association Rules, *Proceedings of the 20th VLDB Conference*, Santiago, Syf. 487-499.
- Aksoy, S., (2015). Sistematik ve Alfabetik Sınıflama Düzenlerinin Bilgiye, Belgeye Konusal Erişim Açısından Karşılaştırılması, **Gece Kitaplığı**, İstanbul.
- Argüden, Y. ve Erşahin, B. (2008). Veri Madenciliği; Veriden Bilgiye, Masraftan Değere. *ARGE Danışmanlık Yayınları*, (10).
- Baradwaj, B. K., and Pal, S. (2011). Mining Educational Data to Analyze Students' Performance, *International Journal of Advanced Computer Science and Applications*, 2(6), 63-69.
- Ceylan, Z., Gürsev, S. ve Bulkan, S. (2017). İki Aşamalı Kümeleme Analizi ile Bireysel Emeklilik Sektöründe Müşteri Profiline Değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(4), 475-485.
- Chen, Y.L., Chen, J.M. and Tung, C.W. (2006). A Data Mining Approach For Retail Knowledge Discovery With Consideration of the Effect of Shelf-Space Adjacency on Sales, *Decisions Support Systems*, Sayı 3, Cilt 42, Syf. 1503- 1520.
- Cios, K. J., Pedrycz, W., Swiniarski, R.W. and Kurgan, L. A., (2007). Data Mining A Knowledge Discovery Approach. New York: **Springer Sciene+Business Media**.
- Cox, L. A., (2002) "Data Mining and Causal Modelling of Customer Behaviours", *Telecommunication Systems*, Volume 21, s. 356
- Damodar N. G. (2001) Temel Ekonometri, İstanbul, *Literatür Yayınları*, s. 16
- Decker, R., and Höppner, M. (2006), "Strategic planning and customer intelligence in academic libraries." *Library Hi Tech*, 24(4), 504-514.
- Deperlioğlu, Ö. ve Sarpkaya, Y. (2009). Öğretim Yönetim Sistemleri İçin Örnek Veri Tabanı Tasarımı. *International Journal of Informatics Technologies*, 2(1),15-21.
- Doedens, C. J. (1994). Text Databases: One Database Model and Several Retrieval Languages. Amsterdam: **Rodopi B. V.**
- Doğaç, A., Özsu, M. T., Biliris, A. ve Sellis, T. (1994). Advances in object oriented database systems. *NATOASI Series Advanced Study institute on Object Oriented Databases* (s. 4-8). İzmir: Springer.

- Doğan, İ. (2000) “Kümeleme Analizi ile Seleksiyon”, *Turkish Journal of Veterinary and Animal Sciences*, <http://journals.tubitak.gov.tr/veterinary/issues/vet-02-26-1/vet-26-1-7-0007-1.pdf>, Son erişim tarihi: 08.05.2019
- Doğan, O. (2015). Bir E-Ticaret Sitesi Kullanıcı Hesaplarında Şifre Yapılarının Birlikte Kuralları ile Güncellenmesi, *İnternet Uygulamaları ve Yönetimi Dergisi*, 6(2), s. 49-61.
- Emel G.G., Taşkın Ç. ve Tok A. (2005) ”Pazarlama Stratejilerinin Oluşturulmasında bir Karar Destek Ağacı, Birlikte Kuralları Madencilik”, *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, <http://www.sbe.deu.edu.tr/adergi/2006/Cilt%207%20Sayi%203%202005/emel-taskin-tok.pdf>, Son erişim tarihi: 09.05.2019
- Garcia-Molina, H., Ullman, J. D. and Widom, J. (2008). *Database Systems: The Complete Book*. New Jersey: **Prentice Hall**.
- Glenn J. Myatt, **Making Sense of Data**, US, John Wiley & Sons Publication, 2007, s. 110
- Guidici, P. (2003) *Applied Data Mining Statistical Methods for Business and Industry*, **Wiley**, West Sussex
- Gülten, A. ve Doğan, Ş. (2008) “Genetik Algoritmalar Yönteminin Biyomedikal Verileri Üzerinde Uygulamaları”, *Doğu Anadolu Bölgesi Araştırmaları Dergisi*, <http://web.firat.edu.tr/daum/docs/71/03>, Son erişim tarihi: 09.05.2019, s.13
- Hajek, P. and Stejskal, J. (2012), “Analysis of User Behavior in a Public Library using Bibliomining”, *Advances in Environment, Computational Chemistry and Bioscience*, Montreux, s.339
- Halevy, A. (2001). Answering queries using views: A survey. *The VLDB Journal*, 10 (4), 270--294.
- Han, J. and Fu, Y. (1995) Discovery of Multiple-Level Association Rules From Large Databases, *Proceedings of the 21st VLDB Conference*, Zurich, Syf. 420-431.
- Han, J. and Kamber, M. (2001). *Data Mining Concepts and Techniques*, **Academic Press**, San Diego
- Han, J. and Kamber, M. (2006) *Data Mining Concepts & Techniques*, **Morgan Kaufmann Publishers**, San Francisco, s. 337
- Hand, D., Mannila. H. and Smyth. P. (2001). *Principles of Data Mining*. **Massachusetts Institute of Technology**.
- Hull, R. (1997). Managing semantic heterogeneity in databases: a theoretical perspective. *PODS '97: Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems* (s. 51--61). New York: ACM.

- Inmon, W. H. (2002). Building the Data Warehouse. Canada: *John Wiley & Sons, Inc.*
- Kantardzic, M. (2001). Data Mining Concepts, Models, Methods and Algorithms. Piscataway, NJ: *IEEE Press.*
- Kaplan, M. S., Caetano, R., Giesbrecht, N., Huguet, N., Kerr, W. C., McFarland, B. H. vd. (2017). The National Violent Death Reporting System: Use of the Restricted Access Database and Recommendations for the System's Improvement. *American Journal of Preventive Medicine*, 53(1), 130-133.
- Kaya, H., and Köymen, K. (2009) "Veri Madenciliği Kavramı ve Uygulama Alanları". *Doğu Anadolu Bölgesi Araştırma ve Uygulama Dergisi*, Şubat 2008 (Çevrimiçi), <http://web.firat.edu.tr/daum/default.asp?id=79>, Son erişim tarihi:07.05.2019.
- KVKK (2016), "Kişisel Verilerin Korunması Kanununa İlişkin Uygulama Rehberi" <https://www.kvkk.gov.tr/Icerik/4197/Kisisel-Verilerin-Korunmasi-Kanununa-Iliskin-Uygulama-Rehberi/> Son erişim tarihi: 15.05.2019
- Mattison, R., (1997) "Data Warehousing and Data Mining for Telecommunications", Norwood, *Artech House*, s. 254
- Nicholson, S. (2003). The bibliomining process: Data Warehousing and Data Mining for Library Decision Making. *Information Technology and Libraries*, 22(4), 146-156.
- Nicholson, S., and Stanton, J.M. (2003). Gaining Strategic Advantage Through Bibliomining: Data Mining for Management Decisions in Corporate, Special, Digital, and Traditional Libraries. *In Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance*. 247-262.
- Ohsuga, L. and Hu L. (2005) Foundations and Novel Approaches in Data Mining, *Springer*, Warsaw, Poland
- Özkan, Y. (2008) Veri Madenciliği Yöntemleri, *Papatya Bilim*, İstanbul
- Pu, H., and Yang, C. (2003) "Enriching user-oriented class associations for library classification schemes." *The Electronic Library*, 21(2), 130-141.
- Pyle, D. (1999). *Data Preperation for Data Mining*. San Diago: *Morgan Kauffman*.
- Rahm, E., and Hai, D. H. (2000). Data cleaning: Problems and current approaches. D. B. Lomet (Dü.), *Bulletin of the technical committee on data engineering*. S. 3. Washington: IEEE Computer Society.
- Rud, O. P. (2001) Data Mining Cookbook Modeling Data for Marketing, Risk and Customer Relationship Management, *John Wiley & Sons, Inc.*, New York, , s. 16 Solved? *Building the Information Society* (3-12). Boston: Springer Boston.
- Takçı, H. ve Soğukpınar, İ. (2002), "Kütüphane Kullanıcılarının Erişim Örüntülerinin Keşfi", *Bilgi Dünyası*, Cilt 3(1), s. 12-26.

- Tatlıdil, H. (2002) Uygulamalı Çok Değişkenli İstatistiksel Analiz, **Ziraat Matbaacılık**, Ankara
- Tozak, E. Y., Ersöz, T. ve Ersöz, F. (2017). Demir-Çelik Sektöründe Meydana Gelen GG Kazalarının Veri Madenciliği Kullanılarak Analizi. **3. Ulusal Sigorta ve Aktüerya Kongresi**, Vol.3, 137-144.
- Tempelman-Kluit, N., and Pearce, A. (2014) “Invoking the User from Data to Design.” **College and Research Libraries**, 75(5), 616–640.
- Uçan, Ö. (2010), “Dijital Kütüphanelerde Veri Madenciliği Uygulamaları: Akdeniz Üniversitesi Merkez Kütüphanesi Örneği”, Yüksek Lisans, **Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü**, Antalya
- Uppal, V. and Chandwani, G. (2013), “An Empirical Study of Application of Data Mining Techniques in Library System”, **International Journal of Computer Applications (0975 – 8887)**, Volume 74– No.11
- Vural, Y. ve Sağıroğlu, Ş. (2010). Veritabanı Yönetim Sistemleri Güvenliği: Tehditler ve Korunma Yöntemleri. **Politeknik Dergisi**, 13(2), 71-81.
- Xiong, H., Pandey, G., Steinbach, M., and Kumar, V. (2006). Enhancing Data Analysis with Noise Removal. **IEEE Transactions on Knowledge and Data Engineering**. 18, s. 304 - 319. Piscataway, NJ, USA: **IEEE Educational Activities Department**.
- Yağcı, M., Ekiz, H. ve Gelbal, S. (2015). Yeni Bir Çevrimiçi Sınav Modeli Geliştirilmesi ve Uygulanması. **Journal of Kırşehir Education Faculty**, 16(1), 296-298.
- Yazıcı, A., Ögüş, E., Ankaralı, S., Canan S., Ankaralı H. (2007) “Yapay Sinir Ağları’na Genel Bakış”, **Tıp Bilimleri Dergisi**, <https://www.turkiyeklinikleri.com/article/tr-yapay-sinir-aglarina-genel-bakis-47277.html>, Son erişim tarihi: 07.05.2019
- Yılmaz, M. (2013). Küreselleşmenin Oluşumuna Zemin Hazırladığı Yeni Ekonomik Anlayış: Bilgi Ekonomisi, **Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi**, 27(1), 241-255.
- Zahedi, F. and Zare-Mirakabad, M. R. (2014). Employing Data Mining to Explore Association Rules In Drug Addicts. **Journal of AI and Data Mining**, 2(2), 135-139.
- Zhang, Q. S., and Wang, X. Y. (2013) “Research of Personalized Information Service Based on Association Rules.” **Advanced Materials Research**, 760-762, 1800–1803.
- Ziegler, P. and Dittrich, K. R. (2004). Three Decades of Data Integration—All Problems
- “Popular Decision Tree: CHAID Analysis, Automatic Interaction Detection (Basic Tree-Building Algorithm: CHAID and Exhaustive CHAID)”, <http://www.statsoft.com/Textbook/CHAID-Analysis#index> Son erişim tarihi: 07.05.2019.

ÖZGEÇMİŞ

1979 yılında Erzincan'da doğan İshak FIRAT ilk ve orta öğrenimini Erzincan'da tamamladıktan sonra 2003 yılında Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Elektrik Elektronik Mühendisliği bölümünden mezun olmuştur. 2005 yılında çalışma hayatına Türk Telekomünikasyon A.Ş. Bilişim Ağları Daire Başkanlığı'nda uzman yardımcısı olarak başlamış olup 2012 yılından itibaren Tarım ve Kırsal Kalkınmayı Destekleme Kurumu Erzincan İl Koordinatörlüğünde uzman olarak çalışmaktadır. Ayrıca 2015 yılından itibaren Erzincan Binali Yıldırım Üniversitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği Anabilim Dalında eğitimini sürdürmektedir. İshak FIRAT evli ve iki çocuk babasıdır.