

**T.C.
MANİSA CELAL BAYAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ
YAZILIM MÜHENDİSLİĞİ ANABİLİM DALI
YÜKSEK LİSANS PROGRAMI**

**APRIORI ALGORİTMASI İLE
KİTAP OYLAMALARI ANALİZİ**

Merve KÖLE

**Danışman
Dr. Emin BORANDAĞ**



MANİSA-2023

Merve
KÖLE

APRİORİ ALGORİTMASI İLE BESİN TÜKETİM ANALİZİ

2023

TAAHHÜTNAME

Bu tezin Celal Bayar Üniversitesi Hasan Ferdi Turgutlu Teknoloji Fakültesi Bölümü'nde, akademik ve etik kurallara uygun olarak yazıldığını ve kullanılan tüm literatür bilgilerinin referans gösterilerek tezde yer aldığını beyan ederim.

Merve KÖLE



İÇİNDEKİLER

İÇİNDEKİLER	I
SİMGELER VE KISALTMALAR DİZİNİ	III
ŞEKİLLER DİZİNİ	IV
TABLO DİZİNİ.....	V
TEŞEKKÜR	VI
ÖZET.....	VII
ABSTRACT	VIII
1. GİRİŞ	1
2. LİTERATÜR ÇALIŞMASI	3
3. YÖNTEMLER	6
3.1. Veri Madenciliği	6
3.2. Birliktelik Kuralları	6
3.2.1 Birliktelik Kuralları Temel Kavramlar ve Matematiksel Model.....	6
3.2.2 Birliktelik Kuralları Madenciliği Süreci	9
3.2.3 Sık Geçen Öge Kümeleri Oluşturma Metotları Alternatifleri.....	10
3.2.4 Sık Öge Küme Madenciliği.....	10
3.3. Apriori Algoritması.....	11
3.3.1 Apriori Algoritması Temel Özelliği.....	11
3.3.2 Apriori Algoritması İşleyişi ve Akış Şeması	14
3.3.3 Apriori Algoritması Özet Kodu ve Örnek Çalışma.....	16
3.4. Feature Selection (Öznitelik Seçimi).....	28
3.4.1. Feature Selection (Öznitelik Seçimi) Süreci ve Akış Şeması	28
3.4.2. Feature Selection (Öznitelik Seçimi) Yöntemleri	30
Öznitelik Seçim Yöntemleri.....	31
4. UYGULAMA.....	33
4.1. Kullanılan Teknolojiler	34
4.2.1. Microsoft Sql Server Management Studio	34
4.2.2. Weka	34
4.2.3. Python	34
4.2.4. Spyder	35
4.3. Veri Seti	35
4.3.1. Veri Seti İçeriği.....	35
4.3.2. Veri Seti Tablolar Arası İlişkiler.....	38

4.4. Kullanıcıların Beğendikleri Kitapların Yazarlarına Göre Birliktelik Kuralları Çıkartımı	39
4.4.1. Veri Madenciliği	39
4.4.2. Sonuçlar: Yazarlara Ait Birliktelik Kuralları , Sık Geçen Kümeler	47
4.5. Kullanıcıların Beğendikleri Kitaplara Göre Birliktelik Kuralları Çıkarımı Çalışması	59
4.5.2 Kitaplara Ait Birliktelik Kuralları, Sık Geçen Öge Kümeleri,.....	71
5. SONUÇLAR	79



SİMGELER VE KISALTMALAR DİZİNİ

I	Veri tabanındaki nesneler
m	Öge Sayısı
D	Veritabanına ait işlemler kümesi
T	İşlemler kümesi
TID	İşlem numarası
k-ögeküme	k ögeli nesne kümeler
n	Nesne sayısı
M	Sık öge küme
M_a	M'nin boş olmayan alt kümeleri
M_k	Sık geçen k ögeli öge kümeler
GÖA	Genişlik Ölçekli Arama
DÖA	Derinlik Ölçekli Arama
C_k	k ögeli aday öge kümeler
ε	Minumum destek değeri
C	Confidence (Güven)
G	Güven
A ⇒ B	Birliktelik Kuralı
A_m	Veri tabanındaki m. Nesne
B_n	Veri tabanındaki n. nesne
Min.	Minumum
ISBN	International Standard Book Number (Uluslararası Standart Kitap Numarası)
T-SQL	Transact-SQL (İşlem-SQL)
ARFF	Attribute-Relation File Format (Öznitelik-İlişki Dosya Biçimi)
CSV	Comma-separated Values (Virgüller Ayrılmış Değerler)

ŞEKİLLER DİZİNİ

Sayfa

Şekil 3.1. Apriori algoritması birleşme özelliği örneği- 1	12
Şekil 3.2. Apriori algoritması birleşme özelliği örneği- 2... ..	12
Şekil 3.3. Apriori algoritması budama özelliği örneği.	13
Şekil 3.4. Apriori algoritması akış şeması	15
Şekil 3.5. Klasik Apriori algoritması özet kodu.....	16
Şekil 3.6. Öznitelik Seçimi akış şeması	28
Şekil 4.1. Book-Crossing veri seti bölümleri	35
Şekil 4.2. Kitap derecelendirmeleri tablosunun görünümü.....	36
Şekil 4.3. Kullanıcılar tablosu	37
Şekil 4.4. Veri seti tablolar arası ilişkiler	38
Şekil 4.5. Açıklaması bulunmayan kitaplara ait oylamalar.....	40
Şekil 4.6. Seçilmiş kolonlar	41
Şekil 4.7. Veri setinin son hali	44
Şekil 4.8. Veri setinin dönüştürülmemiş hali	45
Şekil 4.9. Yazar kurallarına ait destek ve güven değerlerinin dağılımı	54
Şekil 4.10. 25 yazar arasındaki ilişkilerin paralel olarak gösterilmesi.....	56
Şekil 4.11. Yazarlar arası ilişkilerin paralel olarak gösterilmesi	57
Şekil 4.12. Yazarlar arası ilişki gösterimi	58
Şekil 4.13. Kitapların gruplara göre sınıflandırılması.....	61
Şekil 4.14. Puanlara göre oylamaların dağılımı	62
Şekil 4.15. Veri seti puan dağılımı ve grup sınıfı	63
Şekil 4.16. Veri seti puan dağılımı ve grup sınıfı grafiği.....	63
Şekil 4.17. Information Gain algoritma sonuçları.....	64
Şekil 4.18. Classifier Attribute Eval algoritma sonuçları	65
Şekil 4.19. Gain Ratio algoritma sonuçları	66
Şekil 4.20. One-R algoritma sonuçları	67
Şekil 4.21. Relief algoritma sonuçları	68
Şekil 4.22. Symmetrical Uncert Attribute Eval algoritma sonuçları	69
Şekil 4.23. %60 ve üzeri güven değerine sahip birliktelik kuralları- 1.....	73
Şekil 4.24. %60 ve üzeri güven değerine sahip birliktelik kuralları- 2.....	74
Şekil 4.25. %50 ve üzeri güven değerine sahip birliktelik kuralları- 1.....	75
Şekil 4.26. %50 ve üzeri güven değerine sahip birliktelik kuralları- 2.....	76
Şekil 4.27. %50 ve üzeri güven değerine sahip birliktelik kuralları- 3.....	77

TABLO DİZİNİ

	Sayfa
Tablo 3.1. Birliktelik kurallarının gösterimi	7
Tablo 3.2. Notasyonlar	14
Tablo 3.3. İşlemsel veriler.....	17
Tablo 3.4. C_1 kümeleri.....	17
Tablo 3.5. C_1 kümelerinin destek değeri	18
Tablo 3.6. M_1 kümeleri	18
Tablo 3.7. C_2 kümeleri	19
Tablo 3.8. C_2 kümeleri destek değeri	19
Tablo 3.9. M_2 kümeleri (Birleşme öncesi)	20
Tablo 3.10. C_3 kümeleri budama öncesi	21
Tablo 3.11. C_3 kümeleri birleşme ve budama adımları sonrası	25
Tablo 3.12. C_3 küme destek değeri	26
Tablo 3.13. M_3 küme.....	26
Tablo 3.14. Çıkarılan birliktelik kuralları	27
Tablo 3.15. Relief formül notasyonları.....	32
Tablo 4.1. Book-Crossing veri seti	34
Tablo 4.2. Kitap derecelendirmeleri tablosunun sayısal bilgileri	36
Tablo 4.3. Yazar puanı hesaplama öncesi	42
Tablo 4.4. Yazar puanı hesaplama sonrası.....	42
Tablo 4.5. Yazarların veri setindeki yoğunluğu.....	43
Tablo 4.6. Veri seçimlerinin sonucu	43
Tablo 4.7. Sık geçen öge küme satır sayıları	46
Tablo 4.8. M_3 sık geçen öge kümeleri.....	47
Tablo 4.9. M_4 sık geçen öge kümeleri.....	50
Tablo 4.10. Yazarlar arası birliktelik kuralları sayısal bilgiler	50
Tablo 4.11. %100 güvenilirlik değerine sahip yazar ilişkileri	51
Tablo 4.12. %80 ve üzeri güvenilirlik değerine sahip yazar ilişkileri	52
Tablo 4.13. %70 ve üzeri güvenilirlik değerine sahip yazar ilişkileri	53
Tablo 4.14. En iyi 10 kurala ait değerlendirme metrikleri.....	55
Tablo 4.15. Kitap derecelendirmeleri oylamalara ait sayısal bilgiler	59
Tablo 4.16. Veri seti puan sınırına göre eleme bilgileri.....	60
Tablo 4.17. Seyrek görünümlü matris.....	60
Tablo 4.18. Yoğun görünümlü matris.....	60
Tablo 4.19. Feature Selection algoritma sonuçları.....	69
Tablo 4.20. Seçilmiş verilere ait sayısal bilgiler	70
Tablo 4.21. Seçilmiş puan türlerinin Apriori algoritması sonuçları.....	71
Tablo 4.22. Kitaplar arası birliktelik kuralları sayısal bilgiler	71
Tablo 4.23. Kitaplar arası %100 güven değerine sahip birliktelik kuralları	71
Tablo 4.24. Kitaplar arası %70 ve üzeri güven değerine sahip birliktelik kuralları	72

TEŞEKKÜR

Çalışmamın her noktasında bana destek olan, bilgi ve tecrübeleri ile yol gösteren danışman hocam Sayın Dr. Emin Borandağ'a, öğrenim hayatım boyunca beni maddi ve manevi olarak destekleyen ve hep yanımda olan aileme yürekten teşekkür ederim.

Merve KÖLE
Manisa, 2023



ÖZET

Yüksek Lisans Tezi

APRIORI ALGORİTMASI İLE KİTAP OYLAMALARI ANALİZİ

MERVE KÖLE

Manisa Celal Bayar Üniversitesi

Fen Bilimleri Enstitüsü

Yazılım Mühendisliği Anabilim Dalı

Danışman: Dr. Emin BORANDAĞ

Dijital sistemler; okuduğumuz kitapları, satın aldığımız eşyaları ya da beğendiğimiz ürünlere ait verileri kendi yapıları içerisinde, bilgi çıkartımı için kayıt altına alırlar. Bu veriler içerisinde, veri madenciliği yöntemleri ile çıkarılacak bilgiler, dijital sistemler için büyük önem taşımaktadır. Veri madenciliği, matematiksel ve istatistiksel yöntemler yardımı ile veriler içindeki gizli ilişkileri ve bilgileri keşfederek, karar verme sürecini hızlandırmayı amaçlamaktadır. Bu bağlamda çalışmada Apriori algoritması kullanılarak, kitap oylamaları içeren veri seti üzerinde, veri madenciliği yöntemleri kullanılmıştır. Bu sayede veri seti içerisinde bulunan ilişkiler keşfedilmiş, kurallar üretilmiş ve üretilmiş kurallar ile kullanıcı beğenileri arasındaki ilişkiler ortaya çıkartılmıştır. Uygulanan Apriori algoritması ile kitapların diğer kitaplarla birlikte beğenilme ve yazarların başka yazarlarla birlikte beğenilmesini gösteren kurallar üretilmiştir. Çalışmanın sonucunda bir kitabı beğenen kişinin, başka hangi kitapları beğenebileceği ve bir yazarı beğenen kişinin, başka hangi yazarları beğenebileceğine dair kurallar çıkartılmıştır. Çalışmada üretilen kuralların, önerilen kitap listelerinde ve satış stratejilerinde kullanılması beklenmektedir.

Anahtar Kelimeler: Veri Madenciliği, Birliktelik Kuralları, Apriori Algoritması, Kitap Önerileri, Öznitelik Seçimi

ABSTRACT

M.Sc. Thesis

BOOK VOTING ANALYSIS OF APRIORI ALGORITHM

MERVE KÖLE

Manisa Celal Bayar University

Institute of Science

Department of Software Engineering

Supervisor: Dr. Emin BORANDAĞ

Digital systems record the books we read, items we purchase, or the data of the products we enjoy within their own framework for the purpose of information extraction. The information to be derived from these data through data mining methods is of great importance for digital systems. Data mining aims to speed up the decision-making process by uncovering hidden relationships and information in data via mathematical and statistical techniques. In this context, data mining techniques were applied to the dataset containing book votes through use of the Apriori algorithm in this study. In this way, the relationships within the data set were identified, rules were generated, and the connections between the generated rules and user preferences were disclosed. With the application of the Apriori algorithm, rules were produced that show books that were liked together with other books and that authors that were liked together with other authors. As a result of the study, rules were developed regarding which other books a person who liked a particular book and which other authors a person who liked a particular author may also enjoy. The rules produced in the study are expected to be used in book recommendation lists and sales strategies.

Keywords: Data Mining, Association Rules, Apriori Algorithm, Book Recommendations, Feature Selection

1. GİRİŞ

Günümüzde geleceğe yönelik yapılan tahminler çok fazla önem kazanmıştır. Geleceğe yönelik tahminler için bilgisayar teknolojileri kullanılmaktadır. Bilgisayar teknolojileri geleceğe yönelik tahminler yapabilmek için kayıt altına alınmış verileri kullanır [1].

Verilerinin kayıt altında tutulabilmesi, bu verilerin belli amaçlara yönelik kullanılmasını akla getirmektedir. Verilerden yarar sağlamak için, bu veriler üstünde yöntem ve kurallar uygulanarak kıymetli bilginin keşfedilmesi gerekmektedir [2]. Veriden bilgilerin keşfedilmesine veri madenciliği denilmektedir. Veri madenciliği veri setlerinde gizli durumda bulunan örüntü ve eğilimleri keşfetme işlemidir [3].

Veri madenciliğinde kullanılan modellerden biri de birliktelik kuralıdır. Birlikte alınan ürünlerin analizi, yani sepet analizidir. Sepet analizi ile müşterilerin birlikte aldığı ürünlerin satış bilgileri kullanılarak, müşterilerin aldığı ürünlerin birbirleri arasındaki bağıntılar ve kurallar bulunur [4]. Birliktelik kuralları tekniği için birçok algoritma geliştirilmiştir. Birliktelik kuralı algoritmalarından biri de Apriori algoritmasıdır [5].

Bu çalışmada kullanıcıların oylamış olduğu kitapları içeren veri seti üzerinde kitapların ve yazarların analizi, Apriori Algoritması kullanılarak gerçekleştirilmiştir. Bu analiz sonucunda kitaplar arası ilişkiler, bir kitabı beğenen kullanıcının başka hangi kitapları beğenebileceği, bir yazarı beğenen kullanıcının başka hangi yazarları beğenebileceğine dair kurallar çıkarılmıştır. Çıkarılan kurallar kitap öneri sistemlerinde, satış stratejilerinde ve önerilen listelerin kişiselleştirilmesinde fayda sağlanması beklenmektedir.

Çalışmamız dört adımdan oluşmaktadır. Birinci bölümünde çalışmanın amacı ve genel bilgiler verilmiştir. İkinci bölümünde Apriori algoritması ile ilgili yapılmış çalışmalar incelenmiştir. Üçüncü bölümünde veri madenciliği, birliktelik kuralları, birliktelik kurallarının matematiksel modeli ve kavramları, Apriori algoritması ve Apriori algoritmasının özet kodu açıklanmıştır. Oluşturulan model sayesinde, kullanıcıların bir kitabı beğendiğinde başka hangi kitapları beğenebileceği ve bir

yazarı beğendiğinde başka hangi yazarı beğenebileceğine ile ilgili kurallar çıkartılmıştır. Sonuç bölümünde ise elde edilen bulgular sunulmuştur.



2. LİTERATÜR ÇALIŞMASI

Veri madenciliği pazarlama, bankacılık, sigortacılık, veri tabanı analizi, karar destek sistemleri, risk analizi, belgeler arası benzerlik ve kredi risk araştırmaları alanlarında kullanıldığı için oldukça geniş bir literatüre sahiptir [6].

Veri madenciliği çalışmalarında bilgilerin verimli kullanılabilmesi için pek çok yöntem bulunmaktadır. Bu yöntemler içinde yer alan bir yöntem de birliktelik kuralıdır. Birliktelik kuralı, veri tabanı içinde bulunan kayıtlı verilerin birbiriyle olan bağıntıları inceleme işlemi yaparak, hangi olayların eş zamanlı olarak birlikte meydana gelebileceğini belirlemeye çalışan tekniklerinden biridir. Birliktelik kuralı çalışmalarında birçok ilişki tespit eden algoritma geliştirilmiştir [7]. Agrawal ve Srikant'ın 1994 yılında yaptıkları çalışmada önerdikleri Apriori algoritması bu algoritmalar arasında en yüksek performansa sahip algoritmalarından biridir [8].

Özçakır ve ark. (Özçakır, 2007) yapmış oldukları çalışmada bir pastane firmasına ait pastane satış verilerini kullanarak veri madenciliği çalışması yapabilmek için birliktelik kuralları kullanarak Microsoft .NET platformunda Visual Basic programlama dilinde bir yazılım tasarlamışlardır. Veriler 7 satış şubesinin 2002 yılının dokuzuncu ayından 2006 yılının ikinci ayına kadar olan dönemdeki satış verilerini içermektedir. Tasarlanan yazılımda, Apriori algoritması ile veriler analiz edilmiştir. Farklı zaman aralıkları ve satış lokasyonu giriş parametre değerleri göz önüne alınarak birlikte satın alınan yiyecek ve içecekler ile ilgili ilişkiler olduğu belirlenmiştir. Çalışmanın sonucunda bazı kurallar elde edilmiştir. Kurallardan bir tanesinde tuzlu ürünleri tercih eden insanların sıcak içecekleri de tercih ettiği görülmüştür [9].

Erdem ve ark. (Erdem, 2008) belirli bir zaman dilimi boyunca Ege Bölgesi'nde yer alan bir hastanenin acil servisine başvuru için gelmiş 214 bin hastaya ait verileri kullanarak, Veri Madenciliği'nin tekniği olan birliktelik kuralı ile veri setindeki saklı ama kıymetli bilgi içeren ilişkiler keşfedilmeye çalışılmıştır. Apriori algoritması kullanılarak destek değeri 0.1 ve güven değeri 0.5 belirlenerek 65 kural oluşturulmuştur. Çalışma sonucundaki bulgular, bölgelerine ait özellikler taşıyabileceği düşünülen acil servislere hastaların başvuru açma nedenleri ve hasta profilleri açısından içgörü sağlamak ve acil servis bölümlerinin yapılanma

alışmalarına da farklı bir bakış açısından bakılmasını sağlayarak katkıda bulunmaktadır [10].

Ay ve ark. (Ay, 2010) gerçekleřtirdikleri alışmada süpermarket içinde bulunan raf yerleşim yeri düzenlemede yardımcı olacak veri yönelimli olan bir karar destek sistemi uygulaması tasarlamışlardır. alışma market firmalarından biri olan Migros Türk A.Ş’de gerçekleştirilmiştir. 12.077 ürün ve 558 ürün grubundan oluşan veri seti kullanılmıştır. alışmada ilişkisel veri tabanı hazırlanarak, birliktelik kuralı algoritmalarından biri olan Apriori algoritması ile Çok Boyutlu Ölçekleme (ÇBÖ) teknikleri kullanılmıştır. alışma sonunda 518 kural elde edilmiş ve yerleşim düzeni için kullanılmıştır [11].

Gölce (2010) alışmasında, Veri Madenciliğini ve sepet analizi işlemini gerçekleřtirmek amacı doğrultusunda birliktelik kuralları üreten apriori algoritmasını detaylı şekilde ele alınmıştır. Apriori algoritması sepet analizinde yer alan ürün veri setlerinden farklı türde olan bir veri seti üzerine uygulanmıştır. Anket veri seti üzerinde Apriori algoritması kullanılarak birliktelik kurallarını üreten bir uygulama tasarlanmış ve sonuç olarak Apriori algoritmasının anket veri setlerinde de kullanılabileceği anlaşılmıştır [12].

Erpolat (2012) gerçekleřtirdiğı alışmada Türkiye’de otomotiv sektörü bölümünde bulunan bir yetkili servisin 2009 Aralık ayında müşterilerine ait faturalarını analiz etmiştir. Toplam ürün sayısı 50 ve veri setini oluşturan fatura sayısı 40’tır. Apriori ve FP-Growth algoritmaları ile veriler analize sokulmuştur. Algoritmaları alıştırmak için Weka ve veri işlemlerini gerçekleřtirebilmek için Excel programları kullanılarak verilerin analizi gerçekleştirilmiştir. alışma sonunda müşterilerin hangi ürünleri eşzamanlı satın almış oldukları gözlemlenmiştir. Bu gözlemler doğrultusunda karı artırmaya yönelik uygulamaya konulacak kampanya ve promosyonlara karar verilmeye alışılmıştır [13].

Acun ve ark. (Acun, 2015) yaptıkları alışma içerisinde mobilya ve perakende sektörü adına oluşturulmuş kaynak planlama yazılımında bulunan günlük hata loglarını biriktirerek ön işleme işlemleri uygulayarak, elde edilen veriler üzerinde birliktelik analizi algoritmalarından Apriori algoritması ile analiz eden bir yazılım üretilmiştir.

Bu yazılım yardımıyla, bir hatanın oluşmasına sebep olan etkenler arasındaki bağıntılar Apriori algoritması ile ortaya çıkarılarak, bu bağıntıları raporlayan bir yazılım geliştirilmiştir. Çalışma sonunda 1 tane beşli birliktelik, 6 tane dördü birliktelik ve 14 tane üçlü birliktelik kuralı tespit edilmiştir. Hataya en çok sebep olan bağıntılar tespit edilmiştir [14].

Doğrul ve ark. (Doğrul, 2015) yaptıkları çalışmada Çankırı iline ait trafik kazası verilerini analiz etmek için birliktelik kuralları üreten Apriori algoritmasını kullanmışlardır. Çalışmada veri seti içeriği, Çankırı Jandarma ve Polis Timlerinin müdahale ettiği 2013 yılındaki trafik kazalarından oluşmaktadır. Çalışma sonucunda kazaların genellikle yaşandığı yerler ve vakitler keşfedilmiştir. Keşfedilen bu durumlar sayesinde önlemlerin fazlaştırılması ile birlikte kazaların engellenebileceği belirtilmiştir [15].

Kurt Pehlivanoglu ve ark. (Kurt, 2015) yaptıkları çalışmada Kocaeli’nde bir ortaokulda okuyan öğrencilerin sosyal medya kullanma alışkanlıkları üzerine veri toplamış ve sonuçları birliktelik kuralları yardımıyla analiz etmişlerdir. Apriori algoritması destek parametresi 0,9 olarak belirlendiğinde 1130 kural oluşturulmuştur. Çalışmanın sonucunda ortaya çıkan birliktelikler ya da diğer ifadeyle kurallar raporlanmıştır [16].

Çelik ve ark. (Çelik, 2019) yaptıkları çalışmada Kütahya’da faaliyet gösteren bir perakende firmasının satış verileri üzerinde Apriori algoritması uygulayarak birliktelik kuralları üretmişlerdir. Kullanılan veriler üzerinde analizi etkilemeyecek kadar küçük olanlar çıkartılmıştır. Çalışma sonucunda birlikte alınan ürünler kural halinde sunulmuş ve demografik özellikleri de tespit edilmiştir [17].

3. YÖNTEMLER

3.1. Veri Madenciliği

Veri madenciliği, verilerin içerisinde bulunan desenlerin, birlikteliklerin, kuralların ve istatistiksel bakımından kıymetli bilgi taşıyan keşfedilme işlemidir [18]. Veri madenciliği yöntemleri 4 ana alt başlık altında gösterilebilmektedir [19]:

1. Sınıflandırma
2. Ardışık Zamanlı Örüntüler
3. Birliktelik Kuralları
4. Kümeleme

3.2. Birliktelik Kuralları

Birliktelik kuralları Agrawal ve arkadaşları tarafından geliştirilen, veriler veya veri kümeleri arasındaki ilişkileri elde etmek için kullanılan veri madenciliği tekniğidir [20]. Birliktelik kurallarının en yaygın kullanım alanlarından birisi pazar sepet analizidir [21]. Pazar sepet analizi yöntemi, hangi ürünlerin beraber sepette bulunma eğiliminde olduğu hakkında bilgi sunarak; raf yerleşim düzeni, müşterilerinin aldıkları ürünlere ait alışveriş davranışları, satış stratejileri gibi analizlerde faydalı olmaktadır. Sepet analizinin yanı sıra birliktelik kuralları finans, tıp gibi birliktelikler içeren değerli bilgi kazanımının mevcut olduğu alanlarda önem taşımaktadır [22]. Birliktelik kuralları algoritmalarına göre Apriori algoritması nesneler arasındaki ilişkilerin keşfedilmesi konusunda en fazla kullanılan algoritmalarındandır [10].

3.2.1 Birliktelik Kuralları Temel Kavramlar ve Matematiksel Model

1993 yılında Agrawal, Swami ve Imelinskive tarafından birliktelik kuralları incelenmiştir [20]. Birliktelik kuralları bileşenleri destek ve güven olmak üzere iki temel bileşenden oluşur [23]. Literatürde kullanılan diğer bir bileşen ise Lift kriteridir [8]. Birliktelik kurallarını Agrawal ve arkadaşları şu şekilde ifade etmişlerdir:

$I = \{i_1, i_2, \dots, i_m\}$ nesneler kümesi adı verilmektedir.

Öge sayısı m ile temsil edilmektedir. İşlem kümesi D ile ifade edilmektedir. D veri tabanındaki her işlem olmak üzere, $T \subset I$ olarak tanımlanan bir öge kümesidir. T her bir hareketi (işlemi) temsil eder. TID ise, her harekete ait olan tek ayrıçtır. TID, her işlem için benzersiz bir sayıdır [24]. Birliktelik kuralı aşağıdaki şekilde ifade edilmektedir:

$$A_1, A_2, \dots, A_m \Rightarrow B_1, B_2, \dots, B_n \text{ ya da } A \rightarrow B$$

A öncül B izleyen olarak adlandırılmıştır. Bu kural A nesnesi var iken B nesnesinin aynı olay veya hareket içinde olduğunu göstermektedir. Birliktelik kurallarının genel olarak gösterim şekli Tablo 3.1’ de ifade edildiği gibidir. Birliktelik kuralı iki bölümden oluşur, kuralın sol tarafı olan kısım (A) antecedent (öncül) ve sağ tarafı olan kısım (B) consequent (sonuç) olarak ifade edilmektedir [10].

Tablo 3.1. Birliktelik kurallarının gösterimi

Kural Numarası	Öncül 1	Öncül 2		Öncül X	Sonuç
Kural 1	...	...	...	...	...
Kural 2	...	...	...	...	...
...	...	...	...	...	...
Kural N	...	...	...	...	...

Destek: D işlemler kümesi içinde ilgili nesne kümesini içeren hareketlerin yüzdesidir.

Güven: D işlemler kümesi içerisinde yer alan A’yı içeren tüm hareketlerin içinde B’yi içermeye olasılığıdır [10].

Lift: Değişkenler arası kuvvetli birliktelikleri temsil etmektedir [8]. Destek ve Güven değerlerine ait formüller Eşitlik 3.1,3.2 ve 3.3’te ifade edilmiştir. A ve B birbirinden farklı iki öge kümesi olarak ifade edilebilir.

$$\text{Destek (A)} = A \text{ öge küme sayısı} / \text{Toplam Öge Küme Sayısı} \quad (3.1)$$

$$\text{Destek (A, B)} = (A, B) \text{ öge küme sayısı} / \text{Toplam Öge Küme Sayısı} \quad (3.2)$$

$$\text{Güven (A, B)} = (A, B) \text{ öge küme sayısı} / A \text{ Öge Küme Sayısı} \quad (3.3)$$

1 numaralı eşitlik A ögesinin tüm veri kümesinde bulunma oranını vermektedir. 2 numaralı eşitlik A ve B ögelerinin birlikte bulunma durumunda tüm veri kümesine oranını vermektedir [12]. 3 numaralı eşitlik de A ögesinin var olduğu durumlarda B ögesinin de bulunma oranını vermektedir

Birliktelik kuralı üretim aşaması için Conviction (Kanaat) ve Leverage (Kaldıraç) ölçütleri de bulunmaktadır. Eşitlik 3.4 ve 3.5'te bu kavramlara ait formüller verilmiştir.

$$\text{Kanaat}(A, B) = ((1 - \text{Destek}(B)) / (1 - \text{Güven}(A, B))) \quad (3.4)$$

$$\text{Kaldıraç}(A \rightarrow B) = \text{Destek}(A, B) - (\text{Destek}(A) \cdot \text{Destek}(B)) \quad (3.5)$$

Conviction (Kanaat) değeri hesaplanırken, A elemanlarının, B elemanı olmadan bulunma olasılıklarını hesaplamaktadır. Eğer conviction değeri 1 ise A ve B birbirinden uzak yani bağımsız durumda bulunmaktadır. Conviction (Kanaat) değeri 1 değerinden uzak ise birliktelik kuralı oluşturulabilmektedir. Leverage (Kaldıraç) değeri ise bir satış verisi üstünde A ve B ürünlerinin aynı anda sepette bulunmasının A ve B'nin ayrı ayrı olarak satılmasından ne kadar daha fazla olduğunu ifade etmektedir [25].

K-itemset (k-ögeküme): Öge küme veri madenciliği alanında itemset olarak ifade edilmektedir. Bir küme k tane öge barındırıyor ise bu küme 'k-ögeküme olarak ifade edilir. k-ögekümenin sık geçen kümeleri ise M_k şeklinde ifade edilir. Örnek; {a,p,r} 3 elemanlı içeren bir öge kümedir ve 3-ögeküme olarak belirtilir [26].

Frequent Itemset (Sık Öge Küme): Geliştirici tarafından belirlenen minimum eşik(destek) değerini sağlayan öge kümeleri sık geçen öge küme olarak ifade edilebilmektedir. k -öge kümenin sık geçen kümeleri M_k şeklinde ifade edilmektedir [26].

Minumum Destek: Araştırmacı tarafından belirlenir. Birliktelik kurallarında parametre olarak kullanılmaktadır. Belirlenen eşik değeri ile bu eşikğin altındaki değerleri elemektedir [8].

Minumum Güven: Araştırmacı tarafından belirlenir. Birliktelik kurallarında bir parametre olarak kullanılmaktadır. Belirlenen güven eşik değeri ile birlikte bu eşikğin altındaki değerleri elemektedir [27].

Güçlü Birliktelik Kuralları: Yüksek değerde güven ve destek (eşik) değerine sahip olan birlikteliklere (kurallara) güçlü kurallar adı verilmektedir. Bir D işlemler (hareketler) kümesinden beklenen birliktelik kurallarının minimum destek (eşik) ve güven şartını sağlayan $A \Rightarrow B$ kuralların var olmasıdır. Örneğin bir A kitabını alan müşteriler aynı zamanda B kitabını da alıyorsa, bu kural $A \Rightarrow B$ şeklinde gösterilir. $A \Rightarrow B$ [destek=%2, güven=%60] şu şekilde yorumlanır:

İncelenen ilgili alışveriş kayıtlarında, A ile B' nin beraber bulunma olasılığı %2; A kitabını satın alan müşterilerin %60' ı aynı alışverişte B kitabını satın aldığı veya A kitabını satın alan bir müşterinin B kitabını alma ihtimalinin %60 olduğu şeklinde yorumlanabilmektedir [9, 28].

3.2.2 Birliktelik Kuralları Madenciliği Süreci

İlişkilendirme kuralları süreci minimum destek ve asgari güven kavramlarını içermektedir [28]. Sıklıkla iki aşamada birliktelik kuralları oluşmaktadır.

Sık Geçen Nesneler (Ögeler) Kümesi Bulunması: Araştırmacı tarafından önceden karar verilmiş olan minimum destek değerini sağlayan nesne kümeleri sık geçen nesne küme olarak ifade edilmektedir. Sık geçen nesne ögelerini bulabilen etkili yöntemlerin uygulanması gereken adımdır [24]. N tane öge barındıran bir veri seti içerisinde $2^n - 1$ tane boş küme olmayan öge küme üretilir ve üretilen kümelerin hepsi sık geçen öge küme olabilir. Büyük veri setlerinde özellikle de minimum destek değerinin çok düşük olarak belirlendiği veri setlerinde sık geçen öge küme sayısı çok büyük sayılara ulaşabilir. Bunun nedeni, bir sık geçen öge kümenin ilgili bütün alt kümelerinin de sık geçen öge küme olmasından kaynaklanmaktadır. Çok büyük sayıdaki sık geçen öge kümeler herhangi bir bilgisayarın hesaplaması için çok yüksektir. [29].

Güçlü Birliktelik Kurallarının Oluşturulması: Sık geçen nesne kümeleri kullanarak belirlenmiş minimum güven eşik değerini geçen birliktelik kurallarının bulunması işlemidir. Minimum güven eşik değerine göre taranarak bulunur [24]. Kullanıcı tarafından belirlenmiş olan minimum destek (eşik) değerine göre sık geçen öge kümeler üretilir bu adımdan sonra birliktelik (ilişki) kuralları üretilir. Birliktelik kurallarının üretim sürecindeki zorlu süreç sık geçen öge kümelerinin sayısından kaynaklanmaktadır. D tane öge barındıran bir veri setinde mümkün olabilecek birliktelik kurallarının toplamı aşağıda Eşitlik 3.6' da verilmiştir.

$$3^d - 2^{d+1} + 1 \quad (3.6)$$

Bu üssel karmaşıklığa rağmen sık geçen öge kümelerini üretmek çok daha maliyetlidir. Nedeni öge kümelerinin destek (eşik) değerlerinin hesaplama işlemlerinin gerçekleştirilmesi için her bir işlemin (hareketin) tarama ve verileri doğrulama işlemine ihtiyaç duyulmasıdır. Birliktelik kuralları üretiminde kullanılan klasik bir algoritma şu şekilde ifade edilmektedir:

Her bir sık geçen öge küme M ve boş olmayan alt kümeleri M_a için minimum güven değerinden büyük ve eşit olan $M_a \Rightarrow M - M_a$ şekline sahip kurallar üretilir. Tanımlama ifadesine göre üretilen tüm kurallar minimum destek ve güven değerine uymalıdır [30].

3.2.3 Sık Geçen Öge Kümeleri Oluşturma Metotları Alternatifleri

Sık geçen öge kümeleri üretme algoritmaları arama işlemini nasıl gerçekleştirdiğine ve tarama işlemine hangi noktayı seçerek başladığına göre iki farklı gruba ayrılmaktadır [31]. Algoritmalar arama işlemini nasıl uyguladıklarına göre sınıflandırılırsa;

1. Breadth First Search (Genişlik Öncelikli Arama) algoritmaları; aşama aşama kafesi incelemektedir.

2. Depth First Search (Derinlik Öncelikli Arama) algoritmaları; kafesi, kafes içindeki bir düğümden gelecek adımda bulunan bir düğüme ya da mevcut adımda bulunan gelecek bir düğüme geçerek incelemektedir. Diğer bir sınıflandırma yöntemi ise kafesi tarama işlemi için hangi noktadan başlanması üzerinde duran algoritmalara göre iki gruba ayrılmıştır:

1. Spesific to General (Özelden-Genele) arama algoritmaları; kafesi en alt noktasını seçerek aramaya başlamaktadır.
2. General to Spesific (Genelden-özele) arama algoritmaları; kafesi en üst noktasını seçerek aramaya başlamaktadır.

3.2.4 Sık Öge Küme Madenciliği

Sık geçen öge kümeler madenciliğinin hesap sürecinde yaşanan karmaşıklığını azaltmak için birçok yöntemi bulunmaktadır. Bu alanda en çok kullanılan tekniklerden biri Apriori algoritmasıdır. Apriori algoritması bazı aday öge kümelerinin destek

değerlerini tekrar hesaplamadan bu adayları eleme sürecine sokarak ile etkili bir çözüm sunmaktadır [9].

3.3. Apriori Algoritması

1994 yılında Agrawal ve Srikant tarafından geliştirilmiş algoritmadır. Apriori algoritması ismindeki “prior(önceki)” kelimesi, bilgileri bir önceki adımdan aldığından dolayı yer almaktadır [12]. Apriori algoritması pazar sepet analizi, raf düzenlemesi, hastalıklar arası ilişki kurallarını bulmak gibi birliktelik içeren çalışmalarda kullanılmıştır [24].

Genel öge kümelerini ortaya çıkarmak için kullanılan tüm algoritmalar veri setini birçok kez tarar. Gerçekleştirilen ilk tarama işleminde, tek tek ögelerin başlangıçta kullanıcı tarafından belirlenmiş minumum destek değeri ile karşılaştırma işlemi yapılır ve her bir ögenin geniş olup olmadığı kontrol edilir. Bu adımdan itibaren her gerçekleştirilecek tarama bir önceki adımda belirlenmiş geniş ögelerden başlar ve geniş öge kümeleri oluşturulur. Oluşturulmuş geniş öge kümelerine aday öge kümeleri denilmektedir. Tarama işlemleri sona erdiğinde hangi aday öge kümelerinin, kullanıcı tarafından belirlenmiş minimum destek değerinin üzerinde bir destek değerine sahip olup gerçekten geniş aday kümesi olduğu belirlenir. Geniş aday kümeleri belirlenme aşamasından sonra gerçekleştirilen taramada bir önceki taramada belirlenmiş geniş aday kümelerinden başlanır ve veritabanı sonuna kadar destek değerleri hesaplanır. Bu işlem başka bir geniş öge kümeleri bulamadığı anda sona erer [9].

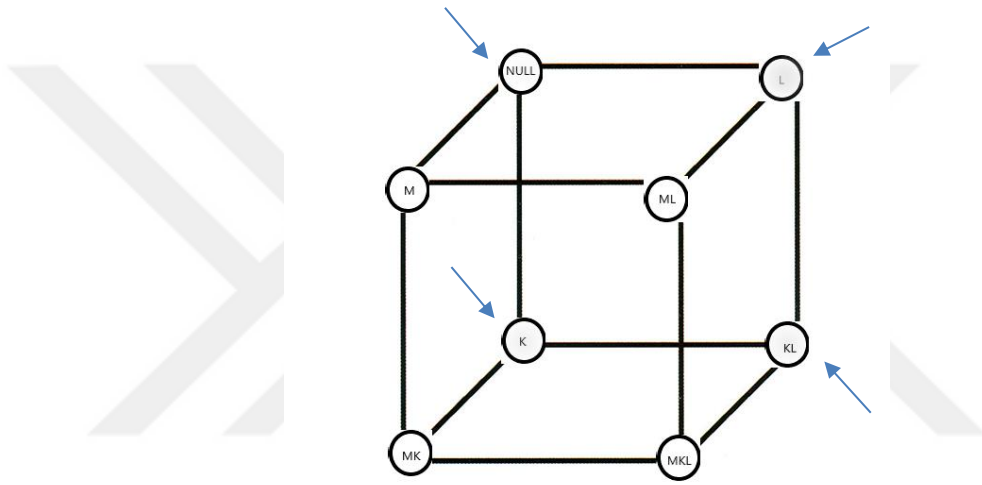
Apriori algoritması ise aday ögeleri oluşturken veri tabanındaki işlemlere dokunmadan, yalnızca bir önceki tarama sonucunda geniş olduğu ortaya çıkarılmış nesne kümelerini kullanarak aday ögelerini oluşturur. Apriori algoritması geniş bir öge kümesine ait herhangi bir alt kümesinin de geniş bir öge kümesi olacağı gerçeğine dayanır. Böylece x adet öge içeren bir öge kümesi, $x-1$ adet ögeyi içeren geniş öge kümelerinin birleştirilme işlemi ve alt kümeleri geniş olmayanların budanma işlemi ile elde edilebilir. Bu birleşme işlemi ve budama işlemi sonucunda aday kümeler daha az sayıda oluşacaktır [32].

3.3.1 Apriori Algoritması Temel Özelliği

Apriori algoritmasının temel özelliği sık geçen ögeler kümesinin alt kümelerinin de sık geçen ögeler kümesi olmasıdır. Sık geçmeyen ögeler kümesinin alt

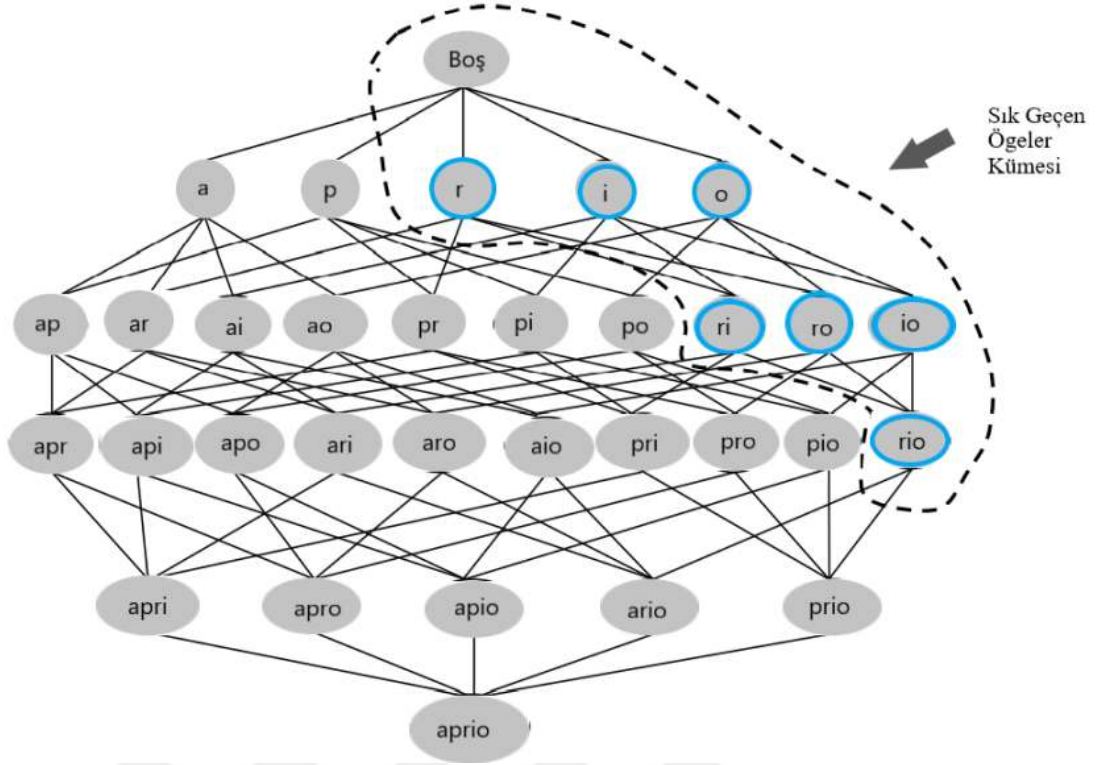
kümeleride sık geçmeyen olarak değerlendirilir. Bazı aday öge kümelerinin destek değerlerini saymadan direk eleyerek etkili bir çözüm sunar [9].

Şekil 3.1’de gösterildiği gibi oklarla gösterilen bölüm sık geçen ögeler kümesi olarak düşünülürse $\{KL\}$ kümesinin de sık geçen öge kümesi olduğu varsayılır. $\{KL\}$ kümelerini içeren her hareketin bu kümenin alt kümeleri olan $\{K\}$, $\{L\}$ kümelerini de içerdiği sonucuna ulaşılır. $\{KL\}$ sık geçen öge kümesi olarak belirlendiği ise alt kümeleri de sık geçen öge kümesi olmalıdır.



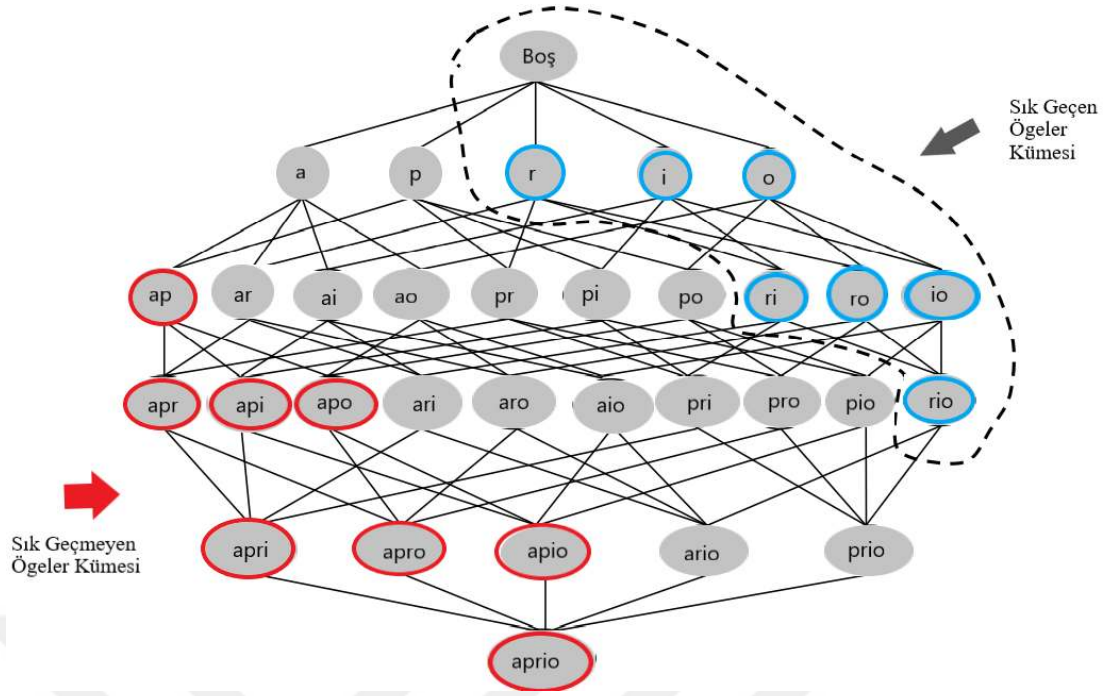
Şekil 3.1. Apriori algoritması birleşme özelliği örneği- 1

Apriori algoritmasının özelliği şu Şekil 3.2’de diğer bir örnekle açıklanmıştır:



Şekil 3.2. Apriori algoritması birleşme özelliği örneği- 2

Apriori özelliği, Şekil 3.2’ deki işaretlenmiş bölüm dikkate alınarak örnekle şu şekilde ifade edilebilir: $\{r, i, o\}$ kümesinin bir sık geçen öge küme olduğu varsayalım. $\{r, i, o\}$ kümesini içeren her hareketin bu kümenin alt kümeleri olan $\{r, i\}$, $\{r, o\}$, $\{r\}$, $\{i\}$ ve $\{o\}$ kümelerini de içerir. Bundan dolayı $\{r, i, o\}$ sık geçen öge kümesi olarak belirlendiğinde bütün alt kümeleri de sık geçen öge kümesi olmalıdır. Şekil 3.3’te budama özelliği örneği gösterilmiştir.



Şekil 3.3. Apriori algoritması budama özelliği örneği

Bu özelliği ters yönden incelediğimizde sık geçmeyen bir öge kümesinin (örneğin $\{a, p\}$) bütün üst kümeleri de sık geçmeyen öge kümesi olarak ifade edilmektedir. Şekil 3.3'te gösterildiği gibi $\{a, p\}$ kümesinin sık olmayan küme olarak belirlendiği anda $\{a, p\}$ kümesini içeren bütün üst kümeler budama işlemine sokulabilir[31].

3.3.2 Apriori Algoritması İşleyişi ve Akış Şeması

Apriori algoritması mantığı, arama olarak bilinen tekrarlamalı bir yaklaşım kullanmaktadır. Bu yaklaşımda k öğeli öge kümeleri $(k-1)$ öğeli öge kümelerin birleştirilmesi ile oluşur. İlk adım olarak 1 öğeli öge kümeleri oluşturmak için öğelerin destek değerleri hesaplanmaktadır, $(k-1)$ ' inci geçiş esnasında bulunan sık öğeler (M_{k-1}), k . geçiş esnasında sık geçen öge aday kümesini (C_k) oluşturmak için kullanılır. Veri tabanı taranır bu adımdan sonra C_k ' ların destek değerleri hesaplanır ve M_k ' lar oluşturulur. Bu şekilde algoritma sık geçen öge kümesi bulamayınca kadar devam eder ve bulamadığında sonlanmaktadır [9]. Tablo 3.2'de Apriori algoritmasının anlatımında kullanılan notasyonlar verilmiştir.

Tablo 3.2. Notasyonlar

Notasyon	Açıklama
E	Minumum destek (eşik) değeri

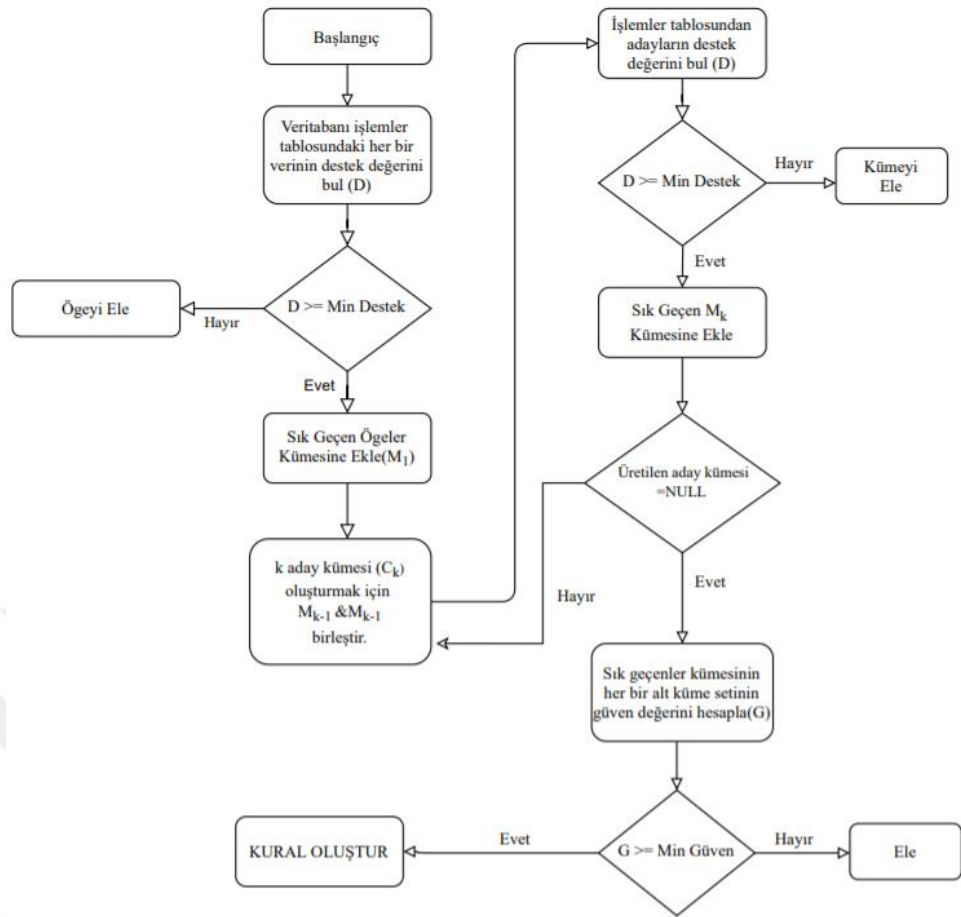
k-öge küme	k öge içeren bir öge kümesi
M_k	Sık geçen k öğeli öge kümeleri
C_k	k öğeli aday öge kümeler (potansiyel sık geçen nesne kümeler)
T	İşlemler (Hareketler) kümesi
D	Veri tabanına ait işlemler kümesi
C	Güven değeri (Confidence)
M	Sık nesne kümesi

M_{k-1} 'in M_k 'nin bulunmasında hangi adımlar ile elde ettiğini anlamak için algoritmanın Join (Birleştirme) ve Prune (Budama) işlemlerini kapsayan adımların işleyişi aşağıda açıklanmaktadır:

1. Birleştirme Adımı: M_k ' yı elde etmek için k öğeli aday öge kümeler (C_k), M_{k-1} 'in kendi arasında birleştirilmesiyle oluşturulmaktadır. m_1 ve m_2 , M_{k-1} ' de bulunan öge kümeler olsun. $m_i[j]$ gösterimi, m_i ' deki j. ögeyi temsil eder (örneğin, $m_1[k-2]$, m_1 ' deki sondan 2.ögeyi temsil etmektedir). Apriori, bir hareketteki ögelerin veya öge kümelerinin alfabetik sıraya göre sıralandığını varsaymaktadır. k-1 öğeli öge kümeler için ögeler $m_i[1] < m_i[2] < \dots < m_i[k-1]$ şeklinde sıralanmaktadır. M_{k-1} in ilk (k-2) ögeleri ortak ise $M_{k-1} \propto M_{k-1}$ birleşmesi gerçekleştirilebilir. Yani, M_{k-1} in m_1 ve m_2 öge kümeleri $(m_1[1] = m_2[1]) \wedge (m_1[2] = m_2[2]) \wedge \dots \wedge (m_1[k-2] = m_2[k-2]) \wedge (m_1[k-1] = m_2[k-1])$ şeklinde ise birleştirilmektedir. $m_1[k-1] < m_2[k-1]$ şartlı aynı adayların oluşturulmamasını sağlar. m_1 ve m_2 ' nin birleştirilmesiyle oluşan öge kümenin sonuçları $m_1[1], m_1[2], \dots, m_1[k-2], m_1[k-1], m_2[k-2]$.

2. Budama Adımı: C_k ' nın elemanları sık geçen olabilir veya olmayabilir fakat bütün sık geçen öge kümelerini içerir ve M_k ' nın üst kümesidir. C_k ' daki her bir adayın destek değerinin hesaplanması için yapılan bir taramanın sonuçları M_k ' yı belirler (örneğin, minimum destek değerinden büyük ve eşit değere sahip bütün adaylar sık geçen öge kümesi olarak tanımlanır ve M_k ' da bulunur.). C_k çok büyük 46 boyutta olabilir ve çok ağır hesaplamalar gerektirebilir. C_k ' nın boyutunu indirgemek için Apriori özelliği kullanılır. Herhangi bir sık olmayan (k-1)- öge kümesi sık geçen k-öge kümenin altkümesi olamaz. Bu sebepten dolayı, aday öge kümenin tüm (k-1) öğeli alt kümeleri M_{k-1} ' de yoksa bu aday sık geçen öge kümesi olamaz ve C_k ' dan çıkarılır [9].

Apriori algoritması akış şeması aşağıdaki Şekil 3.4'te verilmiştir:



Şekil 3.4. Apriori algoritması akış şeması

3.3.3 Apriori Algoritması Özet Kodu ve Örnek Çalışma

Agrawal ve Srikant Apriori algoritmasının çalışma detaylarını ve algoritmanın özet kodunu şu şekilde ifade etmiştir [9]. Verilerin ilk taranması sırasında, geniş öge kümelerinin tespiti için tüm öğeler sayılır. Bir sonraki tarama, k 'nıncı tarama olsun, iki aşamadan oluşur:

- Apriori-gen fonksiyonu kullanılarak, $(k-1)$ inci taramada elde edilen, M_{k-1} nesne kümeleriyle, C_k aday öge kümeleri oluşturulur,
- Sonra veri tabanı taranarak, C_k deki adayların desteği sayılır. Hızlı bir sayım için, verilen bir 1 işlemdeki, C_k 'yi oluşturan adayların çok iyi belirlenmesi gerekir.

Apriori Algoritmasının klasik özet kodu Şekil 3.5'te görülmektedir [33].

Apriori (T, ε)

$M_1 \leftarrow \{1 \text{ elemanlı sık görülen öge(ürün) kümesi}\}$

$k \leftarrow 2$

```

while  $M_{k-1}$  is not empty
     $C_k \leftarrow \text{Apriori\_gen}(M_{k-1}, k)$ 
    for transactions  $t$  in  $T$ 
         $D_t \leftarrow \{c \text{ in } C_k : c \subseteq t\}$ 
        for candidates  $c$  in  $D_t$ 
             $\text{count}[c] \leftarrow \text{count}[c] + 1$ 

     $M_k \leftarrow \{c \text{ in } C_k : \text{count}[c] \geq \epsilon\}$ 
     $k \leftarrow k + 1$ 

return  $\text{Union}(M_k)$ 

```

```

Apriori_gen( $M, k$ )
     $\text{result} \leftarrow \text{list}()$ 
    for all  $p \subseteq M, q \subseteq M$  where  $p_1 = q_1, p_2 = q_2, \dots, p_{k-2} = q_{k-2}$  and
         $p_{k-1} < q_{k-1}$ 
         $c = p \cup \{q_{k-1}\}$ 
        if  $u \subseteq c$  for all  $u$  in  $M$ 
             $\text{result.add}(c)$ 
    return  $\text{result}$ 

```

Şekil 3.5. Klasik Apriori algoritması özet kodu

Apriori algoritması ile sık geçen öge kümelerinin elde edilmesi ve algoritma adımları yukarıdaki maddelerde anlatılmıştır. Şekil 3.5’te kodları verilen ve 3.3.2 başlığında anlatılan birleştirme ve budama adımları aşağıdaki Tablo 3.3’ te verilen örnek veri tabanına göre maddeler halinde açıklanmış ve özetlenmiştir.

Tablo 3.3. İşlemsel veriler

TID	Nesne
1	Yaban, Araba Sevdası
2	Yaban, Acımak, Kuyucaklı Yusuf, Sinekli Bakkal
3	Araba Sevdası, Acımak, Kuyucaklı Yusuf
4	Yaban, Acımak, Kuyucaklı Yusuf, Araba Sevdası

5	Yaban, Araba Sevdası, Kuyucaklı Yusuf, Çalığışu
6	Kuyucaklı Yusuf, Eylül
7	Kuyucaklı Yusuf, Eylül, Sinekli Bakkal
8	Kuyucaklı Yusuf, Araba Sevdası, Sinekli Bakkal
9	Araba Sevdası, Kuyucaklı Yusuf, Çalığışu
10	Acımak, Araba Sevdası, Sinekli Bakkal

Tablo 3.3'te bir D veri tabanına ait işlemsel veriler verilmiştir. Bu işlemsel veriler işlemleri(hareketleri) temsil etmektedir. Bu veri tabanında 10 işlem bulunmaktadır, $|D|=10$ şeklinde gösterilir.

1) Algoritmanın ilk iterasyonunda her bir 1-ögelî kümeler C_1 aday kümesinin elemanıdır. Algoritma, her kitabın destek değerini hesaplamak için bütün hareketleri(işlemleri) basitçe taramaktadır. Tablo 3.4'te C_1 kümeleri gösterilmiştir.

Tablo 3.4. C_1 kümeleri

No	Kitap Adı
1	Acımak
2	Kuyucaklı Yusuf
3	Yaban
4	Çalığışu
5	Araba Sevdası
6	Sinekli Bakkal
7	Eylül

2) Destek eşik değeri (ϵ) 2 olarak belirlenmiştir ($2/10*100=\%2$). 1 ögelî öge kümeler ile sık geçen M_1 kümesi oluşturulmaktadır. Bu örneğe göre C_1 kümesindeki 7 adayın destek değerleri minimum destek değerinin üzerindedir. Tablo 3.5'te C_1 kümelerinin destek değerleri ve Tablo 3.6'da M_1 kümeleri gösterilmiştir.

Tablo 3.5. C_1 kümelerinin destek değeri

No	C_1 Öge Kümesi	Destek Değeri
1	Acımak	4

2	Kuyucaklı Yusuf	8
3	Yaban	4
4	Çalığışu	2
5	Araba Sevdası	7
6	Sinekli Bakkal	4
7	Eylül	2

Tablo 3.6. M_1 kümeleri

No	M_1 Öge Kümesi	Destek Değeri
1	Acımak	4
2	Kuyucaklı Yusuf	8
3	Yaban	4
4	Çalığışu	2
5	Araba Sevdası	7
6	Sinekli Bakkal	4
7	Eylül	2

3) 2 öğeli sık geçen öge kümeleri (M_2) algoritma $M_1 \propto M_1$ birleşmesini kullanır ve 2 öğeli öge kümeleri (C_2) oluşturur. C_2 , M_1 ' in ikili kombinasyonlarından oluşur. Tablo 3.7'de C_2 kümelerinin gösterilmiştir.

Tablo 3.7. C_2 kümeleri

No	C_2
1	Acımak, Kuyucaklı Yusuf
2	Acımak, Yaban
3	Acımak, Çalığışu
4	Acımak, Araba Sevdası
5	Acımak, Sinekli Bakkal

6	Acımak, Eylül
7	Kuyucaklı Yusuf, Yaban
8	Kuyucaklı Yusuf, Çalığışu
9	Kuyucaklı Yusuf, Araba Sevdası
10	Kuyucaklı Yusuf, Sinekli Bakkal
11	Kuyucaklı Yusuf, Eylül
12	Yaban, Çalığışu
13	Yaban, Araba Sevdası
14	Yaban, Sinekli Bakkal
15	Yaban, Eylül
16	Çalığışu, Araba Sevdası
17	Çalığışu, Sinekli Bakkal
18	Çalığışu, Eylül
19	Araba Sevdası, Sinekli Bakkal
20	Araba Sevdası, Eylül
21	Sinekli Bakkal, Eylül

4) D veri tabanındaki işlemler(hareketler) taranır ve C_2 kümesindeki öğelerin destek değerleri hesaplanır. Tablo 3.8’te C_2 kümelerinin destek değerleri gösterilmiştir.

Tablo 3.8. C_2 kümeleri destek değeri

No	C_2	Destek Değeri
1	Acımak, Kuyucaklı Yusuf	3
2	Acımak, Yaban	2
3	Acımak, Çalığışu	0
4	Acımak, Araba Sevdası	3
5	Acımak, Sinekli Bakkal	2
6	Acımak, Eylül	0
7	Kuyucaklı Yusuf, Yaban	3
8	Kuyucaklı Yusuf, Çalığışu	2
9	Kuyucaklı Yusuf, Araba Sevdası	5

10	Kuyucaklı Yusuf, Sinekli Bakkal	3
11	Kuyucaklı Yusuf, Eylül	2
12	Yaban, Çalikuşu	1
13	Yaban, Araba Sevdası	3
14	Yaban, Sinekli Bakkal	1
15	Yaban, Eylül	0
16	Çalikuşu, Araba Sevdası	2
17	Çalikuşu, Sinekli Bakkal	0
18	Çalikuşu, Eylül	0
19	Araba Sevdası, Sinekli Bakkal	2
20	Araba Sevdası, Eylül	0
21	Sinekli Bakkal, Eylül	0

5) C_2 kümesindeki minimum destek değerine sahip olan 2 ögeli öge kümeler L_2 'yi oluşturur. Tablo 3.9'da M_2 kümelerinin destek değerleri gösterilmiştir.

Tablo 3.9. M_2 kümeleri (Birleşme öncesi)

No	M_2
1	Acımak, Kuyucaklı Yusuf
2	Acımak, Yaban
3	Acımak, Araba Sevdası
4	Acımak, Sinekli Bakkal
5	Kuyucaklı Yusuf, Yaban
6	Kuyucaklı Yusuf, Çalikuşu
7	Kuyucaklı Yusuf, Araba Sevdası
8	Kuyucaklı Yusuf, Sinekli Bakkal
9	Kuyucaklı Yusuf, Eylül
10	Yaban, Araba Sevdası
11	Çalikuşu, Araba Sevdası
12	Araba Sevdası, Sinekli Bakkal

6) 3 ögeli adaylar kümesini (C_3) oluşturma işlemi için için $M_2 \propto M_2$ birleşimi kullanılmaktadır. Tablo 3.10'da budama öncesi C_3 kümelerini göstermektedir. Detayları aşağıda gösterilen 3 ögeli aday kümeleri aşağıdaki şekilde bulunmuştur:

Birleşme: $C_3 = M_2 \circ M_2 = \{\{Acımak, Kuyucaklı Yusuf\}, \{Acımak, Yaban\}, \{Acımak, Araba Sevdası\}, \{Acımak, Sinekli Bakkal\}, \{Kuyucaklı Yusuf, Yaban\}, \{Kuyucaklı Yusuf, Çalığışu\}, \{Kuyucaklı Yusuf, Araba Sevdası\}, \{Kuyucaklı Yusuf, Sinekli Bakkal\}, \{Kuyucaklı Yusuf, Eylül\}, \{Yaban, Araba Sevdası\}, \{Çalığışu, Araba Sevdası\}, \{Araba Sevdası, Sinekli Bakkal\}\} \circ \{\{Acımak, Kuyucaklı Yusuf\}, \{Acımak, Yaban\}, \{Acımak, Araba Sevdası\}, \{Acımak, Sinekli Bakkal\}, \{Kuyucaklı Yusuf, Yaban\}, \{Kuyucaklı Yusuf, Çalığışu\}, \{Kuyucaklı Yusuf, Araba Sevdası\}, \{Kuyucaklı Yusuf, Sinekli Bakkal\}, \{Kuyucaklı Yusuf, Eylül\}, \{Yaban, Araba Sevdası\}, \{Çalığışu, Araba Sevdası\}, \{Araba Sevdası, Sinekli Bakkal\}$

Tablo 3.10. C_3 kümeleri budama öncesi

No	C_3
1	Acımak, Kuyucaklı Yusuf, Yaban
2	Acımak, Kuyucaklı Yusuf, Araba Sevdası
3	Acımak, Kuyucaklı Yusuf, Sinekli Bakkal
4	Acımak, Kuyucaklı Yusuf, Çalığışu
5	Acımak, Kuyucaklı Yusuf, Eylül
6	Acımak, Yaban, Araba Sevdası
7	Acımak, Yaban, Sinekli Bakkal
8	Acımak, Araba Sevdası, Sinekli Bakkal
9	Acımak, Araba Sevdası, Çalığışu
10	Kuyucaklı Yusuf, Yaban, Çalığışu
11	Kuyucaklı Yusuf, Araba Sevdası, Yaban
12	Kuyucaklı Yusuf, Yaban, Sinekli Bakkal
13	Kuyucaklı Yusuf, Yaban, Eylül
14	Kuyucaklı Yusuf, Çalığışu, Araba Sevdası
15	Kuyucaklı Yusuf, Çalığışu, Sinekli Bakkal
16	Kuyucaklı Yusuf, Çalığışu, Eylül
17	Kuyucaklı Yusuf, Araba Sevdası, Sinekli Bakkal
18	Kuyucaklı Yusuf, Araba Sevdası, Eylül
19	Kuyucaklı Yusuf, Sinekli Bakkal, Eylül
20	Yaban, Araba Sevdası, Çalığışu
21	Yaban, Araba Sevdası, Sinekli Bakkal

b) Budama, Apriori özelliğini kullanarak:

- {Acımak, Kuyucaklı Yusuf, Yaban}'nın 2 öğeli alt kümeleri {Acımak, Kuyucaklı Yusuf}, {Acımak, Yaban} ve {Kuyucaklı Yusuf, Yaban}' dır. Bütün 2 öğeli öge kümeler M_2 ' nin üyesidir ve C_3 ' te saklanır.

- {Acımak, Kuyucaklı Yusuf, Araba Sevdası}'nın 2 öğeli alt kümeleri {Acımak, Kuyucaklı Yusuf}, {Acımak, Araba Sevdası} ve {Kuyucaklı Yusuf, Araba Sevdası}' dır. Bütün 2 öğeli öge kümeler M_2 ' nin üyesidir ve C_3 ' te saklanır.

- {Acımak, Kuyucaklı Yusuf, Sinekli Bakkal}'ın 2 öğeli alt kümeleri {Acımak, Kuyucaklı Yusuf}, {Acımak, Sinekli Bakkal} ve {Kuyucaklı Yusuf, Sinekli Bakkal}' dır. Bütün 2 öğeli öge kümeler M_2 ' nin üyesidir ve C_3 ' te saklanır.

- {Acımak, Kuyucaklı Yusuf, Çalığışu}'nın 2 öğeli alt kümeleri {Acımak, Kuyucaklı Yusuf}, {Acımak, Çalığışu} ve {Kuyucaklı Yusuf, Çalığışu}' dur. {Acımak, Çalığışu} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. **Bu yüzden C_3 ' ten çıkarılır.**

- {Acımak, Kuyucaklı Yusuf, Çalığışu}'nın 2 öğeli alt kümeleri {Acımak, Kuyucaklı Yusuf}, {Acımak, Çalığışu} ve {Kuyucaklı Yusuf, Çalığışu}' dur. {Acımak, Çalığışu} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Acımak, Kuyucaklı Yusuf, Eylül}'ün 2 öğeli alt kümeleri {Acımak, Kuyucaklı Yusuf}, {Acımak, Eylül} ve {Kuyucaklı Yusuf, Eylül}' dür. {Acımak, Eylül} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Acımak, Yaban, Araba Sevdası}'nın 2 öğeli alt kümeleri {Acımak, Yaban}, {Acımak, Araba Sevdası} ve {Yaban, Araba Sevdası}' dır. {Acımak, Eylül} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Acımak, Yaban, Sinekli Bakkal}'ın 2 öğeli alt kümeleri {Acımak, Yaban}, {Acımak, Sinekli Bakkal} ve {Yaban, Sinekli Bakkal}'dır. {Yaban, Sinekli Bakkal} M_2 'nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 'ten çıkarılır.

- {Acımak, Araba Sevdası, Sinekli Bakkal}'ın 2 öğeli alt kümeleri {Acımak, Araba Sevdası}, {Acımak, Sinekli Bakkal} ve {Araba Sevdası, Sinekli Bakkal}'dır. Bütün 2 öğeli öge kümeler M_2 'nin üyesidir ve C_3 'te saklanır.

- {Acımak, Araba Sevdası, Çalikuşu}'nın 2 öğeli alt kümeleri {Acımak, Araba Sevdası}, {Acımak, Çalikuşu} ve {Araba Sevdası, Çalikuşu}'dur. {Acımak, Çalikuşu} M_2 'nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 'ten çıkarılır.

- {Kuyucaklı Yusuf, Yaban, Çalikuşu}'nın 2 öğeli alt kümeleri {Yaban, Kuyucaklı Yusuf}, {Yaban, Çalikuşu} ve {Kuyucaklı Yusuf, Çalikuşu}'dur. {Yaban, Çalikuşu} M_2 'nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 'ten çıkarılır.

- {Kuyucaklı Yusuf, Araba Sevdası, Yaban}'nın 2 öğeli alt kümeleri {Kuyucaklı Yusuf, Araba Sevdası}, {Kuyucaklı Yusuf, Yaban} ve {Araba Sevdası, Yaban}'dır. Bütün 2 öğeli öge kümeler M_2 'nin üyesidir ve C_3 'te saklanır.

- {Kuyucaklı Yusuf, Yaban, Sinekli Bakkal}'ın 2 öğeli alt kümeleri {Yaban, Kuyucaklı Yusuf}, {Yaban, Sinekli Bakkal} ve {Kuyucaklı Yusuf, Sinekli Bakkal}'dır. {Yaban, Sinekli Bakkal} M_2 'nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 'ten çıkarılır.

- {Kuyucaklı Yusuf, Yaban, Eylül}'nin 2 öğeli alt kümeleri {Yaban, Kuyucaklı Yusuf}, {Yaban, Eylül} ve {Kuyucaklı Yusuf, Eylül}'dür. {Yaban, Eylül} M_2 'nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 'ten çıkarılır.

- {Kuyucaklı Yusuf, Çalikuşu, Araba Sevdası}'nın 2 öğeli alt kümeleri {Çalikuşu, Kuyucaklı Yusuf}, {Çalikuşu, Araba Sevdası} ve {Kuyucaklı Yusuf, Araba Sevdası}'dır. Bütün 2 öğeli öge kümeler M_2 'nin üyesidir ve C_3 'te saklanır.

- {Kuyucaklı Yusuf, Çalığışu, Sinekli Bakkal}' ın 2 öğeli alt kümeleri {Çalığışu, Kuyucaklı Yusuf}, {Çalığışu, Sinekli Bakkal} ve {Kuyucaklı Yusuf, Sinekli Bakkal}' dır. {Çalığışu, Sinekli Bakkal} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Kuyucaklı Yusuf, Çalığışu, Eylül}' nun 2 öğeli alt kümeleri {Çalığışu, Kuyucaklı Yusuf}, {Çalığışu, Eylül} ve {Kuyucaklı Yusuf, Eylül}' dir. {Çalığışu, Eylül} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Kuyucaklı Yusuf, Araba Sevdası, Sinekli Bakkal}' ın 2 öğeli alt kümeleri {Kuyucaklı Yusuf, Araba Sevdası}, {Kuyucaklı Yusuf, Sinekli Bakkal} ve {Araba Sevdası, Sinekli Bakkal}' dır. Bütün 2 öğeli öge kümeler M_2 ' nin üyesidir ve C_3 ' te saklanır.

- {Kuyucaklı Yusuf, Araba Sevdası, Eylül}' ün 2 öğeli alt kümeleri {Araba Sevdası, Kuyucaklı Yusuf}, {Araba Sevdası, Eylül} ve {Kuyucaklı Yusuf, Eylül}' dür. {Araba Sevdası, Eylül} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Kuyucaklı Yusuf, Sinekli Bakkal, Eylül}' nin 2 öğeli alt kümeleri {Sinekli Bakkal, Kuyucaklı Yusuf}, {Sinekli Bakkal, Eylül} ve {Kuyucaklı Yusuf, Eylül}' dür. {Sinekli Bakkal, Eylül} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Yaban, Araba Sevdası, Çalığışu}' nin 2 öğeli alt kümeleri {Yaban, Araba Sevdası}, {Yaban, Çalığışu} ve {Araba Sevdası, Çalığışu}' dur. {Yaban, Çalığışu} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Yaban, Araba Sevdası, Sinekli Bakkal}' nin 2 öğeli alt kümeleri {Yaban, Araba Sevdası}, {Yaban, Sinekli Bakkal} ve {Araba Sevdası, Sinekli Bakkal}' dır. {Yaban, Sinekli Bakkal} M_2 ' nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 ' ten çıkarılır.

- {Çalikuşu, Araba Sevdası, Sinekli Bakkal}'ın 2 ögeli alt kümeleri {Çalikuşu, Araba Sevdası}, {Çalikuşu, Sinekli Bakkal} ve {Araba Sevdası, Sinekli Bakkal}'dır. {Çalikuşu, Sinekli Bakkal} M_2 'nin üyesi olmadığından sık geçen öge kümesi değildir. Bu yüzden C_3 'ten çıkarılır. Tablo 3.11'de C_3 kümelerinin birleşme ve budama adımları sonrası son durumunu göstermektedir.

Tablo 3.11. C_3 kümeleri birleşme ve budama adımları sonrası

No	C_3
1	Acımak, Kuyucaklı Yusuf, Yaban
2	Acımak, Kuyucaklı Yusuf, Araba Sevdası
3	Acımak, Kuyucaklı Yusuf, Sinekli Bakkal
4	Acımak, Araba Sevdası, Sinekli Bakkal
5	Kuyucaklı Yusuf, Araba Sevdası, Yaban
6	Kuyucaklı Yusuf, Araba Sevdası, Çalikuşu
7	Kuyucaklı Yusuf, Araba Sevdası, Sinekli Bakkal

7) D veri tabanı taranır ve C_3 kümesindeki öğelerin destek değerleri belirlenir. Tablo 3.12'de C_3 kümelerinin destek değerini, Tablo 3.13'te ise M_3 kümelerini göstermektedir.

Tablo 3.12. C_3 küme destek değeri

No	C_3	Destek
1	Acımak, Kuyucaklı Yusuf, Yaban	2
2	Acımak, Kuyucaklı Yusuf, Araba Sevdası	2
3	Acımak, Kuyucaklı Yusuf, Sinekli Bakkal	1
4	Acımak, Araba Sevdası, Sinekli Bakkal	1
5	Kuyucaklı Yusuf, Araba Sevdası, Yaban	2
6	Kuyucaklı Yusuf, Araba Sevdası, Çalikuşu	2
7	Kuyucaklı Yusuf, Araba Sevdası, Sinekli Bakkal	1

Tablo 3.13. M_3 küme

No	M ₃
1	Acımak, Kuyucaklı Yusuf, Yaban
2	Acımak, Kuyucaklı Yusuf, Araba Sevdası
3	Kuyucaklı Yusuf, Araba Sevdası, Yaban
4	Kuyucaklı Yusuf, Araba Sevdası, Çalığışu

8) 4 ögeli adaylar kümesini (C₄) oluşturma işlemi için $M_3 \propto M_3$ birleşimi kullanılır. M₃ kümesinden C₄ kümeleri {Kuyucaklı Yusuf, Yaban, Araba Sevdası, Acımak}, {Kuyucaklı Yusuf, Çalığışu, Araba Sevdası, Acımak} ve {Kuyucaklı Yusuf, Yaban, Araba Sevdası, Çalığışu}'dur. Bu kümelerin destek değerleri %10 ve aşağısında (minimum destek değeri %20) olduğundan dolayı M₄ boş küme olmaktadır ve algoritma sık geçen bütün öge kümeleri bulduğu için sonlandırılır. İşlem sonlandırıldıktan sonra kurallar oluşturulur.

Minumum güven değeri 0,7 ve minumum destek (eşik) değeri 0,2 olarak alındığında 13 tane kural bulunmuştur. Oluşturulan birliktelik kurallarından bazı çıkarımlar yapılabilir. Bu kurallar incelendiğinde elimizdeki işlemsel verilere göre Eylül kitabını alanların yanında Kuyucaklı Yusuf alma olasılığı %100'dür. Tablo 3.14'te yer alan 7. birliktelik kuralını incelediğimizde Çalığışu kitabını alan insanların yanında Araba Sevdası ve Kuyucaklı Yusuf alma olasılığı %100 olarak görülmektedir. Apriori algoritması ile oluşturulmuş kurallar kitap öneri sistemlerinde fayda sağlayabilir. Tablo 3.14'te çıkarılan birliktelik kuralları gösterilmiştir.

Tablo 3.14. Çıkarılan birliktelik kuralları

No	A	→	B	Confidence (Güven)	Lift (İlginçlik)
1	Çalığışu	→	Araba Sevdası	1	1.43
2	Çalığışu	→	Kuyucaklı Yusuf	1	1.25
3	Eylül	→	Kuyucaklı Yusuf	1	1.25
4	Yaban, Acımak	→	Kuyucaklı Yusuf	1	1.25
5	Kuyucaklı Yusuf, Çalığışu	→	Araba Sevdası	1	1.43

6	Araba Sevdası, Çalığışu	→	Kuyucaklı Yusuf	1	1.25
7	Çalığışu	→	Araba Sevdası, Kuyucaklı Yusuf	1	2
8	Yaban	→	Araba Sevdası	0.75	1.07
9	Yaban	→	Kuyucaklı Yusuf	0.75	0.94
10	Acımak	→	Araba Sevdası	0.75	1.07
11	Acımak	→	Kuyucaklı Yusuf	0.75	0.94
12	Sinekli Bakkal	→	Kuyucaklı Yusuf	0.75	0.94
13	Araba Sevdası	→	Kuyucaklı Yusuf	0.71	0.89

3.4. Feature Selection (Öznitelik Seçimi)

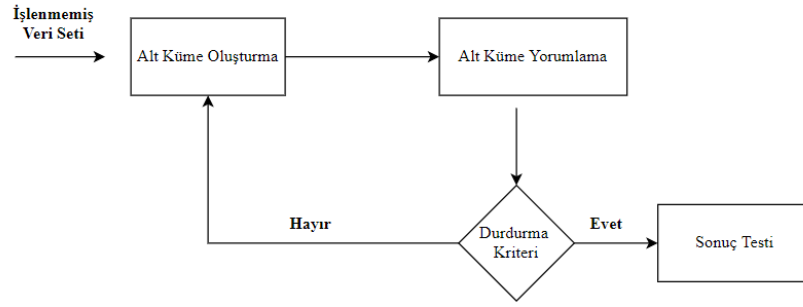
Öznitelik seçimi (feature selection), orijinal veri setini temsil edebilecek en iyi altkümenin belirlenmesi olarak ifade edilmektedir [34]. Feature selection kullanılan algoritmaya göre öznitelikleri değerlendirerek veri setindeki t adet özellik arasından en iyi c adet özelliği seçme işlemi olarak da tanımlanabilmektedir [35]. Bu işlem sayesinde veri setinde yer alan gereksiz özniteliklerin ortadan kaldırılması sağlanmaktadır bunun yanında öğrenme algoritmasının çalışma süresi kısaltılmış olmaktadır. Bu şekilde daha iyi bir sınıflandırma modeli oluşturulmuş olmaktadır [36].

Feature selection yönteminin avantajları [37] :

- Öznitelik kümesinin boyutunu azaltır.
- Algoritma hızını arttırmada etkilidir,
- İlişkili olmayan ve gürültülü veriyi çıkartır.
- Veri seti kalitesini artırır.
- Veri saklamak için gerekli olan hafıza miktarını düşürür.
- Elde edilen modelin başarısını artırır.

3.4.1. Feature Selection (Öznitelik Seçimi) Süreci ve Akış Şeması

Öznitelik seçimi adımları genel olarak Şekil 3.6'daki adımlardan oluşan bir işlemler bütünüdür [36].



Şekil 3.6. Öznitelik Seçimi akış şeması

Bu süreçte ilk olarak işlem görmemiş veri setinden bir alt öznitelik kümesi oluşturulmaktadır. Bu işlemden sonra ele alınan öznitelikler için farklı formüller kullanılarak bir değerlendirme yapılmaktadır. İlgili özniteliğin seçilmesine ya da elenmesine karar verilmektedir. Seçilen öznitelik ilgili alt kümeye dahil edilmektedir. Algoritmanın durdurma kriteri istediğine ulaşıncaya kadar işlemler devam etmektedir.

Alt Küme Oluşturma: Öznitelik alt kümesi oluşturma adımı herhangi bir işlem uygulanmamış öznitelik setinden alt küme yorumlama adımı kullanılmak üzere alt kümeler oluşturan bir arama sürecidir. Öznitelik altkümesi oluşturma işlemi, hiç öznitelik içermeyen, bütün öznitelikleri içeren ya da özniteliklerin rastgele bir alt kümesini içeren bir küme ile başlatılmaktadır. Özniteliklere başlanma durumları aşağıdaki gibidir.

Tümünü ya da hiçbir özniteliği içermeyen kümeler ile başlanması durumunda:

- Öznitelikler her bir tekrarlama eklenmektedir ya da çıkarılmaktadır.

Rastgele bir altküme ile başlatılması durumunda:

- Öznitelikler tekrarlamalı olarak elenebilmektedir, kaldırılabilir ya da rastgele olarak oluşturulabilmektedir [38] .

Alt Küme Yorumlama: Alt küme üretimi aşamasında üretilen her yeni alt küme belirli bir değerlendirme kriterleri ile değerlendirilmektedir. Bu işlemde, oluşturulan öznitelik alt kümelerinin yorumlanması yapılarak en iyi öznitelik alt kümesi bulunmaya çalışılmaktadır.

Bir alt kümenin iyiliği her zaman belirli bir değerlendirme kriterine göre belirlenmektedir [39].

Durdurma Kriteri: Durdurma kriterleri, öznitelik seçimi sürecinin ne zaman bitmesi gerektiğini bildirmektedir. Arama işleminin bitmesi, Minimum öznitelik

sayısına ulaşılması, yeterince iyi alt kümeye ulaşmak gibi durumlar durdurma kriterini örnek olarak verilebilmektedir [39].

Sonuç Doğrulama: Veri hakkındaki önsel bilgiden yararlanarak sonucu ölçmektir. Bazı durumlarda, gerçek hayat problemlerinde önsel bilgi bulunmamaktadır. Bu durumda ise öznitelik değişimine göre algoritma performanslarındaki değişimi denetleyen dolaylı yöntemler kullanılmaktadır. Bu yöntemlerde, performans belirteci olarak sınıflandırma hata oranı tercih edilmektedir [39].

3.4.2. Feature Selection (Öznitelik Seçimi) Yöntemleri

Bu bölümde öznitelik seçimi, öznitelik yöntem türleri ve öznitelik seçim yöntemleri kısaca ele alınmaktadır. Literatürde kullanılan öznitelik seçme yöntemleri bunlarla sınırlı değildir.

Öznitelik Yöntem Türleri: Özellik seçiminde kullanılan yöntemler tür bakımından üç grupta toplanmaktadır [40]:

1. İstatiksel Bilgiyi Temel Alan Filter (Filtreleme) Yöntemleri
2. Öznitelikler Üzerinde Arama İşlemleri Gerçekleştiren Wrapper (Sarmal) Yöntemler
3. En İyi Bölen Ölçütünü Bulmayı Temel Alan Embedded (Gömülü) Yöntemler

Filtreleme Yöntemleri: Filtreleme yöntemleri veri madenciliği çalışmalarında kullanılan en eski öznitelik seçim yöntemleri olarak bilinmektedir. Filtreleme yöntemlerinde herhangi bir sınıflandırıcı kullanılmadan uzaklık, bilgi, bağımlılık ve tutarlılık ölçümleri gibi istatistiksel ölçütleri temel alan fonksiyonlar yardımıyla öznitelik seçimi yapılmaktadır. Veri kümesinde bulunan her bir öznitelik için yorumlama (değerlendirme) fonksiyonu aracılığıyla bir değer hesaplanmaktadır. Fonksiyon sonuçlarına göre bu değerler içerisinde en üst değerleri alan öznitelikler en iyi öznitelik alt kümesine seçilmeye hak kazanmaktadır [34].

Sarmal Yöntemler: Sarmal yöntemlerde, öznitelik seçimi için doğru sınıflandırma oranına bağlı ölçüm kullanılmaktadır. Sınıflandırma oranına bağlı verimliliğin ölçüldüğü ve çeşitli öğrenme algoritmaları kullanılarak en iyi bilen performansını gösteren öznitelikler seçilmektedir [41].

Gömülü Yöntemler: Gömülü yöntemler, yapısında sınıflandırma algoritması ve öznitelik seçimi algoritmasını barındırmaktadır. Sınıflandırma ve öznitelik süreçlerini senkronize olarak gerçekleştirmektedirler [41].

Öznitelik Seçim Yöntemleri

Information Gain (Bilgi Kazancı): Bilgi Kazancı entropiyi temel alan bir öznitelik seçim yöntemidir. Entropi kavramı, bir sistem içindeki belirsizliğin veya tahmin edilemezliğin ölçüsü şeklinde ifade edilmektedir [42]. Bilgi Kazancı skoru, J özniteliğini tanımak için gerekli olan bilgi ile T özniteliğini de kullanılarak J özniteliğini tanımak için gereken bilgi arasındaki farkı göstermektedir. Bilgi Kazancı skorunun hesaplama işlemlerinde entropi modeli kullanılmaktadır. Entropi değeri 0 ve 1 aralığında değerler almaktadır. Bu skor 1 değerine yaklaştıkça belirsizlik artmaktadır. Yüksek entropi skoruna sahip veri daha çok bilgi içermektedir. T özniteliğine bağlı olarak J özniteliğinin entropi değerindeki değişimi gösteren Bilgi Kazancı Eşitlik 3.7, Eşitlik 3.8 ve Eşitlik 3.9'daki gibi hesaplanmaktadır [43].

$$R(J) = - \sum_{j \in J} p(j) \log^2(p(j)) \quad (3.7)$$

$$R(J|T) = \sum_{t \in T} p(t) \sum_{j \in J} p(j|t) \log_2(p(j|t)) \quad (3.8)$$

$$\text{Bilgi Kazancı} = R(J) - R(J|T) \quad (3.9)$$

Classifier Attribute Eval (Sınıflandırıcı Özellik Değerlendirici): Bir öznitelik değeri, kullanıcı tarafından seçilmiş olan bir sınıflandırma algoritması yardımı ile hesaplanmaktadır [44].

Gain Ratio (Kazanç Oranı): Kazanç Oranı simetriye sahip olmayan bir ölçüt tipidir. Bu ölçüt bilgi kazancının çok çeşitli değerlere sahip özniteliklerini seçme eğiliminin önüne geçilmesi amacı ile kullanılmaktadır. Kazanç Oranı değerleri 0-1 aralığında yer almaktadır. Kazanç oranı değeri 1'e eşit olduğunda T bilgisinin tamamen J bilgisini tahmin edebildiğini göstermektedir. Kazanç oranı 0'a eşit olduğunda ise J ile T arasında hiçbir bağıntı olmadığını göstermektedir. Kazanç Oranı Eşitlik 3.10'daki gibi hesaplanmaktadır [45].

$$\text{Kazanç Oranı} = \frac{\text{Bilgi Kazancı}}{R(T)} \quad (3.10)$$

One-R: One-R, Holte tarafından tavsiye edilmiş basit bir öznitelik seçim yöntemidir [46]. One-R algoritmasında, birinci olarak eğitim veri setindeki her öznitelik adına bir kural oluşturulmaktadır. İkinci olarak oluşturulmuş her bir kural adına sınıflandırma doğrulukları hesaplanmaktadır. Daha sonra en az hatalı kurala ait öznitelik seçilmektedir [45].

Bu doğrultuda veri setinde yer alan her bir değişken için bir model oluşturulmaktadır. Oluşturulan her bir model adına veri seti modeli oluşturan değişkenin alabileceği farklı değerler göz önüne alınarak bölünmektedir. Bu adımdan sonra her bir farklı değer için daha sık geçen sınıf temel sınıf olarak kabul edilmektedir ve diğer sınıfa ait örnekler ise hata olarak kabul edilmektedir.

Her bir farklı değer için elde edilen hata toplandığında modelin toplam hatası elde edilmektedir. Modeller toplam hatası küçük olandan büyük olana doğru sıralanmaktadır. Daha sonra önceden belirlenmiş olan eşik değerinin altında kalan modeli oluşturan değişken veri setinden atılmaktadır. Bu şekilde yeni veri seti elde edilir. Birçok öznitelik seçim yöntemi ile karşılaştırıldığında daha basit olmasına rağmen bu yöntemin birçok veri seti için başarılı sonuçlar verdiği görülmüştür [47].

Relief: Kira ve arkadaşları tarafından 1992 yılında geliştirilmiştir. Relief algoritması, öznitelikleri birbirleri ile olan bağlantıya göre ağırlıklandıran bir öznitelik seçim yöntemi olarak ifade edilen bir algoritmadır [48]. Bu yöntem iki sınıfı olan veri setleri için başarılı sonuç verebilmektedir ama çoklu sınıfa sahip veri setleri için yöntem çalışmamaktadır. 1994 yılında bu problemi çözmek adına Kononenko çoklu sınıfı olan veri setleri için çalışan ReliefF algoritmasını geliştirmiştir.

Relief algoritmasını oluşturan üç temel bölüm aşağıdaki gibidir [49].

1. Aynı sınıfa sahip olan en yakın örnekteki ilgili öznitelik değeri ve farklı sınıfa sahip en yakın örnekteki ilgili öznitelik değerinin belirlenmesi,
2. İlgili özniteliğin ağırlığının bulunması,
3. Özniteliğin ağırlıklarına göre sıralama yapılması ve belirli bir eşik değeri veren üstteki u adet özniteliğin seçilmesi.

Relief algoritması ilk aşama olarak tüm özniteliklerin ağırlıkları 0 olarak belirleme işlemini tamamlar. Bu belirleme aşamasından sonra her bir adımda veri setinden tesadüfî bir veri seçilmektedir ve bu veri ile aynı sınıfa ait en yakın k adet veri bulunmaktadır. Ardından her bir farklı sınıfa ait en yakındaki veriler bulunmaktadır. Daha sonra her bir özniteliğe ait ağırlıklar bu verilerden yararlanarak güncellenmektedir. Son adımda belirlenen koşulu sağlayamayan öznitelikler veri

setinden çıkarılarak yeni veri seti oluşturulmaktadır. Relief algoritmasının 2. adımındaki öznitelik ağırlıklarının bulunma işlemi Eşitlik 11'deki formülün n kez tekrarlanması yoluyla bulunmaktadır [50]. Tablo 3.15'te Relief formülüne ait notasyonlar Eşitlik 3.11'de verilmiştir.

$$V_s = V_{s-1} - (x_s - \text{nearHit}_s)^2 + (x_s - \text{nearMiss}_s)^2 \quad (3.11)$$

Tablo 3.15. Relief formül notasyonları

Notasyon	Açıklama
N	Örnek sayısı
V_s	s özniteliğinin ağırlığını
nearHit_s	Aynı sınıfa sahip en yakın örnekteki öznitelik değeri
nearMiss_s	Farklı sınıfa sahip en yakın örnekteki öznitelik değeri

Symmetrical Uncertainty (Simetrik Belirsizlik Katsayısı): Simetrik belirsizlik katsayısı, bilgi kazancının negatif yanını ortadan kaldırmak için kullanılmaktadır. Bilgi kazancını J ve T'nin entropi değerlerinin toplamına bölerek hesaplanmaktadır. Simetrik belirsizlik katsayısı 0-1 aralığında değer almaktadır. Simetrik belirsizlik katsayısı 1'e eşit olduğu durumda T bilgisinin J bilgisini tahmin edebildiğini göstermektedir. 0'a eşit olduğu durumda ise J ile T arasında hiçbir bağıntı olmadığını göstermektedir [45]. Simetrik Belirsizlik Katsayı formülü aşağıdaki Eşitlik 3.12'de verilmiştir.

$$\text{Simetrik Belirsizlik Katsayısı} = 2 \frac{\text{Bilgi Kazancı}}{R(J) + R(T)} \quad (3.12)$$

4. UYGULAMA

Dijitalleşmenin hızla arttığı bugünlerde ürünler arasındaki bağıntıların bulunması satış yöneticilerine, raf dizilimlerine, web sayfası önerilen ürünler listelerine katkı sağlayacaktır.

Bu kapsamda çalışmanın amacı; 0'dan 10'a kadar puanlanmış kitap oylamaları içeren veri setini kullanarak, A yazarını beğenen kişinin başka hangi yazarları beğenebileceği, C kitabını beğenmiş bir kişinin başka hangi kitapları beğenebileceğini tahmin etmek ve aralarındaki ilişkileri tespit etmektir. Çalışmada veri madenciliği, öznitelik seçimi ve birliktelik kuralları yöntemleri kullanılarak yazar ve kitaplara ait kurallar elde edilmiştir ve çıkarılmış olan kurallar değerlendirilmiştir.

4.1. Kullanılan Teknolojiler

Çalışmada kullanılan bilgisayarın özellikleri; işlemci modeli Intel(R) Core (TM) i7-6700HQ, işlemci hızı 2.60 GHz, sistem belleği 16 GB olup işletim sistemi Windows 10'dur.

Veri tabanı işlemleri için Microsoft Sql Server Management Studio v 18.10 programı, öznitelik seçimi (feature selection) işlemleri için ise Weka programı kullanılmıştır. Çalışmada birliktelik kurallarını çıkartmak için kullanılan Apriori algoritması ve görselleştirme kodları Spyder 4.2.5. Python kodlama dilinde yazılmıştır.

4.2.1. Microsoft Sql Server Management Studio

SQL Server Management Studio programı, Microsoft SQL Server içindeki tüm bileşenleri yapılandırmak ve organize etmek için kullanılan, ilk olarak Microsoft SQL Server 2005 ile başlatılan bir yazılım uygulamasıdır. Bu editör yardımıyla T-SQL(Transact-SQL) komutlarını kullanarak veriler üzerinde veri ekleme, düzenleme, raporlama ve analiz gibi işlemler yapılabilmektedir. [51].

4.2.2. Weka

WEKA yazılımı, veri ön işleme araçlarının ve makine öğrenme algoritmalarının bir araya getirildiği bir yazılım varlığıdır. Yeni Zelanda'da Waikato Üniversitesi'nde Java yazılım dili ile oluşturulmuştur. Dağıtık ve büyük ölçekli veri tabanlarında kullanılabilir. WEKA programını kullanarak veri hazırlama işlemleri gerçekleştirilebilir. Sınıflama, birliktelik analizi, kümeleme algoritmaları ile kurallar oluşturulabilmekte ve veri seti üzerinde öznitelik seçim yöntemleri uygulanabilmektedir. ARFF (Attribute-Relation File Format), CSV (Comma-separated values), C4.5 uzantılı dosya tiplerini çalıştırabilmektedir [52].

4.2.3. Python

Python, yorumlanmış üst düzey amaçlı bir programlama dilidir. Tasarım felsefesi, önemli girinti kullanımıyla kod okunabilirliğini vurgulamaktadır. Dilinin yapısı ve nesneye yönelik yaklaşımı, programcıların küçük ve büyük ölçekli projeler için açık, mantıklı kodlar yazmasına yardımcı olmayı amaçlamaktadır [53]. Çalışmada kullanılan paketler aşağıdaki gibidir.

1. `import numpy as np`
2. `import pandas as pd`
3. `import apriori, association_rules`

4. `import preprocessing`
5. `import parallel_coordinates`
6. `import matplotlib.pyplot as plt`

4.2.4. Spyder

Python programlama dilinde bilimsel programlama için açık kaynaklı, platformlar arası entegre bir geliştirme ortamıdır. Spyder, NumPy, SciPy, Matplotlib, pandas, IPython'un yanı sıra diğer açık kaynaklı yazılımlar dahil olmak üzere bilimsel Python yığınındaki bir dizi önemli paketle iş birliği yapmaktadır [54].

4.3. Veri Seti

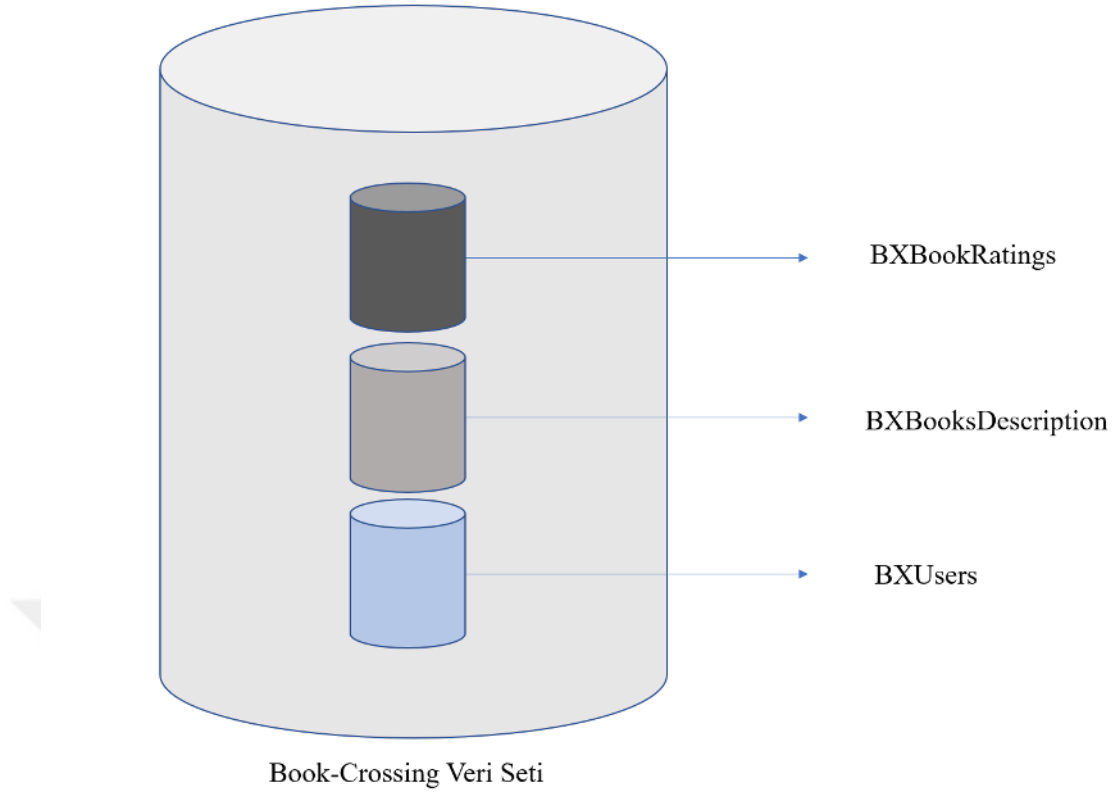
Tez çalışmasında kullanılan Book-Crossing veri seti, kitapların kullanıcılar tarafından derecelendirmelerini içeren bir koleksiyondur. Veriler Cai-Nicolas Ziegler tarafından derlenmiştir [57].

4.3.1. Veri Seti İçeriği

Book-Crossing veri seti, kitaplara ait derecelendirmeler (BXBookRatings), kitapların özellikleri (BXBookDescription), kullanıcıların özellikleri (BXUsers) olmak üzere 3 bölümden oluşmaktadır. Şekil 4.1'de veri setinin bölümleri görülmektedir. 3 bölümü kapsayan veri seti hakkındaki sayısal bilgiler Tablo 4.1'de verilmiştir.

Tablo 4.1. Book-Crossing veri seti

Veri	Sayısal Değeri
Toplam Oy Sayısı	1.149.779
Toplam Kullanıcı Sayısı	278.858
Toplam Kitap Sayısı	271.065
Oy / Kitap	4,24170
Oy / Kullanıcı	4,123170
Kitap / Kullanıcı	0,97205



Şekil 4.1. Book-Crossing veri seti bölümleri

Kitaplara Ait Derecelendirmeler (BXBookRatings Tablosu): Kitaplara ait derecelendirmeler, kullanıcıların kitaplara vermiş olduğu 0-10 arası oylardan oluşmaktadır. Apriori algoritması Book-Crossing veri seti üzerinde çalışacaktır. Bu bölümde UserID (benzersiz kullanıcı numarası), ISBN (benzersiz kitap numarası), BookRating (Oy Puanı) sütunları bulunmaktadır. Şekil 4.2. veri setinin “Kitaplara Ait Derecelendirmeler” kısmını yansıtmaktadır. UserID kullanıcının benzersiz numarasını, ISBN kitabın benzersiz numarasını ve BookRating kullanıcının kitaba verdiği oyu göstermektedir.

	UserID	ISBN	BookRating
1	2	0195153448	0
2	7	034542252	0
3	8	0002005018	5
4	8	0060973129	0
5	8	0374157065	0
6	8	0393045218	0
7	8	0399135782	0
8	8	0425176428	0
9	8	0671870432	0
10	8	0679425608	0
11	8	074322678X	5
12	8	0771025661	0
13	8	0771074670	0
14	8	080652121X	0
15	8	0887841740	5
16	8	1552041778	5
17	8	1558746218	0

Şekil 4.2. Kitap derecelendirmeleri tablosunun görünümü

Veri setini daha iyi anlayabilmek adına aşağıdaki veriler hazırlanmıştır. Tablo 4.2'deki bilgiler işlenmemiş veri setine göre hesaplanmıştır ve her oy puanına göre dağılımlarını göstermektedir.

Tablo 4.2. Kitap derecelendirmeleri tablosunun sayısal bilgileri

Puan Tipi	Toplam Oy	Toplam Kitap	Toplam Kullanıcı
10 Puan	78.609	9.975	7.345
9 Puan	67.541	8.734	7.597
8 Puan	103.736	13.887	11.656
7 Puan	76.457	9.892	9.761
6 Puan	36.924	4.285	5.353
5 Puan	50.974	5.770	6.217
4 Puan	8.904	660	1.268
3 Puan	5.996	404	819
2 Puan	2.759	130	319
1 Puan	1.770	65	197
0 Puan	716.109	99.427	26.968

Kitap Açıklamaları Tablosu (BXBookDescription: Veri setinde BXBookDescription tablosu olarak yer almaktadır. 271.065 farklı kitap bu tabloda yer almaktadır. 33.122 farklı yazara ait kitap bulunmaktadır. Kitaplara ait ISBN (Benzersiz Kitap Numarası), BookTitle (Kitap Başlığı), BookAuthor (Kitap Yazarı), YearOfPublication (Kitap Basım Yılı), Publisher (Yayıncı), ImageURLS, ImageURLL ve ImageURLM bilgileri bulunmaktadır.

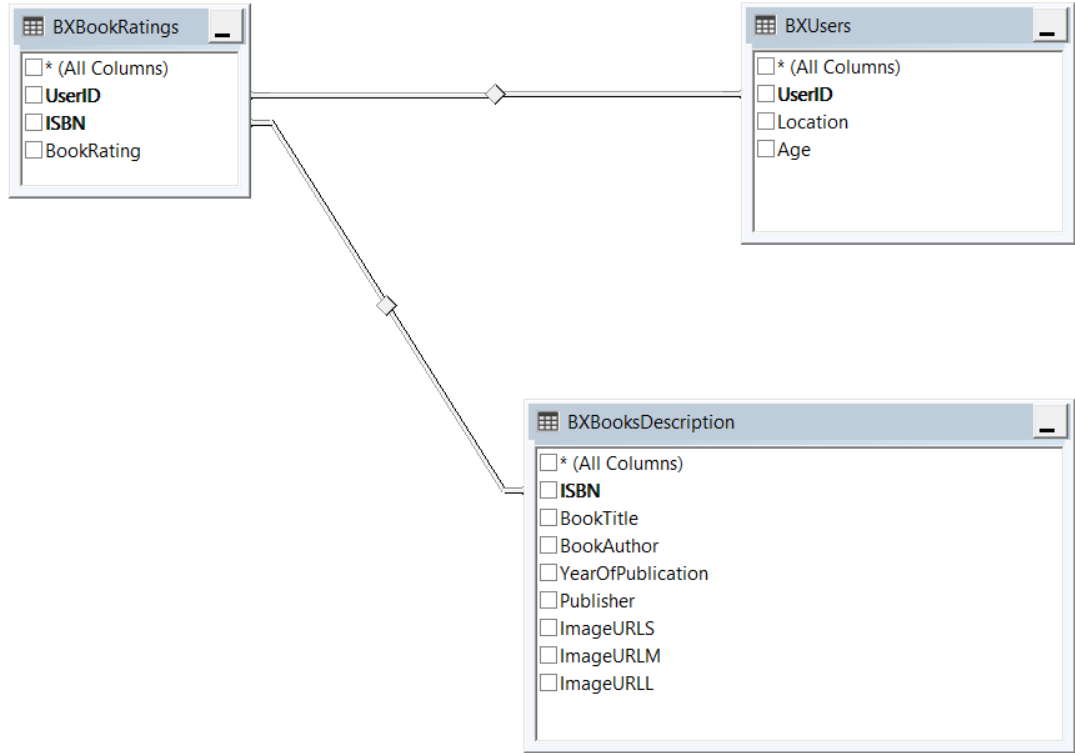
Kullanıcılar Tablosu (BXUsers Tablosu): Veri setinde BXUsers bölümünde 258.758 kullanıcı yer almaktadır. Kullanıcılara ait UserID (Benzersiz kullanıcı numarası), Location (Konum bilgisi), Age (Yaş) bilgisi tutulmaktadır. Şekil 4.3. BXUsers tablosunu yansıtmaktadır.

UserlD	Location	Age
1	nyc, new york, usa	NULL
2	stockton, california, usa	18
3	moscow, yukon territory, russia	NULL
4	porto, v.n.gaia, portugal	17
5	farnborough, hants, united kingdom	NULL
6	santa monica, california, usa	61
7	washington, dc, usa	NULL
8	timmins, ontario, canada	NULL
9	germantown, tennessee, usa	NULL
10	albacete, wisconsin, spain	26
11	melbourne, victoria, australia	14
12	fort bragg, california, usa	NULL
13	barcelona, barcelona, spain	26
14	mediapolis, iowa, usa	NULL
15	calgary, alberta, canada	NULL
16	albuquerque, new mexico, usa	NULL
17	chesapeake, virginia, usa	NULL

Şekil 4.3. Kullanıcılar tablosu

4.3.2. Veri Seti Tablolar Arası İlişkiler

Book-Crossing veri seti kullanıcı bilgilerin yer aldığı BXUsers, kitap özelliklerinin yer aldığı BXBookDescriptions ve oylamaların yer aldığı BXBookRatings olmak üzere 3 tablodan oluşmaktadır. Şekil 4.4'te veri setinde yer alan tablolara ait ilişkiler gösterilmektedir.



Şekil 4.4. Veri seti tablolar arası ilişkiler

Oylamaları içeren tabloda (BXBookRatings) UserID kolununu kullanarak BXUsers tablosundan kullanıcılara ait bilgilere ulaşılabilir. Aynı tablo üzerinden (BXBookRatings) ISBN kolonu kullanılarak kitap bilgilerinin bulunduğu BXBookDescription'a ulaşılabilir. UserID ve ISBN kolonları sayesinde tablolar birleştirilebilir.

4.4. Kullanıcıların Beğendikleri Kitapların Yazarlarına Göre Birliktelik Kuralları Çıkartımı

Çalışmanın amacı, kullanıcıların oy verdiği kitapların yazar bilgilerini ve verdiği oylama puanları kullanarak yazarlar arasındaki bağıntıları bulmaktır. Bu çalışmanın sonucunda, A yazarını beğenen bir kullanıcının başka hangi yazarları beğenebileceği kurallar halinde elde edilecektir. Bu kurallar sonucunda ortaya çıkan kitaplar arası ilişkiler ve yazarlar arası bağıntılar kitap öneri sistemlerinde, önerilen ürün listelerinde fayda sağlanması beklenmektedir.

4.4.1. Veri Madenciliği

Tezde kullanılacak olan veri tabanı, veri madenciliği aşamaları olan aşağıdaki süreçler akışında oluşturulmuştur:

1. Verilerin Oluşturulması
2. Verilerin Temizlenmesi

3. Verilerin Bütünleştirilmesi
4. Veri Seçilmesi
5. Verilerin Dönüşümü
6. Veri Madenciliği Algoritmasının Çalıştırılması

Verilerin Oluşturulması: Çalışmada kullanılacak verilerin oluşturulması için Sql Server Management Studio kullanılmıştır. Book-Crossing veri setine ait SQL dosyaları kullanılmıştır. T-SQL dili kullanılarak veriler veri tabanına eklenmiştir. Veri seti 1.149.779 oydan, 271.065 kitaptan ve 278.858 kullanıcıdan oluşmaktadır.

Verilerin Temizlenmesi: Veri setinde ISBN numaralarındaki hatalar giderilmiştir. BXBookRatings tablosunda oy verildiği halde BXBooksDescription tablosunda yer almayan kitaplara ait oylamalar veri setinden elenmiştir. Bu eleme sonucunda 118.604 oy elenmiştir. Eleme işlemi SQL Server Management Studio’da T-SQL kodlama dilinde gerçekleştirilmiştir. Oylama verilerinin yer aldığı BXBookRatings tablosu ile kitap açıklamalarını içeren BXBooksDescription tabloları birleştirilmiş ve eşleşmeyen oylamalar elenmiştir. Elemenin temel sebebi Şekil 4.5’te gösterilmiştir. Bu eleminin sebebi, oylamalarda yer alan kitaplara ait bilgilerin BXBooksDescription tablosunda bulunmamasından dolayıdır.

	UserID	ISBN	BookRating	ISBN	BookTitle	BookAuthor	YearOfPublication	Publisher	ImageURLS	ImageURLM	ImageURLL
1	52542	2 23405403.6	8	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
2	169728	9722018795	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
3	126074	0271684097	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
4	259004	8701143662	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
5	89891	8489707367	4	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
6	212898	0440141982	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
7	189848	0752859005	10	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
8	2389	342660731X	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
9	142202	8440206879	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
10	156534	0001821040	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
11	95902	8475968279	8	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
12	98436	3763216510	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
13	57412	9724223329	9	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
14	105455	842397233X	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
15	118135	9724223329	7	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
16	259057	842397233X	6	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
17	29169	1554040159	9	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
18	107645	8483310023	9	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
19	22824	9722104071	8	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
20	230522	0061071285	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
21	278418	0061071285	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Şekil 4.5. Açıklaması bulunmayan kitaplara ait oylamalar

Verilerin Bütünleştirilmesi: Bu işlem, Sql Server Management Studio kullanılarak T-SQL kodlama dili ile gerçekleştirilmiştir. Gerekli olan UserID, ISBN, BookRating ve BookAuthor veri kolonlarına ulaşabilmek için BXBookRatings tablosu ile BXBooksDescription tablosu ISBN kolonu kullanılarak birleştirilmiştir.

SELECT * FROM BXBOOKRATINGS INNER JOIN BXBOOKSDESCRIPTION
ON BXBOOKRATINGS.ISBN=BXBOOKSDESCRIPTION.ISBN

Verilerin Seçilmesi: Çalışmada kullanılacak verilerin seçilmesi için kullanılan adımlar aşağıda gösterilmektedir:

Gerekli Kolonların Seçilmesi: Birliktelik kuralları çıkarmak için veri setinden UserID, BookAuthor, BookRating kolonları seçilmiştir. Şekil 4.6’da seçilmiş kolonlar gösterilmektedir.

	UserID	BookAuthor	BookRating
1	2	Mark P. O. Morford	0
2	8	Richard Bruce Wright	5
3	8	Carlo DEste	0
4	8	Gina Bari Kolata	0
5	8	E. J. W. Barber	0
6	8	Amy Tan	0
7	8	Robert Cowley	0
8	8	Scott Turow	0
9	8	David Cordingly	0
10	8	Ann Beattie	5
11	8	David Adams Richards	0
12	8	Adam Lebor	0
13	8	Sheila Heti	5
14	8	R. J. Kaiser	5
15	8	Jack Canfield	0
16	8	Loren D. Estleman	6
17	8	Robert Hendrickson	6
18	8	Julia Oliver	7
19	9	John Grisham	0
20	9	Toni Morrison	6
21	9	The Onion	0
22	10	Celia Brooks Brown	0
23	12	J. R. Parrish	10

Şekil 4.6. Seçilmiş kolonlar

Yazar Ortalama Puanının Bulunması: Seçilmiş kolonlar üzerinden yazar puanı hesaplanmıştır. Kullanıcının herhangi bir yazarın kitabına verdiği oy direkt olarak alınmaktadır ancak aynı yazarın birden fazla kitabını okuduğu durumlarda yazara verdiği puanların ortalaması alınmıştır.

Aşağıdaki Tablo 4.3’te görüldüğü gibi UserID’si 1 olan kullanıcı için B yazarının kitabına verdiği puan 7 kabul edilmiştir. UserID’si 2 olan kullanıcı için B yazarının iki farklı kitabını okuduğu için kitap oylamasının ortalaması alınmıştır.

Sonuçlar Tablo 4.4'te gösterildiği gibidir. B yazarı UserID'si 2 olan kullanıcı için 9 olarak seçilmiştir.

Tablo 4.3. Yazar puanı hesaplama öncesi

UserID	BookAuthor	BookRating
1	B	7
2	B	10
2	B	8
3	C	5
3	D	2

Tablo 4.4. Yazar puanı hesaplama sonrası

UserID	BookAuthor	BookRating
1	B	7
2	B	9
3	C	5
4	D	2

5 ve Üzeri Olan Oylara Ait Verilerin Seçilmesi: Çalışmada kullanıcıların beğendiği yazarlar üzerinden kural oluşturulacağı için veri setinden yazar ortalaması 5 ve üzeri olanlar seçilmiştir. Bu sayede yazarlar kullanıcıların beğendiği yazarlar olacaktır. Çalışmada kullanılan veri seti bir alışveriş sepeti tarzında değil bir oylama olduğundan bu avantajı çalışmaya kazandırmıştır.

Eleme yapmadan önce 797.908 oy bulunmaktadır. 5 ve üzeri yazar ortalaması seçildikten sonra toplam oy sayısı 279.915 olmuştur.

Yoğunluğu Ortalama Altında Olan Seyrek Verilerin Elenmesi : Veri setinde toplamda 1 kez kitabı oylanmış yazarların kural oluşturması imkansızdır. Kural oluşturmak için 1'den fazla oylamada mutlaka geçmek zorundadır bunun nedeni bağıntı kurulması için Apriori algoritmasının gerek duyduğu yoğunluk ilişkisidir. Bu sebepten dolayı tüm yazarların veri setinde bulunma oranları hesaplanmıştır ve yazarların veri setinde ortalama geçme miktarı bulunmuştur. Bu verinin altında kalan yazarlar elenmiştir. Tablo 4.5'te yazarların veri seti içerisindeki yoğunlukları verilmiştir.

Tablo 4.5. Yazarların veri setindeki yoğunluğu

Veri	Sayısal Değeri
Toplam Benzersiz Yazar Sayısı	58.238
Toplam Oy Sayısı	279.195
Yazar / Oy	4,79403

Tablo 4.5'te yer alan sonuçlara göre yazarlar ortalama olarak veri setinde 4.79 kez geçmiştir. Her yazar ortalama olarak 5 kullanıcının beğendiği yazarlar arasındadır. Buradan yola çıkarak veri setinde yoğunluk durumu 3 altında olan değerler elenmiştir. 3 ve üzeri değerler kabul edilmiştir. Veri setinin son hali Tablo 4.6'daki gibidir.

Tablo 4.6. Veri seçimlerinin sonucu

Veri	Sayısal Değeri
Toplam Oy Sayısı	11.622
Toplam Yazar Sayısı	3.067
Toplam Kullanıcı Sayısı	4.240

4240 kullanıcının 3067 farklı yazar hakkında 10.000'den fazla oylama puanı algoritmada değerlendirilmek üzere seçilmiştir. Veri setinin son halinden bir kesit Şekil 4.7'de verilmiştir. UserID numarası 92 olan kullanıcının beğendiği yazar J. R. R. Tolkien'dir. UserID numarası 254 olan kullanıcının beğendiği yazarlar Philip Bulman, J. K. Rowling, Rad Bradbury'dir.

UserID	BookAuthor	BookRating
92	J. R. R. TOLKIEN	8
232	FRANKLIN W. DIXON	5
242	RAY BRADBURY	9
254	PHILIP PULLMAN	5
254	J. K. ROWLING	7
254	RAY BRADBURY	8
507	KAY HOOPER	5
507	SUSAN CONANT	5
507	JULIA SPENCER-FLEMING	6
507	SUSAN COOPER	9
507	SHARYN MCCRUMB	5
638	JAMES PATTERSON	9
638	DAVID SEDARIS	9
638	NICHOLAS SPARKS	8
643	MARY HIGGINS CLARK	7
643	BORIS VIAN	5
709	DIANE DUANE	10
744	BARBARA MICHAELS	7
753	JONATHAN PEARCE	9
801	ADRIAN PAASE	5

Şekil 4.7. Veri setinin son hali

Verilerin Dönüşümü: Veri setini Apriori algoritmasının istediği forma dönüştürme işlemleri için Python ve Excel kullanılmıştır. Apriori algoritmasını Python dilinde kodlarken veri setini matris formatında kabul etmektedir bu amaçla bazı işlemler uygulanmıştır. Bu işlemler aşağıda verilmiştir:

1. Özniteliklerin benzersiz olabilmesi için tüm yazar adları büyük harfe dönüştürülmüş ve aradaki tüm noktalama işaretleri kaldırılmıştır.
2. Veri seti matris formatına dönüştürülmüştür.
3. Encoding işlemi yapılmıştır. Apriori algoritması veri setini 1 ve 0 matris formatında kabul etmektedir. Apriori algoritması 1 ve 0 formatına sahip olduğu için kullanıcı yazara 5 ve üzerinde değerde oy veriyse matriste 1 olarak yer almıştır ya da yazara ait hiç oy vermemişse 0 olarak veri setine işlenmiştir.

Veri dönüşümü önce ve sonra durumu aşağıda gösterilmiştir. Şekil 4.8. veri setinin ilk halini yansıtmaktadır.

Index	UserID	BookAuthor	BookRating
0	92	JRRTOLKIEN	8
1	232	FRANKLINWDIXON	5
2	242	RAYBRADBURY	9
3	254	RAYBRADBURY	8
4	254	PHILIPPULLMAN	5
5	254	JKROWLING	7
6	507	SHARYNMCCRUMB	5
7	507	KAYHOOPER	5
8	507	SUSANCONANT	5
9	507	SUSANCOOPER	9
10	507	JULIASPENCER-FLEMING	6
11	638	DAVIDSEDARIS	9
12	638	NICHOLASSPARKS	8
13	638	JAMESPATTERSON	9
14	643	BORISVIAN	5
15	643	MARYHIGGINSCLARK	7
16	709	DIANEDUANE	10
17	744	BARBARAMICHAELS	7
18	753	JONATHANPEARCE	9
19	901	ADRIANPLASS	5
20	914	JANOSCH	5
21	1131	JANEAUSTEN	5
22	1211	PHILIPKDICK	8
23	1211	LAURENTVICOMTE	9
24	1211	PHILIPKINDREDDICK	8

Şekil 4.8. Veri setinin dönüştürülmemiş hali

Veri setinin son halinde yazarlar matrisin sütunlarını oluşturmaktadır. Kullanıcılar ise satırları oluşturmaktadır. Oylar ise matris içinde 1 ve 0 olarak görülmektedir

Veri Madenciliği Algoritmasının Çalıştırılması: Çalışmada önceki adımlarda temizlenmiş ve matris formatına dönüştürülmüş veri seti kullanılmıştır. Her yazar sepet analizi çalışmalarında yer alan ürünler gibi sütunlarda yer almıştır ve her kullanıcı sepet analizi çalışmalarındaki alışveriş fişi gibi satırlarda yer almıştır. Veri setinde 11.622 oy bulunmaktadır. Veri setinde kullanıcı sayısı, matrisinin satır sayısını, yazar sayısı ise matrisin sütun sayısını ifade etmektedir. Bu oylar 4240 kullanıcı tarafından 3067 farklı yazarın değerlendirilmesine aittir.

Veri seti üzerinde apriori algoritması çalıştırılmıştır. Apriori algoritması çalıştırılması ve sonuçların listelenmesi Python dilinde gerçekleştirilmiştir. Apriori algoritması çalışma esnasında iki parametre almaktadır. Bu parametreler minimum eşik değeri ve güven değeridir. Değerler apriori algoritmasında geliştirici tarafından seçilmektedir.

Apriori algoritmasının minimum eşik değeri 0.0008 ve minimum güven değeri 0.70 seçilmiştir. Sonuçlar, sık geçen kümeler, çıkarılan kurallar ve kural değerlendirilmeleri olmak üzere 4.4.2. başlığında verilmiştir.

4.4.2. Sonuçlar: Yazarlara Ait Birliktelik Kuralları , Sık Geçen Kümeler

Bu başlıkta önce Apriori algoritmasının sonuçlarında yer alan sık geçen öge kümeleri gösterilecektir. Daha sonra apriori algoritmasının oluşturduğu kurallar gösterilip değerlendirilecektir.

Yazarlara Ait Sonuçlar: Sık Geçen K Ögeli Kümeler: Apriori algoritması öncelikle eleme yaparak sık geçen ürünleri bulmaktadır. Toplam sık geçen öge küme sayısı 1229'dur. Apriori algoritması sonuçlarına göre tek elemanlı sık geçen küme sayısı 599'dur. Algoritma sonuçları Tablo 4.7'de verilmiştir.

Tablo 4.7. Sık geçen öge küme satır sayıları

Veri	Sayısal Değer
M ₁ Sık Geçen Öge Küme Sayısı	599
M ₂ Sık Geçen Öge Küme Sayısı	529
M ₃ Sık Geçen Öge Küme Sayısı	89
M ₄ Sık Geçen Öge Küme Sayısı	12

M₁ ve M₂ sık geçen kümelerinin satır sayısı fazla olduğu için tablo olarak gösterilmemiştir. M₁ ve M₂'ye kıyasla satır sayısı daha az olduğu için M₃ ve M₄ tabloları ile ilgili bilgiler Tablo 4.8 ve Tablo 4.9'da gösterilmiştir. M₃ sık geçen öge kümesi 3 ögeli birlikte bulunan öğeleri temsil etmektedir. M₃ sık geçen öğeleri aşağıda bulunan Tablo 4.8'de verilmiştir. Bu kümeler kural çıkartmak için Apriori algoritması tarafından kullanılmaktadır. İlk sırada yer alan M₃ öge kümesi içerisinde birinci öge olarak tarihi romantizm romanlarıyla bilinen Johanna Alindsey, ikinci öge olarak çağdaş romantik gerilim romanlarıyla bilinen Amanda Quick ve üçüncü öge olarak romantizm romanlarıyla bilinen Jude Deveraux yer almaktadır.

Tablo 4.8. M₃ sık geçen öge kümeleri

No	Yazar 1	Yazar 2	Yazar 3	Destek Değeri
1	Johanna Lindsey	Amanda Quick	Jude Deveraux	0.0009433962
2	Johanna Lindsey	Amanda Quick	Julie Garwood	0.0009433962
3	Amanda Quick	Julie Garwood	Jude Deveraux	0.0011792452
4	J.R.R. Tolkien	J.K. Rowling	Amy Tan	0.0009433962
5	Marion Zimmer Bradley	Anne Mccaffrey	Mercedes Lackey	0.0009433962
6	Stephen King	James Patterson	Anne Rice	0.0009433962
7	Julia Quinn	Christine Feehan	Catherine Anderson	0.0009433962
8	Julia Quinn	Julie Garwood	Catherine Anderson	0.0009433962
9	Julia Quinn	Nora Roberts	Catherine Anderson	0.0009433962
10	Nora Roberts	Julie Garwood	Catherine Anderson	0.0009433962
11	Lavyrle Spencer	Nora Roberts	Catherine Anderson	0.0009433962
12	Lavyrle Spencer	Catherine Anderson	Susan Elizabeth Phillips	0.0009433962
13	Nora Roberts	Catherine Anderson	Susan Elizabeth Phillips	0.0009433962
14	Julie Garwood	Jd Robb	Catherine Coulter	0.0009433962
15	Jd Robb	Nora Roberts	Catherine Coulter	0.0009433962
16	Johanna Lindsey	Catherine Coulter	Jude Deveraux	0.0009433962
17	Johanna Lindsey	Catherine Coulter	Julie Garwood	0.0009433962
18	Johanna Lindsey	Nora Roberts	Catherine Coulter	0.0009433962
19	Julie Garwood	Julia Quinn	Catherine Coulter	0.0009433962
20	Catherine Coulter	Julie Garwood	Nicholas Sparks	0.0009433962
21	Julie Garwood	Nora Roberts	Catherine Coulter	0.0011792452
22	Catherine Coulter	Julie Garwood	Susan Elizabeth Phillips	0.0009433962
23	Nora Roberts	Catherine Coulter	Nicholas Sparks	0.0009433962
24	Julia Quinn	Christine Feehan	Julie Garwood	0.0009433962
25	J.R.R. Tolkien	J.K. Rowling	Cs Lewis	0.0011792452
26	Madeleine Lingle	J.K. Rowling	Cs Lewis	0.0009433962
27	J.R.R. Tolkien	Madeleine Lingle	Cs Lewis	0.0009433962
28	Danielle Steel	Nora Roberts	Sandra Brown	0.0011792452
29	V.C. Andrews	Danielle Steel	Nora Roberts	0.0009433962

30	Dean R. Koontz	Stephen King	Dean R. Koontz	0.0014150943
31	Michael Crichton	John Grisham	Dean R. Koontz	0.0009433962
32	John Grisham	Stephen King	Dean R. Koontz	0.0009433962
33	Nicholas Sparks	Stephen King	Dean R. Koontz	0.0009433962
34	Nora Roberts	Sandra Brown	Fern Michaels	0.0011792452
35	John Grisham	James Patterson	Janet Evanovich	0.0009433962
36	James Patterson	John Sandford	Janet Evanovich	0.0009433962
37	Nora Roberts	James Patterson	Janet Evanovich	0.0009433962
38	Stephen King	James Patterson	Janet Evanovich	0.0009433962
39	John Grisham	James Patterson	John Sandford	0.0011792452
40	John Grisham	Patricia Daniel Scornwell	James Patterson	0.0011792452
41	John Grisham	Stephen King	James Patterson	0.0009433962
42	Maryhigginsclark	James Patterson	John Sandford	0.0009433962
43	James Patterson	Patricia Daniel Scornwell	John Sandford	0.0014150943
44	James Patterson	Stephen King	John Sandford	0.0009433962
45	James Patterson	Stuartwoods	John Sandford	0.0009433962
46	James Patterson	Tamihoag	John Sandford	0.0009433962
47	Nora Roberts	Julie Garwood	James Patterson	0.0009433962
48	Janetdailey	Julie Garwood	Janet Evanovich	0.0009433962
49	Janet Evanovich	Julie Garwood	Jennifer Crusie	0.0009433962
50	Nora Roberts	Janet Evanovich	Jennifer Crusie	0.0009433962
51	J.K. Rowling	Janet Evanovich	Lm Montgomery	0.0009433962
52	Suegraston	J.K. Rowling	Janet Evanovich	0.0009433962
53	Nora Roberts	Julie Garwood	Janet Evanovich	0.0011792452
54	Jd Robb	Julie Garwood	Jude Deveraux	0.0009433962
55	Jd Robb	Nora Roberts	Jude Deveraux	0.0009433962
56	Judith Mcnaught	Jd Robb	Nora Roberts	0.0009433962
57	Jd Robb	Nora Roberts	Julie Garwood	0.0011792452
58	Nora Roberts	Julie Garwood	Jennifer Crusie	0.0009433962
59	John Grisham	J.R.R. Tolkien	J.K. Rowling	0.0009433962
60	Johnsteinbeck	J.R.R. Tolkien	J.K. Rowling	0.0009433962
61	Johanna Lindsey	Julie Garwood	Jude Deveraux	0.0011792452
62	Johanna Lindsey	Nora Roberts	Julie Garwood	0.0009433962
63	John Grisham	Patricia Daniel Scornwell	John Sandford	0.0011792452
64	Michael Crichton	John Grisham	Stephen King	0.0014150943
65	Tom Clancy	John Grisham	Stephen King	0.0009433962
66	Maryhigginsclark	Patricia Daniel Scornwell	John Sandford	0.0009433962
67	Perri Oshaughnessy	Michaelconnelly	John Sandford	0.0009433962

68	Perri Oshaughnessy	Robertb Parker	John Sandford	0.0009433962
69	Jonathan Kellerman	Rosamunde Pilcher	Patricia Daniel Scornwell	0.0009433962
70	Nora Roberts	Julie Garwood	Jude Deveraux	0.0011792452
71	Lavyrle Spencer	Nora Roberts	Jude Deveraux	0.0009433962
72	Nora Roberts	Sandra Brown	Jude Deveraux	0.0009433962
73	Nora Roberts	Stephen King	Jude Deveraux	0.0009433962
74	Judith Mcnaught	Julia Quinn	Julie Garwood	0.0009433962
75	Judith Mcnaught	Julia Quinn	Susan Elizabeth Phillips	0.0009433962
76	Judith Mcnaught	Lavyrle Spencer	Julie Garwood	0.0009433962
77	Judith Mcnaught	Nora Roberts	Julie Garwood	0.0014150943
78	Judith Mcnaught	Julie Garwood	Susan Elizabeth Phillips	0.0011792452
79	Judith Mcnaught	Lavyrle Spencer	Nora Roberts	0.0009433962
80	Judith Mcnaught	Nora Roberts	Susan Elizabeth Phillips	0.0009433962
81	Julia Quinn	Nora Roberts	Julie Garwood	0.0009433962
82	Julia Quinn	Julie Garwood	Susan Elizabeth Phillips	0.0011792452
83	Lavyrle Spencer	Nora Roberts	Julie Garwood	0.0011792452
84	Nora Roberts	Julie Garwood	Nicholas Sparks	0.0011792452
85	Sharonsala	Nora Roberts	Julie Garwood	0.0009433962
86	Nora Roberts	Julie Garwood	Susan Elizabeth Phillips	0.0009433962
87	Nora Roberts	Julie Garwood	Susan Wiggs	0.0009433962
88	Lavyrle Spencer	Nora Roberts	Nicholas Sparks	0.0009433962
89	Lavyrle Spencer	Nora Roberts	Susan Elizabeth Phillips	0.0009433962

Tablo 4.9’da ise M₄ sık geçen öge kümelerine yer verilmiştir. Bu tabloda yer alan ilk M₄ sık geçen öge kümesi içerisinde birinci öge olarak tarihi romantizm romanlarıyla bilinen Johanna Alindsey, ikinci öge olarak çağdaş romantik gerilim romanlarıyla bilinen Amanda Quick ve üçüncü öge Amerikalı aşk romanları yazar Julie Garwood ve dördüncü olarak romantizm romanlarıyla bilinen Jude Deveraux yer almaktadır.

Tablo 4.9. M₄ sık geçen öge kümeleri

No	Destek Değeri	Yazar 1	Yazar 2	Yazar 3	Yazar 4
1	9,4x10 ⁻⁵	Johanna Lindsey	Amanda Quick	Julie Garwood	Jude Deveraux
2	9,4x10 ⁻⁵	Julia Quinn	Nora Roberts	Julie Garwood	Catherine Anderson
3	9,4x10 ⁻⁵	Lavyrle Spencer	Nora Roberts	Catherine Anderson	Susan Elizabeth Phillips
4	9,4x10 ⁻⁵	Julie Garwood	Jd Robb	Nora Roberts	Catherine Coulter
5	9,4x10 ⁻⁵	Julie Garwood	Johanna Lindsey	Nora Roberts	Catherine Coulter
6	9,4x10 ⁻⁵	Julie Garwood	Nora Roberts	Catherine Coulter	Nicholas Sparks
7	9,4x10 ⁻⁵	James Patterson	John Grisham	Patricia Daniel Scornwell	John Sandford
8	9,4x10 ⁻⁵	Janet Evanovich	Nora Roberts	Julie Garwood	Jennifer Crusie
9	9,4x10 ⁻⁵	Jd Robb	Nora Roberts	Julie Garwood	Jude Deveraux
10	9,4x10 ⁻⁵	Judith Mcnaught	Julia Quinn	Julie Garwood	Susan Elizabeth Phillips
11	9,4x10 ⁻⁵	Judith Mcnaught	Lavyrle Spencer	Nora Roberts	Julie Garwood
12	9,4x10 ⁻⁵	Judith Mcnaught	Nora Roberts	Julie Garwood	Susan Elizabeth Phillips

Yazarlara Ait Sonuçlar, Yazarlar Arası Birliktelik Kuralları: Apriori algoritmasının oluşturduğu kurallara ait sayısal bilgiler Tablo 4.10’da verilmiştir. Minimum güven filtrelemesine göre kural sayılarını göstermektedir. Kurallar tablosu %50 ve üzeri güven değerine sahip 365 kural oluşturulmuştur. Minimum güven değeri %70 olarak alındığı için %70’in üzerinde kalan kurallar kabul edilen kurallar tablosuna aktarılmıştır.

Tablo 4.10. Yazarlar arası birliktelik kuralları sayısal bilgiler

Veri	Sayısal Değeri
%100 Güvenilir Kural Sayısı	96
%90 ve Üzeri Güvenilir Kural Sayısı	96
%80 ve Üzeri Güvenilir Kural Sayısı	174
%70 ve Üzeri Güvenilir Kural Sayısı	187

Apriori algoritmasının sonuçlarına göre en az %70 güvenilirliğe sahip toplam kural sayısı 187’dir. Bu kuralların 96 tanesi %100 güvenilirliğe sahiptir. Oluşturulmuş kurallar kullanıcıların ilgili yazarın kitaplarına verdiği oylara dayanmaktadır. Bu kurallar sayesinde kullanıcıların beğendiği yazarlar arasındaki bağlantılar bulunmuştur. Toplam 187 kural olduğu için hepsi tabloya yansıtılmamıştır. En güvenilir 10 kural aşağıda Tablo 4.11’de bu kurallara ait değerler verilmiştir.

Tablo 4.11. %100 güvenilirlik değerine sahip yazar ilişkileri (Birliktelik kuralları)

No	Öncül	Sonuç	Destek	Güven
1	Jamie Suzanne	Francinepascal	0.0009433962	1
2	Joseph Conrad	Thomas Hardy	0.0009433962	1
3	Maryjoputney	Julie Garwood	0.0009433962	1
4	Amanda Quick, Johanna Lindsey	Jude Deveraux	0.0009433962	1
5	Amanda Quick, Johanna Lindsey	Julie Garwood	0.0009433962	1
6	Amanda Quick, Jude Deveraux	Julie Garwood	0.0011792452	1
7	Anne Mccaffrey, Mercedes Lackey	Marion Zimmer Bradley	0.0009433962	1
8	Catherine Anderson, Christine Feehan	Julia Quinn	0.0009433962	1
9	Catherine Anderson, Julie Garwood	Julia Quinn	0.0009433962	1
10	Julia Quinn, Nora Roberts	Catherine Anderson	0.0009433962	1

Tablo 4.11’de yer alan ilk kural Jamie Suzanne yazarının kitaplarını beğenen bir kullanıcı Francine Pascal yazarının kitaplarını beğendiğini göstermektedir. Bu kuralın güvenilirlik değeri %100’dür. Jamie Suzanne tercih eden insanlar Francine Pascal yazarını da tercih etmektedir.

3.sırada yer alan kural Mary Jo Putney yazarını beğenen kullanıcı Julie Garwood yazarını da beğendiğini göstermektedir. Bu kuralın güvenilirlik değeri %100’dür.Tablodaki yazarların hangi tür romanlar yazdığını bilmek kuralın daha iyi anlaşılmasını sağlayacaktır çünkü bu veri seti makarna, yoğurt gibi her kullanıcının bildiği günlük ürünlerden oluşmamaktadır. Bağlantının daha iyi anlaşılmasını sağlamak amacıyla yazarlar hakkında birkaç satır bilgi verilmiştir:

Mary Jo Putney yazarı romantik ve fantastik romanlar yazmaktadır. Julie Garwood ise aşk ve romantik konularını içeren romanlar yazmaktadır.

7.sırada yer alan kural ise öncül tarafında iki yazar sonuç tarafında ise bir yazar göstermektedir. Bu kural tipi şu şekilde ifade edebilmektedir:

Hem Anne McCaffrey hem Mercedes Lackey'i beğenen kullanıcılar Marion Zimmer Bradley'i de beğenmektedirler. Bu kuraldaki en önemli nokta kuralın öncülünde iki öge olmasıdır. İki yazarı da beğenenler Mercedes Lackey'i de tercih etmektedirler şeklinde ifade edilebilmektedir.

Diğer güvenilirlik değerlerine ait kuralları gösterebilmek için aşağıdaki tablolar hazırlanmıştır. Tablo 4.12'te %80 ve üzeri güven değerine sahip kurallar aktarılmıştır.

Tablo 4.12. %80 ve üzeri güvenilirlik değerine sahip yazar ilişkileri

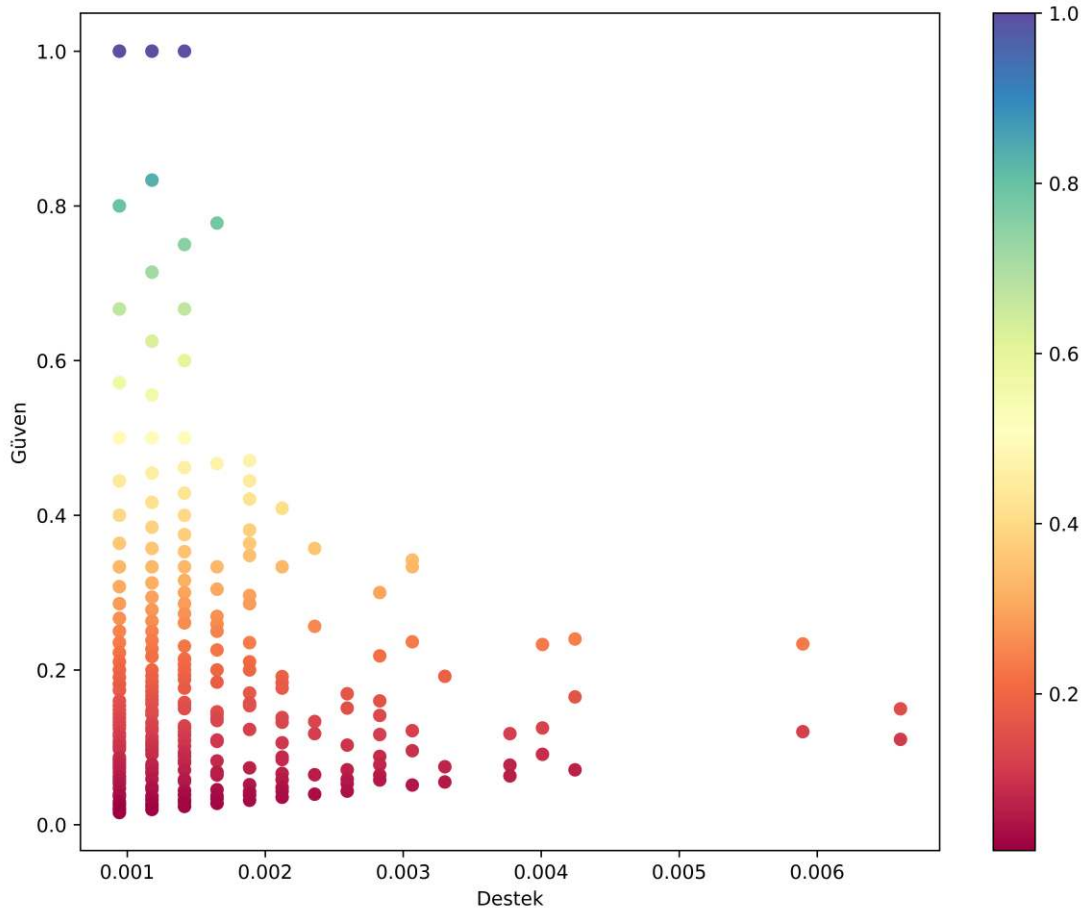
No	Öncül	Sonuç	Destek	Güven
1	Amanda Quick, Julie Garwood	Jude Deveraux	0,0011792	0,83
2	Julie Garwood, Catherine Coulter	Nora Roberts	0,0011792	0,83
3	Nora Roberts, Catherine Coulter	Julie Garwood	0,0011792	0,83
4	Fern Michaels, Sandra Brown	Nora Roberts	0,0011792	0,83
5	Nora Roberts, Fern Michaels	Sandra Brown	0,0011792	0,83
6	John Sandford, John Grisham	James Patterson	0,0011792	0,83
7	Janet Evanovich, Julie Garwood	Nora Roberts	0,0011792	0,83
8	Julie Garwood, Johanna Lindsey	Jude Deveraux	0,0011792	0,83
9	John Sandford, John Grisham	Patricia Daniel Scornwell	0,0011792	0,83
10	Julie Garwood, Susan Elizabeth Phillips	Judith Mcnaught	0,0011792	0,83

Tablo 4.13'te %70 ve üzeri güven değerine sahip kurallar aktarılmıştır. Bu tablolar her çeşit güven değerine ait kuralları gösterebilmek adına verilmiştir. Bir örnek olarak bu tablodan da 5 numaralı kuralın ifadesi verilmiştir. Stephen King ve Brian Lumley yazarlarının bağıntısını göstermektedir. Brian Lumley yazarının beğenen kullanıcılar Stephen King'i de beğenmektedir. Bu kuralın güvenilirlik oran %71'dir.

Tablo 4.13. %70 ve üzeri güvenilirlik değerine sahip yazar ilişkileri

No	Öncül	Sonuç	Destek	Güven
1	Julia Quinn	Julie Garwood	0.00165094	0,77
2	John Sandford, Patricia Daniel Scornwell	James Patterson	0.00141509	0,74
3	Julie Garwood, Judith Mcnaught	Nora Roberts	0.00141509	0,74
4	Nora Roberts, Judith Mcnaught	Julie Garwood	0.00141509	0,74
5	Brianlumley	Stephen King	0.00117924	0,71
6	Catherine Anderson	Julia Quinn	0.00117924	0,71
7	Catherine Anderson	Nora Roberts	0.00117924	0,71

Oluşturulmuş kurallar güven değerine göre yorumlanmıştır. Bunun sebebi algoritma çalıştırılırken metrik olarak güven değişkeninin seçilmesinden dolayıdır. Aşağıdaki Şekil 4.9'da tüm kurallara ait destek ve güven dağılımı verilmiştir.



Şekil 4.9. Yazar kurallarına ait destek ve güven değerlerinin dağılımı

Algoritma sonuçlarında birçok ölçme metriğinin sonucu da alınmıştır. En iyi 10 kurala ait metriklerde aşağıdaki Tablo 4.14’te verilmiştir:

Tablo 4.14’teki kurallar yukarıdaki bölümde destek ve güven açısından incelenmiştir. Bir diğer kural değerlendirme metriği olan Lift ve Leverage değerleride aşağıdaki tabloda verilmiştir.

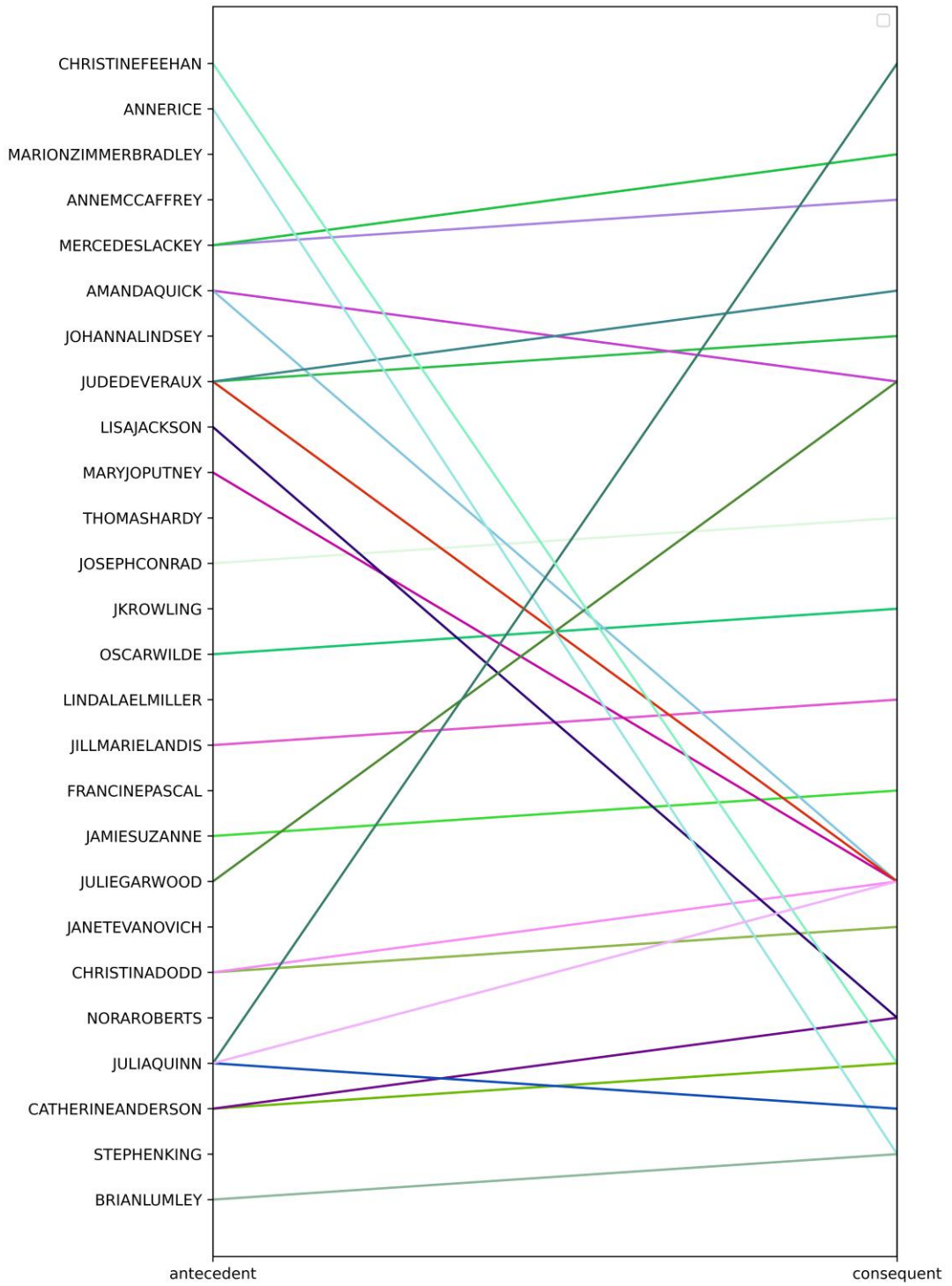
Leverage değeri kuralları ölçmek için kullanılan metriktir. Güven değeri arttıkça leverage değerinin arttığı görülmüştür. Yüksek güven değerine sahip kurallar aynı zamanda yüksek leverage değerini de sahip olmuşlardır.

Lift değeri bir diğer kural ölçüm metriğidir. Bu değer 1’e eşit olduğunda yazarlar arası bağıntı yoktur. Yani A yazarı ile B yazarı ilişkili değildir. 1 ve yukarı değerlerde ise pozitif korelasyon göstermektedir. 1.kurala göre Jamie Suzanne ve Francine Pascal arasında lift değeri 32’dir. Pozitif korelasyona sahiptir. Çıkan en iyi 10 kural arasında lift değeri 1’e eşit olan yoktur. Tüm kuralların birbirleri ile ilişkisi olduğunu kanıtlamaktadır.

Tablo 4.14. En iyi 10 kurala ait değerlendirme metrikleri

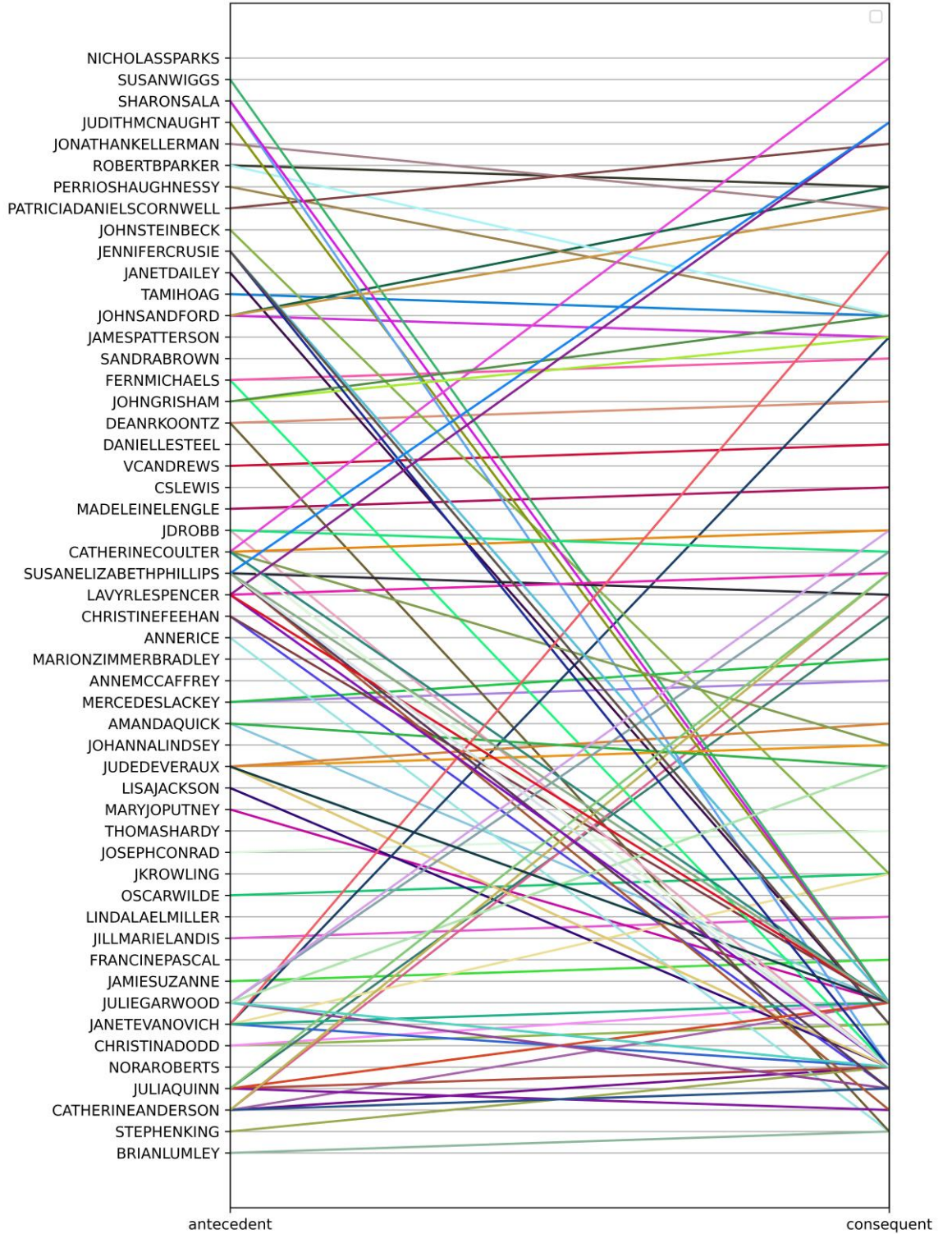
No	Öncül	Sonuç	Destek	Güven	Lift
1	Jamie Suzanne	Francine Pascal	0.000943	1	326
2	Joseph Conrad	Thomas Hardy	0.000943	1	706
3	Mary Jo Putney	Julie Garwood	0.000943	1	151
4	Amanda Quick, Johanna Lindsey	Jude Deveraux	0.000943	1	157
5	Amanda Quick, Johanna Lindsey	Julie Garwood	0.000943	1	151
6	Amanda Quick, Jude Deveraux	Julie Garwood	0.001179	1	151
7	Mercedes Lackey, Anne McCaffrey	Marion Zimmer Bradley	0.000943	1	146
8	Christine Feehan, Catherine Anderson	Julia Quinn	0.000943	1	471
9	Julie Garwood, Catherine Anderson	Julia Quinn	0.000943	1	471
10	Nora Roberts, Julia Quinn	Catherine Anderson	0.000943	1	605
11	Julie Garwood, Catherine Anderson	Nora Roberts	0.000943	1	27
12	Lavyrle Spencer, Catherine Anderson	Nora Roberts	0.000943	1	27
13	Lavyrle Spencer, Catherine Anderson	Susan Elizabeth Phillips	0.000943	1	353

Şekil 4.13'te ve Şekil 4.14'te yazarlara ait birliktelikler görselleştirilmiştir. Şekil 4.14'te tamamı görselleştirilmiştir. Şekil 4.10'da daha net görülmesi için ilk 25 tanesi verilmiştir. %70 ve üzeri güven değerine sahiptirler. Şekil 4.10 incelendiğinde baştan 5.sırada yer alan Mercedes Lackey yazarının Anne McCaffrey yazarı ile bağıntılı olduğu görülmektedir.



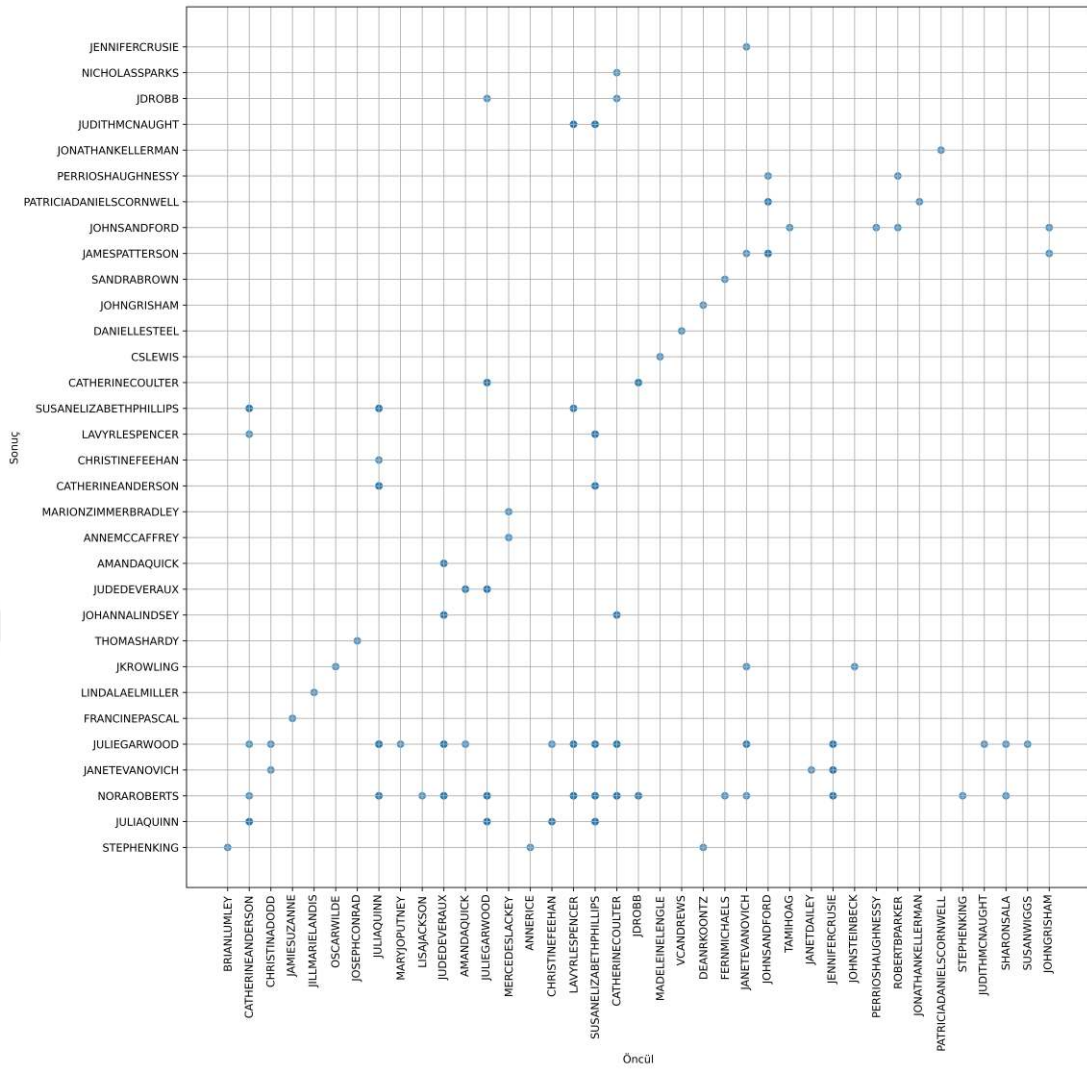
Şekil 4.10. 25 yazar arasındaki ilişkilerin paralel olarak gösterilmesi

Şekil 4.11’de yazarlara ait bağlantılar tespit edilip görselleştirilmiştir.



Şekil 4.11. Yazarlar arası ilişkilerin paralel olarak gösterilmesi

Şekilde çizgiler şu şekilde ifade edilmektedir: Korku romanları yazar Brian Lumley ile gerilim, korku romanları yazar Stephen King’in arasında bağlantı olduğunu göstermektedir. Şekil 4.12’de ise koordinatlar kullanılarak yazarlar arası bağlantılar gösterilmiştir.



Şekil 4.12. Yazarlar arası ilişki gösterimi

Şekil 4.12’de yazarlar arası ilişkiler koordinatlar kullanılarak görselleştirilmiştir. Görsel şu şekilde açıklanabilmektedir: Y ekseninde ilk sırada yer alan Stephen King’in gerilim romanları yazar Brian Lumley, korku ve kurgu yazarı olan Anne Rice ve korku romanı yazarı Dean R. Koontz ile ilişkili olduğu gösterilmiştir.

4.5. Kullanıcıların Beğendikleri Kitaplara Göre Birliktelik Kuralları Çıkarımı Çalışması

Çalışmanın amacı, kullanıcıların oy verdiği kitapları ve verdiği oylama puanları kullanılarak kitaplar arasındaki bağıntıları tespit etmektir. Bu çalışmanın sonucunda, A kitabını beğenen bir kullanıcının başka hangi kitapları beğenebileceği ile ilgili bilgiler kurallar halinde elde edilecektir. Bu kurallar sonucunda ortaya çıkan kitaplar arası ilişkiler, kitap öneri sistemlerinin oluşmasını sağlayacaktır. Çalışmanın

adımları; verilerin oluşturulması, verilerin temizlenmesi, verilerin seçilmesi ve veri madenciliği ile beraber Apriori algoritmasının kullanılması şeklindedir

Verilerin Oluşturulması: Veri seti 1.149.779 oydan, 271.065 kitaptan ve 278.858 kullanıcıdan oluşmaktadır. Çalışmada veri setinde yer alan BXBookRatings tablosu kullanılmıştır. Kullanıcıların kitaplara vermiş olduğu oylardan oluşmaktadır.

Verilerin Temizlenmesi: Veri setinde aynı kitaba ait ISBN numaralarını içindeki “x” harfi bazı oylamalarda küçük x bazı oylamalarda ise büyük X olarak yer almıştır. Aynı öznitelik olarak algılandığı için tüm harfler büyük harfe dönüştürülmüştür. İkinci adım olarak şu işlem uygulanmıştır:

BXBookRatings tablosunda oy verildiği halde BXBooksDescription tablosunda yer almayan kitaplara ait oylamalar veri setinden elenmiştir. Bu eleme sonucunda 118.604 oy veri setinden çıkartılmıştır. Eleme işlemi SQL Server Management Studio’da T-SQL kodlama dilinde gerçekleştirilmiştir. Verilere ilişkin sayısal bilgiler aşağıdaki Tablo 4.15’te verilmiştir.

Tablo 4.15. Kitap derecelendirmeleri oylamalara ait sayısal bilgiler

Veri	Sayısal Değeri
Toplam Oy Sayısı	1.149.779
Toplam Oy Kullanıcı Sayısı	105.282
Toplam Oylanmış Kitap Sayısı	269.862
Eleme Sonrası Oy Sayısı	1.031.175
Eleme Sonrası Kullanıcı Sayısı	92.107
Eleme Sonrası Kitap Sayısı	269.862

Verilerin Seçilmesi: Çalışmada kullanılacak verilerin seçilmesi için şu adımlar uygulanmıştır: 5 ve üzeri değere sahip olan oyların seçilmesi, kitapların oylama yoğunluklarının hesaplanması ve gruplanması, kullanıcı / kitap oranına göre yoğunluk grubu seçimi, puan yoğunluklarına dayalı feature selection (öznitelik seçimi) ile oy seçimidir.

5 ve Üzeri Değere Sahip Olan Oyların Seçilmesi: Çalışmada kullanıcıların beğendiği kitaplar üzerinden kural oluşturulacağı için veri setinden kitap oylaması 5 ve üzeri değere sahip olan oylamalar seçilmiştir. Eleme yapılmadan önce temizlenmiş veri setinde 1.031.175 oy bulunmaktadır. 5 ve üzeri değere sahip olan kitaplar seçildikten sonra toplam oy sayısı 367.261 olmuştur. Tablo 4.16’da yer alan Oy / Kitap

Oranı sütunu bir kitabın ortalama olarak kaç satırda yer aldığını göstermektedir. Bu oranın hesaplanma amacı ileriki adımlarda yer alan veri seçiminde bilgi sunma ihtiyacından kaynaklıdır. Tablo 4.16’da yer alan 5-10 puan arası değere sahip oylamalar veri setinden seçilmiştir.

Tablo 4.16. Veri seti puan sınırına göre eleme bilgileri

Durum	Oy Sayısı	Kitap Sayısı	Kullanıcı Sayısı	Oy / Kitap Oranı
0-10 Puan	1.031.175	269.862	92.107	3,82
5-10 Puan	367.261	145.002	65.859	2,53

Kitapların Oylama Yoğunluklarının Hesaplanması ve Gruplanması: Veri setinde yer alan herhangi bir kitabın veri setinde bulunma sayısı sadece 1 ise bu kitap kural oluşumuna uygun değildir. Bu amaçla veri setinde yer alan her kitabın oylamalarda bulunma durumu hesaplanmıştır.

Kullanıcı / Kitap Oranına Göre Yoğunluk Grubu Seçimi: Her yoğunluk grubuna ait Kullanıcı / Kitap oranları hesaplanmıştır. Kullanıcı / Kitap oranı göz önüne alınarak seçim yapılmıştır. Çalışmada algoritmaya sokulacak olan matriste sütun sayısı kitapları, satır sayısı benzersiz kullanıcıları göstermektedir. Tablo 4.17’de kitap sayısının fazla kullanıcı sayısının az olduğu matris gösterilmiştir. Kullanıcı sayısına göre fazla kitap bulunduğu için seyrek bir görüntü vermektedir.

Tablo 4.17. Seyrek görünümlü matris

	K1	K2	K3	K4	K5	K6	K7	K8	K9
User1	1	0	0	0	1	0	0	0	0
User2	0	1	0	0	0	0	0	1	0

Tablo 4.18’de yoğun görünümlü matris gösterilmiştir. Kullanıcı / Kitap oranının yüksek olması çok satırlı az sütunlu matris görünümü vermektedir. Kitap sayısı az ve kullanıcı sayısı fazla olduğu için oylamaların çakışma ihtimali vardır. Aynı kitaba oy verme oranı artacağı için buradan birliktelik kuralı çıkarımı mümkün olabilmektedir.

Tablo 4.18. Yoğun görünümlü matris

	K1	K2	K3	K4
User1	1	0	0	0
User2	0	1	1	0

User3	1	1	0	0
User4	0	0	0	1
User5	1	0	1	0
User6	0	0	1	1
User7	1	1	0	0

Veri setimizdeki yoğunluklarına göre böldüğümüz grupların Kullanıcı / Kitap ortalaması 96'dır. Kullanıcı / Kitap oranı ortalama ve ortalama üstünde olanlar 48 ile 192 arası değerlere yakın olan 6 ve 16 arası gruplar veri setinden seçilmiştir

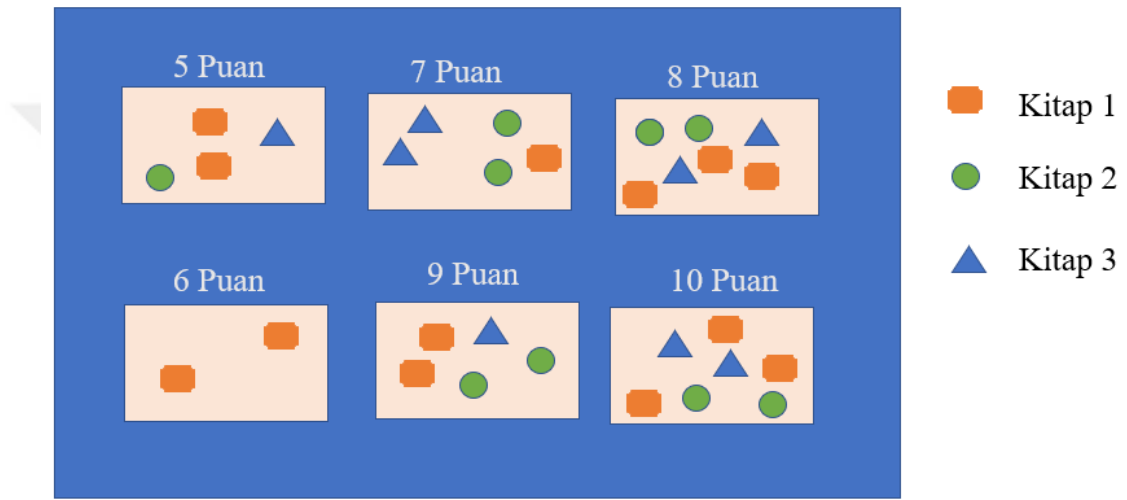
Puan Yoğunluklarına Dayalı Feature Selection ile Oy Seçimi: Aşağıdaki veri seçimi için yapılan yöntemler Şekil 4.13'te verilmiştir. 1. ve 2. Adımlar yukarıdaki başlıklarda anlatılmıştır. Bu bölümde ise "Puanlardaki Yoğunluklara Göre Veri Seçimi" çalışması anlatılacaktır.



Şekil 4.13. Kitapların gruplara göre sınıflandırılması

Bu bölüme kadar seçilmiş olan verilerle kural çıkarımı için Apriori algoritması çalıştırılmış fakat bilgisayarın kaldırabileceği en düşük minimum destek değerine rağmen sadece 10 kural oluşturulmuştur. Daha fazla gruplara göre veri seçimini daraltmak kitap kaybına neden olacağı için (her grupta farklı kitaplar yer alır ve grup elenmesi kitap silinmesine yol açar) her kitabın puan türleri içindeki yoğunluk dağılımına bakılmıştır. Kitabın genel veri setindeki yoğunluk durumunu gösteren grup sınıfı ile 5,6,7,8,9,10 puan oylamaları arasındaki ağırlıkları karşılaştırılarak yoğunluk durumuna en iyi etki eden puan sütunu bulunmuştur ve veri setinde o puan sütunlarındaki oylamalar kullanılmıştır.

Şekil 4.14'te turuncu şekil 15.grupta (çok yoğun) yer alan bir kitabı, yeşil 9.grupta (orta yoğun) yer alan bir kitabı ve mavi 5.grupta (yoğun) yer alan bir kitabı temsil ettiğinde 5,6,7,8,9 ve 10 puan üzerindeki ağırlıkları genel veri setindeki yoğunluk grupları ile karşılaştırılıp en iyi yoğunluğa etki eden puan grubu bulunacak ve veri setinden diğerleri elenecektir. Sonuç yoğunluğa etki eden 8 ve 10 puan türünün etkili olduğunu gösterdiğinde 8 ve 10 puan türüne ait oylamalar üzerinden kural oluşturulacaktır.



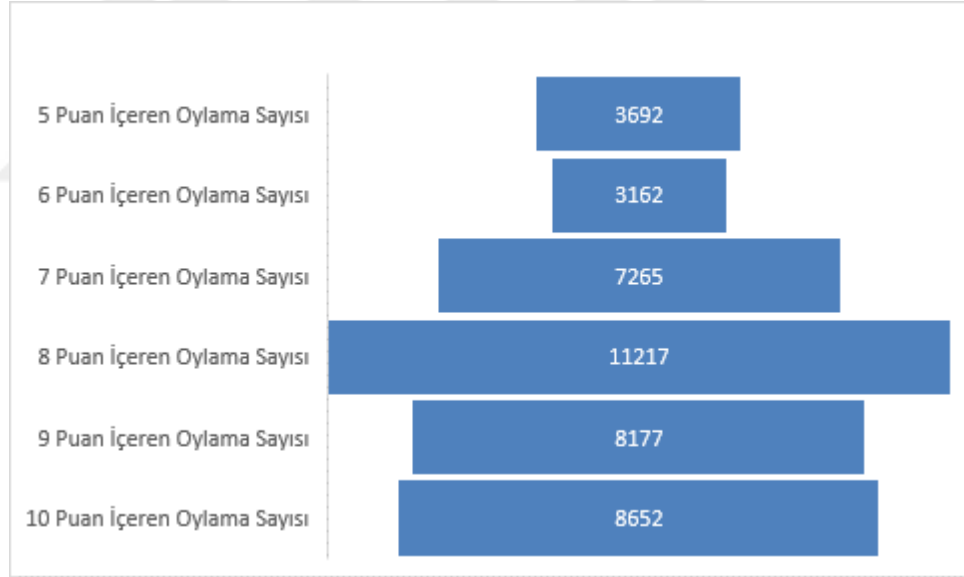
Şekil 4.14. Puanlara göre oylamaların dağılımı

Bu işlem “Feature Selection” ile gerçekleştirilecektir. Öznitelik seçimi için hazırlanmış veri seti Şekil 4.15'te verilmiştir. Tercih sütunu grup sınıfını göstermektedir. ISBN kitabın benzersiz numarasını göstermektedir. BookRating5, BookRating6, BookRating7, BookRating8, BookRating9 ve BookRating10 bu kitaba ait oylamaların puanlara göre ağırlıklarını göstermektedir. 1.sırada yer alan 821854165 ISBN numarasına sahip kitap 5 puanı 2 kez, 6 puanı 1 kez ve 7 puanı 1 kez almıştır. Bu kitabın genel yoğunluk durumu ise Tercih sütunundaki 2 ile gösterilmiştir. Bu veri seti üzerinde öznitelik seçimi algoritmaları kullanılarak tercih durumunu en iyi tahmin eden puan sütununu elde edecektir. Bu puan sütununa ait oylamalar Apriori algoritmasında çalıştırılacaktır.

A	B	C	D	E	F	G	H
ISBN	BookRating5	BookRating6	BookRating7	BookRating8	BookRating9	BookRating10	Tercih
821754165	2	1	1	0	0	0	2
821754173	0	0	0	1	0	0	1
821754181	0	0	0	1	0	0	1
821754246	0	0	0	0	1	0	1
821754394	1	0	0	0	0	0	1
821754408	1	0	1	0	0	0	1
821754416	0	0	2	0	0	0	1
821754661	1	0	0	0	0	0	1
082175467X	0	0	1	0	0	0	1
821754688	0	0	1	0	0	0	1
821754866	0	0	0	0	1	0	1
821755145	0	0	0	0	1	0	1
821755404	1	0	0	0	0	0	1
821755412	1	0	0	0	0	0	1
821755471	0	0	1	0	1	0	1
821755498	1	0	0	1	0	0	1
821755528	0	0	0	1	0	0	1
821755633	1	0	0	0	0	0	1
821755749	0	0	1	0	0	0	1
821755781	1	0	0	0	0	0	1
821755919	0	0	0	1	0	0	1
821755935	1	0	0	0	0	0	1
821755943	0	0	1	1	1	1	2
821755951	0	0	0	2	1	0	1
821756044	0	0	0	0	0	1	1
821756052	0	0	1	0	0	0	1
821756079	1	0	0	0	0	0	1
821756192	1	0	0	0	0	0	1

Şekil 4.15. Veri seti puan dağılımı ve grup sınıfı

Veri setine ait oylama ve puan türü dağılımları Şekil 4.16’da verilmiştir.



Şekil 4.16. Veri seti puan dağılımı ve grup sınıfı grafiği

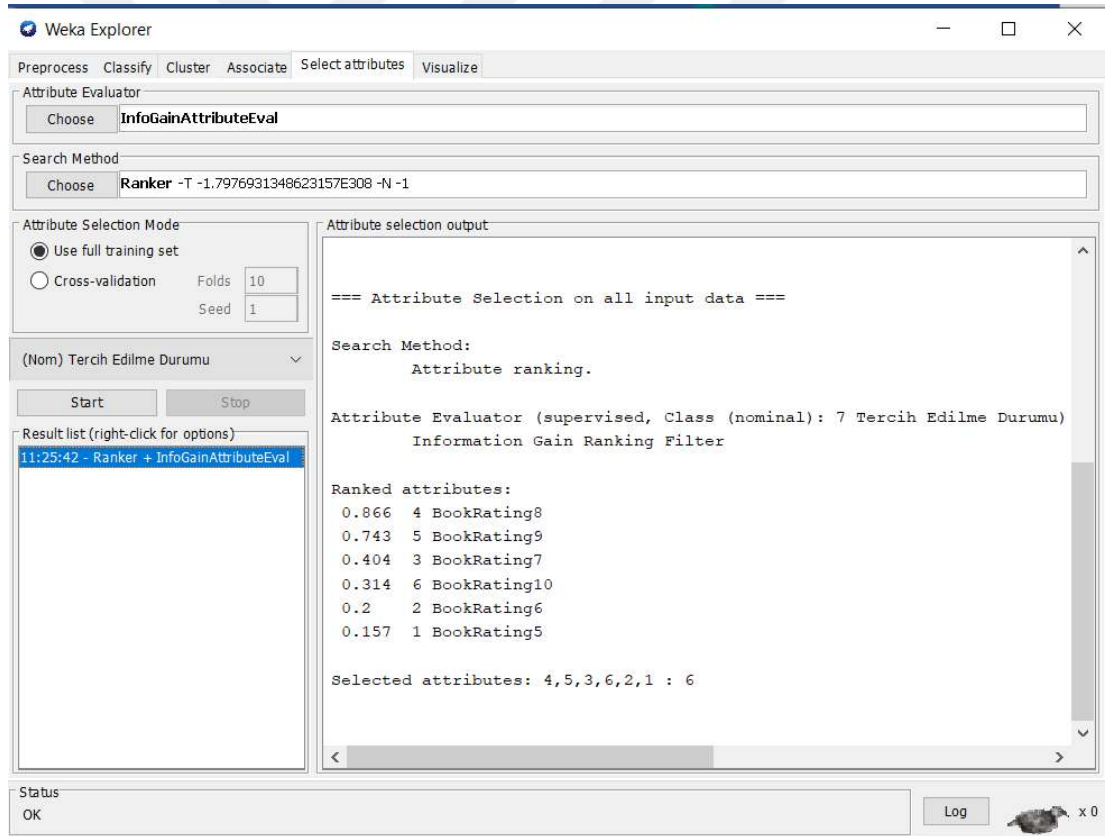
Feature selection işleminin sonuçları, Apriori algoritmasındaki veri seçiminde kullanılacaktır. Algoritmaların seçtiği sonuçlar yani puan türleri bu bölümde anlatılacaktır, kural oluşturmada etkisi ise sonuçlar bölümünde yer alacaktır. Feature selection işlemi için 6 algoritma kullanılmıştır. Algoritmalar aşağıda verilmiştir:

1. Information Gain
2. Classifier Attribute Eval

3. Gain Ratio
4. One-R
5. Relief
6. Symmetrical Uncert Attribute Eval

Information Gain: Information Gain algoritmasının sonuçlarına göre en iyi yoğunluğu belirleyen puan türleri aşağıdaki gibi sıralanmıştır. Şekil 4.17’de Weka programında Information Gain sonuçları gösterilmiştir.

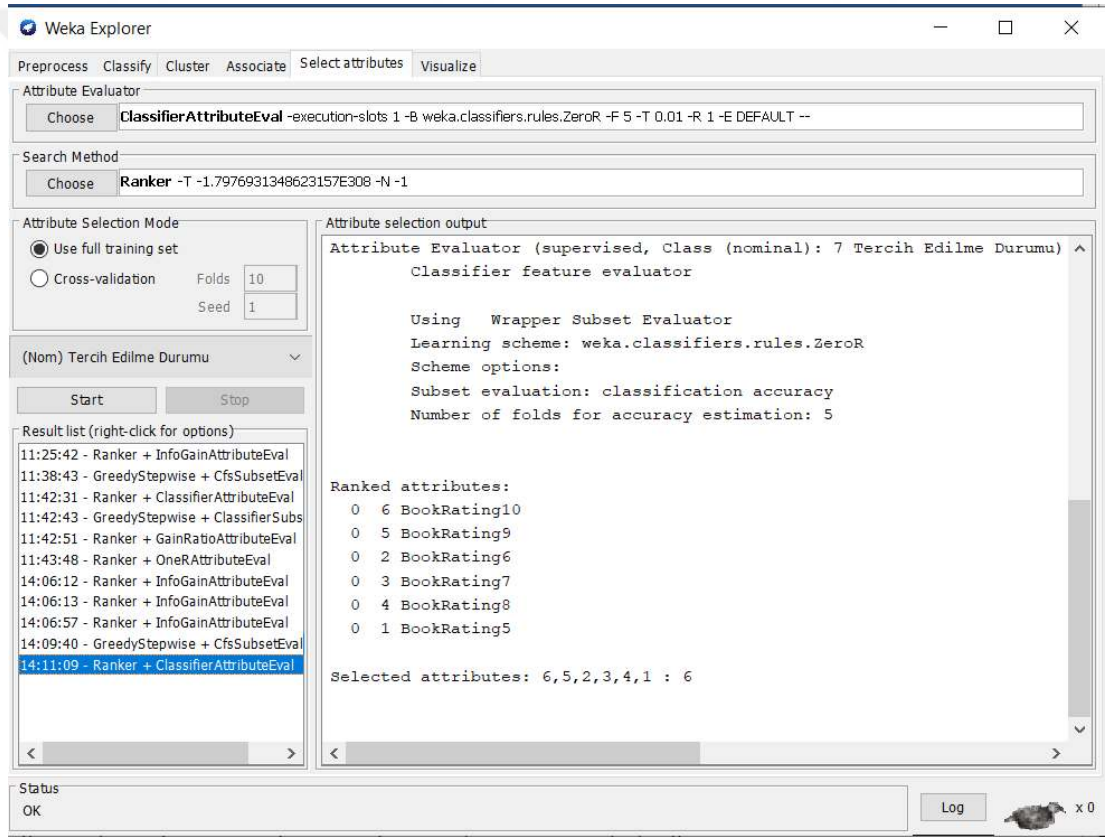
1. BookRating8
2. BookRating9
3. BookRating7
4. BookRating10
5. BookRating6
6. BookRating5



Şekil 4.17. Information Gain algoritma sonuçları

Classifier Attribute Eval: Classifier Attribute Eval algoritmasının sonuçlarına göre en iyi yoğunluğu belirleyen puan türleri aşağıdaki gibi sıralanmıştır. Şekil 4.18'te Weka programında Classifier Attribute Eval sonuçları gösterilmiştir.

1. BookRating10
2. BookRating9
3. BookRating6
4. BookRating7
5. BookRating8
6. BookRating5

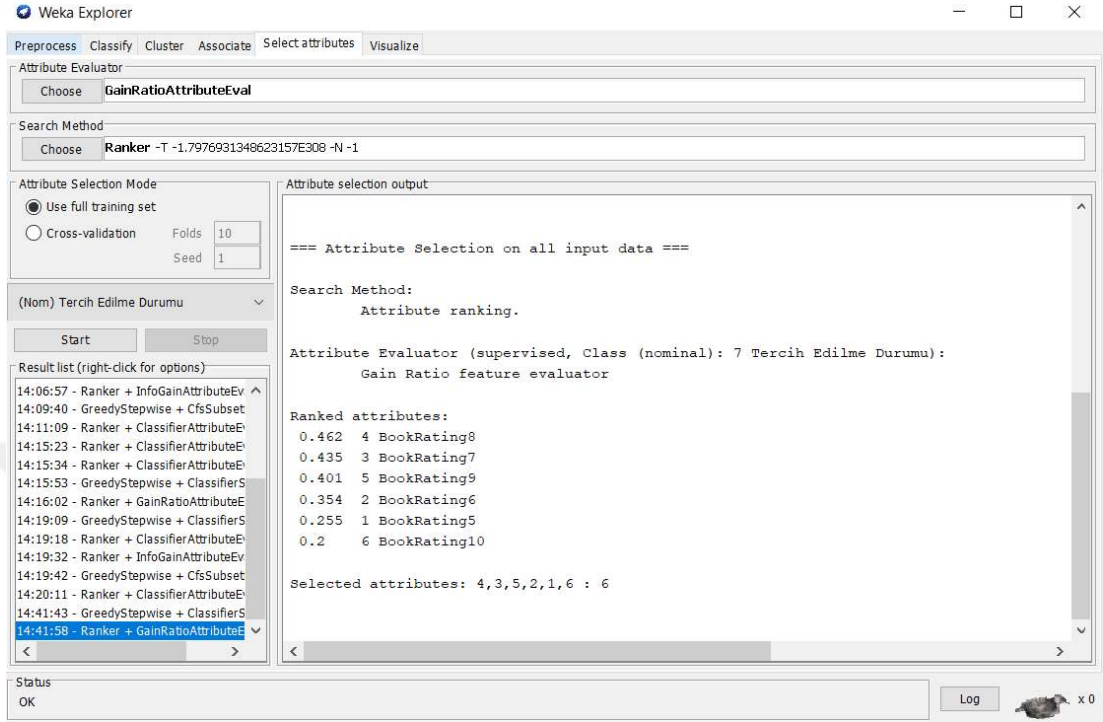


Şekil 4.18. Classifier Attribute Eval algoritma sonuçları

Gain Ratio: Gain Ratio algoritmasının sonuçlarına göre en iyi yoğunluğu belirleyen puan türleri aşağıdaki gibi sıralanmıştır. Şekil 4.19'da Weka programında Gain Ratio sonuçları gösterilmiştir.

1. BookRating8
2. BookRating7
3. BookRating9

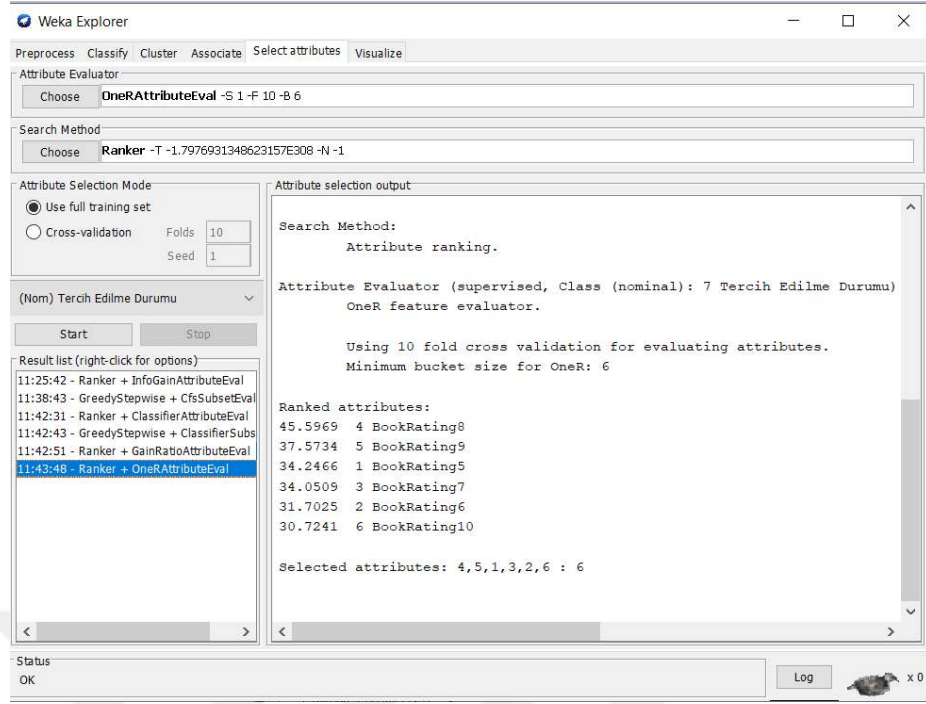
4. BookRating6
5. BookRating5
6. BookRating10



Şekil 4.19. Gain Ratio algoritma sonuçları

One-R: One-R algoritmasının sonuçlarına göre en iyi yoğunluğu belirleyen puan türleri aşağıdaki gibi sıralanmıştır. Şekil 4.20’de Weka programında One-R sonuçları gösterilmiştir.

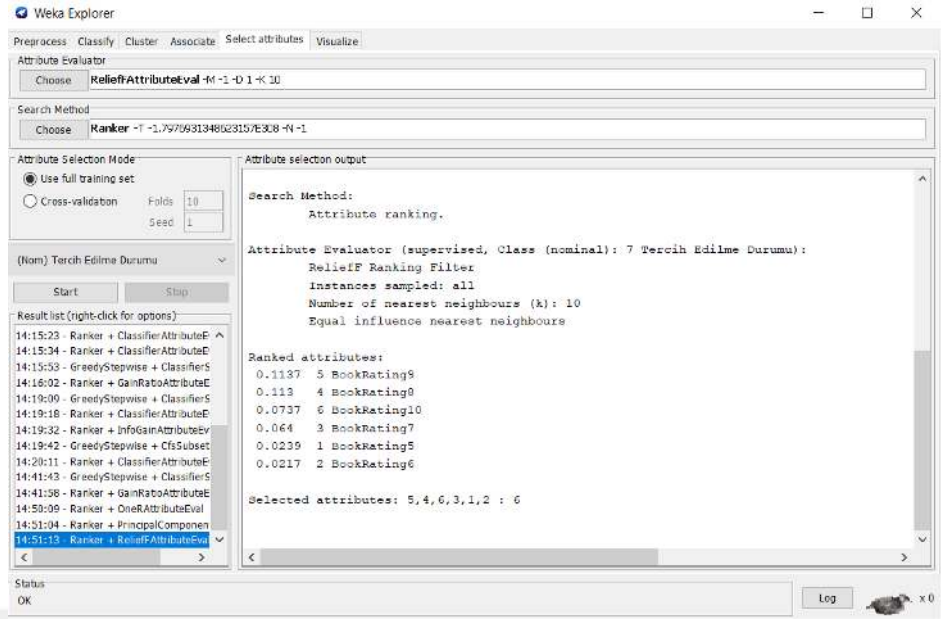
1. BookRating8
2. BookRating9
3. BookRating5
4. BookRating7
5. BookRating6
6. BookRating10



Şekil 4.20. One-R algoritma sonuçları

Relief: Relief algoritmasının sonuçlarına göre en iyi yoğunluğu belirleyen puan türleri aşağıdaki gibi sıralanmıştır. Şekil 4.21’de Weka programında Relief sonuçları gösterilmiştir.

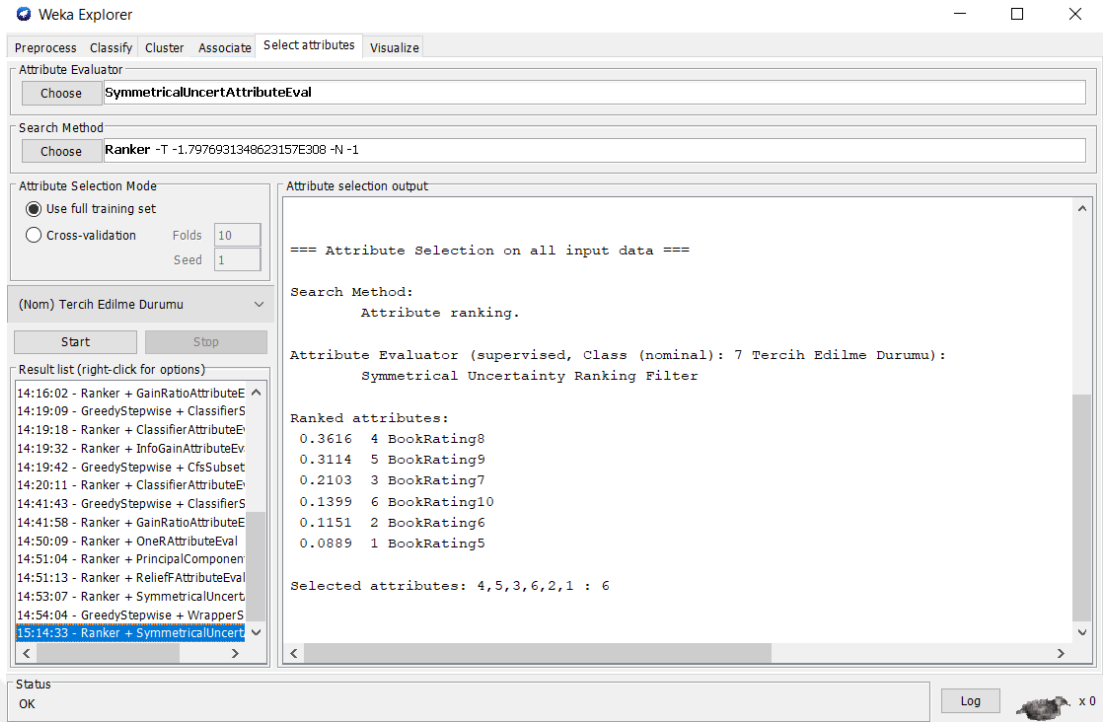
1. BookRating9
2. BookRating8
3. BookRating10
4. BookRating7
5. BookRating6
6. BookRating5



Şekil 4.21. Relief algoritma sonuçları

Symmetrical Uncert Attribute Eval: Symmetrical Uncert Attribute Eval algoritmasının sonuçlarına göre en iyi yoğunluğu belirleyen puan türleri aşağıdaki gibi sıralanmıştır. Şekil 4.22’de Weka programında Symmetrical Uncert Attribute Eval sonuçları gösterilmiştir.

1. BookRating8
2. BookRating9
3. BookRating7
4. BookRating10
5. BookRating6
6. BookRating5



Şekil 4.22. Symmetrical Uncert Attribute Eval algoritma sonuçları

Tüm algoritma sonuçları aşağıdaki Tablo 4.19’da verilmiştir. Apriori algoritmasına sıralamada 1. ve 2. sırada yer alan puan türleri seçilmiştir. Bir sonraki bölümde bu puan türlerine ait oylamalar üzerinde Apriori algoritması çalıştırılmıştır ve kurallar ile ilgili detaylar verilmiştir.

Tablo 4.19. Feature Selection algoritma sonuçları

Yöntem	Öznitelik Sıralaması	Seçilenler
Filtresiz	10,8,9,7,6,5	8,10
Info Gain	8,9,7,10,6,5	8,9
Classifier	10,9,6,7,8,5	10,9
Gain Ratio	8,7,9,6,5	8,7
One R	8,9,5,7,6,10	8,9
Relief	9,8,10,7,5,6	9,8
Syun	8,9,7,10,6,5	8,9

Veri Madenciliği, Apriori Algoritmasının Çalıştırılması: Bu bölümde öncelikle seçilmiş puan türlerine ait veriler, Apriori algoritmasına sokulacaktır. Bu verilerin sonuçlarına göre en iyi kural çıkartan puan türü grubu bulunacak ardından bu gruba ait sonuçlar ve kurallar 4.5.2. bölümünde değerlendirilecektir.

Tablo 4.19’de yer alan seçilenler sütunundaki puan türlerine ait veriler ile algoritma çalıştırılmıştır. Tablo 4.20’de verilere ait feature selection algoritma

sonuçları, seçilen ilk iki puan türü, toplam oy sayısı, destek değeri ve destek*oylama sonucu verilmiştir. İlk sırada filtresiz yani sadece en çok geçen puan türleri alınmıştır. Sonuç kısmında feature selection algoritmasının seçtikleri puan türlerinin kural sonuçları ile feature selection kullanmadan direkt olarak en çok geçen puan türünün kural sonuçları karşılaştırılacaktır. Bu sayede feauture selection işleminin Apriori algoritmasına faydası gösterilecektir.

Tablo 4.20’de yer alan “Öznitelik Sıralaması” puan türlerine ait öznitelik seçimi algoritmaların sıralama sonuçlarını göstermektedir. Seçilenler ise çalışmada kullanılması planlanmış en yüksek iki değeri göstermektedir. Oy sayısı bu puan türlerine ait oylama sayılarını göstermektedir. Destek değeri ise oy sayısı ile çarpımına göre 8 olacak şekilde ayarlanmıştır. Buradaki amaç destek değerini her birinde eşit düzeyde verebilmektir.

Tablo 4.20. Seçilmiş verilere ait sayısal bilgiler

Yöntem	Öznitelik Sıralaması	Seçilenler	Oy Sayısı	Destek Değeri	Destek * Oylama
Filtresiz	10,8,9,7,6,5	8,10	19.869	0,0004	8
Info Gain	8,9,7,10,6,5	8,9	19.394	0,0004	8
Classifier	10,9,6,7,8,5	10,9	16.829	0,0005	8
Gain Ratio	8,7,9,6,5	8,7	18.482	0,0004	8
One R	8,9,5,7,6,10	8,9	19.394	0,0004	8
Relief	9,8,10,7,5,6	9,8	19.394	0,0004	8
Syun	8,9,7,10,6,5	8,9	19.394	0,0004	8

4.5.2 Kitaplara Ait Birliktelik Kuralları, Sık Geçen Öge Kümeleri, Feature Selection Tarafından Sıralanmış Puan Türleri Gruplarının Kural Çıkarım Sonuçları: Tablo 4.20’de bulunan puan türlerine ait oylamalar üzerinde Apriori algoritması çalıştırılmıştır. Tablo 4.21’de bu puan türlerine ait sonuçlar verilmiştir. Feature selection (öznitelik seçimi) algoritmaları kullanmadan direkt olarak veri setinde en çok geçen ilk iki puan türü alındığında %100 güvenilir değere sahip 78 kural elde edilmiştir. Feature selection (öznitelik seçimi) algoritmalarının sonuçlarına göre hareket edildiğinde 786 kural elde edilmiştir. Tabloya bakıldığında en çok %100 güvenilir kural sayısı 8 ve 9 puan türlerinden çıkmıştır. 8 ve 9 sıralamasını Information Gain, One-R, Relief, Symmetrical Uncert Attribute Eval algoritmaları vermiştir. 8 ve 9 puan türlerine ait veriler yüksek sayıda kural çıkartmıştır.

Tablo 4.21. Seçilmiş puan türlerinin Apriori algoritması sonuçları

Yöntem	Öznitelik Sıralaması	Seçilenler	%10 Güvenilir Kural Sayısı	%50 Güvenilir Kural Sayısı	%90 Güvenilir Kural Sayısı	%100 Güvenilir Kural Sayısı
Filtresiz	10,8,9,7,6,5	8,10	1515	593	78	78
Info Gain	8,9,7,10,6,5	8,9	6941	4215	809	786
Classifier	10,9,6,7,8,5	10,9	3365	1630	258	239
Gain Ratio	8,7,9,6,5	8,7	779	263	18	18
One R	8,9,5,7,6,10	8,9	6941	4215	809	786
Relief	9,8,10,7,5,6	9,8	6941	4215	809	786
Syun	8,9,7,10,6,5	8,9	6941	4215	809	786

Kitaplar Arası Birliktelik Kuralları: Apriori algoritmasının çıkarttığı kurallara ait sayısal bilgiler Tablo 4.22’de verilmiştir. Minimum güven filtrelemesine göre kural sayılarını göstermektedir.

Tablo 4.22. Kitaplar arası birliktelik kuralları sayısal bilgiler

Veri	Sayısal Değeri
%100 Güvenilirlik Değerine Sahip Kural Sayısı	786
%90 ve üzeri Güvenilirlik Değerine Sahip Kural Sayısı	809
%80 ve üzeri Güvenilirlik Değerine Sahip Kural Sayısı	1752
%70 ve üzeri Güvenilirlik Değerine Sahip Kural Sayısı	2634

Apriori algoritması sonuçlarına göre %100 güven değerine sahip kural sayısı 786’tır. Oluşturulmuş kurallar ile bir kitabı beğenen kullanıcının başka hangi kitapları beğenebileceği tahmin edilebilecektir. 786 kuraldan bazıları Tablo 4.23’te gösterilmiştir. İlk kurala bakıldığında One for the Money (Bir Para İçin) ve Hard Eight (Sert Sekiz) kitaplarını beğenenler Two fort he Dough kitabını beğenirler durumu ifade etmektedir. Bu kuralın güvenilirlik değeri %100’dür.

Tablo 4.23. Kitaplar arası %100 güven değerine sahip birliktelik kuralları

No	Öncül	Sonuç	Destek	Güven	Lift
1	One for the Money, Hard Eight: A Stephanie Plum Novel	Two for the Dough	0,0004732607	1	170
2	Heart of the Sea, Jewels of the Sun	Tears of the Moon	0,000473261	1	330

3	I Is for Innocent, G Is for Gumshoe	E Is for Evidence: A Kinsey Millhone Mystery	0,00946522	1	364
4	Harry Potter and the Sorcerers Stone, Harry Potter and the Order of the Phoenix	Harry Potter and the Prisoner of Azkaban	0,000662565	1	270
5	A Is for Alibi, G Is for Gumshoe	D Is for Deadbeat	0,000567913	1	320

4.sırada yer alan kurala bakıldığında, Harry Potter and the Sorcerers Stone (Harry Potter ve Felsefe Taşı) ve Harry Potter and the Order of the Phoenix (Harry Potter Zümrüdüanka Yoldaşlığı) kitaplarını beğenenler Harry Potter and the Prisoner of Azkaban (Harry Potter ve Azkaban Tuzağı) kitabını beğenmektedirler. Kurallara lift değeri açısından bakıldığında 1'e eşit olan yoktur. Lift değeri 1 olan kurallar birbirinden bağımsız öğelerden oluşmaktadır. 1'den yüksek değerler ise pozitif korelasyona sahiptir.

Minimum güven değeri sınırı %70 belirlendiği için aşağıdaki tabloda bu güven aralığına ait kurallarda verilmiştir. Tablo 4.24'te %70 güven değerine sahip kurallardan birkaçı gösterilmiştir. Key of Valor kitabını beğenmiş olan kullanıcılar Key of Knowledge kitabında beğenmişlerdir. Bu kuralın güven değeri %72'dir. 2.sırada yer alan kural şunu ifade etmektedir: Gerilim ve polisiye romanları olan A Long Came A Spider ve To Die: A Novel kitaplarını beğenenler John Grisham'ın gerilim romanı olan The Summons kitabını da beğenmektedirler. Bu kuralın güvenilirlik değeri %83'tür.

Tablo 4.24. Kitaplar arası %70 ve üzeri güven değerine sahip birliktelik kuralları

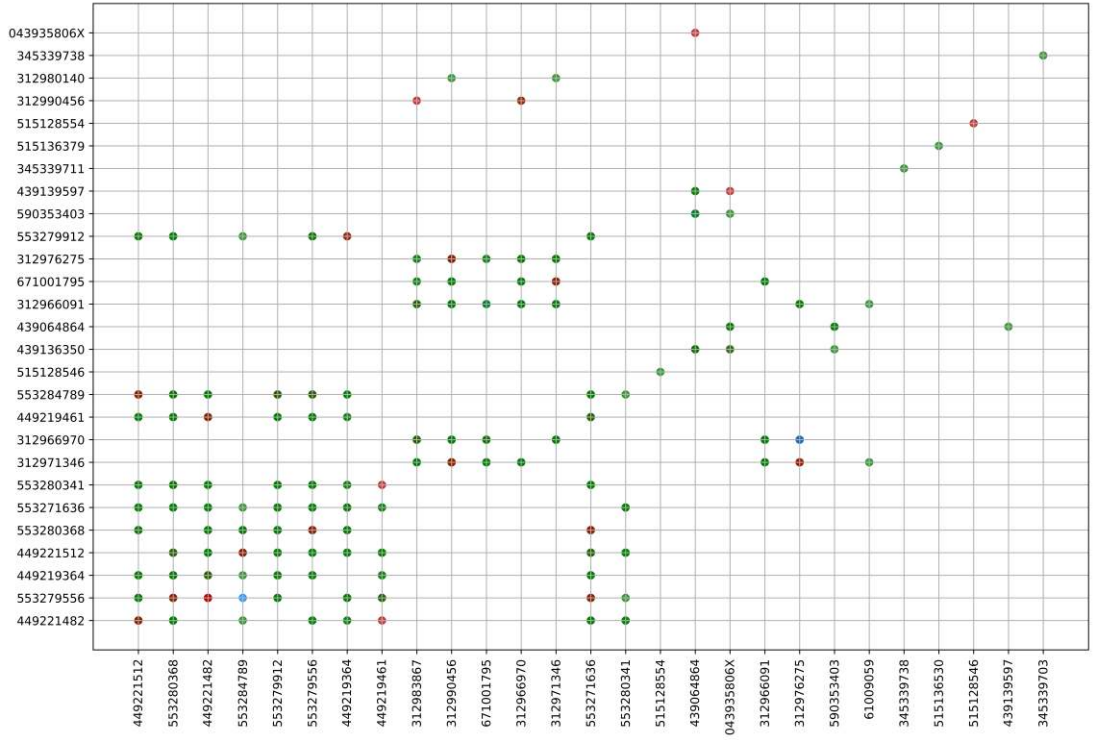
No	Öncül	Sonuç	Destek	Güven	Lift
1	Key of Valor	Key of Knowledge	0,000757217	0,72	480
2	Four To Score, One for the Money	Three To Get Deadly: A Stephanie Plum Novel	0,000757217	0,80	153

3	Four To Score, Three To Get Deadly: A Stephanie Plum Novel	High Five (A Stephanie Plum Novel)	0,00179839	0,77	174
---	--	------------------------------------	------------	------	-----

Veri seti içerisinde seri halinde kitaplar da bulunmaktadır. Harry Potter ve Yüzüklerin Efendisi serisi buna bir örnek olarak verilebilir. Serilerden bazıları kural halinde önerilmiştir. Bu kurallar serinin 1.kitabını okuyan kullanıcının 2.kitabını da okur mantığında kurallar değildir. Serinin 1.kitabını beğenmiş olan kullanıcıların 2.kitabını da beğenmesiyle kural oluşturulmuştur. Eğer kullanıcılar serinin 2.kitabına 5 altında oy verseydi Apriori algoritması bu kitaplara ait kural oluşturmayacaktı. Kural sonuçlarına bakıldığında seriye ait 1.kitap yer alırken serinin diğer kitapları herhangi bir kuralda bulunmayacaktır ve bu bilgi göz önüne alınarak serinin önerilmesinin doğru olmadığı anlaşılabacaktır.

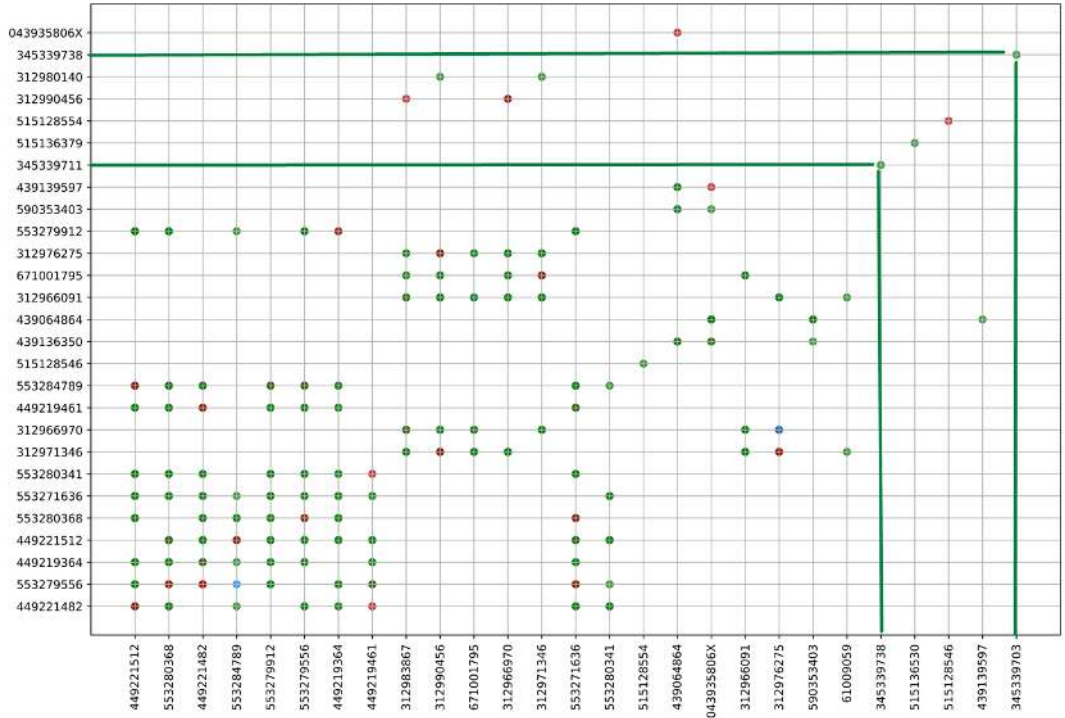
Alışveriş fişleri içeren bir veri seti olsaydı 1.kitap ve 2.kitap satın alındığı için kural oluşacak ve kullanıcılara beğenilmeyen ama satın alındığı için oluşan kural önerilecekti. Çalışmada kullanılan veri seti oylamalara dayandığı için serilere ait tüm kitapların varlığı kurallar içinde bulunup bulunmadığı kontrol edilebilir ve tüm serinin kitapları yer alıyorsa bu serinin kullanıcılar tarafından beğenildiği anlaşılabılır. Serinin 1.kitabını okuyan kullanıcılara, direkt olarak 2.kitabı okuma ihtimali var diyerek öneri sistemleri tarafından kural oluşturulabilir fakat bu kuralın altını dolduracak verilere ihtiyaç vardır. Bu çalışmada ise kullanıcılara ait oylama verileri ile bu kuralın doğruluk ihtiyacı kanıtlanmıştır.

Kullanıcı Yüzüklerin Efendisi serisine ait 1.kitabı okuduğunda 2.kitabı da okur durumu otomatik olarak değil çalışmada yer alan veri setindeki oylamalara dayanarak sunulmuştur. Serideki kitaplara ait kuralların güven değeri ile bu kuralın oluşmasının doğru olduğunu kanıtlamıştır. Veri seti oy derecelendirmelerinden oluştuğu için bu avantajı çalışmaya kazandırmıştır. Şekil 4.23'te ve Şekil 4.24'te kitaplara ait bazı birliktelik kuralları görselleştirilmiştir. Kurallar %60 ve üzeri güven değerine sahiptir.



Şekil 4.23. %60 ve üzeri güven değerine sahip birliktelik kuralları- 1

%60 ve üzeri güven değerine sahip kurallar verilmiştir. Yeşil renkteki bağıntılar %70 ve üzeri güven değerine sahip bağıntıları, mavi renkteki bağıntılar %69 ile %65 arası ve kırmızı %65 altı güven değerine sahip kuralları göstermektedir.

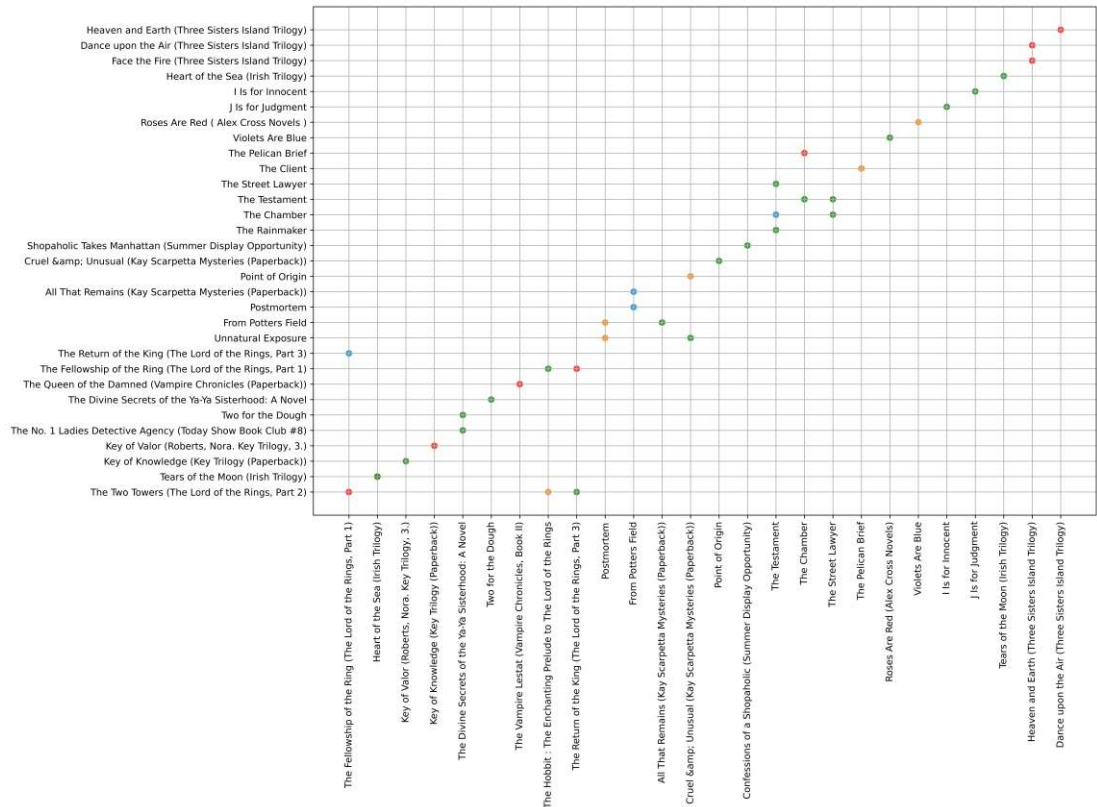


Şekil 4.24. %60 ve üzeri güven değerine sahip birliktelik kuralları- 2

Şekil 4.24'te yeşil ile gösterilen ilk kısım The Fellowship Of The Ring (0345339703) yani Yüzüklerin Efendisi serisine ait 1.kitabı beğenenlerin The Return Of The King (0345339738) Yüzüklerin Efendisi serisine ait 3.kitabı da beğendiğini göstermektedir.

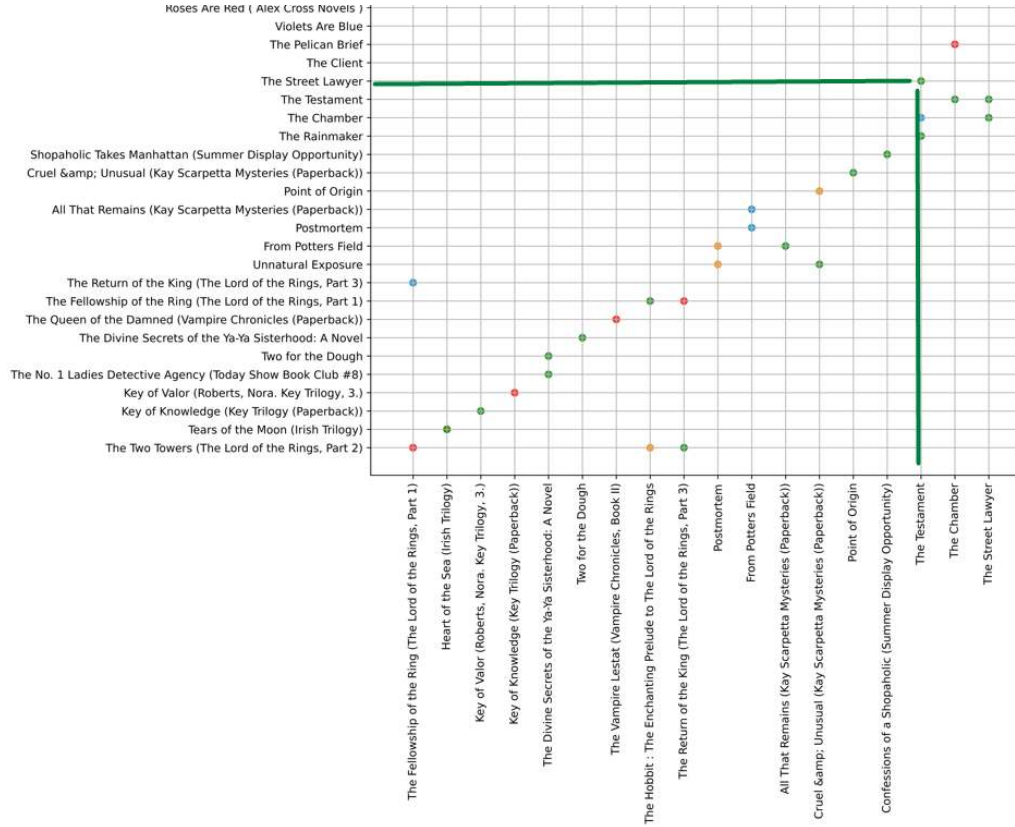
Şekil 4.24'te yeşil ile gösterilen diğer çizgi kısım ise The Return Of The King (0345339738) Yüzüklerin Efendisi serisine ait 3.kitabı beğenenlerin The Two Towers (0345339711) The King Yüzüklerin Efendisi serisine ait 2.kitabı da beğendiğini göstermektedir. Kullanıcılar 1.kitaptan itibaren tüm seriyi beğenerek okumuştur. Bu yüzden öneri sistemlerinde kullanılması mantıklıdır ve serinin tüm kitapları beğenilen kitaplara ait kurallar içinde yer aldığı için sevilen bir seri olduğunu göstermiştir.

Bu seriye ait kitaplar üzerinden çıkan kurallar bu serinin önerilmesinin doğruluğunu kanıtlamıştır. Serilerin önerilmesinde doğruluk ihtiyacı bu bilgilerle sağlanabilmektedir. Kullanılan veri setinde 511 kitap arasından aşırı derecede yer alan üç seriye ait toplam 20 kitap veri setinden çıkartılarak 491 kitap Apriori algoritmasına tekrar sokulmuştur. Buradaki amaç öne çıkan bazı seri kitapların yerine diğer kitaplar arasında oluşmuş kuralları gösterebilmektir. Şekil 4.25'te, Şekil 4.26'da ve Şekil 4.27'de %50 ve üzeri güven değerine sahip kurallar verilmiştir.



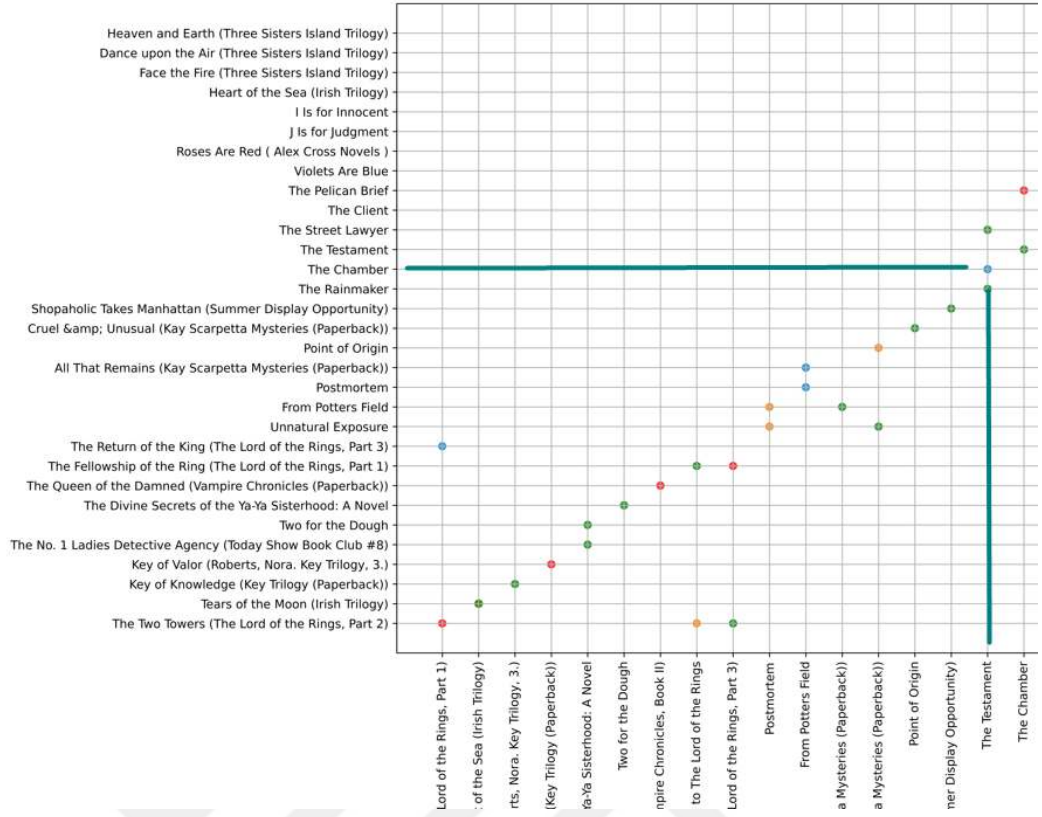
Şekil 4.25. %50 ve üzeri güven değerine sahip birliktelik kuralları- 1

Yeşil renkteki bağlantılar %70 ve üzeri güven değerine sahip bağlantıları, mavi renkteki bağlantılar %69 ile %60 arası güven değerine sahip bağlantıları, sarı %59 ile %55 arası güven değerine sahip bağlantıları ve kırmızı %55 altı güven değerine sahip kuralları göstermektedir.



Şekil 4.26. %50 ve üzeri güven değerine sahip birliktelik kuralları- 2

Yukarıdaki Şekil 4.26'ya bakıldığında The Street Lawyer kitabını beğenen kullanıcıların The Testament kitabını da beğendiğini göstermektedir. Yeşil bağlantı noktasına sahip olduğu güven değeri %70 ve üzeri değere sahiptir. Kitaplara ait bilgiler ile kuralın daha iyi açıklanması için aşağıdaki bilgiler verilmiştir: İki romanda gerilim romanıdır. The Street Lawyer kitabı bir avukatı anlatırken The Testament kitabı vasiyetnamenin şaşırtıcı bilgileri yüzünden 3 kişinin başına gelenleri anlatmaktadır.



Şekil 4.27. %50 ve üzeri güven değerine sahip birliktelik kuralları- 3

Şekil 4.27’de mavi çizgilerin gösterdiği bağıntı ise The Chamber kitabını beğenen kişilerin The Testament kitabını beğendiğini göstermektedir. Güven değeri %70 ile %60 arasındadır. The Chamber kitabı hukuki bir gerilim kitabı olup gerilim kitabıdır. The Testament kitabı da bir gerilim kitabıdır.

5. SONUÇLAR

Bu çalışmada insanların kitap oylamalarından yararlanılarak kitaplar ve kitapların yazarları üzerinden kurallar üretebilmek için Apriori algoritması kullanılmıştır. Veri setinde yer alan kitapların yazarları arasındaki ilişkiler tespit edilmiştir. Bir yazarı beğenen kişinin başka hangi yazarları beğenebileceği tespit edilmiştir.

Seyrek olan kitap oylamaları üzerinde kural çıkartabilmek için kitapların yoğunluğuna göre gruplama, bu gruplamaya ait istatistiki verilere göre eleme, feature selection yöntemi kullanarak en yoğun puan türüne ait oylamaları bulma işlemi yapılmıştır. Ayrıca veri seti bir alışveriş fişinden değil oylama verilerinden oluştuğu için bir seriye ait olan kitapların sadece satın alınma durumuna bakarak değil insanlar tarafından beğenilme durumu göz önüne alınarak kurallar oluşturulmuştur. Sonucunda beğenilen kitap serileri de ortaya çıkmıştır. Bu sayede %70 ve üzeri güven değerine sahip kitap serisine ait çıkarılmış kurallar ile serinin önerilmesinin doğru olduğu da kanıtlanmıştır.

Bir kitabı beğenen kişinin başka hangi kitapları beğenebileceği ve bir yazarı beğenen kişinin başka hangi yazarları beğenebileceğine dair kurallar çıkarılmıştır. Çalışmada üretilen kuralların önerilen kitap listelerinde ve satış stratejilerinde kullanılması beklenmektedir.

KAYNAKÇA

- [1] Alaeddinoğlu, M.F., Aydın, T., Dal, D. Birliktelik kuralları ile mekânsal-zamansal veri madenciliği. Fen Bilimleri Enstitüsü. 2012, 5(2), 192-193.
- [2] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. The KDD process for extracting useful knowledge from volumes of data. Communications of the Association for Computing.1996, 39(11), 27-34.
- [3] Can M.B., Çamur, E., Koru, M., Özkan, Ö., Rzayeva, Z. Veri Kümelerinden Bilgi Keşfi: Veri Madenciliği. Başkent Üniversitesi Tıp Fakültesi XIV. Öğrenci Sempozyumu, 15-16 Mayıs, 2012, Ankara.
- [4] Gürgen, G. Birliktelik Kuralları Uygulaması ve Sepet Analizi. T.C. Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, İstanbul, 2008, 104 s. (Yüksek Lisans Tezi).
- [5] Alan, A., Yeşilyurt, C. Birliktelik kuralları madenciliği ile yatan hasta profilinin çıkarılması. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü. 2019, 23(Özel sayı),1917-1926.
- [6] Baykal, A. Veri madenciliği uygulama alanları. Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi. 2006 ,7, 95-107.
- [7] Ateş, Y., Karabatak, M. Nicel birliktelik kuralları için çoklu minimum destek değeri. Fırat Üniversitesi Mühendislik Bilimleri. 2017, 29(2), 57-65.
- [8] Agrawal, R., Srikant, R. Fast Algorithms for Mining Association Rules. Proceeding of the 20th Int. Conference on Very Large Database, 12-15 September, 1994, San Jose.
- [9] Özçakır, F.C., Çamurcu, A. Birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması.İstanbul Ticaret Üniversitesi Fen Bilimleri.2007, 6(12), 21-37.
- [10] Erdem, S., Özdağoğlu, G. Ege Bölgesi'ndeki bir araştırma ve uygulama hastanesinin acil hasta verilerinin veri madenciliği ile analiz edilmesi. Anadolu Üniversitesi Bilim ve Teknoloji. 2008, 9(2), 261-270.
- [11] Çil, İ., Ay, D. Migros Türk A.Ş.de birliktelik kurallarının yerleşim düzeni planlamada kullanılması. Sakarya Üniversitesi Endüstri Mühendisliği. 2010, 21(2), 14-29.
- [12] Gülce, A.C. Veri Madenciliğinde Apriori Algoritması ve Apriori Algoritmasının Farklı Veri Kümelerinde Uygulanması. Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Edirne, 2010, 82 s. (Yüksek Lisans Tezi).
- [13] Erpolat, S. Otomobil yetkili servislerinde birliktelik kurallarının belirlenmesinde Apriori ve FP-Growth algoritmalarının karşılaştırılması. Anadolu Üniversitesi Sosyal Bilimler. 2012, 12(2),137-146.
- [14] Acun, G., Bilgin T.T. Yazılım hata logları kullanılarak veri madenciliği. Marmara Fen Bilimleri. 2015, 27(1), 14-20.
- [15] Doğrul, G., Akay, D., Kurt, M. Trafik kurallarının birliktelik kuralları ile analizi. Gazi Mühendislik Bilimleri. 2015, 1(2), 265-284.
- [16] Pehlivanoğlu, M.K., Duru, N. Veri madenciliği teknikleri kullanılarak ortaokul öğrencilerinin sosyal ağ kullanım analizi: Kocaeli ili örneği. Düzce Üniversitesi Bilim ve Teknoloji. 2015, 3(2), 508-517.
- [17] Çelik, B., Aytekin, P. Perakende Sektöründe Pazar Sepeti Analizi için Apriori Algoritması ile Birliktelik Kurallarının Oluşturulması. Pazarlama ve Pazarlama araştırmaları Derneği Pazarlama Kongresi, 1-4 Mayıs, 2019, İzmir.
- [18] Vahaplar, A., İncoğlu, M.M. Veri Madenciliği ve Elektronik Ticaret. Türkiye İnternet Konferansları, 1-3 Kasım, 2001, İstanbul.

- [19] Emel, G.G., Taşkın, Ç. Pazarlama stratejilerin oluşturulmasında bir karar destek aracı : birliktelik kuralı madenciliği. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü. 2005, 7(3), 30-59.
- [20] Agrawal, R., Imielinski, T., Swami, A. Mining Association Rules Between Sets of Items in Large Databases. Association for Computing Machinery SIGMOD International Conference on Management of Data, 1 June, 1993, New York.
- [21] Birant D., Kut, A., Ventura, M., Altınok, H., Altınok, B., Altınok, E., Altınok M. İş Zekası Çözümleri için Çok Boyutlu Birliktelik Kuralları Analizi. Akademik Bilişim'10 - XII. Akademik Bilişim Konferansı, 10-12 Şubat, 2010, Muğla.
- [22] Akpınar, H. Veri tabanlarında bilgi keşfi ve veri madenciliği. İstanbul Üniversitesi İşletme Fakültesi Dergisi. 2000, 29(1), 1-22.
- [23] Sharma, S., Bhatia, S. Analysis of Association rule in Data Mining. Second International Conference on Information and Communication Technology for Competitive Strategies, 4-5 March, 2016, Haryana (Bildiri Özetleri Kitabı, 1-4 s.)
- [24] Sağın, A. N., Ayvaz, B. Determination of association rules with market basket analysis: an application in the retail sector. Southeast Europe Journal of Soft Computing. 2018, 7(1), 10-19.
- [25][Çevrimiçi]. Available: <http://bilgisayarkavramlari.sadievrenseker.com/2011/09/09/birliktelik-kurallarini>. [Erişildi: 22 Haziran 2021].
- [26] Han, J., Kamber, M., Pei, J. Frequent Itemsets, Closed Itemsets, Data mining: Concepts and Techniques, 105-259 s., 2006.
- [27] Al-zawaidah, F.H., Jbara, Y.H., Abu-Zanona M.A.-A. An improved algorithm for mining association rules in large databases. World of Computer Science and Information Technology Journal. 2011, 1(7), 311-316.
- [28] Panjaitan, S., Sulindawaty, Amin, M., Lindawati, S., Watrianthos, H., Sihotang, H.T., Sinaga, B. Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns. The Interikjjujkunational Conference on Computer Science and Applied Mathematic, August, 2019, Medan.
- [29] Ay, D., Veri Madenciliği ve Apriori Algoritması., T.C. Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Anabilim Dalı, Sakarya, 2009. (Yüksek Lisans Tezi).
- [30] Gala, D. Analyzing Association Rules Produced By Applying The Apriori Algorithm To Structured Data. University of Nevada, Las Vegas, 2006. (Yüksek Lisans Tezi).
- [31] Tan, P.N., Steinbach, M., Kumar, V. Introduction To Data Mining, Pearson Addison-Wesley Publishing, 2005, 359-369 s.
- [32] Ergün, E. Ürün Kategori Arasındaki Satış İlişkisinin Birliktelik Kuralları ve Kümeleme Analizi ile Belirlenmesi ve Perakende Sektöründe Bir Uygulama. Afyon Kocatepe Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, Afyonkarahisar, 2008, 202 s. (Yüksek Lisans Tezi).
- [33] Wu, X., Kumar, V., Ouinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., Mclachlan, G.J., Ng, A., Liu, B., Yu, P.S. Top 10 algorithms in data mining. 2008, 14(1), 1-37.
- [34] Budak, H. Özellik seçim yöntemleri ve yeni bir yaklaşım. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi. 2018, 22(Özel), 21-31.
- [35] Forman, G. An extensive empirical study. Journal of Machine Learning. 2003, 3, 1289-1305.
- [36] Dash, M., Liu, H. Feature selection for classification. Intelligent Data Analysis. 1997, 1(1-4), 131-156.

- [37] Ladha, L., Deepa, T. Feature selection methods and algorithms. International Journal on Computer Science and Engineering. 2011, 3(5), 1787-1797.
- [38] Langley, P. Selection of Relevant Features in Machine Learning. Proceedings of the AAAI Fall Symposium on Relevance, 4-6 November, 1994, California.
- [39] Liu, H., Yu, L. Towards integrating feature selection algorithms for classification and clustering. IEEE Transactions On Knowledge And Data Engineering. 2005, 17(4), 491-502.
- [40] Larranaga, P., Sayes, Y. A review of feature selection techniques in bioinformatics. Bioinformatics. 2007, 23(19), 2507-2517.
- [41] Guyon, I., Taşkın, Ç. An introduction of variable and feature selection. Journal of Machine Learning Research. 2003, 3, 1157-1182.
- [42] Yazıcı, B., Yaşı, F., Gürleyik, H.Y., Turgut, U.O., Aktas, M.S., Kalıpsız, O. Veri Madenciliğinde Özellik Seçim Tekniklerinin Bankacılık Verisine Uygulanması Üzerine Araştırma ve Karşılaştırmalı Uygulama. 9. Ulusal Yazılım Mühendisliği Sempozyumu (UYMS-15), 15- 17 Eylül, 2015, İzmir.
- [43] Hall, A. Correlation-based Feature Selection for Machine Learning. The University of Waikato, New Zeland, 1999.(Doktora Tezi).
- [44] Şık, M.Ş. Veri Madenciliği ve Kanser Erken Teşhisinde Kullanımı. İnönü Üniversitesi, Sosyal Bilimler Enstitüsü, Malatya, 2014. (Yayınlanmamış Yüksek Lisans Tezi).
- [45] Novaković, J., Strbac, P. Toward optimal feature selection using ranking methods and classification algorithms. Yugoslav Journal of Operations Research. 2011, 21(1), 119-135.
- [46] Holte, R.C. Very simple classification rules perform well on most commonly used datasets. 1993, 11, 63-90.
- [47] Kaynar, O., Arslan, H., Görmez, Y., Işık, Y.E. Makine öğrenmesi ve öznelik seçim yöntemleriyle saldırı tipleri. Bilişim Teknolojileri Dergisi. 2018, 11(2), 175-185.
- [48] Kira, K., Rendell L.A. A Practical Approach to Feature Selection. Ninth International Workshop on Machine learning, 1-3 July, 1992, Scotland (Bildiri Özeti Kitabı 249-256 s.).
- [49] Konenکو, I. Estimating Attributes: Analysis and Extensions of RELIEF. Machine Learning: ECML-94, 6-8 April, 2005, Catania (Bildiri Özeti Kitabı, 171-182 s.).
- [50] [Çevrimiçi]. Available: [https://en.wikipedia.org/wiki/Relief_\(feature_selection\)](https://en.wikipedia.org/wiki/Relief_(feature_selection)). [Erişildi: 25 Şubat 2022].
- [51][Çevrimiçi]. Available: https://en.wikipedia.org/wiki/SQL_Server_Management_Studio. [Erişildi: 1 Mart 2022].
- [52][Çevrimiçi]. Available: https://web.archive.org/web/20120623165941/http://cutter.rexx.com/~dkuhlman/python_book_01.html. [Erişildi: 1 Mart 2022].
- [53][Çevrimiçi]. Available: <https://9kuyruk.com/python-nedir-basit-python-ornekleri.htm>. [Erişildi: 1 Mart 2022].
- [54][Çevrimiçi]. Available: [https://en.wikipedia.org/wiki/Spyder_\(software\)#cite_note-techworld-4](https://en.wikipedia.org/wiki/Spyder_(software)#cite_note-techworld-4). [Erişildi: 1 Mart 2022].

[55] Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G. Improving Recommendation Lists Through Topic Diversification. 14th International World Wide Web Conference (WWW '05), Chiba, 2005.

