

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**E-TİCARET SİTELERİ İÇİN BULANIK MANTIK VE
MAKİNE ÖĞRENMESİ TABANLI BİR ÖNERİ
SİSTEMİ**

YÜKSEK LİSANS TEZİ

Metehan UÇAR

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ**
Tez Danışmanı : Dr. Öğr. Üyesi M. Fatih ADAK

Haziran 2021

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**E-TİCARET SİTELERİ İÇİN BULANIK MANTIK VE
MAKİNE ÖĞRENMESİ TABANLI BİR ÖNERİ
SİSTEMİ**

YÜKSEK LİSANS TEZİ

Metehan UÇAR

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ**

Bu tez 25/06/2021 tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.

Jüri Başkanı

Üye

Üye

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Metehan UÇAR

TEŞEKKÜR

Bilgi ve deneyimleriyle eğitimime katkı sağlayan bütün hocalarıma özellikle de Yüksek Lisans eğitimim boyunca her konuda desteğini esirgemeyen, araştırmanın planlanmasından yazım sürecine kadar tüm aşamalarda değerli tecrübeleriyle çalışmama destek olan ve titizlikle yönlendiren çok değerli danışman hocam Dr. Öğr. Üyesi M. Fatih ADAK'a teşekkürlerimi sunarım.

Ayrıca, eğitim hayatım boyunca her zaman yanımda olan ve desteklerini esirgemeyen aileme teşekkür ederim.

İÇİNDEKİLER

TEŞEKKÜR.....	i
İÇİNDEKİLER	ii
SİMGELER VE KISALTMALAR LİSTESİ	iv
ŞEKİLLER LİSTESİ	v
TABLOLAR LİSTESİ.....	vii
ÖZET.....	viii
SUMMARY	ix

BÖLÜM 1.

GİRİŞ	1
1.1. Öneri Sistemlerine Genel Bakış	2
1.1.1. İçerik tabanlı filtreleme tekniği	4
1.1.2. İşbirlikçi filtreleme tekniği.....	5
1.1.3. Hibrit filtreleme tekniği.....	7
1.2. E-Ticaret’te Öneri Sistemleri	7
1.3. Bulanık Mantık.....	8
1.4. Karar Ağaçları	13
1.5. Çalışmanın Amacı ve Önemi	16

BÖLÜM 2.

LİTERATÜR ARAŞTIRMASI	18
-----------------------------	----

BÖLÜM 3.

MATERYAL VE YÖNTEM.....	22
3.1. Veri Seti.....	22
3.1.1. Amazon reviews veri seti	22

3.1.2. GoodReads veri seti	24
3.1.3. Veri setinin hazırlanması.....	25
3.2. Geliştirilen Modelin Çalışma Prensipleri	29
3.3. Desteksiz Öğrenme ile Kümeleme	30
3.4. Karar Ağacı Modeli.....	32
3.5. Bulanık Model.....	34
BÖLÜM 4.	
BULGULAR VE TARTIŞMA	40
4.1. Desteksiz Öğrenme ile Kümelenmiş Verilerden Elde Edilen Bulgular	40
4.2. Karar Ağacından Elde Edilen Bulgular.....	40
4.3. Değerlendirme Metrikleri.....	46
BÖLÜM 5.	
SONUÇLAR VE ÖNERİLER	48
KAYNAKLAR	51
ÖZGEÇMİŞ	58

SİMGELER VE KISALTMALAR LİSTESİ

Asin	: Amazon Product Id
COA	: Center of Area
COG	: Center of Gravity
CSV	: Comma Separated Values
FOM	: First of Maxima
LOM	: Last of Maxima
MOM	: Mean of Maxima
JSON	: JavaScript Object Notation

ŞEKİLLER LİSTESİ

Şekil 1.1. Öneri Sistemleri Teknikleri	3
Şekil 1.2. İçerik tabanlı filtreleme şeması.....	4
Şekil 1.3. İşbirlikçi filtreleme şeması.....	5
Şekil 1.4. İşbirlikçi filtreleme teknikleri	6
Şekil 1.5. Amazon.com.tr birlikte alınan ürün önerileri	8
Şekil 1.6. Amazon.com.tr müşterilerin aldığı diğer ürünlerin önerilmesi	8
Şekil 1.7. Üçgen üyelik fonksiyonu	9
Şekil 1.8. Yamuk üyelik fonksiyonu.....	10
Şekil 1.9. Gauss üyelik fonksiyonu.....	11
Şekil 1.10. Maksimumlar yöntemi ile durulaştırma.....	13
Şekil 1.11. Karar ağacı örneği	14
Şekil 3.1. Çalışmada kullanılan veri setleri.....	22
Şekil 3.2. JSON dosyasının dataframe veri yapısına aktarılması	23
Şekil 3.3. Sayfa sayısı verilerinin histogram grafiği	26
Şekil 3.4. Fiyat verilerinin histogram grafiği	26
Şekil 3.5. Puan verilerinin histogram grafiği	27
Şekil 3.6. Kategori verilerinin histogram grafiği	28
Şekil 3.7. Benzer kategorilerin birleştirilmesiyle oluşan verilerin histogram grafiği	28
Şekil 3.8. Önerilen yöntemin akış şeması	29
Şekil 3.9. Kategori önerisinin kullanıcıya sunulmasına ait akış şeması	30
Şekil 3.10. Desteksiz öğrenme ile kümelenmiş veriler	32
Şekil 3.11. Karar ağacı modeli.....	33
Şekil 3.12. Küme değişkeninin üyelik fonksiyonu	34
Şekil 3.13. Fiyat değişkeninin üyelik fonksiyonu.....	35
Şekil 3.14. Sayfa sayısı değişkeninin üyelik fonksiyonu.....	35
Şekil 3.15. Puan değişkeninin üyelik fonksiyonu	36

Şekil 3.16. Kategori değişkeninin üyelik fonksiyonu	36
Şekil 3.17. jFuzzyLogic kütüphanesi ile üyelik fonksiyonlarının oluşturulması.....	37
Şekil 3.18. jFuzzyLogic kütüphanesi ile durulaştırma uygulaması	39
Şekil 4.1. Karar ağacının performans ölçümü.....	41
Şekil 4.2. Karar ağacından kural çıkarımı örneği	41
Şekil 4.3. Karar ağacı düğümleri.....	42
Şekil 4.4. Karar ağacının sol kısmı	43
Şekil 4.5. Karar ağacının orta kısmı.....	44
Şekil 4.6. Karar ağacının sağ kısmı.....	45



TABLÖLAR LİSTESİ

Tablo 3.1. Amazon Reviews metadata veri seti.....	24
Tablo 3.2. Amazon Reviews kullanıcı ürün puanlaması veri seti.....	24
Tablo 3.3. GoodReads veri setindeki sayfa sayısı bilgileri.....	25
Tablo 3.4. Karar ağacı modelinin parametreleri	34
Tablo 3.5. Bulanık model kuralları	38
Tablo 4.1. Kümelenmiş verilerin dağılımı	40
Tablo 4.2. Değerlendirme Sonuçları	47

ÖZET

Anahtar kelimeler: Öneri sistemleri, desteksiz öğrenme, bulanık mantık, E-ticaret, karar ağaçları

Öneri sistemleri özellikle e-ticaret sitelerinde, film izleme platformlarında, müzik dinleme platformlarında kullanıcılara en doğru ürünü tavsiye etmek üzerine geliştirilen sistemlerdir. Günümüzde veri üzerine yapılan çalışmaların artması öneri sistemlerinde de farklı yöntemler uygulanmasına yol açmıştır.

Yapılan çalışmada e-ticaret sitelerinde kitap satın almak isteyen kullanıcılara yönelik bulanık mantık tabanlı ürün öneri sistemi sunulmuştur. Alınan ve görüntülenen diğer kitaplar verisi bilgileri ile desteksiz öğrenme (Unsupervised learning) kullanılarak kümeleme yapılmıştır. K-means algoritması kullanılarak yapılan kümeleme işlemi ile altı küme elde edilmiştir. Kitaplara ait diğer değişkenler olan sayfa sayısı, fiyat ve puan bilgilerinin bulunduğu veri seti modele uygun olacak şekilde hazırlanmıştır.

Veri temizleme ve kümeleme işlemleri sonrasında elde edilen veri seti ile karar ağacı modeli oluşturulmuştur. C4.5 yöntemi kullanılan karar ağacı, değişkenlerin kitap kategorisini belirlemek için olan etkisini göstermiştir. Çalışmada kullanılan bulanık modelinin kuralları, bu karar ağacından yararlanarak oluşturulmuştur. Gerçek veri ile yapılan testler sonucunda, karar ağaçları ve bulanık modeller bir arada kullanıldığında başarılı sonuçların alındığı görülmüştür.

Normalde veriye ihtiyaç duyulmayan bulanık modellerde, modelin tasarımı sırasında değişkenlerin ve etkilerinin çok iyi bilinmesi gerekir. Fakat bu çalışmadaki gibi karmaşık ve zor modellerde kuralların oluşturulması oldukça zor olacaktır. Karar ağaçları gibi bir yöntem yardımıyla kuralların hızlı ve doğru bir şekilde oluşturulabileceği, bu çalışmada elde edilen başarılı sonuçlar neticesinde anlaşılmıştır.

A RECOMMENDATION SYSTEM BASED ON FUZZY LOGIC AND MACHINE LEARNING FOR E-COMMERCE SITES

SUMMARY

Keywords: Recommendation systems, unsupervised learning, fuzzy logic, E-commerce, decision tree

Recommendation systems are systems developed to advise on the most accurate product to users, especially on e-commerce sites, movie viewing platforms, and music listening platforms. Nowadays, the increasing number of studies on data has led to the application of different methods in recommendation systems.

In the study, a fuzzy logic-based product recommendation system has been presented for users who want to buy books on e-commerce sites. Clustering has been made using unsupervised learning with information from the “also bought-viewed” book data. Six clusters were obtained by the clustering process using the K-means algorithm. The data set, which includes the page count, price and rating data, which are the other parameters of the books, has been prepared in accordance with the model.

A decision tree model has been created with the data set obtained after data cleaning and clustering. The decision tree using the C4.5 method indicated the effect of the parameters to predict the book category. The rules of the Fuzzy model used in the study have been created by using this decision tree. It has been observed that successful results are obtained when tests are performed with real data and decision trees and fuzzy models are used together.

Usually, in fuzzy models, data is not required. It is necessary to know the parameters and their effects during the design of the model. However, it will be complicated to determine rules in complex and challenging models like as in this study. As a result of the successful results obtained in this study, it has been understood that the rules can be created quickly and accurately with the help of a method such as decision trees.

BÖLÜM 1. GİRİŞ

Günümüzde internet kullanıcılarının artması ve kullanıcıların internet ortamında gerçekleştirdiği olayların neticesinde oluşan ve özellikle son 20 yılda çeşitli alanlarda verilerin büyük ölçüde artmasıyla [1] birlikte veri teknolojisinin önemi ortaya çıkmıştır. 19. yüzyıl başlarından ortaya çıkan bilgi teknolojilerinde veri oluşturma kapasitesi özellikle son yıllarda hiç görülmemiş kadar güçlü bir şekilde sürmektedir [2]. Bankacılık, iletişim, medya ve eğlence sektörü, sağlık hizmetleri, eğitim, üretim, devlet hizmetleri, sigortacılık, perakendecilik ve ticaret, ulaşım, enerji sektörü gibi pek çok alanda büyük veri uygulamaları kullanılmaktadır [3]. Bu derecede artan büyük veri havuzlarından veri elde etmek, anlamlı bilgi ve örüntü çıkarma ve tahminler yapma durumu veri madenciliği uygulamalarını ortaya çıkarmıştır.

İnternet ortamında kullanıcı verilerinin her geçen gün artmasıyla oluşan büyük veri havuzlarından anlamlı bilgiler çıkarılması için çalışmalar yapılmış ve 1990'lı yılların ortalarından itibaren öneri sistemleri veri teknolojisi içerisinde ayrı bir konu olarak araştırılmaya başlanmıştır [4].

Öneri sistemleri özellikle e-ticaret sitelerinin ürün önerilerinde, film izleme ve müzik dinleme uygulamalarında kullanıcıya doğru içerik önermek için kullanılmaktadır. Amazon [5], Netflix [6], Spotify [7] ve Youtube [8] gibi platformların öneri sistemlerinin geliştirilmesi hakkında pek çok çalışmalar bulunmaktadır.

Tez çalışmasında e-ticaret sitelerinde kitap almak isteyen kullanıcılar için karar ağacı tabanlı bulanık mantığı kullanan bir öneri sistemi sunulmuştur. Çalışmada Amazon Reviews veri seti [9] ve bu veri setinde eksik olan sayfa sayısı bilgileri için de GoodReads [10], [11] veri seti kullanılmıştır. Amazon Reviews veri setinden yöntemde kullanılmak üzere fiyat, puan, ürünü alanların aldığı ve görüntülediği diğer

ürünler ve kategorileri içeren veriler alınmıştır. GoodReads veri setinden ise aynı kitaplara ait sayfa sayısı verileri alınmıştır.

Amazon Reviews veri setindeki ürünü alanların satın aldığı ve görüntülediği diğer ürünlerin bilgileri listeler içerisinde kitaplara ait Asin değerleri olarak bulunmaktadır. Listeler içerisinde bulunan Asin değerleri sınıflandırılmış veriler değildir. Bu gibi durumlar için desteksiz öğrenme metodu, doğru sonuçlar veren bir yöntemdir. Veriler desteksiz öğrenme metodu kullanılarak k-means algoritması kümelendirilmiştir. Kümelerin içerisindeki kitaplar incelenerek kitap türlerinin yakınlığına göre kümeler Küme 1'den Küme 6'ya kadar sıralanmıştır. Böylelikle bulanık modelde kullanılmak üzere bir değişken daha elde edilmiştir.

Veri temizleme ve kümeleme işlemlerinin sonrasında elde edilen veri seti ile bir karar ağacı modeli oluşturulmuştur. Kullanım kolaylığı ve eldeki verilere uygunlu nedeniyle C4.5 algoritması kullanılmıştır. Karar ağacı modelinin uygulanmasıyla ortaya çıkan düğüm ve yapraklar yorumlanmış ve bulanık modelin kuralları oluşturulmuştur.

Bulanık model için java ortamında jFuzzyLogic [12], [13] kütüphanesi kullanılmıştır. Sayfa sayısı, puan, fiyat ve desteksiz öğrenme ile oluşturulan kümeler için üyelik fonksiyonları oluşturulmuştur. Üyelik fonksiyonlarının oluşumunda verilerin maksimum, minimum ve ortalama değerleriyle birlikte karar ağacındaki düğüm ve yapraklar gözetilmiştir. Bulanık modelin çıktısı olarak ise kitap kategorileri kullanılmıştır. Üyelik fonksiyonları belirlenen veriler bulanıklaştırılmış ve kuralların uygulanması sonucunda kesin değerleri elde etmek için durulaştırma yöntemi uygulanmıştır. Çalışmanın sonunda durulaştırma yöntemi sonucu elde edilen kategori önerileri gerçek verilerle karşılaştırılmış ve değerlendirmeleri yapılmıştır.

1.1. Öneri Sistemlerine Genel Bakış

Son yıllarda e-ticaret siteleri, film ve müzik platformlarında kullanıcıların ürün ve içeriklere yaptığı yorumlar ve verdikleri puanlar büyük veri havuzları oluşturmuştur. Öneri sistemleri bu veri havuzlarından alınan verilerin analiz edilip belli benzerlik

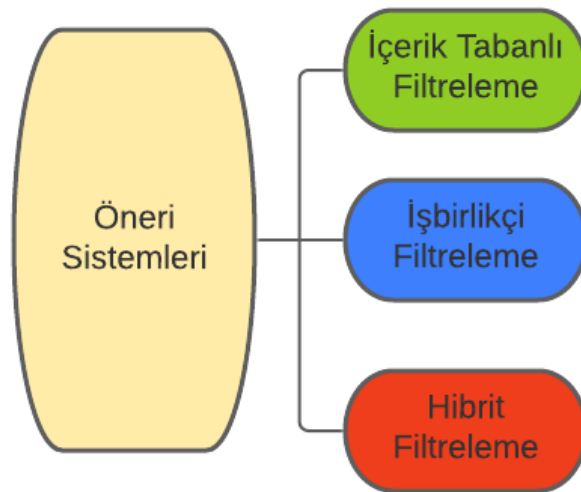
kurallarının oluşturulmasıyla kullanıcıya en doğru ürün ve içeriği tavsiye etmeyi amaçlamaktadır. Öneri sistemlerindeki en belirgin amaç kullanıcının ürün ve içerikleri arama motorlarında aramadan onlara uygun ürün ve içeriği tespit edip sunmaktır. Öneri sistemlerinin satışlar üzerindeki etkisini araştıran çalışmalarda da satışların artmasını sağladığı görülmüştür [9].

Genel olarak kullanıcılara ilgisini çekebilecek ürünleri tavsiye etmeyi hedefleyen öneri sistemleri çalışma prensipleri açısından üç farklı şekilde incelenebilir:

İçerik tabanlı (Content-Based) filtreleme tekniği, kullanıcının daha önceden tercih ettiği ürünlerin benzerlerini kullanıcıya önerir [14]–[16].

İşbirlikçi (Collaborative) filtreleme tekniği, farklı kullanıcıların tercih ettiği ürünler ile kullanıcı için benzer kullanıcılar bularak benzer kullanıcıların tercihlerini kullanıcıya önerir veya tercih edilen ürünlerin arasında benzerlik ilişkisi kurarak kullanıcının tercih ettiği ürünlere benzer ürünleri kullanıcıya önerir [17]–[19].

Hibrit (Hybrid) filtreleme tekniği ise içerik tabanlı filtreleme tekniği ile işbirlikçi filtreleme tekniğini bir arada kullanan öneri sistemi tekniğini temsil eder [20], [21].

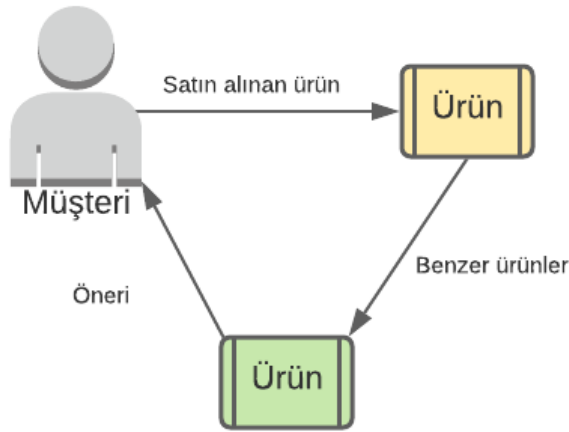


Şekil 1.1. Öneri Sistemleri Teknikleri

1.1.1. İçerik tabanlı filtreleme tekniği

İçerik tabanlı filtreleme tekniği kullanıcının geçmişte tercih ettiği ürünlerin bilgilerine dayanarak kullanıcıya tavsiye sunan öneri sistemlerinin geliştirilmesinde kullanılmaktadır [22]–[24]. Bu filtreleme tekniğinde kullanıcı puanlamasıyla etiketlenen ürün açıklamaları kullanıcıya özel sınıflandırma ve regresyon modeli oluşturmak için eğitim verileri olarak kullanılır [25]. Öneri sistemlerinde en çok kullanılan filtreleme tekniklerinden biridir. Diğer filtreleme tekniklerine göre avantajı elde daha az kullanıcı verisi olmasına rağmen kullanılabilir olmasıdır.

Bu filtreleme tekniğinde genellikle verilerden bir dizi özellik çıkarılarak hesaplamalar yapılır ve önerinin uygunluğunu belirlemek için kullanılır [4]. Amaç kullanıcının ürün hakkında sağladığı geri dönüşler ve tercihler ile o kullanıcıya tavsiye sunabilmek için doğru bir kullanıcı profili oluşturmaktır. Netflix'in kullanıcılar ilk kayıt yaptığı aşamada kullanıcıya beğendiği filmleri sorması ve o kullanıcıya ait bir profil oluşturması buna örnek olarak gösterilebilir.



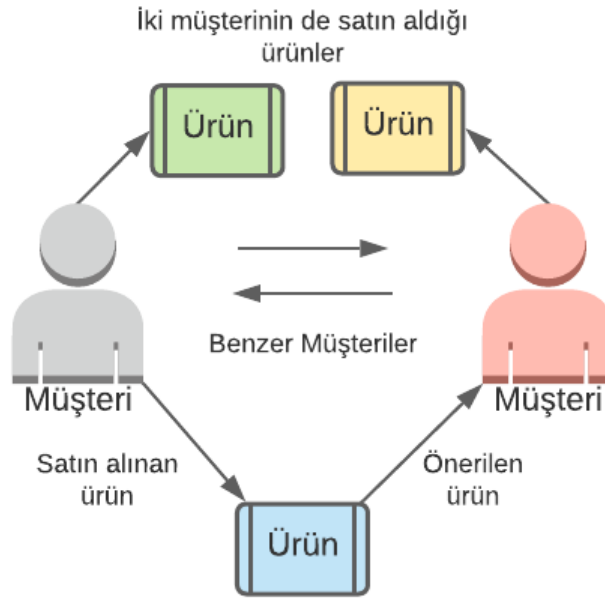
Şekil 1.2. İçerik tabanlı filtreleme şeması

İçerik tabanlı filtreleme tekniğinde anlaşılması kolay olması, güvenilirliğinin yüksek olması, kaynak tüketiminin düşük olması gibi güçlü yönlerinin olmasına karşılık sadece sığ bir analiz üzerinden öneri sunması ve kullanıcının sadece önceki deneyimlerine benzer ürünler görmesi gibi zayıf yönleri de bulunmaktadır [26]. İçerik

tabanlı öneri sistemlerinde doğru ve güvenilir öneriler sunabilmek için kullanıcıların yeterli sayıda ürün değerlendirmeleri bilinmelidir [4].

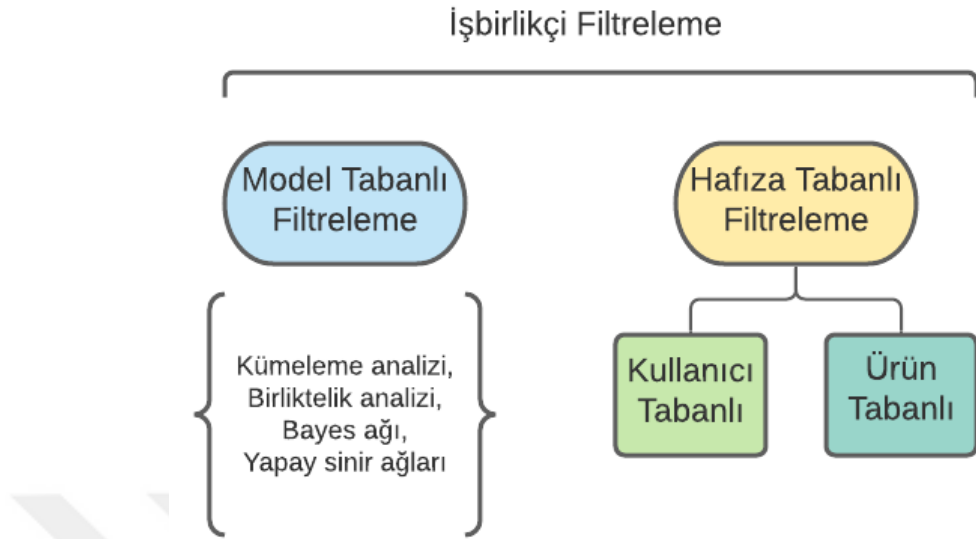
1.1.2. İşbirlikçi filtreleme tekniği

İşbirlikçi filtreleme tekniği öneri sistemlerinde en çok kullanılan filtreleme tekniklerinden biridir. İçerik tabanlı filtreleme tekniğine göre çok daha büyük veri setleri kullanılarak uygulanan bir yöntemdir. Çok fazla kullanıcının oluşturduğu veriler kullanılarak kullanıcıya özel tavsiye metotları uygulanmaktadır. 1992 yılında ilk kez Goldberg tarafından bir öneri sistemi çalışmasında ortaya çıkmıştır [27]. İşbirlikçi filtreleme sisteminde genel mantık geçmişte aynı tercihlerde bulunan kullanıcıların gelecekte de aynı tercihlere yöneleceği üzerinedir.



Şekil 1.3. İşbirlikçi filtreleme şeması

Birçok çalışmada model tabanlı (model-based) ve hafıza tabanlı (memory-based) olarak iki farklı şekilde incelenmiştir.



Şekil 1.4. İşbirlikçi filtreleme teknikleri

İşbirlikçi filtreleme yöntemleri hakkındaki çalışmalarda model tabanlı filtrelemelere nazaran hafıza tabanlı filtreleme yöntemi daha doğru sonuçlar ürettiği için daha yaygın kullanılmaktadır.

Hafıza tabanlı öneri sistemlerinde kullanıcı tabanlı ve ürün tabanlı komşuluk değerlendirmeleri sonucu öneri üretir. Groplens [28] ve Ringo [29] için yapılan çalışmalar bu alandaki ilk çalışmalar olarak gösterilebilir. Kullanıcı tabanlı öneri sisteminde bir kullanıcının kullanıcı-ürün matrisine göre geçmişte oy kullandığı ürünlere bakılarak ona yakın bir kullanıcı grubu (en yakın komşu) bulunur ve sonrasında en yakın komşuların ortalamaları alınarak kullanıcıya ürün önerisinde bulunulur. Ürün tabanlı öneri sisteminde ise kullanıcı-ürün matrisinde ürünler arasındaki benzerlikler kullanılarak kullanıcıya ürün önerisinde bulunulur.

Model tabanlı öneri sistemlerinde ise kullanıcı tercihleri için tanımlayıcı model oluşturulur. Bu modellerin birçoğunda makine öğrenmesi metotları bulunmaktadır [30]. Kümeleme analizi [31], birliklilik analizi [32], bayes ağı [33] ve yapay sinir ağları [34] yaklaşımları ile modeller oluşturulmaktadır.

İşbirlikçi filtreleme sisteminde çalışılabilecek içerik çeşitliliğinin çokluğu ve farklı tarz ürünleri önerebilme olanağı gibi güçlü yönler olmasına karşılık veri eksikliği ihtimalinin en yakın komşuları bulmayı zorlaştırması ve çok fazla kaynak tüketimine yol açabilmesi gibi zayıf yönleri de bulunmaktadır [26].

1.1.3. Hibrit filtreleme tekniği

İşbirlikçi filtreleme ve içerik tabanlı filtreleme tekniklerinin kendi içlerinde güçlü ve zayıf yönleri bulunmaktadır. Önerilerde performansı artırmak için farklı filtreleme tekniklerinin bir arada kullanarak hibrit filtreleme teknikleri oluşturulmaktadır. Böylelikle daha verimli öneri sistemleri ortaya çıkarmak hedeflenmiştir. Yapılan çalışmalarda hibrit filtreleme tekniğinin içerik tabanlı ve işbirlikçi filtrelemelere göre daha doğru sonuçlar verdiği saptanmıştır [35]–[37]. E-ticaret sitelerinde yapılan tavsiyelerde de farklı tekniklerin güçlü yönlerini bir arada kullanmak için de hibrit filtreleme teknikleri kullanılmaktadır [38].

1.2. E-Ticaret’te Öneri Sistemleri

Günümüzde e-ticaret siteleri öneri sistemlerinin başlıca uygulama alanlarından biridir. E-ticaret sitelerinin doğru kullanıcıya doğru ürünü önerme hedefleri satış rakamlarının artmasında önemli yer edinmektedir.

E-ticaret alanında önde gelen Amazon.com ve ülkemizde n11.com, gittigidiyor.com gibi e-ticaret siteleri öneri sistemlerine büyük önem vermektedir. Bu öneri sistemleri sonucu kullanıcıya sunulan öneri iyiye ortalama sipariş artabilir bu neticede gelirlerini arttırabilirler [39].

Geliştirilen öneri sisteminden sonra e-ticaret siteleri satış ekranlarında kullanıcılara bu önerileri farklı şekillerde sunabilirler. Örneğin kullanıcı bir ürün sayfasındayken o ürün bilgilerine benzer ve o ürün ile bağlantılı başka ürünleri Şekil 1.5.’te olduğu gibi “Sıklıkla birlikte alınan ürünler” başlığı altında sunmaktadırlar.

Sıklıkla Birlikte Alınan Ürünler



- ✓ Bu ürün: Xiaomi Mi True Wireless Earphones Basic 2 Bluetooth Kablosuz Kulak İçi Kulaklık, Siyah ₺196,30
- ✓ Xiaomi Mi Band 5 Akıllı Bileklik Siyah ₺227,87
- ✓ Xiaomi Mi Body Composition Scale 2 Tartı, Beyaz ₺162,74 (₺3.537,83 / 100 cm)

Şekil 1.5. Amazon.com.tr birlikte alınan ürün önerileri [40]

Ürün önerilerinin bir başka sunulma yöntemi ise Şekil 1.6.'da olduğu gibi kullanıcının incelediği ürünü alan diğer müşterilerin satın aldığı diğer ürünlerin de kullanıcıya önerilmesidir.

Bu Ürünü Alan Müşterilerin Aldığı Diğer Ürünler



Şekil 1.6. Amazon.com.tr müşterilerin aldığı diğer ürünlerin önerilmesi [40]

1.3. Bulanık Mantık

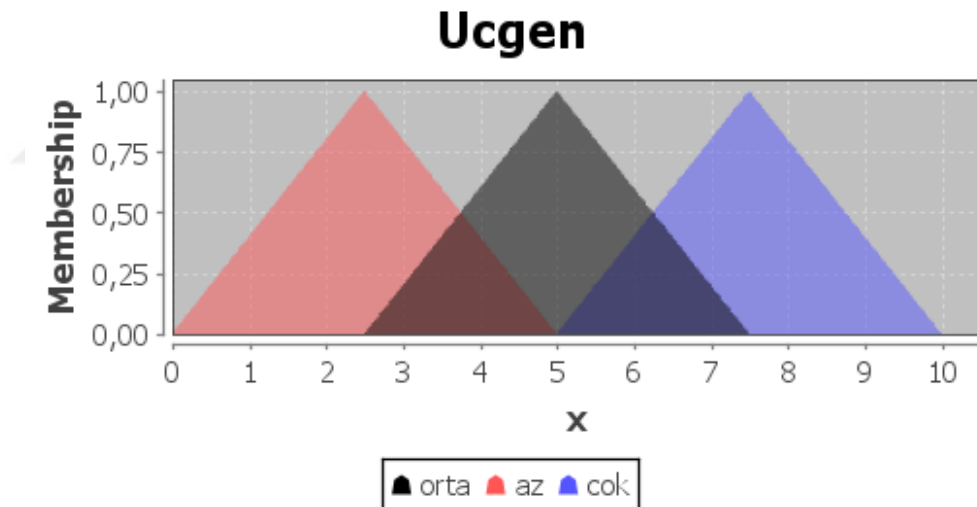
Bulanık mantık 1965 yılında Lotfi A. Zadeh tarafından yayınlanan bir çalışmada ortaya çıkmış, kesin değerler yerine yaklaşık ifadeleri kullanan, dilsel ifadeler arasında tanımlanan kurallar ile gerçekleştirilen bir kavramdır [41]. Bu kavram klasik mantığın aksine yaklaşık durumlar için işlem yapılmasına olanak sağlar. Bu alandaki ilk uygulama 1973 yılında Ebrahim Mamdani'nin buhar makineleri üzerindeki çalışması olmuştur [42]. Günümüze kadar gelen süreçte otomatik kontrol sistemleri, bilgi

sistemleri, görüntü tanımlama ve optimizasyon gibi alanlarda uygulamaları yapılmıştır [43].

Bulanık mantık uygulamalarının oluşturma aşamalarında değişkenlerin bulanık kümelerle olan ilişkileri üyelik fonksiyonları oluşturularak ifade edilir. Üyelik fonksiyonları bütün reel sayılar olarak ifade edilirken içerisindeki değerler kümesi [0,1] arasında süreklilik ifade eden bütün reel sayılar kümesidir. Bu kümeler farklı üyelik fonksiyonları ile ifade edilebilir. Yaygın olarak kullanılan bazı üyelik fonksiyonu türleri [44], [45]:

- Üçgen
- Yamuk
- Gauss

şeklindedir.

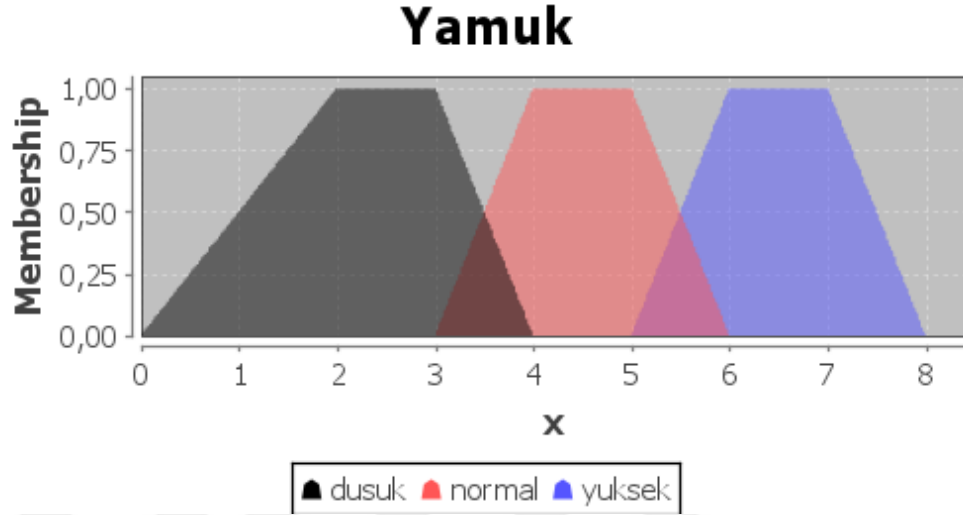


Şekil 1.7. Üçgen üyelik fonksiyonu

Üçgen üyelik fonksiyonunda Şekil 1.7.'de olduğu gibi değerler en düşükten tepe noktasına kadar doğrusal olarak değişkenlik gösterir. Bu üyelik fonksiyonun formülü (Denklem 1.1) aşağıda verilmiştir.

$$\mu_A(x) = \begin{cases} 0, & x < a \\ (x-a)/(b-a), & a < x \leq b \\ (x-c)/(b-c), & b < x \leq c \\ 0, & x \geq c \end{cases} \quad (1.1)$$

Burada değerler a ile c sınırları aralığında ve b tepe noktası olmak üzere süreklilik göstermektedir.

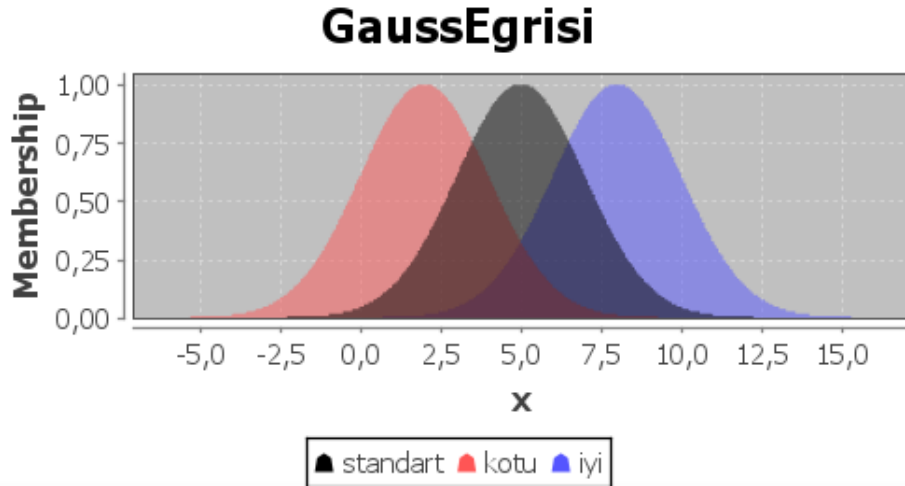


Şekil 1.8. Yamuk üyelik fonksiyonu

Bir diğer üyelik fonksiyonu türü yamuk üyelik fonksiyonlarıdır. Formülü (Denklem 1.2) aşağıda gösterilmiştir.

$$\mu_A(x) = \begin{cases} 0, & x < a \\ (x-a)/(b-a), & a < x \leq b \\ 1, & b < x < c \\ (d-x)/(d-c), & c < x \leq d \\ 0, & x \geq d \end{cases} \quad (1.2)$$

Yamuk üyelik fonksiyonları da üçgen üyelik fonksiyonlarında olduğu gibi en düşükten tepe noktasında kadar süreklilik gösterir. Farklı olarak tepe noktasında birden fazla değerler kümesi bulunmaktadır. Şekil 1.8.'de gösterildiği gibi değerler yamuğun alt köşeleri ve tepe noktaları arasında süreklilik göstermektedir.



Şekil 1.9. Gauss üyelik fonksiyonu

Şekil 1.9.’da gösterilen Gauss üyelik fonksiyonunda değerler doğrusal olmayan bir şekilde süreklilik göstermektedir. Formülde (Denklem 1.3) σ standart sapmayı ifade etmektedir. Grafik c eğri uydurma (curve-fitting) sabiti ile oluşturulmaktadır.

$$\mu_A(x) = e^{-(x-c)^2/\sqrt{(2\sigma)}} \quad (1.3)$$

Üyelik fonksiyonları oluşturulurken çok sayıda yöntem kullanılabilir. Bunlar başlıca; sezgisel, çıkarımsal, sıralama, sinir ağları, genetik algoritma ve tümevarım yöntemleridir [44].

Bulanık mantık modellerinde değişkenlerin üyelik fonksiyonları dilsel terimler ile ifade edilmektedir. Örneğin bir kütle değişkenini sayı ve birimi ile nitelemek yerine o değişkene ait “ağır”, “hafif”, “orta” gibi dilsel terimler ile belirtilir. Bu şekilde oluşturulan üyelik fonksiyonları ile beraber bulanık modelde bulanıklaştırma işlemi gerçekleştirilmiş olur.

Bulanıklaştırılmış giriş ve çıkış verileri arasında istenilen ilişkiyi kurmak için bulanık kurallar oluşturulur. Üyelik fonksiyonlarında olduğu gibi bulanık kurallarda da dilsel terimler ile koşullar belirtilir. Bu kurallar oluşturulurken “EĞER”, “VE-VEYA” ve “O HALDE” ifadeleri kullanılır. “a” ve “b” giriş değerleri, “c” çıkış değeri olarak belirlenmiş bir bulanık modelde, modele ait bir kural:

KURAL: EĞER a A ise ve b B ise O HALDE c C'dir.

Şeklinde ifade edilebilir. "A", "B" ve "C" değişkenlerine ait bulanık kümeleri oluşturmaktadır.

Bulanıklaştırma sonucu çıkış değişkenine ait bulanık küme elde edilir. Çıkış kümesindeki bulanık verileri tekrardan kesin değerlere dönüştürmek için durulaştırma yöntemleri kullanılır. Literatürde çok sayıda durulaştırma yöntemi bulunmaktadır. Bunlardan bazıları; Center of Gravity (COG), Center of Area (COA), First of Maxima (FOM), Last of Maxima (LOM), Mean of Maxima (MOM) gibi yöntemlerdir [46].

COG/COA yöntemi durulaştırma işlemlerinde sıkça kullanılan bir yöntemdir. Çıkış verilerinin üyelik fonksiyonları geometrik alt alanlara bölündükten sonra ağırlık merkezleri bulunur. Bulunan ağırlık merkezlerinin toplamı bulanıklaştırılmış değerlerin sonucunu vermektedir. n sayıda alt alana bölünmüş ve COG/COA yöntemi ile durulaştırılmış Y^* değerlerinin formülü Denklem 1.4 ile gösterilmiştir. A_i alanı, $\bar{y}_i$ ise alan merkezlerini ifade etmektedir.

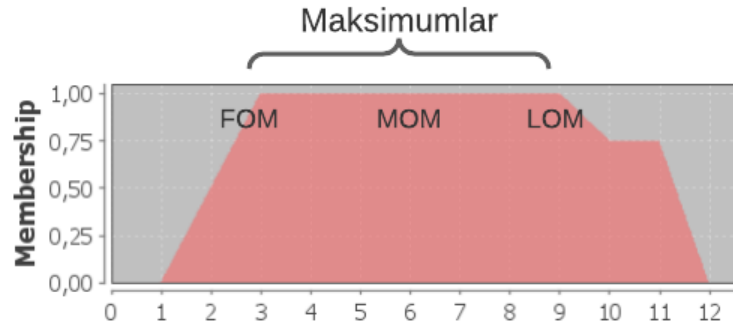
$$Y^* = \frac{\sum_{i=1}^n A_i * \bar{y}_i}{\sum_{i=1}^n A_i} \quad (1.4)$$

FOM/LOM/MOM yöntemlerinde ağırlık merkezi hesaplamalarında değerlerin maksimum karşılıkları göz önüne alınmaktadır. Maksimum değerlerin birden fazla olduğu durumlarda ilk değer, sonuncu değer ve ortadaki değer hesaplanarak işlem yapılır.

$$Y^* = \frac{\sum_{y_i \in M} y_i}{|M|} \quad (1.5)$$

Bu yöntemde durulaştırılmış değer olarak ifade edilen Y^* değeri Denklem 1.5 ile gösterilmiştir. M bulanık kümenin yüksekliğini $|M|$ ise bulanık kümenin eleman sayısını göstermektedir.

Şekil 1.10.'da gösterildiği gibi en yüksek değerin birden fazla olması durumunda ilk değer göz önüne alınırsa FOM, orta değer göz önüne alınırsa MOM, son değer göz önüne alınırsa LOM yöntemi kullanılmış olmaktadır.



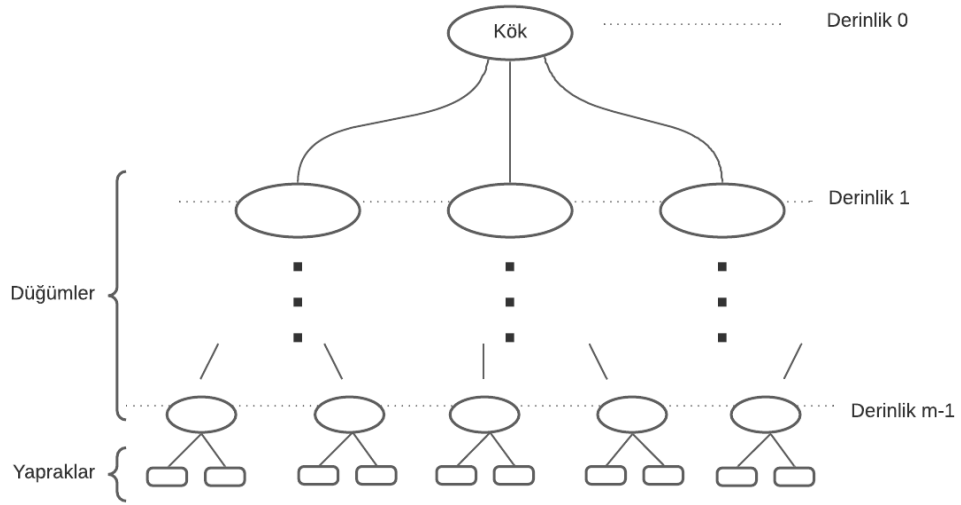
Şekil 1.10. Maksimumlar yöntemi ile durulaştırma

Durulaştırma işlemleri sonucunda bulanık modelin çıkış değerleri kesin değerler olarak elde edilmiş ve sonuca varılmış olur. Çalışmada sunulan bulanık modelde çıkış değerlerini kitap kategorileri oluşturmaktadır. Çıkış değerlerinin üyelik fonksiyonları incelendiğinde simetrik değerler olduğu görülmektedir. Bu nedenle jFuzzyLogic kütüphanesinde uygulanabilecek bir yöntem olan COG yöntemi durulaştırma işleminde kullanılmıştır.

1.4. Karar Ağaçları

Çok aşamalı yaklaşımlarda temel fikir karmaşık bir kararı daha basit kararlardan oluşan modellere ayırarak istenilen çözüme ulaşmayı kolaylaştırmaktır. Karar ağacı modelleri de karmaşık karar verme sürecinde verileri derlenmesini ve basit kararlar sunarak yorumlanmasını kolay hale getiren çözümler sağlar [47].

Örnek bir karar ağacı Şekil 1.11.'de gösterilmiştir [47].



Şekil 1.11. Karar ağacı örneği [47]

Şekil 3.11.'deki karar ağacı modelinde görüldüğü üzere karar ağaçlarındaki ilk hücreye kök adı verilir. Sonraki adımlarda sınıflandırmanın Evet / Hayır olarak koşullandırılması sonucu düğümler oluşur. Düğümler köklerden sonraki hücrelerdir. Köklerde yapılan işlemde olduğu gibi koşullandırma sonucu başka düğümler oluştururlar. En altta bulunan yapraklara ulaşana kadar yapılan koşullandırma sayısı derinlik olarak adlandırılır.

Karar ağaçlarının maliyetinin az olması, anlaşılmasının kolay olması ve güvenilir bir sınıflandırma tekniği olması nedeniye zor ve karmaşık modellerde en yaygın kullanılan sınıflandırma yöntemlerinden biridir [48].

Genellikle sınıflandırma problemlerinde kullanılan karar ağaçlarında ID3 [49], C4.5 [50], Sliq [51] , CART [52], Random Forest [53] gibi başlıca algoritma çeşitleri bulunmaktadır.

Karar ağaçlarında kullanılan bazı algoritmaların özellikleri şu şekildedir [54], [55]:

- CART algoritmasında gini algoritması tabanlı ikili bölme işlemi uygulanır. Her düğüm iki dala ayrılarak devam eder. Budama işleminde karmaşıklık ölçüsü

kullanılır. Verinin hazırlanması gereken CART algoritması sınıflandırma ve regresyonu destekler.

- C4.5 ve C5.0 algoritmaları ID3 algoritmasının geliştirilmiş türleridir. Her düğümden çok sayıda dal oluşturularak işlem devam eder. Tahmin edilmeye çalışılan kategori sayısı kadar dal oluşturulur. Birden fazla karar ağacını birleştirme özelliği olan algoritmaların ayırma işlemleri için bilgi kazancı parametresi kullanılır. Oluşan yapraklardaki hata oranına dayalı budama işlemi yapılır.
- CHAID algoritmasında dal sayıları en az iki en fazla ise tahmin edilmek istenilen kategori sayısı kadardır. Bağımlı değişkenin kategorik olması durumunda “ki-kare” testi uygulanmaktadır. Sürekli bağımlı değişkenlerde ise “F” testi uygulanmaktadır.
- SLIQ algoritması hızlı ve ölçeklenebilir bir yapıda olup, genişlik ilkesi sayesinde birçok yaprağı aynı anda oluşturabilme yeteneğine sahiptir. Dallanma kriterleri gini algoritmasında dayalıdır.

Karar ağacı modellerinde kullanılan algoritma ve verilerin karmaşıklığına göre farklı ayırma kriterleri uygulanmaktadır. Bunların en yaygınları Gini Index (Gini Endeksi) ve Information Gain (Bilgi Kazancı) yöntemleridir.

Bilgi kazancı yönteminde öncelikle rastgele bir değişkenli ilişkili bir belirsizlik ölçüsü olan entropi değeri hesaplanır. Belirsizlik ve rastgelelik ile doğru orantılı olan entropi değeri 0-1 aralığında değişkenlik gösterir. Entropi değeri n durum sayısında i durumunun olasılığı p_i olacak şekilde Denklem 1.6 denkleminde olduğu gibi tanımlanır [56].

$$Entropi(P) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (1.6)$$

Bilgi kazancı hesaplamasında entropi değerleri kullanılır. Bölünmeden sonraki entropi değeri ile bölünmeden önceki entropi değerleri arasındaki fark hesaplanarak bilgi kazancı değeri elde edilir.

Gini endeksi hesaplamasında ise (Denklem 1.7) i sayıda durumun gerçekleşme olasılığının (p_i) karelerinin toplamı hesaplanır. Her sınıf için uygulanır. Ortaya çıkan sonuçlar 0 ile 1 arasında değişkenlik gösterir. Bölünme ne kadar doğru yapılırsa çıkan sonuç sıfıra o kadar yaklaşır.

$$Gini = 1 - \sum_{i=1}^n p_i^2 \quad (1.7)$$

ID3, C4.5 ve C5.0 gibi algoritmalarda bilgi kazancı değerleri hesaplanırken CART algoritması tabanlı karar ağaçlarında gini endeksi hesaplamaları kullanılmaktadır.

Sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan karar ağaçları karmaşık veri setlerinde doğru ve hızlı sınıflandırmalar yapmak için oldukça etkili bir yöntemdir.

1.5. Çalışmanın Amacı ve Önemi

Çalışmada e-ticaret sitelerinden kitap almak isteyen kullanıcılara yönelik bir öneri sistemi geliştirilmiştir. Kullanılan veri setleri kullanıcılara satılmak istenen kitaplar hakkında bilgiler içermektedir. Bunlar fiyat, sayfa sayısı, ürünü alanların satın aldığı ve görüntülediği diğer ürünler ve kullanıcıların ürün puanlamalarıdır. Bu noktada bahsedilen kitap bilgilerinin kitap kategorisini belirlemedeki etkisi belirlenmek istenilmiş ve bu etkiler neticesinde kullanıcılara doğru kategoride ürünler tavsiye etmek amaçlanmıştır.

Çalışmada karar ağacı yönteminin kullanılmasının nedeni, bulanık modelde kullanılacak kuralların böylesine karmaşık verilerde belirlenmesinin zor olmasıdır. Karar ağacı yönteminin zor ve karmaşık değişkenlerin olduğu modellerde belirleyici bir etken olduğu ve bulanık modellerin bu tür öneri sistemlerinde doğru öneriler sunabilmeye elverişli olduğu kanıtlanmaya çalışılmıştır.

Günümüzde kullanıcıların e-ticaret sitelerine ve dijital platformlara yöneliminin artmasıyla beraber bu tür satışlarda kullanıcıya doğru ürünü önermek büyük önem

kazanmıştır. Amazon, Netflix, Spotify, Youtube gibi platformlarda ürün ve içeriklerin kullanıcılara belli koşullarda önerildiği görülmektedir. Kullanıcının doğru ürün ve içeriklere yönlendirilmesi bu tür platformların satışlarının artmasına yol açmaktadır. Bu nedenle öneri sistemlerinin geliştirilmesi önem kazanmaktadır.



BÖLÜM 2. LİTERATÜR ARAŞTIRMASI

Alharthi ve ark. 2017 yılında kitap öneri sistemleri hakkında bir literatür çalışması yapmış ve kullanılan öneri sistemleri yöntemleri ile kullanılan veri setlerini araştırmıştır. Çalışmada yöntem olarak işbirlikçi filtreleme, içerik tabanlı öneri sistemleri ön plana çıkmıştır. Kitap öneri sistemi hakkındaki çalışmalarda ön plana çıkan veri setleri olarak, Book-Crossing veriseti, LitRec, LibraryThing, Amazon reviews, INEX Book Track, Project Gutenberg ve Goodreads hakkında bilgilere yer verilmiştir [57].

Günümüzde e-ticaret alanında öneri sistemlerinin geliştirilmesi için çalışmalar yapılmaktadır. Zhou ve ark. 2018 yılındaki çalışmalarında öneri sistemlerini mikro davranışlar üzerinden incelemiştir. E-ticaret verileri üzerinde yapılan deneysel sonuçlarla mikro davranışların önerilerdeki önemini ortaya çıkarmıştır [58].

Choi ve Kim 2014 yılında yaptıkları çalışmada e-ticaret sitelerinde tekrarlanan satın alma işlemleri için öneri sistemi geliştirmişlerdir. Kullanıcıların satın alma tekrarları verileri geliştirilen öneri sisteminde kullanılmıştır. Yaklaşık 4000 ürün sayılı ve 5000 kullanıcı sayılı veri setine uygulanan öneri sisteminde, yapılan performans analizleri sonucunda kullanıcı tabanlı filtreleme tekniğinin ürün tabanlı filtreleme tekniğinden daha iyi performans verdiği gözlemlenmiştir [59].

2013 yılında Li ve ark. tercih benzerliği, güven analizi modülü ve sosyal ilişki analizi modülünü birleştirerek e-ticaret için kişiselleştirilmiş ürün öneri modülü geliştirmişlerdir. Önerdikleri ürün öneri mekanizması e-ticaret perakendecilerinin hizmet kalitelerini artırmasına, müşteri ilişkilerini geliştirme ve ürün tanıtımının başarılı olmasına yol açtığını belirtmişlerdir [60].

Lian 2011 yılında e-ticarette kişiselleştirilmiş web sayfası üzerine bir çalışma gerçekleştirmiştir. Çalışmada kullanıcıların etkileşimleri, arama verileri, sunucu verileri ve kullanıcıların yapmış olduğu değerlendirmelerin verileri alınarak kümeleme algoritması ile değerlendirmeler yapılmıştır. Öneri sistemi yöntemi olarak işbirlikçi filtreleme tekniği kullanılmıştır [61].

2013 yılında Christidis ve Mentzas yaptıkları çalışmada elektronik ürün satışı yapılan platformlar için bir öneri sistemi geliştirmişlerdir. Geliştirilen öneri sisteminde kosinüs benzerliği yöntemi ile ürünler arasındaki benzerlikler belirlenmiştir. Yapılan değerlendirme analizleri sonucunda geliştirilen yöntemin elektronik ürün satışı yapılan platformlar için kararlı ve yararlı bir öneri sistemi olduğu belirtilmiştir [62].

2020 yılında Sadıkoğlu ve Okkalıoğlu çok kriterli ürün tabanlı işbirlikçi filtreleme tekniğinde ağırlıklandırma yöntemlerini kullanarak tahmin performansını artırmaya yönelik bir çalışma yapmıştır. Çalışmada iki kullanıcı arasındaki benzerlik hesaplamasında kullanılan önem-ağırlıklandırma yönteminin, iki ürün arasındaki benzerlik hesaplamasında kullanılması önerilmiştir. Yapılan deneyler sonucu önerilen metotların diğer mevcut metotlara göre tahmin performansının ve kapsam değerlerinin büyük oranda arttığı belirtilmiştir [63].

Öneri sistemleri günümüzde farklı sektörlerde de uygulanmaya başlanmıştır. 2018 yılında Hou ve ark. işbirlikçi filtrelemeye dayalı güvenilir medikal öneriler sunmuştur. Çalışmada çıkan sonuçlar yöntemin kararlı önerilere ulaştığını ve aynı zamanda da gizliliği sağladığı görülmüştür [64].

2010 yılında Moreno ve ark. kullanıcılar için TV programları önerileri sağlayan bir sistem sunmuştur. Metot olarak içerik filtreleme tekniği ve işbirlikçi filtrelemeyi birleştiren hibrit bir yaklaşım sunulmuştur. Çalışmada belirlenen amaç kullanıcıların izleyeceği programı seçmesini olabildiğince basite indirmektedir [65].

Karar ağacı modelleri öneri sistemi uygulamalarında hedef kullanıcıyı seçmeye olanak sağlar. Cho ve ark. yaptıkları çalışmada tavsiyelerin kalitesini artırmayı amaçlayan bir

öneri sistemi metodolojisi sunmuşlardır. Kullanıcıların meslek bilgisi, yaşı, evlilik durumu gibi bilgiler bir karar ağacında modellenerek kullanıcının satın almaya yatkınlığını belirlemişlerdir. Çalışmada ürün almaya yakın kullanıcıların tespiti için karar ağacı indüksiyonu kullanılmıştır. Böylece müşterileri hayal kırıklığına uğratabilecek yanlış önerilerin üstesinden gelinmiştir [66].

Öneri sistemleri tez çalışmalarına da konu olmuştur. 2018 yılında Ahmet Ayan tez çalışmasında e-ticaret siteleri için bir öneri sistemi yaklaşımı sunmuştur. Çalışmasında bulut bilişim platformunda yapay öğrenme algoritmaları kullanmıştır. İçerik tabanlı öneri sistemi, işbirlikçi filtreleme yöntemi ve hibrit öneri sistemi kullanarak bir öneri sistemi modeli gerçekleştirmiştir. Sunulan modeli e-ticaret siteleri için bir web servis ara yüzüne dönüştürmüştür [67].

2019 yılında Enes Kantepe tez çalışmasında otomatik kodlayıcılar kullanarak bir ürün öneri sistemi oluşturmuştur. MovieLens 1M veri setinin kullanıldığı çalışmada modelin eğitim süresinin azalması için en uygun optimizasyon algoritması bulunmaya çalışılmıştır [68].

2015 yılında Elif Şafak Sivri tez çalışmasında e-ticaret siteleri için birliktelik analizi uygulamalı bir çalışma sunmuştur. Kullanıcılar arasındaki benzerliği hesaplamak için Cosine, Pearson ve Jaccard benzerlik ölçütlerini kullanmıştır. Müşterileri kümeleme analizi kullanarak yaş ve cinsiyetlerine göre kümelemiştir. Weka programında verilerde birliktelik kuralları oluşturmuş ve benzerlik ile kümeleme analizleri yapmıştır [69].

2019 yılında Waleed Abdullah tez çalışmasında e-ticaret için ürün tavsiye sistemi geliştirmiştir. 1023 ürün ve yarım milyon satış bilgilerini içeren veri setinde işbirlikçi filtreleme, Apriori ve FP-Büyüme algoritması ve sık örüntü madenciliği üzerine değerlendirmeler yapılmıştır. Çalışmada benzerlik hesaplama yöntemlerinden Kosinüs, Pearson ve Jaccard kullanılmıştır. Sonuç olarak Jaccard benzerliği ve FP-Büyüme algoritmasının en iyi performansı verdiği belirtilmiştir [70].

2014 yılında Hasan Bulut ve Musa Milli işbirlikçi filtreleme ile ilgili son adım olan tahmin hesaplama işlemleri için yeni bir yaklaşım önermiştir. Önerdiği yaklaşımı iki farklı yöntem olarak değerlendirmiştir. Birinci yöntem olarak tahmin hesaplamalarında aktif kullanıcı ile komşuları arasında yapılan pozitif benzerlik yönteminin tam tersini önermiştir. Yani aktif kullanıcı ile negatif benzerliğe sahip komşuları üzerinden hesaplamalar yapmayı önermiştir. İkinci yöntemde ise kullanıcıların ürün oylamalarında verdikleri değerlerin standart sapmalarını da hesaplamaya katan bir yaklaşım sunmuştur [71].

2013 yılında Zhang ve ark. mobil ürünleri ve hizmetleri için ve bulanık mantık tabanlı kişiselleştirilmiş hibrit bir yaklaşım sunmuşlardır. Çalışmada kullanıcı ve ürün tabanlı filtreleme teknikleri ile bulanık kümeleme teknikleri kullanılmıştır. Önerilerdeki doğruluğu ve performansı artırmak için verilerdeki belirsizlikleri azaltmak için çalışılmıştır. Fuzzy-based Telecom Product Recommender System (FTCP RS) yazılımı üzerinden yapılan çalışma sonucunda kullanıcılara sunulan önerilerin daha iyi sonuçlar verdiği belirtilmiştir [72].

BÖLÜM 3. MATERYAL VE YÖNTEM

3.1. Veri Seti

Çalışmada e-ticaret sitelerinden kitap almak isteyen kullanıcılar için karar ağacı tabanlı bulanık mantık kullanan bir öneri sistemi geliştirmek istenmektedir. Bu nedenle içerisinde kitaplara ait bilgiler bulunan veri setine ihtiyaç vardır. Bulanık model oluşturulurken çok sayıda değişkene ihtiyaç olduğundan değişken sayısı önemlidir. Bu nedenle iki farklı veri seti çalışmada kullanılmıştır.

Bu tür çalışmalarda kullanılmak üzere; Book-Crossing veriseti, LitRec, LibraryThing, Amazon reviews, INEX Book Track, Project Gutenberg ve Goodreads ön plana çıkmaktadır. Bu çalışmada da Amazon Review ve Goodreads veri setlerindeki bilgiler birleştirilerek kullanılmıştır.

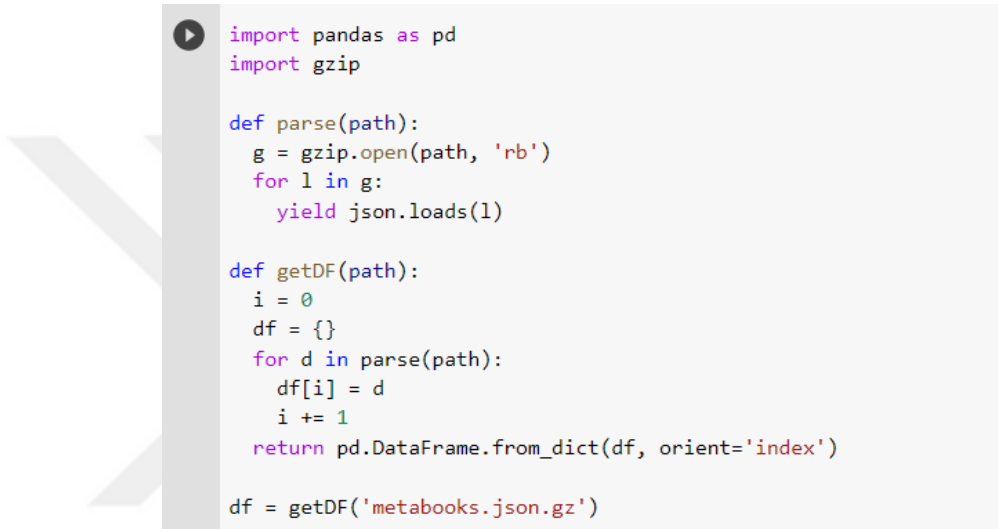


Şekil 3.1. Çalışmada kullanılan veri setleri

3.1.1. Amazon reviews veri seti

Çalışmada kullanılan amazon reviews veri seti 1996-2018 yılları arasında 233,1 milyon kullanıcı geri dönüşü içeren içerisinde ürünlere verilen puanlar, ürün kategorisi bilgileri, fiyat, marka, birlikte alınan-görüntülenen ürünler gibi bilgileri içeren bir veri setidir. Veri setinde amazon sitesinde bulunan ürün kategorilerine ait veriler

bulunmaktadır. Bu veri setinde modelde kullanılacak fiyat, puan ve ürünü alanların aldığı ve görüntülediği diğer ürünlere ait veriler alınmıştır. Modelde kullanılacak değişkenlerden biri olan kitaplara ait sayfa sayısı bilgileri ise bu veri setinde bulunmamaktadır. Bu nedenle GoodReads veri seti de çalışmada kullanılmıştır. Modelde kullanılmak istenen veri setleri JSON dosyası olarak bulunmaktadır. Google Colab ortamında Python programlama dili ile kullanılacak verilerin bulunduğu JSON dosyaları CSV dosyalarına dönüştürülmüştür.



```
import pandas as pd
import gzip

def parse(path):
    g = gzip.open(path, 'rb')
    for l in g:
        yield json.loads(l)

def getDF(path):
    i = 0
    df = {}
    for d in parse(path):
        df[i] = d
        i += 1
    return pd.DataFrame.from_dict(df, orient='index')

df = getDF('metabooks.json.gz')
```

Şekil 3.2. JSON dosyasının dataframe veri yapısına aktarılması

JSON veri depolanmasında kullanılan bir yapıdır. Çalışmada veri analizlerinin Python yazılım dilinin pandas kütüphanesinde yapılması için dataframe veri yapısına dönüştürülmesi gerekmektedir.

Dataframe pandas kütüphanesi içerisinde hızlı ve doğru veri analizleri yapmaya olanak sağlayan çok boyutlu veri yapılarıdır. Ayrıca dataframe içerisinde tutulan veriler yine Python ortamında oldukça kolay bir şekilde “to_csv()” fonksiyonu ile CSV dosyalarına aktarılabilir. Böylelikle çalışmanın sonraki kısımlarında veri ağaçları için kullanılacak Rapidminer ortamında CSV dosyaları entegre olabilecektir.

Amazon reviews veri seti kullanıcı puanlamasını içeren veri seti ve ürün bilgilerini içeren veri seti olarak iki farklı şekilde bulunmaktadır. Ürün bilgilerini içeren veri setinde kitaplara ait Asin, isim, kategori, ürünü alanların aldığı diğer kitaplar, ürünü

alanların görüntülediği diğer kitaplar ve fiyat bilgilerini içermektedir. Bu veri setine ait örnek ürünler Tablo 3.1.'de gösterilmiştir.

Tablo 3.1. Amazon Reviews metadata veri seti

Asin	İsim	Kategori	Ürünü alanların aldığı diğer kitaplar	Ürünü alanların görüntülediği diğer kitaplar	Fiyat
0001061240	The golden treasury of poetry	Children's Books	['0140556125', '0307165221',...]	['B001HDIKYW', 'B000LTQDP6',...]	\$23,98
0002154919	Ignatius Loyola	Christian Books & Bibles	['0140433856', '0898708109',...]	['0898708109', '1945402008']	\$6,96
0002154463	Italy, The Beautiful Cookbook...	Food & Wine	['0002553708', '0002154129',...]	['0681152672', '0002553708',...]	\$79,90
0002005883	Gilead	Literature & Fiction	['0374299102', '0374187614'...]	['1250074843', '0312428545',...]	\$60,18

Amazon Reviews veri setine ait kullanıcıların kitap ürünlerini puanlama verilerini içeren veri setinden örnekler Tablo 3.2.'de gösterilmiştir.

Tablo 3.2. Amazon Reviews kullanıcı ürün puanlaması veri seti

Kullanıcı	Asin	Puan
AH2L9G3DQHHAJ	0000000116	4
A2IIIDRK3PRRZY	0000000116	1
A1TADCM7YWPQ8M	0000000868	4
AWGH7V0BDOJKB	0000013714	4

3.1.2. GoodReads veri seti

Bu veri seti 2,3 milyon kitaba ait bilgiler içermektedir. GoodReads veri seti içerisinde çalışmada kullanılmak üzere yalnızca sayfa sayısı bilgileri alınmıştır. İki ayrı veri setini birleştirmek için Asin bilgileri kullanılmıştır. Böylece çalışmada gerekli olan veri seti oluşturulmuştur. GoodReads veri setinden örnekler Tablo 3.3.'te verilmiştir.

Tablo 3.3. GoodReads veri setindeki sayfa sayısı bilgileri

İsim	Asin	Sayfa Sayısı
The Simple Truths of Service: Inspired by Johnny the Bagger	0000230022	96
Chess for Children	0001061100	128
The Prophet	000100039X	536
Sailing	0001001787	336

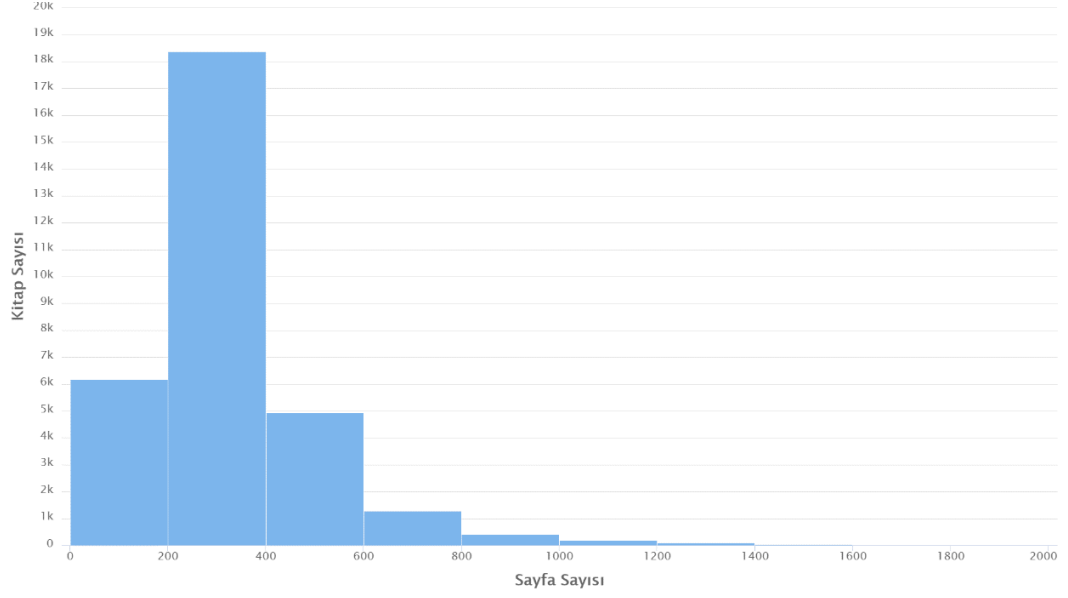
3.1.3. Veri setinin hazırlanması

Veri setinde kitaplara ait puan bilgileri her kitap için ayrı olarak bulunmaktadır. Fakat çalışmada kitaplara ait ortalama puanlar kullanılmak istenmektedir. Bu nedenle aynı kitaplara ait kullanıcı puanlarının ortalaması alınmış ve ortalama puan bilgileri elde edilmiştir.

Kitaplara ait bilgiler içerisinde eksik veriler bulunmaktadır. Modelde doğruluk kaybı olmaması için eksik veriler her değişken için farklı yöntemlerle doldurulmuştur. Örneğin puan değişkeni eksik olan kitaplar için kitabın bulunduğu kategoriye ait kitapların puanlarının ortalaması alınmış ve eksik veri doldurulmuştur. Aynı işlem tüm değişkenler için uygulanmıştır. Veri setinin hazırlanması için yapılan işlemler Python kütüphanesi olan pandas ile gerçekleştirilmiştir [73].

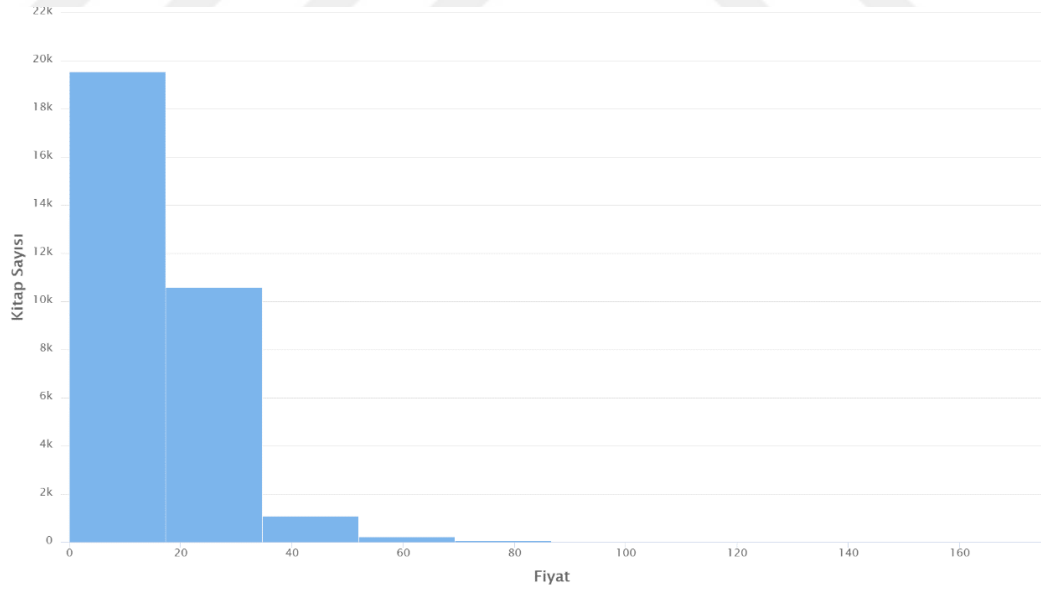
Bulanık modelde değişkenlerin üyelik fonksiyonları oluşturulurken değişkenlerin hangi aralıklarda çoğunlukta olduğunu bilmek gerekmektedir. Bu nedenle her değişkene ait histogramları hazırlanmıştır. Histogram bir veri kümesinin toplam dağılımının sütun grafikleri ile gösterilebildiği bir yöntemdir. Rapidminer ortamına veri setinin entegre edilmesiyle oluşturulan histogramlar bulanık modeldeki üyelik fonksiyonlarının oluşumunda yorumlanmış ve göz önüne alınmıştır.

Sayfa sayısı değişkeni kitapların kategorisini belirlemede ayırt edici özelliklerden biridir. Veriler incelendiğinde Şekil 3.3.'te olduğu gibi verilerin ağırlıklı olarak 0-600 aralığında olduğu görülmektedir. Bulanık modelde bu değişkene ait üyelik fonksiyonu oluşturulurken minimum, maksimum ve ortalama değerler gözetilmiştir.



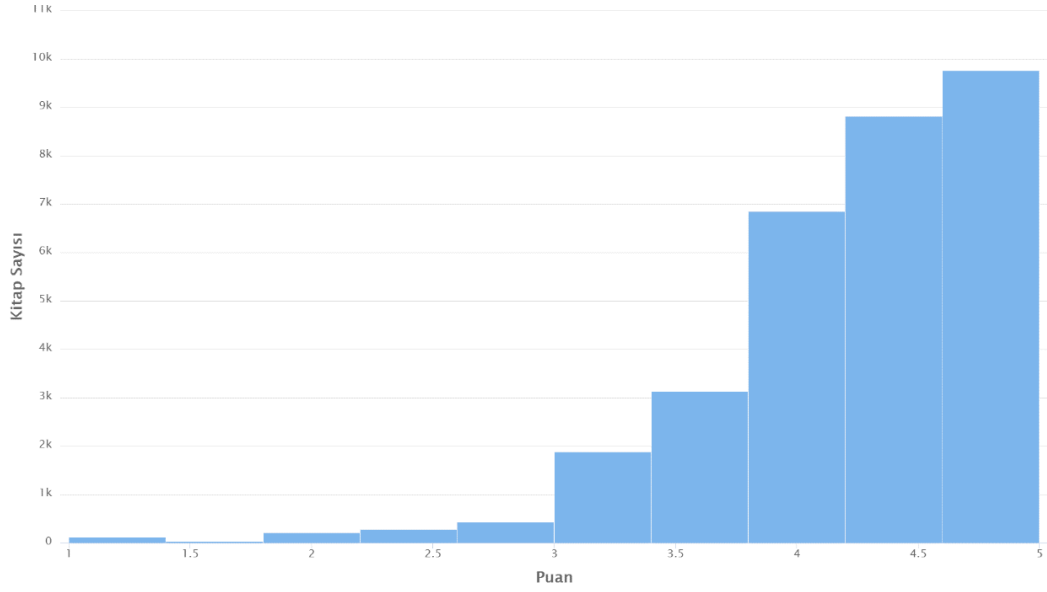
Şekil 3.3. Sayfa sayısı verilerinin histogram grafiği

Kitapların fiyat verileri incelendiğinde yığılmanın 0-20 aralığında olduğu gözlenmiş olup bu değişkene ait üyelik fonksiyonu oluşturulurken bu değerler gözetilmiştir. Fiyat değişkenine ait histogram grafiği Şekil 3.4.'te gösterilmiştir.



Şekil 3.4. Fiyat verilerinin histogram grafiği

Amazon Reviwes veri setindeki kullanıcı puanlamalarının her kitaba ait olanlarının ortalaması alınmış ve ortaya çıkan ortalama puan değerleri Şekil 3.5.'te gösterilmiştir.

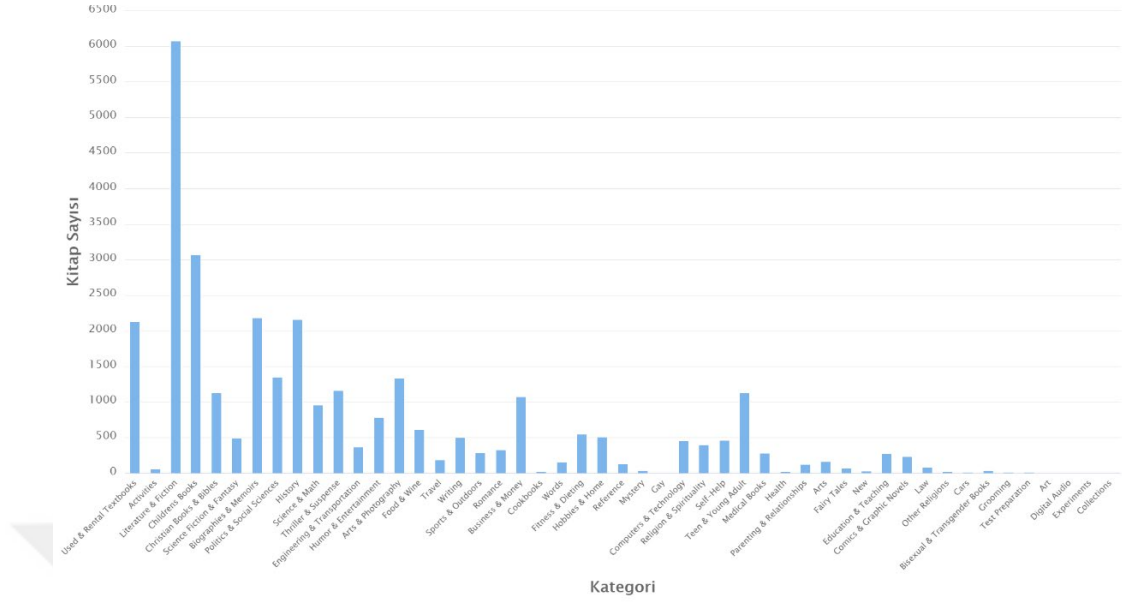


Şekil 3.5. Puan verilerinin histogram grafiği

Tez çalışmasında kullanıcılara önerilmek istenilen kitaplara ait kategoriler veri setinde 48 ayrı kategori şeklinde bulunmaktadır. Karar ağacı yöntemiyle desteklenen bulanık model oluşturulurken her kategoriye ait karar yöntemini belirlemek zor olacaktır.

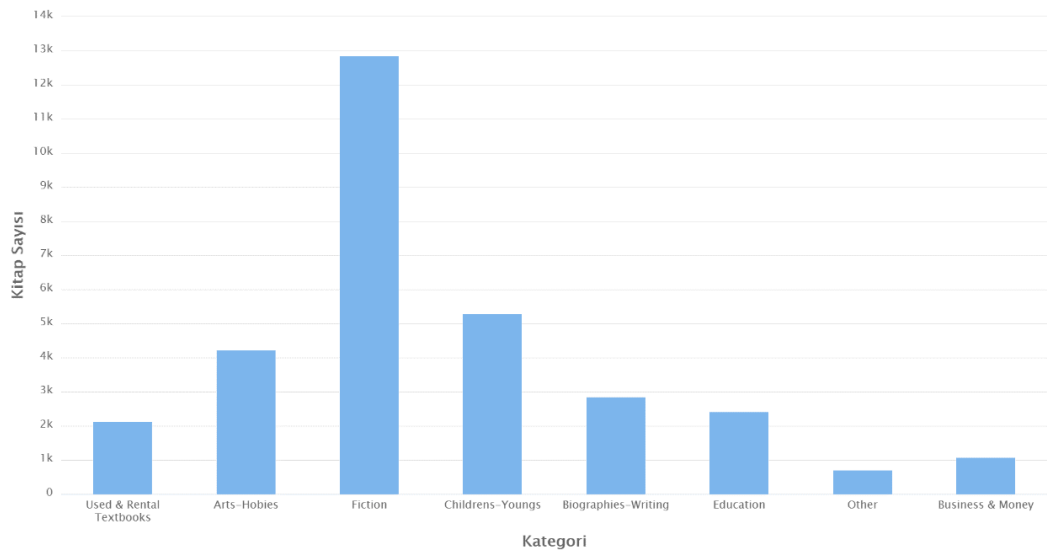
Değişkenlerin çok sayıda farklı kategori için belirleyici olamayacağı düşünülerek veri setindeki benzer kategoriler birleştirilerek Şekil 3.7.’deki ana kategoriler belirlenmiştir. Örneğin Şekil 3.6.’da gösterilmiş olan kategorilerden “Scince-Math”, “Computer-Tech”, “Engineering”, “Education-Teaching” ve “Medical Books” kategorileri “Education” ana kategorisinde toplanmıştır.

Aynı şekilde “Arts-Photo”, “Food-Wine”, “Fitness”, “Hobbies”, “Self-Help”, “Sports-Outdoors”, “Travel”, “Arts”, “Activities”, “Cook Books”, “Health” ve “Art” kategorileri “Arts-Hobbies” ana kategorisinde; “Biographies”, “Writing”, “Words” ve “Write” kategorileri “Biographies-Writing” ana kategorisinde; “Literature-Fiction”, “Science Fiction-Fantasy”, “History”, “Politics”, “Religion”, “Other Religion”, “Christian-Bibles”, “Thriller-Suspense” ve “Mystery” kategorileri “Fiction” ana kategorisinde; “Children”, “Teen-Young”, “Humor-Entertainment” ve “Fairy Tales” kategorileri ise “Children-Young” ana kategorisinde toplanmıştır.



Şekil 3.6. Kategori verilerinin histogram grafiği

Verilerin modele uygun bir şekilde hazırlanmasından sonra 31,563 kitap verisi gerekli Değişkenlerle birlikte elde edilmiştir. Histogram grafiklerinde değişkenlerin genel anlamda parametrik bir düzende olmadığı görülmüş ve bu nedenle karar ağacı modelinin kural çıkarımında doğru bir yöntem olacağı saptanmıştır.

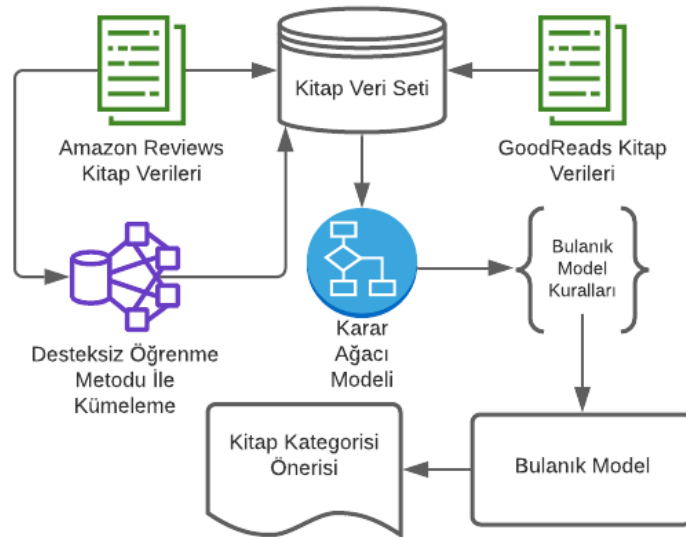


Şekil 3.7. Benzer kategorilerin birleştirilmesiyle oluşan verilerin histogram grafiği

3.2. Geliştirilen Modelin Çalışma Prensibi

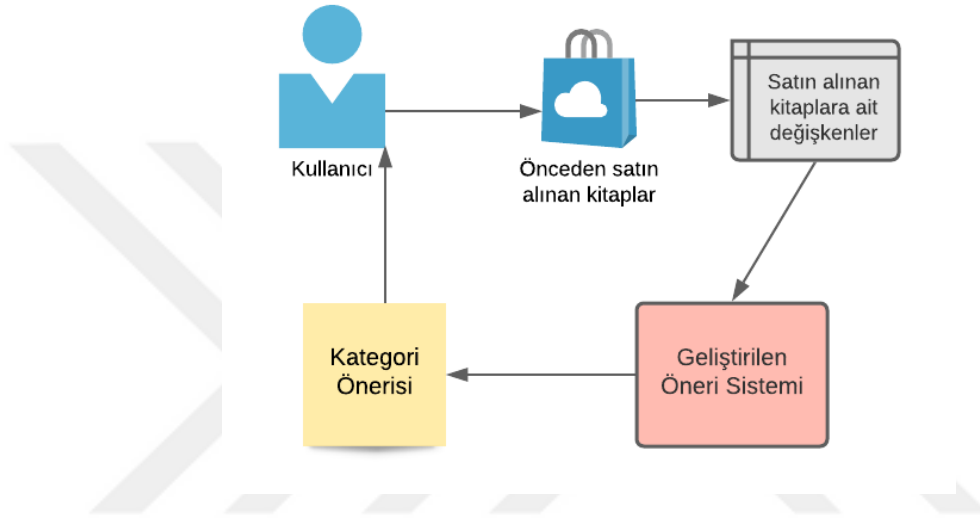
Çalışmada sunulan modelin akış şeması Şekil 3.8.'de gösterilmiştir. Modele ait akış şu adımlardan oluşmaktadır:

- Amazon Reviews veri setindeki bilgilerden modelde kullanılmak üzere değişkenler oluşturulur.
- GoodReads veri setindeki bilgilerden sayfa sayısı değişkeni elde edilir ve iki veri seti aynı Asin değerine sahip kitaplar üzerinden birleştirilir.
- Veri setindeki ürünü alanların satın aldığı ve görüntülediği diğer kitaplar desteksiz öğrenme yöntemi ile kümelendir.
- Değişkenlerin kategori belirlenmesindeki etkisini gözlemlemek için karar ağacı modeli oluşturulur.
- Karar ağacı modeli incelenerek bulanık modelin kuralları oluşturulur.
- Her değişkene ait üyelik fonksiyonları ve kurallar ile bulanıklaştırma yapılır.
- Elde edilen değerler ile durulaştırma yapılır ve kitap kategorisi önerisi kullanıcıya sunulur.



Şekil 3.8. Önerilen yöntemin akış şeması

Normalde veriye ihtiyaç duymayan bulanık modellerde model oluşturulurken değişkenlerin model üzerindeki etkilerinin çok iyi bilinmesi gerekir. Özellikle karmaşık ve zor modellerde değişkenlerin etkilerini öngörmek zor olacaktır. Bu nedenle bulanık model için kullanılan kurallar oluşturulurken karar ağacı yöntemi ile yardım alınmıştır. Çalışmanın bundan sonraki kısımlarında desteksiz öğrenme metodu ile oluşturulan kümeler, karar ağacı modeli ve bulanık model anlatılmıştır.



Şekil 3.9. Kategori önerisinin kullanıcıya sunulmasına ait akış şeması

3.3. Desteksiz Öğrenme ile Kümeleme

E-ticaret sitelerinde kullanıcıya sunulan ürün sayfalarında genellikle her ürüne ilişkin birlikte alınan-görüntülenen ürünler yer alır. Kullanılan veri setinde de her kitaba ait birlikte alınan-görüntülenen kitapların bilgisi yer almaktadır. Önerilen bulanık modelde kitap kategorisini belirleyen bir değişken belirlemek için bu bilgiler kullanılarak desteksiz öğrenme yöntemi ile kitaplar kümelendi.

Belirli sınıf sayıları ve eğitilmiş verilerde destekli öğrenme yöntemi kullanılmaktadır. Kümeleme yöntemleri ile belirli kümeler oluşturmak için ise desteksiz öğrenme kullanılmaktadır. Desteksiz öğrenme uygulanan verilerde ise belirli sınıflar ve eğitilmiş veriler bulunmamaktadır. Bu öğrenme yönteminde verilerin birbiriyle olan

ilişkileri belirlenmeye çalışılır. Böylelikle elde edilen verilerle belirli çıkarımlar yapılmasına olanak sağlar.

Kümeleme algoritmaları elde edilen belirli yöntemlerle sınıflandırmaya ve kümelenecek veriler elde etmeyi sağlamaktadır. Çalışmada kullanılan denetimsiz öğrenme ile yapılan kümelemede k-means algoritması kullanılmıştır. K-means algoritması n adet veriden oluşan bir veri kümesini belirlenen k küme sayısı kadar kümeye bölmeye olanak sağlar. Algoritmanın uygulanması sonucunda veri kümesinde benzerlikleri bulunan verilerin aynı kümede bulunması beklenmektedir.

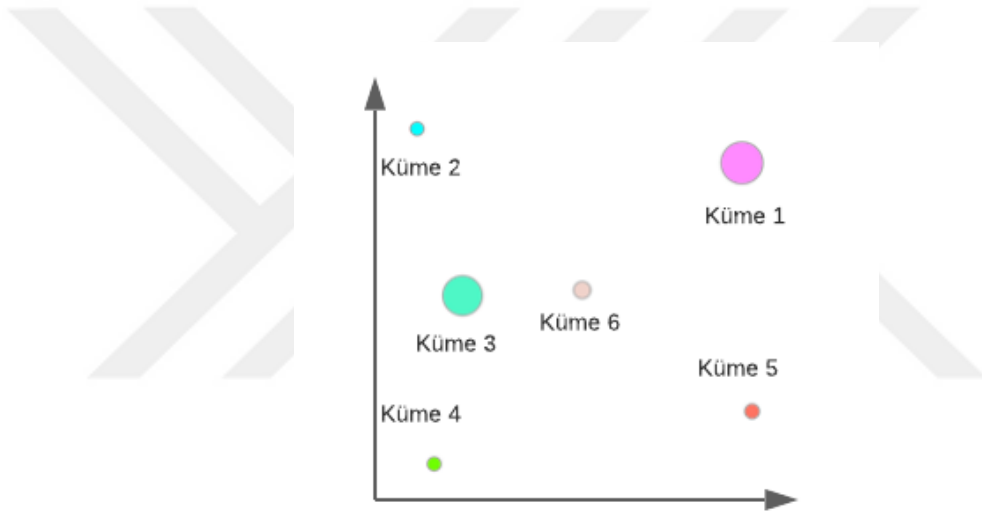
Tez çalışmasında verilerin gözetimsiz öğrenme kullanılarak k-means algoritması ile kümelenecek yöntemi BigML platformunun web arayüzünde gerçekleştirilmiştir. BigML veri analizinde kullanılan bir makine öğrenmesi platformudur. Çeşitli makine öğrenmesi yöntemleriyle kümeleme, sınıflandırma ve regresyonlar oluşturmaya olanak sağlamaktadır. Web arayüzüne yüklenen veri setleri ile veri analizleri yapılabilir. Kullanılan k-means algoritması şu aşamalardan oluşmaktadır;

- Gerekli veriler yüklenir.
- Küme sayısı (k) belirlenir.
- Küme sayısı (k) kadar rastgele noktalar oluşturulur.
- Her nesnenin noktalara olan uzaklıkları hesaplanır.
- Her nesne en yakın olduğu noktanın bulunduğu kümeye dahil edilir.
- Nesnelerin yeri değişmeyene kadar işlem devam edilir ve kümeler oluşturulur.

Veri setindeki birlikte alınan-görüntülenen kitaplar değişkeni içerisinde kitapların Asin bilgilerini içeren listelerden oluşmaktadır. Fakat bu listeler belirli bir şekilde sınıflandırılmamış ve eğitilmemiş ham verilerden oluşmaktadır. Desteksiz öğrenme yöntemi ile ham veriler belirli bir ayırım yapmadan algoritma ile birliktelikleri öğrenilmiş ve kümelenebilir. K-means algoritmalarında optimum k sayısını belirlemek için Elbow yöntemi, ortalama silhouette yöntemi, Gaussian ortalamalar (G-means) algoritması gibi yöntemler bulunmaktadır. Bu çalışmada kullanılan BigML platformunun desteklediği G-means algoritması optimum k sayısını belirlemek için

kullanılmıştır. G-means algoritması az sayıda k-means merkezi ile az sayıda k-means merkezi ile başlayıp her iterasyonda merkezleri bölerek k sayısını artırır ve gauss dağılımından veri gelmeyene kadar merkezlerin bölünme işlemleri devam eder. [74] BigML platformunda uygulanan G-means algoritması sonucunda K-means algoritmasında küme sayısı 6 olarak belirlenmiş ve 6 farklı nokta oluşturulmuştur. Veri setinde içerisinde asin değerleri bulunan listeler her noktaya olan uzaklığına göre kümelerle dahil edilmiş ve nesnelerin yeri değişmeyene kadar işleme devam edilmiştir.

K-means algoritması ile yapılan kümeleme işlemi sonucunda 6 farklı küme elde edilmiştir. Elde edilen kümeler Şekil 3.10.'da gösterilmiştir.



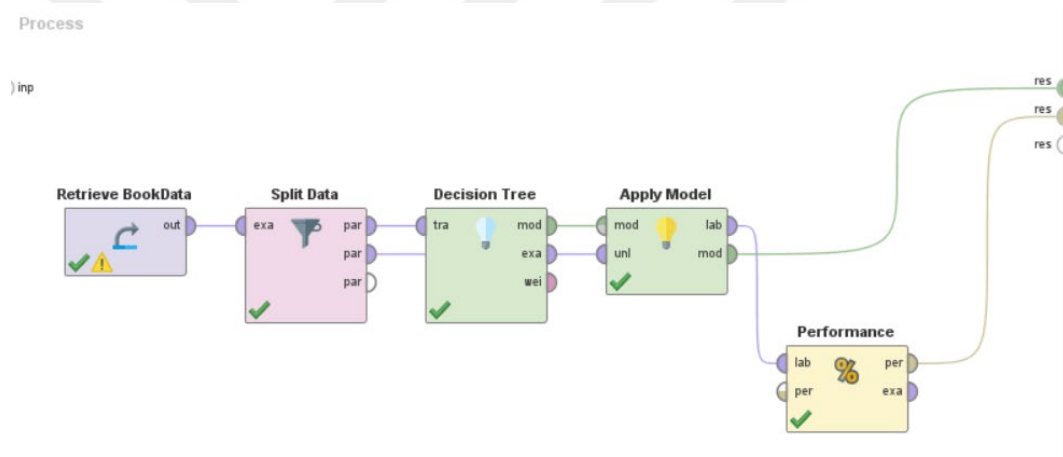
Şekil 3.10. Desteksiz öğrenme ile kümelenmiş veriler

Desteksiz öğrenme yöntemi ile kümelenmiş veriler bulanık modelde kullanılmak üzere birer değişken haline gelmiştir.

3.4. Karar Ağacı Modeli

Karar ağaçlarında yalnızca ikili dallanmaya izin veren metotlar da vardır [75]. Ancak, bulanık modelin kuralları oluşturulmaya çalışıldığında daha fazlasına ihtiyaç duyulmaktadır.

Çalışmada C4.5 algoritmasına dayalı karar ağacı modeli kullanılmıştır. C4.5 algoritması uygulanmasının kolay olması, kolayca yorumlanılabilen modeller oluşturması ve eksik değerlerle başa çıkabilmesi gibi avantajları bulunmaktadır. ID3 algoritmasının bazı eksik ve sorunlarının giderilmiş hali olan C4.5 algoritması özellikle eksik verilerle başa çıkabilme özelliği sebebiyle bu çalışmada tercih edilmiştir. C4.5 algoritmasını ID3 algoritmasından ayıran özelliklerden biri de normalizasyon kullanıyor olmasıdır. C4.5 algoritmasında hedeflenen sınıflar için entropi değeri hesaplanır. Bu çalışmada bu hedef değerleri kitap kategorileri oluşturmaktadır. Karar ağacı modelinde veriler 0,20-0,80 olarak oranlanıp eğitim ve test verilerine ayrılmıştır. Rapidminer ortamında gerçekleştirilen karar ağacı modelinin tasarımı Şekil 3.11.'de gösterilmiştir



Şekil 3.11. Karar ağacı modeli

Karar ağacı modelinin uygulanması için gerekli parametreler Tablo 3.4.'te gösterilmiştir. Tüm niteliklerin entropi değerlerinin hesaplandığı bilgi kazancı yöntemi çok sayıda değere sahip niteliklerde avantaj sağladığı için tercih edilmiştir. Maksimum derinlik değeri ayırmaların maksimum sayıda yapılması ve oluşan karar ağacından kurallar çıkartabilecek düzeyde olması açısından 6 olarak sınırlandırılmıştır. Minimum kazanç değeri bölünmeden önce hesaplanır ve kazanç minimum kazançtan büyükse düğüm bölünür. Daha yüksek minimum kazanç değerleri bölünme sayısını azaltacaktır. Bu nedenle minimum kazanç değeri düşük tutulmuştur.

Tablo 3.4. Karar ağacı modelinin parametreleri

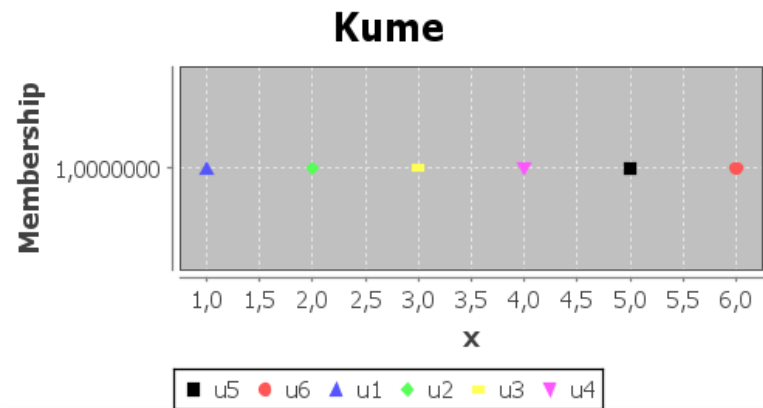
Parametre	Değer
Ayırma kriteri	Bilgi Kazancı
Maksimum derinlik	6
Güven değeri	0,1
Minimum kazanç	0,01
Minimum yaprak sayısı	2

Modelin uygulanması sonucunda elde edilen karar ağacı sayesinde veri setindeki değişkenlerin kitap kategorileri üzerindeki etkisi incelenmiştir. Elde edilen bulgular bulanık modeldeki kuralların belirlenmesini sağlamıştır. Modelde her değişken için oluşan dallar farklı kategoriler için bir kural olarak kullanılmıştır.

3.5. Bulanık Model

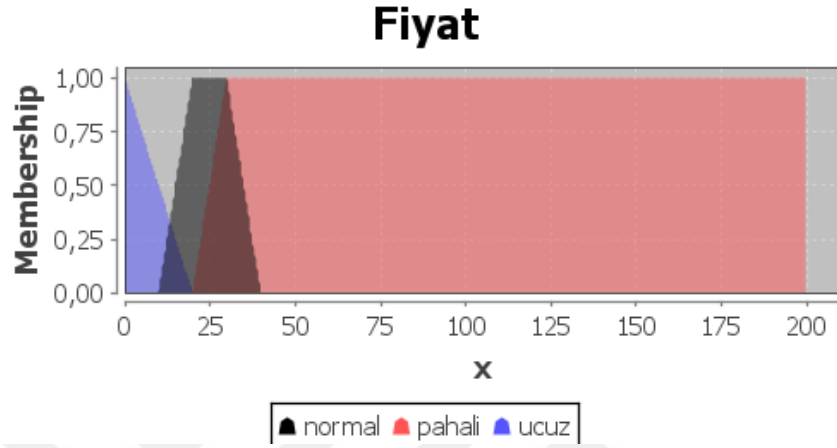
Önerilen metotta veri setinde bulunan değişkenlerin kesin bir karşılığı olması yerine bulanık değerler üzerinden tanımlanması gerçekleştirilmiştir. Veri setindeki puan, fiyat, sayfa sayısı ve desteksiz öğrenme ile kümelenmiş kitaplar için üyelik fonksiyonları oluşturulmuştur. Üyelik fonksiyonları oluşturulurken değişkenlerin minimum, maksimum ve ortalama değerleri göz önüne alınmıştır.

Desteksiz öğrenme ile yapılan kümeleme sonucunda veri setindeki kitaplar 6 farklı kümeye ayrılmıştır. Her küme içerisinde bulunan kitapların türleri gözetilerek kümeler 1'den 6'ya kadar sıralanmış ve üyelik fonksiyonları oluşturulmuştur.



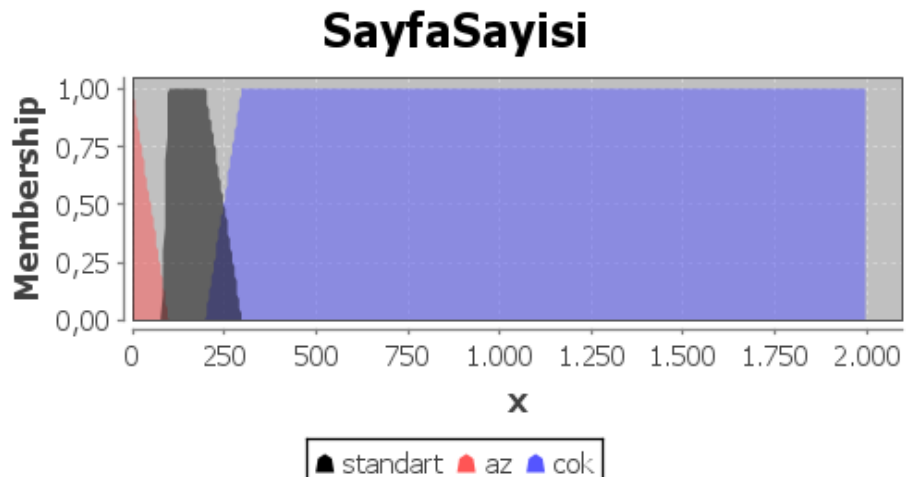
Şekil 3.12. Küme değişkeninin üyelik fonksiyonu

Fiyat değişkeni için oluşturulan üyelik fonksiyonunda kitapların fiyatları ucuz, normal ve pahalı olarak dilsel ifadelerle ayrılmıştır. Alt ve üst limitler belirlendikten sonra ortalama fiyat bilgisi gözetilerek üyelik fonksiyonu belirlenmiştir.



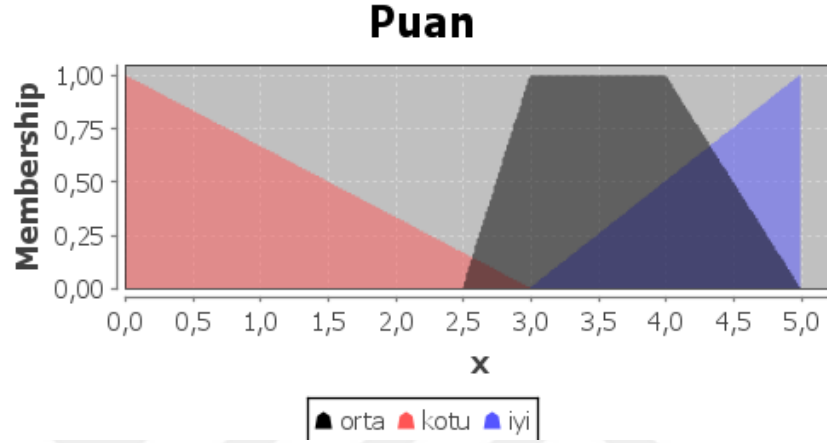
Şekil 3.13. Fiyat değişkeninin üyelik fonksiyonu

Sayfa sayısı değişkeni kitapların kategorilerini ayırt edebileceğimiz değişkenlerden biridir. Veri setinde bulunan kitapların minimum, maksimum ve ortalama değerleri gözetilerek kitaplar az, standart ve çok sayfalı ifadelerine ayrılmış ve üyelik fonksiyonları oluşturulmuştur.



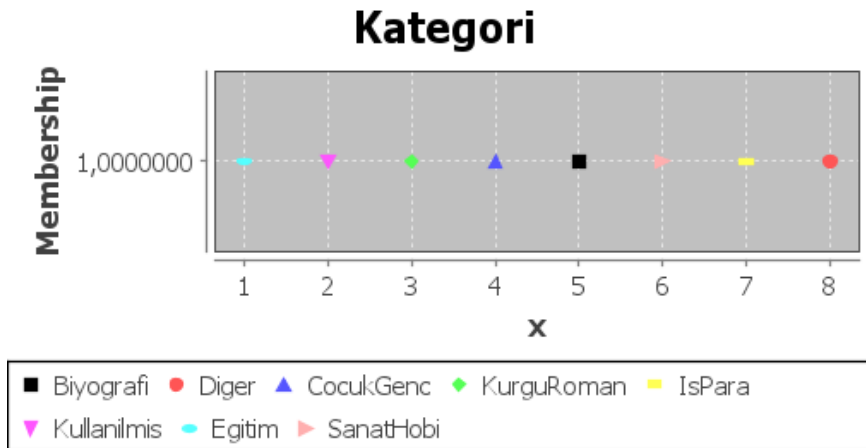
Şekil 3.14. Sayfa sayısı değişkeninin üyelik fonksiyonu

E-ticaret sitelerinde kullanıcılar satın aldıkları ürünleri beğenme derecelerine göre puanlarlar. Kullanılan veri setinde her kitaba ait ayrı ayrı kullanıcıların puanları bulunmaktadır. Bu puanlamaların ortalaması alınıp her kitaba ait ortalama puan bilgisi elde edilmiştir. Elde edilen ortalama puan bilgisi verileri kötü, ortalama ve iyi olarak dilsel ifadelerle ayrılmış ve üyelik fonksiyonları oluşturulmuştur.



Şekil 3.15. Puan değişkeninin üyelik fonksiyonu

Önerilen metotta kullanıcıların önceden tercih ettiği ürünlere ait değişkenler üzerinden bir kitap kategorisi ve o kategoriye ait kitapların önerilmesi amaçlanmaktadır. Veri setinde bulunan kategorilerden birbiriyle ilişkili olanlar birleştirilmiş ve ana kategoriler elde edilmiştir. Bulanık modelin uygulanmasıyla önerilecek kategorilerin üyelik fonksiyonları nümerik olarak üyelik fonksiyonlarda belirtilmiştir.



Şekil 3.16. Kategori değişkeninin üyelik fonksiyonu

Üyelik fonksiyonları ve kuralları içeren bulanık modelin oluşturulmasında jFuzzyLogic kütüphanesi kullanılmıştır. jFuzzyLogic Java ortamında bulanık mantık algoritmaları gerçekleştirmeye olanak sağlayan bir kütüphanedir [12], [13].

FUZZIFY Fiyat

TERM ucuz := (0, 1) (20,0) ;

TERM normal := (10, 0) (20,1) (30,1) (40,0);

TERM pahali := (20, 0) (30, 1) (200, 1);

END_FUZZIFY

FUZZIFY Puan

TERM kotu := (0, 1) (3,0) ;

TERM orta := (2.5,0) (3,1) (4,1) (5,0);

TERM iyi := (3,0) (5,1);

END_FUZZIFY

FUZZIFY SayfaSayisi

TERM az := (0, 1) (100,0) ;

TERM standart := (80,0) (100,1) (200,1) (300,0);

TERM cok := (200,0)(300,1) (2000,1);

END_FUZZIFY

FUZZIFY Kume

TERM u1 := 1 ;

TERM u2 := 2 ;

TERM u3 := 3 ;

TERM u4 := 4 ;

TERM u5 := 5 ;

TERM u6 := 6 ;

END_FUZZIFY

Şekil 3.17. jFuzzyLogic kütüphanesi ile üyelik fonksiyonlarının oluşturulması

Eclipse platformunda jFuzzyLogic kütüphanesi kullanılarak Şekil 3.17.'deki gibi her değişkene ait üyelik fonksiyonu oluşturulmuş ve değişken değerleri bulanıklaştırılmıştır. Bu türdeki karmaşık değişkenlerin olduğu bulanık modellerde jFuzzyLogic kütüphanesinin çok hızlı başarılı sonuçlar verdiği görülmüştür. Kullanılan jFuzzyLogic kütüphanesi ile üçgen, yamuk, gauss, genelleştirilmiş çan, sigmoidal, singleton ve parçalı doğrusal üyelik fonksiyonları oluşturulabilmektedir. Çalışmada sayfa sayısı, fiyat ve puan değişkenleri için yamuk ve küme değişkeni için singleton üyelik fonksiyonları kullanılmıştır.

Bulanık modelin son aşamasında karar ağacından elde edilen bilgilerle kurallar oluşturulmuştur. Bu kurallar dilsel ifadelerle ifade edilmiştir. Bulanık model için belirlenen kurallar Tablo 3.5.'deki gibidir.

Tablo 3.5. Bulanık model kuralları

NUMARA	KURAL
KURAL 1	EĞER (Fiyat normal VEYA Fiyat pahalı) VE (Sayfa Sayısı çok VEYA Sayfa Sayısı standart) O HALDE Kategori Kurgu-Roman
KURAL 2	EĞER Sayfa Sayısı standart VE Küme u3 VE Fiyat Pahalı O HALDE Kategori Sanat-Hobi
KURAL 3	EĞER Sayfa Sayısı standart VE Küme u2 VE (Fiyat ucuz VEYA Fiyat normal) O HALDE Kategori Kurgu-Roman
KURAL 4	EĞER Sayfa Sayısı standart VE Küme u2 VE Puan iyi O HALDE Kategori Çocuk-Genç
KURAL 5	EĞER Sayfa Sayısı standart VE Küme u2 VE (Puan orta VEYA Puan kötü) O HALDE Kategori Kurgu-Roman
KURAL 6	EĞER Sayfa Sayısı standart VE Küme u4 O HALDE Kategori Kurgu-Roman
KURAL 7	EĞER Sayfa Sayısı standart VE Küme u1 VE Puan iyi O HALDE Kategori Sanat-Hobi
KURAL 8	EĞER Sayfa Sayısı standart VE Küme u1 VE (Puan orta VEYA Puan kötü) O HALDE Kategori Kurgu-Roman
KURAL 9	EĞER Sayfa Sayısı az VE Fiyat ucuz O HALDE Kategori Çocuk-Genç
KURAL 10	EĞER Sayfa Sayısı çok O HALDE Kategori Kurgu-Roman
KURAL 11	EĞER Fiyat normal O HALDE Kategori Kullanılmış

Bulanık modelin kurallara karar ağacından elde edilen bulgular neticesinde oluşturulmuştur. Karar ağacı yönteminin hızlı ve doğru sonuçlar vermesi birçok kategori için kurallar oluşturmayı sağlamıştır.

```

DEFUZZIFY Kategori
    TERM KurguRoman := 3;
    TERM SanatHobi := 6;
    TERM CocukGenc := 4;
    TERM Egitim := 1;
    TERM Kullanilmis := 2;
    TERM Biyografi := 5;
    TERM IsPara :=7;
    TERM Diger :=8;
    METHOD : COGS; //'Center Of Gravity'
    DEFAULT := 3;
END_DEFUZZIFY

```

Şekil 3.18. jFuzzyLogic kütüphanesi ile durulaştırma uygulaması

Üyelik fonksiyonları ve bulanık modelin kuralları belirlendikten sonra sonuçları kesin değerlere dönüştürmek için jFuzzyLogic kütüphanesi kullanılarak Center of Gravity metodu ile Şekil 3.18.'de gösterildiği gibi durulaştırma yöntemi uygulanmıştır. Java ortamında gerçekleştirilen yöntem sonucunda kitap kategorilerini içeren veriler elde edilmiştir.

BÖLÜM 4. BULGULAR VE TARTIŞMA

Bu bölümde kullanılan metotlardan elde edilen bulgular tartışılmış ve sonuçlar sunulmuştur.

4.1. Desteksiz Öğrenme ile Kümelenmiş Verilerden Elde Edilen Bulgular

Veri setinde bulunan ürünü alanların satın aldığı ve görüntülediği diğer ürünler bilgileri kullanılarak desteksiz öğrenme ile kümeleme gerçekleştirilmiştir. Yapılan kümeleme işlemi sonucu kitaplar Tablo 4.1.'deki gibi 6 farklı kümeye ayrılmıştır. Veri setindeki bu değişkene ait çok sayıda eksik değer olması kümelere ait kitap sayılarının birbirinden çok farklı olmasına neden olmuştur. Genel olarak verilerin 3 kümede daha çok bulunduğu gözlemlenmiştir. Bununla birlikte kümeleme işlemi sonucu elde edilen küme değişkeni karar ağacından kural çıkarma aşamasında ayırt edici değişkenlerden biri olmuştur.

Tablo 4.1. Kümelenmiş verilerin dağılımı

Küme	Yüzde(%)	Kitap Sayısı
Küme 1	2,5	774
Küme 2	50,7	16003
Küme 3	19,5	6143
Küme 4	26,2	8281
Küme 5	0,9	287
Küme 6	0,2	75
Toplam	100	31563

4.2. Karar Ağacından Elde Edilen Bulgular

Bulanık modelde kuralların oluşturulması için veri setinde bulunan değişkenlerin etkilerinin çok iyi bilinmesi gerekmektedir. Karar ağacı yöntemiyle bu değişkenlerin

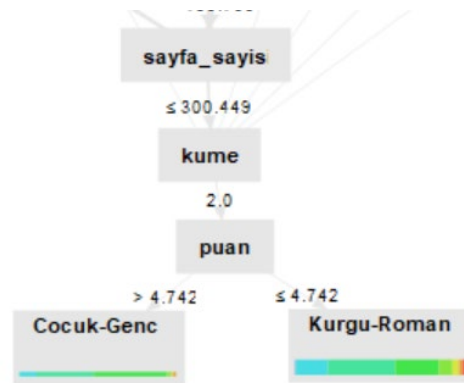
etkileri gözlemlenmiş ve bu tür modellerde karar ağacı yönteminin başarılı sonuçlar verdiği anlaşılmıştır. Karar ağacında her kategorinin ne oranda tahmin edilebilir olduğu Şekil 4.1.'de görülmektedir.

Özellikle “Kurgu-Roman”, “Kullanılmış” ve “Çocuk-genç” kategorisine ait kitaplarda değişkenlerin etkisi çok doğru bir oranda olduğu gözlemlenmiştir. Böylelikle bulanık modelin kuralları oluşturulurken bu kategorilere ait kurallar daha fazladır. “Diğer” kategorisine ait kitaplarda ise değişkenlerin etkisi gözlemlenememiştir.

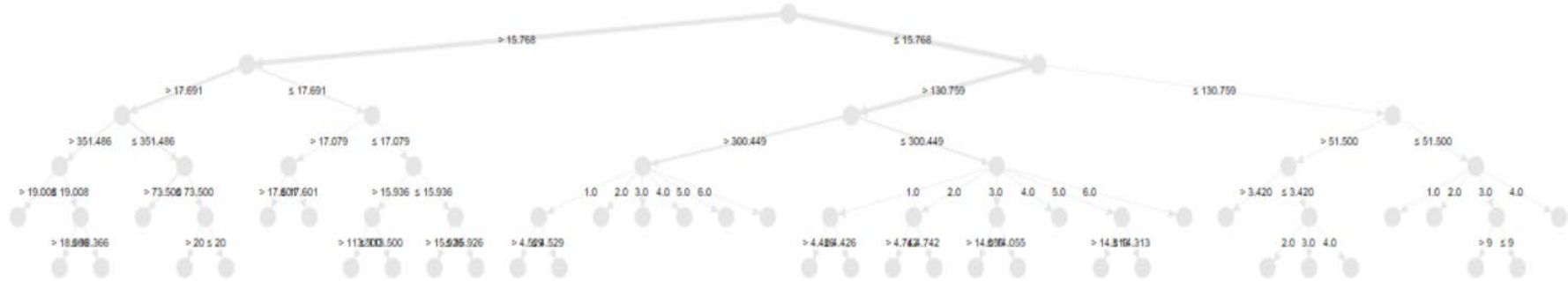
	true Kullanılmış	true Sanat-Hobi	true Kurgu-Ro...	true Çocuk-Ge...	true Biyografi	true Eğitim	true Diğer	true Is-Para	class precision
pred. Kullanıl...	1543	65	353	37	81	100	85	14	67.73%
pred. Sanat-H...	0	54	111	32	27	5	2	19	21.60%
pred. Kurgu-R...	116	2935	9286	2272	1867	1573	452	752	48.23%
pred. Çocuk-G...	1	271	380	1857	53	79	30	16	69.11%
pred. Biyografi	0	10	48	9	221	4	2	0	75.17%
pred. Eğitim	0	25	30	11	11	170	0	8	66.67%
pred. Diğer	0	0	0	0	0	0	0	0	0.00%
pred. Is-Para	44	20	71	17	18	7	2	53	22.84%
class recall	90.55%	1.60%	90.34%	43.85%	9.70%	8.77%	0.00%	6.15%	

Şekil 4.1. Karar ağacının performans ölçümü

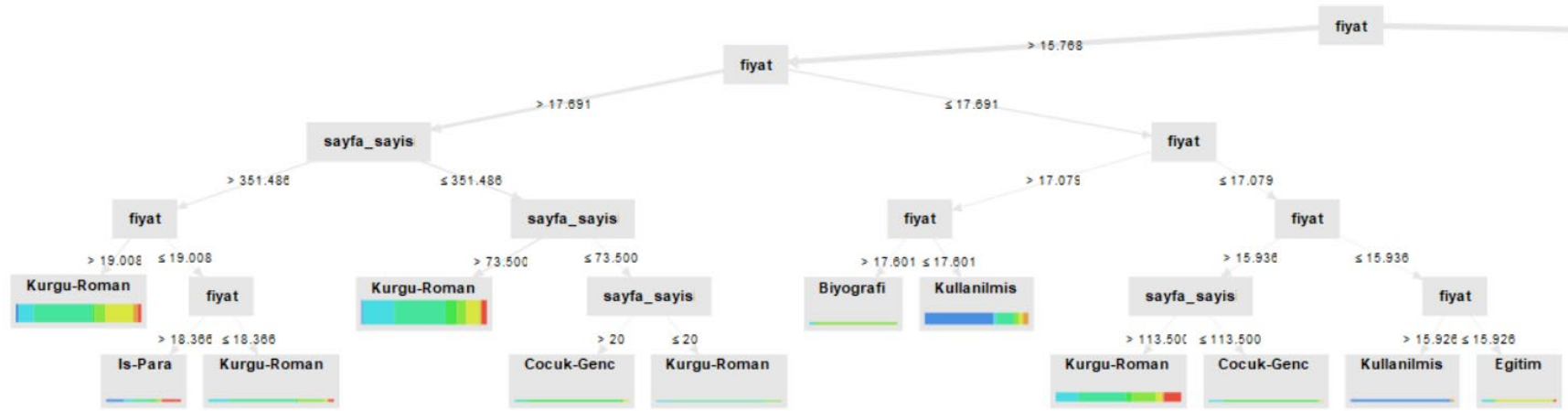
Karar ağacı modelindeki düğümler bulanık modeldeki kuralları oluşturulmasına yardımcı olmuştur. Örneğin Şekil 4.2.'deki gibi düğümde sayfa sayısı, küme ve puan değişkenlerinin etkisi gözlemlenmiş ve yapraklardaki kategorilere ait kurallar ortaya çıkmıştır.



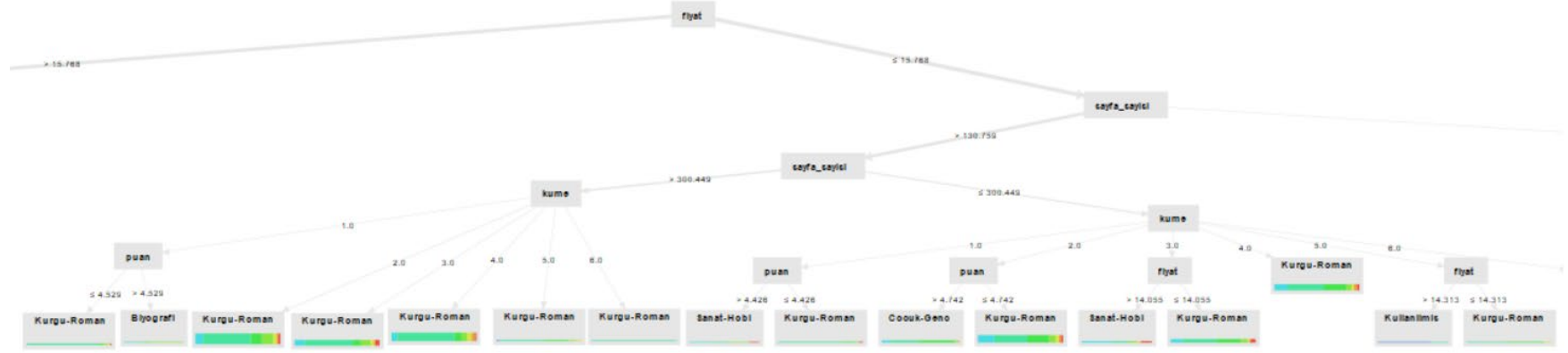
Şekil 4.2. Karar ağacından kural çıkarımı örneği



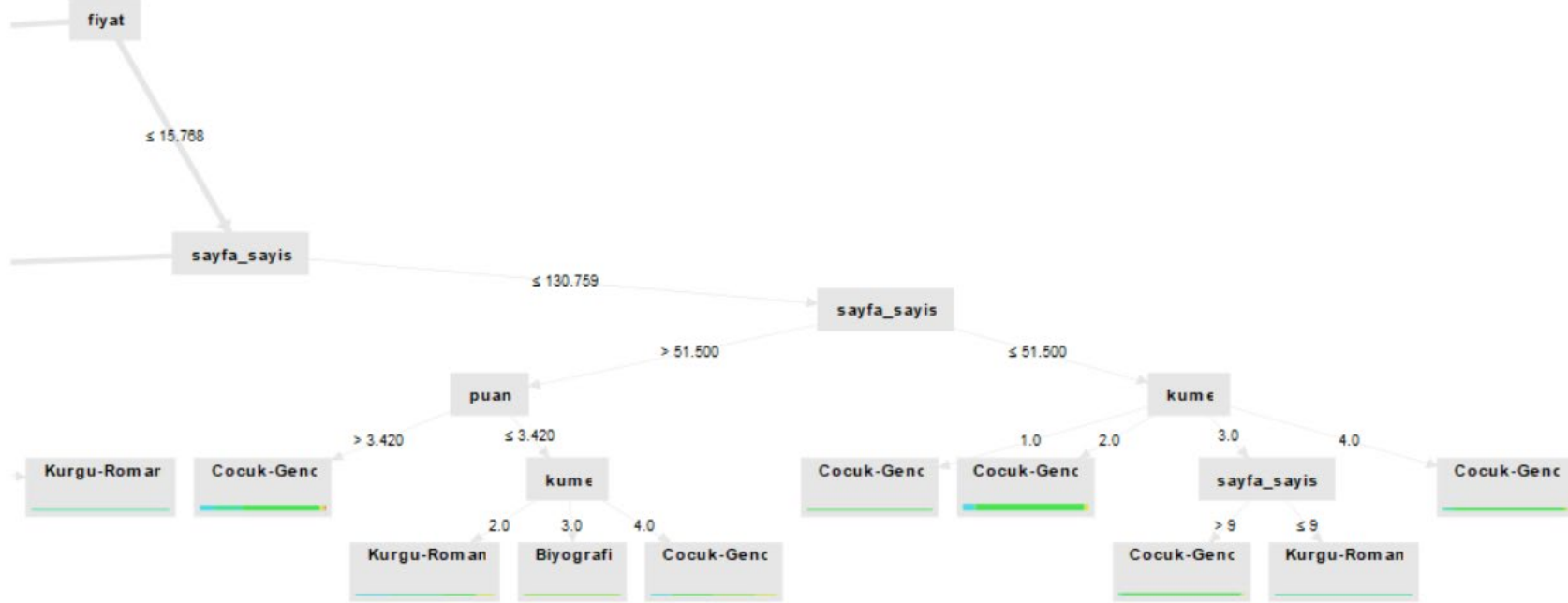
Şekil 4.3. Karar ağacı düğümleri



Şekil 4.4. Karar ağacının sol kısmı



Şekil 4.5. Karar ağacının orta kısmı



Şekil 4.6. Karar ağacının sağ kısmı

4.3. Değerlendirme Metrikleri

Önerilen metodun performansını test etmek amacıyla öneri sistemlerinde sıklıkla kullanılan MAE (Mean Absolute Error), recall, precision ve F1 metrikleri kullanılmıştır. MAE n kadar veride gerçek değerler ile hesaplanan değerler (r_a, r_p) arasındaki ortalama mutlak hatayı hesaplayan bir değerlendirme metriğidir.

$$MAE = \frac{\sum_{i=1}^n |r_a - r_p|}{n} \quad (4.1)$$

Recall tercih edilen değer in önerilen değere oranı, precision ise önerilen değerin tercih edilen değere oranı olarak tanımlanır.

$$recall = \frac{|\{tercih\ edilen\} \cap \{\önerilen\}|}{|\{tercih\ edilen\}|} \quad (4.2)$$

$$precision = \frac{|\{tercih\ edilen\} \cap \{\önerilen\}|}{|\{\önerilen\}|} \quad (4.3)$$

F1 metriği ise recall ve precision metriklerinin harmonik ortalaması olarak tanımlanır.

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \quad (4.4)$$

Değerlendirme metriklerinin modele uygulanması için Python yazılım dilinde bir makine öğrenmesi kütüphanesi olan scikit-learn kütüphanesi kullanılmıştır [76]. Hesaplamalarda hızlı ve doğru bir yöntem olduğu görülmüştür.

Önerilen metottan elde edilen sonuçlar ve gerçek değerlere bu değerlendirme metrikleri uygulandıktan sonra ortaya çıkan sonuçlar Tablo 4.2.'deki gibidir.

Tablo 4.2. Değerlendirme Sonuçları

Değerlendirme Metriği	Sonuç
MAE	1,18347
Recall	0,44450
Precision	0,91480
F1	0,57456

Değerlendirme sonuçları çok sayıda değişken ve veri için oluşturulan bulanık modellerde karar ağacı yönteminin kuralları hızlı ve doğru bir şekilde oluşturulabileceğini göstermiştir.

BÖLÜM 5. SONUÇLAR VE ÖNERİLER

Öneri sistemleri özellikle e-ticaret siteleri ve dijital içerik sunan platformlarda kullanıcılara doğru ürün önerisinde bulunmayı amaçlayan yöntemlerdir. Yapılan tez çalışmasında e-ticaret sitelerinden kitap satın alan kullanıcılara yönelik kitap kategorisi önerisi sunan bir öneri sistemi modeli tasarlanmıştır. Veri setlerinden elde edilen verilere ait çok sayıda değişken olması bulanık model tabanlı bir uygulama yapmaya olanak sağlamıştır. Çalışmada kümeleme yapmak için k-means algoritması kullanan denetimsiz öğrenme yöntemi, değişkenlerin kitap kategorisini belirlemeye etkisini öğrenmek için karar ağacı yöntemi, giriş değişkenleri kullanılarak oluşturulan kurallar ile öneri üretmek için bulanık mantık yöntemi kullanılmıştır.

Veri seti olarak kullanılan Amazon Reviews ve GoodReads kitaplarla ilgili çok sayıda değişken ve veri içeren veri setleridir. Bu gibi modeller ile oluşturulan öneri sistemlerinde veri analizin ayrıntılı olarak yapılması için büyük verilere ihtiyaç vardır. Çok sayıda kullanıcı değerlendirmesi ve kitap bilgisi bulunan veri setleri kapsamlı bir model oluşturmaya olanak sağlamıştır.

Sayfa sayısı ve fiyat değişkenlerinin yanı sıra kullanıcı puanlarını içeren veriler ortalamaları alınarak ortalama puan bilgilerine dönüştürülmüştür. Veri setinde bulunan ürünü alanların satın aldığı ve görüntülediği diğer ürünler bilgileri ise desteksiz öğrenme metoduyla k-means algoritması kullanılarak kümelenebilir. Kümeleme işlemi ile kitapların benzerliklerine göre ayrılması amaçlanmıştır. Böylelikle modelde kullanılabilecek bir değişken daha elde edilmiştir. 6 farklı kümeye ayrılan veriler karar ağacı modelinde kural oluşturmak için etken bir değişken olmuştur. Değişken sayısı arttıkça karar ağacından elde edilen bulgular ile belirlenen bulanık model kurallarının daha kapsamlı olacağı görülmüştür.

Normalde veriye ihtiyaç duyulmayan bulanık modeller için belirlenen değişkenler çok sayıda ve karmaşık olduğu için belirli karar verme yöntemleri kullanmak gerekir. Bu çalışmada karar ağacı yönteminin bulanık modellerin kurallarının belirlenmesi için ne kadar etkili bir yöntem olduğu anlaşılmıştır. Karmaşık veriler ve çok sayıda değişkenler ile geliştirilen modellerde kural çıkarmak zor olacaktır. Fakat karar ağacı modeli ile bu zorluğun üstesinden gelinmiş, hızlı ve doğru bir şekilde kurallar belirlenmiştir.

Öneri sistemlerinde bulanık modellerin kullanımının doğru öneriler sunmak için etkili bir yöntem olduğu görülmüştür. Veri setlerinden elde edilen değişkenlerin üyelik fonksiyonları oluşturulan histogramlar dahilinde oluşturulmuştur. Karar ağacında elde edilen bulgular ile belirlenen kurallar ile bulanıklaştırılan veri kümeleri sık kullanılan bir durulaştırma yöntemi olan COG ile durulaştırılmış ve kesin değerlere dönüşen kitap kategorisi önerileri elde edilmiştir. Çalışmada kullanılan jFuzzyLogic kütüphanesinin bulanık modellerin tasarlanmasında hızlı ve etkili bir yöntem sunduğu görülmüştür.

Çalışmada kullanılan puan değişkeninin kural oluşturmada diğer değişkenlere göre daha az etkisinin olduğu anlaşılmıştır. Bunun nedeni puanların kitaplar arasında belirleyici olmasına rağmen tüm kategoriye etki etmemesidir. Diğer değişkenler ise kategori belirlemek için etkin bir şekilde kullanılmıştır. Özellikle sayfa sayısı değişkeni karar ağacı modelinden çıkarılan kurallar için en belirleyici değişken olmuştur. Kümeleme işlemi sonucunda oluşturulan 6 farklı küme ile elde edilen değişkenin ve fiyat değişkeninin de kitap kategorisini belirlemek için etkili değişkenler olduğu görülmüştür.

Karar ağacı modeli için belirlenen değişkenler arttıkça daha net kurallar belirlenebilecektir. Çalışmada puan, fiyat, ürünü alanların satın aldığı ve görüntülediği diğer ürünler ve sayfa sayısı bilgileri yer almaktadır. Başka çalışmalarda kitabın çıkış tarihi, benzer ürünler bilgisi ve yazım dili gibi değişkenler de eklenebilir. Değişken sayısı arttıkça kategoriye belirleyici kurallar artacak ve model daha doğru sonuçlar verecektir.

Tez çalışmasında kitap kategorisine ait bilgiler kullanılmıştır. Farklı kategorilere ait daha çok bilgiye sahip veri setleri çalışmanın doğruluğunu artıracaktır. Örneğin elektronik ürünler, tekstil, film, müzik gibi ürünlere ait çok sayıda belirleyici değişken ile geliştirilen bir modelde de etkili sonuçlar elde edilebilir.

Öneri doğruluğunu artırmak için sunulan metoda işbirlikçi filtreleme tekniği eklenerek hibrit bir yaklaşım oluşturulabilir. Böylelikle kitap kategorisi önerisiyle beraber o kategoriye ait kitap önerileri de kullanıcıya sunulabilir. Hibrit bir yaklaşım öneri doğruluğunu artırmak için etkili bir yöntem olacaktır.



KAYNAKLAR

- [1] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," in *Mobile Networks and Applications*, 2014, vol. 19, no. 2, doi: 10.1007/s11036-013-0489-0.
- [2] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, 2014, doi: 10.1109/TKDE.2013.109.
- [3] E. Aktan, "Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu," *Bilgi Yönetimi*, vol. 1, no. 1, 2018, doi: 10.33721/by.403010.
- [4] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, 2005, doi: 10.1109/TKDE.2005.99.
- [5] B. Smith and G. Linden, "Two Decades of Recommender Systems at Amazon.com," *IEEE Internet Comput.*, vol. 21, no. 3, 2017, doi: 10.1109/MIC.2017.72.
- [6] C. A. Gomez-Urbe and N. Hunt, "The netflix recommender system: Algorithms, business value, and innovation," *ACM Trans. Manag. Inf. Syst.*, vol. 6, no. 4, 2015, doi: 10.1145/2843948.
- [7] M. Schedl, H. Zamani, C. W. Chen, Y. Deldjoo, and M. Elahi, "Current challenges and visions in music recommender systems research," *Int. J. Multimed. Inf. Retr.*, vol. 7, no. 2, 2018, doi: 10.1007/s13735-018-0154-2.
- [8] S. Qin, R. Menezes, and M. Silaghi, "A recommender system for youtube based on its network of reviewers," 2010, doi: 10.1109/SocialCom.2010.53.
- [9] J. Ni, J. Li, and J. McAuley, "Justifying recommendations using distantly-labeled reviews and fine-grained aspects," 2020, doi: 10.18653/v1/d19-1018.
- [10] M. Wan and J. McAuley, "Item recommendation on monotonic behavior chains," 2018, doi: 10.1145/3240323.3240369.

- [11] M. Wan, R. Misra, N. Nakashole, and J. McAuley, "Fine-grained spoiler detection from large-scale review corpora," 2020, doi: 10.18653/v1/p19-1248.
- [12] P. Cingolani and J. Alcalá-Fdez, "jFuzzyLogic: A robust and flexible Fuzzy-Logic inference system language implementation," 2012, doi: 10.1109/FUZZ-IEEE.2012.6251215.
- [13] P. Cingolani and J. Alcalá-Fdez, "jFuzzyLogic: A Java Library to Design Fuzzy Logic Controllers According to the Standard for Fuzzy Control Programming," *Int. J. Comput. Intell. Syst.*, vol. 6, no. SUPPL1, 2013, doi: 10.1080/18756891.2013.818190.
- [14] Z. Erkin, M. Beye, T. Veugen, and R. L. Lagendijk, "Privacy-preserving content-based recommender system," 2012, doi: 10.1145/2361407.2361420.
- [15] M. N. M. Adnan, M. R. Chowdury, I. Taz, T. Ahmed, and R. M. Rahman, "Content based news recommendation system based on fuzzy logic," 2014, doi: 10.1109/ICIEV.2014.6850800.
- [16] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, "A content-based recommender system for computer science publications," *Knowledge-Based Syst.*, vol. 157, 2018, doi: 10.1016/j.knosys.2018.05.001.
- [17] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," 2001, doi: 10.1145/371920.372071.
- [18] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," *IEEE Internet Comput.*, vol. 7, no. 1, 2003, doi: 10.1109/MIC.2003.1167344.
- [19] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, no. 1. 2004, doi: 10.1145/963770.963772.
- [20] K. Y. Jung, D. H. Park, and J. H. Lee, "Hybrid collaborative filtering and content-based filtering for improved recommender system," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3036, 2004, doi: 10.1007/978-3-540-24685-5_37.
- [21] G. Geetha, M. Safa, C. Fancy, and D. Saranya, "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System," in *Journal of Physics: Conference Series*, 2018, vol. 1000, no. 1, doi: 10.1088/1742-6596/1000/1/012101.

- [22] K. Lang, "NewsWeeder: Learning to Filter Netnews," in *Machine Learning Proceedings 1995*, 1995.
- [23] M. Pazzani and D. Billsus, "Learning and Revising User Profiles: The Identification of Interesting Web Sites," *Mach. Learn.*, vol. 27, no. 3, 1997, doi: 10.1023/A:1007369909943.
- [24] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," *Proc. ACM Int. Conf. Digit. Libr.*, 2000, doi: 10.1145/336597.336662.
- [25] C. C. Aggarwal and C. C. Aggarwal, *Recommender Systems: The Textbook*. 2016.
- [26] S. Ahn and C. K. Shi, "Exploring movie recommendation system using cultural metadata," 2008, doi: 10.1109/CW.2008.13.
- [27] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to Weave an Information tapestry," *Commun. ACM*, vol. 35, no. 12, 1992, doi: 10.1145/138859.138867.
- [28] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," 1994, doi: 10.1145/192844.192905.
- [29] U. Shardanand and P. Maes, "Social information filtering," 1995, doi: 10.1145/223904.223931.
- [30] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, and H. P. Kriegel, "Probabilistic Memory-Based Collaborative Filtering," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 1, 2004, doi: 10.1109/TKDE.2004.1264822.
- [31] M. Phanich, P. Pholkul, and S. Phimoltares, "Food recommendation system using clustering analysis for diabetic patients," 2010, doi: 10.1109/ICISA.2010.5480416.
- [32] S. Parvatikar and B. Joshi, "Online book recommendation system by using collaborative filtering and association mining," 2016, doi: 10.1109/ICCIC.2015.7435717.
- [33] K. Miyahara and M. J. Pazzani, "Collaborative filtering with the Simple Bayesian Classifier," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2000, vol. 1886 LNAI, doi: 10.1007/3-540-44533-1_68.

- [34] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural collaborative filtering," 2017, doi: 10.1145/3038912.3052569.
- [35] M. Balabanović and Y. Shoham, "Content-Based, Collaborative Recommendation," *Commun. ACM*, vol. 40, no. 3, 1997, doi: 10.1145/245108.245124.
- [36] M. J. Pazzani, "Framework for collaborative, content-based and demographic filtering," *Artif. Intell. Rev.*, vol. 13, no. 5, 1999, doi: 10.1023/a:1006544522159.
- [37] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. User-Adapted Interact.*, vol. 12, no. 4, 2002, doi: 10.1023/A:1021240730564.
- [38] T. Tran and R. Cohen, "Hybrid Recommender Systems for Electronic Commerce," *AAAI Tech. Rep.*, 2000.
- [39] J. Ben Schafer, J. Konstan, and J. Riedl, "Recommender systems in e-commerce," 1999, doi: 10.1145/336992.337035.
- [40] "Amazon Türkiye Ürün Satın Alma Sayfası." www.amazon.com.tr, Erişim Tarihi: 20.02.2021.
- [41] L. A. Zadeh, "Fuzzy Sets-Information and Control-1965," *Inf. Control*, vol. 8, 1965.
- [42] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. J. Man. Mach. Stud.*, vol. 7, no. 1, 1975, doi: 10.1016/S0020-7373(75)80002-2.
- [43] D. İsmail and H. Altaş, "Bulanık Mantık : Bulanıklılık Kavramı," *Bilesim yayıncılık A.Ş*, vol. 62, 1999.
- [44] T. J. Ross, *Fuzzy Logic with Engineering Applications: Third Edition*. 2010.
- [45] A. Jain and A. Sharma, "Membership Function Formulation Methods for Fuzzy Logic Systems: a Comprehensive Review," vol. 7, no. 19, p. 2020, 2020.
- [46] A. Talon and C. Curt, "Selection of appropriate defuzzification methods: Application to the assessment of dam performance," *Expert Syst. Appl.*, vol. 70, 2017, doi: 10.1016/j.eswa.2016.09.004.

- [47] S. R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," *IEEE Trans. Syst. Man Cybern.*, vol. 21, no. 3, 1991, doi: 10.1109/21.97458.
- [48] A. Çalış, S. Kayapınar, and T. Çetinyokuş, "Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama," *Endüstri Mühendisliği*, vol. 25, no. 3, 2014.
- [49] M. Umano *et al.*, "Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems," in *IEEE International Conference on Fuzzy Systems*, 1994, vol. 3, doi: 10.1109/fuzzy.1994.343539.
- [50] S. Ruggieri, "Efficient C4.5," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 2, 2002, doi: 10.1109/69.991727.
- [51] M. Mehta, R. Agrawal, and J. Rissanen, "SLIQ: A fast scalable classifier for data mining," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1996, vol. 1057 LNCS, doi: 10.1007/bfb0014141.
- [52] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda, "The CART decision tree for mining data streams," *Inf. Sci. (Ny)*, vol. 266, 2014, doi: 10.1016/j.ins.2013.12.060.
- [53] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, 2001, doi: 10.1023/A:1010933404324.
- [54] E. R.-R. Catherine Bounsaythip, "Overview of Data Mining for Customer Behavior Modeling," *VTT Inf. Technol.*, 2001.
- [55] G. G. EMEL and Ç. TAŞKIN, "Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması," *Eskişehir Osmangazi Üniversitesi Sos. Bilim. Derg.*, vol. 6, no. 2, pp. 221–239, 2005, doi: 10.17494/ogusbd.06798.
- [56] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, "Analysis of Various Decision Tree Algorithms for Classification in Data Mining," *Int. J. Comput. Appl.*, vol. 163, no. 8, 2017, doi: 10.5120/ijca2017913660.
- [57] H. Alharthi, D. Inkpen, and S. Szpakowicz, "A survey of book recommender systems," *J. Intell. Inf. Syst.*, vol. 51, no. 1, 2018, doi: 10.1007/s10844-017-0489-9.

- [58] M. Zhou, Z. Ding, J. Tang, and D. Yin, "Micro behaviors: A new perspective in E-commerce recommender systems," in *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, 2018, vol. 2018-February, doi: 10.1145/3159652.3159671.
- [59] Y. K. Choi and S. K. Kim, "An auxiliary recommendation system for repetitively purchasing items in E-commerce," 2014, doi: 10.1109/BIGCOMP.2014.6741415.
- [60] Y. M. Li, C. Te Wu, and C. Y. Lai, "A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship," *Decis. Support Syst.*, vol. 55, no. 3, 2013, doi: 10.1016/j.dss.2013.02.009.
- [61] R. Lian, "The construction of personalized Web page recommendation system in e-commerce," 2011, doi: 10.1109/CSSS.2011.5974890.
- [62] K. Christidis and G. Mentzas, "A topic-based recommender system for electronic marketplace platforms," *Expert Syst. Appl.*, vol. 40, no. 11, 2013, doi: 10.1016/j.eswa.2013.01.014.
- [63] E. Sadıkoğlu and B. Demirelli Okkalıoğlu, "Çok Kriterli Ürün-Tabanlı İşbirlikçi Filtrelemede Ağırlıklandırma Yöntemlerini Kullanarak Tahmin Performansının Arttırılması," *Eur. J. Sci. Technol.*, 2020, doi: 10.31590/ejosat.779171.
- [64] M. Hou, R. Wei, T. Wang, Y. Cheng, and B. Qian, "Reliable medical recommendation based on privacy-preserving collaborative filtering," *Comput. Mater. Contin.*, vol. 56, no. 1, 2018, doi: 10.3970/cmc.2018.02438.
- [65] D. Sánchez-Moreno, A. B. Gil, and M. N. Moreno, "TV-seriesrec: A recommender system based on fuzzy associative classification and semantic information," in *Advances in Intelligent Systems and Computing*, 2013, vol. 221, doi: 10.1007/978-3-319-00563-8_25.
- [66] Y. H. Cho, J. K. Kim, and S. H. Kim, "A personalized recommender system based on web usage mining and decision tree induction," *Expert Syst. Appl.*, vol. 23, no. 3, 2002, doi: 10.1016/S0957-4174(02)00052-0.
- [67] A. Ayan, "E-Ticaret Sitelerinin Satışlarının Artırılması İçin Bulut Bilişim Tabanlı Yapay Öğrenme," Necmettin Erbakan Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Bölümü, 2018.
- [68] E. Kantepe, "Derin Öğrenme Yöntemleri Kullanılarak Ürün Önerme Sistemi Tasarımı," Marmara Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü, 2019.

- [69] E. Ş. Sivri, “Veri Madenciliği/ E-Ticaret İçin Ürün Tavsiye Sistemi Geliştirilmesi,” İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü, 2015.
- [70] W. Abdullah, “E-Ticaret İçin Ürün Tavsiye Sistemi Geliştirilmesi,” İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü, 2019.
- [71] H. Bulut and M. Milli, “İşbirlikçi filtreleme için yeni tahminleme yöntemleri,” *Pamukkale Univ. J. Eng. Sci.*, vol. 22, no. 2, 2016.
- [72] Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang, and J. Lu, “A hybrid fuzzy-based personalized recommender system for telecom products/services,” *Inf. Sci. (Ny)*, vol. 235, 2013, doi: 10.1016/j.ins.2013.01.025.
- [73] W. McKinney, “Data Structures for Statistical Computing in Python,” 2010, doi: 10.25080/majora-92bf1922-00a.
- [74] G. Hamerly and C. Elkan, “Learning the K in K-Means,” *Adv. Neural Inf. Process. Syst.*, vol. 17, Mar. 2004.
- [75] M. F. Adak and N. Yurtay, “Gini Algoritmasını Kullanarak Karar Ağacı Oluşturmayı Sağlayan Bir Yazılımın Geliştirilmesi,” *Int. J. INFORMATICS Technol.*, vol. 6, pp. 1–6, Jan. 2013.
- [76] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, 2011.

ÖZGEÇMİŞ

Adı Soyadı : Metehan UÇAR

ÖĞRENİM DURUMU

Derece	Eğitim Birimi	Mezuniyet Yılı
Yüksek Lisans	Sakarya Üniversitesi / Fen Bilimleri Enstitüsü / Bilgisayar ve Bilişim Mühendisliği	Devam ediyor
Lisans	Sakarya Üniversitesi / Bilgisayar ve Bilişim Bilimleri Fakültesi / Bilgisayar Mühendisliği	2018
Lise	Sakarya Mithatpaşa Şükrü Ayna Anadolu Lisesi	2013

İŞ DENEYİMİ

Yıl	Yer	Görev
2019-Halen	Yalova Üniversitesi	Araştırma Görevlisi

YABANCI DİL

İngilizce