



T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



Yüksek Lisans Tezi

VERİ MADENCİLİĞİ ALGORİTMALARI İLE KREDİ KARTI
KULLANIM ALIŞKANLIKLARININ İNCELENMESİ ve KİŞİYE
ÖZGÜ KAMPANYA TEKLİFİ

Tuğçe Süheyla ÇETİN

Enformatik Anabilim Dalı

Enformatik Programı

DANIŞMAN
Prof. Dr. Sevinç GÜLSEÇEN

Temmuz, 2019

İSTANBUL


Uygundur
10.07.2019
Prof. Dr. S. Gülseçen

Bu çalışma, 2.07.2019 tarihinde ařağıdaki jüri tarafından Enformatik Anabilim Dalı, Enformatik Programında Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi


Prof. Dr. Sevinç GÜLSEÇEN(Danışman)
İstanbul Üniversitesi
Fakülte


Doç. Dr. Çiğdem EROL
İstanbul Üniversitesi
Enformatik


Dr. Öğr. Üyesi Ayşe ÇINAR
Marmara Üniversitesi
İşletme Fakültesi



20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi’nin aboneli olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

ÖNSÖZ

Yüksek lisans tez çalışmam boyunca gösterdiği her türlü destek ve yardımdan dolayı tez danışmanım Prof. Dr. Sevinç GÜLSEÇEN'e,

Tez hazırlama sürecimde yardımlarını esirgemeyen Dr. Murat GEZER'e,

Hayatım boyunca desteğini benden esirgemeyen, ellerini hep üzerimde hissettiğim aileme,

Bana inanıp yardımlarını eksik etmeyen eşime ve onlar için hep daha iyisini yapmak istediğim oğullarıma teşekkürü bir borç bilirim.

Temmuz, 2019.

Tuğçe Süheyla ÇETİN

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	iv
İÇİNDEKİLER.....	v
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ.....	x
SİMGE VE KISALTMA LİSTESİ	xi
ÖZET	xii
SUMMARY	xiii
1. GİRİŞ	1
2. GENEL KISIMLAR	3
2.1. VERİ MADENCİLİĞİ.....	3
2.1.1. Veri Madenciliği Modelleri	5
2.1.1.1. Tahmin Edici Modeller (Predictive)	5
2.1.1.2. Tanımlayıcı Modeller (Descriptive).....	5
2.1.2. Veri Madenciliği Yöntemleri.....	6
2.1.2.1. Sınıflandırma ve Regresyon	6
2.1.2.2. Kümeleme.....	6
2.1.2.3. Birliktelik Kuralı	7
2.1.3. Veri Madenciliği Uygulama Alanları	7
2.2. TAVSİYE SİSTEMLERİ	8
2.2.1. İçerik Tabanlı Filtreleme Yöntemleri	10
2.2.2. İş Birlikçi Filtreleme Yöntemleri	13
2.2.2.1. Bellek Tabanlı İş Birlikçi Filtreleme Yöntemleri	15
2.2.2.2. Model Tabanlı İş Birlikçi Filtreleme Yöntemleri	16
2.2.3. Melez (Hibrid) Filtreleme Tavsiye Sistemleri.....	16
2.2.3.1. Özellik Kombinasyonu (Feature Combination).....	17
2.2.3.2. Özellik Arttırma (Feature Augmentation).....	18
2.2.3.3. Ağırlıklı Yöntem (Weighted method).....	18
2.2.3.4. Yarı Seviye (Meta Level)	18
2.2.3.5. Kademeli Sistemler (Cascade)	18
2.2.3.6. Anahtarlama Melezi (Switching).....	19

2.2.3.7.	<i>Karma Melez Sistemler (Mixed)</i>	19
2.3.	TAVSİYE SİSTEMLERİNDE BENZERLİK HESAPLAMA TEKNİKLERİ	19
2.3.1.	Kosinüs Benzerliği	20
2.3.2.	Pearson Korelasyon Katsayısı	20
2.3.3.	Jaccard Benzerliği.....	21
2.3.4.	Öklid Uzaklığı	21
2.4.	TAVSİYE SİSTEMLERİNDE KARŞILAŞILAN SORUNLAR	22
2.4.1.	Soğuk Başlangıç (Cold Start)	22
2.4.2.	Seyreklik (Sparsity)	23
2.4.3.	Ölçeklenebilirlik (Scalability)	23
2.4.4.	Aşırı Uzmanlaşma (Overspecialization).....	24
2.4.5.	Gecikme (Latency)	24
2.4.6.	Benzerlik (Synonymy).....	24
2.4.7.	Kararsız Kullanıcılar (Gray Sheep)	25
2.4.8.	Yanlış Yönlendirme (Shilling Attacks)	25
2.4.9.	Gizlilik (Privacy)	26
2.5.	TAVSİYE SİSTEMLERİ DEĞERLENDİRME ÖLÇÜTLERİ	27
2.5.1.	Doğruluk (Accuracy).....	28
2.5.2.	Kapsama Oranı (Coverage)	31
2.5.2.1.	<i>Öğe Kapsamı</i>	31
2.5.2.2.	<i>Kullanıcı Alanı Kapsamı</i>	31
2.6.	TAVSİYE SİSTEMLERİ İLE İLGİLİ ARAŞTIRMALAR.....	32
3.	MALZEME VE YÖNTEM	35
3.1.	ÇALIŞMANIN AMACI VE HEDEFİ.....	35
3.2.	ARAŞTIRMANIN MODELİ	35
3.3.	VERİNİN ELDE EDİLMESİ VE HAZIRLANMASI.....	35
3.4.	VERİLERİN ANALİZİ VE MODELİN SEÇİLMESİ.....	40
3.5.	TAVSİYE SİSTEMİNİN OLUŞTURULMASI	41
4.	BULGULAR	42
4.1.	MODELLERİN UYGULANMASI VE KARŞILAŞTIRILMASI.....	42
4.1.1.	Öğe Benzerliği Modeline Kosinüs Benzerliği Uygulanması	43
4.1.1.1.	<i>Temel Veri Sonuçları</i>	43
4.1.1.2.	<i>Etkileşime Dayalı Veri Seti için Sonuçlar</i>	45
4.1.1.3.	<i>Normalize Edilen Veri için Sonuçlar</i>	48

4.1.2.	Öge Benzerliği Modelinde Pearson Korelasyonunun Hesaplanması	50
4.1.2.1.	<i>Temel Veri Seti için Sonuçlar</i>	50
4.1.2.2.	<i>Etkileşime Dayalı Veri için Sonuçlar</i>	53
4.1.2.3.	<i>Normalize Edilen Veri için Sonuçlar</i>	55
4.1.3.	Modellerin Karşılaştırılması	57
4.2.	OLUŞTURULAN TAVSİYELERİN YORUMLANMASI.....	59
5.	TARTIŞMA VE SONUÇ	64
	KAYNAKLAR	66
	ÖZGEÇMİŞ	72



ŞEKİL LİSTESİ

	Sayfa No
Şekil 2.1: CRISP-DM Veri Madenciliği Modeli [7].	4
Şekil 2.2: Tavsiye Sistemi Teknikleri [21].	10
Şekil 2.3: İçerik Tabanlı Filtrelemenin Genel Yapısı [27].	12
Şekil 2.4: İş Birlikçi Filtreleme Tavsiye Sistemlerinin Genel Yapısı [35].	14
Şekil 2.5: Melez Tavsiye Sistemlerinin Yapısı [43].	17
Şekil 3.1: Çalışmada Kullanılan Örnek Veri seti.	36
Şekil 3.2: Örnek Sektör Kodları [82].	37
Şekil 3.3: Derecelendirme Veri Seti.	38
Şekil 3.4: Temel Veri Seti Derecelendirme Dağılımı.	39
Şekil 3.5: Etkileşime Dayalı Veri Seti.	40
Şekil 3.6: Normalize Veri Seti.	40
Şekil 3.7: Çalışma Akış Şeması.	41
Şekil 4.1: Örnek Python Kodu.	42
Şekil 4.2: Temel Veri Kosinüs Benzerliği Hassasiyet ve Duyarlılık Değerleri.	43
Şekil 4.3 : Temel veri Hassasiyet Değerler Grafiği.	44
Şekil 4.4: Temel Veri Duyarlılık Verileri.	44
Şekil 4.5: Etkileşime Dayalı Veri Seti için Hassasiyet ve Duyarlılık.	46
Şekil 4.6: Etkileşime Dayalı Veri Seti Hassasiyet.	46
Şekil 4.7: Etkileşime Dayalı Veri Seti Duyarlılık.	47
Şekil 4.8: Normalize Veri Seti Hassasiyet ve Duyarlılık Değerleri.	48
Şekil 4.9: Normalize Veri Seti için Hassasiyet.	49
Şekil 4.10: Normalize Veri Seti için Duyarlılık.	49
Şekil 4.11: Temel Veri için Pearson Hassasiyet ve Duyarlılık Değerleri.	51

Şekil 4.12: Temel Veri Pearson Korelasyonu Hassasiyet Grafiği.....	51
Şekil 4.13: Temel Veri Pearson Duyarlılık Grafiği.....	52
Şekil 4.14: Etkileşime Dayalı Veri Seti Pearson Hassasiyet ve Duyarlılık Değerleri.....	53
Şekil 4.15: Etkileşime Dayalı Veri Seti Pearson Korelasyonu ile Hassasiyet Grafiği.....	53
Şekil 4.16: Etkileşime Dayalı Veri Seti Pearson Benzerliği Duyarlılık Grafiği.....	54
Şekil 4.17: Normalize Veri Seti için Pearson Hassasiyet ve Duyarlılık.....	55
Şekil 4.18: Normalize Veri Seti için Pearson Hassasiyet Grafiği.....	55
Şekil 4.19: Normalize Veri Seti için Duyarlılık Grafiği.....	56
Şekil 4.20: Modellere göre Hassasiyet Değerlerinin Karşılaştırılması.....	57
Şekil 4.21: Modellere göre Duyarlılık Değerlerinin Karşılaştırılması.....	58
Şekil 4.22: Modellere göre RMSE Değerlerinin Karşılaştırılması.....	58
Şekil 4.23 : Pfm Türlerine göre İşlem Sayıları.....	59
Şekil 4.24: Müşterilere İlk Önerilen Kampanya Grupları.....	60
Şekil 4.25: Müşterilere İkinci Olarak Önerilen Kampanya Grupları.....	61
Şekil 4.26: Üretilen Tavsiye Sayıları Dağılımı.....	62
Şekil 4.27: Kullanıcı Pfm- Derecelendirme Değerleri.....	63

TABLO LİSTESİ

	Sayfa No
Tablo 1: Çevrim İçi Haber Öneri Sonuçları.	30
Tablo 2: Temel Veri Kosinüs Benzerliği RMSE Değerleri.....	45
Tablo 3:Etkileşime Dayalı Veri Kosinüs Benzerliği RMSE Değerleri.....	47
Tablo 4: Normalize Veri Seti Kosinüs Benzerliği RMSE Değerleri.....	50
Tablo 5: Temel Veri Pearson RMSE Değerleri.....	52
Tablo 6: Etkileşimli Veri Seti Pearson Modeli RMSE Değerleri.....	54
Tablo 7: Normalize Veri Seti İçin Pearson RMSE Değerleri.....	56

SİMGE VE KISALTMA LİSTESİ

Kısaltmalar	Açıklama
BKM	: Bankalar Arası Kart Merkezi
CRISP – DM	: Endüstriler Arası Standart İşleme (Cross Industry Process for Data Mining)
CRM	: Müşteri İlişkileri Yönetimi
FİBS	: Adil Bilgi Uygulaması
İBF	: İşbirlikçi Filtreleme
İTF	: İçerik Tabanlı Filtreleme
MAE	: Ortalama Mutlak Hata
MCC	: Sektör Kodu (Merchant Category Code)
OECD	: Ekonomik İş Birliği ve Kalkınma Örgütü
PFM	: Gruplandırılmış Kampanya Kodu
RMSE	: Ortalama Kök Kare Hatası
TS	: Tavsiye Sistemleri

ÖZET

VERİ MADENCİLİĞİ ALGORİTMALARI İLE KREDİ KARTI KULLANIM ALİŞKANLIKLARININ İNCELENMESİ ve KİŞİYE ÖZGÜ KAMPANYA TEKLİFİ

YÜKSEK LİSANS TEZİ

Tuğçe Süheyla ÇETİN

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman : Prof. Dr. Sevinç GÜLSEÇEN

Bu tez çalışmasında özel bir bankadan alınan 4497 adet müşterinin 3 aylık süreç içerisindeki kredi kartı hareketleri incelenerek kendilerine en uygun kampanyaya yönlendirilmesi hedeflenmiştir. Veriler ön işlemeden geçirildikten sonra Python programlama dili ile yazılan bir tavsiye sistemi ile oluşturulan iki model uygulanmış, başarılı olan modele göre bir tavsiye sistemi yazılmıştır. Bu tez çalışması hayatımızın her alanında olan tavsiye sistemlerinin kampanya yönetiminde kullanılmasını ele alması sebebiyle güncel ve özgün bir çalışma olmayı hedeflemektedir. Ayrıca ülkemizde bu alanda yapılmış olan araştırmaların az olması nedeniyle bu çalışma ilgili alan yazına katkıda bulunacaktır.

Temmuz 2019, 85 sayfa.

Anahtar kelimeler: Veri Madenciliği, Tavsiye Sistemleri, Kampanya Yönetimi

SUMMARY

ANALYZING CREDIT CARD USAGE BEHAVIOURS WITH DATA MINING ALGORITHMS AND PEARSON SPECIFIC CAMPAIGN OFFER

M.Sc. THESIS

Tuğçe Süheyla ÇETİN

İstanbul University

Institute of Graduate Studies in Sciences

Department of Informatics

Supervisor : Prof. Dr. Sevinç GÜLSEÇEN

In this thesis, 4497 customers received from a private bank were aimed to be directed to the most suitable campaign by examining their credit card movements within 3 months. After the data was pre-processed, two models were implemented with a recommendation system written in the Python programming language and a recommendation system was written according to the successful model. This thesis aims to be an up-to-date and original study as it addresses the use of referral systems in all areas of our lives in the campaign management. Also, this study will contribute to the related literature due to the small number of studies conducted in this field in our country.

July 2019, 85 pages.

Keywords: Recommendation Systems, Data Mining, Campaign Management

1. GİRİŞ

Günümüzde aktif biçimde kullandığımız kredi kartları, Türkiye’ de ilk olarak 1968 yılında Koç grubuna bağlı olarak kullanılmaya başlanmış olup ülkemizde kredi kartlarının yaygın olarak kullanımı 1990’lı yıllarda başlamıştır [1]. Bankalar arası Kart Merkezi A. Ş.’nin (BKM’nin) verilerine göre Türkiye’de 2016 yılında toplam 58.795.476 kredi kartı bulunurken 2018 yılında bu sayı 66.304.603 olarak belirtilmiştir [2]. Bu kredi kartları ile 2018 yılında 3.945.245.693 işlem yapılmış ve toplam işlem tutarı 794.305,40 (Milyon TL) olarak hesaplanmıştır. Bu yüklü miktarda işlem adedi günümüzde popüler olan başka bir konu olan büyük veri (big data) ve veri madenciliğinin de başlıca konularından birini oluşturmaktadır.

Veri Madenciliği en sade haliyle “büyük ölçekli veriler arasından değerli olan bilgiyi elde etme işidir [3]. Bilgisayar sistemlerinin her geçen gün ucuzluyor olması, hem de güçlerinin artıyor olması, bilgisayarlarda daha büyük miktarlarda verinin saklanabilmesine imkân vermektedir. Bu yüzden, büyük miktardaki verileri işleyebilen teknikleri kullanabilmek büyük önem taşımaktadır [4].

Veri madenciliği sağlık, telekomünikasyon, finans gibi birçok sektörde yaygın olarak kullanılmaktadır. Finans sektöründe bankalar sahip oldukları büyük veri tabanları ile veri madenciliğini bu alanda kullananların başında gelmektedir. İlk başlarda kredi risk analizi, sahtekarlık tespiti gibi alanlarda kullanılsa da zamanla müşteri memnuniyetini artırma çalışmalarının temelini oluşturmuştur.

Bankacılık sektöründe müşteri ilişkileri yönetimi (CRM) uygulaması müşterilere ilişkin verilerin farklı müşteriyle temas noktalarından elde edilmesi ve bu veriler doğrultusunda müşterilerin tanımlanması, müşterilerin sınıflandırılması (demografik özelliklerine veya satın alma tiplerine göre) ve hedef pazarlarını hangi müşteri grubuna ne hizmet vereceğinin saptanması ve her bölüm için karlılık analizlerinin yapılması, müşteri önceliklerinin belirlenmesi ve bu önceliklere uygun pazarlama karmasının tasarlanması ve buna uygun müşterilerin yönetimidir. Böylelikle bankalar var olan mevcuttaki müşterilerin hangi ürünü yada hizmeti seçtiğini tahmin ederek, kişiye özgü pazarlama yaklaşımı ile müşterilere daha iyi hizmet sunmaktadır.

Bu tez çalışmasında özel bir bankadan alınan müşteri kitlesinin yaklaşık üç aylık kredi kartı hareketleri incelenecek olup, buradan her bir müşteri için bir profil oluşturulması hedeflenmektedir. Böylece her bir profile kendine en uygun kampanya tavsiye edilecek olup müşteri memnuniyetinin artırılması amaçlanmaktadır.

Bu tez çalışması, günümüzde aktif olarak kullanılan kredi kartları ile yapılan işlemlerden yola çıkarak, kişiye en uygun modeli bulup uygun kampanyayı sunmayı hedeflemektedir. Bu konuda yapılan araştırmaların az olması sebebiyle literatüre katkıda bulunacaktır.



2. GENEL KISIMLAR

2.1. VERİ MADENCİLİĞİ

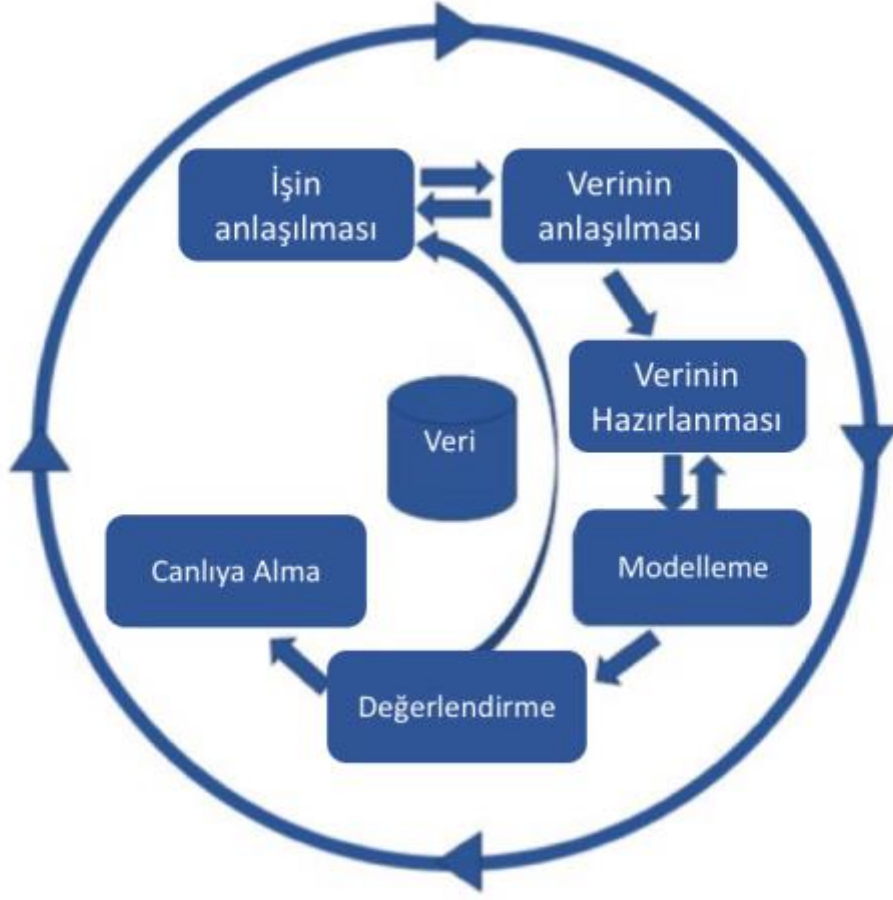
Veri madenciliği, büyük veri setlerinden yararlı bilgi çıkarım sürecidir. Ham veri tek başına değersizdir, verinin işlenmesiyle bilgi elde edilir. Diğer bir tanımla veri madenciliği, verileri bazı analiz teknikleriyle analiz ederek büyük veri tabanlarından veri öznitelikleri, ilişkiler veya modeller gibi gizli bilgileri çıkarmayı ve faydalı bilgilerle özetlemeyi amaçlayan süreçler zinciridir [5]. Veri madenciliği istatistik, makine öğrenmesi, veri yönetimi, örüntü tanıma, yapay zekâ gibi alanların kesiştiği bir disiplindir.

Veri madenciliği süreci temel olarak 4 adım ile açıklanır.

- Problemin tanımlanması,
- Verilerin hazırlanması (Ön İşleme)
- Modelin Seçilmesi
- Değerlendirme

Daha detaylı bir süreç tanımını, veri madenciliğinin uluslararası standardı olarak kabul edilmiş, Endüstriler Arası Standart İşleme (Cross Industry Process for Data Mining) yapmıştır. CRISP-DM süreç modeli büyük veri madenciliği projelerinin hız, güven ve tekrarlı bilirlilik gibi özelliklerini artırıp yönetilmeyi kolaylaştırmayı amaçlamaktadır. Diğer bir amacı ise veri madenciliği projelerinin maliyetini düşürmektir. CRISP-DM veri madenciliği sürecini altı aşamada tanımlamıştır [6].

- İşin Anlaşılması,
- Verinin Anlaşılması,
- Verinin Hazırlanması,
- Modelleme,
- Değerlendirme,
- Canlıya Alma/ Yayılım



Şekil 2.1: CRISP-DM Veri Madenciliği Modeli [7].

Burada işin anlaşılması aşamasında, problemin tanımı yapılır, çözümden beklentiler planlanır. İkinci adımda probleme uygun veriler seçilir. Burada ilk adımda problemin doğru tanımlanması çok önemlidir, yanlış tanımlanan bir problem verinin yanlış seçilmesine ve sonraki adımlarında probleme uygun olmayan veriler üzerinden çalışmasına sebep olur. Verinin hazırlanması yani ön işleme adımında ise veri üzerinde yapılacak işlemlerdir. Örneğin veri setinde bulunan eksik verilerin modele alınmaması ya da eksik verilerin veri setine uygun olarak tamamlanması örnek verilebilir. Modelleme aşamasında probleme en uygun model seçilir. Değerlendirme aşamasında, diğer adımların ve modelin işin anlaşılması aşamasında tanımlanan problemin çözümüne ne kadar katkı sağladığı ölçülüp açıkta kalan bir konu olup olmadığı araştırılır. Yayılım aşamasında bir ürün ortaya çıkar örnek olarak; çıktının bir rapor

olarak sunulması ya da veri madenciliği sürecinin programlama dilleri aracılığıyla bir uygulama haline getirilmesi verilebilir.

2.1.1. Veri Madenciliği Modelleri

Veri madenciliğinde kullanılan modeller tahminleyici ve tanımlayıcı olmak üzere iki ana başlık altında toplanmaktadır.

2.1.1.1. Tahmin Edici Modeller (Predictive)

Tahmin edici modellerde, sonuçları bilinen veri setlerinden yola çıkılarak bir model tanımlanır ve bu model ile sonuçları bilinmeyen veri setleri ile hedef niteliğin tahmin edilmesi amaçlanır. Örnek olarak bir bankanın daha önce vermiş olduğu kredi bilgilerini veri seti olarak alırsak; burada hedef nitelik kredinin ödenip ödenmediği bilgisidir. Bu veri seti üzerinden kurulan model aracılığı ile daha sonraki müşterilerin özelliklerine göre kredinin ödenme durumunu tahmin etmekte kullanılabilir [4]. Tahmin edici modeller “Ne olacak?” ve “Neden olacak?” sorularına cevap ararlar. Tahmin edici model, tanımlanmış değerleri kullanarak tanımlanmış değerleri kullanarak tahminde bulunur. Örneğin, Sınıflandırma, regresyon, zaman serileri analizi tahmin edici modellere örnektir. Veri madenciliği uygulamalarının çoğu, verilerin gelecekteki durumunu tahmin etmek içindir [8].

Heinzelmann yaptığı araştırma ile tahmin edici analiz yöntemleri ile pazarlama yöntemlerini incelemiş ve bunların en popülerini tavsiye üretmek için kullanılan iş birlikçi filtreleme yöntemleri olduğu sonucuna varmıştır [9].

2.1.1.2. Tanımlayıcı Modeller (Descriptive)

Tanımlayıcı model verideki tasarımları veya ilişkileri tanır ve çalışılan veri setinin özelliklerini keşfeder. Tanımlayıcı modeller “ne oldu?” sorusuna cevap aralar. Tahmin edici modeller geleceğe odaklanırken tanımlayıcı modeller bize geçmişle ilgili analizler verir ve gelecek sonuçları nasıl etkileyebileceklerini öngörmeye yardımcı olurlar [9]. Tanımlayıcı modeller, algoritmanın kullanıcı yönlendirmesi olmadan ilişkileri tanımladığı denetimsiz öğrenmeye bir örnektir. Hedef değişkenin değerlerini tahmin etmek yerine veri yapısına, ilişkilere ve bağlantıya ilişkin ipuçları vermeyi amaçlamaktadır [10]. Örneğin geliri X-Y aralığında ve arabası olan çocuklu aileler ile geliri X-Y aralığından düşük çocuğu olmayan

ailelerin satın alma özelliklerinin arasındaki benzerliğin bulunması tanımlayıcı modele örnektir[4]. Birliktelik kuralları, kümeleme tanımlayıcı modele örnektir.

2.1.2. Veri Madenciliği Yöntemleri

Veri madenciliği yöntemleri sınıflandırma ve regresyon, kümeleme ve birliktelik kuralı olmak üzere üç başlıkta incelenebilir.

2.1.2.1.Sınıflandırma ve Regresyon

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

Sınıflandırma ve regresyonda kullanılan temel teknikler şunlardır [11]:

- Karar Ağaçları,
- Yapay Sinir Ağları,
- Genetik Algoritmalar
- K- En Yakın Komşu
- Naive Bayes
- Karar Destek Makineleri
- Bellek Temelli Nedenleme
- Lojistik Regresyon

Karar ağaçları veri tabanları sistemine kolay uygulanabilmesi ve yorumlamasının basit olması sebebiyle sınıflandırma tekniklerinde en çok kullanılan yöntemdir.

2.1.2.2.Kümeleme

Kümeleme, veri setini birbirine benzer verileri sınıflara ayırarak verileri gruplandırır. Bir küme birbirlerine benzer ve diğer kümelerin öğelerinden farklı öğelerin birleşimi olarak tanımlanabilir. Kümeleme sınıflandırma yönteminden farklı olarak hedef nitelik içermez[12]. Kümeleme istatistik, biyoloji ve makine öğrenmesi gibi birçok alanda kullanılır [4]. Çevrimiçi alışveriş sistemlerinde müşterilerin gruplandırılması ve grupların satın alma ilişkilerinin çıkarılması kümelemeye örnek olarak verilebilir.

Kümeleme yöntemleri aşağıdaki şekilde sınıflandırılır [13]:

- Bölümlenme Metotları,
- Hiyeraşik Metotlar,
- Yoğunluk Tabanlı Metotlar,
- Model Tabanlı Metotlar

En çok kullanılan yöntem bölümlenme metotlarıdır. Burada n veri setindeki nesne sayısı, k ise bölünecek küme sayısıdır. En yaygın bölümlenme yöntemi K- Means yöntemidir.

2.1.2.3. Birliktelik Kuralı

Birliktelik kuralları madenciliği algoritmaları, bir veri setindeki diğer öğelerin varlığına dayanarak bir öğenin oluşumunu öngören kuralları çıkarır. Burada amaç, belirli bir veri setinde birlikte olan öğeleri belirlemektir.

Verilen veri seti için üretilen kuralların sayısı çok büyük olabileceğinden ve keşfedilen birlikteliklerin hepsi anlamlı olmadığından, veri tabanındaki daha az ilişkili kurallar belirlenerek göz ardı edilir [14].

Birliktelik kurallarının sınıflandırmadan farkı, bir değişkenin iki yönlü olabilmesi, bazı durumlar için girdi, bazıları için çıktı değişken gibi tanımlanabilmesidir. En yaygın kullanılan birliktelik kuralı Apriori, GRI ve CARMA algoritmalarıdır [15].

2.1.3. Veri Madenciliği Uygulama Alanları

Veri madenciliği i istatistik, makine öğrenmesi, yapay zekâ gibi alanların kesiştiği bir disiplin olduğundan uygulama alanları oldukça geniştir. Temel kullanım alanları aşağıdaki şekilde gruplandırılır.

- Sağlık,
- Telekomünikasyon,
- Bankacılık ve Finans,
- Eğitim
- Pazarlama

Veri madenciliğinin en yaygın kullanım alanlarından biri bankacılık sektörüdür. Bankacılıkta kredi kartı dolandırıcılıklarının tespiti, müşteri ilişkileri yönetimi için müşterilerin sınıflandırılması, kredi taleplerinin değerlendirilmesi en temel veri madenciliği uygulamalarına örnektir.

Tavsiye sistemleri genel olarak insan- bilgisayar etkileşimi veya bilgi edinme (IR) gibi komşuluk tabanlı teknikler ve metodolojiler uygulamaktadır. Bununla birlikte, bu sistemlerin çoğu temelinde veri madenciliği tekniği olarak anlaşılabilir bir algoritmaya sahiptir. Aslında, veri madenciliğindeki zorlukların çoğu aynı zamanda tavsiye sistemlerinde de karşılaşılan zorluklardır [16].

2.2. TAVSİYE SİSTEMLERİ

Tavsiye sistemleri karmaşık bilgi ortamlarındaki kullanıcılar için bir karar verme stratejisi olarak tanımlanmaktadır [17]. Diğer bir tanımla; alternatifler hakkında yeterli kişisel bilgi ya da deneyim olmadığında seçim yapmak için başkalarının tavsiyelerini kullanmanın bir aracıdır [18].

Tavsiye sistemleri (TS), büyük miktardaki veri (kitap, resim, film, müzik) içerisinden kullanıcıların ilgisini çekebilecek olanları bulup öneren tahminleme yazılımları olarak da tanımlanabilirler [19]. TSleri sınıflandırma, kullanıcı modelleme ve bilgi filtreleme gibi birçok veri madenciliği tekniğini içinde barındırır. TSleri bilişsel bilim, insan-bilgisayar etkileşimi, makine öğrenmesi gibi birçok araştırma dalında ele alınmaktadır.

Tavsiye sistemlerinin genel yapısı aşağıdaki gibidir [20];

- **Arka plan verisi:** Sistemin öneri işlemine sahip olmadan önce sahip olduğu bilgiler
- **Veri Girişi:** Kullanıcıya bir öneri oluşturulması için sisteme iletilmesi gereken bilgiler
- **Algoritma:** Arka plan veri ile iletilen verileri birleştirip kullanıcıya tavsiyeleri oluşturur

Verinin toplanması aşamasında, kullanıcının bilgileri, eriştiği içerik vb. bilgiler kullanıcı profili ya da modeli oluşturmak için toplanır. Hiçbir tavsiye sistemi modeli tam olarak oluşturamadan çalışmaz. Veriler tavsiye sisteminin yapısına göre kapalı bildirim, açık bildirim ve melez bildirim yöntemleri ile toplanırlar [21].

- **Açık Bildirim (Explicit Feedback):** Açık bildirim kullanıcılara oylama gibi tercihlerini net bir şekilde ifade etme imkanı sağlar. Bu şekilde oluşturulan sistemlerin tavsiyelerinin doğruluğu kullanıcıların derecelendirmesine bağlıdır. Açık bildirimlerin en önemli açığı kullanıcıların öğelerle ilgili fikirlerini paylaşmak da çekingen olmalıdır[22].
- **Kapalı Bildirim (Implicit Feedback):** Kullanıcıların fikirlerini direk olarak beyan etmelerine gerek yoktur. Kapalı bildirimler kullanıcının alışveriş geçmişine, bir sayfada ne kadar zaman geçirdiği, mailin içeriği, kullanıcının takip ettiği linkler gibi veriler ile beslenir [23].
- **Melez Bildirim (Hybrid Feedback):** Açık ve kapalı bildirimlerin birleştirilmesiyle oluşturulur. Tavsiye sistemi ne kadar iyi geri bildirim alırsa model o kadar iyi oluşur, sistem de daha başarılı tavsiyelerde bulunur.

Tavsiye sistemleri internetteki içeriğin kişiselleştirilmesinde inanılmaz bir etkiye sahiptir. En yaygın kullanım alanı olan e-ticarette uygulanmaya başlanması sonrasında satışları %8 ile %10 arasında arttırdığı gözlemlenmiştir [24]. Bu sistemler karar verme, kar oranını artırma ve risk azaltma gibi bir çok konuda önemli role sahiptirler. Bugün Google, LinkedIn, Twitter, ve Netflix gibi bir çok firma tavsiye sistemlerini aktif olarak kullanmaktadır [25].

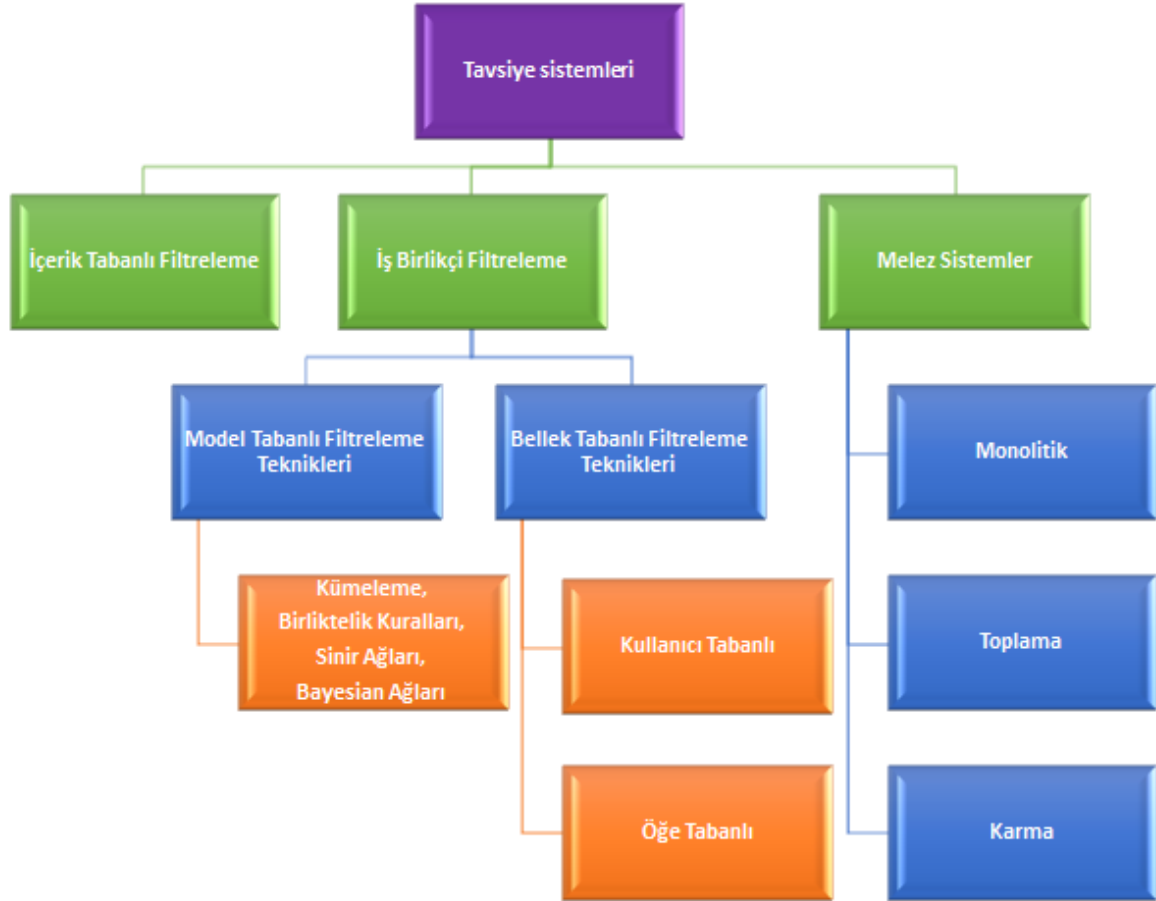
Tavsiye sistemleri hem servis sağlayıcılar için hem de kullanıcılar için avantajlara sahiptir [26]. Örneğin e-ticaret gibi satışa yönelik sistemlerde kullanıcılar ilgisini çekebilecek veya ihtiyacı olan ürünleri bulmakta kolaylık yaşarken, servis sağlayıcı ise müşteri memnuniyeti ve sadakati sağlamanın yanında karlılık oranını da arttırdığından her iki taraf içinde bir kazanım söz konusudur. Bilimsel kütüphanelerde tavsiye sistemleri katalog taramasında kullanıcıların işini kolaylaştırabilmektedir [21]. Bir öneri sisteminin üreteceği tavsiyenin kalitesi kullanılan verinin çeşidi ve miktarıyla alakalıdır.

TSleri aslında kullanıcıların geri bildirimlerine dayanmaktadır, bu geri bildirimlerin elde edilme ve kullanılma şekilleri ise öneri sisteminin tipine göre değişmektedir, Kullanıcıdan alınan geri bildirimler bir filme verilen oy olabildiği gibi bir sayfada ne kadar kaldığı ya da bir şarkıyı kaç kere dinlediği olabilir. Tavsiye sistemleri kendi arasında üç ana grupta toplanırlar;

- İçerik tabanlı tavsiye sistemler,
- İş birlikçi filtreleme tavsiye sistemleri

- Melez sistemler.

Bunlardan en popüler olanı ve en yaygın olarak kullanılan öneri sistemi iş birlikçi tavsiye sistemleridir.



Şekil 2.2: Tavsiye Sistemi Teknikleri [21].

2.2.1. İçerik Tabanlı Filtreleme Yöntemleri

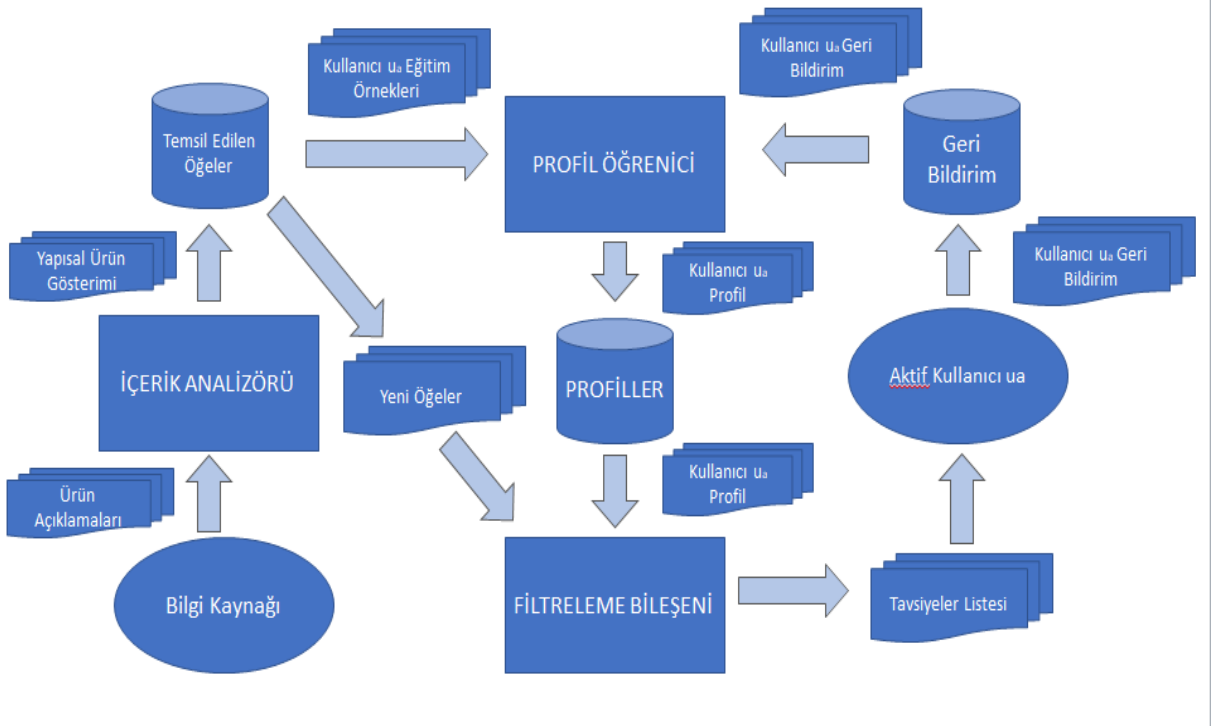
İçerik tabanlı filtreleme yöntemleri (İTF) kullanıcının daha önceki tercihlerine dayanır ve algoritmalar ile potansiyel olarak önerilecek ürünler kullanıcının daha önce etkileşimde bulunduğu/oyladığı ürünler ile karşılaştırılarak en iyi eşleştirilenler önerilir. Web sayfası, yayın ve haber tavsiyelerinde İTF yöntemleri çok başarılıdır [20].

İTF yöntemleri iki doküman arasındaki benzerliği hesaplamak için çeşitli modelleme yöntemleri kullanır. Vektör uzay modeli ile terim sıklığı (Term Frequency-TF) ya da Ters Döküman Sıklığı (Inverse Document Frequency – IDF) hesaplanabilir ya da Naive Bayes sınıflandırma yöntemleri, karar ağaçları algoritmaları kullanılarak da öneriler de bulunulur. Bu teknikler ile ya istatistiksel analizler yapılır ya da makine öğrenmesi algoritmaları kullanılarak tavsiyelerde bulunulur [21].

İTF yöntemleri sadece aktif kullanıcının geçmiş etkileşimleriyle çalışır, diğer kullanıcıların tercihlerini dikkate almaz. Mevcut yaklaşım yakın zamandaki beğenilerin nitelik olarak benzerliğini hesaplayıp benzer öğeleri tavsiye etmektir. İçerik tabanlı tavsiye sistemlerinin genel yapısı üç bileşenden oluşur [27].

- **İçerik Tahlil Modülü:** Bilginin yapısı olmadığında (metin vb.) içerisinden yapılandırılmış veriyi çıkarmak için bir takım ön işleme çalışmaları gerekir. Bu aşamanın asıl amacı veri içeriğini temsil eden yapının diğer alt kademelere uygun biçimde aktarılmasıdır
- **Profil Oluşturma Modülü:** Bu bileşen kişinin tercihlerini oluşturan verileri toplar ve genelleştirmeye çalışır. Çoğunlukla genelleme stratejisinde makine öğrenmesi teknikleri kullanılır.
- **Filtreleme Modülü:** Profil gösterimini öneride bulunulacak öğelerinkiyle eşleştirerek ilgili öğeleri tavsiye etmek için kullanıcı profilinden yararlanır.

İçerik tabanlı filtreleme yöntemlerinin yapısını şekildeki gibidir.



Şekil 2.3: İçerik Tabanlı Filtrelemenin Genel Yapısı [27].

İTF yöntemleri kullanıcı tarafından verilen bir derecelendirme olmasa bile öneride bulunabilirler. Aynı zamanda kullanıcı tercihleri değişirse, tavsiyeleri yeni tercihe göre kısa zamanda ayarlama kapasitesine sahiptir. Kullanıcılar kişisel bilgilerini paylaşmadan tavsiye alabilirler bu da gizlilik açısından önemlidir [28]. Ayrıca İTF yöntemleri önerilerin nasıl oluştuğuna dair kullanıcılara bilgilendirmeler sunabilir. Bununla birlikte İTF yöntemlerinin de literatürde tartışılan bazı problemleri vardır.

Bu tekniğin en büyük dezavantajı; profildeki öğelerin özellikleri hakkında derinlemesine bilgi ve açıklama sahibi olma gereksinimidir [21]. Örneğin bir kullanıcının çevrimiçi mağaza üzerinden bir telefon incelediğini düşünürsek; tavsiye sistemi bu telefonun fiyat, bellek kapasitesi, kamera özellikleri hakkında bilgi toplar ve veri tabanında bu özelliklere sahip diğer telefonları kullanıcıya öneri olarak döner. Bundan dolayı içeriğe ait bilgilerin azlığı tavsiye sisteminin başarısını etkilemektedir. Diğer bir problem ise iki farklı öğe aynı özellikler ile sistemde tanımlanmışsa birbirinden ayırt edilemeyecektir. Örneğin bilimsel makaleler genellikle anahtar kelimeler ile temsil edildiğinden, aynı anahtar kelimeler ile tanımlanan iki makale arasında İTF yöntemleri ile ayırım yapılamayacaktır.

Aşırı uzmanlaşma (Overspecialization) sorunu İTF yöntemleri için de geçerlidir. Sistem sadece kullanıcı profiline göre öneride bulunduğundan; tavsiyeler sadece kullanıcının daha önce değerlendirdiklerine benzer öğeler ile sınırlıdır [29]. Örneğin Türk mutfağı hakkında hiçbir değerlendirmesi olmayan biri bir Türk restoranı ile ilgili bir öneri asla alamaz.

Yeni kullanıcılar ise herhangi bir değerlendirmeleri olmadığından, İTF yöntemleri ile öneri alamazlar.

Pandora.com, içerik tabanlı filtreleme yöntemine güzel bir örnektir. Pandora.com şarkıları özelliklerine göre kataloglar ve ardından dinleyiciye melodileri bu daha önce sevdiğinize benziyor diyerek tavsiyede bulunur. Bu tavsiyeyi oluştururken kullanıcı profilleri yerine şarkıların özelliklerini kullanır [30].

2.2.2. İş Birlikçi Filtreleme Yöntemleri

İş birlikçi filtreleme yöntemlerin (İBF) de, sistem kullanıcılar hakkında bilgileri (aktiviteleri, değerlendirmeleri vb.) toplayıp analiz eder ve diğer kullanıcılar ile benzerliklerine göre tavsiyede bulunur [31].

İş birlikçi filtreleme yöntemleri, benzer ilgi alanlarına sahip kullanıcıları bularak, benzer kullanıcının yüksek puanla değerlendirdiği seçimleri hedefteki kullanıcıya tavsiye etmesi mantığı ile çalışır [32]. Burada en önemli faktör kullanıcının değerlendirmesidir (rating).

İçerik tabanlı filtrelemede kullanıcının daha önce değerlendirdiği öğelerin profillerine bakılarak kullanıcının daha önce değerlendirmedeği bir ürüne olan ilgisi tespit edilmeye çalışılırken, iş birlikçi filtrelemede ise kullanıcı değerlendirmelerinin benzerliklerinden yararlanılarak yeni ürünler önerilir [33].

İş birlikçi Filtreleme ilk defa 1992 yılında Goldberg tarafından Tapestry [34] olarak adlandırılan tavsiye sisteminin ortaya atıldığı bir çalışmada geçmektedir. Tavsiye sistemleri çalışılmaya başladığından beri de en gözde tavsiye sistemi İBF yöntemidir.

İBF öneri sistemlerinin genel yapısı Şekil 2.4 de gösterilmiştir.



Şekil 2.4: İş Birlikçi Filtreleme Tavsiye Sistemlerinin Genel Yapısı [35].

İBF yöntemi, filmle, müzik gibi yapısı kolayca ve yeterince tanımlanamayan içerik için alandan bağımsız bir tahmin yöntemidir. İBF, kullanıcılar tarafından tercih edilen öğeler için bir veri tabanı (kullanıcı-madde matrisi) oluşturularak çalışır. Daha sonra tavsiyelerde bulunmak için profiller arasındaki benzerlikleri hesaplayarak kullanıcıları ilgi alanlarına ve tercihlerine göre eşleştirir. Bu tür kullanıcılar komşuluk denilen bir grup oluşturur. Bir kullanıcı daha önce derecelendirmedeği, ancak grubundaki kullanıcılar tarafından zaten olumlu olarak derecelendirilen öğeler için tavsiyeler alır [21]. Yani işbirlikçi tavsiye yönteminin tavsiyeleri kullanıcının daha önceki beğenisiyle sınırlı değildir, çok çeşitli tavsiyelerde bulunabilir.

İş birlikçi filtreleme içerik tabanlı tavsiye yöntemlerine göre daha avantajlıdır çünkü içerik hakkında bilgisi olmamasına rağmen tavsiyelerde bulunabilir.

Literatürde yapılan araştırmalar incelendiğinde işbirlikçi filtreleme yönteminde kullanıcı profili, kullanıcı ve etkileşim kurdukça artan bir derecelendirme vektöründen oluşur. Derecelendirme vektörü çok değerli bir ölçüğe sahip olabileceği gibi (1-5 veya 1-10 vb.) ikili ölçekte de (beğendi/beğenmedi) olabilmektedir [20].

İş birlikçi filtreleme yöntemleri iki grupta toplanmaktadır:

- Model tabanlı iş birlikçi filtreleme,
- Hafıza tabanlı iş birlikçi filtreleme.

2.2.2.1. Bellek Tabanlı İş Birlikçi Filtreleme Yöntemleri

Bellek tabanlı algoritmalar tahmin oluşturmak için tüm kullanıcı beğeni veri tabanını kullanır. Bu sistemler hedefteki kullanıcıya en yakın olan komşuyu bulmak için bazı istatistiksel yöntemler kullanır. Bu model ayrıca en yakın komşu tavsiye sistemleri olarak da bilinmektedir [36]. Bellek tabanlı filtreleme yöntemleri benzerlikleri bir model oluşturmaya gerek duymadan hesaplayabilmektedir.

İlk tavsiye sistemleri olarak geçen GroupLens [37] ve Ringo [38] bu kategorideki tavsiye sistemlerindedir. Bellek tabanlı kullanıcı sistemleri; geçmişte benzer değerlendirmeleri olan kullanıcıların gelecekte de benzer beğenileri olacağı görüşünü benimser.

Kullanıcı tabanlı ve öğe tabanlı olmak üzere iki türlü bellek tabanlı tavsiye sistemi vardır. Kullanıcı tabanlı iş birlikçi filtreleme de tavsiyeler kullanıcının komşularının tercihlerine göre üretilir. Bu tavsiye sisteminde önce kullanıcıya benzer tercihleri olan kullanıcılar seçilir, daha sonra aktif kullanıcıya henüz değerlendirmedeği ancak ona benzeyen kullanıcıların değerlendirdiği ürünler tavsiye olarak sunulur. Öğe tabanlı iş birlikçi filtreleme tavsiye sistemlerinde ise tavsiyeler öğelerin komşuluklarına göre hesaplanır. Kullanıcı tabanlıdan farklı olarak öncelikle öğeler arasındaki benzerlikler hesaplanır daha sonra kullanıcının daha önce oyladığı öğelerin benzerleri kullanıcıya sunulur [39].

Öğe tabanlı filtreleme yöntemlerinin en kritik adımı, öğeler arasındaki benzerlikleri hesaplayıp birbirine en benzer öğeyi bulmaktır. Aralarındaki benzerliği bulacağımız öğelerin i ve j olduğunu varsayalım. Buradaki temel fikir öncelikle i ve j ye puan veren kullanıcıları ayırmaktır. Daha sonra $s_{i,j}$ benzerliğini bulmak için benzerlik bulma yöntemleri uygulanır. Bunlar arasında en popüler olanları kosinüs temelli benzerlik, korelasyon temelli benzerlik ve düzeltilmiş kosinüs benzerliğidir [40].

Kullanıcı tabanlı filtreleme yöntemlerinde ise, bir test kullanıcısının benzer kullanıcı profillerinden gelen derecelendirme bilgilerine dayanarak bir test öğesine ilgi duyabileceğini tahmin. her bir kullanıcı profili (sırt vektörü), test kullanıcısının profiline olan benzerliği ile sıralanır. Daha benzer kullanıcılar tarafından yapılan derecelendirmeler, test öğesi derecelendirmesini öngörmede daha fazla katkıda bulunur. Benzer kullanıcılar, bir eşik kullanılarak veya üst-N seçilerek tanımlanabilir [41].

2.2.2.2. Model Tabanlı İş Birlikçi Filtreleme Yöntemleri

Daha önceki değerlendirmeleri yeni tavsiye üretmek kullanan bellek tabanlı tavsiye sistemlerinin aksine model tabanlı tavsiye sistemleri bu bilgileri tahminlemeyi yapan bir model oluşturmakta kullanır. Genel fikir, kullanıcı-öge etkileşimlerini, kullanıcıların tercih sınıfı ve ögelerin kategori sınıfı gibi, kullanıcıların ve sistemdeki ögelerin gizli özelliklerini temsil eden faktörlerle modellenmesidir [42].

Model tabanlı yöntemler, denetlenen (supervised) veya denetlenmeyen (unsupervised) makine öğrenim yöntemlerinde olduğu gibi, önceden özetlenmiş bir veri modeli oluşturulur. Bu nedenle, eğitim (veya model oluşturma aşaması) tahmin aşamasından açıkça ayrılmıştır. Bu tür yöntemlere örnekler geleneksel makine öğrenmesinde karar ağaçları, kurala dayalı yöntemler, Bayes sınıflandırıcıları, regresyon modelleri, destek vektör makineleri ve sinir ağları olabilir. İlginç bir şekilde, neredeyse tüm bu modeller iş birlikçi filtreleme tavsiye sistemlerine genelleştirilebilir, tıpkı k-en yakın komşu sınıflandırıcılarının iş birlikçi filtreleme için bellek tabanlı modellere genelleştirilebilmesi gibi. Bunun nedeni, geleneksel sınıflandırma ve regresyon problemlerinin iş birlikçi filtreleme yöntemlerinin özel durumları olmasıdır [43].

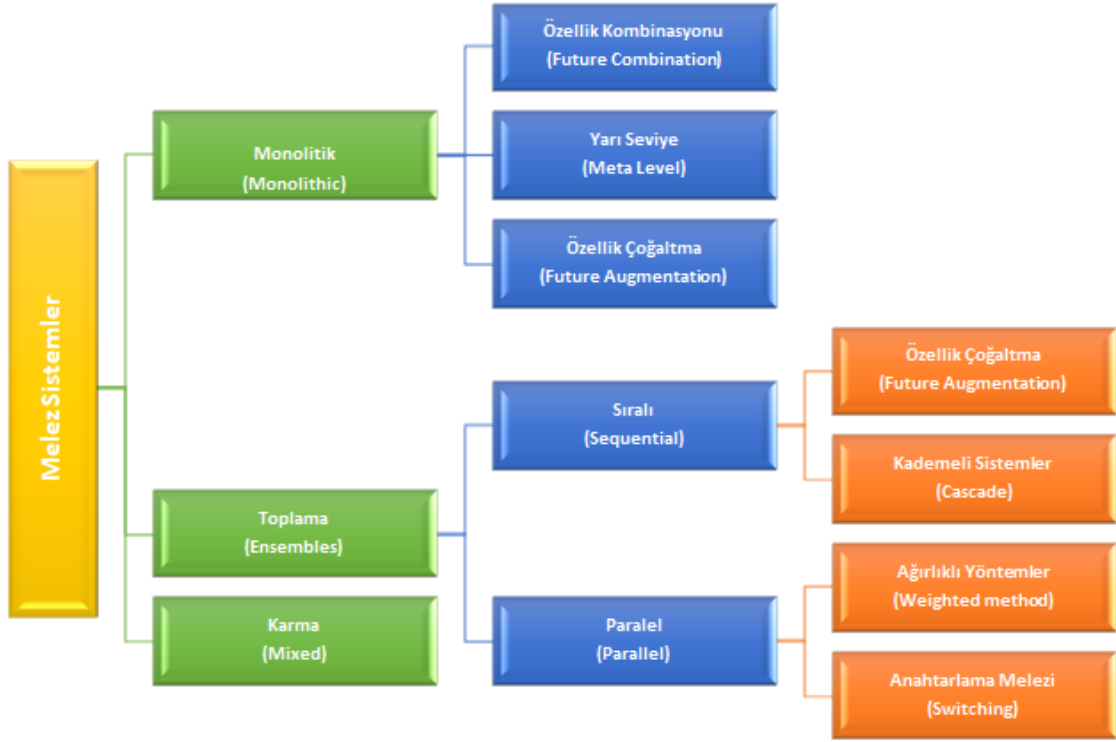
2.2.3. Melez (Hibrid) Filtreleme Tavsiye Sistemleri

Her tavsiye sisteminin güçlü olduğu ve güçsüz olduğu yönleri vardır. Bu sebeple birden fazla tavsiye sistemlerini birlikte kullanarak tavsiyede bulunmak mümkündür. Örneğin, yeni ögelerin derecelendirmesi olmadığından modelin başarısız olduğunu bildiğimiz işbirliğine dayalı filtreleme yöntemlerini, ögelerle ilgili özellik bilgilerinin mevcut olduğu içerik tabanlı sistemlerle birleştirerek yeni ögeler için daha doğru ve verimli tavsiyeler üretebiliriz. Burada amaç daha başarılı tavsiyelerde bulunmaktır [44].

Melez yaklaşımların kombinasyonu iki şekilde yapılabilir. Birincisi algoritmaların ayrı ayrı oluşturulması ve sonuçların birleştirilmesi diğeri ise iş birlikçi filtreleme de bazı içerik tabanlı algoritmaları kullanarak ya da içerik tabanlı filtreleme de bazı iş birlikçi filtreleme algoritmalarının kullanılmasıdır [21].

Melez sistemlerin en büyük avantajı, diğer tekniklerin tek tek uygulanmasından çok daha ölçeklenebilir olmasıdır. Yeni kullanıcılar için soğuk başlangıç problemi ve veri seyrekliği

problemleri daha kolay çözümlenir. Tüm bu etkenler sayesinde tavsiyelerin doğruluğu diğer sistemlere göre daha başarılıdır [44].



Şekil 2.5: Melez Tavsiye Sistemlerinin Yapısı [43].

Monolitik melez sistemler, farklı yaklaşımları aynı algoritmaya dahil eder, topluluk melez sistemleri farklı tavsiye sistemlerini çalıştırır ardından sonucu bir öneride birleştirir, karma ise bir dizi tavsiye sistemini çalıştırıp hepsinin kombinasyonunu döner [45].

2.2.3.1. Özellik Kombinasyonu (Future Combination)

Bu tür melez tavsiye sistemleri bir tavsiye sisteminin sonuçlarını ek özellik verileri olarak görür ve diğer sistemi yeni genişletilmiş veriler üzerinde kullanır. Bir iş birlikçi ve içerik tabanlı melez durumunda, sistem sadece iş birlikçi filtreleme iş birliğine dayalı veri çıkışına dayanmaz.

2.2.3.2. *Özellik Arttırma (Future Augmentation)*

Bu teknik önceki tavsiye sistemlerinden üretilen verileri kullanır. Örneğin, Libra, Amazon.com'da bulunan veriler üzerine, bir Bayes metin sınıflandırıcısı kullanarak içerik temelli önerilerde bulunur. Özellik büyütme melezleri, birincil tavsiye de bulunan ürüne daha az sayıda özellik eklediklerinden özellik kombinasyonu yöntemlerinden üstündür [21].

2.2.3.3. *Ağırlıklı Yöntem (Weighted method)*

Ağırlıklı yöntemler melez tavsiye sistemleri içinde en sık kullanılan yaklaşımdır. Ağırlıklı doğrusal fonksiyonlar kullanarak her öneri tekniğinin çıktı puanlarını toplayarak, önerdikleri öğelerin puanlarını hesaplar. İlk ağırlıklı tavsiye verenlerden biri, çevrimiçi gazeteleri önermek için iş birlikçi tabanlı ve içerik tabanlı filtreleme derecelendirme puanlarını doğrusal ağırlıklı bir şekilde birleştiren P-Tango [46] idi. P-Tango'da, her bir değerlendirmeye eşit başlangıç ağırlıkları veren ve daha sonra kullanıcıların geri bildirimleriyle adapte olacak şekilde toplama yapılmıştır. İşbirlikçi ve içerik tabanlı filtreleme tavsiye sistemlerinin ağırlıkları, sistemin her kullanıcı için en uygun karışımı belirlemesi ve kararsız kullanıcı sorununu hafifletmesi için kullanıcı bazında ayarlanmıştır [44]. Ağırlıklı bir melezin avantajı, tüm öneri sisteminin güçlerinin, öneri sürecinde basit bir şekilde kullanılmasıdır [21].

2.2.3.4. *Yarı Seviye (Meta Level)*

Bir öneri tekniği ile oluşturulan iç model başkaları için girdi olarak kullanılır. Oluşturulan model, tek bir derecelendirmeye kıyasla bilgi bakımından her zaman daha zengindir. Yarı-seviye melezleri, işbirlikçi filtreleme tekniklerinin seyreklik problemini, birinci teknikte öğrendiği modeli ikinci sisteme girdi olarak vereceğinden büyük oranda çözecektir.

2.2.3.5. *Kademeli Sistemler (Cascade)*

Bir tekniğin önerileri, başka bir tavsiye tekniğiyle iyileştirilmiştir. İlk öneri tekniği, bir sonraki öneri tekniği ile arıtılmış olan kaba bir tavsiye listesi çıkarır. EntreeC [47] kademeli bilgi tabanlı ve iş birlikçi tavsiye sistemleri kullanan kademeli melez sistem örneğidir.

2.2.3.6. Anahtarlama Melezi (Switching)

Sistem, iyi bir derecelendirme için öneri tekniklerinin birinden diğerine geçer. Anahtarlama melezi bir yöntemle özgü sorunları önleme özelliğine sahiptir; örneğin içerik tabanlı tavsiye sistemlerinin yeni kullanıcı sorununu iş birlikçi filtreleme yöntemlerine geçerek çözebilir. Bu stratejinin avantajı, sistemin kurucu danışmanlarının güçlü ve zayıf yönlerine duyarlı olmasıdır. Anahtarlama melezlerinin en büyük dezavantajı, genellikle öneri işlemine daha fazla karmaşıklık getirmesidir, çünkü normal olarak öneri sistemine parametre sayısını artıran anahtarlama ölçütü belirlenmelidir.

2.2.3.7. Karma Melez Sistemler (Mixed)

Karma melezler her bir öğeye tek bir tavsiye sunmak yerine farklı tavsiyeleri birleştirerek dönerler. Böylece her öğe kendi tercihlerine uygun, farklı tekniklerle oluşturulmuş birden fazla tavsiyeye sahiptir. Karma melez sistemlere örnek olarak, içerik tabanlı ve iş birlikçi sistemlerin tavsiyelerini birleştirerek kullanıcıya televizyon için program öneren PTV sistemini verebiliriz [21].

2.3. TAVSİYE SİSTEMLERİNDE BENZERLİK HESAPLAMA TEKNİKLERİ

Benzerlik ağırlıkları, komşuluk tabanlı tavsiye yöntemlerinde önemli rol oynar. Benzerlik ağırlıkları ile derecelendirme tahmininde kullanılan komşular seçilir ve tahminde bu komşulara az ya da çok önem verilmesi konusunda sistemi yönlendirirler. Benzerlik ağırlıklarının hesaplanması, hem tavsiye sisteminin doğruluğunu hem de performansını, etkilediğinden komşuluk tabanlı bir tavsiye sistemi kurmanın en kritik adımlarından biridir [42]. Tavsiye sistemleri ya sistemdeki kullanıcıların benzerliğini ya da sistemdeki öğelerin benzerliğini kullanarak öneride bulunurlar.

Benzerlik hesaplama teknikleri, iki kullanıcı veya öğe arasındaki mevcut bilgiyi iki vektör olarak kabul eder ve aralarındaki mesafeyi ölçmek için basit matematiksel hesaplamalar yapılırlar [44]. Pearson Korelasyonu, Kosinüs, Düzeltilmiş Kosinüs Benzerliği ve Öklid uzaklığı temel benzerlik hesaplama teknikleridir.

2.3.1. Kosinüs Benzerliği

İki madde arasındaki benzerliği hesaplamamanın bir yolu, sistemde bulunan her bir öğeyi vektör olarak ele alıp, bu vektörler arasındaki açının kosinüsünü bir benzerlik ölçüsü olarak kullanmaktır. Kosinüs benzerliği hesaplamak için gerekli olan açı vektörlerin skaler çarpımlarının, normlarına bölünmesiyle elde edilir. İki vektör arasındaki açı ne kadar küçük ise öğelerin benzerlik oranı o kadar fazladır. Kosinüs benzerliği aşağıdaki gibi hesaplanır.

$$\cos \theta = \frac{d \cdot d^*}{|d||d^*|} = \frac{\sum_{i=1}^n d_i d_i^*}{\sqrt{\sum_{i=1}^n (d_i)^2} \sqrt{\sum_{i=1}^n (d_i^*)^2}} \quad (2.1)$$

d ve d^* birbirinden farklı iki öğeyi temsil eden çok boyutlu vektörleri ve “•” vektörlerin iç çarpımını, $|d|$ ise vektörün boyunu ifade etmektedir[48]. Vektör boyutundan bağımsız olması kosinüs benzerliğinin güçlü bir yönüdür.

2.3.2. Pearson Korelasyon Katsayısı

Pearson korelasyon katsayısı, vektörler arasında benzerliği bulmak yerine bu yaklaşımda vektörler arasındaki korelasyon kullanılır. Pearson korelasyon katsayısı aşağıdaki gibi hesaplanır.

$$r = r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) \quad (2.2)$$

Burada r korelasyon katsayısı, n toplam veri sayısı, x_i x vektörünün vektör noktası y_i y vektörünün vektör noktasıdır. \bar{x} ve \bar{y} vektörlerin ortalaması, s_x ve s_y standart sapma değerleridir.

İki değişken arasındaki korelasyon katsayısını hesaplamamanın diğer bir yolu, iki değişkenin kovaryansının standart sapmalarının çarpımı ile bölünmesidir.

$$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_x \sigma_y} \quad (2.3)$$

Yapılan çalışmalar Pearson katsayısının, kullanıcı tabanlı iş birlikçi filtreleme yöntemleri uygulanan sistem için diğer benzerlik hesaplama yöntemlerinden daha iyi performans gösterdiğini göstermiştir [39].

2.3.3. Jaccard Benzerliği

Tavsiye sistemlerinde kullanılan diğer bir benzerlik hesaplama türü Jaccard Benzerliğidir. Jaccard benzerlik katsayısı, iki kullanıcı veya öge aralarındaki özelliklerin benzerliklerinin kesişiminin birleşimine oranıdır. Matematiksel olarak ifade edersek; A ve B iki vektörü olmak üzere

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2.4)$$

Jaccard benzerlik katsayısı, örnek setlerdeki benzerliği ve çeşitliliği bulmak için kullanılan bir istatistiktir. Kullanıcılar ve öğeler vektörler veya kümeler olarak temsil edilebildiğinden, kullanıcılar veya öğeler arasında benzerliği bulmak için Jaccard katsayısını öneri sistemlerinde kullanabiliriz[44]. Jaccard benzerlik katsayısı, yüksek uzaklık değerlerinde kosinüs benzerlik katsayısına, düşük uzaklık değerlerinde ise Öklid'e benzer özellik göstermektedir[48].

2.3.4. Öklid Uzaklığı

Öklid uzaklık ölçütü, iki nokta veya iki vektör arasındaki mesafeyi hesaplamak için kullanılan en yaygın benzerlik yöntemlerinden biridir. Öklid Uzaklığı, çok boyutlu uzaydaki öğelerin birbirlerine geometrik uzaklığıdır. Öğelerin konumları üzerinden ne kadar farklı oldukları incelenir. Veri seti yoğun ve birbirlerinden iyi ayrılmış kümeler içeriyor ise iyi sonuç verir. İki öge birbirine ne kadar yakın ise Öklid uzaklığı da o kadar sıfıra yaklaşır [48]. Matematiksel olarak aşağıdaki gibi ifade edilir;

$$\text{Öklid Uzaklığı}(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (2.5)$$

Öklid uzaklığının tavsiye sistemlerinde nasıl kullanıldığını anlamak için bir kullanıcı- öge derecelendirme matrisini varsayalım. İki satır arasındaki Öklid uzaklığı bize kullanıcı benzerliğini, iki kolon arasındaki Öklid uzaklığı ise bize ögeler arasındaki benzerliği verir.

2.4. TAVSİYE SİSTEMLERİNDE KARŞILAŞILAN SORUNLAR

Bir tavsiye sistemi oluşturulurken üç temel aşama vardır. Bunlar verinin toplanması, öğrenme aşaması ve tahmin veya tavsiyelerin oluşturulmasıdır. Her bir aşamada uygulanan tekniğe göre belirli problemler ile karşılaşır.

Tavsiye sistemleri manipülasyona açıktır. Bir kullanıcı grubu eğer organize olup belirli ögeleri yüksek oylarlarsa tavsiye sistemi bundan etkilenecektir. Ayrıca bir kullanıcının kendi ürünlerine pozitif değerlendirmelerde bulunup, rakiplerine kötü değerlendirmeler de bulunması da olası bir durumdur.

Veriler toplandıktan sonra, verilerden bir model oluşturulur daha sonra tavsiye ve tahminler oluşturulur. Her tavsiye sisteminin güçlü ve güçsüz olduğu yönler vardır ancak genel olarak her bir tavsiye sistemini oluştururken karşılaşılan genel sorunlar şunlardır:

2.4.1. Soğuk Başlangıç (Cold Start)

Bu problem sisteme yeni kullanıcı yada yeni öge eklendiği zaman yaşanır. Soğuk başlangıç problemini çözenin bir çok yolu vardır. İlk başta kullanıcıların bazı ögeleri oylaması istenebilir, en popüler ürünler önerilebilir yada kullanıcıların demografik özelliklerine göre yeni ürünler önerilir [49]. Örneğin, MovieLens datası için bazı başlangıç derecelendirmelerine sahip olmadan yeni filmler öneremez. Yeni kullanıcı problemini çözmek kolay değildir, çünkü benzer kullanıcıları bulmak veya bir kullanıcının tercihleri olmadan iş birlikçi filtreleme yöntemine profil oluşturmak mümkün değildir [50]. MovieLens gibi bazı sistemler bu sorunu aşmak için ilk sisteme girişte kullanıcıların tercihlerini sorarlar. İşbirlikçi filtreleme tekniğiyle oluşturulan sistemlerde yeni kullanıcı değerlendirmelerinin olmaması tek sorun değildir. Eklenen yeni ögelerin de değerlendirmeleri olmayacağından kimseye tavsiye edilmeyecektir. Bu durum “Early Rater” olarak adlandırılır [30].

2.4.2. Seyreklik (Sparsity)

Kullanıcı deęerlendirmelerine baęlı tavsiye sistemlerinde yařanan en byk sorunlardan biridir. Kullanıcıların çoęu fikirlerini belirtmek istemezler ve bu sebeple kullanıcı oęe derecelendirme matrisi seyrekleřir. Bu da benzer derecelendirmeye sahip kullanıcıları bulmayı zorlařtırır. İř birlikçi filtreleme teknięine dayalı tavsiye sistemlerinin en byk dezavantajı da budur [50].

Derecelendirme matrisindeki seyreklik arttıka; öneri sistemlerinin tavsiyede bulunmasının zorlařması yanı sıra, tahminlerin kalitesi iin de problem teřkil etmektedir. Tavsiyelerin az veriyle hesaplanması sebebiyle sistemin bařarı oranı sorgulanmaya aıktır [51]. Ayrıca yksek seyreklik problemi sebebiyle bir ok derecelendirmemiř oęe ve aktif olmayan kullanıcı vardır, bunlar da iř birlikçi tabanlı tavsiye sistemlerinde, öneri hesaplamasında gz ardı edilebilirler [52]. Bu sorunun zm olarak eksik derecelendirmeleri, derecelendirme aralıęının orta deęeri, varsayılan deęeri ya da kullanıcı/ oęe derecelendirmesinin orta deęerini kullanarak zebiliriz. Ya da daha gvenilir bir yntem olarak ierik bilgilerini kullanarak seyreklik problemini ařabiliriz. Ancak bu yntemler bařka sorunlara yol aabilirler. Örneęin eksik derecelendirmelere varsayılan deęer atamak tavsiyeler de nyargıya sebep olabilir ve ya eksik derecelendirmelere tahmin de bulunmak iin oęe ierikleri her zaman temin edilemeyebilir. Bu durumda ise boyut azaltma(dimension reduction) ya da grafik tabanlı yntemler kullanılmalıdır [42]. Temel olarak, boyut azaltma yaklařımları, sadece en alakalı kullanıcıları ve oęeleri dikkate alan daha yoęun bir kullanıcı-oęe derecelendirme matrisi oluřturarak seyreklik sorunuyla ilgilenir. Tahminler daha sonra bu azaltılmıř matris kullanılarak yapılır [53].

2.4.3. leklenebilirlik (Scalability)

Tavsiye sistemlerinin dięer bir sorunu da leklenebilirliktir nk hesaplama normalde kullanıcı ve oęe sayısıyla doęrusal olarak byr. Veri kmesi sayısı sınırlı olduęunda etkili olan bir öneri teknięi, veri kmesi hacmi artırıldıęında bařarılı öneri retemeyebilir. Bu nedenle, veri seti hacmi arttıka lekleyerek tavsiye tekniklerini uygulamak ok nemlidir. leklenebilirlik problemini zmek ve önerileri hızlandırmak iin Tekil Deęer Ayrıřtırma (SVD) ve ya boyut azaltma teknikleri kullanılır [21]. Son yıllarda, kullanıcılardan ok sayıda derecelendirme ve rtk geri bildirim toplamak giderek daha kolay hale geldi. Bu gibi

durumlarda, veri kümelerinin boyutları zamanla artmaya devam eder. Bundan dolayı büyük verilerde etkili ve düzgün tavsiyelerde bulunan sistem tasarlamak daha önemli hale gelmiştir. Bir sistemin ölçeklenebilirliğini belirlemek için eğitim süresi, tahmin süresi ve bellek gereksinimi gibi ölçütler göz önüne alınır. Derecelendirme matrisinin boyutu büyüdükçe, bellekte tutmak zorlaşır bu durumda bellek ihtiyacını en aza indirecek algoritmayı tasarlamak gerekir. Büyük veri önem kazandıkça ölçeklenebilirliğin önemi de artacaktır [43].

2.4.4. Aşırı Uzmanlaşma (Overspecialization)

Tüm sınıflandırma temelli yaklaşımların karşılaştığı diğer bir sorun da aşırı uzmanlaşmadır. Uzmanlaşmanın etkisi bir kullanıcıya zaten bildiği daha önce değerlendirdiği öğelere çok benzeyen tavsiyeler almasıdır. Bu sorunu aşmak için çok benzer öğeleri filtreleyebiliriz [54]. Aşırı uzmanlaşma için diğer çözümler ise tavsiye sistemi sürecine raslantısallığı ekleyerek ya da bireysel tavsiye çözümlerini arttırmaktır [55].

2.4.5. Gecikme (Latency)

Özellikle iş birlikçi filtreleme tabanlı tavsiye sistemleri veri tabanına yeni öğeler daha sık eklendiğinde gecikme sorunuyla karşı karşıya kalmaktadır; yeni eklenen ürünler derecelendirilmemiş olduğundan sistem sadece mevcuttaki öğeleri önermektedir. İçerik tabanlı filtrelemeyi kullanmak bekleme sürelerini azaltabilir ancak aşırı uzmanlaşmaya neden olabilir. Bu durumla başa çıkmak için, kullanıcı profili ile birlikte kategori temelli yaklaşım kullanılabilir [49].

2.4.6. Benzerlik (Synonymy)

Gerçek hayat senaryolarında aynı ya da çok benzer ürünler sistem üzerinde farklı şekilde tanımlanmış olabilir [56]. Birçok tavsiye sistemi bu gizli ilişkiyi bulamaz ve bu ürünleri farklı ürünlermiş gibi değerlendirir. Örneğin “children movie “ ve “children film” aslında aynı öğedir ancak sistemde farklı iki ürün olarak ele alınıyor. Ancak bellek temelli iş birlikçi filtreleme sistemleri benzerliği hesaplamak için aralarında bir eşleşme bulamazlar. Benzer öğelerin yaygınlığı, tavsiye sistemlerinin performansını azaltır. Tam otomatik yöntemlerin dezavantajı, bazı ilave terimlerin amaçlanandan farklı anlamlara sahip olabileceği ve bu nedenle tavsiye performansının hızlı bir şekilde düşmesine yol açmasıdır [57].

2.4.7. Kararsız Kullanıcılar (Gray Sheep)

İnsanların zevkleri genel olarak kolay tahmin edilemez. Bir şeyi değerlendirirken çok farklı düşünebilirler. İş birlikçi filtreleme yöntemlerinde korelasyon katsayısı iki kullanıcı arasındaki benzerliği hesaplarken benzeyenleri 1, farklı olanları ise -1 olarak hesaplar. Bu korelasyon katsayısına göre kullanıcılar iki ana gruba ayrılabiliriz. Kararlı kullanıcılar (White sheep) diğer kullanıcılar ile yüksek korelasyon katsayısına sahip olanlar ve kararsız kullanıcılar (gray sheep) çoğunluk fikrine katılmayan düşük korelasyon katsayısına sahip olanlar [58]. Kararsız kullanıcıların düşük veya orta büyüklükte bir sistemde olması hem onların doğru tavsiyeleri almamasına hem de ileri vadede tüm sistemdeki tavsiyelerin başarı oranının azalmasına sebep olmaktadır [59]. Bu durumu çözmek için güvene bağlı sistemleri kullanabiliriz. İçerik tabanlı filtrelemeler bu sorunu kullanıcının özelliklerine ya da öğelerin içerik özelliklerinden faydalanarak çözebilir. Kararsız kullanıcılar yani kararsızlar, k-ortalama kümeleme gibi çevrimdışı kümeleme tekniklerini uygulayarak tanımlanabilir ve diğer kullanıcılardan ayrılabilir. Bu şekilde sistemin performansı ve başarı oranı bu sorundan etkilenmez [49].

2.4.8. Yanlış Yönlendirme (Shilling Attacks)

Bir tavsiye sisteminin güvenilirliği çok önemlidir, sistem ne kadar tutarlı tavsiyede bulunursa o kadar başarılı sayılır. Ancak bazı kişiler sistemlere saldırıda bulunarak kullanıcıların yanlış tavsiyeler almasına sebep olmaktadır. İş birlikçi filtreleme yöntemlerinde saldırılarla aslında kullanıcı için uygun olmamasına rağmen öğeler tavsiye edilir ya da bir öğenin kullanıcıya önerilmemesine sebep olarak sistem zayıflatılır. Tavsiye sistemleri üzerinde yaygın olarak ürün itme (product push) ve ürün nükleer saldırı (product nuke) stratejileri uygulanır. Ürün itme saldırısı ile hedeflenen öğelerin tahmin değeri arttırılır, ürün nükleer saldırı altındayken hedefteki ürünlerin tahminleri azaltılır [60]. Özellikle e-ticarette kullanıcılar kendi ürünlerini sahte profiller oluşturularak yüksek oylar vererek sistemi manipüle ederler.

Rastgele saldırıda dolgu maddesine rastgele derecelendirmeler atanırken, hedef maddeye önceden belirlenmiş derecelendirme verilir. Eğer bu bir ürün itme saldırısı ise öğeye en yüksek puan verilir, ürün nükleer saldırısı ise en düşük puan atanır. Bu saldırıyı gerçekleştirmek için az bilgi gerektiğinden düşük bilgi saldırısıdır.

Ortalama saldırıda her bir dolgu ögesi ortalama olarak derecelendirilir, enjekte edilen profiller mevcut kullanıcılara benzerlerdir. Bu modelin uygulanması için az bilgi gerekmesine rağmen rastgele saldırıdan daha etkilidir [61].

Çoğunluk saldırısında, saldırgan hedef ögeyi sık sık oy alan öğelerle ilişkilendirmeye çalışır. Saldırgan genellikle diğer öğelere rastgele derecelendirme yapar ve çok popüler öğelere maksimum puan verir. Bu da enjekte kullanıcılarının diğer kullanıcılarla olan benzerliğini artırır.

Segment saldırısı sistem hakkında az bilgi gerektirir. Buradaki asıl amaç belirli öğeleri hedeflenen kullanıcı grubu için (örneğin macera filmlerini seven kullanıcılar) önerilir yapmaktır. Saldırı ile birlikte kullanıcılara ilgi alanı olmayan öğeler önerilebilir. Bu da sisteme olan güveni sarsmaktadır. Segment saldırısı özellikle öge tabanlı iş birlikçi filtreleme yöntemleri için kullanılır. Burada benzer kullanıcılar yerine benzer öğelerin komşulukları hesaplanır [62].

Rastgele ve ortalama saldırı teknikleri, hedef ögeye maksimum derecelendirme yerine minimum derecelendirmeyi vererek bir maddeyi indirgemek veya iptal etmek için zorlanabilir. Ancak bir öğenin itilmesinde etkili olan teknikler, nükleer saldırılar için mutlaka etkili değildir. Bu nedenle, bazı saldırı modelleri araştırmacılar tarafından özellikle bir ögeyi indirgemek için tasarlanmıştır. Örneğin, beğeni / nefret saldırısında ise hedef ögeye en düşük puan, diğer dolgu öğelere en yüksek puanlar verilir [60].

2.4.9. Gizlilik (Privacy)

Bazı tavsiye sistemleri, öneride bulunmak için sistem üzerinde kullanıcı bilgilerine ihtiyaç duymaktadırlar. Kullanıcıların bazıları bilgilerini paylaşırken bazıları ise paylaşmak istemezler. Genel olarak kullanıcılar bu konuda üç gruba ayrılırlar. Gizliliğe önem verenler kişisel bilgilerini paylaşmaya karşıdırlar. Pragmatik olanlar ise kişisel verilerini paylaşma konusunda kontrollü davranırlar ve belirli kişisel verileri koruma ölçülerine dikkat ederler. Marjinal olanlar ise genel olarak kişisel verileri paylaşmak konusunda çekinmezler [63]. Son zamanlarda birçok çevrimiçi gizlilik ihlali vakası olmuştur. Örneğin Facebook 2010 yılında bazı uygulamalarının kullanıcıların kişisel verilerini reklam verenler ile paylaştığını açıkladı. Mart 2011'de, Kaliforniya merkezli sigortacı HealthNet, yaklaşık 2 milyon kullanıcısının

adlarını, adreslerini, sağlık ve finansal verilerini ortaya çıkaran bir gizlilik ihlali olduğunu açıkladı [60].

Tavsiye sistemlerinde gizlilik riski temel olarak, sistemlerin kullanıcı bilgilerini üzerinde tutmasından kaynaklanır. Sistem kişiye özel tavsiyede bulunabilmek için kişi hakkında bazı bilgilere ihtiyaç duyar bu aynı zamanda verinin üçüncü şahıslardan korunması yükümlülüğünü getirir. Bu gizlilik- kişiselleştirme takası olarak anılır [64].

Modern bilgi çağında insanların kişisel bilgilerini, bazı kişisel kazanımlar karşılığında paylaşmaya açık olduğu biliniyor. Tavsiye sistemleri bunun için uygun bir örnektir çünkü girdi olarak kullanıcı bilgilerini kullanır ve onlara hizmet veya ürün sunar. Ekonomik İş Birliği ve Kalkınma Örgütü (OECD), veri toplayıcıların topladığı verilerin güvenliğinden emin olmak için Adil Bilgi Uygulaması (FIPS) diye bir uygulama tanımlamıştır [65]. Gizlilik ihlalleri için ilk başta incelenmesi gereken konu ihlalin var olan verilere doğrudan erişim ile mi ilgili olduğu yoksa yeni verilere erişirken mi oluştuğudur. FIPS de tanımlanan ilkelere saldırı üç şekilde yapılır. İlk olarak tavsiye sistemleri kullanıcı ile etkileşime girer ancak sistem kullanıcının gizlilik beklentilerini karşılamayabilir (veri toplama ve kullanım sınırı ihlali). İkinci olarak sistemin kullanıcılarının, diğer kullanıcıların özel bilgilerine erişimi yoktur ancak tavsiye edenin çıktısından hedef kullanıcının bilgilerini açığa çıkarabilirler (Güvenlik Önlemleri ilkesini ihlali). Ve son olarak dış etkenler sistemin kullanıcısı değildir, ancak sistem tarafından tutulan bilgilere erişmeye çalışabilir veya bu bilgilere erişebilmek için sistem ile kullanıcıları arasındaki etkileşime müdahale edebilir (Korunma ilkesi) [64]. Bu sebeple tavsiye sistemleri ile paylaşılan bilgiler sınırlı gibi görünse de hem sistemden kaynaklı sorunlar hem de sistemi suiistimal eden dış etkenler sebebiyle çeşitli riskler içermektedir. Bir tavsiye sistemi için veri toplamak ne kadar önemliyse verilerin güvenliğini sağlamak da bir o kadar önemli olup sistem tasarımına güvenlik maddesinin de eklenmesi gerekmektedir.

2.5. TAVSİYE SİSTEMLERİ DEĞERLENDİRME ÖLÇÜTLERİ

Bir tavsiye sisteminin tavsiyelerinin doğruluk ve kapsama oranı, farklı ölçüm türleri kullanılarak değerlendirilebilir. Kullanılan metrik türü, tavsiye sisteminin tekniğinin türüne bağlıdır. Tavsiye sistemlerinin doğruluğunu ölçmek için kullanılan ölçütler istatistiksel ve

karar destek doğruluğu ölçütlerine bölünmüştür. Her metriğin uygunluğu, veri setinin özelliklerine ve tavsiye veren sistemin türüne bağlıdır [66].

Tavsiye sistemleri çevrimiçi ve çevrim dışı olmak üzere iki şekilde değerlendirilirler. Çevrimiçi bir sistemde, sunulan tavsiyeye göre kullanıcı tepkileri ölçülür. Bu nedenle, çevrimiçi sistemlerde kullanıcı katılımı çok önemlidir. Örneğin, bir alışveriş sitesinin sisteminin çevrimiçi değerlendirmesinde, önerilen ürünlere tıklayan kullanıcıların dönüşüm oranı ölçülebilir. Bu test yöntemleri A / B testi olarak adlandırılır ve tavsiye sisteminin doğrudan etkisini ölçer. Bununla birlikte, çevrimiçi değerlendirmeler aktif kullanıcı katılımı gerektirdiğinden, bu değerlendirmeleri karşılaştırma ve araştırmada kullanmak genellikle mümkün değildir. Büyük ölçekli kullanıcı kitlesi olan sistemlerden kullanıcı dönüşüm verilerine erişimde ciddi zorluklar vardır. Diğer taraftan, birden fazla veri seti üzerinde test yapılması, tavsiye edilen sistemin daha fazla genelleme gücünün sağlanması yani eğitimi için özellikle önemlidir. Algoritmanın çeşitli ayarlar altında çalıştığı bu gibi durumlarda, geçmiş veri setleri ile çevrimdışı değerlendirmeler kullanılır. Çevrimdışı yöntemlerle çalışırken, doğruluk ölçütü tek başına yetersiz kalabilir, bu sebeple ikincil bir değerlendirme ölçütü de önemli rol oynamaktadır. Çevrimdışı testlerde, derecelendirme gibi geçmiş veriler kullanılır. Bu şekilde kullanılan veri setlerinin en bilinen örneklerinden biri Netflix Prize veri setidir[67]. Bu veri seti ilk olarak bir çevrimiçi yarışma kapsamında yayınlandı ve o zamandan beri birçok algoritmayı test etmek için standart bir kıstas olarak kullanılmaktadır. Geçmiş veri setlerinin kullanımının ana avantajı, geniş bir kullanıcı tabanına erişim gerektirmemeleridir. Bir veri seti toplandıktan sonra, çeşitli algoritmaları çeşitli ortamlarda karşılaştırmak için standartlaştırılmış bir ölçüt olarak kullanılabilir. Ayrıca, tavsiye sisteminin geliştirilebilirliğini test etmek için çeşitli alanlardan (örneğin, müzik, filmler, haberler) elde edilen çoklu veri setleri kullanılabilir. Çevrimdışı değerlendirmeler en popüler değerlendirme şeklidir. Çevrimdışı değerlendirmelerin en büyük dezavantajı, kullanıcının gelecekte danışman sisteme tepki verme gerçek eğilimini ölçmemeleridir. Örneğin, veriler zaman içinde değişebilir ve mevcut öneriler durumu yansıtmayabilir [43].

2.5.1. Doğruluk (Accuracy)

Model performansı için en yaygın kullanılan ölçü, doğru tahminlerin toplam tahminlere olan oranı olarak tanımlanan doğruluktur. Tavsiye sistemlerinin doğruluğunu ölçmeye yönelik metrikler istatistiksel ve karar destek doğruluğu olmak üzere ikiye ayrılır. İstatistiksel

doğruluk metrikleri, öngörülen derecelendirmeleri doğrudan gerçek kullanıcı derecelendirmeleri ile karşılaştırarak filtreleme tekniğinin doğruluğunu değerlendirir. Ortalama Mutlak Hata (MAE), Ortalama Kök Kare Hatası (RMSE) ve Korelasyon genellikle istatistiksel doğruluk ölçütleri olarak kullanılır. MAE en popüler ve en çok kullanılanıdır; tavsiyenin, kullanıcının verdiği değerden sapmasının bir ölçüsüdür [40]. RMSE metriği, tavsiye yöntemlerinin etkinliğini değerlendirmek için kullanılan bir başka iyi bilinen metriktir. Bu metrik, mutlak hataların gerçek ve öngörülen derecelendirme değerleri arasındaki katkısını gösterir [68]. MAE ve RMSE aşağıdaki gibi hesaplanır.

$$MAE = \frac{1}{N} \sum_{u,i} |p_{u,i} - r_{u,i}| \quad (2.6)$$

$$RMSE = \sqrt{\frac{\sum_{u,i} (p_{u,i} - r_{u,i})^2}{N}} \quad (2.7)$$

Burada $p_{u,i}$; u kullanıcısının i ögesine olan tahmini değerlendirme, $r_{u,i}$ mevcuttaki değerlendirme değeri N ise öge setindeki toplam değerlendirme sayısıdır.

MAE ve RMSE değerinin düşük olması tavsiye değerinin doğruluğu açısından önemlidir. Yaygın olarak kullanılan karar destek doğruluğu ölçümleri, Tersine Dönme Hızı, Ağırlıklı Hatalar, Alıcının Çalışma Karakteristikleri (ROC) ve Hassasiyet - Duyarlılık Eğrisi (PRC), Hassasiyet, Duyarlılık ve F ölçümüdür. Bu ölçütler, kullanıcıların sistemdeki öge grubundan yüksek kalitede ürünler seçmelerine olanak sağlar [69]. Hassasiyet tavsiye edilen öğelerin tüm öğelere oranıdır. Duyarlılık kullanıcının seçtiği öğelerin kendisine sunulma oranıdır. F ölçütü ise duyarlılık ve hassasiyetin tek bir ölçütte hesaplanmasını sağlar.

$$Doğruluk = \frac{\text{Doğru tahmin edilen öge sayısı}}{\text{Toplam öge sayısı}} \quad (2.8)$$

$$Hassasiyet = \frac{\text{Doğru tahmin edilen sayısı}}{\text{Toplam tahmin sayısı}} \quad (2.9)$$

$$Duyarlılık = \frac{\text{Doğru tahmin edilen öge sayısı}}{\text{Seçilen öge sayısı}} \quad (2.10)$$

$$F \text{ ölçütü} = \frac{2 * \text{hassasiyet} * \text{duyarlılık}}{\text{hassasiyet} + \text{duyarlılık}} \quad (2.11)$$

Hata matrisini incelemek için çevrimiçi haber öneren bir sistemi içerelim. Sistem toplamda 50 web sayfası içeriyor ve X kullanıcıasına 35 tane tavsiyede bulunuluyor. Kullanıcı 25 tane linke tıklarken kalan 10 linke tıklamıyor.

Tablo 1: Çevrim İçi Haber Öneri Sonuçları.

		Tercih Edilen	
		Doğru	Yanlış
Tavsiye Edilen	Pozitif	25	10
	Negatif	10	5

Burada tabloda alt satırda bulunan 5 değeri X kullanıcıasına önerilmeden incelediği web sayfası sayısı, 10 değeri ise kullanıcıya henüz önerilmeyen kullanıcının da incelemediği web sayfası sayısıdır. Tablodaki değerlerin karşılıklarını incelersek;

- **Doğru Pozitif (TP):** Sistemce pozitif olarak tavsiye edilen ve gerçek değer ile uyuşan durumları ifade eder örnekte kullanıcıya önerilen ve X kullanıcısının ilgisini çeken 25 değeri,
- **Doğru Negatif (TN):** Sistemce negatif olarak tavsiye edilen ve gerçekte değeri negatif olan değerlerdir, tabloda kullanıcıya önerilmeyip kullanıcının da ilgilenmediği 10 değeridir,
- **Yanlış Pozitif (YP):** Sistemce olumlu değerlendirilip gerçekte negatif olan durumlardır. Tabloda kullanıcıya önerilen ama kullanıcının ilgilenmediği 10 web sayfasına karşılık gelir.
- **Yanlış Negatif (YN):** Sistemce olumsuz değerlendirilip gerçekte olumlu olan durumlardır. Tabloda kullanıcıya önerilmeyen ama kullanıcının ilgisini çekmeyen 5 değerini ifade eder.

Yeni terimler ile formüller aşağıdaki gibi ifade edilir.

$$\text{Hassasiyet} = \frac{TP}{TP + FP} \quad (2.12)$$

$$\text{Duyarlılık} = \frac{TP}{TP + YN} \quad (2.13)$$

Bir tavsiye sistemini değerlendirirken hem hassasiyet hem de duyarlılık göz önüne alınmalıdır. Yüksek hassasiyet ile tavsiye üreten sistemimizin duyarlılık değeri çok başarılı olmayabilir. Bu durumda sistem gereksinimlerine uygun olarak hassasiyet ve duyarlılık arasında ideal dengeyi sağlamak gerekir [44].

2.5.2. Kapsama Oranı (Coverage)

Bir tavsiye sistemi yüksek hassasiyete sahip olsa bile bazı öğeleri önermeyebilir yada bazı kullanıcılara tavsiyede bulunmayabilir. Bu metrik kapsama oranı olarak tanımlanır. Bu aslında derecelendirme matrislerinin seyrek olmasının bir sonucudur. Örneğin bir derecelendirme matrisinin her satırında tek bir değer olduğu durumlarda anlamı bir tavsiye mümkün değildir. Çoğu tavsiye sisteminde boş olan değerlere ortalama değer verildiğinden tavsiye sistemlerinin kapsama oranı %100 dür [70]. Kullanıcı ve öğe bakımından olmak üzere şekilde hesaplanır.

2.5.2.1. Öğe Kapsamı

Genel olarak tavsiye sisteminin önerebileceği öğelerin oranı olarak tanımlanır ve katalog kapsamı diye de ifade edilir. En basit ifadeyle öğe kapsamı ölçüsü tavsiye edilebilecek öğelerin yüzdesidir. Öğe kapsamının diğer bir ölçütü de bir tavsiye sistemi kullanılarak farklı öğelerin nasıl seçildiğini gösteren satış çeşitliliğidir [42].

$$\text{Katalog Kapsamı} = \frac{\text{Tavsiye edilen öğe sayısı}}{\text{Toplam öğe sayısı}} \quad (2.14)$$

2.5.2.2. Kullanıcı Alanı Kapsamı

Kullanıcı alanı kapsamı, sistemin en az k derecelendirmesini önerebileceği kullanıcıların oranını ölçer. K değeri tavsiye listesinin beklenen büyüklüğüne göre ayarlanmalıdır. Sistem

bir kullanıcı için k değerinden daha az derecelendirme öngörülebiliyorsa, kullanıcıya k boyutlu anlamlı bir tavsiye listesi sunmak mümkün değildir. Bu durum, bir kullanıcı diğer kullanıcılarla ortak olarak çok az derecelendirmeye sahipse ortaya çıkabilir. Bir kullanıcı tabanlı komşuluk algoritması kullandığımızı varsayarsak; diğer kullanıcılarla karşılaştırdığımızda az sayıda derecelendirmesi olması nedeniyle, bu kullanıcının eşlerini hesaplamak zordur. Bu nedenle, bu kullanıcı için yeterli tavsiyede bulunmak genellikle zordur. Farklı algoritmalar farklı kapsama seviyelerine sahip olabilir ve bir kullanıcının kapsamı, her algoritmayı çalıştırarak ve tahmin yapılan öğelerin sayısını belirleyerek tahmin edilebilir [42].

2.6. TAVSİYE SİSTEMLERİ İLE İLGİLİ ARAŞTIRMALAR

Bu bölümde araştırma konusu ile ilgili olarak daha önceden yapılan çalışmalara ve çalışmaların sonuçlarına yer verilecektir.

Kulkarni ve arkadaşları (2017), Apache Spark aracılığıyla model tabanlı iş birlikçi filtreleme yöntemi kullanarak bir kitap tavsiye sistemi oluşturuyor. Çalışmanın amacı matris faktörizasyon yöntemi ile hızlı ve başarılı tavsiyelerde bulunmaktır. Bu çalışmada yeni kullanıcı problemi popülerlik modeli uygulanarak çözülüyor [71].

Panigrahi, Lenka ve Stitipragyan (2016) MovieLens veri setini kullanarak Apache Spark ile kullanıcı odaklı iş birlikçi filtreleme yöntemi için yeni bir algoritma denemiştir. Alternatif en küçük kare gibi boyutsallık azaltma teknikleri ve K-Means gibi kümeleme teknikleri ile ölçeklenebilirlik ve soğuk başlangıç probleminin etkilerini azaltmaya çalışmışlardır [72].

Lee ve arkadaşları (2005) yaptıkları çalışma ile derecelendirme değerlerinin kolay hesaplanmadığı market sepet analizini ele almışlardır. Konuyu iki sınıflı bir sınıflandırma problemi olarak düşünerek lojistik regresyon modelini uygulamışlardır. Gerçek veri setini kullanarak yaptıkları gözlemlerde tavsiyelerde başarı sonuçları artarken, hesaplama süresi ile ilgili sorunlar yaşadıklarından çevrimdışı sistemlerde kullanılmasını önermişlerdir [73].

Aggarwal ve Bahuguna (2017), kullanıcı değerlendirmeleri kullanarak öneri üreten sistemlerin hepsinin yaşadığı yeni kullanıcı probleminin çözümü için kullanıcıların demografik (yaş, cinsiyet, meslek vb.) özelliklerini kullanmayı önermiştir. Bu çalışma için MovieLens veri seti

kullanılmıştır. Sonuç olarak, MovieLens veri kümesindeki demografik verilerin kullanıcıların derecelendirmelerini farklı şekilde etkilemediği gözlemlenmiştir [74].

Aynı şekilde Safoury ve Salah (2013) da yeni kullanıcılar için soğuk başlangıç problemi için kullanıcı demografik özelliklerini kullanmayı denemişlerdir. MovieLens datasıyla yaptıkları çalışmada kullanıcı özelliklerini kullanarak yaptıkları önermelerin kullanıcı oyları ile benzer sonuçlar gösterdiğini ortaya koymuştur [75].

Sedhain ve arkadaşları (2014) soğuk başlangıç problemine çevrimiçi satış ortamında henüz hiç bir ürün almamış ancak site ile belli bilgilerini paylaşmış (yaş, cinsiyet vb.) veya sosyal ağ bilgilerini paylaşmış kişiler için çözüm önerisinde bulunmaktadır. Yapılan çalışmalar sonun da Facebook sayfası beğenilerinin soğuk başlangıç önerisi için kullanılmasının ortalama hassasiyette üç kat artış sağladığını görmüşlerdir [76].

Majó ve Révész (2011) yaptıkları çalışma ile iş birlikçi filtreleme yöntemlerinin pazarlama üzerindeki etkilerini araştırmışlardır. Tüketici tercihlerinin ölçülmesi için uygun algoritmaların kullanılmasının önemini vurgulamışlardır. Ayrıca gizlilik sorunları gibi, pazarlamanın hala cevap bulması gereken bazı soruları gündeme getirmişlerdir [77].

Sarwar ve arkadaşları (2001) farklı öge tabanlı algoritmaları analiz ederek öge-öge benzerliğini hesaplamak için çeşitli teknikler kullanmışlardır. Bu çalışmalardan elde edilen sonuçları K- en yakın komşu algoritması ile karşılaştırmışlardır. Sonuç olarak öge tabanlı algoritmaların kullanıcı tabanlı algoritmalarından çok daha iyi performans sunmuştur [40].

İşbirlikçi filtreler kullanıcıların tercihlerinin olduğu ya da ilgili öğelerin bilgilerinin olduğu veri tabanlarını kullanırlar. Breese, Heckerman ve Kadie (1998) vektör tabanlı benzerlik hesaplama yöntemleri, istatistiksel Bayes metodları ve korelasyon katsayılarına dayanan teknikleri karşılaştırmışlardır. En iyi performansı Bayesian ağlarından almışlardır. Tavsiye sistemlerinin en önemli noktalarından biri de veri tabanı boyutları ve tavsiye sisteminin ne kadar sürede tahmin ürettiğidir. Çalışma sonucunda Bayesian ağlarının diğer bellek tabanlı tavsiye sistemlerine göre (korelasyon vb.) hem daha küçük boyutlu veri tabanı kullandığını hem de daha hızlı tahmin ürettiğini görmüşlerdir [78].

Dror, Koenigstein ve Koren (2011), Yahoo Müzik'in on yıldan fazla bir süredir toplanan bir milyondan fazla kullanıcı, 600 bin müzik ve 250 milyondan fazla derecelendirmeden oluşan

veri setini inceleyerek seyrek ama büyük ölçekli veri analizine örnek bir çalışma yapmışlardır. Oluşturdukları modelle matris faktörizasyonu ile öge popülerliğini birleştirmişlerdir [79].

Matris faktörizasyonu kullanılarak yapılan birçok tavsiye sistemi kapalı bildirimleri kullanarak kurgulanır. Kapalı bildirimlerde kullanıcılar direk olarak belirli bir yelpaze içinde belirtirler. Fakat gerçek dünyada direk kullanıcı oyları genel olarak bulmak zordur. Bunun yerine genel olarak açık bildirimler kullanılır, bu bildirimler; görüntülenme sayısı, görüntüleme geçmişi vb. bilgilerdir. Pacula (2009) bu çalışmada LastFm verisini kullanarak kullanıcıların şarkıları dinleme sıklıklarını açık bildirim olarak ele alıp bir tavsiye sisteminin başarısını incelemiştir [80].



3. MALZEME VE YÖNTEM

Bu bölümde tez çalışmasında kullanılan araştırma modeli, çalışma sırasında kullanılan veri toplama araçları, verilerin toplanması ve analizi konularında bilgiler verilmiştir.

3.1. ÇALIŞMANIN AMACI VE HEDEFİ

Bu çalışmanın amacı, özel bir bankanın müşterilerinin kredi kartı alışverişlerinden yola çıkarak kişiyi, iş birlikçi filtreleme yöntemiyle oluşturulan tavsiye sistemiyle en uygun kampanyaya yönlendirmektir. Çalışmanın sonucunda ise müşteri memnuniyetinin artması hedeflenmektedir.

3.2. ARAŞTIRMANIN MODELİ

Bu araştırma kapsamında müşterilerin kredi kartı işlemleri, harcama yaptıkları sektöre göre incelenip oluşturulan tavsiye sistemi ile kendilerine en uygun kampanyalar önerilmiştir.

Bankadan alınan veriler araştırmaya uygun olarak analiz edilmiş, iş birliği tabanlı filtreleme yöntemi ile oluşturulan tavsiye sistemine girdi olarak verilmiştir.

3.3. VERİNİN ELDE EDİLMESİ VE HAZIRLANMASI

Araştırma kapsamında özel bir bankadan temin edilen 4997 adet müşterinin 5 aylık süre içerisinde yaptıkları 662.088 kredi kartı işlemi analiz edilmiştir. Bankadan excel formatında alınan ve müşteri numarasının şifrelenerek verildiği dosyada yer alan bilgiler aşağıdaki gibidir;

- Şifrelenmiş Müşteri Numarası
- Yaş
- Çalışma Durumu
- Medeni Hal
- Eğitim Durumu
- Cinsiyet
- Müşteri Segmenti

- İşlem Tarihi
- İşlemin yapıldığı sektör kodu
- İşlem tutarı
- Şehir

Şekil 3.1’de örnek veri verilmiştir.

348170214871	40 özel Sektör Evli	Üniversite Erkek	Varlıklı	2016-07-2	5411	23.9	ANKARA	8	Market Harcamaları
3581836315881	35 özel Sektör Evli	Üniversite Kadın	Kitlesel	2016-07-2	7230	570	ISTANBUL	10	Sağlık ve Kişisel Bakım
348290154892	45 özel Sektör Bekar	Üniversite Kadın	Kitlesel	2016-07-2	3047	303.98	ISTANBUL	13	Tatil ve Seyahat
3237795762377	30 özel Sektör Bekar	Üniversite Kadın	Kitlesel V	2016-07-2	5411	18.75	LEFKOSA	8	Market Harcamaları
3301271373021	38 özel Sektör Bekar	Üniversite Kadın	Kitlesel	2016-07-2	7298	77.5	ANKARA	10	Sağlık ve Kişisel Bakım
389812348918	39 özel Sektör Bekar	Üniversite Erkek	Kitlesel	2016-07-2	5541	30	BURSA	9	Otomobil ve Ulaşım
3763236467623	41 özel Sektör Evli	İlkokul Erkek	Varlıklı	2016-07-2	5411	6.55	TEKIRDAG	8	Market Harcamaları
3751666007561	32 Kamu çalıştı: Evli	Üniversite Erkek	Kitlesel V	2016-07-2	5411	63.63	KIBRIS	8	Market Harcamaları
3184397331834	34 özel Sektör Bekar	Yüksek Lisans Erkek	Kitlesel V	2016-07-2	5812	30	ANKARA	15	Yeme ve İçme
3875659298765	28 Üniversite Evli	Üniversite Kadın	Kitlesel	2016-07-2	5813	34	ISTANBUL	7	Hobi ve Eğlence
314230471432	44 özel Sektör Evli	Üniversite Kadın	Kitlesel	2016-07-2	5411	23.9	MUGLA	8	Market Harcamaları
3911672989161	38 özel Sektör Bekar	Lise Kadın	Kitlesel V	2016-07-2	5945	139.9	ISTANBUL	1	Çocuk
3922969969292	42 özel Sektör Evli	Üniversite Erkek	Kitlesel V	2016-07-2	5411	5.9	ISTANBUL	8	Market Harcamaları
35933575933	47 özel Sektör Evli	Üniversite Erkek	Kitlesel	2016-07-2	5411	5	IZMIR	8	Market Harcamaları
3498033214908	34 özel Sektör Bekar	Üniversite Erkek	Kitlesel	2016-07-2	5411	100.4	IZMIR	8	Market Harcamaları
3342224083422	38 özel Sektör Evli	Lise Erkek	Kitlesel	2016-07-2	7011	117.4	KOCAELI	13	Tatil ve Seyahat
3519625745169	32 özel Sektör Evli	Üniversite Erkek	Kitlesel	2016-07-2	5812	36	IZMIR	15	Yeme ve İçme
3352805403582	35 özel Sektör Evli	Lise Kadın	Kitlesel	2016-07-2	5651	64.75	ISTANBUL	6	Giyim ve Aksesuar
388518815	40 özel Sektör Evli	Lise Kadın	Kitlesel	2016-07-2	5651	16.99	ISTANBUL	6	Giyim ve Aksesuar
321491882194	38 özel Sektör Evli	Üniversite Erkek	Kitlesel	2016-07-2	7230	65	ISTANBUL	10	Sağlık ve Kişisel Bakım
3571750225771	37 özel Sektör Evli	Üniversite Erkek	Kitlesel	2016-07-2	5411	37.14	AYDIN	8	Market Harcamaları

Şekil 3.1: Çalışmada Kullanılan Örnek Veri seti.

Veriler içerisinde bulunan sektör kodu exceldeki MCC kolonuna denk gelmektedir. Merchant Category Code (MCC), kredi kartı kullanmayı kabul eden her ticari işletmeye kredi kartı ağları (Visa, MasterCard, American Express) tarafından atanan dört haneli bir numaradır. Bu sayı içinde bulunduğu iş veya hizmet türünü belirtir. Ticari kategori kodları, belirli satın alma türlerini kategorilere ayırmak, izlemek veya sınırlamak için kredi kartı sağlayıcıları tarafından kullanılır[81].

SEKTÖR	SEKTÖR (MCC) KODLARI
EĞİTİM SEKTÖRÜ	8241
EĞİTİM SEKTÖRÜ	8299
OTOMOTİV SEKTÖR	5511
OTOMOTİV SEKTÖR	5521
OTOMOTİV SEKTÖR	5533
OTOMOTİV SEKTÖR	5532
OTOMOTİV SEKTÖR	7535
OTOMOTİV SEKTÖR	7538
OTOMOTİV SEKTÖR	7531
OTOMOTİV SEKTÖR	5940
OTOMOTİV SEKTÖR	7542
OTOMOTİV SEKTÖR	5046
OTOMOTİV SEKTÖR	5983
OTOMOTİV SEKTÖR	5085
OTOMOTİV SEKTÖR	5599
OTOMOTİV SEKTÖR	5571
OTOMOTİV SEKTÖR	5013
OTOMOTİV SEKTÖR	5969
OTOMOTİV SEKTÖR	7534
KIRTASIYE HOBİ OYUNCAK	5970
KIRTASIYE HOBİ OYUNCAK	8675

Şekil 3.2: Örnek Sektör Kodları [82].

Excel dosyası ile alınan bilgiler Microsoft SQL Server Management Studio 2016 ya excelden aktarım yöntemi ile eklenerek bir veri tabanı oluşturulmuştur. Bu veri tabanı üzerinde oluşturulan iki tablo ile exceldeki verilerin müşteriler ve işlemler olmak üzere iki ayrı tabloya aktarımı sağlanmıştır. Çalışma sırasında kolaylık olması açısından şifrelenen müşteri numaralarına 1 - 4997 arası değerler set edilmiştir. Veri tabanındaki birçok sektör kodu (MCC) gruplandırılarak bankadan temin edilen kampanya başlıkları altında toplanmıştır. Gruplandırılmış MCC kodları (PFM) aşağıdaki gibidir;

- Çocuk
- Diğer Harcamalar
- Eğitim
- Ev,
- Fatura Ödemeleri,

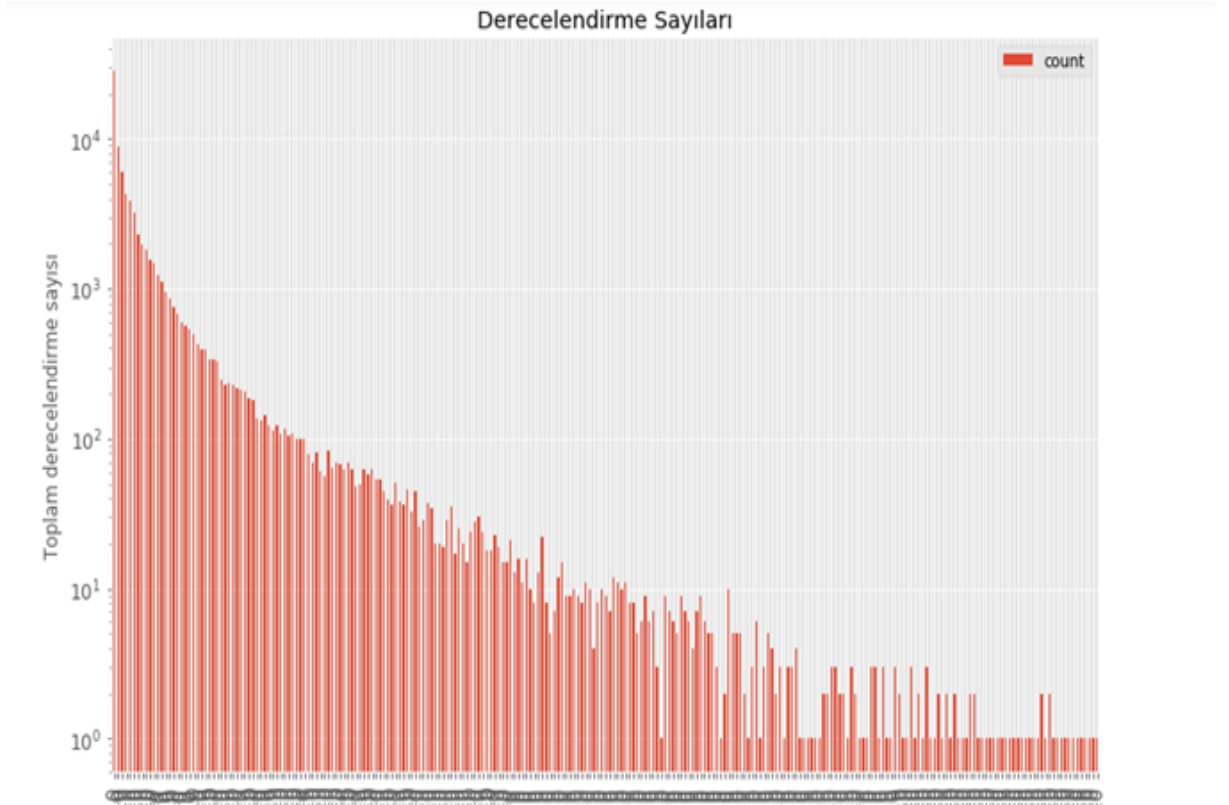
- Giyim ve Aksesuar
- Hobi ve Eğlence
- Market Harcamaları,
- Otomobil ve Ulaşım
- Sağlık ve Kişisel Bakım
- Sınıflandırılmamış Harcamalar
- Sigorta
- Tatil ve Seyahat
- Vergi ve Yasal Ücretler
- Yeme ve İçme
- Yatırım ve Tasarruflar

SQL de yazılan kod ile veri tabanında oluşturulan işlemler tablosunda her mcc kodunun dahil olduğu pfm ile eşleştirilmesi sağlanmıştır. Çalışma kapsamında tavsiye sistemini oluşturmak için işbirlikçi filtreleme yöntemleri kullanılacaktır. İşbirlikçi filtreleme tavsiye sistemlerinde gerekli olan kullanıcı – öge derecelendirme veri setini oluşturmak için SQL de yazılan kod ile her müşteri pfm bazında gruplanarak toplamı alınmıştır. Tablonun çıktı sonucu cvs formatında alınarak çalışmada kullanılan veri setinin son hali hazırlanmıştır.

userid	pfm	rating	
240	8	25	
1731	6	5	
792	9	1	
3069	3	2	
2564	4	3	
1397	5	7	
2354	10	14	
176	3	5	
2826	1	1	
1424	13	1	
2482	10	4	
722	13	2	
4581	3	1	
3031	7	1	
2145	4	13	
4218	15	8	
1815	2	24	
3756	8	10	

Şekil 3.3: Derecelendirme Veri Seti.

Çalışmada karşılaşılan en temel zorluk, derecelendirmeye dayalı tavsiye sistemlerinde derecelendirmeler genel olarak bir değer aralığındadır. Örneğin film veya kitap öneri sistemlerinde derecelendirmeler 0 – 5 aralığındadır. Ancak çalışma kapsamında kullanılan satın almaya dayalı veri setlerinde bir üst sınır yoktur. Çalışmamızdaki veri setinde en düşük derecelendirme oranı 1 iken en yüksek derecelendirme 2579 olarak hesaplanmıştır. Bu şekilde elde edilen veri seti çalışmada kullanıldığında elde edilen RMSE değerleri literatürde bulunan çalışmalar ile benzerlik göstermemektedir.



Şekil 3.4: Temel Veri Seti Derecelendirme Dağılımı.

Bu sorunu aşmak için öncelikle bir tane değişken tanımlıyoruz ve eğer kullanıcı o pfm' in ait olduğu sektör kodlarında bir alışveriş yaptıysa 1, yapmadıysa 0 atıyoruz [83]. Bu bizim çalışmada kullanacağımız ikinci veri setimizdir çalışma kapsamında etkileşimli veri seti olarak adlandırılacaktır.

userid	pfm	rating	etkilesim_degeri	
0	240	8	25	1
1	1731	6	5	1
2	792	9	1	1
3	3089	3	2	1
4	2564	4	3	1

Şekil 3.5: Etkileşime Dayalı Veri Seti.

Diğer bir çözüm yöntemi ise tüm pfm'leri kullanıcıların alışveriş sıklığına göre normalize etmektir. Normalizasyon sonrasında 1 o pfm için en yüksek derecelendirme 0 ise en düşük derecelendirme olacaktır. Çalışma kapsamında bu veri seti normalize veri seti olarak adlandırılacaktır. Bu şekilde oluşturulan üçüncü veri setine de tüm modeller uygulanıp sonuçlar karşılaştırılacaktır.

userid	pfm	normalize_degeri	
0	1	1	0.15
1	2	1	0.00
3	4	1	0.00
5	6	1	0.00
6	7	1	0.05

Şekil 3.6: Normalize Veri Seti.

3.4. VERİLERİN ANALİZİ VE MODELİN SEÇİLMESİ

Çalışma kapsamında yazılacak işbirlikçi tavsiye sistemi için veri seti olarak; bankadan alınan 4997 adet müşterinin 662.088 adet kredi kartı hareketi pfm bazında gruplandırılmıştır. Son hali ile veri seti 4997 müşteri ve toplam 51.527 adet kullanıcı – pfm değerlendirmesinden oluşmaktadır.

Verilerin analizi için, Jupyter Notebook ile Python programlama dili kullanılarak bir tavsiye sistemi geliştirilmiştir. Tavsiyelerin oluşturulması için açık kaynaklı bir paket olan Turi Create [84] kullanılmıştır. Verilere Turi Create aracılığı ile önce popülerlik modeli uygulanmış daha sonra Pearson Korelasyonu ve Kosinüs benzerlikleri üzerinden oluşturulan modeller ile tavsiyeler üretilmiştir. Çalışmanın akış diyagramı Şekil 3.7 deki gibidir.



Şekil 3.7: Çalışma Akış Şeması.

Veri seti hazırlandıktan sonra veriler test ve eğitim olmak üzere iki gruba ayrılırlar. Çalışmada veri seti %80 eğitim - %20 test ve %70 eğitim - %20 test olarak iki gruba ayrılır ve modeller uygulanır. Analiz sonucunda tavsiyeler RMSE, hassasiyet ve duyarlılık değerlendirme ölçütleri üzerinden karşılaştırılıp uygun model seçilir.

3.5. TAVSİYE SİSTEMİNİN OLUŞTURULMASI

Değerlendirme sonuçlarına göre en başarılı sonuçları veren modele tüm veri girdi olarak verilerek Turi Create aracılığı ile tavsiyelerin oluşturulması sağlanır ve verilen müşteri numarası için kişiye en uygun 10 pfm tavsiye edilir. Çalışma kapsamında işbirlikçi filtreleme tavsiye sistemlerinde yaşanan yeni kullanıcı sorunu müşteriye en popüler 10 pfm önerilmesiyle aşılmıştır.

4. BULGULAR

Bu bölümde, hazırlanan veri setlerine modellerin uygulanması sonucu ortaya çıkan tavsiyelere ve değerlendirme sonuçlarına yer verilmiştir. Bankadan alınan verilerin pfm bazından gruplanarak oluşturulan derecelendirme veri seti temel olmak üzere toplamda üç veri seti incelenecektir. Bunlar aşağıda belirtilmiştir.

- Derecelendirmeleri içeren temel veri seti,
- Temel veri seti üzerinde müşterilerin o pfm için etkileşimde bulunup bulunmadığı değişkenini içeren etkileşime dayalı veri seti,
- Temel veri setinin normalize edilmesiyle oluşan normalize veri seti

4.1. MODELLERİN UYGULANMASI VE KARŞILAŞTIRILMASI

Çalışma kapsamında incelenmek üzere üç veri seti %80 eğitim ve %20 test ve %70 eğitim - %30 test olarak ayrılmıştır. Turi Create “item_similarity_recommender” fonksiyonu Pearson Korelasyonu ve Kosinüs Benzerliği yöntemleri ile çağırılarak oluşturulan modeller incelenmiştir.

```
In [10]: def model(train_data, name, user_id, pfm, target, user_list, n_rec, n_display):
    if name == 'popularity':
        model = tc.popularity_recommender.create(train_data,
                                                user_id=user_id,
                                                item_id=pfm,
                                                target=target)

    elif name == 'cosine':
        model = tc.item_similarity_recommender.create(train_data,
                                                    user_id=user_id,
                                                    item_id=pfm,
                                                    target=target,
                                                    similarity_type='cosine')

    elif name == 'pearson':
        model = tc.item_similarity_recommender.create(train_data,
                                                    user_id=user_id,
                                                    item_id=pfm,
                                                    target=target,
                                                    similarity_type='pearson')

    recom = model.recommend(users=user_list, k=n_rec)

    return model
```

Şekil 4.1: Örnek Python Kodu.

4.1.1. Öge Benzerliği Modeline Kosinüs Benzerliği Uygulanması

Bu model ilk önce, her iki öğeyle etkileşime geçen kullanıcıların gözlemlerini kullanan öğeler arasındaki benzerliği hesaplar. Çalışma kapsamına öncelikle üç veri seti üzerine Kosinüs benzerliği modeli uygulanmıştır.

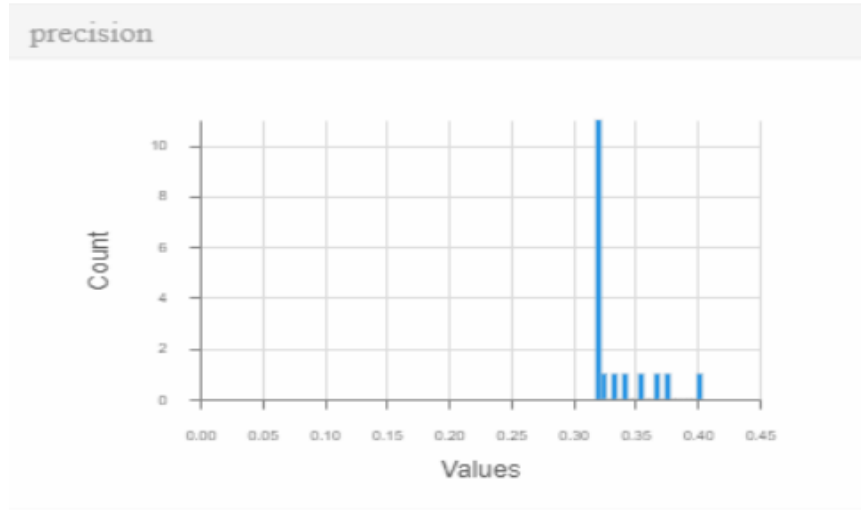
4.1.1.1. Temel Veri Sonuçları

Bankadan alınan veriler test ve eğitim olmak üzere ikiye ayrılmış ve turicreate aracılığı ile benzerliğin hesaplanması sağlanmıştır.

```
Precision and recall summary statistics by cutoff
```

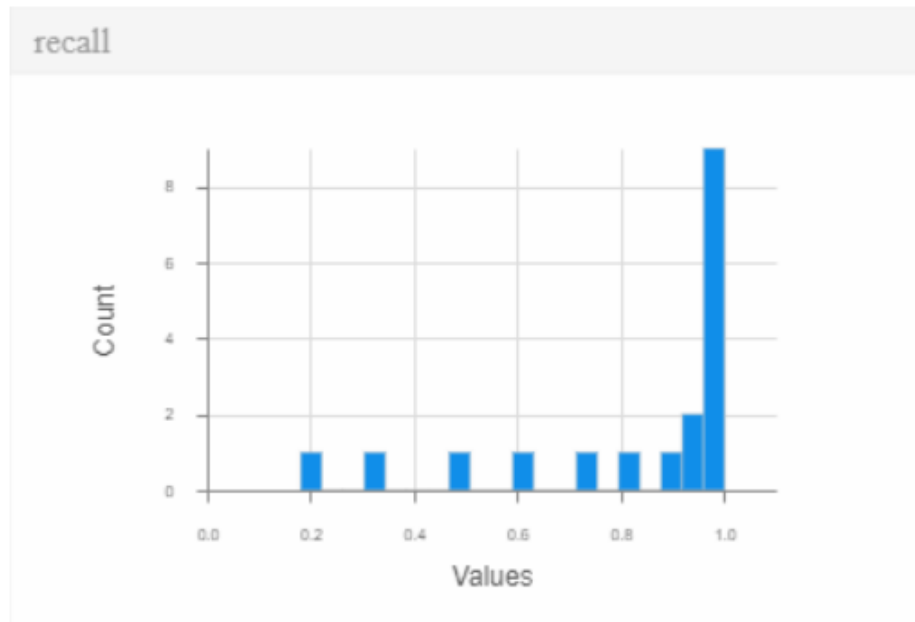
cutoff	mean_precision	mean_recall
1	0.40355272147574733	0.18266159496517062
2	0.37542700979275745	0.33346853050792996
3	0.3679116374402181	0.4848068560150098
4	0.35379184695969007	0.6113097497406312
5	0.34173309041220784	0.7236776704153167
6	0.33263873073711325	0.811249200203882
7	0.3257089718146431	0.8798761355297453
8	0.32101593085423635	0.9264425249337556
9	0.3193869675707512	0.9579108421627208
10	0.3187619517555756	0.9765409405240884

Şekil 4.2: Temel Veri Kosinüs Benzerliği Hassasiyet ve Duyarlılık Değerleri.



Şekil 4.3 : Temel veri Hassasiyet Değerler Grafiği.

Şekil 4.3 de hassasiyet ve duyarlılık değerlerine yer verilmiştir. Temel veri seti için ortalama hassasiyet değeri 0.334 olarak hesaplanmıştır. Grafikte de en yoğun değerlerin bu aralıkta olduğu görülmektedir.



Şekil 4.4: Temel Veri Duyarlılık Verileri.

Temel veri seti duyarlılık grafiğini incelediğimizde 0.80 - 0.90 arası değerlerin çoğunlukta olduğu görülmektedir. Bu model için ortalama duyarlılık da 0.826 olarak hesaplanmıştır.

Bu model için ortalama RMSE değerleri Tablo 2 de verilmiştir.

Tablo 2: Temel Veri Kosinüs Benzerliği RMSE Değerleri.

RMSE Değerleri	%80 Eğitim - %20 Test	%70 Eğitim - %30 Test
Ortalama Değer	24.27727055130467	23.753812912216784
Kullanıcı Bazlı En İyi Değer	0.015044957399368286	0.007005350930350263
Kullanıcı Bazlı En Kötü Değer	266.0	274.42446244911093
Öge Bazlı En İyi Değer	5.668492658496454	5.759840390660107
Öge Bazlı En Kötü Değer	60.31603882780518	59.680733818533156

Bu veri seti için RMSE değerleri incelendiğinde çok yüksek olduğu görülmektedir. Derecelendirme aralığımız geniş olduğundan modelimiz temel veri seti için istediğimiz başarı oranını yakalayamamıştır.

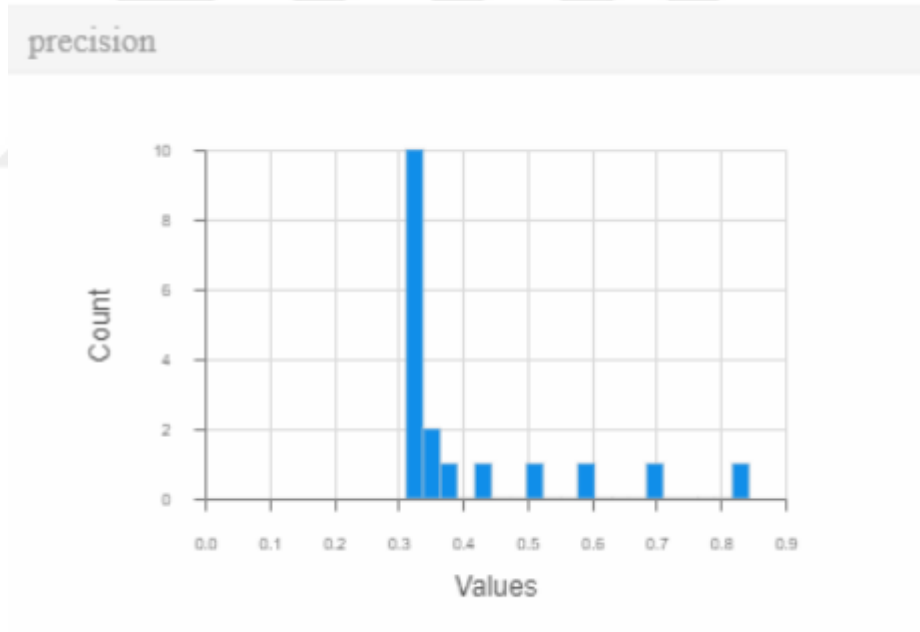
4.1.1.2.Etkileşime Dayalı Veri Seti için Sonuçlar

Etkileşime dayalı veri seti için Kosinüs benzerliği metrik olarak uygulandığında ortalama hassasiyet, duyarlılık değerleri Şekil 4.5 de belirtilmiştir.

```
Precision and recall summary statistics by cutoff
```

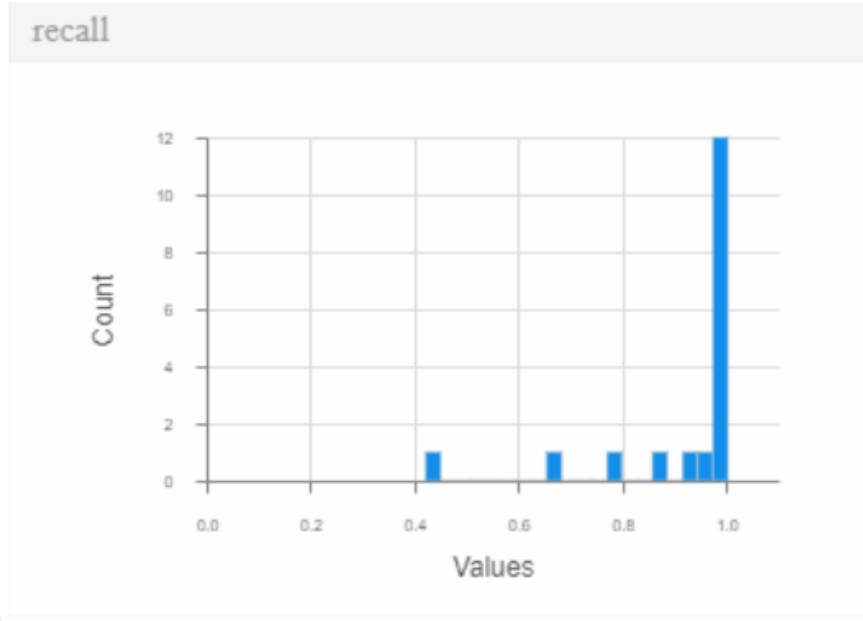
cutoff	mean_precision	mean_recall
1	0.8362147406733381	0.4297275120239187
2	0.6975659690627837	0.6590607262013077
3	0.5884137094328192	0.797519660730532
4	0.5018387928419782	0.8789283547814032
5	0.43791704579921176	0.932165875904502
6	0.3915491355777984	0.9644593028294124
7	0.3594691061137832	0.9793648988257744
8	0.3400967871224927	0.9904312361887422
9	0.32911458107659264	0.9959901317214773
10	0.3232708847870352	0.9979770570648626

Şekil 4.5: Etkileşime Dayalı Veri Seti için Hassasiyet ve Duyarlılık.



Şekil 4.6: Etkileşime Dayalı Veri Seti Hassasiyet.

Etkileşime dayalı veri seti için ortalama hassasiyet değeri 0.408 olarak hesaplanmıştır. Bu veri seti için hassasiyet değerleri temel veri setinden daha yüksek çıkmıştır.



Şekil 4.7: Etkileşime Dayalı Veri Seti Duyarlılık.

Bir kullanıcının pfm ile daha önce etkileşimi olup olmadığına dayalı veri setimiz için duyarlılık ortalaması 0.849 olarak hesaplanmıştır.

Bu model için ortalama RMSE değerleri Tablo 3 ' de gösterilmiştir.

Tablo 3: Etkileşime Dayalı Veri Kosinüs Benzerliği RMSE Değerleri

RMSE Değerleri	%80 Eğitim - %20 Test	%70 Eğitim - %30 Test
Ortalama Değer	0.4005069811146648	0.45914859263858143
Kullanıcı Bazlı En İyi Değer	0.22103551030158997	0.21772301197052002
Kullanıcı Bazlı En Kötü Değer	1.0	1.0
Öğe Bazlı En İyi Değer	0.3197340399770914	0.3904906310813116
Öğe Bazlı En Kötü Değer	0.7798909406200679	0.7987882168968724

Etkileşime dayalı veri setinde derecelendirme aralığımız ikili değere sahiptir, pfm için işlemi bulunanlara 1 bulunmaya 0 değeri verilmiştir. Bu şekilde RMSE değerinin düştüğü görülmüştür. İki eğitim seti karşılaştırıldığında %80 eğitim olarak ayrıldığında daha başarılı RMSE değerleri elde edildiği görülmüştür.

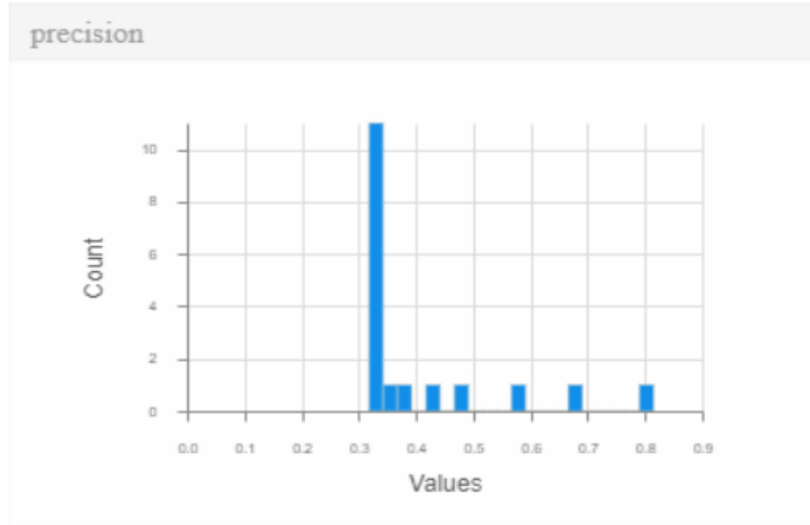
4.1.1.3. Normalize Edilen Veri için Sonuçlar

Normalize edilen veri setine öge benzerlik modelinde Kosinüs Benzerliği metrik olarak uygulandığında elde edilen hassasiyet, duyarlılık ve RMSE değerleri Şekil 4.8 de gösterilmiştir.

Precision and recall summary statistics by cutoff

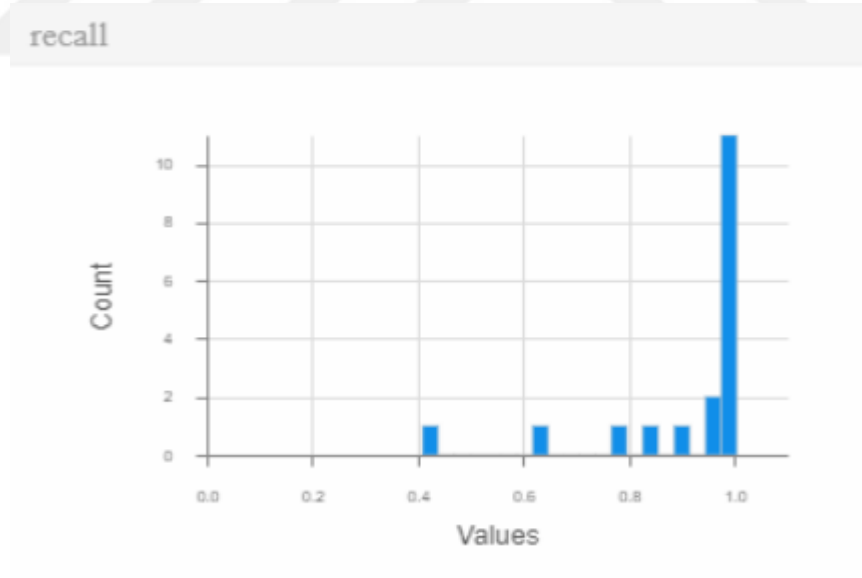
cutoff	mean_precision	mean_recall
1	0.8125429159990841	0.41334483209626294
2	0.677386129549093	0.6323927236264142
3	0.5720988784618901	0.7713247555831687
4	0.48638132295719816	0.850597554196776
5	0.4272030212863357	0.9080788346466997
6	0.38519493400473137	0.9467323894538344
7	0.3567205092153597	0.9706795169429637
8	0.33956637129559925	0.9855352646895321
9	0.3295589961016828	0.9925151772771352
10	0.3242793745254257	0.9958893284940449

Şekil 4.8: Normalize Veri Seti Hassasiyet ve Duyarlılık Değerleri.



Şekil 4.9: Normalize Veri Seti için Hassasiyet.

Normalize veri seti için ortalama hassasiyet 0.404 olarak hesaplanmıştır. Etkileşime dayalı veri seti çok az da olsa hassasiyet metriğinde daha başarılı sonuçlar vermiştir.



Şekil 4.10: Normalize Veri Seti için Duyarlılık.

Normalize veri set için duyarlılık 0.914 olarak hesaplanmıştır. Kosinüs benzerliğinin metrik olarak kullanıldığı öge benzerlik modelinde duyarlılık olarak en yüksek sonuç normalize veri setinde elde edilmiştir.

Bu model için RMSE değerleri Tablo 4 de verilmiştir. %70 eğitim olarak ayrılan veri setinde ortalama RMSE değerinin daha düşük olduğu görülmüştür.

Tablo 4: Normalize Veri Seti Kosinüs Benzerliği RMSE Değerleri.

RMSE Değerleri	%80 Eğitim - %20 Test	%70 Eğitim - %30 Test
Ortalama Değer	0.09504442985161031	0.09374523864202362
Kullanıcı Bazlı En İyi Değer	0.0	0.0
Kullanıcı Bazlı En Kötü Değer	0.7951094309488933	0.8812616308530172
Öge Bazlı En İyi Değer	0.01871007108455467	0.020480392124514208
Öge Bazlı En Kötü Değer	0.1464413563080115	0.15605141207480666

4.1.2. Öge Benzerliği Modelinde Pearson Korelasyonunun Hesaplanması

Pearson korelasyonunun da, Kosinüs Benzerliğindeki gibi benzerliğin hesaplanmasında iki vektör arasındaki açı hesaplanır. Kosinüs benzerliğinden farklı olarak iki vektörün çarpımı yapılmadan önce her birinin ayrı ayrı ortalama değerleri hesaplanır ve her ortalama değer ait olduğu vektörün tüm elemanlarından çıkarılır [48]. Bu bölümde üç veri seti için ayrı ayrı sonuçlar verilmiştir.

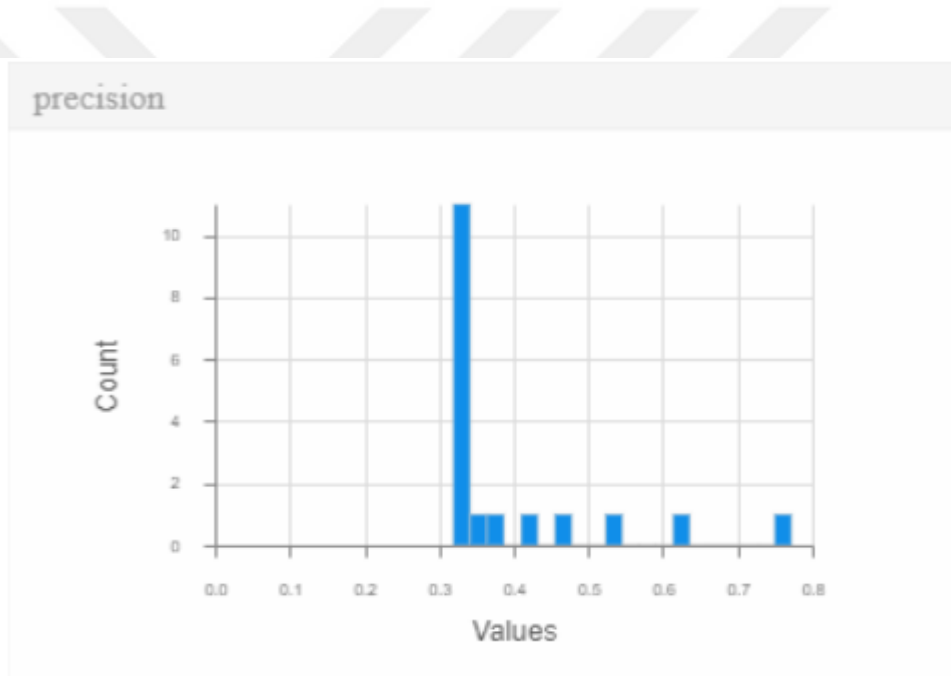
4.1.2.1. Temel Veri Seti için Sonuçlar

Temel veri üzerine modelin uygulanması ile ele edilen sonuçlar Şekil 4.11 ' de belirtilmiştir.

Precision and recall summary statistics by cutoff

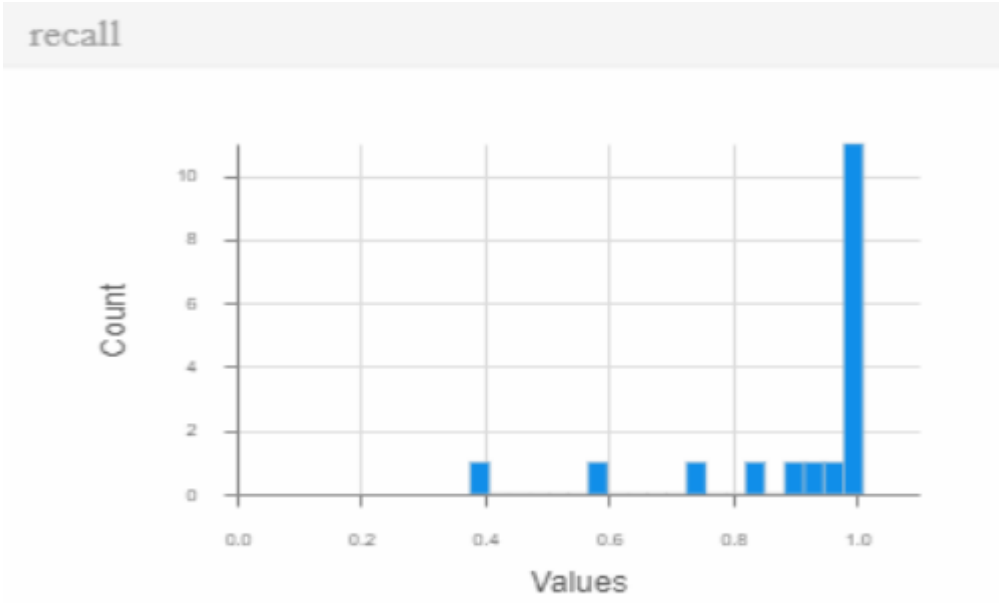
cutoff	mean_precision	mean_recall
1	0.7715782281940315	0.3849471140464076
2	0.6226372124800743	0.5832207292694659
3	0.531427920746984	0.7269243907993636
4	0.4679457982236401	0.8292685615960513
5	0.42025734456843616	0.900287293996016
6	0.38338647232976597	0.945281654755578
7	0.35612833609872996	0.9702890652958985
8	0.3386087343158635	0.9840895699356194
9	0.3285977269523165	0.9922630344174396
10	0.32331672287832613	0.9961929704699006

Şekil 4.11: Temel Veri için Pearson Hassasiyet ve Duyarlılık Değerleri.



Şekil 4.12: Temel Veri Pearson Korelasyonu Hassasiyet Grafiği.

Temel veri seti için benzerliğin hesaplamasında Pearson ilişkisi kullanıldığında ortalama hassasiyet 0.394 olarak hesaplanmıştır. Kosinüs Benzerliği ile karşılaştırdığımızda Pearson Korelasyonunun hassasiyet bakımından daha başarılı sonuçlar verdiği görülmüştür.



Şekil 4.13: Temel Veri Pearson Duyarlılık Grafiği.

Temel veri seti için duyarlılık grafiği incelendiğinde yoğunluğun 0.8-1.0 aralığında olduğu görülmüştür. Ortalama duyarlılık 0.906 olarak hesaplanmıştır.

Bu model için RMSE değerleri Tablo 5' de verilmiştir.

Tablo 5: Temel Veri Pearson RMSE Değerleri.

RMSE Değerleri	%80 Eğitim - %20 Test	%70 Eğitim - %30 Test
Ortalama Değer	17.79617799636283	17.71848919559406
Kullanıcı Bazlı En İyi Değer	0.003231676474007017	0.00783021329459288
Kullanıcı Bazlı En Kötü Değer	220.74066003584247	265.6561286069583
Öge Bazlı En İyi Değer	1.802833398730371	2.5399434542389914
Öge Bazlı En Kötü Değer	39.458359534689784	38.5626692535023

Kosinüs Benzerliğine göre daha iyi bir sonuç elde edilse de; bu model için elde edilen RMSE değeri de çok yüksektir.

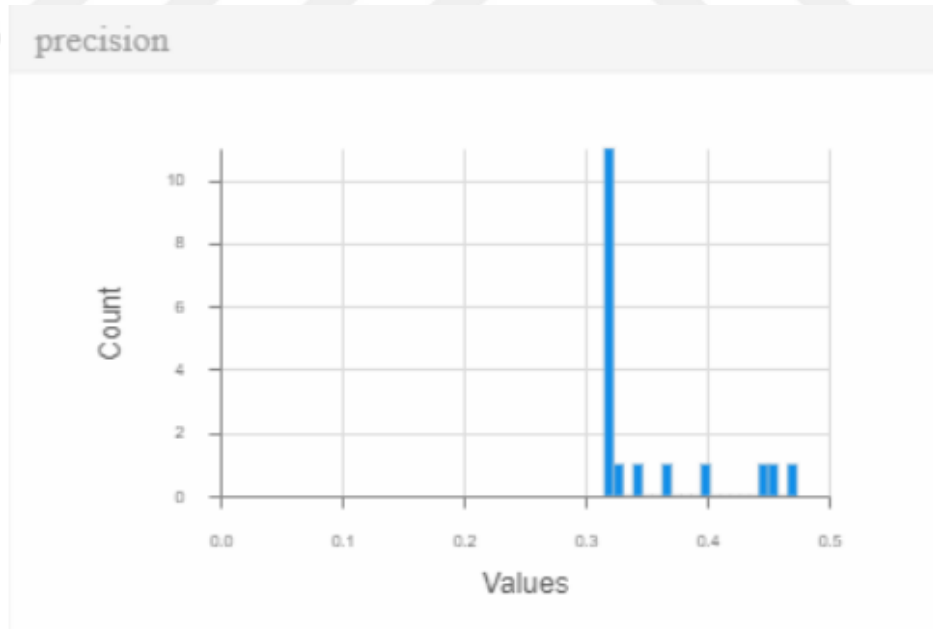
4.1.2.2. Etkileşime Dayalı Veri için Sonuçlar

Etkileşime dayalı veri setine Pearson modelinin uygulanmasıyla elde edilen sonuçlar Şekil 4.14 de gösterilmiştir.

```
Precision and recall summary statistics by cutoff
```

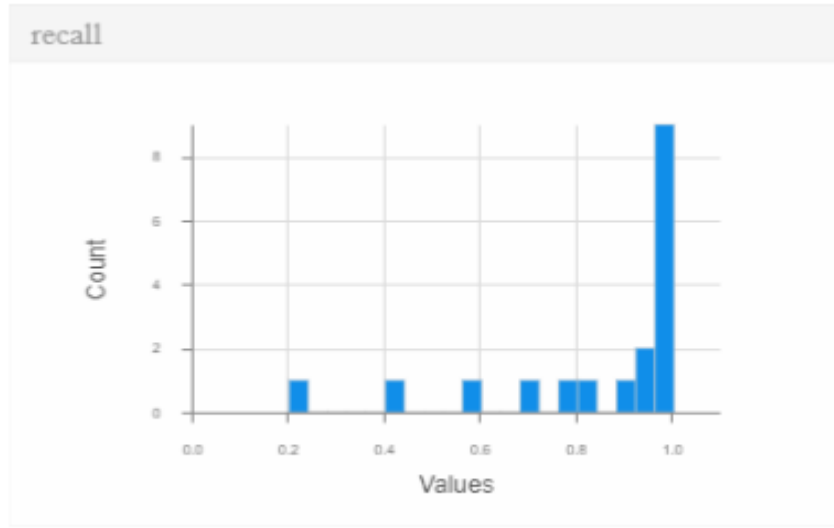
cutoff	mean_precision	mean_recall
1	0.4560964513193812	0.2095197473894019
2	0.470655141037307	0.4277118809307155
3	0.444191689414619	0.5928406235105511
4	0.4012928419775555	0.7005077668009907
5	0.3648051258720039	0.7750874712942512
6	0.3403662420382171	0.8361846808787207
7	0.32667955717318803	0.8868064582520893
8	0.3212444213354139	0.9296124182156938
9	0.31933296142236	0.9607689891243124
10	0.3188122860609214	0.9800050912084572

Şekil 4.14: Etkileşime Dayalı Veri Seti Pearson Hassasiyet ve Duyarlılık Değerleri.



Şekil 4.15: Etkileşime Dayalı Veri Seti Pearson Korelasyonu ile Hassasiyet Grafiği.

Etkileşime dayalı veri seti için Pearson Korelasyonu ile ortalama hassasiyet 0.351 olarak hesaplanmıştır.



Şekil 4.16: Etkileşime Dayalı Veri Seti Pearson Benzerliği Duyarlılık Grafiği.

Bu modeli duyarlılık bakımından incelediğimizde ortalama değerin 0.850 olarak hesaplandığı görülmüştür.

Bu veri seti için Pearson modeli uygulandığında RMSE değerleri Tablo 6 'da gösterilmiştir.

Tablo 6: Etkileşimli Veri Seti Pearson Modeli RMSE Değerleri.

RMSE Değerleri	%80 Eğitim - %20 Test	%70 Eğitim - %30 Test
Ortalama Değer	1.0	1.0
Kullanıcı Bazlı En İyi Değer	1.0	1.0
Kullanıcı Bazlı En Kötü Değer	1.0	1.0
Öğe Bazlı En İyi Değer	1.0	1.0
Öğe Bazlı En Kötü Değer	1.0	1.0

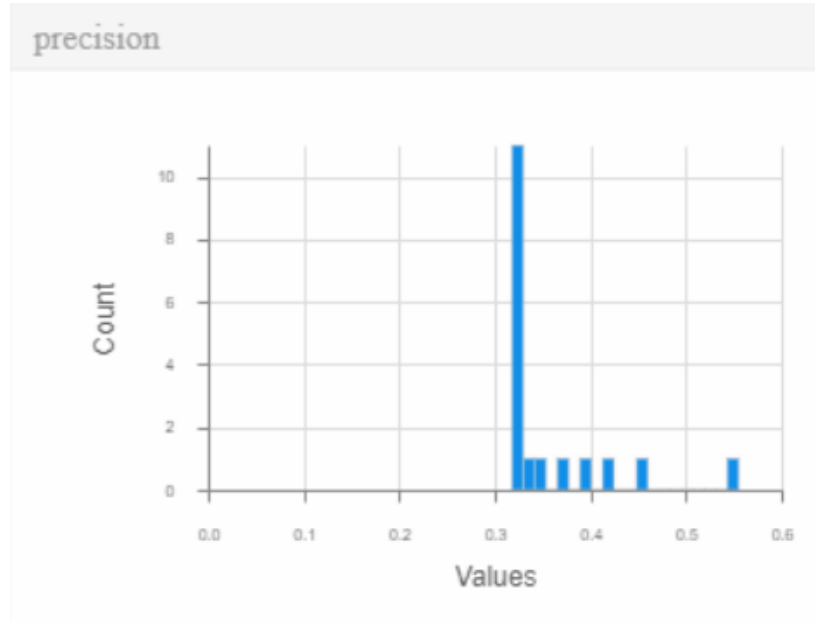
4.1.2.3. Normalize Edilen Veri için Sonuçlar

Normalize edilen veri seti için Pearson Modeli uygulandığında elde edilen sonuçlar Şekil 4.17 'de belirtilmiştir.

```
Precision and recall summary statistics by cutoff
```

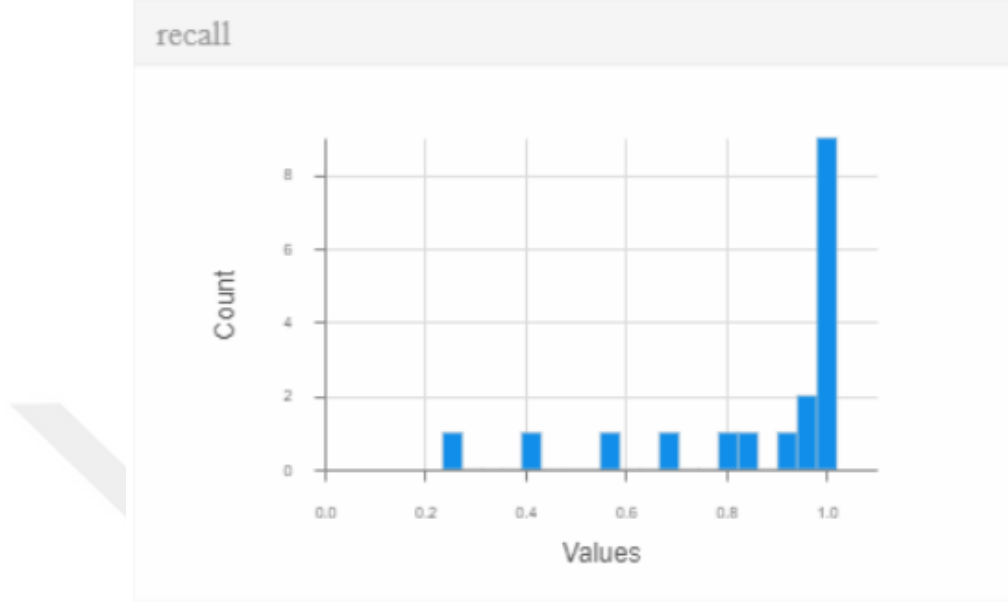
cutoff	mean_precision	mean_recall
1	0.552300297550925	0.25446408135238524
2	0.4501029983978023	0.4078349627788853
3	0.4182879377431899	0.565762297136752
4	0.39665827420462374	0.6994904576616643
5	0.3710345616845967	0.7944857709620812
6	0.3497558556496532	0.8622704334652161
7	0.33638241288733484	0.9149704628933282
8	0.32835099020152775	0.9494059880761649
9	0.32368427267145555	0.9716642688203679
10	0.32180741297816146	0.9857965754395138

Şekil 4.17: Normalize Veri Seti için Pearson Hassasiyet ve Duyarlılık.



Şekil 4.18: Normalize Veri Seti için Pearson Hassasiyet Grafiği.

Normalize edilen veri seti için Pearson Korelasyonu hesaplandığında ortalama hassasiyet 0.356 olarak hesaplanmıştır.



Şekil 4.19: Normalize Veri Seti için Duyarlılık Grafiği.

Bu model için ortalama duyarlılık 0.855 olarak hesaplanmış, dağılımında en düşük değerlerin 0.25 oranında olduğu görülmüştür.

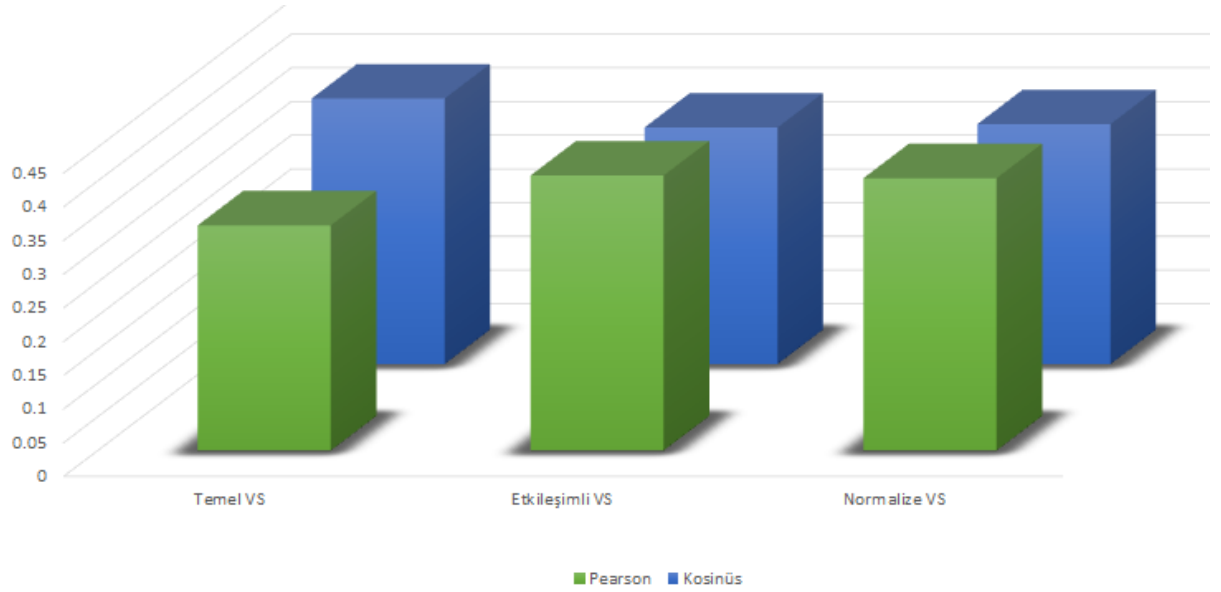
Bu model için RMSE değerleri Tablo 7' de gösterilmiştir.

Tablo 7: Normalize Veri Seti İçin Pearson RMSE Değerleri.

RMSE Değerleri	%80 Eğitim - %20 Test	%70 Eğitim - %30 Test
Ortalama Değer	0.0802199457845637	0.07852996352182436
Kullanıcı Bazlı En İyi Değer	0.7517778492159004	9.378082135400034e-06
Kullanıcı Bazlı En Kötü Değer	8.583010734215107	0.8066486922540457
Öge Bazlı En İyi Değer	0.018058252934240765	0.018916775821790285
Öge Bazlı En Kötü Değeri	0.140029173348152	0.13960040141117105

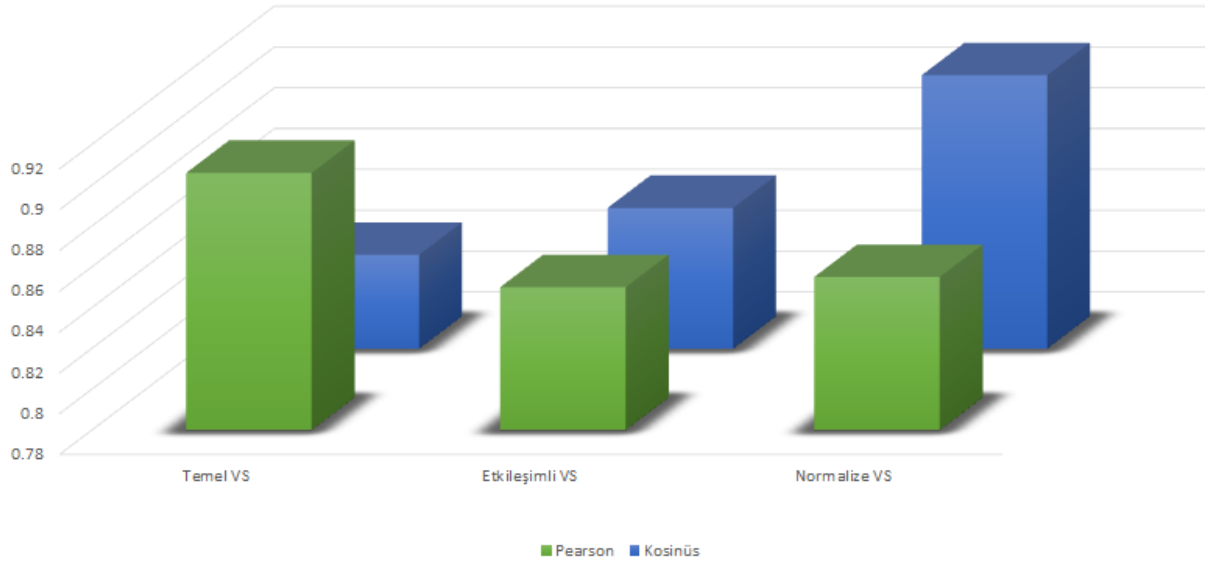
4.1.3. Modellerin Karşılaştırılması

Bankadan alınan temel veri seti, etkileşime dayalı veri seti ve normalize edilmiş veri setlerine Turi Create ile öge benzerliği modeli Kosinüs ve Pearson korelasyonu aracılığıyla uygulanmış ve sonuçlar incelenmiştir. Sonuçları hem hassasiyet ve duyarlılık hem de RMSE açısından değerlendirmek gerekmektedir.



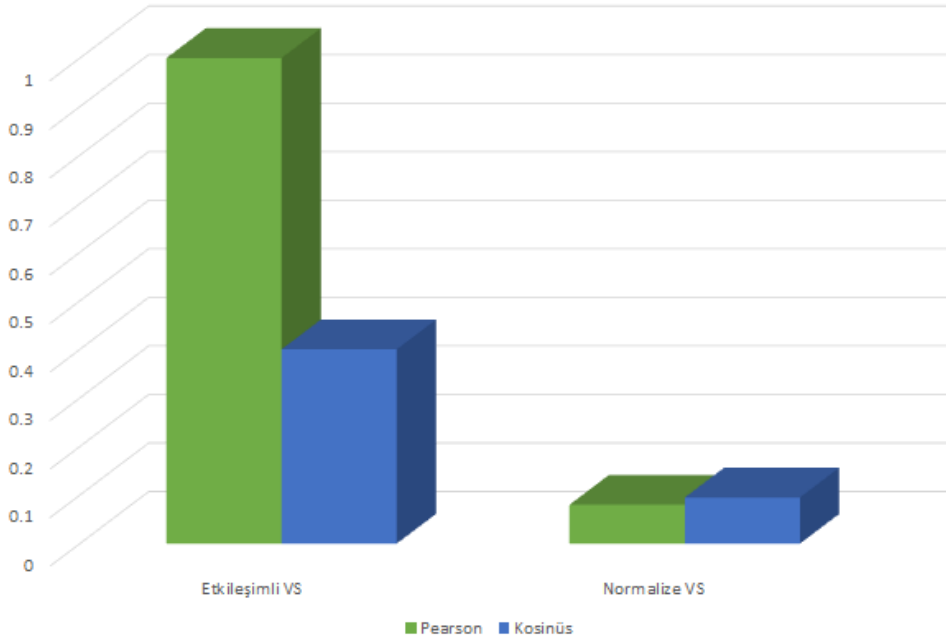
Şekil 4.20: Modellere göre Hassasiyet Değerlerinin Karşılaştırılması.

İlk olarak iki model için hassasiyet değerlerini karşılaştırırsak; en başarılı sonucu etkileşimli veri için kosinüs benzerliği vermiştir. Kosinüs benzerliği normalize veri seti için de etkileşimli veri setine çok yakın bir sonuç vermiştir. En düşük sonuç ise etkileşimli veri seti için Pearson korelasyonu ile elde edilmiştir.



Şekil 4.21: Modellere göre Duyarlılık Değerlerinin Karşılaştırılması.

Duyarlılık değerleri incelendiğinde ise en başarılı sonuç normalize veri seti için Kosinüs benzerliği ile elde edilmiştir. İkinci başarılı sonucu temel veri seti için Pearson Korelasyonu verirken, en başarısız sonucu temel veri seti için kosinüs benzerliği vermiştir.



Şekil 4.22: Modellere göre RMSE Değerlerinin Karşılaştırılması.

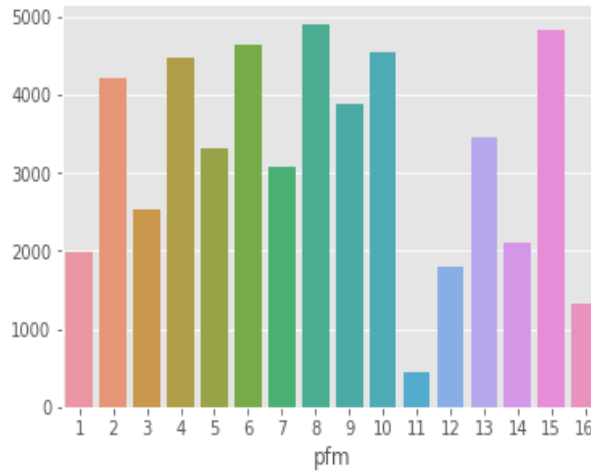
Şekil 4.22 incelendiğinde en yüksek RMSE değerlerinin temel veri seti için alındığı görülmüştür. Tavsiye sistemleri için RMSE değerleri ortalama olarak 0-1 aralığındadır. RMSE değeri ne kadar düşük ise tavsiye sistemi o kadar başarılıdır. Çalışmamız da en düşük sonuçlar normalize veri seti ile elde edilmiştir.

Model seçimimizde hassasiyet ve duyarlılığı yüksek, RMSE oranı düşük olan modeli seçeceğiz. En düşük RMSE Pearson korelasyonu normalize veri seti ile elde edilmiştir ancak kosinüs benzerliği ile elde edilen sonuçla değerlerin birbirine çok yoğun olduğu görülmüştür. Bu sebeple hassasiyet ve duyarlılık sonuçları daha yüksek olan kosinüs benzerliği seçilmiştir.

Çalışmamızda tavsiyelerin üretilmesinde veri seti olarak normalize veri seçilmiştir. Öğe benzerliğinin uygulanmasında ise metrik olarak kosinüs benzerliği kullanılmıştır.

4.2. OLUŞTURULAN TAVSİYELERİN YORUMLANMASI

Bir tavsiye sisteminin oluşturulmasındaki en önemli adım modelin seçimidir. Veri setimize göre en uygun sonucu veren modeli seçtikten sonra model, tüm veri setine uygulanarak tüm kullanıcılar için tavsiyeler üretilmiş, tavsiyelerin çıktısı csv formatında kaydedilmiştir. Bu şekilde ürettiğimiz tavsiyeler sadece sistemde daha önceden etkileşimi olan kullanıcılar içindir. Yani iş birlikçi filtreleme tavsiye sistemlerinin genel sorunu olan yeni kullanıcılar (soğuk başlangıç problemi) için tavsiye üretilmemektedir. Bu sorunu aşmak için kullanıcılara en çok işlem yapılan pfm grubuna ait kampanyalar önerilmiştir.

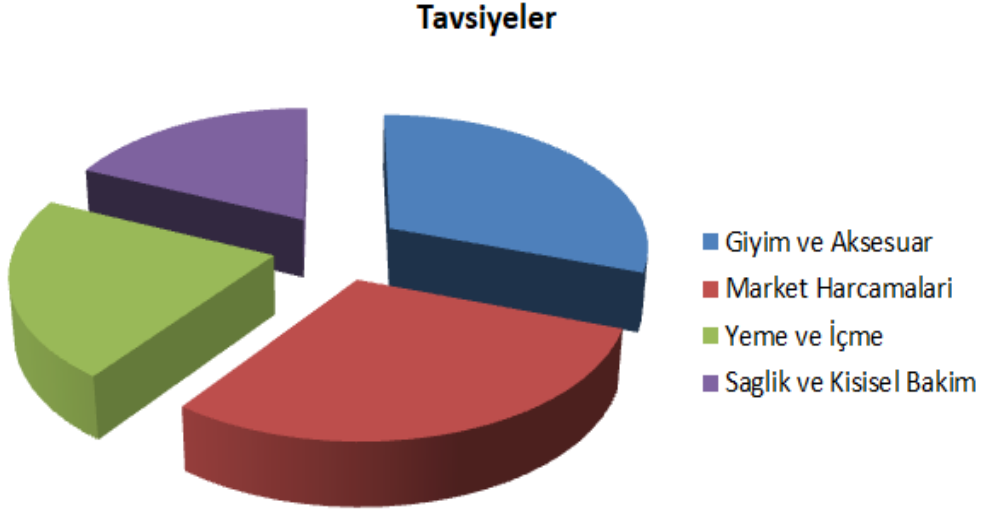


Şekil 4.23 : Pfm Türlerine göre İşlem Sayıları.

Temel verilere göre pfm bazında işlem sayılarını gösteren grafik incelendiğinde en çok işlem yapılan pfm aşağıdaki şekilde gruplanmıştır;

- Market harcamaları
- Yeme ve İçme
- Giyim e Aksesuar
- Otomobil ve Ulaşım
- Sağlık ve Kişisel Bakım

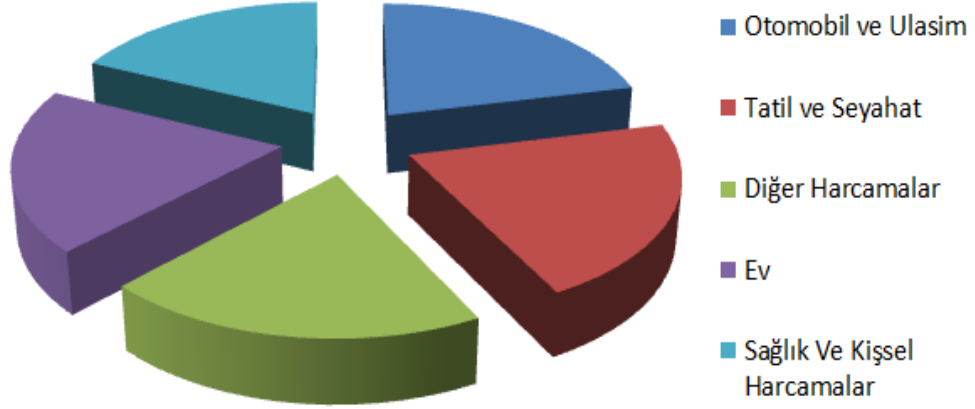
Mevcuttaki müşterilere önerilen tavsiyeler incelendiğinde ise, müşterilere ilk olarak en çok giyim ve aksesuar grubundaki kampanyalar tavsiye edilmiştir.



Şekil 4.24: Müşterilere İlk Önerilen Kampanya Grupları.

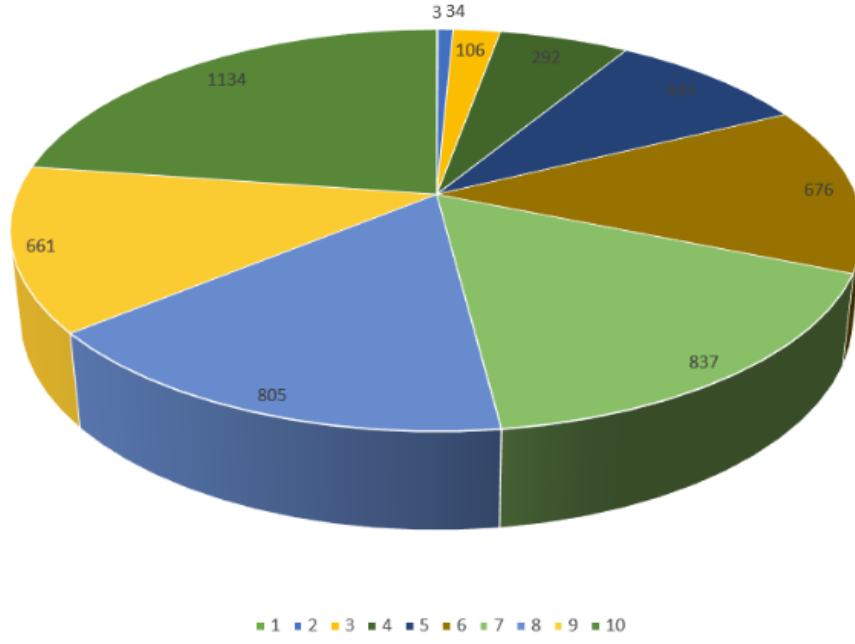
Müşterilere ikinci olarak en çok otomobil ve ulaşım grubundaki kampanyalar tavsiye edilmiş olup diğer pfm grupları ise Şekil 4.25 de gösterilmiştir.

İkinci Tavsiye Grubu



Şekil 4.25: Müşterilere İkinci Olarak Önerilen Kampanya Grupları.

Genel olarak üretilen tavsiyeler incelendiğinde bazı müşteriler için 10 adet tavsiyenin üretilmediği görülmüştür. Mevcut 4997 kullanıcı için çalıştırılan tavsiye sistemimiz de 1135 adet kullanıcıya 10 adet tavsiye üretilmiştir. Kişilere göre tavsiye sayısı dağılımı Şekil 4.26 da gösterilmiştir.



Şekil 4.26: Üretilen Tavsiye Sayıları Dağılımı.

Kullanıcılara en çok 10 tavsiye üretilirken azalan değerler ile sistem tarafından 3 kişiye yalnızca bir tavsiye üretilmiştir. Bu kullanıcılardan bir tanesinin normalize veri seti için değerleri Şekil 4.27 de verilmiştir.

userId	pfm	normalize_degeri
1150	1	0.3
1150	2	0.12962963
1150	3	0.114285714
1150	4	0.3
1150	5	0.001163693
1150	6	0.181818182
1150	7	0.051282051
1150	8	0.339832869
1150	9	0.101769912
1150	10	0.174757282
1150	11	0
1150	12	0
1150	13	0.009328358
1150	14	0.048387097
1150	15	0.188732394

Şekil 4.27: Kullanıcı Pfm- Derecelendirme Değerleri.

1150 numaralı kullanıcı için mevcutta 15 pfm de işlemi olduğu görülmüştür ancak 11 ve 12 numaralı pfmler için, bu pfm derecelendirmeleri arasında düşük değere sahip olduğundan normalize veri setinde 0 değeri verilmiştir. Oluşturulan tavsiye sisteminde kullanıcının kendisine benzer kullanıcılar arasında hiç işlemi bulunmayan “Yatırım ve Tasarruflar” kategorisi önerilmiştir. Kullanıcılara yeteri kadar tavsiye üretilmemesinin sebebi, veri setimizdeki öge sayısının kısıtlı olmasıdır. Daha çeşitli ürün yelpazesine sahip veri setleri için tavsiye sistemi daha başarılı tavsiyeler üretecektir.

5. TARTIŞMA VE SONUÇ

Bu çalışma kapsamında özel bir bankadan alınan veriler ile bir tavsiye sistemi oluşturularak müşterilere en uygun kampanya önerilerek müşteri memnuniyetinin artırılması amaçlanmıştır. Oluşturulan üç veri setine Turi Create aracılığıyla benzerlik modeli üzerinden Pearson korelasyonu ve Kosinüs benzerliği uygulanmıştır. Birbirine en yakın kullanıcılar bulunarak, hedefteki kullanıcıya kendisine benzeyen kullanıcıların kullandığı kampanyaların önerilmesi sağlanmıştır.

Sonuçlar veri seti açısından karşılaştırıldığında aşağıdaki sonuçlar gözlenmiştir:

- Temel veri seti için her iki metrik için de RMSE değerlerinin çok yakın olduğu, hassasiyet ve duyarlılık değerlerinin derecelendirme aralıklarının geniş olmasından etkilenmediği diğer iki veri seti ile yakın sonuçlar elde edilmiştir.
- Etkileşime dayalı veri seti için, RMSE değerleri karşılaştırıldığında kosinüs benzerliğinin daha başarılı olduğu, duyarlılık için iki metrik de birbirine çok yakın sonuçlar vermesine rağmen hassasiyet için kosinüs benzerliğinin daha başarılı sonuçlar elde edilmiştir.
- Normalize veri seti için RMSE değerleri birbirine yakın sonuçlar verirken, hassasiyet ve duyarlılık için kosinüs benzerliği daha başarılı sonuçlar gözlemlenmiştir.

Sonuçlar uygulanan metrikler açısından değerlendirilirse, en başarılı sonuçlar Pearson korelasyonu ile elde edilmiş tavsiye sistemi de bu modele göre yazılmıştır.

Veriler üzerinde işbirlikçi filtreleme yöntemleri uygulanarak veri setinin derecelendirme aralığının geniş olması sorunu verileri normalize ederek aşılmıştır. Uygulanan iki modelde kosinüs benzerliğinin veri setimiz için daha başarılı sonuçlar verdiği görülmüş ve tavsiyeler bu modele göre üretilmiştir.

İşbirlikçi tavsiye sistemlerinde yaşanan yeni kullanıcı problemi, kullanıcılara popüler olan pfm grubunun önerilmesiyle aşılmıştır. Öğe sayısının az olması sebebiyle bazı kullanıcılara hedeflenen 10 tavsiye üretilememiştir. Yeni kullanıcı yani soğuk başlangıç problemi için kullanıcıların demografik özelliklerinin kullanılması ileriki çalışmalarda kullanılmak üzere açıkta bırakılmıştır.

Kredi kartı işlemlerinin yanı sıra bankamatik kartları ile yapılan işlemlerin de sürece dahil etmesi, tavsiye sisteminin eğitime katkı sağlayacağından, sistemin daha başarılı tavsiyeler oluşturması beklenmektedir.

Yapılan işlemlerin zaman diliminden yapılacak incelemeler ile müşterilerin gün içerisindeki alışveriş analizi yapılabilir. Bu verilerin kampanya yönetiminde kullanılması hem müşteri yönetimi hem de müşterilerin tekrar alışverişe yönlendirilmesi açısından önemlidir.



KAYNAKLAR

- [1] “Kredi Kartlarının Türkiye ’deki Tarihsel Gelişimi”. [Çevrimiçi]. Available at: <https://www.ekodialog.com/kredi-karti-piyasasi/kredi-kartlarinin-turkiyedeki-gelisimi.html>. [Erişim: 25-May-2019].
- [2] “Bankalar Arası Kart Merkezi”. [Çevrimiçi]. Available at: <https://bkm.com.tr/yerli-ve-yabanci-kredi-kartlarinin-yurt-ici-kullanimi/>. [Erişim: 10-May-2019].
- [3] Y. Özkan, *Veri madenciliği yöntemleri*. Papatya Yayıncılık Eğitim, 2008.
- [4] S. Özekes, “Veri Madenciliği Modelleri ve Uygulama Alanları”, *İstanbul Ticaret Üniversitesi Derg.*, c. 2, sayı 3, ss. 65–82, 2003.
- [5] J. L. Anderson, L. D. Jolly, ve A. E. Fairhurst, “Customer relationship management in retailing : A content analysis of retail trade journals”, c. 14, ss. 394–399, 2007.
- [6] R. Wirth, “CRISP-DM : Towards a Standard Process Model for Data Mining”, sayı 24959.
- [7] Ş. E. Şeker, “CRISP-DM : Endüstriler Arası Standart İşleme – Veri Madenciliği için (Cross Industry Standard Processing – Data Mining)”, 2018.
- [8] K. B. Agyapong, “An Overview of Data Mining Models (Descriptive and Predictive)”, c. 4, sayı 5, ss. 53–60, 2016.
- [9] M. Heinzlmann ve P. P. Bug, “Predictive Analytics Models and Collaborative Filtering”, 2016.
- [10] D. L. Olson, *Descriptive Data Mining*. 2016.
- [11] H. Akpınar, “Veri tabanlarında bilgi keşfi”, *İstanbul Üniversitesi İşletme Fakültesi Derg.*, c. 1, ss. 1–22, 2000.
- [12] C. Bah ve M. M. Yen, “Data Mining Techniques and a Banking Application”, 2011.
- [13] H. Kaya ve K. Köymen, “Veri madenciliği kavramı ve uygulama alanları.” Doğu Anadolu bölgesi araştırmaları, ss. 154–169, 2008.
- [14] V. Aggelis, “Association Rules and Predictive Models for e-Banking Services”, 2003.
- [15] M. Çelik, “Veri Madenciliğinde Kullanılan Sınıflandırma Yöntemleri ve Bir Uygulama”, İstanbul Üniversitesi, 2009.
- [16] X. Amatriain, A. Jaimes, N. Oliver, ve J. M. Pujol, “Data Mining Methods for Recommender Systems”, içinde *Recommender Systems Handbook*, Boston: Springer, 2011, ss. 39–71.

- [17] A. M. Rashid *vd.*, “Getting to Know You: Learning New User Preferences in Recommender Systems”, ss. 127–134, 2002.
- [18] P. Resnick, H. R. Varian, ve G. Editors, “Recommender Systems”, c. 40, sayı 3, ss. 56–58, 1997.
- [19] Z. G. Çabuk, “Anlamsal web üzerinde çoklu etmen tabanlı öneri sistemi”, Ege Üniversitesi, 2013.
- [20] R. Burke, “Hybrid Recommender Systems : Survey and Experiments”, 2002.
- [21] F. O. Isinkaye, “Recommendation systems : Principles , methods and evaluation”, ss. 261–273, 2015.
- [22] G. Jawaheer, M. Szomszor, ve P. Kostkova, “Comparison of Implicit and Explicit Feedback from an Online Music Recommendation Service”, ss. 47–51, 2010.
- [23] D. W. Oard ve J. Kim, “Implicit Feedback for Recommender Systems”, 1997.
- [24] B. Kupisz ve O. Unold, “Collaborative filtering recommendation algorithm based on Hadoop and Spark”, *Proc. IEEE Int. Conf. Ind. Technol.*, c. 2015–June, sayı June, ss. 1510–1514, 2015.
- [25] P. Alencar ve D. Cowan, “The use of machine learning algorithms in recommender systems : A systematic review”, *Expert Syst. Appl.*, c. 97, ss. 205–227, 2018.
- [26] P. Pu ve L. Chen, “A User - Centric Evaluation Framework for Recommender Systems”, ss. 157–164, 2011.
- [27] P. Lops, M. De Gemmis, ve G. Semeraro, *Content-based Recommender Systems : State of the Art and Trends*. 2011.
- [28] S. K. T. Lam, D. Frankowski, ve J. Riedl, “Do You Trust Your Recommendations ? An Exploration of Security and Privacy Issues in Recommender Systems”, ss. 14–29, 2006.
- [29] G. Adomavicius ve A. Tuzhilin, “Toward the Next Generation of Recommender Systems : A Survey of the State-of-the-Art and Possible Extensions”, c. 17, sayı 6, ss. 734–749, 2005.
- [30] A. L. Claudio, “Master of Science Thesis An Analysis of Memory Based Collaborative Filtering Recommender Systems with Improvement Proposals”, sayı September, ss. 1–89, 2014.
- [31] T. Kutuzova ve M. Melnik, “Market basket analysis of heterogeneous data sources for recommendation system improvement”, *Procedia Comput. Sci.*, c. 136, sayı January, ss. 246–254, 2018.
- [32] H. Lu, “Recommendations Based on Purchase Patterns”, c. 4, sayı 6, ss. 4–7, 2014.
- [33] H. Bulut ve M. Milli, “İşbirlikçi filtreleme için yeni tahminleme yöntemleri”,

Pamukkale Üniversitesi Mühendislik Bilim. Derg., c. 22, sayı 2, ss. 123–128, 2016.

- [34] D. Goldberg, D. Nichols, ve B. M. Oki, “Using Collaborative Filtering to Weave an Information Tapestry”, sayı 12, 1992.
- [35] L. Chen, F. Hsu, ve M. Chen, “Developing recommender systems with the consideration of product profitability for sellers”, c. 178, ss. 1032–1048, 2008.
- [36] A. L. Claudio, “An Analysis of Memory Based Collaborative Filtering Recommender Systems with Improvement Proposals”, 2014.
- [37] M. Suchak, N. Iacovou, P. Resnick, P. Bergstrom, ve J. Riedl, “GroupLens”, ss. 175–186, 2004.
- [38] P. Maes ve M. S. Sharadanand, “Social information filtering: algorithms for automating”, *Word Mouth*”, *CHI Proc*, ss. 210–217, 1995.
- [39] S. K. Gorakala ve M. Usielli, *Building a Recommendation System with R*. 2015.
- [40] B. Sarwar, G. Karypis, J. Konstan, ve J. Riedl, “Item-based Collaborative Filtering Recommendation Algorithms”, ss. 1–15, 2001.
- [41] J. Wang, A. P. De Vries, ve M. J. T. Reinders, “Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion Categories and Subject Descriptors”, sayı January, 2006.
- [42] F. Ricci, L. Rokach, B. Shapira, P. B. Kantor, ve F. Ricci, *Recommender Systems Handbook*. 2015.
- [43] Charu C. Aggarwal, *Recommender Systems*. 2017.
- [44] S. K. Gorakala, *Building Recommendation Engines*. Packt Publishing, 2016.
- [45] K. Falk, *Practical Recommendation Systems*. 2019.
- [46] L. M. De Campos, J. M. Fernández-Luna, J. F. Huete, ve M. A. Rueda-Morales, “Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks”, *Int. J. Approx. Reason.*, c. 51, sayı 7, ss. 785–799, 2010.
- [47] U. Modeling ve R. Burke, “Hybrid Recommender Systems : Survey and Experiments Hybrid Recommender Systems ”:, sayı November 2002, 2014.
- [48] M. Işık ve A. Y. Çamurcu, “Web Belgeleri Kümelemede Benzerlik Ve Uzaklık Ölçütleri Başarılarının Karşılaştırılması”, c. 20, sayı 2008, ss. 35–49, 2009.
- [49] S. Khusro, Z. Ali, ve I. Ullah, “Recommender Systems : Issues , Challenges , and Research Opportunities Recommender Systems : Issues , Challenges , and Research Opportunities”, sayı March, 2016.
- [50] B. Kumar, “Approaches , Issues and Challenges in Recommender Systems : A Systematic Review”, sayı January, 2017.

- [51] A. Bessa, A. H. F. Laender, A. Veloso, ve N. Ziviani, “Alleviating the sparsity problem in Recommender Systems by exploring underlying user communities”, *CEUR Workshop Proc.*, c. 866, ss. 35–47, 2012.
- [52] M. Grčar, D. Mladenič, B. Fortuna, ve M. Grobelnik, “Data Sparsity Issues in the Collaborative Filtering Framework”, ss. 58–76, 2006.
- [53] Y. Chen, C. Wu, M. Xie, ve X. Guo, “Solving the sparsity problem in recommender systems using association retrieval”, *J. Comput.*, c. 6, sayı 9, ss. 1896–1902, 2011.
- [54] J. J. P. Arias, *Recommender Systems for the Social Web*, c. 32. 2012.
- [55] P. Adamopoulos ve A. Tuzhilin, “On over-specialization and concentration bias of recommendations”, ss. 153–160, 2014.
- [56] B. Sarwar, G. Karypis, J. Konstan, ve J. Riedl, “TR 00-043 Application of Dimensionality Reduction in Recommender System - A Case Study”, 2000.
- [57] X. Su ve T. M. Khoshgoftaar, “A Survey of Collaborative Filtering Techniques”, c. 2009, sayı Section 3, 2009.
- [58] M. A. Ghazanfar ve A. Prügel-Bennett, “Leveraging clustering approaches to solve the gray-sheep users problem in recommender systems”, *Expert Syst. Appl.*, c. 41, sayı 7, ss. 3261–3275, 2014.
- [59] M. Ghazanfar ve A. Prugel-Bennett, “Fulfilling the needs of gray-sheep users in recommender systems, a clustering solution”, 2011.
- [60] T. Kumari ve P. Bedi, “A Comprehensive Study of Shilling Attacks in Recommender Systems”, *Int. J. Comput. Sci. Issues*, c. 14, sayı 4, ss. 44–50, 2017.
- [61] W. Bhebe ve O. P. Kogeda, “Shilling attack detection in Collaborative Recommender Systems using a Meta Learning strategy”, *Proc. 2015 Int. Conf. Emerg. Trends Networks Comput. Commun. ETNCC 2015*, ss. 56–61, 2015.
- [62] R. Burke, B. Mobasher, C. Williams, ve R. Bhaumik, “Classification features for attack detection in collaborative recommender systems”, s. 542, 2006.
- [63] M. S. Ackerman, L. F. Cranor, F. Park, ve J. Reagle, “Privacy in E-Commerce : Examining User Scenarios and Privacy Preferences”, ss. 1–8, 1999.
- [64] A. Friedman, B. Knijnenburg, ve K. Vanhecke, *Privacy Aspects of Recommender Systems Privacy Aspects of Recommender Systems*, sayı January. 2015.
- [65] Organisation for Economic Co-operation and Development, “OECD guidelines governing the protection of privacy and transborder flows of personal data of 23 September 1980 (C(80)58/FINAL as amended on 11 July 2013 by C(2013)79)”, ss. 11–37, 2013.
- [66] N. Friedman, D. Geiger, ve M. Goldszmit, “Bayesian Network Classifiers Overfitting

- and Underfitting With Machine Learning Algorithms (no date). Available at: <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/> (Accessed: 1 July 2018).”, *Mach. Learn.*, c. 29, ss. 131–163, 1997.
- [67] W. Kaye-Blake, F. . Li, A. M. Martin, A. McDermott, H. Neil, ve S. Rains, “A review of Multi-Agent Simulation Models in Agriculture Paper presented at the 2009 NZARES Conference”, *NZARES Conf.*, sayı August, ss. 1–10, 2009.
- [68] H. Parvin, P. Moradi, ve S. Esmaeili, “TCFACO : Trust-aware collaborative filtering method based on ant colony optimization”, *Expert Syst. Appl.*, c. 118, ss. 152–168, 2019.
- [69] N. Smar, J. A. Konstan, N. Borchers, J. Heriocker, B. Wer, ve J. Wew, “Using FilteringAgents to ImprovePrediction Qualityin the GroupLens Research CollaborativeFilteringSystem”, 1998.
- [70] A. Felfernig, L. Boratto, M. Stettinger, ve M. Tkalcic, *Group Recommender Systems An Introduction Series editors*. 2018.
- [71] I. Kulkarni, P. Gandhi, ve P. Karlekar, “Book Recommendation System Using Apache Spark”, ss. 7982–7987, 2017.
- [72] S. Panigrahi, R. Ku, ve A. Stitipragyan, “A Hybrid Distributed Collaborative Filtering Recommender Engine Using Apache Spark”, *Procedia - Procedia Comput. Sci.*, c. 83, sayı BigD2M, ss. 1000–1006, 2016.
- [73] J. Lee, C. Jun, J. Lee, ve S. Kim, “Classification-based collaborative filtering using market basket data”, c. 29, ss. 700–704, 2005.
- [74] G. Agarwal, H. Bahuguna, ve A. Agarwal, “Solving Cold-Start Problem in Recommender System Using User”, c. 8, sayı 1, ss. 55–61, 2017.
- [75] L. Safoury ve A. Salah, “Exploiting User Demographic Attributes for Solving Cold-Start Problem in Recommender System”, c. 1, sayı 3, ss. 1–5, 2013.
- [76] S. Sedhain, S. Sanner, N. Anu, ve D. Braziunas, “Social Collaborative Filtering for Cold-start Recommendations”, ss. 3–6, 2014.
- [77] Z. Majó ve B. Révész, “The role of collaborative filtering in marketing”, ss. 59–71, 2011.
- [78] “Empirical Analysis of Predictive Algorithms for Collaborative Filtering”, ss. 43–52, 1998.
- [79] G. Dror ve Y. Koren, “Yahoo ! Music Recommendations : Modeling Music Ratings with Temporal Dynamics and Item Taxonomy”, 2011.
- [80] M. Pacula, “A Matrix Factorization Algorithm for Music Recommendation using Implicit User Feedback”, 2009.

- [81] “<https://www.creditcards.com/credit-card-news/glossary/term-merchant-category-codes-mcc.php>”. [Çevrimiçi]. Available at: <https://www.creditcards.com/credit-card-news/glossary/term-merchant-category-codes-mcc.php>. [Erişim: 21-Şub-2019].
- [82] “VISA MERCHANT CATEGORY CLASSIFICATION (MCC) CODES DIRECTORY”. [Çevrimiçi]. Available at: <https://www.albaraka.com.tr/pdf/SEKTOR-MCC-KODLARI.pdf>. [Erişim: 18-Nis-2019].
- [83] “Medium.com”. [Çevrimiçi]. Available at: <https://medium.com/datadriveninvestor/how-to-build-a-recommendation-system-for-purchase-data-step-by-step-d6d7a78800b6>. [Erişim: 28-Mar-2019].
- [84] “Turi Create”. [Çevrimiçi]. Available at: <https://apple.github.io/turicreate/docs/api/>. [Erişim: 20-Oca-2019].



ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Tuğçe Süheyla ÇETİN
Doğum Yeri	İstanbul
Doğum Tarihi	25.04.1987
Uyruğu	<input checked="" type="checkbox"/> T.C. <input type="checkbox"/> Diğer:
Telefon	
E-Posta Adresi	tugces_kaya@outlook.com
Web Adresi	



Eğitim Bilgileri	
Lisans	
Üniversite	İstanbul Üniversitesi
Fakülte	Fen Fakültesi
Bölümü	Fizik
Mezuniyet Yılı	2009

Yüksek Lisans	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Enformatik
Programı	Enformatik