



AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



Ömer UÇAN

ÖNERİ SİSTEMİ MODELLERİNDE KULLANILAN MAKİNE ÖĞRENMESİ
TEKNİKLERİNİN MÜŞTERİ SATIN ALMA TERCİHLERİ DOĞRULTUSUNDA
KARŞILAŞTIRILMASI

İşletme Ana Bilim Dalı
Doktora Tezi

Antalya, 2020



AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



Ömer UÇAN

ÖNERİ SİSTEMİ MODELLERİNDE KULLANILAN MAKİNE ÖĞRENMESİ
TEKNİKLERİNİN MÜŞTERİ SATIN ALMA TERCİHLERİ DOĞRULTUSUNDA
KARŞILAŞTIRILMASI

Danışman

Prof. Dr. Can Deniz KÖKSAL

İşletme Ana Bilim Dalı

Doktora Tezi

Antalya, 2020

Akdeniz Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğüne,

Ömer UÇAN'ın bu çalışması jürimiz tarafından İşletme Ana Bilim Dalı Doktora Programı tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Mustafa Zihni TUNCA (İmza)

Üye (Danışmanı) : Prof. Dr. Can Deniz KÖKSAL (İmza)

Üye : Prof. Dr. Gökhan AKYÜZ (İmza)

Üye : Doç. Dr. Sezgin IRMAK (İmza)

Üye : Doç Dr. Meltem KARAATLI (İmza)

Tez Başlığı: Öneri Sistemi Modellerinde Kullanılan Makine Öğrenmesi Tekniklerinin Müşteri Satın Alma Tercihleri Doğrultusunda Karşılaştırılması
--

Onay: Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Tez Savunma Tarihi : 27/01/2020

Mezuniyet Tarihi : 20/02/2020

(İmza)
Prof. Dr. İhsan BULUT
Müdür

AKADEMİK BEYAN

Doktora Tezi olarak sunduđum “Öneri Sistemi Modellerinde Kullanılan Makine Öğrenmesi Tekniklerinin Müşteri Satın Alma Tercihleri Doğrultusunda Karşılaştırılması” adlı bu çalışmanın, akademik kural ve etik değerlere uygun bir biçimde tarafımda yazıldığını, yararlandığım bütün eserlerin kaynakçada gösterildiğini ve çalışma içerisinde bu eserlere atıf yapıldığını belirtir; bunu şerefimle doğrularım.

Ömer UÇAN





T.C.
AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
TEZ ÇALIŞMASI ORJİNALLİK RAPORU
BEYAN BELGESİ



SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

ÖĞRENCİ BİLGİLERİ	
Adı-Soyadı	Ömer UÇAN
Öğrenci Numarası	20118604101
Enstitü Ana Bilim Dalı	İşletme
Programı	Doktora
Programın Türü	() Tezli Yüksek Lisans (X) Doktora () Tezsiz Yüksek Lisans
Danışmanın Unvanı, Adı-Soyadı	Prof. Dr. Can Deniz KÖKSAL
Tez Başlığı	Öneri Sistemi Modellerinde Kullanılan Makine Öğrenmesi Tekniklerinin Müşteri Satın Alma Tercihleri Doğrultusunda Karşılaştırılması
Turnitin Ödev Numarası	1256659506

Yukarıda başlığı belirtilen tez çalışmasının a) Kapak sayfası, b) Giriş, c) Ana Bölümler ve d) Sonuç kısımlarından oluşan toplam 134 sayfalık kısmına ilişkin olarak, 13/01/2020 tarihinde tarafımdan Turnitin adlı intihal tespit programından Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nda belirlenen filtrelemeler uygulanarak alınmış olan ve ekte sunulan rapora göre, tezin/dönem projesinin benzerlik oranı;

alıntılar hariç % 2

alıntılar dahil % 2'dir.

Danışman tarafından uygun olan seçenek işaretlenmelidir:

(X) Benzerlik oranları belirlenen limitleri aşmıyor ise;

Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporu'nun doğruluğunu onaylarım.

() Benzerlik oranları belirlenen limitleri aşıyor, ancak tez/dönem projesi danışmanı intihal yapılmadığı kanısında ise;

Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporu'nun doğruluğunu onaylar ve Uygulama Esasları'nda öngörülen yüzdeler sınırlarının aşılmasına karşın, aşağıda belirtilen gerekçe ile intihal yapılmadığı kanısında olduğumu beyan ederim.

Gerekçe:

Benzerlik taraması yukarıda verilen ölçütlerin ışığı altında tarafımda yapılmıştır. İlgili tezin orijinallik raporunun uygun olduğunu beyan ederim.

13/02/2020

Prof. Dr. Can Deniz
KÖKSAL

İÇİNDEKİLER

ŞEKİLLER LİSTESİ	iv
TABLOLAR LİSTESİ	v
KISALTMALAR.....	vi
ÖZET	vii
SUMMARY	viii
TEŞEKKÜR.....	ix
GİRİŞ.....	1

BİRİNCİ BÖLÜM

ÖNERİ SİSTEMLERİ VE MODELLEMELERİNE YÖNELİK TEMEL BİLGİLER

1.1. Öneri Sistemlerinin Temel Bileşenleri ve Kullanım Hedefleri	5
1.2. Öneri Sistemlerinin Özellikleri	7
1.2.1. Kullanıcı Tercihleri	8
1.2.2. Tahmin Tutarlılığı	8
1.2.3. Tahmin Değerleri	9
1.2.4. Ölçümlerin Sıralanması.....	11
1.3. Öneri Sistemlerinin Kapsamı	11
1.3.1. Ürün Kapsama Alanı	12
1.3.2. Kullanıcı Kapsama Alanı	12
1.3.3. Soğuk Başlangıç Problemi	12
1.4. Öneri Sistemlerinde Kullanılan Veri Yapıları.....	13
1.5. Öneri Sistemleri Uygulama Alanları.....	15
1.6. Öneri Sistemlerinde Temel Modeller	18
1.6.1. İşbirlikçi Filtreleme Modelleri	18
1.6.2. İçerik Temelli Öneri Modelleri	22
1.6.3. Bilgi Temelli Filtreleme Modelleri	23
1.6.4. Topluluk Temelli Filtreleme Modelleri.....	24
1.6.5. Hibrit Öneri Modelleri	25
1.7. Öneri Sistemi Modellerinin Gelişimi ve Literatür Taraması.....	25

İKİNCİ BÖLÜM

İŞBİRLİKÇİ FİLTREME YÖNTEMLERİ VE MODEL OLUŞTURMA SÜRECİ

2.1. İşbirlikçi Filtreleme Süreci	34
---	----

2.2. Değerlendirme Matrisleri	35
2.3. Veri Hazırlık ve İşleme Süreçleri	37
2.4. İşbirlikçi Filtreleme Yöntem Çeşitleri	37
2.4.1. Hafıza Temelli İşbirlikçi Filtreleme Modelleri	38
2.4.2. Model Temelli İşbirlikçi Filtreleme Teknikleri	44
2.4.3. Model Temelli İşbirlikçi Filtreleme ve Naive Bayes Sınıflandırıcısı	47
2.4.4. Model Temelli İşbirlikçi Filtreleme ve Boyut İndirgeme Yöntemleri	48
2.4.5. Model Temelli Komşuluk Modelleri ve Birliktelik Kuralları	53
2.5. Hafıza ve Model Temelli Modellerinin Karşılaştırılması	57

ÜÇÜNCÜ BÖLÜM

METODOLOJİ VE UYGULAMA

3.1. Araştırma Metodolojisi	59
3.2. Araştırmanın Konusu ve Önemi	60
3.3. Araştırmanın Amacı	61
3.4. Araştırmanın Kısıtları	63
3.5. Araştırmanın Kapsamı	64
3.6. Araştırma Modeli	66
3.7. Örneklem Süreci ve Kullanılan Veri Seti	67
3.7.1 RFM Analizi	71
3.7.1.1. Satın Alma Matrisi	74
3.7.2. Veri Setinin Eğitilmesi ve Test Verilerinin Belirlenmesi	75
3.7.3. Değerlendirme Ölçümleri	77
3.8. Araştırmada Kullanılan Yöntemler	79
3.8.1. Benzerlik Ölçümü Uygulaması	81
3.8.2. Matris Çarpanlara Ayırma Metodu	85
3.8.3. Birliktelik Kuralları	89
3.9. Yöntemlerin Karşılaştırılması	94
SONUÇ	101
KAYNAKÇA	104
EK-1 BENZERLİK ÖLÇÜMÜ MODELİ HATA MATRİSLERİ ORTALAMA DEĞERLERİ	117
EK-2 MATRİS AYRIŞTIRMA MODELİ HATA MATRİSLERİ ORTALAMA DEĞERLERİ	118

EK-3 BİRLİKTELİK KURALLARI MODELİ HATA MATRİSLERİ ORTALAMA DEĞERLERİ.....	119
ÖZGEÇMİŞ.....	120



ŞEKİLLER LİSTESİ

Şekil 1.1.1 İşbirlikçi Filtreleme Çalışma Modeli	19
Şekil 1.2 Öneri Sisteminin Literatürde Kullanım Alanları.....	29
Şekil 1.3 Literatürde Öneri sistemlerinde Kullanılan Metotların Dağılımı.....	30
Şekil 2.1 Fp-Growth İçerisinde Fp-tree ağaç yapısı örneği.....	55
Şekil 3.1 Veri Seti İçerisindeki Tekrarlanan Satış ile Ürün Çeşidi Sayısı	69
Şekil 3.2 Veri Seti İçerisindeki Tekrarlanan Alımlar ile Müşteri / Kümülatif Müşteri Sayısı..	72
Şekil 3.3 Veri Seti İçerisindeki Tekrarlanan Satın Alımlar ile Satış/Kümülatif Satış.....	73
Şekil 3.4 Jaccard ve Kosinüs Benz. Ölçümleri Kesinlik/Hassasiyet Değerlerinin Gösterimi..	85
Şekil 3.5 TDA ve NMF Metotları Kesinlik/Hassasiyet Değerlerinin Gösterimi	88
Şekil 3.6 PSO ve FpGrowth Metotları Kesinlik/Hassasiyet Değerlerinin Gösterimi	94
Şekil 3.7 Araştırmada Kullanılan Metotların Kesinlik Değeri Karşılaştırması.....	98
Şekil 3.8 Araştırmada Kullanılan Metotların Hassasiyet Değeri Karşılaştırması	99
Şekil 3.9 Araştırmada Kullanılan Metotların F1 Değeri Karşılaştırması	99
Şekil 3.10 Araştırmada Kullanılan Metotların Sürelere Göre Skor Karşılaştırması	100

TABLOLAR LİSTESİ

Tablo 1.1 Hata Matrisi Standart Tablosu.....	10
Tablo 3.1 Referans Günlere Göre Oluşturulan Alt Veri Setleri Özellikleri	77
Tablo 3.2 Model Değerlendirmesinde Kullanılan Hata Matrisi Tablosu	78
Tablo 3.3 Benzerlik Ölçüm Model Algoritması.....	82
Tablo 3.4 Jaccard ve Kosinüs Benz. Ölçümü 45 Günlük Metrik Değerlendirmesi	83
Tablo 3.5 Jaccard ve Kosinüs Benz. Ölçümü 90 Günlük Metrik Değerlendirmesi	83
Tablo 3.6 Jaccard ve Kosinüs Benz. Ölçümü 365 Günlük Metrik Değerlendirmesi	83
Tablo 3.7 Jaccard ve Kosinüs Benzerlik Ölçümü Metrik Değerlendirmesi	83
Tablo 3.8 TDA ve NMF Metotlarının 45 Günlük Metrik Değerlendirmesi.....	87
Tablo 3.9 TDA ve NMF Metotlarının 90 Günlük Metrik Değerlendirmesi.....	87
Tablo 3.10 TDA ve NMF Metotlarının 365 Günlük Metrik Değerlendirmesi.....	87
Tablo 3.11 TDA ve NMF Metotlarının Metrik Değerlendirmesi.....	87
Tablo 3.12 Fp-Growth Model Oluşturma Algoritması.....	90
Tablo 3.13 PSO Model Oluşturma Algoritması	91
Tablo 3.14 PSO ve FpGrowth Metotlarının 45 Günlük Metrik Değerlendirmesi.....	92
Tablo 3.15 PSO ve FpGrowth Metotlarının 90 Günlük Metrik Değerlendirmesi.....	92
Tablo 3.16 PSO ve FpGrowth Metotlarının 365 Günlük Metrik Değerlendirmesi.....	92
Tablo 3.17 PSO ve FpGrowth Metotlarının Metrik Değerlendirmesi.....	92
Tablo 3.18 Kesinlik, Hassasiyet ve F1 Skorlarının Toplu Gösterimi.....	95

KISALTMALAR

DP	: Doğru Pozitif
EM	: Beklenti Maksimizasyon Algoritması
IDF	: Inverse Document Frequency (Ters Doküman Frekansı)
KL	: Kullback-Leibler Uzaklığı
KNIME	: Konstanz Information Miner
MÖ	: Makine Öğrenme
NDCG	: Normalized Cumulative Discounted Gain (Normalleştirilmiş Kümülatif İskontolu Kazanç)
NMF-NNMF	: Non Negative Matrix Factorization (Negatif Olmayan Matris Ayırıştırması)
PCA	: Principle Component Analysis (Temel Bileşen Analizi)
PSO	: Parçacık Sürü Optimizasyonu (Particle Swarm Optimization)
RFM	: Recency, Frequency and Monetary (Yenilik, Sıklık ve Mali)
ROC	: Receiver Operating Characteristic (Alıcı İşletim Karakteristik Eğrisi)
SPSS	: Statistical Package for the Social Sciences
SVD	: Singular Value Decomposition (Tekil Değer Ayırıştırması)
TBA	: Temel Bileşen Analizi
TDA	: Tekil Değer Ayırıştırması
WEKA	: Waikato Environment for Knowledge Analysis
WWW	: World Wide Web
YN	: Yanlış Negatif
YP	: Yanlış Pozitif

ÖZET

E-ticarete öneri sistemleri, müşterilerden elde edilen veriler ile ürün özelliklerini değerlendirerek, müşterilerin tercih edebileceği ürünlerin ortaya çıkarılmasını amaçlayan modeller bütünüdür. Öneri sistemlerinde, kullanıcı ve nesne ilişkisi değerlendirilerek kullanıcıların davranışları doğrultusunda yönelim gösterebilecekleri nesnelerin tahmini sağlanmaktadır. Araştırmada e-ticaret sektöründeki uygulamalar ile bu uygulamaların yarattığı çeşitlilik incelenerek, kullanıcıların satın alma davranışları üzerinden öneri modelleri kıyaslanmaya çalışılmıştır.

Araştırmada, öneri sistemleri için kullanılan yöntemler irdelenmiş ve bu tür sistemlerin gelişimi hakkında bilgiler verilmiştir. Farklı kullanıcı ürün ilişkileri ve model oluşturma tekniklerinin yer aldığı araştırmalar doğrultusunda kullanıcıların satın alma davranışlarının, kullanıcıların doğrudan ürün hakkında değerlendirmesi olarak kabul edildiği bir uygulama gerçekleştirilmiştir. Bu doğrultuda literatürde kullanılan öneri sistemi yöntemlerinden bazıları seçilerek öneri sistemi modellerinin karşılaştırılması yapılmıştır. Ürün temelli komşuluk modelleri ile kullanılan Jaccard ve Kosinüs benzerlik ölçümü, matris ayrıştırma yöntemlerinden TDA (Tekil Değer Ayrıştırması) ve NMF (Negatif Olm. Matris Ayrıştırması) yöntemleri ile birliktelik kuralları içerisinde yer alan Fp-Growth ve PSO (Parçacık Sürü Optimizasyonu) yöntemleri hem kendi içlerinde hem de birbiri ile olan performansları karşılaştırılmıştır.

Bu yöntemler sonucu elde edilen çıktılar karşılaştırılırken belirli zaman aralıkları da sonuçların değerlendirilmesinde kullanılmıştır. Belirtilen yöntemlerle elde edilen çıktılar, belirlenen referans günü doğrultusunda kullanıcıların 45, 90 ve 365 günlük dönemler ile tüm veri seti kapsamında tekrar alımları ile ilişkilendirilerek skor ölçümleri hesaplanmıştır. Daha sonra sonuçların tutarlılığını test etmek için 10'ar günlük zaman dilimleri için modeller tekrarlanarak çalıştırılmış ve elde edilen sonuçların sağlamaları yapılmıştır.

Bu sonuçlar neticesinde hem kısa zaman aralıklarında hem de geniş zaman aralıklarında kullanıcıların satın alma davranışlarında oluşan ikili veri tipine sahip modellerde birliktelik kuralları Fp-Growth ve PSO'nun daha başarılı sonuçlar elde ettiği gözlemlenmiştir. Birliktelik kuralları diğer metotlara kıyasla daha iyi bir sonuç elde etmesine rağmen kendi içerisindeki performansı kısa zaman aralıklarında daha yüksek olduğu da zaman aralığı kriteri sayesinde ortaya çıkarılmıştır.

Anahtar Kelimeler: E-Ticaret, Öneri Sistemleri, İşbirlikçi Filtreleme, Benzerlik Ölçümleri, Birliktelik Kuralları

SUMMARY

COMPARISON OF MACHINE LEARNING TECHNIQUES USED IN THE RECOMMENDER SYSTEM MODELS IN ACCORDANCE WITH CUSTOMER PURCHASE PREFERENCES

Recommender systems in e-commerce are the collection of models that aim to reveal products that customers would prefer after evaluating product features via information obtained from customers. In the recommender systems, the efforts made in order to guess the objects which users have a tendency to prefer, based on the evaluation of user and object relationship. In this research, applications in e-commerce sector and the diversity created by them have been analyzed and recommender systems are tried to be compared based on users' purchasing behaviors.

In the study, the methods that are used for recommender systems are investigated and information regarding the development of these kinds of systems have been presented. In the light of many studies about different user – product relations and model establishment techniques, an application that acknowledges users' purchasing behaviors as direct product reviews of users, has been developed. In this regard, by selecting some of the recommender system methods from the literature, developed models have been compared. The model performances such as Jaccard and Cosine similarity measurement used with product-based neighborhood models, SVD and NMF methods from matrix separation methods and the Fp-Growth and PSO methods included in the association rules were evaluated by comparing within each other and also with each other.

While comparing the results obtained through these results, some time periods also used in evaluation of findings as well. Score measurements have been calculated by associating re-purchases of users in the scope of all data sets per 45, 90, 365 days periods in line with determined reference day and the findings obtained via these methods. After that, in order to test the consistency of the results, the models were run repeatedly for 10 day periods and crosscheck is sustained.

According to the results obtained, both in short term and in long term, in the binary data type models belonging to users' purchase behaviors, association rules of Fp-Growth and PSO obtained more successful results. Although the association rules have a better outcome relative to other methods, the time interval criterion has disclosed that their performance in itself is higher in short time intervals.

Keywords: E-commerce, Recommender Systems, Collaborative Filtering, Similarity Measures, Association Rules

TEŞEKKÜR

Uzun tez çalışma sürecimde her daim destek olan tez danışmanım Prof. Dr. Can Deniz KÖKSAL'a, Tez İzleme Komitesinde olup görüş ve önerileriyle yardımcı olan Prof. Dr. Gökhan AKYÜZ ve Doç. Dr. Sezgin IRMAK hocalarıma teşekkür ederim. Ayrıca, bu süreçte maddi ve manevi olarak her türlü kolaylığı gösteren ailem ve arkadaşlarıma anlayışları için teşekkür ederim.

Ömer UÇAN
Antalya, 2020



GİRİŞ

Günlük hayatın hızına erişmenin, teknolojiye ayak uydurmakla paralellik gösterdiği bu günlerde, yaşanan anları zaman algısıyla anımsamak yerine veri parçalarına dönüştürerek akıllı cihazlar üzerinden erişmenin daha popüler olduğu bir dönemde yaşamlar sürdürülmektedir. İnsanların kendileri için herhangi bir anı boş geçirme korkusu, topluma ayak uydurma içgüdüğü insanoğluna her anın birer veri izdüşümünün oluşmasına razı olmaktan başka çare bırakmamaktadır. Günlük hayattaki tüm hareket ve davranışları kanıtlamak için artık kalp atışlarından göz hareketlerine, konuşulan kelimelerden, refleksif hareketlerin zaman aralığına kadar ölçen ve analiz eden makinelerle ihtiyaç duyulmaktadır. Bu durum kötü bir senaryoyu işaret ettiği düşünülmemeli, aksine hayat kalitesini arttırmak için bir fırsat olarak ilgiyle peşine düşülmelidir. Her organizmanın, her anının değerinin veri ile ölçülmesi ve zamanla benlik üzerindeki tüm bağımsız haklara ek olarak oluşturulan veri yığınları da insanoğlunun ayrılmaz bir parçası haline gelmesi kaçınılmaz bir durumdur.

Tüm bu süreç içerisinde gündelik ihtiyaçlar konusunda teknolojiye bağımlılık hızla artarken, teknolojinin de kişiliğe bürünerek insanı tanıma iç görüşü gittikçe büyümektedir. Şirketler tarafından müşterilerine en kolay ve hızlı erişim aracı olarak görülen ve tartışma götürmez bir gerçek olan internet ile yan teknolojileri, bambaşka bir pazarlama algısının oluşmasına yol açmaktadır (Huang, Z. vd., 2007). Müşterilerin gelişimlere hızlı adapte olması, şirketlerin de hantal yapılarının dışına çıkarak anlık ve değişkenlik gösterebilen bir adaptasyon süreci edinmelerine mecbur bırakmıştır. İnsanlığın kısıtlı algısının ötesinde yeni bir bakış açısı ile ortaya çıkarılması gerekmektedir. Her davranışın analiz edildiği ve bu analizleri insanı taklit edercesine yorumlayabilen makinelerin oluşturulması, teknoloji şirketlerinin en büyüğünden küçüğüne kadar hedeflerinin en üst basamaklarında yer almaktadır. Makine öğrenme teknikleri, istatistiksel analize dayalı bir hesaplama birikimi üzerinden oluşturulmakta, deneysel çalışmalarının başarısı sistematik bir kontrol mekanizmasıyla doğrulanmakta ve gerçek hayat uygulamalarında başarıya ulaşmasında yol gösterici olmaktadır (Portugal vd., 2018: 210).

Bu doğrultuda e-ticaret sektörünün hızlı ve ön görülmesi oldukça zor olan büyüme alanı, ister istemez müşterilerinin gerçekten neye ihtiyacı olduğunun tespiti ve müşterilere neler satabilirim öngörüsünün somutlaştırılması için şirketleri yeni yollar arayışına sürüklemiştir (Schafer vd., 1999: 158). Kullanıcıların satın almak istedikleri ürünle ilgili kusursuz bir bilgi birikimi ile alışveriş yapmadığının, bunun da kullanıcı tercihlerinin değişkenlik gösterebileceğinin bilinmesi, e-ticaret şirketlerinin müşterilerinin yönlendirebilmesi için makine öğrenme tekniklerine ihtiyaç duymasına neden olmuştur. Müşterinin ihtiyaç duyduğu

ancak henüz farkında olmadığı ürünlere yönlendirilmesi, yeni bir ürün segmentinin doğru müşteri gruplarına ön yargı oluşturulmadan sunulmasının sağlanması ve kullanıcılarının yeni öğrenme teknikleri ile satın alma işlemini kusursuz bir süreçle tamamlayabilmesi istatistiksel analizlere dayalı bir makine öğrenmesi sisteminin çıktıları ile başarılabilir.

E-ticaret şirketleri için tüm bu değişim içerisindeki sürecin beslenebilmesi açısından literatürdeki farklı varsayım ve algoritmalar çeşitlendirilerek yeni perspektiflerden konular tekrar ele alınmaktadır (Park, D. H. vd., 2012). Bu doğrultuda müşterilerin başka ürünlerin farkına varabilmesi açısından öneri sistemleri önemli bir araç olarak e-ticaret şirketlerince sıklıkla kullanılabilmektedir.

Öneri sistemleri ile kullanıcı davranışları ve ürün özellikleri analiz edilerek kullanıcının satın alma ihtimali olan ürün yelpazesi ortaya çıkarılabilmektedir. Bu oldukça çetrefilli ve yüksek işlem gücü gerektiren bir süreç sonucunda ortaya çıkmaktadır. Bilgisayarlardaki işlem gücü sorunun günbegün ortadan kalkması, istatistiksel analizler aracılığıyla daha tutarlı çıkarımlar elde edilmesi konusunda analistler ve araştırmacılara daha fazla sorumluluk yüklemekte ve literatürün gelişmelerin hızlanarak artmasına olanak sağlamaktadır.

Öneri sistemlerinde genel çıkarımların yapıldığı araştırmaların oldukça fazla olması, son yıllardaki çalışmalarda farklılaşmalara gidilmesine uygun bir ortam oluşturmuş ve farklı veri tipleri, sistemler ve nesneler için çeşitli çalışmalar hız kazanmıştır (Alyari ve Navimipour, 2018: 205). Araştırma bu varsayımları kabul ederek öneri sistemleri için kullanıcıların satın alma davranışları inceleyerek gerçekleştirilmiştir. Bu yönelimin sebeplerinden birisi satın alma davranışlarının basit ve kolay gözlemlenebilir bir veri yapısına sahip olmasına rağmen, müşterinin doğrudan ürün hakkındaki net tavrını herhangi bir değerlendirmeye ihtiyaç duymadan kestirilmesine olanak vermesidir. Müşteri ürünü satın alarak bu ürünün kendisi için yararlı olabileceğini kabul ettiği ve bu ürün hakkında yeterli bir bilgi birikimine sahip olduğu varsayımı üzerinden analizler gerçekleştirilmiştir.

Kullanıcıların satın alma geçmişi verilerinin, öneri sistemleri içinde verimli kestirimlerin gerçekleştirilebilmesi için araştırmada birden fazla metot üzerinden model oluşturma yolu seçilmiştir. Böylece elde edilen çıktı değerlerinden çok, bu metotlar birbiri ile kıyaslanarak bu tür veri tipleri için başarılı bir modelleme seçeneğinin geliştirilmesi üzerine durulmuştur.

Bir diğer kriter olarak da zaman aralıkları modellemeye dahil edilmiştir. Bu durum tahmin çıktılarının başarılarının ölçümü için belirli zaman aralıkları ile kısıtlanarak tahmin aralığı oluşturma performansına etkilerinin test edilmesine olanak vermiştir. Böylelikle hangi metodun hangi zaman aralığında daha verimli olduğunun gözlemlenebilmesinin önü açılmıştır.

Bu tür arařtırmaların gerekleřtirilmesi ve bařarılı sonular elde edilmesi, sadece akademik aba ile mmkn deėildir. Akademiden beslenen sektrlerin sadece ıktılar zerine yoėunlařarak elde edilen kazanımları kullanması da srdrlebilir bir durum deėildir. Bu srecin devamlılıėının saėlanması iin kendi ierisinde bir dngnn oluřturulması řarttır. Bilime ilginin ve dolayısıyla yařam seviyesi standartlarının yksek olduėu lkelerde, sz konusu dng kısmen saėlıklı iřleyebilmektedir. Ancak bilime olan inancın daha ařaėıda olduėu ve lke řartlarının daha ok hızlı tketime ynlendirdiėi lkelerde bu geri beseleme iřlemlerinin gerekleřtirilmesi zorlukla saėlanabilmektedir. Arařtırma iin veri elde edininim sreci, arařtırma ve ėrenme srecinden ok daha zor ve yer yer umutsuzluėa kapılmaya msait bir dnem oldu. Bu durum sektr ierisindeki karar vericilerin veri paylařma konusunda rızasız olması, arařtırmaların verimli olmadığına inancı ve lkedeki veri gizliliėi konusundaki hukuki bořlukların karar vericilerin gzn korkutmasından kaynaklanmakta olduėu sıklıkla gzlemlenmiřtir. Bu konuda herhangi bir eėitim ve destekleme projelerinin olmaması gerek anlamda bir arařtırma yapmanın zorluėunu ciddi bir biimde arttırmaktadır.

Deėiřen insan doėası ve gnmzde buna baėlı oluřan teknoloji hkimiyeti, akademik dnyanın bu tr alıřmalara olan alıėının hibir zaman bitmeyeceėinin bir gstergesi olarak kabul edilmiřtir. alıřma, makine ėrenme teknikleri ile elde edilebilecek bir tahmin sistemine farklı bir bakıř aısından bakarak ve geliřtirilerek kullanılabilecek farklı bir yolu desteklemek zerine kurgulanmıřtır. alıřma ile elde edilen sonular farklı modellerle bir araya getirilerek geliřime aık bir kaynak olması umuduyla sunulmuřtur.

Bu tez alıřması  blm olarak planlanmıřtır. Birinci blmde neri sistemlerinin temel yapısı ve geliřimi hakkında bilgilere yer verilmiř ve kısa srede oluřan metodolojinin genel atısı iřlenmiřtir. İkinci blm ise arařtırma iin belirlenen spesifik alana ynelik detaylı bilgiler verilerek, ilgili makine ėrenme teknikleri anlatılmıřtır. nc blmde ise tez arařtırması kapsamında gerekleřtirilen arařtırma hakkında bilgilere yer verilerek uygulama adımları ve sonuları gsterilmiř ve elde edilen bulgulara yer verilmiřtir.

BİRİNCİ BÖLÜM

ÖNERİ SİSTEMLERİ VE MODELLEMELERİNE YÖNELİK TEMEL BİLGİLER

Bir bilgi filtreleme tekniği olarak uzun bir süredir kullanılan öneri sistemleri; bilgi erişim sistemleri, makine öğrenme ve veri madenciliği yöntemlerinin çeşitlendirilmesi ile ortaya çıkarılan ve istatistiksel metotları temel alarak çeşitli modellemeler oluşturulmasıyla gerçekleştiren yöntemlerden oluşmaktadır. Veri kavramının oluşması, depolama alanlarının elverişliliğinin artması ve bilgilerin sayısal verilere dönüşümünün kolaylaşması, araştırmacıları veri kavramını yeniden yorumlamaya zorlamış ve dijital ortamların ham veri üreten sistemler haline getirmiştir. “World Wide Web” (kısaca WWW veya Web) kavramının gündelik yaşamın her alanına sızarak temel yaşam ihtiyaçlarının birçoğunu karşılayarak veya gündelik alışkanlıklara yardımcı olacak düzeye gelmesi, kullanıcıların teknolojiye adaptasyonunu hızlandırarak, güven duygularının hızlı bir şekilde olgunlaşmasını hızlandırmış ve veri oluşumuna gönüllü katkı vermesini sağlamıştır. Bunun yanında kullanıcılar bilerek veya farkında olmadan kendi kişisel verilerini paylaşmaları, ürün ve kullanıcı arasındaki bağın oluşturulması konusunda çıkarım ve öğrenme temelli modellemeler için eşsiz bir kaynak haline gelmiştir. Araştırmacılar, araştırmalarında “Büyük Veri (Big Data)” kavramın kullanmaya başlamaları şirketlerin en küçük bir kullanıcı verisini bile kayıt altında tutarak uzun yıllar saklamanın önemini anlaşılmamasını sağlamıştır. Karar vericiler, kullanıcı davranışlarının direkt olarak değerlendirmek yerine kullanıcının farkında olmadan bıraktığı izler arasında bağıntı kurmanın kullanıcı alışkanlıklarını tespiti açısından ciddi bir avantaj sağladığının farkına varmışlardır.

Öneri sistemleri, kullanıcıların bilgi arayış süreçlerinde aranan nesneler ile ilgili kullanıcıların ihtiyaç ve tercihleri hakkında öneriler oluşturmalarına yardımcı olan akıllı uygulamalardır (Resnick ve Varian, 1997: 56; Mahmood ve Ricci, 2009: 100). Kullanıcılar için sağlanan bu öneri mekanizması genel bir karar verme sürecinin bir parçası olabilirken, özelleştirilmiş bir sistemin yardımcı mekanizmaları olarak da kullanılabilirler. Bu sistemin ana değişkeni olan temel veri, çoğu kez ürün sıfatı ile betimlense de farklı uygulamalarda çeşitli tanımlarla da ortaya çıkabilmektedir.

Öneri Sistemleri, bireylerin belirli bir sistem üzerindeki kişisel deneyimlerinin ürün seçimindeki tercihlerine yansımada aracılık etmesinde yardımcı olur. Bu sistem içerisinde bireyin kendi iradesi doğrultusunda yapacağı tercihlerin yanı sıra kullanıcının ilgi alanına giren

ürünlerin de kullanıcıya sunumu sistemin esasını oluşturmaktadır. Öneri Sistemleri, kullanıcı ürün için aradığı özelliklere konusunda hâkim olmasına rağmen hangi ürünü tercih edeceği konusunda kararsız kalması durumunda da öneri sistemleri kullanıcıyı doğru seçenekleri yönlendirmek amacıyla oluşturulmuştur.

1.1. Öneri Sistemlerinin Temel Bileşenleri ve Kullanım Hedefleri

Öneri sistemlerinin temel özellikleri oldukça fazla çeşitlilik gösterse de literatürde dört ana başlık altında toplanmaktadır (Kaminskas ve Bridge, 2016: 2);

- **Bağıntılılık:** Öneri sistemlerinde en açık ve doğrudan kullanılan sistem özelliği olan bağıntılılık, kullanıcılara herhangi bir aracı metoda ihtiyaç duymadan öneri sistemi modellemeleri ile etkileşim kurabilmesini sağlamaktadır. Kullanıcı öneriyi dikkate değer bulup bulmadığının izlenmesi ile sistemin başarısı ölçülebilmektedir. İlgililik, öncelikli operasyonel hedef olarak tanımlansa da tek başına bir başarı kriteri olarak değerlendirilmemektedir.
- **Yenilik:** Kullanıcının tercih etme eğiliminde olduğu bir ürünün benzeri veya tamamlayıcı nitelikte olan başka bir ürün zincirinin, öneri sistemleri yoluyla kullanıcı karşısına getirilmesi sistem özelliklerinden biri olmalıdır. Kullanıcının ihtiyaç duyduğu ürün ile birlikte, öneri sistemleri tarafından yakın ilgisi olduğu tespit edilen diğer ürünlerin kullanıcıya teklif edilmesi, karar vericiler tarafından çoğunlukla arzulanan bir durum olarak görülmektedir. Ancak satış derinliğini etkilememesi açısından tekrarlanan önerilerinden kaçınılması gerekmektedir (Fleder ve Hosanagar, 2007: 198).
- **Keşfetme Yeteneği:** Şans eseri değerli bir şeyler keşfetme yeteneği anlamına gelen “serendipity” kavramı (Good vd., 1999; Herlocker, Jonathan vd., 2004: 43), öneri sistemlerinin birbirleri arasında fark yaratabildikleri özelliklerden bir tanesidir. Kullanıcının henüz farkında olmadığı ya da dolaylı yoldan kullanıcının ilgi alanına girebilecek olan bir ürünün bu kullanıcıya sunulması, öneri sistemlerinin verimliliği ve başarısı açısından yenilikçi bir fonksiyondur. Yenilik tanımından farklı olarak kullanıcının o ürün hakkında bir fikri olmaması veya yeni karşılaştığı bir kavram/işleyişe sahip olma özelliğini barındırmasına önem verilmektedir. Bu nedenle kullanıcının tercihleri, ürün özelliklerinin iyi tanımlanmış olması ve öneri sistemi için oluşturulan algoritmalar için bu bilgilerin iyi belirlenmiş olmasına göre farklılık gösterebilmektedir.
- **Öneri Derinliğinin Arttırılması:** Günümüzde oldukça farklı öneri sistemi algoritması kullanımı olsa da elde edilen sonuçlar açısından fark yaratmak için bahsedilen diğer

özelliklerin yanında, kullanıcı/ürün bilgileri doğrultusunda sunulabilecek ürün derinliği ve çeşitliliğinin sağlanması gerekmektedir (Vargas ve Castells, 2011: 109). Bu nedenle belirli sınırlar çerçevesinde karar vericiler tarafından oluşturulan ürün kategorizasyonunu tek başına kullanmak yerine kendi oluşturulacağı sınıflandırma veya gruplama yöntemlerini kullanmalıdır.

Tüm bu özelliklerin yanı sıra, özelleştirilmiş/kişiselleştirilmiş öneri sistemleri uygulama alanına özel tasarlanmış kurallar ve hedefler döngüsünü oluşturabilmelidir. Kullanıcı ilgisinin karar vericinin kurguladığı sistem üzerinde tutmak, tüm bu hedefler için çatı hedef olarak belirlenmelidir. Öneri sistemleri kullanıcı odağını merkez noktası olarak belirleyerek sınırları belirlenmiş bir projenin strateji ve yol haritasını kullanıcıyı tanımlayarak öneri merkezi sistemi çekirdeğini bu merkez çevresinde tutarak kurmalıdır.

Belirtilen özellikler çerçevesinde literatürde yaygın olarak betimlenen öneri sistemlerinin gerçekleştirebileceği görevlerden bazıları şu şekilde oluşturulmuştur (Ricci vd., 2011: 7);

- **İyi Ürünlerin Seçici Olarak Tespiti:** Kullanıcının ürüne olan ilgisini puanlama yoluyla ölçülmesi amacıyla, proje içerisinde bir ara yüz vasıtasıyla tanımlanmasıdır. Öneri sistemleri içinde yaygın kullanılır.
- **Tüm iyi ürünlerin tespiti:** Kullanıcıyı tatmin etmesi düşünülen tüm ürünlerin kullanıcıya herhangi bir filtre uygulanmaksızın sunulmasıdır. Ürün çeşitliliğinin nispeten az veya öneri sistemlerinin özelleştirilmiş veya kritik öneme sahip bir alanda kurgulandığı durumlarda kullanılmaktadır. Bu gibi durumlarda ayrımın kullanıcı tarafından hassas bir şekilde yapılması tercih edilerek, kullanıcıya ürünün ayırt edici özellikleri açık ve net olarak sunulmalıdır.
- **İçerik Çıkarımı:** Belirli bir içerikten kullanıcının uzun dönem tercihleri dikkate alınarak yapılan öneri odaklanmalarıdır. Örneğin bir makale öneri sistemi içerisinde kullanıcının okuduğu makaleler içerisinden metin-filtreleme yöntemi kullanımı ile doğrudan bağ içermeyen bağlantılar keşfedilerek yeni ve kullanıcının devamlı olarak takip etmediği ilgi alanları ile ilintili makaleler sunulmasıdır.
- **Etkisel Önerim:** Tek bir öneri seçeneği sunmak yerine belirli bir kategorik unsur kullanarak kullanıcıya aynı kategori içerisinde farklı seçim şansı tanıyan araçlardır.
- **Yığın Önerim:** Birden fazla ürün bir araya getirilerek yeni bir ürün halinde kullanıcıya alternatif çözüm olarak sunulan öneri paketleri bu kategoriye girmektedir. Tatil paketleri, hediye bulma siteleri ve birlikte işlevsellik kazanan elektronik donanım malzemeleri ve kozmetik ürünler için bu yola başvurulabilmektedir.

- **Sadece Gözetim (Just Browsing):** Kullanıcının herhangi bir ürün alma eğiliminde olmadığı bilinerek, kullanıcıya belirli bir katalog içerisindeki ürün gamından oluşturulan kataloglar vasıtasıyla gerçekleştirilir. Bu süreç çoğunlukla adaptif bağımlı medya teknikleri üzerinden oluşturulmaktadır (O'mahony vd., 2004: 215). Belirli bir kullanıcı tercihi dikkate alınarak, kullanıcının orta ve uzun vadede ürün tercih yönelimini belirlemesinde kullanılmaktadır. Market katalogları, tanıtım kitapçıkları bu kategori altında yer alabilmektedir.
- **Güvenilir Öneri Bulma:** Öneri sistemlerine karşı önyargılı olan ve bu sistemlerin güvenilirliği konusunda tereddüt yaşayan kullanıcılar için farklı bir yol izlenerek söz konusu grup için sistemin test edilebilirliğinin sağlanması amaçlanmaktadır. Kullanıcı ürünü tercih etmese bile bu öneri ile ilgili geri bildirim yapması beklenir.
- **Profil Artırımı:** Öneri sisteminin veri kaynağı olarak belirlenen kullanıcı profili için daha net ve amaca uygun veri elde edilmesine yönelik yapılan çalışmalardır. Kişiselleştirilmiş öneri sistemlerinin çekirdek verisinin oluşturulması için önemli bir aşamadır. Bu durumda öneri sistemi genel bir kullanıcı kitlesi yerine seçilmiş bir kullanıcı kitlesine yönelik çalıştırılmalıdır.
- **Kendini betimleme:** Bazı kullanıcıların öneri sistemi hakkında bilgi sahibi olmaması ya da öneri sistemlerini önemsememesi gibi durumlarda bu kullanıcı özelliklerine benzer grubun sisteme dâhil edilmesi için gerçekleştirilen çalışmalarda öneri sistemleri içinde yer alır. Kullanıcı özelliklerinin çeşitliliğinin artırılması için kendilerini sistem dışı olarak gören grupların da kendilerini tanımlaması sağlanmalıdır.
- **Diğerlerine Yardım:** Bazı kullanıcı grupları geri bildirim konusunda daha istekli ve katkıya açık bir tutum sergileyebilmektedir. Verdikleri yorumlar veya puanlamalarla sisteme direkt olarak veri sağlarlar.

Öneri sistemleri, kullanıcı veya ürün odaklı kişiselleştirilerek her kullanıcı veya kullanıcı grubu için farklı ürün seçenekleri sunabilmektedir. Bu kişiselleştirme işlevi e-ticaret vasıtasıyla kullanıcıya alışveriş imkânı sağlayan sistemlerle birlikte, kitap, müzik gibi kişisel zevklerin ön planda olduğu tavsiye sistemlerinde sıklıkla uygulanmaktadır.

1.2. Öneri Sistemlerinin Özellikleri

Öneri sistemleri, kendi içinde oldukça farklı disiplinleri barındırmakta ve bu sistemle de her durum için özelleştirilmiş algoritmaların uygulanabildiği modellerdir. Bu durum öneri sistemi geliştirmek isteyen uygulayıcılar ve araştırmacılar için çeşitlilik sağlasa da aynı zamanda karmaşık ve detaylı bir inceleme ile sonuca varabileceği bir sistem ortaya

çıkarmaktadır. Öncelikli olarak oluşturulması planlanan öneri sistemleri için önem arz eden kriterlerin tespiti ve belirlenen kriterlerin uygulanabileceği metot ve yöntemlerin ortaya çıkarılması önemli ve kritik bir süreçtir. Bu nedenle sistemin performansını etkileyebilecek her türlü etken düşünülerek, deneysel çalışmalara ağırlık verilerek doğru yöntemin bulunması öncelikli bir durum olmalıdır.

Öneri sistemlerinin genel yapısı oluşturulurken taşınması gereken özelliklerin bir kısmı aşağıda sıralanmıştır (Aamir ve Bhusry, 2015: 27);

Kullanıcı Tercihleri

Temel işleyişi kullanıcılar tarafından gerçekleştirilen işlemler üzerine kurgulanmış öneri sistemi mekanizmaları oluşturmak için doğrudan veya dolaylı olarak kullanıcı davranışlarının gözlemlenerek bu davranışların veri haline dönüştürülmesi şarttır. Ancak bir model oluşturulmadan önce kullanıcı faktörü ile ilgili dikkat edilmesi gereken bazı unsurlar mevcuttur. Sistem içerisinde değerlendirmeye dâhil edilecek kullanıcıların hepsinin eşit olarak kabul edilip edilmeyeceği durumu bunlardan bir tanesidir. Örneğin bir e-ticaret web sitesinde sadece tek ürün alan kullanıcı grubunun ile devamlı olarak alışveriş yapan kullanıcı grubunun model içerisinde eşit değerlendirilip değerlendirilmeyeceği belirlenmelidir. Bu gibi durumlarda, verilerin farklılaşması durumunda ağırlıklandırma yönteminin devreye girmesi gibi yeni seçenekler ortaya çıkabilmektedir.

Bazı durumlarda ise kullanıcıların sistemi kullanım sebebi ve tercihlerinde farklı varsayımlarla hareket etmesi, modelden beklentinin değişmesine yol açabilmektedir. Kullanıcıların dönemsel ihtiyaçları veya kişiye özel belli sebeplerle bir ürünü tercih etmesi ve tercihlerini bu şekilde oluşturduğunun tespit edilmesi modelin başarısı için önemli bir kıstas olarak kabul edilmektedir. Modeli oluşturmadan önce kullanıcı davranış standartlarının belirlenerek bunların dışında hareket eden durumların model dışında bırakılması değerlendirilmelidir.

Tahmin Tutarlılığı

Öneri sistemleri içerisinde oluşturulan modellerin yarattığı çıktılar ile sistemin performansının ölçümü için birden fazla metodun ortaya konması ve bunların yarattığı sonuçlar doğrultusunda geliştirmelerin yapılması beklenmektedir (Jahrer vd., 2010: 693). Öneri sistemlerinde kullanılan makine öğrenme teknikleri yöntemleri farklı veri türleri için hedefe yönelik çıktılar oluşturabilmelerine rağmen veri kalitesi ve kullanılan yöntemin tutarlılığı elde edilen sonuca doğrudan etkili olmaktadır. Bu nedenle öneri sistemlerinin temel yapısında önemli bir özellik olarak kabul edilebilmektedir (Aamir ve Bhusry, 2015: 27). Temel varsayım,

öneri modeli ile üretilen sonuçlar içerisinde tahmin doğruluğunun geri bildirimlerle tutarlılığının yükseltilmesi üzerine kurgulanmalıdır. Böylece kendi kendini besleyen bir model ile başarılı bir sistem elde edilebilir.

Kök Ortalama Kare Hatası (Root Mean Squared Error) yöntemi bir model tarafından öngörülen değerlerin bir örneklem içerisindeki mevcut değerler arasındaki farkın ölçümü ile elde edilen sonuçları göstermektedir. Öneri sistemlerinde tahmin doğruluğu için sıklıkla kullanılan yöntemle test veri seti (T) içerisinde (u,i) kullanıcı ürün çifti için oluşturulan $\hat{r}_{ui}$ modeli ile r_{ui} çıktıları ile kıyaslayarak tahmin değeri oluşturur. r_{ui} genellikle önceden belirlenmiş bir çevrimdışı işlemle ya da eğitim veri seti üzerinde oluşturulmuş çıktılarla ortaya konulmaktadır. Gerçekleşen ve tahmin edilen değerleri arasındaki kök ortalama kare hata değeri şu şekilde hesaplanır (Melville ve Sindhvani, 2010: 841);

$$\sqrt{\frac{\sum_{(u,i) \in T} (\hat{r}_{ui} - r_{ui})^2}{T}}$$

Kök ortalama kare hatasına alternatif olarak kullanılman başka bir metot ise ortalama mutlak hata yöntemidir. Ortalama mutlak hata ise şu şekilde hesaplanır;

$$\sqrt{\frac{\sum_{(u,i) \in T} |\hat{r}_{ui} - r_{ui}|}{T}}$$

İki yöntem karşılaştırıldığında kök ortalama kare hatası ile daha büyük hata değerlerinin tespiti daha net elde edilebilmekte ve başarı oranını daha iyi kestirebilmektedir. Daha net sonuçlar elde edebilmek için bu yöntemler için normalizasyon işlemlerine başvurulabilir. Böylece ölçeklendirilmiş bir sonuç elde edileceği için bir sıralama ölçütü ile değerlendirme imkânı ortaya çıkarılabilmektedir.

Eğer veri setleri içerisindeki veri dağılımında problemler kestirilirse ve dengelenmemiş bir dağılımın farkına varılırsa bu durumda kök ortalama kare hatası ve ortalama mutlak hata değerleri genel bir ölçümü göstermek yerine, belli uç noktalardan etkilenmiş bir hesaplama ile hatalı bir sonuç elde edimine neden olabilir. Bu durumda her nesne için ayrı bir hesaplama yaparak ortalama değerler üzerinden bir değerlendirme yapmak da doğru bir sonuç verebilmektedir.

Tahmin Değerleri

Öneri sistemlerinin kalitesi değerlendirilirken çoğu yaklaşım sadece tahmin tutarlılığını göz önüne almaktadır (Ge vd., 2010: 257). Tutarlılıkla birlikte kullanım tahmin ölçümlerinin gerçekleştirilmesi için de farklı yöntemler ortaya çıkarılmıştır. Modelin eğitim veri seti üzerinde çalıştırılması ve çevrimdışı fazda test veri seti üzerinde çıktı sonuçlarının

karşılaştırılması veya çevrimiçi fazda anlık olarak kullanıcılara oluşturulan öneri değerlendirilmelerinin sunulması; sistemin olağan akışı içinde beklenen bir adımdır. Bu değerlendirmelerin doğru ve tutarlı olup olmadığının ölçümü ise performans skorlarının tespiti ile mümkündür (Schröder vd., 2011: 53).

Kullanıcıya önerilen ürünün kullanıcı tarafından kabul edilmesi veya edilmemesi, kullanıcıya önerilmemiş bir ürünün kullanıcı tarafından tercih edilmesi performans skorunun oluşturulmasında kullanılan kriterlerdir.

Tablo 1.1 Hata Matrisi Standart Tablosu

	Gerçek Pozitif	Gerçek Negatif
Pozitif Değerlendirilen	DP	YP
Negatif Değerlendirilen	YN	DN

Bu tür skorlar oluşturulurken çıktıların veri seti üzerinde yapılan karşılaştırma ile bir değer elde edilmesi amaçlanmaktadır. Hata Matrisi (Confession matrix) olarak adlandırılan tablodaki yapı ile veri seti üzerindeki model çıktı değerlendirmeleri için bir sonuç kümesi oluşturulabilmektedir. Hata matrisinin genel anlamda üç kullanım yolu vardır. Sınıflandırma içerisinde genel tutarlılığın tespit edilmesi, kategorileştirilmiş veriler içindeki özel hataların tespit edilmesi ve kullanılan yorumlama tekniklerinin kategoriler hakkında fazla ya da az katkı sağlayıp sağlamadığının tespiti olarak belirlenmiştir (Hay, 1988: 1395). Tablo 1.1’de temel hali gösterilen hata matrisinde karar verici tarafından pozitif değerlendirilen ve gerçekleşen değeri pozitif olan sonuçlar DP (Doğru Pozitif), karar verici tarafından yine pozitif değerlendirilen ancak gerçekleşen değer negatif olduğu sonuçlar YP (Yanlış Pozitif), karar verici tarafından negatif değerlendirilen ancak gerçekleşen değer pozitif olduğu durumlar YN (Yanlış Negatif) olarak gösterilmiştir. Bu gözlemler doğrultusunda bir kriter belirlemek için hassasiyet ve kesinlik değerleri bilgi edinim sistemlerinde sıklıkla başvuru bir değerlendirme yöntemidir (Raghavan vd., 1989: 206). Bu değerlerin hesaplamaları aşağıda gösterilmiştir.

$$Hassasiyet = \frac{DP}{DP + YP}$$

$$Kesinlik = \frac{DP}{DP + YN}$$

Kesinlik değerlerini geçerli sonuçların tüm sonuçlar içindeki oranı, hassasiyeti ise tahmin edilenler içerisindeki geçerli sonuçları olarak değerlendirilir (Buckland ve Gey, 1994). Sonuçlar incelendiği zaman değerler arasında bir ters ilişki olabileceği anlaşılmaktadır. Model sonucunda fazla sayıda bir tahmin çıktısı kesinlik değerini arttırırken, hassasiyet değerinde de azalmaya neden olacaktır. Eğer belirlenmiş bir sayı üzerinden çıktı elde edilmesi planlanmışsa hassasiyet değerinin öncelikli değerlendirmeye alınması kaçınılmazdır.

Farklı durumlar için modellerin sınırlandırılmış bir liste ile tahmin çıktısı oluşturulmadığı durumlarda hassasiyet ve kesinlik değerleri arasında grafiksel olarak bir eğri ile gösterim sağlanabilmektedir. Hassasiyet-kesinlik eğrisi olarak gözlemlenebilecek bu eğri daha da geliştirilerek literatürde “Receiver Operating Characteristic (ROC)” ismiyle anılmaktadır (Beck vd., 1986). Kesinlik-hassasiyet eğrileri sadece tercih edilen ürünlerin orantısal durumunu gösterirken, ROC eğrisi ile önerilen ancak tercih edilmeyen ürünler hakkında da fikir edinilebilmektedir.

Ölçümlerin Sıralanması

Öneri sistemlerinin amacı doğrultusunda kullanıcılar için oluşturulan çıktılar, modelin başarısının ölçümü için ana etken olarak görülür. E-ticarette veya dijital tabanlı bir sistemde çıktıların gösteriminde yatay veya dikey bir biçimde bir sıralama kullanılması mümkündür.

Modellemelerin belirli bir sıralama ile öneri sistemi çıktısı oluşturabilmesinin yanı sıra herhangi bir kriteri dikkate almadan rastgele sonuçların listelendiği yöntemler de mevcuttur. Bu durum kullanıcılar için oluşturulan öneri sisteminin herhangi bir ağırlıklandırma yapmadan eşit şekilde değerlendirdiği öneriler için karar vericilerin ek bir işleme ihtiyaç duymasına neden olur. Bu durumda da genelde kullanıcı tercihleri düşünülerek yeni yapılar ortaya çıkarılabilmektedir. Sıralama uygunluğunun ölçülmesi ve tespiti için genelde iki önermeden yararlanılır. Her kullanıcı için ortak bir yapı üzerinden standardize edilmiş bir sıralama kuralı belirlenebilir ya da genel bir sistemsal kuralı herhangi bir kullanıcı ayırt etmeksizin tümüne uygulamanın yolları araştırılabilir. Her iki durumda da modelin yeni algoritmalarla adaptasyon sürecine ihtiyaç duyması kaçınılmazdır (Kelly, 2009: 108).

1.3. Öneri Sistemlerinin Kapsamı

Öneri sistemlerinde farklı modelleme teknikleri ve veri yapıları kullanılsa da temel olarak çoğu zaman belirli bir kısıtlama altında algoritmaların çalıştırılması ile oluşturulmaktadır. Bu durum spesifik bir iş için programlanmış bir yazılımla çalışılmasına ve karar vericilerin genel bir yapının verimli bir çıktı sağlamayacağına olan inancına paralel olarak yürütülmektedir. Verinin çoğalması ile birlikte öneri tutarlılığının artacağına olan inanç model

oluştururken belirli bir standardın tutturulması ile gerçeğe dönüşebilmektedir. Bu durum iki aşamada incelenebilir

Ürün Kapsama Alanı

Öneri sistemlerinde ürün kapsama alanı olarak modele dâhil edilen ve öneri sistemlerinin çıktı olarak sunabileceği nesneler, sistemi besleyen temel veri olarak kabul edilir. Katalog olarak da adlandırılabilir bu alan tüm havuz içerisindeki öneri sistemine dâhil edilebilecek nesnelerin yüzdesi ile gösterilmektedir. Bu hesaplama oluşturulan veri setlerinin amacı doğrultusunda olabileceği gibi, çevrimiçi, çevrimdışı faz hesaplamaları için de farklılık gösterebilir. Bazı çalışmalarda bu kapsam sadece nesnelerin tekil sayımı ile değil ağırlıklandırma yöntemi ile de analizlere dâhil edildiği gözlemlenebilmektedir. Satış Derinliği, Shannon Entropisi ve Gini Katsayısı gibi yöntemlerle daha özelleştirilmiş sonuçlar elde etmek de mümkündür (Brusilovski vd., 2007: 679).

Kullanıcı Kapsama Alanı

Ürün temelli kapsamanın değerlendirmeye alınabileceği gibi kullanıcı açısından da öneri sisteminin başarı ölçütü olarak kapsam derinliği analizleri gerçekleştirilebilir (Aamir ve Bhusry, 2015: 27). Öneri sistemleri çoğu zaman modellerin oluşturulması ve çıktıların alması esnasında yeterli ürün yelpazesine sahip olmasına rağmen kullanıcıların yetersizliği nedeniyle kullanıcılara sunulabilecek tutarlı bir öneri çıktısı sunulmayabilmektedir. Ürün kapsama alanının genişlemesi tahmin tutarlılığı açısından avantaj sağlamasına olanak sağlasa da kullanıcı çeşitliliği açısından aynı durumun söz konusu olmadığı aşîkârdır. Kullanıcı çeşitliliğinin artması sisteme dâhil edilen ürün ile ilgili bilgi ile paralel olması durumunda fayda sağlaması mümkün iken, kullanıcı ve ürün arasındaki ilişkiyi ortaya koyan verilerin kullanıcılar arasında homojen dağılmaması öneri sisteminin sağlıklı çalışmasına engel olacaktır.

Soğuk Başlangıç Problemi

Öneri sistemlerinde oluşturulan metotların neredeyse genelinde kronik olarak karşılaşılabilecek sorunların başında soğuk başlangıç problemi gelmektedir. Bu problem ürün kapsam alanı içerisinde seyreklik problemi ile ilgili olduğu kabul edilir (Lika vd., 2014: 2066). Sistemin kapsadığı veri seti içerisinde yer alan yeni kullanıcı veya ürünlerin değerlendirmeye alınmasında bu nesneler ilgili veri eksikliğinden kaynaklı sonuç üretememe ya da hatalı bir çıktı oluşturma problemleriyle karşılaşılabilmektedir. Model içerisinde bu tür sorunlar öneri tutarlılığını ve modelin başarısını ölçmek için kullanılan skorlarda yanıltıcı senaryoların oluşmasına neden olabilmektedir.

Bu gibi durumların önüne geçebilmek için modelde yer alan yeni nesneler ile var olan nesneler arasındaki boşluklar için köprü mahiyetinde alternatif tekniklerin kullanımı mümkündür (Schein vd., 2002: 257; Lam vd., 2008). Yeni eklenen bir ürün beş kullanıcı tarafından değerlendirilmedikten sonra analizlere dâhil edilmemesi buna örnektir. Bu basit örnek ile yola çıkılarak daha gelişmiş ve ürün/kullanıcı için “soğukluk” tanımının yapılması ve genel bir kuralın tüm sistem içerisinde uygulanması, sorunun önüne geçmek için uygulanması gereken metotlardandır.

1.4. Öneri Sistemlerinde Kullanılan Veri Yapıları

Ürün-kullanıcı arasındaki ilişkiyi açıklayabilecek her türlü verinin kullanılabildiği Öneri Sistemlerinde, en basitinden en karmaşığına kadar tüm veri çeşidi üzerine uygulamalar gerçekleştirilmektedir. En temelde ürün özellikleri, kullanıcının ürün hakkında deneyim ve bilgi paylaşımlarından yola çıkılarak kullanıcının ilgili kişisel bilgileri, sistem üzerinden çıkarımı yapılan işlem kayıtları (sayfa görüntüleme sayısı ve süreleri, yaptığı arama kayıtları, vb.), sosyal medya hesapları gibi ikincil veri kaynakları da sıklıkla modellere dâhil edilmektedir.

Tüm bu veri tipleri genellikle üç taraf üzerinden elde edilebilmektedir. Öneri sistemleri çift taraflı bir etkileşim üzerine kurgulanmaları nedeniyle bu durum sisteme konu olan nesne özellikleri ile bu nesneler arasındaki etkileşim sonucu ortaya çıkan bağlantılı veriler olarak kabul edilir. E-ticaret özelinde değerlendirildiğinde ürün, kullanıcı ve kullanıcı ve ürün arasındaki bağlantıyı sağlayan özelleştirilmiş veriler olarak kategorize edilmektedir (Ricci vd., 2015: 8). Söz konusu etkileşim genellikle kurulan model üzerinden kullanıcının isteyerek veya bilerek bıraktığı izler olabileceği gibi, modelin kurduğu dolaylı bağlantılarla da sağlanabilmektedir.

Sistem içerisinde ürün özellikleri, ürün için üretici tarafından belirlenmiş temel özelliklerin yanında kullanıcı tarafından tek taraflı oluşturulan sübjektif düşünce ve görüşlerden oluşur. Ürün için değer kavramı, kullanıcının kişisel olarak o ürün hakkında elde ettiği bilgiler veya kendisi veya başkasından kaynaklanan tecrübelerle dayalı ürün verimi hakkındaki düşünceler olarak kabul edilmektedir. Kullanıcı bu değeri elde etmek için piyasa değerinin yanında o ürün için harcadığı alternatif maliyetleri (cognitive cost) de düşünerek karar verir. Bu karar nihayetinde kullanıcı tercihinin pozitif veya negatif olarak şekillenmesi beklenmektedir. Örneğin bir ücretsiz metin tabanlı içerik hizmeti sunan web/mobil projesinde kullanılması planlanan öneri sistemi de kullanıcılarının o içeriğe ulaşması için katlandığı zaman

ve kaynak gibi parametrelerin değerlendirilmesinin yanında, kullanıcının aradığı bilgi ile önsizde bulunarak öneri sonuçlarını iyileştirmesi gerekmektedir.

Değer algısı biçiminde düşük karmaşıklığa sahip ürünler haber, blog/mikroblog yazıları, kitaplar, filmler/diziler/görseller vb. içerikler olarak tanımlanabilirken, elektronik donanımlar, bilgisayarlar, cep telefonları gibi daha yüksek bir karmaşıklığa sahip ürünler için farklı bir algı olduğu hesaba katılmalıdır. Sigortacılık, finansal araçlar, seyahatler, iş fırsatları ise öneri sistemleri için en yüksek karmaşıklığa sahip veri kaynağı olarak nitelendirilmektedir (Montaner vd., 2003: 285).

Bilgi edinim sistemlerinde kullanıcı modelleme ihtiyacı, insan bilgisayar iş birliğinin ortaya çıkarılması amacıyla ortaya çıkmıştır. Bu işbirliği ortak bir amacı gerçekleştirmek için iki veya daha fazla temsilcinin birlikte çalışma içgüdü'sü' olarak tanımlanmaktadır (Fischer, 2001: 66). Öneri sistemlerinde kullanıcı kavramı, sistemin yapısı ile paralel olarak belirli kriterler doğrultusunda gruplanmış hedef kitle bireylerini tanımlamaktadır. Önerilerin ve insan-bilgisayar etkileşiminin veriye dönüştürülmesi için öneri sisteminin kendi yapısı doğrultusunda kullanıcı bilgilerinin keşfi sağlanmalıdır. Bu işlem kullanıcının doğrudan paylaştığı bilgiler olabileceği gibi, sistem tarafından devamlı olarak gerçekleştirilen izlemelerle sistem içerisinde kullanıcı davranışlarının kayıt altına alınması ile de ortaya çıkarılabilmektedir. Tüm bu veriler kullanıcı modellerinin olgunlaşmasını sağlamaktadır. Öneri sistemleri, belirli bir sistem içerisinde kişiselleştirme yapılabilmesi için kullanıcı modelleri oluşturulmasında ve devamlı olarak güncellenmesinde kullanılan bir araçtır. Kişiselleştirme kullanıcının tercih ve davranışları hakkındaki bilgiler göz önüne alınarak kullanıcıya içerik ve hizmet sağlama yeteneğidir (Adomavicius ve Tuzhilin, 2005: 83). Hangi öneri sistemi modeli kullanılırsa kullansın, kullanıcı tercih ve ihtiyaçları ile oluşturulan kullanıcı modelleri veri setinin merkezinde yer alması gerekmektedir.

İşlem kayıtları, genellikle kullanıcı ve öneri sistemleri arasında doğan etkileşim sonucu elde edilen ham veriler için kullanılan bir terimdir. Bu kayıtlar yazılım dilinde “log” (işlem kayıtları) kelimesi ile tanımlanmakta, kullanıcının hareketlerinin herhangi bir ön işlemden geçirilmeden kayıt altına alındığı ve işlemin anlık durumunu gösterir bilgi kümesidir (Peters, 1993: 41). Bu tür kayıtlarda genellikle bir ön çalışma yapılarak öneri sisteminde kullanılabilecek bilgilerin çıkarımı ve temizlenmesi işlemleri gerçekleştirilmelidir. Diğer taraftan sistem tarafından sadece öneri sistemi için oluşturulan kayıtlar doğrudan öneri modellerine dâhil edilebilir (Schafer vd., 1999: 162).

Kullanıcı ve ürün arasındaki etkileşim kaynaklı veri elde etme aracı olan ve yaygın olarak da kullanılan puanlama sistemleri, işlem kayıtları içinde sıklıkla kullanılmaktadır.

Puanlama mekanizması, bir ürünü daha önce satın alarak tecrübe etmiş kişilerin o ürün hakkındaki düşüncelerinin nicelikle ölçülmesi yoluyla elde edilmektedir.

Günümüzde metin tabanlı içerik yönetiminde kullanılmaya başlanarak, kullanıcıya oluşturulan içeriği kategorize etme olanağı tanıyan etiketleme yöntemi, işlem kaydı verisi olarak değerlendirilen yapılardan bir başkasıdır. Kullanıcının ürün ile ilgili fikirlerini bir veya birden fazla kelimelerle belirlediği etiketleme sisteminde, ürünün temel ve başlıca özellikleri sübjektif açıdan kullanıcı tarafından tanımlanabilmektedir. Örneğin bir film öneri sistemi için kullanıcı filmleri “sanatsal”, “korku”, “çok uzun” kelimeleriyle etiketleyerek film hakkındaki görüşlerini belirtebilir.

Kullanıcı ve ürün ile ilgili veri kaynaklarının çeşitlilik göstermesi ile birlikte bu verilerin birçoğu kullanıcı veya ürün kategorisi altında keskin çizgilerle ayrılamamaktadır. Sistem içerisinde bu iki kategorideki veriler sıklıkla etkileşim içerisinde olması nedeniyle birbirleri ile bir bağ içerisinde olduğu değerlendirilmelidir.

1.5. Öneri Sistemleri Uygulama Alanları

Öneri sistemleri kişiselleştirilmiş haber portallarından turist yönlendirme servislerine, borsa/bahis tahmin platformlarından sosyal medyaya kadar; kullanıcı ürün etkileşiminin olduğu çoğu ara yüzde karşılaşılan bir yöntemdir. 1990’lı yılların ortalarında bağımsız bir araştırma alanı olarak ortaya çıkmaya başlayan öneri sistemi modelleri, diğer klasik bilgi sistem araçları ve tekniklerinin dışında ayrı bir disiplin olarak kabul edilmektedir (Hill vd., 1995: 201). Teknolojinin hızlı bir şekilde ilerlemesi ticaretin dijital platformlar üzerinde kurulmasına olanak sağlamış, bu yeni düzende yeni pazarlama stratejilerine ortaya çıkmıştır. Netflix, Amazon, Youtube, IMDB ve benzeri dijital içerik üreten ve dağıtan web siteleri ürün tanıtım ve satış stratejilerini dijital algı üzerine şekillendirmeye odaklanan ilk internet tabanlı şirketlerin başında gelmektedirler.

İlk çalışmalardan biri olan GroupLens Öneri sistemi, o yıllarda yaygın olarak kullanılan bilgi paylaşım platformlarından biri olan Usenet haber ağı üzerine yapılan çalışmalar sonucu kurulmuştur (Resnick vd., 1994: 183). Kullanıcıların Usenet üzerinden haberlere verdiği puanları toplayarak kullanıcının okumayı planladığı haberin ilgisini çekip çekmeyeceği konusunda tahmin modelleri oluşturulmaya çalışılmıştır. “The Better Bit Bureaus” adı verilen bir puanlama sistemi sayesinde oluşturulan kullanıcı puan matrisi üzerinden bir korelasyon katsayısı hesaplanarak benzer ilgiler dikkate alınarak modeller oluşturulmuştur. İşbirlikçi filtreleme modellerinin temellerinin atıldığı çalışma daha sonra “Movielens” ve “Booklens” adlarıyla veri setleri üzerinden geliştirmeye devam edilmiştir. Bu veri setleri genişletilerek altı

gruba ayrılarak halka açık şekilde hâlihazırda kullanıma açık olup, temel araştırmalar için kaynak niteliği taşımaktadır (Resnick ve Varian, 1997: 56).

1990'lı yılların sonunda Amazon şirketinin perakende satış sistemlerinde öneri sistemlerini kullanmaya başlaması ile e-ticaret yapılarında yeni bir öneri modeline ihtiyaç duyulmuştur. Amazon ürün temelli işbirlikçi filtreleme modelini 1998 yılında kullanmaya başlamış ve daha önce görülmemiş şekilde milyonlarca müşteri ve ürünleri için öneri çıktıları oluşturmaya başlamıştır (Smith ve Linden, 2017: 12). Öneri sistemi yöntemlerinden biri olan “İşbirlikçi öneri sistemi” ile sadece şirketin kendi satışlarını arttırmaya yönelik bir çalışmanın yolu açılmamış, aynı zamanda diğer e-ticaret sitelerine de yol gösterici bir görev üstlenilmiştir. Bu yöntem, standart kâr amacı güden şirketlerin temel motivasyon kaynağı olan satış hacminin nasıl arttırabileceği konusunda yeni bir anlayışı ortaya çıkarmıştır. Araştırma ekibi bu veri setlerini genişleterek akademik çalışmalar için de kaynak oluşturabilecek halka açık veri setleri oluşturulmuştur.

Amazon bu sistemi kurgularken daha önceki birikimlerinden faydalanmıştır. Oluşturduğu öneri sistemi aşağıda belirtilen temel çalışmalar üzerine kurgulanmıştır (Schafer vd., 1999: 159).

Bu ürünü satın alan müşterilerin tercihi: Günümüzde klasikleşmiş çoğu e-ticaret sitesinin, satın alma süreci ile bütünleşmiş şekilde kullanmakta olduğu “Bunu alan bunu da aldı”, “Benzer kullanıcıların tercih ettikleri” şeklinde başlıklar altında sunulan öneri sistemleri, işbirlikçi filtreleme tekniğini kullanarak kullanıcılara, kendileri gibi benzer alışveriş davranışları olan kullanıcıların daha önce tercih ettiği ürünleri önerdiği bir sistemdir. Amazon o yıllarda odaklanmış olduğu kitap satışları için öncelikli olarak kitap adı daha sonra da yazarları kategorize ederek ürün tercihlerini yansıtmaya çalışmıştır.

“The Eyes” Listeleri: Amazon, kullanıcı tercihleri doğrultusunda oluşturduğu e-posta listelerine “The Eyes” (Gözler) adını vermiştir. Kullanıcı kitap özellikleri ile ilgili (kitap adı, yazar adı, konu vb.) belirli aralıklarla sistem tarafından e-posta aracılığıyla bilgilendirilmektedir. Kullanıcıya, geliştirdiği ikili veri temelli algoritmalar sayesinde uyarı sorguları oluşturmaya olanak sağlanmakta, bu sorgular neticesinde tercihlerine yönelik yeni ürün tavsiyesinin yapılmaktadır. E-posta listelerinin sağladığı yardım kullanıcılar tarafından puanlanmaktadır. Bu puanlama sayesinde standart bir e-posta listelemesinden daha ikna edici bir yapıya sahip olduğu ve geçerli bilgi erişimi için sezgisel üstünlük sağladığı gözlemlenmiştir (Zhang, J. vd., 2010: 1341).

Amazon.Com Editör Seçimleri: Editör seçimleri, Amazon tarafından oluşturulan uzman bir ekip tarafınca üretilen ve “The Eye” e-posta listelerinin özelleştirilmiş hali olarak

değerlendirilebilecek sistematik kullanıcı bildirimleridir. Kullanıcının belirli kategori ve/veya türlerde yaptığı direkt tercih istemleri doğrultusunda Amazon editörlerince manuel olarak hazırlanan kitap listeleri kullanıcıya ulaştırılması ile elde edilen bir yönlendirmedir (Schafer vd., 2001: 123).

Kitap Eşleştiricisi: Kitap eşleştiricisi sistemiyle kullanıcının satın almış olduğu kitap hakkında direkt olarak puanlama yapması istenilmektedir. 1 ile 5 arasındaki bir ölçekte kullanıcının kitap hakkındaki beğenisini gösteren bu puan ile yine “Bu Ürünü satın alan müşterilerin tercihi” özelliğindeki gibi kullanıcının puanlamış olduğu kitaplara benzer puan veren kullanıcıların tercihleri dikkate alınarak, kullanıcıya yeni seçkiler sunulmaktadır.

Kullanıcı Yorumları: Ürünü alarak deneyimlerini paylaşan kullanıcı yorumları dikkate alınarak metin odaklı bir öneri sistemi oluşturulmasıdır. Kullanıcılar ürün hakkındaki deneyimlerini olumlu veya olumsuz etiketiyle siteye yazmaları teşvik edilerek günlük hayatta ürünün sağladığı pratik özelliklerin vurgulanmasına ya da kullanıcıların özel durumlarda karşılaştığı olumsuzlukları sipariş vermeyi düşünen kullanıcılarla paylaşımının sağlanmasına olanak sağlamaktadır.

Bu özelliklerin kullanıma açılması ve web sitesinde kullanıcı ile daha interaktif bir ortamın oluşturulması öneri sistemlerinin çeşitlendirilmesi konusunda mihenk taşlardan biri olmuştur. Oluşturulan sistemle elde edilen başarı e-ticaret şirketleri için yeni bir pazarlama anlayışı oluşturmaya ışık tutmuştur.

Öneri sistemlerinin en önemli örneklerinden biri literatürde “Netflix Prize” olarak geçen ve bir milyon dolarlık ödülle o tarih itibarıyla çevrimiçi DVD ve video kiralama hizmeti veren Netflix firmasının açmış olduğu öneri sistem yarışması gösterilebilir (Hallinan ve Striphas, 2016: 117). Bu yarışmayla Netflix’in öneri sistemleri iyileştirilerek tahmin oranını %10,06 oranında arttırmayı başarmıştır (Bennett ve Lanning, 2007: 35). Netflix 1 Ekim 2006 tarihinde başlayan yarışma için 17,770 film, 480,000 kullanıcı ve yüz milyondan fazla puanlama verilerini içeren bir veri seti ile şirketin hâlihazırda sunmakta olduğu öneri sisteminin doğruluğunu %10 arttırma şartı ile yaklaşık 3 yıl boyunca devam ettirilmiştir (Bell ve Koren, 2007). Bu yarışma ile veri madenciliği uygulamalarının e-ticaret sektöründeki etkisi, kalıcı bir şekilde yerleşmiştir.

Belirtilen uygulama alanları öneri sistemlerinin gerçek zamanlı kullanımı için temel oluşturmuş ve bu uygulamalar hızlı bir şekilde benzer hizmet sağlayan e-ticaret sistemlerinde yaygınlaşmıştır. Öneri sistemleri işletmelere ellerindeki veriler dikkate alınarak sistemlerinin uyarlama pratiği oluşturabilmektedir (Felfernig vd., 2007: 20). Birçok sektörde farklı uygulamalarla karşılaşılmaktadır. Özellikle internet üzerinden hizmet veren şirketler için e-

ticarete emlak sektöründe geleceğe yönelik fiyat tahminleri (Schneider, 2018: 42), doğrudan pazarlama tekniklerinde verimliliği arttırmak adına doğru müşteri seçimi (Omondi ve Mbugua, 2017: 1) otel rezervasyonlarında kullanıcılar için yönlendirme servisinin açılması (Xiang ve Nan, 2016: 404) farklı uygulama alanlarına örnek olarak gösterilebilmektedir.

1.6. Öneri Sistemlerinde Temel Modeller

Kullanıcının tanımlanmış olduğu sisteme çekirdek bir modelleme atayabilmek için modellemenin en azından tanımlama, karşılaştırma ve karar verme (kestirim gücü) yeteneklerinin oluşturulması gerekmektedir. Kullanıcı ve ürün nesnelerinin karakteristik özellikleri ve aralarındaki etkileşimden doğan veriler vasıtasıyla tanımlama işlemi gerçekleştirilirken, bu veriler vasıtasıyla kullanılan yöntemler ile kullanıcı-kullanıcı, ürün-ürün veya kullanıcı-ürün arasındaki kıyaslama öneri sistemleri için çıktı oluşturmaktadır. Karar verme aşaması, modelin tüm aşamalarını birleştirerek ve sonuca ulaştırılarak modelin başarısının ölçülebileceği yer olarak kabul edilmelidir (Adomavicius ve Tuzhilin, 2005: 736).

Öneri sistemleri için modellemeler oluşturulurken veri türleri ön plana çıkarılmıştır. Kullanıcı-ürün ile ilişkili olan kullanıcı puanlamaları ve satın alma davranışları ile kullanıcı ve ürünün özellikleri ile metinsel profil bilgileri veya doğrudan olmayan farklı kullanıcı-ürün davranışları ana ayrışmayı sağlamaktadır (Aggarwal, 2016: 8).

Belirlenen veri türünün kullanımına göre öneri sistemlerinde farklı modeller kullanılabilmektedir. Kullanıcı ürün ilişkisinden faydalananarak “İşbirlikçi Filtreleme Öneri Modeli” veya kullanıcı ve ürünlerin karakteristik özelliklerinden yola çıkarak “İçerik Tabanlı Öneri Modeli” seçilebilmektedir. Literatürde sıklıkla kullanımı gözlemlenen bu iki modelleme türünün yanı sıra “Bilgi tabanlı Öneri Modeli” de bu sınıflandırmada yer alabilmektedir.

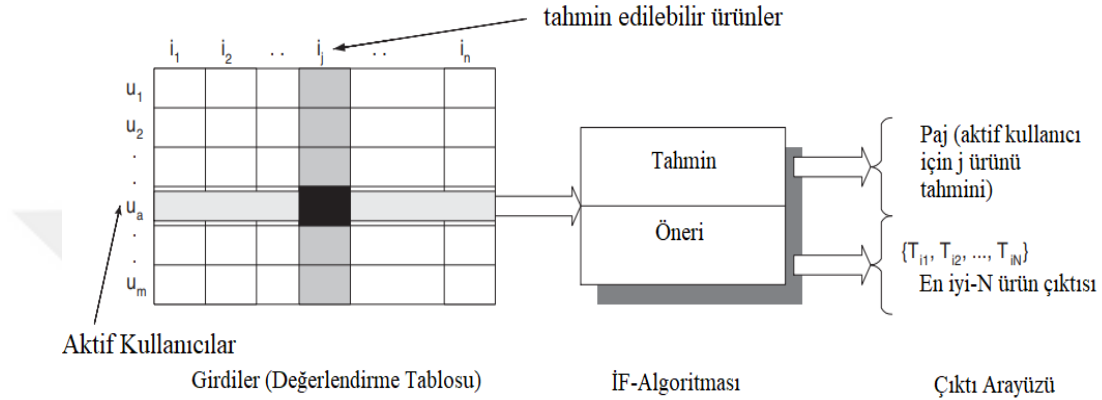
Farklı öneri sistemi modelleri için Burke (2007: 379) tarafından yapılan sınıflandırma sistemi kabul görmüş bir yaklaşım olarak literatürde yerini almıştır. Kullanıcı ve ürün verilerinin kullanımı ve değerlendirmeye alınış şekillerine göre incelenmektedir.

İşbirlikçi Filtreleme Modelleri

Öneri Sistemleri içerisinde temel ve ilk uygulamalardan biri olan işbirlikçi filtreleme modeli, kullanıcıya aynı sistem içerisinde kendisiyle benzer özelliklere sahip diğer kullanıcıların tercih ettiği veya sevdiği ürünleri kullanıcıya tavsiye etmesi olarak özetlenebilmektedir (Breese vd., 1998: 45). Birden fazla kullanıcı arasındaki tercih veya zevklerdeki benzerlik, kullanıcıların sistem içerisindeki ürünlere yönelik benzer davranışları dikkate alınarak ortaya koyulmaktadır. “İnsandan İnsana Korelasyon” olarak da tanımlanan

işbirlikçi modeller öneri sistemleri içerisinde en fazla kullanılan tekniklerin başında gelmektedir (Schafer vd., 2001: 141).

İşbirlikçi filtreleme modelleri, kullanıcı ürün etkileşimi sonucu ortaya çıkan ve genellikle ürün ile ilgili kullanıcı değerlendirmesi üzerinden tanımlanan veri kaynaklarına ihtiyaç duymaktadır. Kullanıcılar tarafından üretilen bu veriler model tarafından değerlendirme (puanlama) matrislerine dönüştürülmektedir.



Şekil 1.1.1 İşbirlikçi Filtreleme Çalışma Modeli

Kaynak: (Sarwar, B. vd., 2001)

Bu tür filtreleme modelleri ile kullanıcı tanımlarına veya ürün hakkında detaylı bir bilgi kümesine sahip olunmadan kullanıcı değerlendirmeleri ve bağıntılı veriler kullanılarak kullanıcı odaklı öneri seçenekleri üretilmesini sağlar. Bu filtreleme yöntemi ile kullanıcılar ve ürün arasındaki sistematik ilişki kümelenmeleri ortaya çıkarılarak ürünün karakteristik bilgilerine ihtiyaç duyulmadan ve dolayısıyla ürün kimliği doğrudan kullanılmadan öneri seçenekleri ortaya koyulabilmektedir (Hu vd., 2008: 265).

İşbirlikçi filtreleme modelinde diğer modellerde olduğu gibi kullanıcı, ürün veya her ikisinden elde edilen verilerle oluşturulan matris yapı Şekil 1.1’de gösterildiği üzere girdi olarak değerlendirilmekte ve bu girdilerin tamamı veya bir kısmı ile model oluşturulmaktadır. Girdi verilerin bir kısmının bölünerek modellenmenin yapılması modelin geçerliliğinin ve tutarlılığının test edilmesi amacıyla gerçekleştirilmektedir. Elde edilen sonuçlar bir çıktı arayüzü vasıtasıyla ileride oluşturulması planlanan tahmin sonuçları için kullanılmaktadır.

Bu işlemler sırasında karşılaşılan sorunların en başında veri seyrekliği gelmektedir (Grčar vd., 2005: 58). Sistem içerisinde her kullanıcının her ürün için bir puan/değer atamadığı varsayıldığında, oluşturulacak matrislerde boşluklar oluşması kaçınılmaz olacaktır. Modelin verimli ve istenildiği gibi çalışabilmesi için bu seyreklikleri elimine edebilecek algoritmalara

yer verilmelidir. İşbirlikçi filtreleme modellerinde temel fikir henüz puanlanmamış/değerlendirilmemiş boşlukları, kullanıcılar tarafından sağlanan diğer gözlemlenebilir veriler vasıtasıyla tahmin edebilmektir.

Bu modellemeleri oluşturabilmek için iki tür yöntem üzerinde yoğunlaşmıştır (Breese vd., 1998: 43);

Hafıza Temelli Yaklaşımlar

“Komşu temelli filtreleme” olarak da adlandırılana hafıza temelli metotlar kullanıcıların komşuluk ilişkileri doğrultusunda kullanıcı-ürün puanlama tahminlerini oluştururlar (Aggarwal, 2016: 9).

Bu tür sınıflandırmada modele dahil edilen veriler olduğu gibi kullanılabilmesi için tek bir parça halinde saklanır. Tahmin oluşturma aşamasında saklanan veriler doğrultusunda belirlenmiş benzer kullanıcı veya ürünler sıralanır. Bu sıraya göre tahmin çıktıların üretimi gerçekleşir. Hafıza temelli işbirlikçi filtreleme yöntemlerinde daha az parametreye ihtiyaç duyulur ve bu sayede daha fazla parametre ayarı gerekmektedir. Ancak soğuk başlangıç ya da seyreklik problemlerinin aşılabilmesi için ek yöntemlere ihtiyaç duyar.

Komşuluk prensibine göre hafıza temelli metotlar iki grupta incelenmektedir (Wang vd., 2006: 107);

Kullanıcı Temelli İşbirlikçi Filtreleme: Kullanıcılar için oluşturulması planlanan öneri sistemi benzer özelliklere sahip kullanıcıların yardımıyla hedef kullanıcı için çıktı oluşturur. Temel prensip; bir x kullanıcısına benzer puanlamayı yapmış kullanıcı grubu tespit edilerek, bu denk grup içerisinde bir değer oluşturulması ve kullanıcının henüz puanlamadığı ürünler hakkında öneriler elde etmesini amaçlamaktadır. Bu öngörü için makine öğrenme tekniklerine başvurulmaktadır.

Ürün Temelli İşbirlikçi Filtreleme: Ürün dikkate alınarak oluşturulan bir işbirlikçi filtreleme modelinde kullanıcının bir ürün için yapabileceği değerlendirmeyi tahmin edebilmek için, model tarafından belirlenen kriterler çerçevesinde ürüne benzer gruplar oluşturularak bu grup içinde kullanıcılar tarafından yapılmış değerlendirmeler incelenir ve bir öneri mekanizması oluşturulur. Böylece kullanıcının benzer ürünler hakkında beğenileri yeni ürün için de geçerli olacağı düşünülür.

Hafıza temelli modelleri ile sade ve uygulaması basit algoritmamalar oluşturmaları mümkün olmasına rağmen, değerlendirme matrislerindeki seyreklik sorunu nedeniyle doğru sonuçlar elde edilmeyebilmektedir.

Model Temelli Yaklaşımlar

Makine öğrenme, derin öğrenme ve veri madenciliği yöntemlerinin sıklıkla kullanıldığı bu metotta, parametrelerle ifade edilmiş modellerle optimize bir öneri çıktısı elde edilmeye çalışılmaktadır. Karar ağaçları, örtük faktör modelleri, bayieysen metotlar oluşturulan bu model içerisinde kullanılarak bir öneri sistemi algoritması oluşturulur. Model temelli metotlar seyrek matris yapılarında iyi sonuçlar verebilmektedirler.

Tüm bu metotlarda kullanılan temel veri kaynağı olan kullanıcılar tarafından ürüne biçilen değeri gösteren değerlendirme kriteri, birçok formatta ortaya çıkabilmektedir. Ürün, kullanıcı tarafından dolaylı veya doğrudan bir değerlemeye tabi tutulabilmektedir. Örneğin Avi Nathanson, Ephrat Bitton, ve Ken Goldberg. Tarafından geliştirilen “Jester” Şaka Öneri Sistemi, kullanıcılar arasında oldukça sübjektif bir durum olan şaka/espri algısını ölçmek için kullanılmıştır. Bu sistemde tanımlanan 1 ile 10 arasında bir gösterge çizelgesi arasında devamlı değerler ile kullanıcıya sunulan esprileri değerlendirmesi istenir (Nathanson vd., 2007: 493). Kullanıcı rakamsal değerleri görmeden bir ölçek yardımıyla değerlendirmesini yaparak öznel görüşünü değerlendirilebilir bir veriye dönüştürebilmektedir.

Öte yandan aralıklı ölçek üzerinden puanlama uygulamaları daha sık rastlanan bir ölçme metodu olarak karşılaşılmaktadır. 5, 7 ve 10 ölçekli puanlama uygulamaların sıklıkla kullanılmakta olup kullanıcı tarafından değerlendirilmesi için rakamsal, metinsel veya görsel ifadelerle kullanıcının ürün hakkında görüşlerini ölçek üzerindeki bir puanla tanımlaması istenir.

Bir diğer kullanım alanı ise sıralı kategorik değerlerdir. Kullanıcıyı kalıplaşmış ölçekler içerisinde iki uç değerlendirme arasında seçim yapmasına olanak sağlayan kategorik ölçekli puanlama, “Kesinlikle katılmıyorum, Katılmıyorum, Katılıyorum, Kesinlikle Katılıyorum” gibi ifadelerle sınırlandırılarak kullanıcının görüşünü belirli bir kategori altında belirtmesi sağlanabilmektedir. Bu kategorik yapı ikili (binary) yapıda da olabilmektedir. “Evet-Hayır”, “Beğendim-Beğenmedim” gibi kullanıcının keskin ve net bir fikir belirtilmesi istenen yapılar 0, 1 ile kodlanarak kullanıcının ürün hakkındaki görüşü hakkında kestirim yapılabilir. Çok sık rastlanmasa da birli (tekli) yapılar da günümüz sistemlerinde kullanılmaktadır (Rossetti vd., 2013: 164). Örneğin sosyal medyada sıklıkla kullanılan “Beğenme”, “Favorilere Ekleme” eylemi birli puanlama sistemlerine iyi birer örnektir. Kullanıcı tek taraflı ve istemli olarak beğenisini belirtebilmekte ancak sistem beğenmeme fikrini belirtebileceği bir araç sunulmamaktadır. Bu nedenle kullanıcı beğendiğine dair fikrini diğer insanların da görebileceğini bilerek ve isteyerek kesin bir şekilde ortaya koyabilmektedir.

Birli puanlama yapısında kullanılan bir diğer yöntem ise satın alma işlemleri ile elde edilen veri yapısının bu yapıya dönüştürülmesidir. Kullanıcılar bir sistem içerisinde yaptıkları alışverişlerde doğrudan olmasa bile ürün hakkında bir fikir belirttiği varsayılarak bu fikrin kullanıcının ürün hakkındaki eğilimi olduğu varsayılmaktadır. Bu durumda kullanıcı ürünü satın almışsa o ürünü başkalarına da tavsiye edebilir algısı üzerinden hareket edilmektedir. Bu durumda açık puanlama sisteminden çok kullanıcının yaptığı her alışverişinden örtülü olarak bir puanlama sistemi benzeri yapı oluşturulması gerekmektedir.

Kullanıcının birden fazla seçenek arasından tercihini belirtebilmesi, pozitif veya negatif yönde ürün tercihini belirtmesi bir avantaj olsa da Birli bir puanlama tercihi ile kullanıcının ürün hakkında net ve ayırt edici bir görüş belirtebilmesi avantaj sağlayabilmektedir. Oluşturulan öneri matrisinde tıpkı ikili yapıda olduğu gibi kullanıcının olumlu davranış sergilediği ürünler için 1 diğer ürünler için 0 değeri kullanılarak oluşturulur (Bodapati, 2008: 78).

İçerik Temelli Öneri Modelleri

İçerik temelli öneri sistemlerinde ürünlerin tanımlayıcı bilgileri, öneri oluşturmak için kullanılan başlıca veri kaynaklarıdır İçerik temelli öneri modellerinde ürün tanımları kullanıcı için ilgili içerik olup olmadığına göre analiz edilmektedir (Pazzani ve Billsus, 2007: 325). Bu nedenle içerik sözcüğü ürünü tanımlayabilecek tüm veriler için kullanılan genel bir tabirdir İçerik temelli filtreleme ürün içeriği ile kullanıcı tercihlerinin işbirlikçi filtrelemenin aksine ürün odaklı korelasyonu tespit eden uygulamalardır (Van Meteren ve Van Someren, 2000: 50). Bu tür öneri sistemlerinde ürüne verilen puan ve yorumlar da dahil olmak üzere ürün hakkında kategorize edilebilir tüm veriler, öneri sistemi için kullanılabilir yöntem ve algoritmalar için kullanılabilirler. Kullanıcı bir ürün hakkında fikir belirttiği zaman diğer kullanıcıların o ürün hakkındaki fikri içerik temelli sistemlerde kullanılmaktadır.

Diğer öneri sistemi modellerinde olduğu gibi bu modelde de kullanılan ürün açıklamaları eğitime verileri (training data) içerisinde kullanıcı temelli sınıflandırma veya regresyon modelleme problemlerinin oluşturulması için kullanılır (Aggarwal, 2016: 15). Her kullanıcı için, eğitime verileri kullanıcının satın aldığı ya da ürün hakkında herhangi bir görüş bildirdiği ürün özelliği olarak betimlenir. Yapılan tüm hesaplamalar kullanıcı özelinde satın alma davranışı veya ürünle olan etkileşimi dikkate alınarak başka ürünler için satın alma eğiliminin ölçümüne olanak sağlar.

İçerik temelli öneri sistemleri kullanıcının geçmişte diğer kullanım profilleriyle gösterdiği benzerlikleri dikkate alarak, kullanıcı-ürün eşleşmesinin sağlamayı amaçlamaktadır.

Hedef kullanıcının diğer kullanıcılardan farklı olarak ürün hakkında yaptığı değerlendirmenin aksine kullanıcının olumlu ya da olumsuz değerlendirme yaptığı ürünün özellikleri anahtar rol oynamaktadır.

Kullanıcılar arası değerlendirme kriterini ön planda tutan işbirlikçi filtreleme modelinin aksine, hedeflenen kullanıcının kendi değerlendirmeleri ve değerlendirmenin yapıldığı ürünün özellikleri dikkate alınarak bir model oluşturulur. Bu nedenle diğer modelleme uygulamalarından farklı veri tiplerine ihtiyaç duyabilmektedir (Van Meteren ve Van Someren, 2000).

Veri tiplerini belirlerken hedeflenen kullanıcı ve kullanıcının etkileşim sağladığı ürünlerin özellikleri dikkate alınır. Kullanıcı için sistem içerisinde sistem tarafından oluşturulacak kullanıcı profili ile elde edilecek model, kullanıcının doğrudan sisteme verdiği bilgilerden elde edilebileceği gibi sistemin dolaylı yoldan kullanıcı davranışlarını inceleyerek ortaya çıkarabileceği bilgileri de içerebilir. Hedef kullanıcı ilgi alanlarını sisteme giriş yapabilir kullanıcının sistem içerisindeki aktiviteleri ile elde edilebilen bilgiler dolaylı veri setlerine dahil edilebilir. Ürün için toplanan bilgiler ise genellikle doğrudan sisteme dahil edilmiş veri setlerinden oluşmaktadır.

Öte yandan bu sistem, sadece belirgin olan ve derinlemesine oluşturulmuş bir sonuç üzerinden öneriler oluşturmayabilmektedir. Eğer bir kullanıcı sistem üzerinde herhangi bir iz (yorum, puan, alışveriş) bırakmamışsa sistemin kullanıcı için öneri oluşturması mümkün olmamaktadır. Bunun temel sebebi oluşturulan sistemin kullanıcı merkezinde sonuç üretebilmesi ve sistem içerisindeki diğer kullanıcı bilgilerinin tespit edilememesinden kaynaklanmaktadır. Yeni kullanıcı girişleri sistemin derinliğini azaltmakta ve sistemin tamamı üzerinde istenmeyen öneriler oluşmasına neden olmaktadır.

Bu sistem; henüz kullanıcılar tarafından hiçbir etkileşim içine girmemiş veya diğer öneri sistemleri vasıtasıyla sonuç elde edebilecek bir bilgi kaynağı oluşturulamaması durumunda, kullanıcının satın alma davranışının kestiriminde iyi sonuçlar verebilmektedir. Yeni bir ürün için öneri mekanizması oluşturulmak istenmesi durumunda hızlı ve tutarlı sonuçlar elde edilebilmektedir.

Bilgi Temelli Filtreleme Modelleri

Bilgi temelli filtreleme, ürün özelliklerinin kullanıcı ihtiyaçlarını ve tercihlerini nasıl karşılayacağına dair özel bir uzmanlık alanı dikkate alınarak oluşturulan öneri sistemleridir. Bu tür öneri sistemleri özel bir kurgu odağı merkezinde oluşturulabilmektedir (Smyth, 2007: 342).

Problem tanımını açıklayabilmek için benzerlik fonksiyonları, öneri sisteminin ne kadarlık bir kısmını kullanıcı ihtiyacını karşıladığını tespit edebilmek için benzerlik skorları oluşturur.

Belirlenen veri tipleri ile karar vericiler tarafından oluşturulan kısıtlar çerçevesinde farklı öneri sistemi modelleme tekniklerinden de yararlanarak yeni modeller kullanılabilir. Kullanıcı davranışları sisteme önceden tanımlanan kriterler çerçevesinde değerlendirilerek ihtiyaç duyulmayan bilgiler analizlere dâhil edilmemektedir. Diğer modellerden farklı olarak Kurgu odaklı bir sistemin tersine, bir olay döngüsü yerine net ve sistemli bir karmaşa içine sokmayan çapa noktaları oluşturmasıdır.

Bu tür sistemler, farklı öğrenme algoritmaları ile beslenmesi durumunda sağlıklı sonuçlar ortaya çıkarabilmektedirler. Bu bilgiler karar vericiler tarafından da sağlanabildiği gibi sistemde oluşturulan kayıt dosyalarının yorumlanabildiği algoritmalarca da geri bildirim oluşturulabilmektedir.

Topluluk Temelli Filtreleme Modelleri

Bu tür filtreleme modellerinde kullanıcının dâhil edildiği topluluk üzerinden ürün önerisi geliştirilir. “Sana arkadaşını söyle, sana kim olduğunu söyleyeyim” özlü sözü üzerinden kurulan sistem kullanıcıların daha önce güven ilişkisi kurduğu kişiler tarafından ürün hakkındaki görüşlerinin değerlendirme açısından önemi üzerine kurgulanmıştır (de Mello Neto ve Nowé, 2012: 37) topluluk temelli öneri sistemi, kullanıcılar ve öneri kaynakları arasında bireyin sosyo-ekonomik olarak kendisine benzer olanlarla ilişki kurma eğilimi, bağlılık gücü, güven ve saygınlık gibi kriterlerle ilişki göstergelerini dikkate alarak çıktı üretir. (Arazy vd., 2009: 34). Son yıllarda yapılan çalışmalarda “arkadaş” kavramı kişilerin kendi aralarında kurulan kişisel bağlantıların yanı sıra sosyal medya aracılığıyla kurulan bağlantıları da kapsadığı görülebilmektedir (Golbeck, 2006: 96). Kullanıcılar arasında kurulan bu bağ ister kişisel arkadaşlık ilişkileri olsun, isterse sosyal medya aracılığıyla kurulan bir bağlantı olsun “güven” ilişkisi çerçevesinde gerçekleştirilebilmektedir. Özellikle belirli bir alan içerisinde kullanıcıların ortak ilgi alanı veya zevkleri doğrultusunda kurulan (hobi, mesleki gelişim, akrabalık ilişkileri) birliktelikler kullanıcıların tercih öncesi ilk danıştıkları bilgi kaynakları olmaktadır. Bu durum hem karar vericiler tarafından uygulanabilecek öneri sisteminin inşasında kolay ve güvenilir bir yapı oluşmasına olanak verirken, kullanıcının öneri sistemi tarafından sunulan bir ürünü tercih ederken daha kararlı ve güvenerek tercih yapması imkânını da sunmaktadır.

Topluluk temelli filtreme ile oluşturulan ödeme sistemleri 2000’li yıllar itibarıyla sıklıkla başvuru ve yaygın bir şekilde kullanılmaya başlanılan bir sistemdir. Bu tür öneri

sistemleri olumlu sonuçlar elde edilmesine olanak sağlasa da geleneksel işbirlikçi model ile üretilen öneri sistemleri kadar verimli çalışmadığı da görülebilmektedir (Groh ve Ehmig, 2007: 127). Daha önce benzer ürünler hakkında yapılan fikir alışverişleri, yeni ürünler veya kullanıcılar için öneri sisteminin ilk etapta çıktı elde edememesi (soğuk-başlangıç problemi) gibi ön yargı yaratabilecek çıkarımlar sistemin verimli çalışmamasına neden olabilmektedir. Bu tür durumlarda işbirlikçi model ile oluşturulan hibrit çözümler ön plana çıkmaktadır

Hibrit Öneri Modelleri

Hibrit öneri sistemleri iki veya daha fazla öneri tekniğinin tek başına kullanımının sebep olabileceği eksikliklerinin azaltılarak daha iyi bir performans kazanımı sağlamak amacıyla uygulanan yöntemlerdir. Çoğu zaman işbirlikçi filtreleme tekniklerinde karşılaşılan sorunların önüne geçmek için kullanılır (Burke, 2002: 337). Belirli bir öneri sistemi tekniği yerine birden fazla filtreleme tekniğinin bir arada kullanıldığı yöntemlere genel olarak hibrit sistemler denmektedir. Bir öneri sisteminin kullanılan teknik nedeniyle oluşturduğu dezavantajları başka bir filtreleme tekniği kullanarak ortadan kaldırılması bu tür yöntemin kullanılmasının en başlıca sebeplerinden biridir. Örneğin kullanılan yöntem nedeniyle sisteme dahil edilen yeni kullanıcılar için soğuk-başlangıç probleminin ortaya çıkması ve bu kullanıcılar için herhangi bir öneri çıktısı oluşturulamaması durumunda içerik-temelli filtreleme tekniklerinin yardımına sıklıkla başvurulabilmektedir. Bu gibi durumlarda standart filtreleme teknikleri genişletilerek ortaya çıkan dezavantajlı durumları ortadan kaldırılması mümkündür (Burke, 2007: 26).

1.7. Öneri Sistemi Modellerinin Gelişimi ve Literatür Taraması

Öneri sistemleri üzerine 1990'lı yılların ortalarından bu yana devam eden akademik çalışmalar, günümüzde hızla değişen teknolojik değişimlerle birlikte çeşitlilik ve hız kazanmış durumdadır (Wei vd., 2007: 1). Bu durumun başlıca sebepleri bilgisayar işlem gücünün artması, bilgi depolama alanlarının genişlemesi ve kullanıcıların bilgi erişimi ve internet alışkanlıklarının hızla değişmesi olarak gösterilebilmektedir. Özellikle e-ticaretin hızlı bir şekilde teknoloji çağına uyum sağlama kabiliyeti kullanıcılara yeni ürün çeşitliliğinin sunması açısından öneri sistemlerinin yoğun kullanılmasına, dolayısıyla da akademik alanda da hızlı bir gelişim sağlamasına olanak sunmuştur.

Dijital ortamda son kullanıcıya erişimin internet yoluyla kolaylaşması, şirketlerin pazarlama stratejilerini farklı bir perspektifle yorumlama ihtiyacı duymasına yol açmıştır. Tamamen internet üzerinden perakende ürün satışı yapan şirketlerin artmasıyla müşteriye nasıl ulaşılabileceği hususunda yeni stratejilere ihtiyaç duyulmuştur. Bu doğrultuda akademik anlamda da 1990'lı yılların sonu, 2000'li yılların başındaki mevcut çalışmalar, sıklıkla bilişim

dünyasında yapılan çalışmalarla desteklenmeye başlamıştır. E-ticaretin kullanımının giderek yaygınlaşması kullanıcılarına dünya çapında ürün satışı yapan web sitelerinin devamlı olarak öneri sistemleri üzerinden pazarlama stratejisi arayışlarını arttırmasına yol açmış, bununla paralel olarak bu konudaki bilimsel çalışmalar için veri seti kaynağı olmaya başlamıştır.

1970’lerde başlayan bilgi sistemi araştırmaları ile Vektör Uzay Modeli (Salton vd., 1975: 613) ve SMART sistemleri (Rocchio, 1971: 313) ile bilgi erişim metotlarının gelişimi ile başlamıştır. Araştırmacılar Bilgi Edinim modellerini geniş ve asimetrik bir şekilde vektörler arası kosinüs benzerliği kullanarak çok boyutlu bir uzayda modellemişlerdir. Kesinlik (precision) ve recall (hassasiyet) terimleri bu modeller üzerinden oluşturulmuş ve öneri sistemlerinde kullanılan istatistiksel analizler için de temel olmuştur. 1970 sonlarında akademik çalışmalar ile elektronik haberleşme araçlarına yoğunlaşılması, bu modellerin derinleşmesine olanak sağlamış, kelime işlemcilerinin ve hesaplama araçlarının yaygınlaşması veri setlerinin çeşitliliğini arttırmıştır. Bu durum Denning (1982: 163) tarafından “Electronic Junk (Elektronik Çöplük)” terimiyle literatürdeki yerini almıştır. Dijital veri üretiminin bu gelişmeler karşısında artması bilgi sistemi araştırmalarına da yeni bir yön katmıştır. Araştırmacılar elde edilen bilgileri doğrudan işlemek yerine zorunlu olarak alakasız veya uygunsuz verilerinin öncelikli olarak elimine edilmesi konusunda çalışmamalara yön verme mecburiyetinde kalmıştır. Bunun yanında bilgi kategorize etme ve filtreleme teknikleri veri temizliğinin kaçınılmaz yan ürünler olmuştur.

1991 yılında düzenlenen workshop sonucu yayımlanan “Communication of the ACM” dergisinde bulunan makalelerle, Bilgi filtreleme, bilgi sistemleri çerçevesinde bir alt çalışma konusu olarak kabul görmüştür (Loeb ve Terry, 1992: 26). Bilgi edinim sorgu vb. yöntemlerle kısa dönem için ilişkili bilgileri ortaya çıkarma yöntemi olarak tartışılırken, bilgi filtreleme modelleri ise genelleştirilmiş veya alakasız bilgilerin orta veya uzun dönemler içerisinde elimine edilmesi olarak şekillenmiştir. Bilgi filtreleme modelleri zaman içinde içerik-temelli filtreleme sistemlerinin temelini oluşturmuş ve gerçek dünya verileri ile uygulamalar gerçekleştirilmesi sağlanmıştır.

Bu çalışmaların yanı sıra Foltz ve Dumais (1992: 51) tarafından yazılan bir makalede örtülü semantik endeksleme (latent semantic indexing) tekniği kullanılarak kavram-metin matrisleri için boyut indirgeme teknikleri ortaya koyulmuştur. Öneri sistemleri için daha önemli olan bir gelişme ise işbirlikçi filtreleme modelinin Tapestry adında bir öneri sistemi üzerinde uygulanmasıdır (Goldberg vd., 1992: 61). Bu uygulamayla belirli bir doküman arşivi içerisinde kullanıcı etkileşimlerini kullanarak yeni kullanıcılara arama önerilerinin sunulduğu bir sistem

oluşturulmuştur. Bu uygulama literatürdeki ilk işbirlikçi filtreleme modelinin kullanıldığı öneri sistemi olarak kabul edilmektedir (Resnick ve Varian, 1997: 56).

İnternet ağının genişlemesi, geniş kullanım alanının gün geçtikçe yayılması ve popüleritesinin artması internetin insan yaşamının bir parçası kabul edilerek günlük hayatta içselleştirilmesi, veri derinliğinin artmasına imkân tanımış ancak bu veriler içerisinde gerekli ve kullanılabilir verilerin ortaya çıkarılmasını zorlaştırmıştır. İnternet kullanıcılarının günlük yaşam ve davranışları hakkında izleri bu veri yığını içerisinde yer alması ile birlikte “Sosyal bilgi filtreleme” teknikleri ortaya çıkmaya başlamıştır (Shardanand ve Maes, 1995: 211).

1997 yılında yeniden yapılan “Communication of the ACM” konferansı ile öneri sistemleri olgunlaşma aşamasına girerek işbirlikçi filtreleme modelleri konusundaki çalışmalarda yoğunlaşma gerçekleşmiştir. Resnick ve Varian (1997: 56) tarafından yazılan makalede işbirlikçi filtreleme modellerinde önericilerin, açık bir şekilde öneri alıcılarının işbirliğine ihtiyaç duymayabileceğini bu nedenle bir sistemin bütünü olarak değerlendirerek “Öneri Sistemleri” kalıbını kullanmıştır. Böylece “öneri sistemleri” terimi literatürde ilk kez yerini almıştır.

1990’lı yıllarda hız kazanan akademik çalışmalar e-ticaret sektörü tarafından da elde edilen yöntemlerin gerçek zamanlı veriler üzerinde de uygulanabilirliği test edilerek öneri sistemlerinin gelişimi desteklenmiştir.

Amazon ve Netflix ile başlayarak perakende ürün ve hizmet satışı yapan e-ticaret sitelerinin yarattığı gelişim, eğlence sektöründe iTunes, last.fm, Pandora ve IMDB, rottentomatoes, Metacritic gibi kullanıcının zevklerinin ve tercihlerinin ön planda olduğu tavsiye sistemlerinin ortaya çıkmasına olanak sağlamıştır. Oluşan bu sistemde bilimsel çalışmalar büyük veri setlerini kullanım imkânı sağlayan bilişim şirketleri ile karşılıklı bir fayda çizgisi çerçevesinde gelişim göstermiştir.

Amazon şirketi 1994 yılında kurulması ile birlikte sektörde öncü bir konum almış ve 1998 yılında işbirlikçi öneri sistemleri üzerine algoritmalar geliştirmiştir (Linden vd., 2003: 78) Geniş ürün yelpazesi ve kullanıcı profili ile çok sayıda ürün içerisinde her kullanıcı için anlık olarak küçük ölçekte ürün öneri sunumunu gerçekleştirmesi ile birlikte elde ettiği başarıyı 2003 yılında yayımlamıştır. (Smith ve Linden, 2017: 12).

Netflix ise en büyük atılımlardan birini gerçekleştirerek büyük ödüllü bir yarışma düzenlemiş ve bu yarışma sonucunda %10 oranının üzerinde bir başarı ile bu tür sistemlerin önemini göstermiştir (Bell ve Koren, 2007: 75).

Literatür taramasında, tez çalışması uygulamasında kullanılan veriler için üst sınır olan 2019 yılı aralık ayı dikkate alınarak akademik çalışmalar derlenmiştir. Taramada

çalışmalardaki araştırma metodolojisi, sınıflandırma kriterleri, kullanılan öneri sistemi filtreleme yöntemleri ve kullanılan araçlar üzerine yoğunlaşmıştır.

Çalışmada bilgi teknolojileri konusunda yapılan 2012 – 2019 yılları arasında öneri sistemleri üzerine yayımlanmış akademik çalışmamalar incelenmiştir. Belirlenen kriterler doğrultusunda uygulamalı olarak gerçekleştirilen 96 çalışma incelenerek bir derleme ortaya konulmuştur.

Yapılan araştırma internet ortamında çevrimiçi akademik dergiler üzerinden elde edilerek gerçekleştirilmiş olup, “öneri sistemleri”, “ürün öneri sistemleri”, “kişiselleştirilmiş öneri sistemleri”, “öneri filtreleme” gibi çeşitli anahtar kelime öbekleri dikkate alarak bulunmuştur. Yapılan çalışmada teorik çalışmalar yerine öneri sistemlerinde makine öğrenme teknikleri ile gerçekleştirilen uygulamalar göz önüne alınmıştır.

Literatür taraması yapılırken çalışmanın amaç ve uygulama alanı dikkate alınarak 3 ana başlık üzerinden değerlendirilmiştir.

- Kullanılan Öneri Sistemi filtreleme teknikleri
- Uygulama Alanları
- Kullanılan yazılımlar

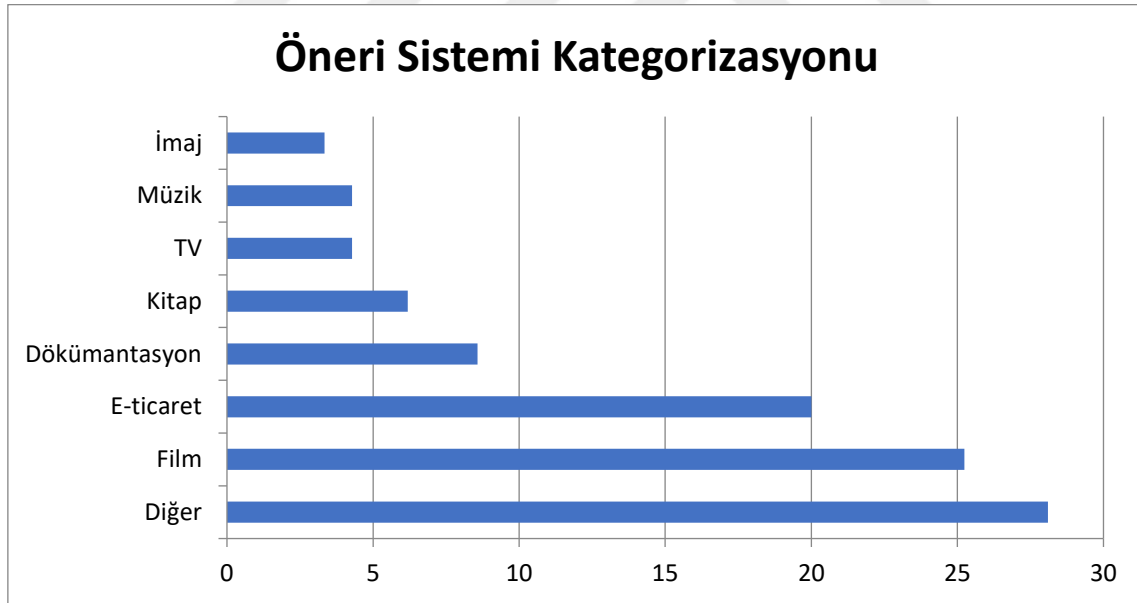
Son yıllarda kullanılan farklı yöntem çeşitliliğine rağmen akademik çalışmalarda temel çıkış noktaları bu bölümde bahsi geçen filtreleme tekniklerine bağlı kalmış, adaptasyonlar ve farklı algoritma denemeleri literatürde standartlaşmış bir filtreme tekniği üzerinden gerçekleştirilen çalışmalara yer verilmiş ve bu çizginin dışında yer alan filtreleme teknikleri “Diğer” seçeneği ile belirtilmiştir. Yine taramaya dâhil edilen çalışmalarda konu edilen filtreleme tekniği için herhangi bir uygulama alanı üzerinde çalışma yapılarak sonuçların paylaşılması hususuna dikkat edilmiştir. Literatür taraması bir dizi ön araştırma ışığında gerçekleştirilmiştir (Park, D. H. vd., 2012: 10059; Lu vd., 2015: 12; Alyari ve Navimipour, 2018: 985).

Gerçekleştirilen uygulama çalışmaları öncelikli olarak öneri sisteminin konu edildiği ürün pazarı ile ilgili bir sınıflandırma yapılmıştır. Bu sınıflandırmada dijital ortamda ürün olarak tanımlanan sistem girdilerinin belirli kategoriler altında toplanarak öneri sistemlerinin akademik boyutta kullanım alanları listelenmiştir.

Çalışmalarda özel kapalı sistemlerden elde edilen verilerin yanı sıra halka açık ve akademik çalışmalarda sıklıkla kullanılan veri setleri ile karşılaşılmaktadır. Yapılan literatür taramasında ortaya konulan çalışmaların yaklaşık yüzde 30’unda veri seti ile ilgili bilgi paylaşılmamış ya da spesifik veya herhangi bir kategoriye sokulamayan verilere yer verilmiştir. Öte yandan diğer çalışmalar sinema filmi tanıtım ve içerik sağlayıcı siteler, belirli bir

dokümantasyon sistemi üzerinden metin öneri sistemler barındıran sistemler, kitap, müzik, televizyon programları ve diğer multimedya içerik sağlayan sistemlerin kullandığı öneri sistemleri de kullanılmaktadır.

Çeşitli özelleştirilmiş veya açıklanmamış çalışma alanları dışında Sinema filmi, televizyon dizisi ve benzeri yapımların kullanıldığı çalışmalar yapılan taramanın %27'sini oluşturmaktadır. Benzer özelliklere sahip Kitap ve müzik eserlerinin %7 ve %4'lük oranları ile birlikte değerlendirildiğinde yaklaşık %40'lık bir çalışmanın kullanıcıların sanatsal zevk ve tercihlerinin yönlendirilmesi için öneri sistemlerinin sıklıkla kullanıldığının göstergesi olarak kabul edilebilmektedir. Öte yandan metin/kelime temelli öneri sistem yöntemlerinin kullanıldığı dokümantasyon kategorisinde mevcut taramanın %8'ini kapsamakta olup, bu tür çalışmaların yeni ve geliştirme aşamasında olmasından dolayı akademik çalışmalarda sonuca ulaştırılmış yeterli çalışma sayısına henüz ulaştırılmamıştır. Benzer çalışmalar “Diğer” kategorisinde de yer almakta olup, özellikle sosyal medya aracılığıyla kullanıcılar tarafından yapılan anlık paylaşımların incelemelerine de sıklıkla rastlanılmaktadır. Bahse konu kullanım alanları Şekil 1.2’de gösterilmiştir.

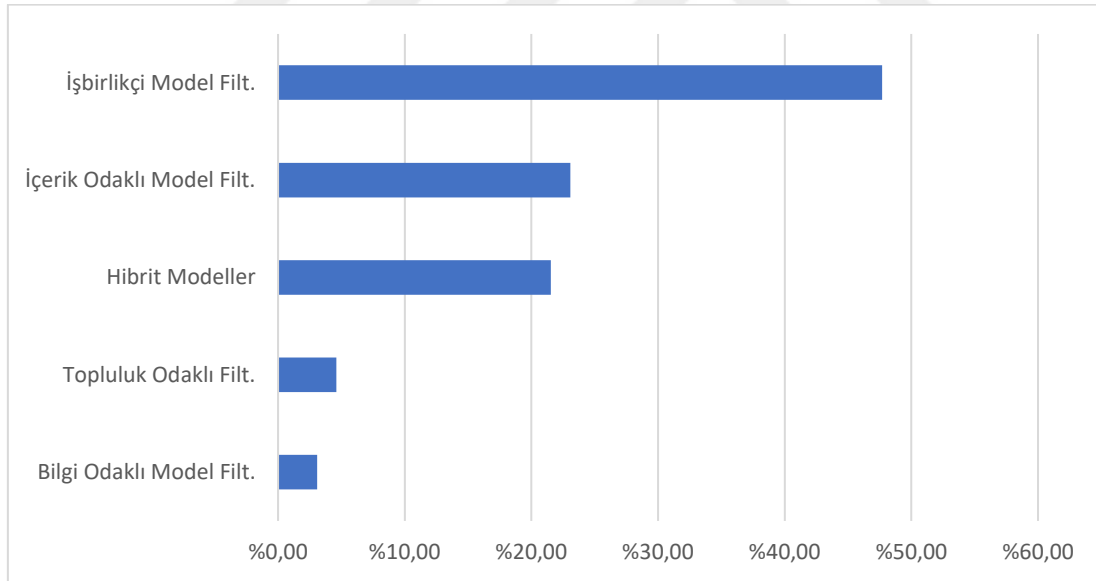


Şekil 1.2 Öneri Sisteminin Literatürde Kullanım Alanları

Literatür taraması sonucu yapılan çalışmalarda, yukarıda bahsi geçen öneri sistemleri incelendiğinde ise işbirlikçi filtreleme modellerine daha sık başvurulduğu gözlemlenmektedir. 2000’li yılların başından itibaren ortaya atılan bilgi temelli filtreleme teknikleri ve hibrit temelli modeller son yıllarda ivmelenerek kullanımı hızla artmaktadır (Eirinaki vd., 2018: 414). Bu durum yapılan taramada da gözlemlenmiştir (Şekil 1.3). İncelenen çalışmalarda işbirlikçi filtreleme tekniği 68 çalışmada kullanılmış, bu modellerin ardından içerik temelli filtreleme

teknikleri %15'lik 16 adet bir oranla gelmiştir. Diğer taraftan topluluk temelli filtreleme teknikleri %11 10 adet, Bilgi temelli filtreleme teknikleri ise 2 adet çalışmaya rastlanmıştır. Bu çalışmalar literatür taramasının kapsadığı son 3 yıl içinde gerçekleşmiştir. İşbirlikçi ve içerik temelli filtreleme tekniklerinin ardından en sık karşılaşılan filtreleme tekniği olan hibrit modellerin sayısı 14 olmasına rağmen, bu çalışmaların çoğunun son yıllarda gerçekleştirildiği gözlemlenmiştir. Önümüzdeki yıllarda gerçekleştirilecek çalışmalarda işbirlikçi ve içerik temelli modeller kadar popüler olması muhtemel olarak düşünülmektedir.

Günümüz çalışmaları veri kaynaklarının erişilebilir ve açık olması, şirketlerin büyük veri ve dolayısıyla öneri sistemlerinin önemini fark etmesini sağlamış ve böylece çalışmaların hızlı bir şekilde çoğalmasına ve yeni modellerin farklılaştırılarak çeşitlenmesine olanak sağlamıştır. Öneri sistemlerinde tüm bu modelleme teknikleri oluştururken karşılaşılan farklı zorluklar aşikâr bir durum olarak ortaya çıkmıştır. Zorlukların sebebine inilirken yapay zeka açısından yeni düşünce yapıları olgunlaştırılmalı ve çoklu yapılara olanak tanınmalıdır (Burke vd., 2011: 14). Bu durum günümüz çalışmalarında tek bir filtreleme modelinin kullanımından ziyade farklı modellerin ayrı ayrı ya da birleştirilerek kullanımına olanak sağlamıştır.



Şekil 1.3 Literatürde Öneri sistemlerinde Kullanılan Metotların Dağılımı

Bahsi geçen tüm öneri sistemi tekniklerinin artısı ve eksileri çoğu kez literatürde tartışılarak açıklığa kavuşturulmuştur (Burke ve Robin, 2002: 338). Çoğu model diğer farklı bir modele üstünlük sağlayamadığı açıkça görülmesi hibrit modellerin oluşumunu tetiklemektedir. Hibrit modeller genellikle öneri sisteminin uygulanacağı ana sistem ve sistemin ürettiği verilerin özelliklerine göre farklı öneri sistemi tekniklerinin birleşimi ile (Burke ve Robin, 2007: 377). Bu da özelleştirilmiş bir yapı oluşturması nedeniyle karar vericiler için avantajlı bir

çözüm olarak ortaya çıkar. Hibrit çözümler son yıllarda popülaritesi artmaya başlanan bir yöntemdir. Öneri sistemi bu çözümlerin uygulanabilmesi için gerekli esnekliği sağlayabilmektedir.

Yapılan akademik çalışmalarda, gelişen yazılım teknolojisinin etkileri sıklıkla gözlemlenebilmektedir. Açık kaynak kodlu yazılımların gün geçtikçe artması çalışmaların gerçekleştirilebildiği veri yığınlarının artmasının yanında bu yığınlar içerisinde verilerin temizlenerek analizlere uygun hale getirilmesinin yarattığı zorluklar nedeniyle yeni yazılımlar ortaya çıkarmıştır. Bunun ötesinde veri biliminin gelişimi ile birlikte bu amaca uygun yazılım dillerini de oluşturulmasına olanak sağlamıştır. Verilerin karmaşıklığı ve yönetiminin karmaşık hale gelmesi tek bir komut ile analizlerin gerçekleştirilmesini zorlaştırmış, belirli kurallar çerçevesinde oluşturulmuş algoritma komut dosyalarını ihtiyaç haline getirmiştir.

İstatiksel analizlerin yer aldığı akademik çalışmalarda oldukça geniş bir yazılım çeşitliliği bulunmaktadır. Ancak literatür taraması yapılan yıllar içerisinde, araştırmanın ilk dönemlerinde SPSS ve MATLAB, KNIME, WEKA gibi kullanıcı ara yüzü üzerinden işlemleri gerçekleştiren yazılımlar üzerine bir çeşitlilik mevcuttur (Dwivedi vd., 2016: 1). Araştırmanın yapıldığı aralığın son dönemlerinde gerçekleştirilen çalışmalarda, analiz edilen veri çeşitliliğinin ve büyüklüğünün artması verilerin istatiksel analizlere dâhil edilmeden önce verinin temizlenmesi, düzenlenmesi veya uyarlanması işlemleri gibi ön hazırlık gerektiren işlemler nedeniyle çeşitli yazılım dilleri aracılığıyla oluşturulan yazılımlara başvurulmuş, bu yazılımlarda da özellikle Python yazılım dilinin ön plana çıktığı gözlemlenmiştir. Bunun yanında R dilinin sağladığı esneklik ve modüler yapısı veri analizleri ve dolayısıyla öneri sistem modellemelerinde R program dilinin kullanımının da sıklıkla kullanımına olanak sağlamaktadır.

Büyük veri kaynaklarının çoğalması, şirketlerin bu tür çalışmalar için veri kaynaklarını paylaşma isteklerinin gün geçtikçe artması ve bu veri depolama olanaklarının çeşitlenmesi, Büyük veri ile istatistiksel analizlerin gerçekleştirilmesini mecburi kılmakta ve yazılım dünyasının buna yönelik olanakları sunmasına fırsat vermektedir. Ancak bu çeşitlilik tek bir ara yüz üzerinden istatiksel analizleri gerçekleştirmesini karmaşıklaştırmakta ve komut sistemi ile çalışan programları veya doğrudan yazılım dili vasıtasıyla analizlerini gerçekleştirmeye zorlamaktadır.

Açık kaynak programlama dili olan R ve Python, istatistiksel hesaplamalar için sağladığı ortam istatistikçiler ve veri analistleri tarafından sıklıkla kullanılmakla beraber kullanım alanı hızlı bir şekilde artmaktadır (Ozgur vd., 2017: 356).

Modül odaklı geliştirme seçenekleri ile kullanıcılar tarafından yapılan geliştirmelerle istatistiksel açıdan geniş bir kullanım olanağı sunabilmektedir. R her türlü veri analizini gerçekleştirmeye olanak sağlamakta ve her geçen gün dünyanın dört bir yanından bilim adamları tarafından sağlanan katkılarla oluşturulan modüller vasıtasıyla ihtiyaçları mümkün olduğunca fazla karşılamaya yetmektedir.



İKİNCİ BÖLÜM

İŞBİRLİKÇİ FİLTREME YÖNTEMLERİ VE MODEL OLUŞTURMA SÜRECİ

Öneri sistemlerinin kullanımının yaygınlaşması ve veri elde ediniminin kolaylaşması, bu sistemler üzerinde uzmanlaşma ve sistemleri özelleştirme ihtiyacının ortaya çıkmasına yol açmıştır. Bu durum ürün temelli ve kullanıcı temelli olmak üzere iki ana düğüm noktasından başlayarak, farklılaştırılmış istatistiksel analizler ile makine öğrenmesi yoluyla çeşitlenmesini sağlamıştır.

Öneri sistemlerinde en sık kullanılan metotlardan biri olan İşbirlikçi Filtreleme yaklaşımı, mevcut kullanıcı topluluğu içerisinde geçmiş deneyim, davranış ve görüşler çerçevesinde, topluluğa dahil olan yeni ve mevcut kullanıcılardan elde edilen veriler ile bilgi havuzu oluşumunu sağlar. İşbirlikçi Filtreleme modelleri geçilen dönemler içerisinde çok çeşitli algoritmalarla öneri modelleri oluşturulmak için geliştirilmiş araçlardır. Her algoritmik yaklaşım bir önceki benzer yapıdan üstün bir gelişme göstermektedir (Herlocker, Jonathan vd., 2004: 6). Bir öngörü mekanizması ile oluşturulan bu sistemler, veri çeşitliliğini beslemeleri sebebiyle günlük hayatın her alanında kullanıma uygun hale gelmektedir. Ekonomik sistem içerisinde sınırlı sayıda çeşitliliğe karşı her biri birbirinden farklı davranışlara sahip müşteri ve ürün profilleri, işbirlikçi filtrelemeyi diğer araçlardan farklı kılmaktadır.

Akademik araştırmalar açısından da geniş bir kaynak sağlayan bu metodun yıllardır yapılan çalışmalarda keşfedilen dezavantajları günbegün ortadan kaldırılmakta, performansı iyileştirilerek, algoritmaları çeşitlendirmekte ve gerçek dünya uygulamaları, yapay zekâ çalışmalarına katkı sağlamaktadır (Gunawardana ve Shani, 2009: 2936).

Temel işbirlikçi filtreleme modelleri, kullanıcı-ürün etkileşimi sonucu elde edilen veri matrislerinin girdi olarak kabul edildiği ve sayısal bir çıktının sunulduğu tahmin denklemleriyle oluşturulmaktadır. Oluşturulan bu model kullanıcının bir ürün listesi içerisinde bulunan herhangi bir ürün ile ilgili satın alma isteği veya ürün hakkındaki fikrinin oluşturulma süreci ile ilgilidir.

İşbirlikçi filtreleme teknikleri diğer bilgi filtreleme seçeneklerine göre üç temel avantaja sahiptir. (i) otomatik işlem süreçlerinde içerikleri kolay bir şekilde analiz edilemeyen nesneler için filtreleme desteği. (ii) kalite ve deneyime göre filtreleme yeteneği. (iii) tesadüfen keşfedebilme yeteneği (Herlocker, Jon vd., 2002). Bu özellikler genel bir varsayım altında tahmin oluşturma yöntemi için model oluşturma süreci başında dikkate alınması gereken

hususlardır. İçerik temelli ve işbirlikçi filtreleme yöntemleri girdi bilgi açısından farklı bir veri seti yapısına ihtiyaç duyarlar. İçerik temelli filtreleme teknikleri için veri seti Kullanıcının ilgi alanındaki bilgi havuzu içerisinde ilgi duyulan nesnelerin özelliklerinin kıyaslanması ile elde edilmektedir. İşbirlikçi filtreleme tekniği ise öncelikli olarak kullanıcı davranışlarından elde edilen bilgi birikimini veri seti olarak kabul etmektedir. Bu birikim kullanıcının karakteristik özelliğini ortaya çıkarmakta ve diğer kullanıcılarla kurulabilecek etkileşimi ortaya koymaktadır. İşbirlikçi filtreleme ile oluşturulan modelde kullanıcılar analitik davranışları ile ilgili çıktıları paylaşarak etkileşim içinde bulundukları nesnelerin hangi özelliğinin ön plana çıkması konusunda veri oluştururlar.

İşbirlikçi filtreleme metotlarının ilk ortaya atıldığı dönemlerde çıkan komşuluk öneri sistemlerinin temeli, kullanıcılardan elde edilen bilgilerden oluşan bir veri havuzu içerisinde, hedef kullanıcının ürünü tercih edip etmeyeceği konusunda tahminlerin modellenmesi ile oluşur (Herlocker, Jon vd., 2002: 288). Burada oluşturulan varsayımlar şu şekildedir;

- Kullanıcı davranışlarının tutarlı olduğu ve bu davranışlar için gelecekte de aynı yönde bir beklentinin varlığı
- Sisteme dahil olan yeni kullanıcıların, mevcut kullanıcılarla ortak yönlerinin bulunabileceğine dair inanç

Bu varsayımlar, sistematik ve tutarlı bir öneri mekanizmasının oluşması için model öncesi kabul edilmiş yol göstericilerdir. Filtreleme yönteminde tahminlerin oluşturulması için modelin kurgulanarak sisteme adapte edilmesinin amaçlanması, kullanıcı ürün ilişkisinin kuvvetli olduğu ve sistemin ana dinamiği çerçevesinde bu döngünün bozulmayacağı, geçmişe dönük verilerin devamlı olarak birbirini besleyen mekanizmaların oluşturulmasıyla ortaya çıkarılacaktır.

2.1. İşbirlikçi Filtreleme Süreci

Öneri sistemi literatüründe sistemin kurgulanması ve değerlendirilmesinde çevrimdışı algoritmik değerlendirme ana rol oynamaktadır. Bu tür analizler algoritmaların test öncesi eğiterek kullanıcı özellikleri doğrultusunda algoritmanın nasıl davranacağını gözleme şansı yaratır (Ekstrand vd., 2011: 114).

İşbirlikçi filtreleme metotları sınıflandırma ve regresyon modellerinin genellemesi olarak incelenebilir (Aggarwal, 2016: 13). Bu tür modellerde bağımlı değişken sınıfsal özellik taşıyan bir sütun olarak, bağımsız değişkenler ise bağımlı değişkenin konusuna giren nesnelerin özelliklerini gösteren veriler olarak sınıflandırılmaktadır. Öneri sistemlerinde amaç önceden kestirilmiş olsa bile kullanılacak veriler ve dolayısıyla değişkenlerin konum

ayrımı keskin çizgilerle ayrılmamaktadır. Sınıf değişkenleri ve özellik değişkenleri bağımlı ve bağımsız değişken olarak ayırt etme konusunda çift taraflı görevleri vardır (Aggarwal, 2016: 204).

Bu süreç, standart bir işbirlikçi filtreleme senaryosunda n kullanıcı bir $U = \{u_1, u_2, \dots, u_n\}$ kümesi ve m adet üründen oluşan $P = \{p_1, p_2, \dots, p_m\}$ listesi üzerinden kurgulanır. Her u_i kullanıcısının, p_{ui} ürünü ile ilgili kısmen veya tamamen bilgiye sahip olduğu varsayılmalıdır. Belirtilen ürün grubu için kullanıcıların değerlendirmede bulunmuş veya değerlendirmeye konu bir davranış sergilediği düşünülerek kullanılacak makine öğrenmesi tekniğine göre veri tekrar yapılandırılmalı ve belirlenen yöntemlerle modelin yapısı ortaya konulmalıdır.

2.2. Değerlendirme Matrisleri

İşbirlikçi filtreleme metotlarında veri seti olarak kullanılan kullanıcı değerlendirmeleri genel olarak bir puanlama skoru kapsamında kullanıcı tarafından açık bir şekilde yapıldığı düşünülür. Bu skor genellikle bir ölçek içerisinde sistem tarafından sınırlandırılmış bir numerik sıra içinde gerçekleştirilir (Lee, S. K. vd., 2010: 2143). Bu durumun dışında kullanıcının örtülü davranışları ile yaptığı hareketler sistem tarafından farklı bir algoritma ile bir skor oluşturularak da kullanıcı değerlendirilmesi oluşturulabilir.

İşbirlikçi filtreleme modelinde $m \times n$ kullanıcı-ürün değerlendirme matrisini kullanılır. Bu matristeki her girdi a_{ij} , i kullanıcısının j ürünü hakkındaki değerlendirme skorunun temsilini gerçekleştirir. Her bireysel değerlendirme belirlenen ölçek içerisindeki numerik değer ile gösterilebileceği gibi 0 değerli verilerek ya da boş bırakılarak da gösterimi gerçekleştirilebilir.

Kullanıcılardan elde edilen değerlendirmeler farklı yapılar vasıtasıyla ortaya çıkarılabilmektedir (Aggarwal, 2016: 31);

- **Sürekli Ölçekli Puanlama:** Çok sık kullanım örneği olmayan bu yöntemle kullanıcıya açık bir alan bırakarak sübjektif ve geniş bir ölçek içerisinde kullanıcı değerlendirmesi talep edilir. Kullanıcı çoğu zaman değerlendirmeye yatkın olsa bile tutarlı bir puan, yorum ve görüş bildirmekte zorlanır.
- **Aralık Temelli Puanlama:** Bu puanlama yaygın olarak kullanılan 5 veya 7 puanlı ölçekler yoluyla gerçekleştirilir. Kullanıcı ürünle ilgili görüşünü bir puan cetveli üzerinde belirtilen numerik değerler vasıtasıyla belirtir. Değerlendirme aracı ile belirlenen puan kriterleri kullanıcının ürün veya hizmet ile ilgili bilgi ve düşüncelerini net bir şekilde belirtebileceği bir aralık içerisinde eşit bir mesafe ile temsil edilebilir

şekilde hazırlanmalıdır. Aralık temelli puanlama bir hesaplama nitelikli bir puanlama ile farklı içsel skor ölçeklerine izin vererek kullanıcının niyetini belirlemede daha doğal bir yansıma olarak kabul edildiği düşünülmektedir (Koren ve Sill, 2011: 117).

- **Sıralı Puanlama:** Aralık temelli puanlamaya benzemesine rağmen bu puanlama türünde sıralı kategorik değerlendirmeler kullanılabilir. Çoğu zaman karşılaştırma net nicel ifadelerle sınırları belirlenmiş ve kullanıcının tek kelimeyle kendini ifade etmesine yardımcı olabilecek kriterlerce sınırlandırılmıştır. “Erken teslimat”, “Zamanında teslimat”, “gecikmeli teslimat” veya “Çok memnun kaldım”, “Memnun kaldım”, “Kararsızım”, “Memnun Kalmadım”, “Hiç memnun kalmadım” gibi ölçekler belirlenebilmektedir.
- **İkili Değerlendirme:** Kullanıcıların doğrudan görüşlerini, bir puanlama mekanizması ile belirtmediği ancak ürün hakkındaki dolaylı davranışları ile gözlemlenebilen hareketlerin sistematik bir değerlendirme kriteri ile ürün üzerinde olumlu veya olumsuz düşüncesi olduğu hususunda çıkarım yapılmaya çalışıldığı bir metottur. İkili değerlendirme ölçeği iki farklı sınıfa kullanıcı tarafından atama yapılması istenmesi durumunda kullanılan bir ölçüm metriğidir (Schröder vd., 2011: 80). Bu sistem ile iki seçenek arasında karar vermesi istenen kullanıcıyı kesin ve net bir ayrım ile fikrinin kayıt altına alınması hedeflenmektedir. Bu durum doğrudan kullanıcıya sorulan sorularla elde edilebileceği gibi dolaylı olarak ürün ile ilgili bıraktığı izlerden elde edilebilecek çıkarımlarla da öneri sistemi tarafından da otomatik oluşturulacak bir değerlendirme olabilir. Facebook/Twitter kullanıcılarının, oluşturulan içerikler hakkındaki tek taraflı değerlendirmelerini gösteren sembolik ifadeler bu tür geribildirim seçeneklerine örnek gösterebilir.

Bunların yanında karar verici tarafından belirlenen kıstaslar çerçevesinde de dolaylı olarak geribildirim toplama seçenekleri oluşturulabilmektedir. Örneğin kullanıcı tarafından ürünün satın alınması, sepetinde bulundurulması, bir makaleyi belirlenen standart okuma süresinden daha uzun süre incelemesi, belirli bir ürün ile ilgili sistem içerisinde yapılan anahtar kelime sorgulamalarının yoğunluğu ürüne olan ilginin tespiti için dolaylı bir değerlendirme mekanizması olarak kullanılabilir.

İşbirlikçi filtreleme modellerinde değerlendirme matrislerinin yapıları kullanılan yöntemlere göre değişebilmektedir. Ürün öneri sistemleri verilerin işlenerek standardize edilmiş bir yapı içerisinde analiz edilmesi üzerine kurgulanmıştır. Önceden belirlenmiş metotlar üzerinde gerçekleştirilecek öğrenme teknikleri ile veri yapısının uyumu gözetilmeli ve değerlendirme matrisi buna göre kurgulanmalıdır. Her ne kadar literatürde çoğunlukla

değerlendirme ya da puanlama matrisi olarak anılsa da bu metrisi oluşturan veriler daha önce bahsi geçtiği gibi doğrudan bir değerlendirme verisi olarak kabul edilmemeli ve model öncesi tespit edilen kurgu doğrultusunda uygulayıcı ve karar vericilerin tercihleri doğrultusunda bir matris oluşturulmalıdır.

2.3. Veri Hazırlık ve İşleme Süreçleri

Verileri kullanarak gerçekleştirilen çoğu analizde olduğu gibi komşuluk temelli modellerde de verilerin algoritmalara dahil edilmeden önce belirli işlemlerden geçirilmesine ihtiyaç duyulabilmektedir. Modelin tutarlılığı ve modelden elde edilecek çıktılar verilerin kalitesi ile doğru orantılıdır. Komşuluk temelli metotlarda verilerin belirli analiz içerisinde standart veri temizleme, dönüştürme, eksik verilerin tamamlanması yanında benzerlik ölçümleri ve boyutsal indirgeme araçları da kullanılarak modellerin doğru ve tutarlı kurgulanması sağlanabilmektedir (Gorakala ve Usueli, 2015: 9).

Çoğu veri madenciliği yönteminde olduğu gibi veri ön işleme teknikleri öneri sistemlerinde kullanılan tekniklerde de uygulanmaktadır. Veri ön işleme, ham verilerin işlenerek anlaşılabilir bir kalıba dönüştürülmesi süreci olarak kabul edilmektedir. Bu süreç, gerçekleştirilmesi planlanan veri analizi öncesi veriler üzerinde gerçekleştirilen tüm işlemleri kapsamaktadır (Famili vd., 1997: 5). Çoğu istatistiksel yöntemden tutarlı ve doğru sonuçların elde edinimi için veri temizleme, veri entegrasyonu, veri dönüşümü ve veri ayrıklaştırılması işlemlerinin yapılması neredeyse zorunlu bir durum olarak ortaya çıkar.

Öneri sistemlerinin çevrimiçi veya çevrimdışı uygulamalarının kullanımı bu sürecin sistematik bir şekilde işletilmesini ön şart koşar. Özellikle verilerin anlık olarak sistemden alınması ve analize dahil edilmesi için veri yapılarının net bir süzgeçten geçirilerek önceden belirlenmiş kurallar ile döngünün sağlanması gerekmektedir. Ancak bu durum da tutarlı ve verimliliği artarak devam eden öneri sonuçları elde edilebilmektedir.

2.4. İşbirlikçi Filtreleme Yöntem Çeşitleri

İşbirlikçi filtreleme sistemlerinde genellikle elde edilen verilerin örtülü veya doğrudan olup olmadığına göre bir ayırım yapılır (Herlocker, Jonathan vd., 2004: 44). Doğrudan veriler kullanıcıların bilinçli olarak tercihlerini belirttiği ve sayısal bir değere karşılık gelen değerlendirmeler olurken dolaylı veriler kullanıcı davranışlarının ve seçimlerinin yorumlanarak bir değerlendirme değeri veya tercihe atfedilmesidir.

Tüm bu verilerin sınıflandırmasının yanı sıra kullanılan sistemin eksik verileri nasıl değerlendirdiği de işbirlikçi filtreleme modeli için önemli bir olgudur (Marlin vd., 2011: 2686).

Kullanıcı ürün ilişkisi içerisinde tüm verilerin kusursuz bir şekilde değerlendirmeye tabi olmuş olması pek beklenebilecek bir durum değildir. Standart bir sistem içerisinde kullanıcı sadece tercih ettiği ya da ilgilendiği ürün için bir değerlendirmede bulunur. Bu durum kullanılan işbirlikçi filtreleme yöntemi türü ile birlikte sistemin performansı için de belirleyici bir faktördür.

Hafıza Temelli İşbirlikçi Filtreleme Modelleri

Kullanıcı temelli işbirlikçi filtreleme modelleri, belirlenen kullanıcı profili ile ikili alanların belirli ürün üzerinde değerlendirmeleri dikkate alarak benzer kullanıcı profili oluşturması için tahmin oluşturma modeli elde edilir (Breese vd., 1998: 44). Hafıza temelli işbirlikçi filtreleme algoritmalarında hedef kullanıcının tercihi ile veri seti içerisindeki diğer kullanıcı bilgileri dikkate alınarak aktif kullanıcının muhtemel değerlendirmeleri tahmin edilmeye çalışılır.

Komşuluk temelli algoritmalar olarak da adlandırılan bu yaklaşım tüm kullanıcı-ürün değerlendirme matrisini kullanarak öneri oluşturabilmektedir (Aggarwal, 2016: 29). Belirlenen hedef kullanıcı ile değerlendirme matrisi içerisinde ortak veriler dikkate alınarak “komşular” olarak adlandırılabilen kullanıcı grubunu tespit edilebilmesi için istatistiksel tekniklerin kullanımı ile oluşturulur. Ortak veriler aynı ürünler için benzer değerlendirmeler üzerinden hesaba katılır. Komşuların tespiti ile bir tahmin oluşturulabilmesi için farklı algoritmalar komşu tercihlerini birleştirilir. Bu teknikler “en yakın komşu” başlığı altında olarak da literatürde yer almakta ve sıklıkla kullanılmaktadır (Sarwar, Badrul M vd., 2002: 293).

Hafıza temelli işbirlikçi filtreleme modelleri literatürde iki farklı grupta incelenebilmektedir: kullanıcı temelli ve ürün temelli işbirlikçi filtreleme (Wang vd., 2006: 501). Kullanıcı temelli işbirlikçi filtreleme tekniklerinde hedef kullanıcı için veri seti içerisindeki diğer kullanıcıların değerlendirmeleri dikkate alınır. Kullanıcının etkileşim içerisinde bulunduğu kullanıcı grubu içerisindeki ağırlık ortalama değerleri dahil edilir. Ürün temelli işbirlikçi filtreleme tekniklerinde ise hedef olarak belirlenen ürüne benzer ürün gruplarının oluşturulması gerekmektedir. Bir kullanıcının hedef ürün hakkındaki değerlendirmesini tahmin edebilmek için kullanıcının oluşturulan ürün grubu içindeki değerlendirmelerinin ağırlıklı ortalamaları dikkate alınır.

Hafıza temelli yaklaşımda belirlenen hedef ile, kullanıcının benzer kullanıcı grupları arasında eşleştirilmesi amaçlanmaktadır. Hedef kullanıcı için tüm kullanıcılarla olan benzerlik değerlendirme yapısı ortaya çıkarılmalıdır. Bu nedenle bu ölçümün temel yapısı değerlendirme verileri üzerine kurulmuştur. Yapı, kullanıcıların sezgisel düşünceleri üzerine kurulu olduğu

için modelin başarılı bir şekilde oluşturulması, çeşitliliğin arttırılarak farklı varsayımlara açık olmasına bağlıdır.

Hafıza temelli yaklaşımlarda kullanılan makine öğrenmesi metotları kullanıcı veya ürün temelli bir yaklaşımı göz önüne alarak farklı uygulamalarda kullanılabilir. Araştırma konusu dikkate alınarak bu bölümde ürün temelli yaklaşımlarda kullanılan yöntemler üzerinde durulmuştur.

Tahmin Fonksiyonları

Tahmin fonksiyonu standart komşuluk temelli tahmin modellerinde farklı değerlendirme kriterlerinin ağırlıklandırılması ile elde edilmektedir. Kullanıcı v 'nin ürün j için merkezi ortalama değeri s_{uj} , hedef kullanıcı u 'nun kullanıcı v ile belirli benzerlik değeri sayısı ile ağırlıklandırılmasıdır (Aggarwal, 2016: 38).

$$Sim(u, v) = Pearson(u, v)^2$$

$Sim(u, v)$ hedef kullanıcı u 'nun her v kullanıcısı ile olan benzerliğini göstermektedir. Pearson korelasyon katsayısı olarak tespit edildiğinde α ile üssel olarak kuvvetlendirilebilir. $\alpha > 1$ olduğu durumlarda benzerliğin önem derecesi yükselir.

Tahmin fonksiyonlarının birçok varyasyonunda ham verilerin, orta değere göre merkezi ortalaması hesaplanması yerine z-skoru kullanımına literatürde sıklıkla karşılaşılmaktadır.

Z skoru hesaplamalarında kullanıcılar için sunulan değerlendirme ölçeği dışında bir değer elde edilmesi de mümkündür. Bu durum literatürde çeşitli tartışmalara yol açsa da tutarlı sonuçlar elde edilebilmektedir (Herlocker, J. L. vd., 1999).

Bahse konu modeller bir regresyon modeli olarak kabul edilecek oluşumlardır. Tahmin değerleri sürekli değişkenler olması ve kategorik değişken olarak kullanılması modeli diğer süreçlerden ayırmaktadır.

Uzun Kuyruk Problemi

Değerlendirme sistemlerinde hangi yöntem kullanılırsa kullanılsın sıkça karşılaşılan sorunlardan bir tanesi uzun kuyruk problemidir (Park, Y.-J. ve Tuzhilin, 2008). Sınırlı sayıda ürün hakkında belirli bir şekilde değerlendirme yapılabildiği ancak ürünlerin büyük çoğunluğunda kullanıcılar verileri ile oluşturulan değerlendirmelerin seyrek veya yetersiz olduğu gözlemlendiği durumlarda bu sorunla karşılaşılmaktadır. Bu gibi durumlarda karar vericilerin nasıl bir sonuç elde etmeyi planladıklarının belirliliği önem kazanmakta ve model ile elde edilmek istenen sonuca göre birden fazla model üzerinden en iyi sonucu bulmaya yönelik çalışmalar gerektirmektedir. Birçok e-ticaret sitesi öneri sistemleri için oluşturdukları kaynakları öncelikli olarak bu tür az ve standart model içerisinde sonuç elde edemedikleri

ürünlerin sisteme dahil edilmesi için ayırabilmektedirler (Yin vd., 2012: 897). Öneri sisteminden elde edilecek sonuçlara göre hareket etmeyi planlayan karar vericiler için bu tür ürünler genellikle yüksek karlılık veya erken dönem ürünleri içerebilmekte ve standart ürünlerden daha çok öneme sahip olduğu kabul edilmektedir.

Benzer durumda karar vericilerin tercihi doğrultusunda kuyruk dışı ürünler için kuyrukta bulunan ürünlerin sisteme benzer ürün statüsünde dahil edildiği durumlar mevcuttur. Ancak gerçek veri ile yapılan çalışmalarda her zaman yeterli düzeyde sonuç elde edilememiştir (Cremonesi vd., 2010: 42).

E-ticaret sistemlerinden elde edilen gerçek veriler incelendiği zaman, karar vericiler müşterilerine geniş bir ürün yelpazesi sunmak istemelerine rağmen, kullanıcıların genellikle kategorik olarak belirli ve sınırlı bir ürün yelpazesi içinde alışveriş yapmayı tercih ettiği gözlemlenmektedir. Bu durumda belli ürünler için değerlendirme geri bildirimi yoğun alınabilirken, birçok üründe değerlendirme verileri ile sağlıklı bir matris oluşturmak mümkün olmamaktadır. Benzer bir durum metin/içerik madenciliği ile oluşturulan modellerde de karşımıza çıkmaktadır. Belirli bir metin içerisinden kelime ağırlıklarının tespitinin yapılması esnasında bağlaçlar ve konuyla doğrudan ilgili olmayan kelimeler sağlıklı bir çıkarım elde etmeye engel olmaktadır.

Bu gibi durum Ters Doküman sıklığı (IDF-Inverse Document Frequency) yöntemiyle giderilmektedir (Manning vd., 2010: 117). Yöntem kullanıcı temelli modeller üzerinde uygulanmakta ürün j için yapılan değerlendirmeler m_j ile ifade edilirse model için bu ürün ağırlığı ile gösterilmektedir;

$$w_j = \left(\log \frac{m}{m_j} \right) \quad \forall u \in \{1, \dots, m\}$$

Örneğin pearson korelasyonunun kullanıldığı durumda aşağıdaki formülle ile değerlendirilir;

$$Pearson(u, v) = \frac{\sum_{k \in I_u \cap I_v} w_k \cdot (r_{uk} - \mu_u) \cdot (r_{vk} - \mu_v)}{\sqrt{\sum_{k \in I_u \cap I_v} w_k \cdot (r_{uk} - \mu_u)^2} \cdot \sqrt{\sum_{k \in I_u \cap I_v} w_k \cdot (r_{vk} - \mu_v)^2}}$$

Uzun kuyruk sorunu sadece işbirlikçi filtreleme modellerinde bulunmamakla beraber benzer çözümler öneri sistemlerinde içerik temelli modellerde sıklıkla görülmektedir.

Hafıza Temelli Komşuluk Modelleri ve Benzerlik Ölçümleri

Öneri sistemleri içerisinde ürünler veya kullanıcılar arası bir benzerlik üzerine modellemeler gerçekleştirilmektedir. Belirli bir düzen içerisindeki veri yapıları belirli bir özelliğe göre gruplandırılmak istendiğinde benzerlik ölçülerine başvurulmaktadır. Belirlenen

nesnelerin birbirine mesafesi benzerlik ölçüleri ile tanımlanabilirken bu benzerlik katsayılar üzerinden oluşturulmaktadır (Campbell, 1978: 103).

Öneri sistemlerinde veri yapılarına göre de farklı benzerlik ölçü yöntemlerine başvurulabilmektedir. Yaygın olarak kullanılan pearson, öklidyen ve kosinüs yanı sıra uygulama da yararlanılan jaccard uzaklık ölçüsüne de değinilmiştir.

Öklidyen Benzerliği

İki nesne arasındaki benzerliğin ölçümü için kullanılan en basit tekniklerden biridir. Bu hesaplama ile bir veri seti içerisindeki iki nesne/nokta arasındaki (x ve y) temel uzaklık özet olarak şu denklem ile hesaplanır (Danielsson, 1980: 228);

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

Bu hesaplamada x ve y ardışık veri noktaları olarak kabul edilir ve n veri setinin içerisindeki özellik sayısı olarak belirlenmiştir. Öklidyen uzaklık düz bir çizgisel uzunluğu mesafe olarak alır ve iki veya üç boyutlu bir uzayın çok boyutlu uzay ile olan mesafesinin ölçümü için kullanılır.

Kosinüs Benzerliği

Kosinüs benzerliği bilgi erişim sistemlerinde ve en çok da vektör benzerliği üzerinde kullanılan klasik bir ölçüm olarak kabul edilir (Gerard Salton ve McGill, 1983: 121). Kosinüs Benzerlik uzaklığı iç ürün uzayındaki iki vektörün arasındaki açının kosinüs değerinin ölçümü olarak tanımlanmaktadır. Kosinüs benzerlik denklemi iki vektör x ve y arasındaki ölçümü şu şekilde tanımlanmaktadır.

$$\text{kosinus}(\theta) = \frac{x^T \cdot y}{\|x\| \|y\|}$$

Kosinüs benzerliği ile elde edilen sonuç her zaman -1 ile +1 arasında yer almaktadır. Ürün veya kullanıcı, vektör olarak tanımlanırsa vektörler arası korelasyondan yola çıkarak vektörler arasındaki açının kosinüs değeri benzerlik ölçütü olarak kabul edilir. Kosinüs benzerliği özellikle metin madenciliği uygulamalarında kümeleme ve bilgi tarama modellerinde oldukça sık kullanılmaktadır.

Pearson Benzerliği

Pearson korelasyon katsayısı bu iki değişkenin kovaryans değerinin standart sapmalarına bölümü ile ortaya çıkarılmaktadır (Van Dongen ve Enright, 2012: 1). İki ürün veya

kullanıcı arasındaki benzerlik bu öğelerin sahip veya ilgili oldukları değişkenler arasındaki benzerliklere de işaret edebilir.

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

Pearson korelasyon katsayısı, +1 ile -1 arasında bir değer almakta, 1 değeri pozitif lineer bir korelasyonu, -1 değeri negatif lineer bir korelasyonu, 0 ise lineer olmayan bir korelasyonu göstermektedir.

Ampirik çalışmalar pearson katsayısının kullanıcı temelli öneri sistemlerinde diğer benzerlik hesaplamalarından daha iyi olduğu durumları ortaya çıkarmaktadır (Massa ve Avesani, 2004: 492).

Jaccard Benzerliği

Tanimoto katsayısı olarak da anılan Jaccard katsayısı elemanların birleşiminin kesişimlerine bölümü sonucu elde edilmektedir (Hamers, 1989: 315). Bu katsayı sınırlı bir örneklem içinde örneklem setlerinin birleşimi ile kesişimleri arasında belirlenen boyut ile ölçümü şeklinde gerçekleşir.

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

Jaccard uzaklığı ise veri seti içerisindeki benzersizliğin ölçümü ile hesaplanabilmektedir. Bu uzaklık, Jaccard katsayısının tersi olarak kabul edilerek 1 değerinden çıkarılarak elde edilen sonuçtur. Tespit edilen tüm ortak özelliklerin tersi bir değer ile ifade edilir.

$$j_\delta = 1 - J(A,B) = 1 - \frac{A \cap B}{A \cup B}$$

Bu denklem sonucunda, asimetrik ikili özellikler arasındaki benzerlik ölçümü elde edilebilmektedir. Bir nesnenin özellikleri ikili bir formatta incelenirken karar vericinin kolaylıkla sonuca erişmesinde uygulanabilecek bir uzaklık ölçüm mekanizması sunmaktadır. Jaccard benzerliği A ve B nesnelerinin ikisi için de ortak olan özelliklerinin yanı sıra A ve B'nin kendine has durumlarını da değerlendirme içinde kullanabilmektedir.

Çoğu vakada, Jaccard ölçümlerinin kosinüs ölçümlerine göre monoton bir yapıda olması beklenir. Ancak kosinüs iki vektör arasındaki metrik uzunluğu açılar vasıtasıyla ölçerken iki nesne setinin birleşim kümesine kıyasla kesişimlerinin göreceli büyüklüğüne odaklanmaktadır. Bunun yanında oluşan eş dizilim ya da asimetrik görünümlü matrisler üzerinden marj toplamaları üzerinden normalizasyon işlemlerinde farklı bir yapı kullanılmaktadır (Leydesdorff, 2008).

Ortalaması Alınmış Kullback-Leibler Uzaklığı

Ürün-ürün veya kullanıcı-kullanıcı ilişkilerinde iki olasılık dağılımı arasındaki uzaklığı ölçmek için kullanılan araçlardan bir tanesi de Kullback-Leibler Uzaklığı (KL Uzaklığı)'dır. Nesnelerin sistem içerisinde kullanabilecekleri özellikler arasından belirlenmiş olanların, diğer özellikler üzerindeki önemini Kullback-Leibler Uzaklığı ile ölçmek mümkündür (Lee, C.-H. vd., 2011: 1146). Bu ölçüm birçok açıdan geliştirilmesi ile birlikte nesnelerin özelliklerinin ağırlıklandırılması ve bu ağırlıklandırmanın uzaklık ölçümü formülünde dahil edilmesi önem arz etmektedir.

Göreceli entropi uzaklığı olarak da adlandırılan bu yöntem, dağılımlar arasındaki mesafeyi ölçtüğü için simetrik bir yapıya sahip olması beklenmemektedir (Huang, A., 2008: 9). P ve Q iki dağılım olarak kabul edildiğinde P 'den Q 'ya olan uzaklık şu şekilde formüle edilir;

$$D_{KL}(P||Q) = P_{\log} \left(\frac{P}{Q} \right)$$

Söz konusu dağılımlar arasındaki uzaklık ölçümleri pek çok alanda kullanımı mevcuttur. Görsel ve konuşma tanılamada iki akustik modelin benzerliğinin tespiti, dil biliminde kelime ayrımlarının ortaya çıkarılmasında, histogram görsel modellerin oluşturulmasında bu ölçümün kullanılabilirliği mümkündür (Hershey ve Olsen, 2007: 317). Ancak KL uzaklığı çoğu zaman analitik izlenebilirliği sağlayamamakta ve algoritma hesaplanabilirliğinde olumsuz sonuçlara ortaya çıkarabilmektedir. Bu nedenle özelleştirilmiş modeller üzerinde kullanımı daha uygundur.

Adamic Adar Uzaklığı

Adamic Adar uzaklığı sosyal ağlar üzerindeki bağlantıların tahmini için oluşturulan ve düğümler arası mesafelerin ölçümü için kullanılan metotlardan biridir (Adamic ve Adar, 2003). Ters logaritmik açı merkezliği toplamaları hesaplanarak iki komşu düğüm arasındaki uzaklığın ölçümü sağlanmaktadır.

$$AdamicAdar(i, j) = \sum_{k \in S_i \cap S_j} \frac{1}{\log(|S_k|)}$$

Formülde i ve j düğümlerinin arasındaki uzaklık ortak komşu kümesindeki düğümlerin ağırlıklandırılmış değerleri ile ifade edilmektedir. Komşu düğüm k 'nın ağırlık katsayısı $1/\log(|S_k|)$ ile ifade edilmiştir. Tüm düğüm eşleşmeleri için tutarlı bir model oluşturulduğu sürece logaritmik taban değeri konusunda bir sınırlandırma getirilmemektedir.

Katz Uzaklığı

Komşuluk temelli benzerlik ölçüm metotları nesne veya düğümler arasındaki benzerliğin belirlenmesinde keskin bir hesaplama sağlamasına rağmen komşuluk bağı kurulabilecek nesne sayısının azlığında etkili bir sonuç verememektedirler. Bu gibi durumlarda eşleştirilmiş farklı tahmin modelleri için kıyaslama ve ayırt etme özelliği konusunda zorluklarla karşılaşılması ve az sayıda ortak eşleşmenin uzun yollar (uzaklıklar) ile tanımlanan dolaylı bağlantıların ayırt edilememesi farklı bir ölçüm metodunun ortaya çıkarmıştır. Katz uzaklığı ile yürüme-temelli ölçümler arasında kabul edilen bağlantı, tahmin bağ gücünü ortaya çıkararak bu gibi durumlarda da tutarlı sonuçlar elde edinilmesi sağlamaktadır (Katz, 1953: 39). Bu metot şu şekilde formüle edilmiştir (Aggarwal, 2016: 325);

$$Katz(i, j) = \sum_{t=1}^{\infty} \beta^t \cdot n_{ij}^{(t)}$$

$n_{ij}^{(t)}$ i ve j düğümleri arasında oluşturulan adımların t uzunluğunu göstermektedir. Karar verici tarafından tanımlanan $\beta < 1$ değişkeni için hesaplama bu şekilde olacaktır. β değeri belirli bir adım mesafesinin dengeleme değeri olarak denklemde yer almaktadır. Bu değer çok küçük değerler için denklemin sonsuz toplamı olarak yakınsanacaktır. Eğer denkleme yönlendirilmemiş ağırlıklı simetrik yakınlık matrisi (A) eklenirse, $n \times n$ eşleştirmeli Katz katsayısı matrisi (K) aşağıdaki gibi formüle edilebilir;

$$K = \sum_{t=1}^{\infty} (\beta A)^t = (I - \beta A)^{-1} - I$$

A özvektörlerinin k 'ninci üssü olan A^k özvektörü β değeri A 'nın en büyük özvektörünün tersinden küçük bir değer olarak seçilmesi zorundadır. Bu işlem sonsuz toplamın yakınsaklığından emin olabilmek adına yapılmaktadır. Denklemin ağırlıklandırılmış sürümü de bir ağırlık matrisi eklenerek gerçekleştirilebilir. Katz katsayısı net sonuçlar yakalamakta oldukça başarılı işler çıkartabilmektedir. Bir düğümün diğer düğümlere göre belirlenen Katz katsayıları toplamı katz merkezi olarak tanımlanır. İnternet arama sistemlerinde kullanılan PageRank, DMOZ gibi yapıların temel çıkış noktası bu ölçüm aracıdır (Zaihrayeu vd., 2007:653).

Model Temelli İşbirlikçi Filtreleme Teknikleri

Model temelli işbirlikçi filtreleme modelleri öncelikli olarak kullanıcı değerlendirmelerinin modellenmesi ile ürün önerisi oluşturmaktadır. Bu kategoride yer alan algoritmalar olasılıksal yaklaşım ile kullanıcının diğer ürünler üzerindeki değerlendirmeleri ile beklenen bir kullanıcı tahmin değeri öngörmeyi hedeflemektedir.

Model temelli modellerde olay odaklı bir anlayışla, oluşturulan model öncelikli olarak bir çıktı üretmek yerine veri üzerinde bir ön işleme süreci yürütür. Bu sayede modelin adapte edilebilirliği ve güvenilirliği konusunda güvenilir bir yol oluşturulur (Aggarwal, 2016: 71).

Öneri sistemleri için çıktı elde edilmeden önce uygulanan yöntemler gözetimli veya gözetimsiz algoritmalarından oluşabilmektedir. Bu nedenle model oluşturma süreci bu tür model oluşturma tekniğine giren yöntemlerde farklı bir sürecin işletilmesini gerektirir (Amershi ve Conati, 2009: 18). Öneri sistemi içerisinde Model temelli bir algoritmada kullanıcı değerlendirmeleri üzerine bir model oluşturma amacıyla hareket edilir. Bu algoritmalar olasılıksal yaklaşım ve kullanıcının beklenen tahmin değeri üzerine hesaplamaları dikkate alır (Sarwar, Badrul Munir vd., 2001: 287).

Hafıza temelli yaklaşımlarda benimsenen değerlendirme matrisi yapıları öneri sisteminde tanımlanan probleme ilişkin bir veri seti içerisinde bağımsız ve bağımlı değişken tanımları ile elde edilir. Model temelli yaklaşımlarda da benzer bir mantık bulunmasına rağmen, hafıza temelli modellerden farklı olarak oluşturulan matris içerisindeki eksik verilerin minimum düzeye indirilmesi de amaçlanmaktadır. Hafıza temelli model için girdi kabul edilen bir veri setinde bağımlı ve bağımsız değişkenler keskin bir şekilde belirlenirken, model temelli bir yapıda tahmin modeli oluşturulmadan önce elde edilen veri setinde bu ayrım net değildir. Her sütun bağımlı veya bağımsız değişken olma potansiyeline sahiptir.

Bu durum model oluşturma ve test etme süreçlerinde kullanılan eğitim ve test veri seti ayrımını da etkilemektedir. Model temelli bir uygulamada veri setinin tamamen farklılaştırılması net bir ayrımın ortaya koyulmasını önlemektedir. Sütunlarda olduğu gibi satırlar da ayrımlar ortadan kaldırılmış olabilmektedir.

Model temelli yaklaşımların hafıza temelli yaklaşımlara kıyasla bazı avantajları mevcuttur (Sarwar, Badrul Munir vd., 2001: 288);

Seyreklik: Pratik hayatta öneri sistemleri genelde büyük ve dağınık bir veri seti üzerinden işlemlerini gerçekleştirir. Bu sistemler anlık olarak öneri motoru üzerinden aktif kullanıcıya bir öneri çıktısı sunmak zorunda ve bunu kullanıcı için sıfır bekleme süresi içinde gerçekleştirmelidir. Ancak kullanıcı ürün ilişkisi üzerinden elde edilen veri setleri tam ve kusursuz bir yapıya sahip olamazlar. Böyle bir sistemin oluşturulabilmesi için kullanıcının sistemdeki tüm ürünler hakkında bir değerlendirmede bulunmuş olması gerekmektedir. Bu doğrultuda veri seti içinde değerlendirmede bulunulamayan ürünler için boşluk oluşacak ve seyreklik gözlemlenecektir.

Model temelli öneri sistemlerinde bu tür sorunlarla daha az karşılaşılmaktadır. Bu tür modeller öneri modeli oluşturmadan önce bir ön işlem neticesinde veri setini dönüştürürler.

Böylece farklı bir veri seti yapısı ile öneri modeli oluşturulmasını sağlarlar. Seyrekliği yoğun olan bir veri setinin uygulanması hem maliyet açısından hem de öneri çıktıtlılarının performansı açısından aşılması gereken bir engeldir (Suryavanshi vd., 2005: 618).

Ölçeklenebilirlik: Hafıza temelli bir öneri modelinde karar vericiler tarafından oluşturulmuş veri seti çoğu zaman sistem içerisindeki tüm veri setini kapsayıcı şekilde olmaktadır. Bu durumda büyük veri setlerinin analizlerde kullanılmasında zorluklar yaşanmasına sebep olabilir. hafıza temelli yaklaşımların aksine model temelli yaklaşımlar eğitim veri setlerindeki farklı kullanıcıları değerlendirme desenlerine göre daha küçük sınıflar içerisinde gruplandırır (Han, P. vd., 2004: 203). Hafıza temelli modeller tüm veri seti içerisinde nesnelerin birbiri ile etkileşimi üzerine öneri çıktısı oluştururlar. Büyük bir veri seti içerisinde nesneler arası etkileşimin tespit edilmesi yüksek işlem gücü ve saklama kapasiteleri ihtiyacını oluşturacaktır. Model temelli yaklaşımlarda veri seti bilgi kaybı olmaksızın öneri sistemi modellemesi için hazırlanabilmektedir. Böylece kullanılabilecek veri setleri ölçeklenmiş bir şekilde analizlere dahil edilir.

Model temelli yaklaşımda bu özellikler gerçek zamanlı uygulamalarda avantaj sağlayabilmektedir. Eğitim ve tahmin hızı gerçekleştirilen ön işleme sebebiyle yüksek olmaktan aşırı uyum (overfitting) sorunlarıyla daha az karşılaşılmaktadır. Aşırı uyum veri içerisindeki rasgele oluşan yapay olgulardan tahmin değerlerinin fazla etkilenmesi durumudur (Aggarwal, 2016: 73). Bu tür durumlarda oluşan modellerde, ihtiyaç duyulandan daha fazla nesneye yer verilmesi ya da olması gerekenden daha fazla karmaşık yaklaşımlarla modelin kurgulanması neticesinde sınırlılık ilkesinin aşılması durumu gözlenir (Hawkins, 2004: 1).

Model Temelli İşbirlikçi Filtreleme ve Karar Ağaçları

Karar ağaçları tahmin ve sınıflandırma teknikleri arasında en yaygın araçlardan biridir. (Hastie vd., 2009: 266). Karar ağaçları girdileri planlayarak girdilerin öznitelikleri dikkate alınarak tahmin modelleri oluşturabilmektedir. Ağaçlardaki her iç düğüm bir özniteliğe denk gelmekte ve üst düğümden alt düğüme oluşan yay özniteliklerin değerler kümesinin olası değerini simgelemektedir. Nesnenin niteliği oluşturulan alt düğümlere ve kök yapılarına atanır. Girdi veri seti değerlere bölünerek nesneden bir özellik parçasının bilgilerini taşır. Bu süreç üst ve alt düğüm noktaları oluşturacak şekilde tekrarlayarak devam eder.

Karar ağaçlarında bağımsız değişkenlerde ayırım kriteri olarak belirlenen hiyerarşik kararlarla bir bölümlenme işlemi gerçekleştirilir. Çoğu durumda oluşturulan ayırma fonksiyonları tek değişkenlidir. Tek değişkenli olması içsel düğümün sadece tek bir öznitelik dikkate alınarak bölünebilmesini ifade etmektedir. Sonuç olarak düğüm işleminin başlaması

bölünebilecek en iyi niteliklerin tespiti ile ortaya çıkmaktadır (Rokach ve Maimon, 2005: 168). Tek değişkenlik kriterine göre bölünmenin oluşturulabilmesi için öznitelik bir kez kullanılabilir.

Bölme işleminin kalitesi gini indeksi ile belirlenebilmektedir. Gini indeksi olasılık dağılımı ile hedef öznitelik değerleri arasındaki farkı ölçen katışıklık temelli bir kriterdir. Bir S düğümünde r sınıfları arasında veri setinde p fraksiyonları için aşağıdaki formülle hesaplanır (Aggarwal, 2016: 74).

$$Gini(S) = 1 - \sum_{i=1}^r p_i^2$$

Belirlenen katsayı ağaç içerisinde belirli bir seviyedeki uygun özneliğin belirlenmesi için kullanılmaktadır. En düşük gini indeksine sahip öznitelik bölünme için seçilir. Yukarıdan aşağıya doğru her düğüm belirli bir sınıfa ait veriyi kapsayınca kadar devam eder.

Karar ağaçlarının öneri sistemlerinde kullanımına literatürde sıklıkla karşılaşılmaktadır (Cho vd., 2002; Bouza vd., 2008; Gershman vd., 2010: 170) . Benzer çalışmalarla birlikte işbirlikçi filtreleme yönteminde karar ağaçlarının kullanımı tahmin edilen girdiler ile gözlemlenen girdiler arasındaki fark özellikle model temelli çalışmalarda net değildir. Bu durum bir ağaç yapısı oluşturmada problemlere neden olabilir. Yine veri seti içerisindeki bağımlı ve bağımsız değişken ayrımı her zaman net çizgilerle ayrılmamaktadır. Bu durum genellikle bağımlı değişken için öznitelikler sabitlenerek farklı bir ağaç yapısı, bağımsız değişkenler içinse farklı bir yapı oluşturmaları ile çözülmektedir. $m \times n$ değerlendirme matrisi içerisinde n adet ağaç oluşturulması gerekmektedir. Hedef ürün için oluşturulan ağaç yapısı üzerinden tahmin çıktıları oluşturularak sonuçlar kıyaslanabilmektedir.

Tahmin modeli oluştururken karar ağaçların kullanımında en önemli zayıflıklardan biri çok büyük sayılarda ağaç yapısına ihtiyaç duyulabilmesidir (Gershman vd., 2010: 171). Model aynı anda sadece bir ürün için beklenen değerlendirme skorunu hesaplayabilir. Kullanıcı için öneri oluştururken her bir ürün için değerlendirme önerisi oluştururken kök üzerinden farklı geçişlerin oluşturulması gerekmektedir. Kullanıcılar tarafından oluşturulan tüm değerlendirmeler yapıldıktan sonra öneri sistemi bir tahmin modeli oluşturabilir. Karar ağacı yapılarında Ürün sayısı arttıkça öneri sistemlerinde ölçeklenebilirlik problemi baş gösterecektir.

Model Temelli İşbirlikçi Filtreleme ve Naive Bayes Sınıflandırıcısı

Naive Bayes sınıflandırıcıları, Bayes teoremi üzerine kurgulanmış ve güçlü bağımsız varsayımlar oluşturan bir teoremdir (Ghazanfar ve Prugel-Bennett, 2010: 3). Naive Bayes

sınıflandırıcıları model oluşturmada basit teknikleri kullanarak ilerlemektedir. Bunlar; sınıf etiketlerini problem örneklerine atar, bu etiketlerin sınırlı bir veri seti içinde olması durumunda özellik değerlerini vektörlerle gösterir.

Naive bayes sınıflandırıcıların temel avantajları, gürültülü noktaları ve alakasız özellikleri sınırlayabilmesi ve olasılık kestirim hesaplamalarında eksik değerleri göz ardı edebilmesidir. Bayesyen sınıflandırıcıları öneri sistemlerinde popülaritesi yüksek bir öğrenme tekniği olarak kullanılmaktadır (Isinkaye vd., 2015). İçerik temelli öneri sistemlerinde kullanımı gözlemlenmekle birlikte model temelli işbirlikçi filtreleme yöntemlerinde de karşılaşılabilmektedir (Miyahara ve Pazzani, 2000: 679; Ghani ve Fano, 2002: 27).

Bu tür bir model genelleyici bir yaklaşım sergilerken sınıflandırma amaçlı kullanılmaktadır. Öneri sistemi gibi yapılarda veri seti içindeki ürünler özellik niteliği taşıırken kullanıcılar sınıf örneği olarak değerlendirilerek eksik değerler konusunda sınıflandırma modelinden yararlanılabilmektedir (Aggarwal, 2016: 82). Bu doğrultuda Tanımlanmamış eksik verilerle uyumlu bir şekilde çalışabilecek ve özellik olarak nitelendirilen ürünlerin hedef sınıf üzerinde işbirlikçi filtreleme modeli ile tahmin çıktıları oluşturulmasında kullanılabilmektedir (Robles vd., 2003: 168).

Öneri sistemlerinde ürün değerlendirmesinin şartlı olasılık hesabının diğer ürün değerlendirmelerine göre yapıldığı modellemeler oluşur. Bu tür modeller ürün temelli yaklaşım bazında değerlendirilirken geleneksel sınıflandırma metodolojisinde doğrudan (açık) adaptasyon olarak nitelendirilir. Bu durumun tek istinası hedef ürün işbirlikçi filtrelemede değişkenlik gösterebilmektedir. Ancak bu tür sınıflandırma yaklaşımlarında sabitlenmiş bir boyut oluşturulmaktadır. Benzer bir yaklaşım kullanıcı temel için de geçerli olmakla birlikte iki yaklaşımının hibrit türevleri de işbirlikçi filtreleme modellerinde kullanılabilmektedir.

Model Temelli İşbirlikçi Filtreleme ve Boyut İndirgeme Yöntemleri

Verilerin asgari veri kaybıyla öznitelik çıkarımı, kompakt kodlama veya hesaplama etkinliği için boyutsal indirgeme tekniği kullanılmaktadır. Verinin "iyi" bir temsiliyete bürünerek ileriki aşamalarda kullanımı için dönüşümün sağlanması, öznitelik değişkenleri arasında sabitlerin belirlenerek gereksiz çoğunluğun atılmasıdır (DeMers ve Cottrell, 1993: 580). $m \times n$ boyutunda X matrisinin olduğunu ve t adet veri vektörü x_i ($i \in \{1, 2, \dots, n\}$) ile D boyutluluğunu oluşturduğunu varsayalım. Veri seti X matrisi içindeki noktaların D boyutlu uzay içerisinde d içsel boyutları ($d < D$) barındırdığı düşünülür. Bu noktalar boyut içerisinde yakın ve dağıtılmış vaziyette bulunduğu kabul edilir. Boyutsal indirgeme tekniği, D boyutluluğunda olan X matrisini yeni bir veri seti olarak d boyutunda Y veri setine dönüştürür. Bu dönüşüm

sırasında verinin geometriğini korumayı amaçlar. Genel yapıda ne dağıtılan verilerin ne de içsel boyutun geometrik yapısı bilinmemektedir. Bu nedenle boyutsal indirgeme 'kötü konumlanmış' bir probleme dönüşmemesi için sadece belli özelliklere sahip verilerde geçerlidir (Van Der Maaten vd., 2009: 68).

Boyut indirgeme, komşuluk temelli metotlarda modelin hem kalitesinin artırılması hem de etkinliğinin sağlanması açısından kullanılmaktadır. İkili benzerliklerin seyrek değerlendirme matrislerinde hesaplanması oldukça güç olmasıyla birlikte boyut indirgeme yöntemi örtülü/gizli faktörlere dayanarak yoğun düşük boyutlu gösterimleri sağlayabilmektedir. Bu nedenle modeller “örtülü faktör modeli” olarak da isimlendirilmektedir. Bir modele içerisinde iki kullanıcı ortak değerlendirme kriteri çok az sayıda olsa bile, düşük boyutsal örüntülü vektörler arasındaki uzaklıklar hesaplanabilmektedir. Bununla birlikte eşleştirilmiş grupların belirlenmesinde de düşük boyutu örtülü vektörlerin kullanımı daha etkin bir yol olarak kabul edilir.

Öneri sistemlerinde iki farklı yönden örtülü faktör modeli uygulanabilir (Aggarwal, 2016: 47):

- Verilerin indirgenmiş gösterimi satır veya sütun üzerinden oluşturulabilmektedir. Bir başka deyişle indirgenmiş gösterim kullanıcı boyutluluğu ile ürün boyutluluğunu sıkıştırarak örtülü faktörleri ortaya çıkarır. Bu indirgenmiş boyut komşuluk-temelli modellerde rastlanılan seyreklik sorununun giderilmesi için kullanılan yollardan biridir. Hangi boyutun sıkıştırılarak örtülü faktörü elde edildiğine bağlı olarak kullanıcı veya ürün temelli modellere dahil edilir.
- Örtülü faktör gösteriminde satır veya sütun verileri eş zamanlı olarak sınırlandırılır. Olayın örtülü faktör satırlarını tüm değerlendirme matrisinin tek bir seferde tekrar oluşturulması için kullanılır. Bu durumda komşuluk-temelli modelden elde edilen matrisin yerine geçer. Örtülü faktörün oluşumunda eş zamanlı uygulamalar komşuluk temelli metotlarda sıklıkla kullanılır.

Model temelli işbirlikçi filtreleme modellerinde temel hedef, $m \times n$ değerlendirme matrisinin (R) düşük boyutsal alana dönüştürülmesi temel bileşen metodunun da içinde yer aldığı birtakım yöntemler vasıtasıyla gerçekleştirilebilir (Sarwar, B. vd., 2000: 6). Oluşturulan matris (R') $d \ll n$ olduğu ve $m \times d$ boyutuna indirgeneceği kabul edilir. Böylelikle her n boyutlu değerlendirme vektörü indirgenmiş boyutsal alana denk gelecektir. Standart değerlendirme vektörünün aksine her d boyutu tamamen belirlenerek özelleştirilir. Örtülü gösterimde indirgenmiş boyutta benzerlik hesapları daha verimli sonuçlar ortaya koyabilmektedir.

Temel Bileşen Analizi

Temel Bileşen Analizinde (TBA – Principal Component Analysis (PCA)) bir veri seti içerisindeki gözlemleri analiz ederek birbiriyle bağlantılı niceliksel bağımlı değişkenleri açıklar (Abdi ve Williams, 2010: 433). Veri seti içerisinde önemli bilgilerin ortaya çıkarımı amaçlanırken, temel bileşen adı verilen ortogonal değişkenlerin sunumunu hedefler. Bu işlemi gerçekleştirirken bir düzlem üzerinde gözlemlerle değişkenler arasındaki benzerlik desenlerini haritalandırır.

TBA hedef olarak kendine iki ana gösterge belirlenmiştir (Abdi ve Williams, 2010: 433); (i) veri setinden en önemli verilerin çıkarımı, (ii) önemli bilgilerin kaybı yaşanmadan veri setini sıkıştırmak. (iii) veri setinin tanımını ve açıklamasını basitleştirmek (iv) elde edilen gözlemlerin ve değişkenlerin yapısını analiz etmek. Bu hedefleri gerçekleştirebilmek için orijinal değişkenlerden elde edilen lineer kombinasyonlarla yeni değişkenler olan temel bileşenleri elde eder. İlk elde edilecek muhtemel en geniş varyansı elde etmek için gereklidir. Daha sonraki ise ilk hesaplama dikey olan sabit değerler ile hesaplanır. Böylece muhtemel her en geniş hareketlilik değeri elde edilir. Gözlemler için elde edilen yeni değişkenler faktör skoru olarak adlandırılır ve faktörler temel bileşen üzerinde gözlemlerin projeksiyonu olarak geometriksel yorumlanır.

TBA ile gerçekleştirilen öneri modellerinde sapma problemleri ile karşılaşabilmektedir (Sarwar, B. vd., 2000: 3). R_f matrisi, elde bulunan orijinal R matrisindeki sütun veya satır değerlerinin farklı ortalama değerleri ile doldurulması sonucu oluşmaktadır. Bu durum tarafsızlığın bozulmasına neden olabilmekte, objektif kriterlerin göz ardı edilmesine yol açabilmektedir. Örneğin bir öneri sisteminde TBA kullanılarak boyut indirgeme işlemi yapılması planlanmakta ve bu doğrultuda değerlendirme matrisini kovaryans metrise dönüştürülmesi hedeflenmektedir. Eksik değerler sütun ortalamaları dikkate alınarak tamamlanmıştır. Bu durumda kullanıcı ve ürünler için aşağıdaki tablonun oluşturulduğu varsayılmıştır

Tekil Değer Ayırıştırması (TDA)

TDA gizli anlamsal indeksleme tekniğinin bir parçası olarak bilgi edinim metotları arasında eş ve çok anlamlılık çözümlemesi için geniş bir kullanıma sahiptir (Vozalis ve Margaritis, 2007: 3017). Tekil değer ayrıştırması (Singular value decomposition) ile hesaplama, öncelikli olarak R matrisindeki $m \times n$ boyutlarındaki risk girdilerinin doldurulmasından başlanarak oluşturulur. Eksik girdiler, girdinin bulunduğu alanda yer alan değerlendirme puanlarının ortalama değeri oluşturularak tamamlanır. Bu çözüme alternatif olarak sütun

başında ortalamasının alınması da eksik girdinin tamamlanabilmesi sağlayan yollardan biridir. Sonuç matrisi R_f olarak ifade edildiği varsayıldığında hesaplamaya esas olan nesneler arasında benzerlik matrisi $n \times n$ olarak kabul edilir. Bu matris pozitif ve yarı kesindir. R_f matrisindeki baskını temel vektörün tanımlayabilmek için benzerlik matrisi aşağıdaki gibi olacaktır. TDA bir matris çarpanlara ayırma tekniği olmakla birlikte $m \times n$ boyutundaki bir R matrisini şu şekilde çözümler;

$$R = UXSV^t$$

U ve V ortogonal matrisler olmak üzere $m \times m$ ve $n \times n$ boyutlarına sahiptir. S Tekil matris ise diagonal girdileri negatif olmayan $m \times n$ boyutundadır. S matrisinin ilk r diyagonal girdileri $S(s_1, s_2, \dots, s_r)$ ve $s_1 > 0, s_1 \geq s_2 \geq \dots \geq s_r$ özelliklerine sahiptir. Buna göre U matrisinin ilk r sütunları R matrisinin sol tekil vektörleri kabul edilmektedir. V matrisinin ilk r sütunları ise sağ tekil vektörleridir. Sadece sıfır olmayan tekil değerlere bakıldığında U , S ve V için boyutlar $m \times r$, $r \times r$ ve $r \times n$ olarak görülecektir.

TDA en düşük R orijinal matrisin en düşük dereceli gösterimini elde ederek öneri sistemlerinde olduğu gibi yaygın bir kullanım sağlayabilmektedir. $k \leq r$ tekil değer denklemini koruyarak S içindeki girdilerin sıralamasını gerçekleştirmekte k en büyük tekil değerlerini oluşturabilmektedir. Bu işlemde veri boyutunun indirgenmesi düşürülmekte ve orijinal matris içindeki önemi veri bileşenlerini korumaya çalışılmaktadır (Gong vd., 2009: 770). Sonuç olarak U ve V matrislerine de S matrisi ile denk indirgeme işlemi uygulanmaktadır.

$$R_k = U_k x S_k x V_k^T$$

Yukarıdaki denklemden görüleceği üzere S_k diagonal matrisi oluşurken U ve V matrisleri de aynı şekilde $r-k$ sütunlarını ortadan kaldırarak boyut indirgeme işlemlerine tabi tutulmuş U_k ve V_k gösterimleri ile temsil edilmiştir.

Söz konusu metodoloji değerlendirme matrisleri için de TDA ile uygulanabilirliği ortaya konulmuştur. Jameson (2004: 48) ile Aggarwal and Parthasarathy (2001: 227) gibi çalışmalarda farklı kullanım alanlarında temel bileşen analizi ile birlikte birleştirilmiş boyutsal indirgeme tekniği olarak kullanımı gözlemlenmekte ve iyi sonuçlar elde edilebilmektedir.

Negatif Olmayan Matris Ayırışması

Negatif olmayan matris ayrışması (NMF veya NNMF) olarak adlandırılan bu yöntem negatif olmayan vektörler barındıran veri setlerinin lineer boyut indirgeme teknikleri ile karakteristik özelliklerinin ortaya çıkarılması işlemidir (Lee, D. D. ve Seung, 1999: 788). Özellik çıkarımı işlemi için NMF yöntemi ile gürültü verileri temizlerken verilerin sıkıştırılmasını gerçekleştirmektedir. Verilerden elde edilen öncül bilgiler performans artırımı

sağlanmakta ve seyreklik, pürüzsüzlük ve ortogonallik durumlarını ele almaktadır (Hoyer, P. O., 2004: 1452).

NMF makine öğrenme metotları içerisinde çok sayıda uygulama alanında kullanılmaktadır. NMF çok değişkenli veri setlerinin kör kaynak ayrışması tekniği ile anlamlı negatif olmayan faktörlere ayrılmasını sağlar (Bobadilla vd., 2017: 3549).

Bir negatif olmayan V matrisi içerisinde NMF sayesinde uygun ayrıştırma boyutlarını $V \approx WH$ ile ortaya çıkarmaktadır. W ve H negatif olmayan faktörlerdir. negatif olmayan sabitler, matris gösterimini yalın halde eklenebilir (çıkarmaya izin vermeden) şekle sokarak, bağımsız bileşen analizi veya PCA metoduna göre pozitif değerler üzerinde daha iyi çıkarımlar elde edebilmektedir (Hyvärinen ve Oja, 2000: 411).

NMF'nin en kullanışlı özelliklerinden biri tutarlı ama seyrek bir gösterim şekli sunmasıdır. Bu şekilde bir gösterim az sayıda aktif bileşen kodlanması yorumlanmasını daha kolay ve gerçekleştirilebilir kılmaktadır. Bu işlemi gerçekleştirirken veri seti içerisindeki önemli bilgileri elimine etmeden kullanıcıya sunabilmesi önemli bir avantajdır.

Veri seti içerisindeki bilgilerin önemli bileşenlerini tespitinde ne kadarlık bir seyrekliğe izin verileceği konusunda NMF kullanıcıya esneklik sağlar. Böylece gösterim hususunda daha fazla kontrol elde edilebilmesi sağlanabilmektedir (Hoyer, P. O. J., 2004: 1457).

V matrisinin $m \times n$ boyutlarında olduğu varsayıldığında, negatif olmayan $m \times r$ boyutunda W matrisi ve $n \times r$ boyutunda H matrisi, $V \approx WH$ yapısına göre oluşmaktadır. r değeri genellikle $(m + n)r < mn$ koşulunu sağlaması kriteriyle seçilerek WH çıktısı V matrisinin sıkıştırılmış bir hali olarak gösterilir (Tsuge vd., 2001: 961). Bu denklem sütunlar dikkate alarak yazıldığı durumda $v \approx Wh$, her v veri vektörü W matrisinin sütunlarının lineer kombinasyonlarının h bileşenleri ile ağırlıklandırılarak yakınlştırılması ile elde edilen değeri temsil eder. W bu nedenle V matrisindeki verinin lineer yakınlştırılmasının optimize edilmiş hali olarak kabul edilmektedir. Göreceli olarak sadece temel vektörlerin temsil yetkisi olduğu düşünüldüğünde, bu vektörlerin doğru ve tutarlı bir temsili gerçekleştirebilmesi için elimine edilen diğer vektörlerdeki gizli ve örüntülü yapıları analiz etmesi beklenen bir durumdur. NMF bu temsiliyeti iyi sağladığı durumlarda SVD gibi diğer modellerden iyi sonuçlar elde edebilmektedir (Boutsidis ve Gallopoulos, 2008: 1350).

Öneri sistemlerinde negatif olmayan matris ayrıştırması negatif değer bulundurmeyen herhangi bir değerlendirme matrisinde kullanılabilmektedir. Bu yöntemin öneri sistemlerinde sağladığı avantajları genellikle, kullanıcının puanlama yöntemi yerine beğeni, satın alma davranışı gibi tek bir davranışını gösterdiği yapılar üzerinden elde edilen

değerlendirme matrislerinde görülebilmektedir. Bu tür matrisler birli veya ikili veri yapısı içinde değerlendirilmekte negatif olmayan girdiler aktivite sıklığına göre yapılandırılmaktadır.

Negatif olmayan ayırıştırma yönteminin önemli avantajlarından biri olan kolay ve detaylı yorumlanabilirlik durumu sayesinde matris yapısındaki önemli bileşenlerin ayrımının yapılabilmesidir. Bu durum öneri sistemlerinde de avantaj sağlamaktadır.

Model Temelli Komşuluk Modelleri ve Birliktelik Kuralları

Birliktelik kuralları Hajek ve arkadaşları (1966: 293) tarafından bağımsız olarak ortaya atılan bir istatistiksel analizdir. Bu analizler makine öğrenme metotlarında kullanımı While (1977: 415) tarafından gerçekleştirilmeye başlamış, temel metotlardan karmaşık ve özelleştirilmiş durumlara kadar geniş bir çapta öneri sistemlerin içine girmiştir.

Birliktelik kuralları, satın alma değerleri matrisinde olduğu gibi benzer yapılarda nesnelerin birlikte gözlemlenebilme durumlarının ortaya çıkarılabildiği bir metottur. Eğer sınırlandırılan bir veri havuzu içerisinde ürün B'ye, ürün A ile aynı sıklıkla satın alma eylemine rastlanması durumunda kural $A \Rightarrow B$ olarak belirtilir. İki ölçüm için güven desteği kural keskinliğini ve kullanışlılığı ölçmek için kullanılır (Kim ve Kim, 2003: 524).

Birliktelik kuralları öneri sistemlerinde sıklıkla kullanılmakla beraber özellikle birli ve ikili veri yapılarının temel alındığı öneri sistemlerinde sıklıkla kullanılmaktadır. İşbirlikçi filtreleme modellerinde birliktelik kurallarının uygulamasındaki ilk adım minimum destek ve güven noktalarının tespitidir. Bu değerler analiz öncesi belirlenerek tutarlılığı en yüksek seviyede tutacak değerlere göre seçilir. Elde edilen sonuçlar doğrultusunda öneri sistemlerinde kullanılan top-N öneri algoritması ile öneri çıktıları oluşturulabilmektedir (Demiriz, 2001: 1).

Literatürde IS, SETM, Apriori gibi çeşitli yöntemlerle uygulanan birliktelik kuralı, bu araştırmada kullanılan yöntemler nedeniyle bu bölümde sadece FpGrowth ve PSO anlatımlarına yer verilmiştir.

Fp-Growth

Birliktelik kuralları veri madenciliğinde önemli bir kullanım alanına sahip olmakla birlikte, büyük veri setleri içerisinde isabetli bağlantılar ve korelasyonların tespitinde sıklıkla kullanılmaktadır. Sık tekrarlayan nesne setlerinin tespiti, birliktelik kuralları uygulamaları içerisinde anahtar yöntemlerin başında gelmektedir. En ünlü algoritmalarından biri olan ve Agrawal ve arkadaşları (1994: 490) tarafından ortaya konan Apriori algoritması ile sık tekrarlanan veriler arasındaki düzenin ortaya çıkarılmasında büyük fayda sağlanmıştır. Apriori algoritması veri seti üzerinde devamlı bir tarama gerçekleştirmekte ve veride nesne tekrarlanmalarının tespiti için tekrarlanmayan nesneleri veri setinden ayırmaktadır. Apriori bu

işlemleri gerçekleştirirken her defasında veri setini baştan taramakta, büyük veri setleri için verimsiz ve sistem kaynaklarını tüketerek yavaşlayan bir yapıya dönüşmektedir.

Birliklilik kuralları içerisinde en hızlı ve sıklıkla kullanılan yöntemlerden biri Fp-growth algoritmasıdır (Li vd., 2008: 107). Bu algoritma belirli bir işlevsel veri setinin önek ağacı oluşturarak tüm işlem veri setinin değerlendirilmesi yerine algoritmaya dahil edilecek bölümleri belirli kısımlar halinde analize dahil edilmesini sağlar. Bu durum diğer birliklilik kuralı algoritmalarından olan Apriori ve Eclat ile benzerlik göstermektedir. Bu ağaç literatürde FP-tree adıyla anılmaktadır. FP-growth algoritmasındaki temel prensip tekrarlamalı elimine aşamaları ile açıklanabilir. Verilerin algoritmaya dahil edilmeden önce veri hazırlık sürecinde bireysel olarak tekrarlama göstermeyen nesneler veri setinden çıkarılır. Örneğin kullanıcı tarafından belirli sayıda kriter konulan bir veri seti içerisinde bu koşulu sağlamayan veriler algoritmaya dahil edilmez (Han, J. vd., 2000: 2).

Ön işleme sürecinde her işlem kaydı belirli bir sıraya göre yapılması sağlanır. Böylelikle veri seti içerisindeki sıklıkları dikkate alınarak azalan bir sıraya göre konumları yeniden belirlenir. Algoritmanın bu özelliği diğer algoritmalarındaki rastgelelik durumuna göre daha verimli bir işleyiş sağlamaktadır (Hong vd., 2006: 2167). İşlem kayıtlarının artan sıklıkla gerçekleştirilmesi ise performansı aşağı yönlü bir seyre yöneltmekte ve sağlıklı sonuçlar elde edilememektedir. Tekrarlanan süreçte, tekrar derinliği k olarak kabul edilirse; işlem veri setindeki k 'ncı nesne veri setini belirli bölümlere ayırmada yardımcı olacaktır.

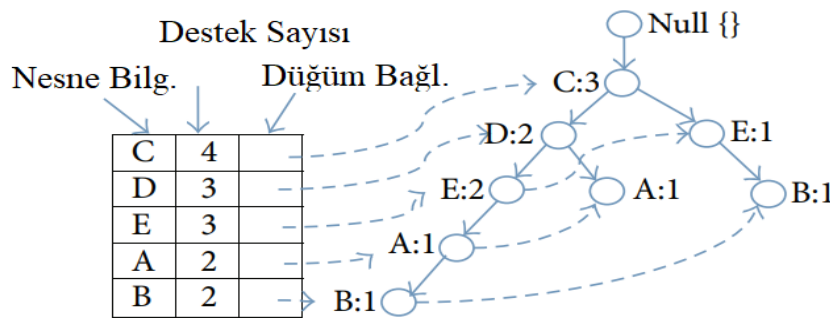
FP-tree uygulaması, hafızadaki işlem kaydı veri setinin gösterimini numerik bir dizin ile gösterimini sağlayabilmektedir. Sözlüksel (leksikografik) şekilde gösterim tekrarlı bir prosedüre uygun zemin hazırlamakta ve işlemsel hız konusunda daha başarılı bir sürecin oluşmasına yardımcı olmaktadır (Tanbeer vd., 2008: 1064). Oluşturulan her bölüm nesnenin FP-tree içerisindeki uyumuna göre oluşturularak etiketlenmek ve tekrarlanarak tüm bölümlerin kendi içlerinde alt bölümler oluşturulması yeni düğüm katmanların ortaya çıkarılması sağlanmaktadır.

FP-tree ağacının elde edinimi ön işleme sürecinden geçirilen veri seti üzerinden gerçekleştirilmektedir. FP-tree tüm işlem kayıtları için bir önek ağacı olarak kabul edilir. Her yol bir işlem kayıt setini paylaşan önekleri, her düğüm ise bir nesneyi simgelemektedir (Kumar ve Rukmani, 2010: 400). Bunun yanında tüm düğümler bir liste içerisindeki aynı ürünü göstermekte, böylece belirlenmiş bir nesneye ait tüm işlem kayıtları kolay bir şekilde listelerin ters çevrilmesiyle sayılabilmektedir.

FP-growth algoritmasının temel görevi öngörülen veri seti için FP-tree hesaplamalarının gerçekleştirilmesidir. Bu hesaplamalar hem belirlenmiş bir nesnenin veri seti içerisinde

bulunması hem de bulunmaması durumları için gerçekleştirilir. Bu veri setindeki her nesne için tekrarlanarak gerçekleştirilen bir süreçtir.

Algoritma sıklık desen uzunluğu 1 olan değerden başlayarak alt veri seti olarak kabul edilen ve sonek desenleri ile birlikte gözlemlenen önek yolları üzerinden şartlı desen düzeni oluşturur. Bu şartlı desen düzeni alt veri setlerinde tekrarlanarak algoritmanın temel yapı taşı olan fp-tree yapısını oluşturur. Nesneler arasındaki düzenin tespitinin başarılı olabilmesi için FP-tree üzerindeki bağlamların öncül ve ardıl nesnelerle sık ve anlamlı bir tekrarlanma ile gerçekleşmesi beklenmektedir.



Şekil 2.1 Fp-Growth İçerisinde Fp-tree ağaç yapısı örneği

Örneğin Şekil 2.1’de gösterilen A, B, C, D ve E nesnelerinden oluşan bir veri seti mevcut olsun. Algoritmanın veri setini ilk taramasında tekrarlanan nesneler ve destek değerini elde edebiliriz. Elde edilen nesne dağılımının destek değerlerinin azalan şekilde bir sıralamaya tabi tutulduğunda {C:6, D:3, E:3, A:2, B:2} olduğunu varsayalım. Bu aşamadan sonra algoritma her işlem kaydı için bir dal oluşturur. Örneğin “01: A, B, C,D,E” işlem kaydı beş nesne içermekte ve sıralı liste dikkate alınarak {C:1, D:1, E:1, A:1, B:1} dal oluşturulur. Oluşturulan dalın 5 düğüm noktası atanacaktır. Bu düğüm noktaları içerisinde “null” etiketli kök düğümün alt düğümü C, D ile C, E ve D, A ve E ile B ve A arasında bağlantılar oluşması beklenmektedir. İkinci işlem kaydı “002: B, C, E” olduğu düşünülürse yine sıralanmış şekilde {C,E,B} üç nesne tespit edilecektir. Kök düğüm ile C, C ile E, E ile arasında bağlantı oluşacaktır. Öncül C nesnesinin daha önceki işlem kaydından kaynaklanan bir bağlantısı olması nedeniyle düğüm sayısını 1 arttırarak [(E:1)(B:1)] düğüm grubunu [C:2] ile bağlayacaktır. Genel olarak ortak öneklerin tespitinde dal sayısının artırımını sağlamayı amaçlamaktadır.

Algoritmanın takibinin kolaylığı açısından bir başlık tablosu ağaç yapısı ile birlikte tutulmaktadır. Bu ağaç yapısı ile düğüm bağlantıları destek sayısı ve nesne bilgileri saklanmaktadır (Sidhu vd., 2014: 7). Böylelikle bir kurallar havuzu elde edilmesine olanak

sağlayan algoritma diğer birliktelik kurallarında olduğu gibi belirlenmiş veri seti içerisinde nesneler arasındaki bağlantılardan yola çıkarak nesne gruplarının tespitini sağlar.

Parçacık Sürü Optimizasyonu (PSO)

Parçacık Sürüsü Optimizasyonu 1995 yılında J.Kennedy ve R.C.Eberhart (1995: 39) tarafından; kuş sürülerinin hareketleri üzerine oluşturulmuş popülasyon tabanlı stokastik optimizasyon tekniğidir (Zhang, L. vd., 2003: 134). PSO matrisleri parçacık olarak tanımlayarak problem için her birini ayrı ayrı birer çözüm olarak kabul etmektedir. Her parçacık için geliştirilen çözüm, bir sonraki parçacık için bilgi birikimi olarak kabul edilerek en optimum sonucun elde edini mi amaçlanmaktadır.

Parçacık sürü konsepti basitleştirilmiş bir sosyal sistem simülasyonu olarak kabul edilir. Kuşların içgüdüsel koreografilerinden yola çıkılarak en yakın komşu hızlanma eşleşmesini bağımlı değişkenleri elimine ederek çok boyutlu ve uzaklık hız ilişkisi çerçevesinde ortaya çıkarılması amaçlanmıştır. Genetik algoritmalarla benzerlik gösteren bu yöntemle elde edilen sonuçlar, rastgele hale getirilmiş hız ve parçacık adı verilen problem uzayında yayılan potansiyel çözümlerce ortaya çıkarılmaktadır. Her parçacık gerçekleştirdiği en iyi çözüm veya uygunluk değeri ile problem uzayında koordinasyon takibini gerçekleştirir. Bu değerler “pbest” ismi ile anılır. Diğer bir en iyi değer çıkarımı daha global bir versiyon olan parçacık sürüm optimizasyonu ile tüm parçacıklardan toplanır (gbest). Bu konsept her zaman diliminde belirlenen pbest ve gbest konumları arasında hız değişimlerini değiştirilmesi ile oluşturulmaktadır. Hızlanma, pbest ve gbest konumları arasında rastgele belirlenmiş ağırlıklandırılmış değerlerle elde edilir (Ujjin ve Bentley, 2003: 128).

PSO üst sezgisel bir metot olarak kabul edilir. Optimize edilen problemle ilgili varsayımlara ihtiyaç duymaz ve belirlenen çözümler üzerinde büyük alanlarda optimum aramayı gerçekleştirebilir. Ancak elde edilen sonucun optimum olup olmadığı konusunda doğrudan yorum oluşturmaz. Diğer klasik optimizasyon metotlarında olduğu gibi Optimize edilen problemin türevlenebilir versiyonlarını analiz etmeye ihtiyaç duymaz.

PSO yöntemi üzerinde yıllara yaygın olarak birçok farklılaştırma gerçekleştirilerek, karşılaşılan çeşitli sorunlar için modifikasyonlar oluşturulmuştur. Böylelikle daha hızlı ve buna bağlı olarak daha kaliteli bir yakınsama ortaya çıkarılmıştır.

PSO devamlı optimizasyon problemleri için uygulanan bir yöntem olması nedeniyle ikili veriler için İkili PSO kavramı oluşturulmuştur (Sarath ve Ravi, 2013). Böylelikle satın alma geçmiş değeri matrisi gibi ikili verilere sahip matris yapılarına da uygun sonuçlar üretebilecek duruma getirilmiştir.

Araştırma için İkili PSO modeli şu şekilde oluşturulmuştur:

$$V_{k+1} = W * G_k + C_1 * r_1 * (P_k - X_k) + C_2 * r_2 * (g - X_k)$$

W değeri atalet faktörü olarak kabul edilmekte, C_1 ve C_2 değerleri ise pozitif sabit olarak düşünülmektedir. Her parçacık pozisyonu 0 ile 1 değerleri alacağından X_i değeri G_k değeri doğrultusunda olasılıklar dahilinde sınırlandırılacaktır. g değeri en iyi pozisyon değeri olarak ortaya çıkmaktadır.

PSO algoritmaları genellikle açık bir seçim mekanizması yürütmediği için diğer meta sezgisel yaklaşımlardan ayrı tutulmaktadır (Cheng ve Jin, 2015). Ancak çoğu PSO değişkenleri tüm sezgisel yaklaşımların içerisinde algoritmalarla bütünleşmede başarılı bir adaptasyon geçirebildiği görülebilmektedir. Bu algoritmalar dolaylı olarak kişisel en iyi sonuçları güncelleyerek global bir test sonucu üzerinden seçimi gerçekleştirmeleri bu özelliği kazanmalarında önemli bir faktördür. Diğer benzer algoritma yapılarında karşılaştırmalı yöntem kullanılmasına rağmen PSO ikili gruplar halinde sonuçları ele alır. İkili mukayese hafıza kullanımında tasarrufun yanında verimli bir çözümleme işlemine olanak vermektedir.

2.5. Hafıza ve Model Temelli Modellerinin Karşılaştırılması

Hafıza temelli yaklaşımlar tüm veri setini kullanarak öneri çıktıları üreten tekniklerdir. Bu algoritmalar istatistiksel tekniklerle hedef olarak belirlenen bir kullanıcı ile 'komşu' olarak adlandırılan kullanıcı gruplarının geçmiş davranışları inceleyerek aralarındaki benzerlikten yola çıkmaktadır. Komşuluk ilişkisinin tespitinin ardından farklı algoritmalar kullanarak kullanıcı davranış veya tercihlerini birleştirerek doğrudan bir tahmin yürütebilir ya da top-N öneri modelini hedef kullanıcı için üretebilir (Sarwar, Badrul Munir vd., 2001: 266). Hafıza temelli yaklaşımda algoritmalar veri setini bütün bir şekilde alarak kullanıcı ürün ilişkisi çerçevesinde tüm verilerle etkileşim içindedir. Bu nedenle herhangi bir optimizasyon işlemi gerçekleştirilmemişse yüksek işlem gücü gerektiren bir altyapı çerçevesinde algoritma çalışması gerçekleşir (Aditya vd., 2016: 304).

Model temelli yaklaşımda ise öneri algoritmalarının çalıştırılabilmesi için veri setinin geliştirilmesini ve dönüştürülmesini ön planda tutulmaktadır. Bu kategoride yere alan algoritmalar olasılıksal yaklaşımları dikkate alabilmekte ve kullanıcı tahmini için beklenen bir değer hesaplanabilirliği üzerine kurgulanmaktadır. Model oluşturma aşamaları farklı algoritmalarla oluşabilir Model temelli metotlarda kullanılan algoritmalar manipülasyon amaçlı saldırılara daha dirençlidir (Zhou vd., 2018). Öte yandan hafıza temelli metotlarda kullanılan algoritmalar daha fazla tahmin çeşitliliği sağlayan yöntemler ürün temelli metotlarda kullanılan algoritmalarla daha geniş bir perspektif içerisinde sonuca ulaşılmasını

sağlamaktadır. Tahmin çeşitliliği geniş bir çalışma alanının yanı sıra olumlu ve yeni kaynakların ortaya çıkmasını sağlayan tesadüfi sonuçların da çalışma içerisinde yer almasına olanak sağlayabilmektedir. Yenilik, çeşitlilik ve olumlu yönüyle tesadüfîlik yaratıcı ve yeni çözümlerin ortaya çıkarılmasında kullanıcı temelli modellerde farklı alternatiflerin sunulmasını sağlamaktadır.

Hafıza temelli metotların karşılaştığı en büyük sorunların başında seyreklik problemi gelmektedir (Martín-Vicente vd., 2014: 3801). Seyreklik, kullanıcı ürün veri seti içerisinde matrisin doğası gereği bilgi eksikliklerini tanımlar. Bu bilgi eksikliği nedeniyle algoritmalarda doğruluk ve tutarlılık problemleri yaşanabilmektedir (Huang, Z. vd., 2004). Seyreklik problemi ile ilgili farklı önleyici uygulamalara literatürde yer verilmekte ve genelde boyutsal indirgeme teknikleri ile grafik temelli modellemelerle çözüm üretilebilmektedir (Aggarwal, 2016: 67).

Model temelli metotlar seyreklik problemi ile daha iyi mücadele edebilmektedir (Ghazarian ve Nematbakhsh, 2015: 3804). Öneri modelinin oluşturulmasından önce veri seti üzerinde ön işlem gerçekleştirilebilen modeller kullanıcı ürün arasındaki eksik bilgileri elimine edebilmektedir. Ancak bu durum hafıza temelli modele göre daha fazla zaman ve işlem gücüne ihtiyaç yaratabilmektedir. Tüm veri tabanı üzerinde anlık işlem gerçekleştirmemesine rağmen model için veri setinin işlenmesi gerekliliği zaman ve işlem yoğunluğu açısından değerlendirmeye alınmalıdır.

ÜÇÜNCÜ BÖLÜM

METODOLOJİ VE UYGULAMA

3.1. Araştırma Metodolojisi

Öneri sistemleri ile ilgili e-ticaret üzerindeki uygulamalarının çeşitliliğine ve bu sektördeki hızlı büyümenin getirdiği yeni pazarlama seçeneklerine olan ihtiyaç, kullanıcı için farklı alternatiflerin devamlı ve hızlı bir şekilde oluşturulması talebini yaratmaktadır. E-ticarete öneri sistemleri, istatistiksel uygulamalarında gerçek zamanlı bir kullanıcı ürün etkileşimini oluşturma probleminin çözümü için istatistiksel ve bilgi keşfi yöntemlerinin toplu bir şekilde uygulandığı yöntemlerdir (Sarwar, B. vd., 2000: 158). Bu nedenle farklı algoritmik yapıların ortaya çıkarılmasına farklı disiplinlerden faydalanarak yorumlanabilmeye uygun bir metodolojiye takip edilmesi gerekmektedir. Bu doğrultuda literatürde birçok yönden kullanıcılar için öneri sistemleri incelenmiş ve farklı metotlarda yeni model ve yapılar oluşturulmaktadır.

Çalışmanın birinci bölümünde E-ticaret gelişimi ve veri kavramının değişimi üzerinde durularak, e-ticaret sistemlerinde verilerin yorumlanması adına öneri sistemlerinin nasıl bir gelişim gösterdiği açıklanmıştır. Öneri sistemlerinin geliştirilmesi için veri yapılarının oluşturulması ve bu veriler doğrultusunda ortaya çıkan öneri sistem modelleri (İşbirlikçi, içerik temelli, hibrit modeller vs.) üzerine tanımlar açıklanarak, mevcut e-ticaret sistemleri üzerindeki uygulanabilirliği üzerine çalışmalar incelenmiştir. E-ticaret sistemlerinde farklı uygulama teknikleri üzerine tarihsel süreç irdelenerek, modellerin oluşturulması ile ilgili bugüne kadarki gelişim, literatür taraması ile akademik boyutta incelenmiştir.

İkinci bölümde ise uygulamada da kullanılan işbirlikçi filtreleme modelinde, kullanıcı ve ürün temelli komşuluk model oluşumu konusunda teorik yapılar incelenmiştir. Bu bağlamda hem kullanıcı hem de ürün temelli e-ticaret öneri sistemlerinde komşuluk temelli istatistiksel uygulamaların nasıl işlediği ve öneri sistemlerine nasıl adapte edildiği konusunda detaylara yer verilmiştir. Bunun yanında modellerin getirdiği avantaj ve dezavantajlar ortaya çıkarılarak, karşılaşılabilecek problemler hakkında bilgilere yer verilmiştir.

Çalışmanın bu bölümünde ise metodolojik uygulamalara yer verilmiştir. İlk olarak yapılan araştırmanın konusu, kapsamı ve akademik açıdan önemi vurgulanarak, karşılaşılan kısıtlar anlatılmıştır. Sonra ise oluşturulan model ile ilgili hipotezlere yer verilerek örneklem süreci ve veri toplama yöntemleri, analizleri konusunda açıklamalar yapılmıştır. Elde edilen tüm veriler doğrultusunda analizler gerçekleştirilmiştir.

3.2. Araştırmanın Konusu ve Önemi

Dünya genelinde 2,8 trilyon dolarlık bir perakende satış hacmine ulaşan e-ticaret sitelerinin sayıları gün geçtikçe artmakta ve tüm internet kullanıcılarına ulaşmak için çeşitli yollarda yeni yöntemler geliştirmeye çalışmaktadırlar. Rekabetle birlikte hızlı büyüme, yeni e-ticaret fırsatları ortaya çıkarmakta ve kullanıcının ilgi alanını genişletecek çözümler üretilmektedir. 2018 yılı itibarıyla dünya nüfusunun %59,9'una internet erişimi sağlandığı düşünüldüğünde Pazarın katlanarak büyümesi e-ticaret sistemlerinin çeşitliliğinin artmasına olanak sağlaması beklenmelidir (Statista, 2018).

Ürün tercihinin belirlenmesinde kullanıcı davranışlarının hızlı bir değişimle birlikte ön plana çıkması, zaman içerisinde yapılan araştırmalar ile akademik çalışmaların çeşitlenmesine de vesile olmuştur. Kullanıcıların reklam, tek taraflı kullanıcı tanıtımları gibi kullanıcının isteği dışında gerçekleştirilen yönlendirmeler yerine, kullanıcının özellikleri analiz edilerek ona uygun bir tahmin mekanizması oluşturmak kullanıcının internet gezinmesindeki doğal davranışlarının önemli bir gösterge olabileceğinin farkına varılmasını sağlamıştır (Glance, 2005: 1). Bu durumda da kullanıcı ürün ilişkisinin sağlıklı bir şekilde kurularak başarılı bir yönlendirme döngüsü elde edilebilmiştir.

Çalışmalarda oldukça sık kullanılan ve tezin konusu dahilinde burada da anlatımı yapılan işbirlikçi filtreleme modeli de farklı algoritmik yapılarla çeşitlendirilmiş ve ihtiyaca uygun olarak yeniden yorumlanmıştır. Genellikle belirli bir ölçek üzerinden gerçekleştirilen kullanıcı değerlendirme puanları ile kurgulanan model yapıları, kullanıcıların doğrudan ürün hakkındaki fikirlerini almak üzere kurgulanmıştır. Bu doğrultuda kullanıcının görüşleri sınıflandırılıp kullanıcı tercihleri dikkate alınarak modele dahil edilmiş ve hedef kullanıcılar için sonuçlar üretilmiştir.

Öte yandan kullanıcının belirli bir yapı hakkında geniş bir ölçek içerisinde düşüncelerinin belirtilmediği ancak olumlu veya olumsuz algının tespit edilmesi için kullanıcının bazı davranışlarının dikkate alınabildiği yapılarla ilgili de çalışmalar gerçekleştirilmiştir. Kullanıcının ürünü beğenmesi, yararlı olarak tanımlaması gibi olumlu ifadeler üzerinden görüş bildirmesi de işbirlikçi filtreleme modellerinde kullanılmıştır.

Doğrudan kullanıcı davranışlarının incelendiği çalışmaların yanında, kullanıcının ürün hakkındaki dolaylı görüş bildirimleri de modellerde kullanılabilmektedir (Koren vd., 2009: 44). Bu durum kullanıcıya doğrudan bir ürün hakkındaki düşüncelerini sormak yerine, kullanıcının sistem içerisindeki davranışları incelenerek, o ürün hakkındaki olumlu/olumsuz tavrının kestirilmesi üzerine kurgulanmıştır. Böylelikle kullanıcının ürün ile ilgili görüşleri için

oluşturulan değerlendirme ölçeğinin dışına çıkılarak kullanıcının belirlenmiş bir davranışının ürün hakkındaki düşüncesi olarak kabul edilmiştir.

Çalışmanın uygulama bölümünde bahse konu dolaylı kullanıcı davranışı üzerinden ürün hakkındaki değerlendirmesinin tespit işlemi ile tahmin oluşturma modelleri yapılandırılması amaçlanmıştır. Bu süreç kullanıcının bir ürünü satın alma durumu üzerinden kurgulanmıştır. Sisteme kayıtlı bir kullanıcı, belirli bir ürünü satın almışsa o ürün hakkında olumlu bir fikre sahip olduğu varsayılmış, aynı ürün için bu davranışı sergileyen kullanıcılarla eşit bir puan verdiği varsayılarak analizlere dahil edilmiştir. İkili puanlama sistemi olarak kabul edilen bu yapıda, eğer kullanıcı sisteme girme amacını yerine getirerek o ürünü satın almışsa, ürünün amacına uygun olarak alınabilirliğine onay vermiş ve benzer özelliklere sahip diğer kullanıcılar için de olumlu bir yapıya sahip olduğu kabul edilmiştir.

Kullanıcıların satın alma tercihi üzerinden oluşturulan bir ürün öneri tahmin modeli ile işbirlikçi filtreleme metotları vasıtasıyla benzer kullanıcı kümeleri içinde yer alan hedef kullanıcılar için farklı ürün tercihlerine yönlendirilme başarısının karşılaştırılması, bu çalışmanın ana konusu olarak belirlenmiştir. Kullanıcının standart öneri sistemlerinde olduğu gibi doğrudan bir görüş bildirmesine gerek kalmaksızın ürün satın alma davranışının olumlu bir değerlendirme olarak kabul edildiği bir öneri mekanizması kurgulanmıştır. Böylece kullanıcının tek taraflı ve anlık olarak değişme ihtimali olan değerlendirmeleri yerine, nihai karar süreci üzerine bir temel oluşturulmaya çalışılmıştır.

Araştırma, satın alma davranışları üzerinden işbirlikçi filtreleme yöntemleri ile öneri sistemi modelinin kurgulanması üzerine olsa da veri setinde kullanılan veri türlerinin kullanılacak metotlar üzerindeki performans etkisinin ölçümü için de yol gösterici olmuştur. Bu durum araştırmada genel bir yöntem arayışının yerine veri türüne göre yöntemleri özelleştirerek farklı uygulamaların yaratabileceği sonuçları ortaya koymuştur. Böylelikle geçmişte aynı tür veri yapısına sahip veri setleri ile yapılan çalışmalarla kıyaslayarak bu gibi özel durumlar için kullanılabilecek en verimli metotların tespitinin yapılmasına katkı sağlanmaya çalışılmıştır.

3.3. Araştırmanın Amacı

Yapılan çalışmanın genel amacı e-ticaret sitelerinde gerçekleştirilen satın alma işlemleri üzerinden, benzer kullanıcı profillerinin ortaya çıkarılarak kullanıcılara en uygun ürün önerilerinin sunulması için işbirlikçi filtreleme modeli içindeki çeşitli yöntemlerin kıyaslamasının yapılmasıdır. Araştırmada, kullanıcıların ürünler hakkında kendi puanlamalarından elde edilen değerlendirmelerin aksine, sistem içerisinde gerçekleştirdiği satın

alma işlemleri üzerinden modellemelerin oluşturulması tercih edilmiştir. Kullanıcıların ürün hakkında tercihini belirleyen önemli bir kriter olarak kabul edilebilecek satın alma davranışı incelenerek, kullanıcıların bu eylemi ile ürün hakkında bir kanaat belirttiği kabul edilmiş ve bu veriler üzerinden modeller kurgulanmıştır.

Bu tercihin sebeplerinden biri, kullanıcıların düşüncelerini belirli bir ölçek çerçevesinde değerlendirmeleri nedeniyle, ister istemez sistem tarafından bir yönlendirmeye maruz kalmakta ve gerçek düşüncesini sınırlı bir çizelge üzerinden belirtmeye zorlanmasının neden olabileceği hatalı bilgi girişlerinin önüne geçmektir. Ancak kullanıcı kendi ihtiyaç analizi doğrultusunda yaptığı ön araştırma ve değerlendirme ile bir ürünü almaya karar vermesi, o konuda dolaylı olarak bir değerlendirme yaptığının varsayılarak daha net ve keskin bir sonuç ile değerlendirmesini yaptığı düşünülebilir.

Bir diğer husus ise kullanıcının tek taraflı yapacağı değerlendirmeler ile ancak sınırlı sayıda ve belirli ürünler üzerinde yoğunlaşacağı gerçeğidir. Kullanıcılar genelde popüler hale gelmiş ya da kullanıcının kendi tecrübe edinimleri doğrultusunda değerlendirme yapmaya meyillidirler (Qiu vd., 2015: 429). Kullanıcılar, ürüne sahip olmaları ya da üçüncü kaynaklar vasıtasıyla elde ettiği bilgiler doğrultusunda bir puanlama ve değerlendirme gerçekleştirebilmektedirler. Bu durum kullanıcının sınırlı bir ürün yelpazesi içerisinde değerlendirmede bulunmasına neden olmakta, kullanıcı kendi tecrübeleri dışında elde ettiği bilgiler ile tutarsız puanlamalara yönelebilmektedirler.

Belirtilen bu durumlar dikkate alındığında kullanıcının hem tecrübe ve araştırma verileri doğrultusunda net kararını gösteren hem de nihai bir sonuç içeren satın alma tercihi, iyi bir değerlendirme kriteri olarak kabul edilebilir.

E-ticaret sistemlerinde kullanıcı sayılarına kıyasla ürün çeşitliliği oldukça fazla olabilmektedir. Her kullanıcı için ürünler hakkında bir değerlendirme oluşturma imkânı mevcut olabilirken, her ürün için bu durum her zaman söz konusu olamayabilmektedir. Kullanıcıların her ürün hakkında değerlendirme yapması veya bir değerlendirme yapmaya yönlendirilmesi genellikle kullanıcı deneyimi açısından problemlı bir veri elde edinimine neden olacaktır. Bu yüzden ikili veri tipinin kullanımı, değerlendirmeye alınan ürün yelpazesini genişletecek ve daha fazla çıkarıma imkân sağlayabilecektir.

Tez kapsamında yapılan uygulamada bahse konu ikili değerlendirmeler çerçevesinde kullanıcı satın alma davranışlarının kestiriminin yapılıp yapılamayacağı hangi yöntemlerin bu kestirim için elverişli olduğu incelenmiş, zaman kısıtlamaları da araştırmanın içine katılarak bir karşılaştırma yapılmıştır. Kullanıcı puan enstrümanı olarak satın alma tercihi dikkate alınarak,

tespit edilen küme içerisindeki benzer kullanıcıların bu ürünü alıp almama kararları üzerine bir model oluşturulmuştur.

Araştırma gerçekleştirilirken daha önce yapılan çalışmalardan yola çıkarak öneri sisteminin üreteceği çıktılarının başarısının yükseltilebilmesi için farklı kısıtlamalara gidildiği gözlemlenmiştir. Bu doğrultuda araştırmada sadece kullanıcıların satın alma geçmişi incelenerek genel bir zaman dilimi içerisinde çıktılarının karşılaştırılması yerine, birden fazla satın alma işlemi yapan kullanıcıların bu davranışı tekrar yapma sıklıkları incelenerek, ardından gelen satın alma eylemlerinin süreleri ile ilgili kriterler belirlenmiştir. Bu nedenle tahmin modelleri 45, 90 ve 365 gün sınırları ile birlikte tüm zaman dilimi içerisindeki karşılaştırmalarına analizlerde yer verilmiştir.

Bu doğrultuda, Araştırma öneri sistemlerinin kullanımını özelleştirerek farklı metotlarla elde edilebilecek sonuçları kıyaslamak, belirtilen kısıtlara ek olarak zaman aralıkları açısından da öneri sistemlerinin performansının ölçümünü hedeflemektedir. Belirli zaman aralıkları ile elde edilecek öneri sistem performansının süre açısından kıyaslaması yapılarak sadece ikili bir veri yapısına sahip veri seti üzerinden değerlendirmeler yapılmıştır.

3.4. Araştırmanın Kısıtları

Araştırma çerçevesinde gerçekleştirilen uygulamada, hem tez çalışmasında uyulması gereken kurallar hem de veri seti içerisinde homojen yapının bozulmaması adına tek bir kaynaktan alınan veriler kullanılmıştır.

Kişiselleştirilmiş bir öneri sistemi modeli oluşturmak için ihtiyaç duyulan veri setine, öncelikli olarak yerel e-ticaret şirketleri üzerinden ulaşılmaya çalışılmıştır. Ülkemizde son yıllarda ivmelenerek artmaya başlayan küçük ve orta büyüklükteki e-ticaret siteleri, altyapılarını henüz sistematik ve düzenli bir yapı üzerine kurmayı başarmakta zorlanmaktadır ve veri saklama konusunda yeterli özen gösterilmeyerek, işlemler dağınık bir sistem üzerinde gerçekleştirilmektedir. Bu durum genellikle verinin saklanması ve kayıt altına alınmasında karşılaşılan maliyetler nedeniyle işletmelerin bu maliyetlerden kaçınması için kısıtlamaları görmesinden kaynaklanmaktadır. Veri saklama maliyetlerindeki düşüşler ile bu verileri kayıt altına alma yöntemlerindeki gelişmelere rağmen, küçük ve orta ölçekteki e-ticaret işletmeleri önceliklerini veri üzerine kurgulamamaktadırlar. Bunun yanında küçük ve orta büyüklükteki şirketlerin günümüzdeki veri anlayışının hızlı değişimine anlayarak, ayak uydurma konusunda sorunlar yaşamaktadır. Veriler geçmişe yönelik bir bilgi birikimi olmalarının yanında geleceğe yönelik şirket politikalarını etkileyecek ve yönlendirebilecek bir araç olduğu bilişim sektöründe

artık kabul edilen bir gerçektir. Ancak Türkiye’de sektörün belirli bir kültür çerçevesinde gelişimine olanak sağlayacak çevresel faktörlerin bulunmamasından kaynaklanmaktadır.

Veri biliminin hızla ilerlediği son yıllarda büyük ve küresel çapta hizmet veren şirketlerin aralarındaki rekabette bir adım öne geçmek adına farklı metotlar ve stratejiler gerçekleştirmelerine rağmen, bu yeniliklerin kendilerinden daha küçük ölçekteki e-ticaret sistemlerinde uygulanması belirli bir zaman geçmesinin ardından gerçekleşebilmektedir. Pratik uygulamalarda akademik çalışmaların ışında gerçekleştirilmesi nedeniyle literatürdeki gelişmeler öncelikli olarak yeniliklere açık ve araştırma-geliştirme yatırımlarının yoğun olduğu şirketlerce adapte edilmektedir. Bu adaptasyon süreci büyük şirketlerin öncülüğünde gerçekleştirilse de hızlı bir şekilde daha alt kademelerde sistematik bir yayılım gerçekleşebilmektedir.

Söz konusu verileri ve verilerin elde edildiği zaman dilimi şirketin veri yapısı ve saklama politikası kaynaklı belirli kısıtlamalar dahilinde elde edilebilmiştir. Ülkemiz e-ticaret ekosistemi içerisinde yasal düzenlemelerin yeterince hızlı ve dinamik bir şekilde gerçekleştirilememesi ve e-ticaret sitelerinde karar vericilerin veri paylaşımı konusunda önyargı ve çekinceli tutumları akademik çalışmaların gerçekleştirilebilmesi açısından engelleyici ve aşılması en zor adımların başında gelmektedir. Verilerin gizliliği konusunda şirket politikalarının oluşturulmaması, gizlilik ile ilgili taahhüt ve karşılıklı gizlilik sözleşmelerinin oluşturulmasının yaygın olmaması araştırmacılar tarafından yapılacak olan çalışmalar için ciddi engeller oluşturabilmektedir.

3.5. Araştırmanın Kapsamı

Öneri sistemleri, literatür oldukça geniş ve her çalışma için özelleştirilebilir bir model çeşitliliği sunmaktadır. Farklı türlerdeki veri yapıları için istatistik yöntemlerinin uygulanabilirliği, mevcut modellere çeşitli algoritmalar ekleyerek, sistemsel yapılara uyarlanabilirliği sağlanmış ve öneri mekanizmasının daha verimli çalışması için fazla sayıda metoda yer verilmiştir.

Literatürde kullanıcı değerlendirmeleri ile alışveriş esnasında elde edilen meta veriler üzerinden yapılan araştırmalara sıklıkla rastlanılmaktadır. Bu doğrultuda çalışmada öneri sistemlerinin özelleştirilerek müşterilerin sadece alışveriş alışkanlıkları dikkate alınması ve öneri sistemleri modellerine uyarlanması üzerine özelleştirilmiş bir çerçeve belirlenmiştir. Böylece kullanıcılardan elde edilen çok yönlü veri setlerinden beslenen bir modelden daha çok sınırlı bir çerçevede ele alınmış ve hedefe yönelik tutarlı bir sonuç elde edilmesi planlanmıştır.

Araştırma genel bir öneri sistemleri uygulaması modeli oluşturmak yerine veri tipi ve kullanılabilir yöntemler açısından değerlendirmeler yaparak 3 ana metot ve zaman kriterini dikkate alınmıştır. Kullanıcıların kişisel bilgileri, sitede oluşturduğu diğer meta veriler ile ürün özellikleri ve ürün hakkındaki betimleyici bilgiler analizlere dahil edilmemiştir. Böylelikle sadece satın alma verileri ve ürünün kendisi değerlendirilerek kullanıcıların satın alabileceği ürünlerin tespitinin gerçekleştirilmesinin sağlanması amaçlanmıştır.

Oluşturulan modellerle ikili veri tipine sahip satın alma geçmiş veri seti için belirlenen üç metot üzerinden karşılaştırma yapılmıştır;

- Ürün Temelli İşbirlikçi Filtreleme Modeli ve Benzerlik Ölçüm Metotları
 - Kosinüs Benzerlik Ölçümü
 - Jaccard Benzerlik Ölçümü
- Kullanıcı Temelli İşbirlikçi Filtreleme Modeli ve Matris Ayırıştırma Metotları
 - TDA (Tekil Değer Ayırışımı)
 - NMF (Negatif Olmayan Matris Ayırıştırma)
- Kullanıcı Temelli İşbirlikçi Filtreleme Modeli ve Birliktelik Kuralı Metotları
 - Fp-Growth
 - PSO (Parçacık Sürü Optimizasyonu)

Söz konusu metotlar uygulama alanları göz önüne alınarak hafıza temelli, model temelli ve bunlarla birlikte birliktelik kuralları başlığı altındaki yöntemlerden literatürdeki kullanımları doğrultusunda belirlenerek araştırmaya dahil edilmiştir. Böylelikle araştırma için kullanılacak veri tipi ile uyumluluğu denenerek performans kıyaslaması ile verimliliklerinin ölçümü sağlanmıştır.

Uygulanabilecek metotların kısıtlanarak sadece üç temel metot üzerinde yoğunlaşılması, sağlıklı bir performans kıyaslamasının yapılabilmesine olanak sağlamaktadır. Mümkün olduğu kadar dar bir alanda net bir sonuç çıkarabilmek adına farklı yöntemlerin incelenmesi ve bu yöntemlerin kıyaslanması literatüre doğrudan bir katkı vermeye olanak tanımaktadır.

Bunların yanında zaman kriteri ile öneri sistemleri modelleri ile elde edilen çıktıların zaman dilimleri ile sınırlandırılarak performans farklılıkları ortaya çıkarılmaya çalışılmıştır. Bu durumda tüm veri seti üzerinde yapılan değerlendirme ile kullanıcıların alışveriş yaptıkları zaman aralığı değerlendirilerek belirli dilimlerde daha iyi bir öneri skoru elde edilemeyeceği araştırılmıştır.

3.6. Araştırma Modeli

Araştırma, e-ticaret sistemlerinde kullanıcıların satın alma davranışları dikkate alınarak belirli bir zaman aralığına dayalı öneri sistemi modellerinin performans ölçümünün yapılması üzerine kurgulanmıştır. Model oluşturulurken, literatürde sıklıkla da kullanılan genel kullanıcı ve ürün bilgilerinin tümünden derlenerek karma bir veri seti üzerinden gitmek yerine kullanıcının yorumunun doğrudan dahil edilmediği, özelleştirilmiş ve kişiselleştirilmiş bir veri havuzundan yararlanılmıştır.

Araştırmanın konusunu kapsayan uygulama metodu literatürde öneri sistemleri için geliştirilen temel birkaç ana modelden biri olan “İşbirlikçi Filtreleme Modeli” üzerine kurgulanmıştır. Temel prensip olarak, kullanıcıların sağlamış olduğu doğrudan değerlendirmeleri üzerine kurulu olan bu model, duyulan ihtiyaç ile sınırlandırılarak kullanıcı ve ürün bilgilerinin doğrudan veya dolaylı değişken olarak algoritmalarla, dolayısıyla modele dahil edilmesi ile farklılaştırılmıştır. Araştırmada da kullanıcı ve ürün özelliklerinin kullanımı yerine kullanıcının doğrudan ürün üzerindeki tercihi yansıtan ürünü satın alma davranışı analizlerde temel değişken olarak kullanılmıştır.

Kullanıcının ürün üzerindeki temel düşüncesi olarak kabul edilen satın alma eylemi, veri yapısı olarak ikili veri tipi olarak kabul edilmiş ve kullanıcının ürün hakkındaki yorumunu satın alma eylemini gerçekleştirmesi ile yaptığı varsayılmıştır. Bu varsayım klasik işbirlikçi filtreleme yöntemlerinde kullanılan kullanıcı değerlendirme verilerine benzer bir yöntemle puanlama/değerlendirme matrisi yerine, satın alma geçmişi matrisleri oluşturularak analizlerin gerçekleştirilmesi sağlanmıştır. Bunun dışında ürün ve kullanıcı temelli diğer veriler, yapısına göre matrise dahil edilmiş ve analizde oluşan farklar açısından da değerlendirmeye alınmıştır. Oluşturulan satın alma matrisi bahsedildiği üzere ikili bir yapıya sahip olması nedeniyle komşuluk temelli işbirlikçi filtreleme modellerinin özelleştirilmesi ile analiz edilmiştir.

Literatürde, e-ticarette öneri sistemleri ile ilgili yapılan çok sayıda uygulama olmasına rağmen kullanıcıların sadece satın alma eğilimleri dikkate alınarak oluşturulmuş model yapınınması ile çok fazla karşılaşmamaktadır. İşbirlikçi filtreleme ve içerik temelli filtreleme modellerinin tek başına veya kombinasyonları ile oluşturulan hibrit modellerde, kullanıcıdan elde edilen verilerin yanı sıra ürün ile ilgili veriler de analizlere sıklıkla dahil edilmiştir (Çano ve Morisio, 2017: 1487). Bu durum çok yönlü bir çıkarım edilmesine olanak tanırken oluşturulan modellerdeki hata paylarının da minimize edilmesine yardımcı olmaktadır.

Öte yandan veri setindeki çeşitlilik, modeller için belirlenen hedef çıktıların yorumlanması konusunda tutarsızlıklara neden olabilmektedir (Jamiy vd., 2015: 1). Çok fazla değişkenin modellerde yer alması sonuç ile çıktının edilen niceliksel veriler üzerindeki etkisini

ölçmek ve yorumlamak çoğu zaman daha fazla işlem gücüne ve modelin genişlemesine sebep olabilmektedir. Araştırmanın Amacı ve Araştırmanın Kısıtlarında belirtilen hususlarla birlikte bu etkenler de düşünülerek araştırma modeli dar ve özelleştirilmiş bir yönde kurgulanmasına karar verilmiştir.

Araştırma modeli, belirtilen e-ticaret sitesinde gerçekleştirilen alışverişler içinde birden fazla satın alma işlemi sahibi sadık kullanıcıların, belirlenen zaman aralıkları içerisinde oluşturulan öneri sistemi modellerinin başarısı üzerine kurgulanmıştır. Model oluşturulurken kullanılan sadık kullanıcı terimi belirlenen süre içerisinde iki veya daha fazla satın alma işlemi gerçekleştiren kullanıcıları tanımlamaktadır.

Karar vericiler ile yapılan görüşmeler ve veri seti yapısı dikkate alınarak belirlenen zaman dilimleri doğrultusunda 3 yıllık bir dönem içerisinde 40,000'den fazla kullanıcı içerisinde sadık kullanıcılar belirlenerek, ikili veri yapısına uygun işbirlikçi filtreleme modellerinden matris ayrıştırması, birliktelik kuralı algoritmaları ve ürün bazlı öneri sistemleri üzerinden özelleştirilmiş öneri sistem modelleri oluşturulmuştur. Seçilen yöntemler literatürde kullanım sıklıkları değerlendirilerek rasgele bir seçim ile belirlenmiştir.

Model için temel veri olarak kullanıcıların satın alma davranışı olarak belirlenmiştir. Araştırmada bilgi alanı belirlenen yüksek satın alma aktivite aralıkları çerçevesinde gruplandırılarak ayrı ayrı modeller oluşturulmuştur. Araştırma bu yöntemler doğrultusunda oluşturulan modellerin çıktılarının karşılaştırılarak en tutarlı sonuç veren yöntemlerin belirlenmesi üzerine kurgulanmıştır.

Araştırma modeli kurgulamasında Python dilinden faydalanılmış ve model Amazon SageMaker üzerinde oluşturulmuştur. Çalışmada elde edilen verilerin ön işlem ve iterasyon tek bir yerden yürütülmüştür.

3.7. Örneklem Süreci ve Kullanılan Veri Seti

Araştırma için kullanılan veri seti Türkiye'de popüler olarak b2c (İşletmeden Tüketicieye) kategorisinde kozmetik ve ıtriya malzemesi satışı yapan Turuncukasa e-ticaret sistemi üzerinden elde edilmiştir. Şirket yetkililerinden edinilen bilgilere göre, günlük 300.000'den fazla sayfa gösterimi gerçekleştiren ve sektöründe öncü olan firmalardan biri olan bu şirket, Antalya merkezli ulusal bazda perakende satış gerçekleştirmekte olup anne ve bebek, cilt bakımı, makyaj ve kişisel bakım gibi birçok kategori de binden fazla ürün üzerinden hizmet vermektedir.

Araştırma için gerekli veriler, söz konusu e-ticaret sitesinden elde edilmiş olup 2013-2016 yılları arasını kapsamaktadır. Bu yıllar arası verilerin tercih edilme sebepleri düzenli bir

alışveriş sisteminin kurulmuş olması ve şirketin kurumsal kimliğinin oturması nedeniyle müşteri sadakati konusunda iyi bir yol kat edildiğinin şirket karar vericileri tarafından belirtilmesidir. Bunun yanında veri yapısı, ulaşılabilirliği ve gizlilik ile ilgili ilkesel tutumlar, ideal veri aralığı olarak bu dönemin seçilmesinde etkili olmuştur.

Söz konusu sitemden veriler ham haliyle elde edilmiş ve kullanıcı ürün kategori ayrımı için yapılandırılarak bir veri tabanı üzerinde yeniden oluşturulmuştur. Anlamsız, hatalı ve kullanıcı/kategori tabloları arasında bağdaştırılamayan kayıtlar ön temizleme işlemine tabi tutulmuş ve veri seti tüm bu süreç öncesi hazır hale getirilmiştir. E-ticaret sitesi sağlık ürünleri ile birlikte ıtriya ve güzellik/bakım ürünlerine yoğunlaşmış bir ürün yelpazesine sahiptir. Bu durum aynı ürünlerin farklı gramajlar veya adetlerle sunulmasına imkân vermekte ve özünde aynı olan ürünlerin veri seti içerisinde farklı ürün gibi algılanmasına neden olmaktadır. Bu durum da yapılan karşılaştırmalarla tespit edilerek benzer tür bir yapıya sahip ürünler tek bir ürüne indirgenmiştir. Yine veri seti içerisindeki kategorik yapı irdelenmiş, bazı ürünlerin birden fazla kategoriye ait olma durumları değerlendirmeye alınmıştır. Uygulama her ne kadar kullanıcı – ürün ilişkisi üzerine kurgulansa da araştırma öncesi kullanıcı- kategori/alt kategori ilişkileri irdelenmiş ve analize dahil edilip edilmeyeceği test edilmiştir. Bu durumun uygunluğu tespit edilemediği için kullanıcı ürün ilişkisi üzerinden devamına karar verilmiştir.

Veri seti üzerinde yapılan ön işlemler neticesinde sınırlı bir biçimde modele entegre edilen veri seti içerisinde, yanlış veya eksik olarak bulunan veri girdilerinin yer almasının önüne geçilmiştir. Özellikle analizlerde kullanılan kategorik verilerde tespit edilen eksiklikler, verilerin temin edildiği şirketten tamamlanabilecek ikincil kaynaklardan elde edilmeye çalışılmış, temin edilememesi durumunda ise veri setinden çıkarılmıştır. Verilerin tamamlanması sırasında satış kaydının tamamlayıcı bilgileri olarak tanımlanan kullanıcı bilgileri veya ürün detayları, belirtilen 2013-2016 yılları arasında temin edilememesi durumu olduğunda yazılımla düzenli olarak şirketin tüm veri tabanından sorgulanarak tamamlanmaya çalışılmıştır. Bunun yanında tüm kayıtlardaki ikincil veriler tutarlılığın sağlanması amacıyla tekrar zaman kısıtı olmadan veri tabanından kontrol edilerek belirsizliklerin önüne geçilmiştir.

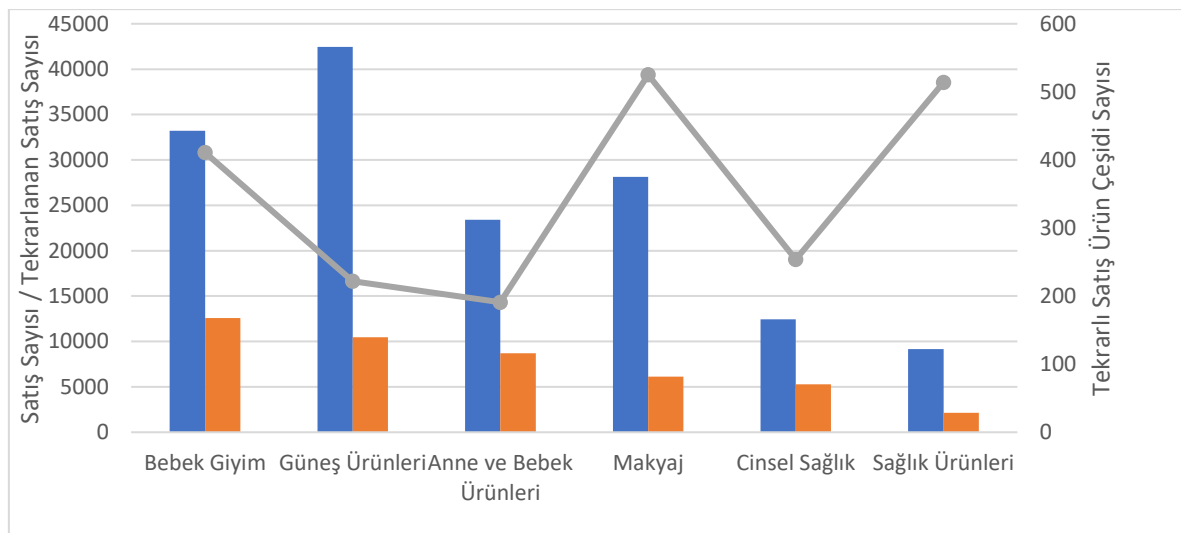
Veri seti incelendiğinde 321.421 üzerinde satış kaydının, 49.212 kullanıcı ve 4429 ürün 12 üst kategori altında 98 alt kategorinin bulunduğu tespit edilmiştir. Bu satış kayıt bilgileri satışın gerçekleştirildiği zamanı, gerçekleştiren kullanıcı, satın alınan ürün, ürün teslimat aşamaları, ödeme yöntemleri ve kısmi olarak kullanıcı detay bilgilerini kapsamaktadır. Şirket, kullanıcı gizliliğini sağlamak amacıyla kişisel verileri maskeleyerek erişime açılmasına izin verilmiştir. Maskeleyme işlemi ile kullanıcıların kimliklerinin tespitine imkân vermeyecek şekilde veriler yeniden yapılandırılmıştır. Bununla birlikte doğum tarihleri bilgisi giren

kullanıcıların ay ve yıl bilgileri alınmış, alışveriş adresleri içinde şehir bilgisi veri tabanından aktarılmıştır.

Veri seti içerisinde gerçekleştirilen tüm bu temizleme ve yeniden oluşturma işlemleri sonrasında 312.775 kayıt analizlere dahil edilebilir olduğu tespit edilerek, çalışma bu veriler ışığında ilerletilmiştir. Araştırma kullanıcıların satın alma geçmişleri dikkate alınarak öneri modelleri oluşturmaya yönelik olması nedeniyle veri setlerinde bazı sınırlamalara gidilmiştir. Bunlardan en önemlisi veri seti içerisinde tekrarlanan alışverişlerin oluşturduğu dizilimlere yer verilmesidir. Bu nedenle analizlerde birden fazla kez alışveriş yapmış kullanıcılar ve ürünlerle sınırlamaya gidilmiştir. Böylelikle veri seti içerisindeki seyreklik problemine öncül bir önlem alınmıştır. Bu sınırlama sonucu, satış kayıtları 203.566 adete, satış yapılan ürün çeşidi sayısı ise 2389 adete düşmüştür. Veri setindeki kullanıcı sayısı ise 20.176 olarak belirlenmiştir.

Ölçeklendirilmiş veri seti incelendiğinde kullanıcı ve ürün bilgi dağılımında ciddi farklılıklar gözlemlenmemiştir. Veri seti içerisinde eksik verilerin tamamlanarak hatalı kayıtların çıkarıldığı hali ile kıyaslandığında düşüş gösteren satış işlemine konu olan ürün çeşitliliğinde kayda değer bir azalma gözlemlenmemiştir. Bu durum, birden fazla ürün tercihi yapan kullanıcıların genelde belirli bir ürün skalası içerisinde tercihlerini yinelediği, bu doğrultuda ürün sadakatinin yüksek olabileceğinin değerlendirilmesi gereken konulardan biri olduğunun göstergesidir.

Araştırmada kullanılan verilerin yapısal ve sayısal özellikleri incelendiğinde Türk e-ticaret ekosistemi içerisindeki dağılımlara yakın bir durumun temsil edildiği gözlemlenmiştir (Tusiad Raporu). Bu duruma şirketin sektör içerisinde oturmuş bir yapısının olması ve kendi kategorisi altında güvenilir ve marka bilinirliğinin yüksek olmasının katkısı olduğu açıktır.



Şekil 3.1 Veri Seti İçerisindeki Tekrarlanan Satış ile Ürün Çeşidi Sayısı

Ürün bazında veri seti içerisindeki çeşitlilik Şekil 3.1’de en fazla tekrarlanan satış işleminin gerçekleştirildiği beş kategori gösterilmiştir. “Bebek Giyim”, “Güneş Ürünleri” ve “Anne ve Bebek Ürünleri” kategorileri veri seti içerisindeki satışların yaklaşık %34,21’ini kapsamaktadır. Bu kategorilerde ise ortalama %26,44 oranında kullanıcıların daha önce satın aldıkları ürünleri tekrar alma eğiliminde olduğu gözlemlenmiştir. En çok tekrar satın alma işlemi ise oransal olarak Bebek giyim, güneş ürünleri ve cinsel sağlık kategorilerinde gerçekleşmektedir. Oranlar yüksek olsa bile kullanıcı ihtiyaçlarının tekrarlanan bir dönu içerisinde sistem üzerinden giderildiği ve kullanıcıların bir bölümünün alışverişlerini tekrarladığı gözlemlenebilmektedir. Burada belirtmek gerekir ki ürünlerin çeşitliliği veri tabanı içerisinde boyut, adet veya ürün için belirlenen standart özellikler halinde gruplandırılarak tek ürün gibi düşünülmüştür. Gözlemlenen bu durum kategoriler altındaki ürün çeşitliliği açısından da benzer bir yapıyla örtüşmektedir.

Kategori verileri oluşturulurken sistem içerisinde oluşturulan alt kategoriler analizlere dahil edilmemiştir. Bunun başlıca nedeni alt kategori kırımlarının çok fazla sayıda olması ve verilerin alt kategorilere inildikçe seyrekleşerek analizlere dahil edilmesinin zorluğundan kaynaklanmaktadır.

Verilerde satılan ürün yelpazesi açısından çeşitlilik göstermesine rağmen satışlarda %71 oranında kadın müşterilerin satın alma işlemi gerçekleştirdiği tespit edilmiştir. Bu durum kozmetik, bebek ve kişisel sağlık genel kategorilerinin kadınlar tarafından daha fazla talep edilmesi ve belirtilen kategoriler için satın alma işlemi ile güven unsurunun daha yüksek olması ile ilgili olduğu kabul edilebilmektedir (Chou vd., 2015: 3).

Belirtilen müşteri portföyü çerçevesinde kadın müşterilerinin alışveriş sadakati ve aldıkları risk konusunda daha cesur oldukları konusunda araştırmalar mevcuttur (Rodgers ve Harris, 2003: 322; Garbarino ve Strahilevitz, 2004: 768). Bu durum modelde veri kaynağı olarak seçilen birden fazla alışveriş yapan sadık müşteri portföyü üzerinden öneri sistemi model oluşturma fikrine uygun ve tutarlı bir altyapı oluşturma imkânı sunmaktadır.

Araştırma modeli için belirlenen kriterler doğrultusunda birden fazla alışveriş yapan kullanıcıların tespit edilerek veri setini daha dar bir ölçekte yeniden oluşturulması sağlanmıştır. Belirli dönemler içerisinde birden fazla satın alma eylemi gerçekleştiren kullanıcılar için oluşturulacak öneri sistemi için temel alınacak matris yapısı bu yeniden ölçeklendirilmiş veri seti üzerinden gerçekleştirilecektir.

Zamansal faktörler açısından veri incelendiğinde hem yıllara hem de aylara göre orantısal bir dağılım içinde olduğu gözlemlenmektedir. 2013 – 2016 yılları arasındaki tekrarlanan satışların zaman içinde dalgalanmaması modelde kurgulanan zamansal kısıtlamalar

nedeniyle önemli bir faktör olarak belirlenmiştir. Bu nedenle belirtildiği üzere şirketin zaman aralığı içinde yapmış olduğu kampanyalarda oluşan satış kayıtları değerlendirmelere alınmamıştır. Bu durum oluşturulan her zaman dilimi içerisinde homojen bir yapının oluşmasını sağlamıştır.

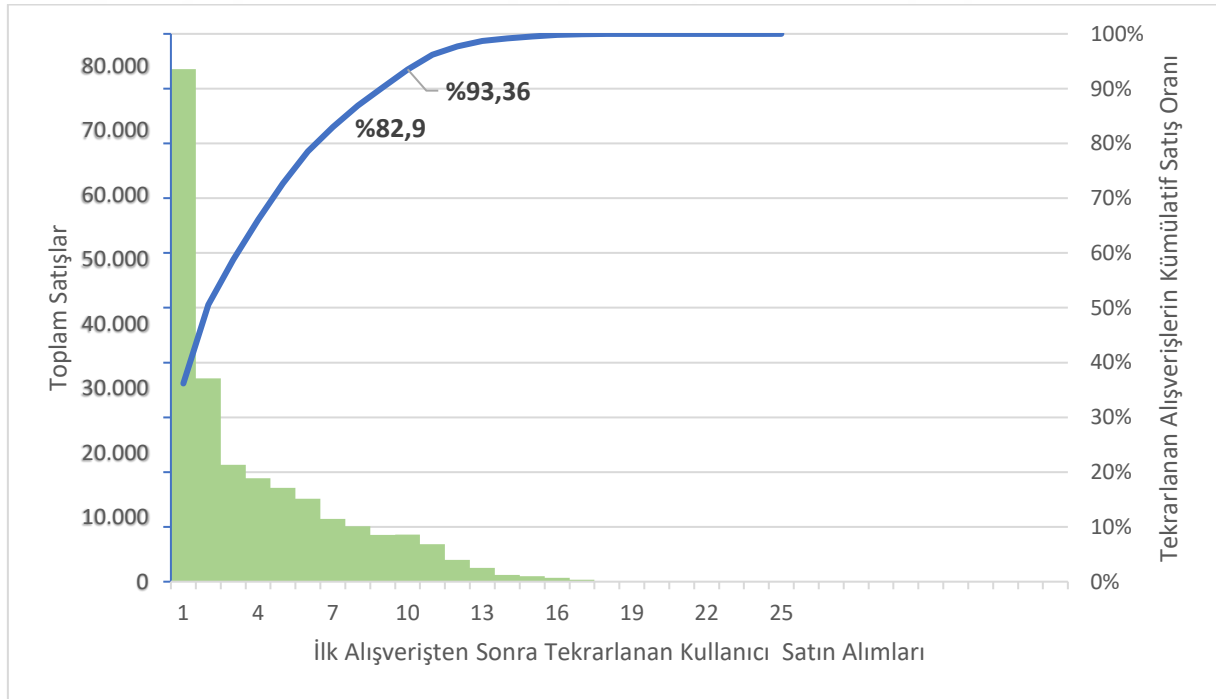
Verilerin araştırmanın konusunu teşkil eden kullanıcıların tekrar eden satın alma işlemlerinin öneri sistemleri üzerine etkisi üzerine yorumlama yapılabilmesi için veri setinde kullanıcıların ürün alımlarındaki tekrarlı siparişleri ile ilgili bir düzen olup olmadığından emin olunmalıdır. Bu durum öneri sisteminde kullanıcıların ürün beğenisi ve bu beğenin sistem üzerinden tekrar alışverişi ile tutarlı bir değerlendirmeye olanak sağlayacaktır. Belirtilen veri kriterlerinin modelin kurgusal ve düşünce yapısına uygun olup olmadığını incelemek için RFM analizinden faydalanılmıştır.

3.7.1 RFM Analizi

RFM (Recency (R), Frequency (F), and Monetary (M)) Analizi ilk olarak Bult ve Wansbeek (1995: 378) tarafından ortaya koyulmuş ve pazarlama araştırmalarında kullanılan bir yöntem haline getirilmiştir (Yeh vd., 2009: 5866). RFM İngilizce recency (yakınlık), frequency (sıklık) ve monetary (parasallık) bağıntısını ölçerek müşteri - satın alma ilişkisi hakkında geleceğe yönelik çıkarımlar oluşturmaya yardımcı olan bir araç olarak kabul edilir (Blattberg vd., 2008: 13). Yakınlık terimi müşteri davranışlarının belirlenen zaman dilimleri çerçevesinde gösterimi olarak değerlendirilebilmektedir. Pazarlama alanındaki çoğu karar verici en son zaman diliminde bulunan müşterilerin diğer zaman diliminde bulunanlara göre daha fazla tekrar alışveriş yapma potansiyeline sahip olduğunu düşünmektedirler. Sıklık ise yine belirlenen zaman dilimi içerisinde müşteri satın alma işleminin gösterimi ile ilgilidir. Sıklık kavramında oluşturulan varsayım en çok satın alma işlemi gerçekleştiren müşterinin tekrar satın alma eylemini gerçekleştirme yatkınlığının daha fazla olduğunun düşünülmesi üzerine kurgulanmıştır. Tutar ise belirlenen müşteri veya müşteriler için kümülatif harcama tutarını göstermektedir.

RFM analizi üç boyutta sınıflandırma veya sıralama yaparak belirli bir segmentasyonda en iyi %20'lik dilimi belirlemek için kullanılır. 20/80 prensibi olarak kabul edilen ve ölçülen değişkenin belirleyicisi olan ana etken içerisinde %20'lik kısım ölçümün %80'ini karşılar varsayımı üzerine kurgulanmıştır (Aggelis ve Christodoulakis, 2005: 1). Pazarlama müşteri sınıflandırmada ve segmentasyon oluşturmada benzer piramit modelleri kullanmakta; böylelikle müşterilerden elde edilen gelirin yapısı hakkında detaylı bir bilgi ağı

Tüm satışlar içerisinde kullanıcıların ilk alışverişlerinden sonra yaptıkları satın alma adetleri Şekil 3.3 ile incelendiği zaman tüm satışların %82,9'unu kapsayan satın alımları gerçekleştiren kullanıcıların ilk alışveriş sonrası satın alma işleminin 8 olduğu gözlemlenmiştir. %93,36'lık satın alım işlemleri ise en fazla 12 adet satın alma işlemi gerçekleştiren bir kitleyi temsil ediyor. Elde edilen bu veriler veri seti içerisindeki derinliği gözlemleyebilmemize yardımcı olmaktadır. Kullanıcı satın alma işlemlerinin tekrarlanması kullanıcı ürün arasındaki tercih hiyerarşisinin ortaya çıkarılmasına yardımcı olmakta ve tahmin kalitesini arttırmaktadır. Gerçekleştirilen bu analiz kapsamında düşünüldüğünde %80 barajını geçen tekrarlanan ürün satın alım sayısı ne kadar yüksek olursa veri setinin analize uygunluğunun o kadar iyi olduğu varsayımı ile açıklanabilmektedir. Burada %80'lik baraj için 8, %90'lık baraj içinse 12 rakamları analizlerin gerçekleştirilmesi için uygun değerler olarak yorumlanmıştır.



Şekil 3.3 Veri Seti İçerisindeki Tekrarlanan Toplam Satın Alımlar ile Toplam Satış/Kümülatif Toplam Satış

Elde edilen veriler ışığında her iki RFM analizinde uzun kuyruk etkisinin RFM analizlerinde etkili olduğu gözlenmektedir. Görüldüğü üzere veri seyrekliği söz konusu olup kuyruk kısmı bölümü için bir öneri sistemi modeli oluşturulması güç olacağının varsayılması muhtemeldir. Ancak grafiğin kuyruk kısmı yerine baş ve orta kısmı dikkate alındığında veri yoğunluğunun arttığı ve ürün çeşitliliği ile beraber satın alma kayıtlarının fazlalaştığı gözlemlenmektedir. Bu durum daha tutarlı bir tahmin modelinin oluşturulmasına olanak sağlayacaktır. Yapılan çalışmada bölümlere göre öneri sistem mekanizmasının verimliliği işlenecektir.

Bu analiz veri setindeki seyreklik durumu hakkında da bilgi vermektedir. Veri seti içerisinde ikili bir veri yapısı kullanılması ve 0 ile 1'den oluşan bir matris halinde analizlere dahil edilmesi analizlerin performansı üzerinde etkileri olabilmektedir. Her ne kadar sistemde birden fazla alışveriş yapan kullanıcılar analize dahil edilse de ürün çeşitliliğinin yüksek olması veri seti içinde satın almama verileri tarafından seyreklik problemi ortaya çıkarmaktadır. Bu durum analiz sonuçlarında elde edilen performans skorlarının beklenenden düşük seyretmesinin nedenlerinden biridir.

Satın Alma Matrisi

İşbirlikçi filtreleme modellerinde, kullanıcıların değerlendirmelerinin ve değerlendirmeleri yapan kullanıcıların tek bir matris altında toplanarak uygun modellerin oluşturulması esastır. Literatürde öneri sistemleri için üç veri modeli üzerinden değerlendirme matrisleri oluşturulmaktadır (Bodapati, Anand 2008: 78):

- Puanlama veri modeli
- İkili veri modeli
- Tekli veri modeli

Veri türünün doğrudan ve dolaylı olmasının yanında bu üç seçenek ile elde edilen puanlama ve değerlendirmeler öneri sistemlerinde sıklıkla kullanılan veri çeşitleridir (Schröder vd., 2011: 80). Örneğin, oluşturulması planlanan kullanıcı-ürün matrisinde değerlendirme verileri kullanıcıların ürün hakkında tek taraflı olarak yaptığı yorumlardan oluşabileceği gibi dolaylı ya da doğrudan elde edilen diğer veriler de dahil edilebilmektedir.

Modelin oluşturulması ve önemli bir aşama olan eğitim veri setinin algoritmaları ile eğitilmesi sürecinde, girdi olarak satın alma geçmişi matrisi kullanılmıştır. Bu matris her sütunun ürün olarak tanımlandığı ve her satırın kullanıcıların satın alma işlemi olarak gösterildiği bir yapıya sahiptir. Satın alma işlemi ile belirtilen ve belirlenmiş bir zaman dilimi içerisinde kullanıcı tarafından yapılan alışveriş temel veri olarak kabul edilmelidir. İşbirlikçi filtreleme yöntemlerinde oluşturulan modeller kurgulanırken genellikle matrisin satırını oluşturan bilgiler belirli bir kullanıcının ürün hakkındaki değerlendirmelerinden veya ürün üzerindeki davranışlarının gösterildiği sayısal değerlerden oluşmaktadır. Çalışmada da satırda zaman aralığında kullanıcı tarafından yapılan tüm satış bilgilerinin yer aldığı verileri içerir. Zaman kıstası araştırmanın konusu olan farklı modellerin zaman döngüleri içerisindeki performansının ölçümü konusu çerçevesinde değişkenlik göstermesi nedeniyle kurgu, algoritma yapılarına göre dinamik bir biçimde yeniden yapılandırma işlemine tabi tutulmaktadır. Tüm alternatif senaryolar dahilinde araştırmanın bundan sonraki kısmında G ile

gösterilecek satın alma matrisi sadece ikili değerleri içerecektir. Örneğin matris içerisindeki (i,k) satırı içinde k ürünü/ürün grubunu i 'inci oturumda en az bir kere almışsa 1 ile, almadıysa 0 ile temsil edilir. Araştırmada bahsi geçecek olan R değeri ise G matrisi içinde yer alan satır sayısını simgeleyecektir.

Veri Setinin Eğitilmesi ve Test Verilerinin Belirlenmesi

Veri seti için örneklem oluşturulurken, literatürde yer alan eğitim/test veri seti ayrımı için kullanılan yollardan faydalanılması düşünülmüştür. Bilindiği üzere eğitim veri seti, modeller için oluşturulan analizlerde parametre tespiti ve algoritma konfigürasyonlarının gerçekleştirilmesi ve nihayetinde modelin yapısının ortaya çıkarılması için oluşturulmaktadır. Test veri seti ise oluşturulan modelin değerlendirilmesinin yapılması ve eğitim veri setinde oluşturulan konfigürasyonların verimliliğini ve doğruluğunun tespiti için oluşturulmuştur.

Araştırmanın konusunda olduğu gibi tahmin sistemleri oluşturulmak üzere belirlenen algoritmik yapılarda model eğitim/test veri setleri üzerinden değerlendirme yapılır. Bununla ilgili genel uygulama; mevcut veri setinin, %80'i eğitim veri seti, %20'si ise test veri seti için kullanıldığı gözlemlenmektedir. Bu durum, veri örnekleme içerisinde rastgele olarak yerine yeniden koyulma işlemi olmadan %20'lik bir bahse konu verinin çekilerek test veri seti olarak tanımlanması ve geriye kalan %80'lik kısım içinse eğitim veri seti olarak kullanılması işlemi olarak özetlenir. Belirtilen 20/80 dağılımı çok keskin bir şekilde sınırlandırılması her zaman geçerli olmayıp ihtiyaç duyulması halinde eğitim veri seti için 2/3 veya 7/10 gibi oranların kullanımı da literatürde kabul edilmektedir. (Ricci vd., 2011: 43)

Modellerin çalıştırılabilmesi için bazı durumlarda veri örnekleme içerisinde test/eğitim veri grupları için farklılaştırma çalışmaları denenmiştir. Model için doğru veri yapısının tespiti için bu doğrultuda belirtilen veri gruplama işlemleri tekrarlanarak en optimum yapı elde edilmeye çalışılır. Orijinal veri örnekleme üzerinden oluşturulan eğitim veri grubu, model çalıştırılarak belirlenen test veri grubu üzerinde sonuçlar değerlendirilir. Eğer karar vericiler farklılaştırma düşüncesine sahipse grupların oluşturulmasında bazı metotlara başvurulabilirler.

Çapraz Sorgulama Yöntemi: Bu yöntem, veri bölme işlemlerinde en fazla kullanılanlardan biridir. Eldeki veri setini k farklı parçaya bölerek her parçayı bir doğrulama/test veri seti olarak kullanır (Aggarwal, 2016: 239). Veri seti kalan $k-1$ parça ile eğitilir. Bu eğitim ile elde edilen model k parçasına uygulanarak performans ölçümü gerçekleştirilir. Bu işlem her k parçası için ayrı ayrı gerçekleştirilir ve elde edilen performans değerleri ortalaması kıyaslanır. En iyi ortalamaya sahip değer belirlenerek modelin çalıştırılmasında uygulanır.

Çapraz doğrulama tekniği ise k-katlı çapraz sorgulama tekniğinin farklılaştırılmış halidir. Oluşturulan algoritma veri noktaları için her seferinde bir tanesini dışarıda bırakarak performans değerlemesi yapmakta bu değerleri kıyaslayarak bir veri gruplaması gerçekleştirmektedir.

Araştırmada belirli dönemleri kapsayan kullanıcı satın alma davranışları ile ilgili bir tahmin oluşturma modeli elde edilmek istenmesi nedeniyle, modelin uygulanacağı veri seti üzerinde literatürde yer alan farklı eğitim/test veri seti ayrımı varyasyonlar denenmiştir. Araştırma belirli zaman dilimleri içerisinde seçilen makine öğrenme metotlarının tekrarlanan hesaplamaları nedeniyle basit ve daha karmaşık grup ayrımı yapan yöntemler arasında sonucu etkileyen farklılıklar tespit edilmemiştir. Bu nedenle analizleri daha karmaşık hale getirmemek için temel ayırım yöntemlerinden olan veri setinin %80'inin eğitim veri seti, %20'sinin test veri seti olarak kullanımına karar verilmiştir. Ancak modellerde kullanılan zaman dilimlerinin sağlıklı çalışabilmesi için test veri setinde bulunan kullanıcıların eğitim veri seti içerisinde de yer almasına dikkat edilmiştir. Bu koşulun sağlanabilmesi için setler rastgele olarak tekrarlanarak uygunluğu değerlendirilmiştir.

Araştırma için eğitim ve test veri setlerinin oluşturulmasının ardından zaman dilimlerine göre tekrar bir alt eğitim veri setleri oluşturulmuştur. Bunun sebebi daha önce de belirtildiği üzere zaman dilimlerinin oluşturmanın analiz sonuçlarında yaratacağı etkilerin ölçümü için düşünülmüştür. Eğitim veri seti içerisinde referans günler tespit edilerek bu referans günlerden başlayarak belirlenen süreler içindeki satın alma işlemleri üzerinden elde edilen modeller, test veri setinde uygulanarak elde edilen değerlendirme skorlarının ortalaması ile sonuca gidilmiştir.

Araştırmada öncelikli olarak ilk etapta 45. 90. ve 365. referans noktaları üzerinden araştırmada kullanılan metotların verimliliği ölçülmüştür. Birden fazla kez satın alma işlemi gerçekleştiren her kullanıcı için satın alma işleminin olduğu gün referans günü olarak belirlenmiş ve tanımlanan referans noktaları gün olarak bu referans değere eklenmiştir. Böylelikle oluşan zaman aralığı içerisinde kullanıcı aynı ürünü tekrar alması durumunda havuz veri setine dahil edilmiştir. Örneğin eğitim veri seti içerisinde 1. referans günü olarak belirlenen tarihte alışveriş yapmış ve bu tarihi takip eden referans noktası olan 45 gün içerisinde de tekrar alışveriş yapmışsa, oluşturulan alt veri setinde bu kullanıcının alışverişlerinin yer alması uygun olduğu varsayılmıştır. Bu şekilde eğitim veri seti içerisindeki her referans günü ve belirlenen her referans noktası için bu alt veri setleri oluşturulmuş, algoritmalar çalıştırılarak eğitim ve test veri seti üzerinden değerlendirmeler yapılmıştır. Söz konusu veri ayrıştırma yönteminde kullanıcıların birden fazla kez satın aldıkları ürünler analizlere dahil edildiği de belirtilmelidir.

Test veri seti ise sabit tutularak tüm modellerin aynı veri üzerindeki etkisinin ölçümüne dikkat edilmiştir.

Alt veri setlerinin oluşturulması için geliştirilen bir yazılımla bu gruplandırma işlemleri gerçekleştirilerek Tablo 3.1’de gösterilen özelliklerde analizlere dahil edilmiştir;

Burada belirlenen gün sayısına ilk alışverişin yapıldığı günün de dahil edildiği not edilmelidir. Bunun yanında verinin temin edildiği şirketin özel olarak gerçekleştirdiği ve sonucu etkileyeceği düşünülen özel kampanya günleri belirtilen gün aralıklarına dahil edilmemiştir.

Tablo 3.1 Test Veri Seti İçinde Referans Günlere Göre Oluşturulan Alt Veri Setleri Özellikleri

Zaman Aralık Tanımı	Veri Seti Sayısı	Eğitim Veri Seti Büyüklüğü								
		Satış Kaydı			Ürün Çeşidi			Müşteri Çeşidi		
		Min.	Maks.	Ort.	Min.	Maks.	Ort.	Min.	Maks.	Ort.
45 Gün	754	11.058	15.365	13.452	74	91	82	171	214	189
90 Gün	431	29.751	36.522	34.583	112	229	173	302	548	463
365 Gün	299	73.245	75.872	74.942	651	1025	799	1494	2654	1851

İkinci etapta belirlenen referans noktalarının genişletilerek genel trendin yönünün belirlenmesi sağlanmıştır. 45. ve 90. günleri kısa dönem, 365 ve tüm veri seti üzerinde gerçekleştirilen çalışmaları uzun dönem olarak nitelendirilmiş ve bu süreler arasında bu modellerin oluşturduğu çıktıların değişiminin gözlemlenmesi amacıyla 30. günden başlayarak 540. güne kadar her günü referans noktası kabul ederek görselleştirilmesi amaçlanmıştır. Böylelikle kullanılan yöntemlerin zaman içindeki değişimi tespit edilmiş ve modellerin verimliliği bu trende göre yorumlanmıştır.

Değerlendirme Ölçümleri

Araştırmada, bahsi geçen kriterler ve sınırlamalar doğrultusunda elde edilen algoritmalar belirli bir ürün listesini oluşturarak bu listeleri değerlendirme için modele yeniden dahil etmektedirler. Test veri setinde elde edilen sonuçları değerlendirebilmek içinse literatürde yaygın olarak kesinlik (precision), hassasiyet (recall) ve F1-skor değerleri kullanılmaktadır.(Herlocker, Jonathan vd., 2004: 27; Gunawardana ve Shani, 2009: 2945)

Kesinlik ve hassasiyet değerleri tespiti için hata matrisi (confusion matrix) kullanılır (Townsend, 1971: 40). İkili bir karar verme probleminde sınıflandırma genellikle pozitif ve negatif olarak gerçekleştirilir. Bu sınıflandırmada verilen kararlar belirtilen matris çerçevesinde görselleştirilerek, değerlendirme yapma aşamasında yardımcı bir araç olarak kullanılır. Hata

matrisinde dört kategori bulunur. Doğru pozitifler (DP), yapılan gözlem sonucunda doğru olarak işaretlenmiş pozitif değerleri simgeler. Yanlış pozitifler (YP) ise hatalı olarak pozitif işaretlenmiş negatif değerleri gösterir. Öte yandan doğru negatifler (DN) negatif değerlerin doğru bir şekilde negatif olarak, yanlış negatiflerin (YN) pozitif olmasına rağmen hatalı olarak negatif değerlendirilen çıkarımları gösterir;

Model değerlendirmelerinde kullanılan hata matrisinin örneği Tablo 3.2’de gösterilmiştir. Araştırma çerçevesinde oluşturulan model değerlendirmesinde, kullanıcı bir kez alışveriş yaptıktan sonra belirtilen zaman dilimi içerisinde modelin önerdiği ürünü tekrar alması durumu DP, almaması durumu YP olarak, değerlendirme sonucunda almayacağı düşünülen ve alınması durumu YN, almaması durumu ise DN olarak açıklanabilir.

Tablo 3.2 Model Değerlendirmesinde Kullanılan Hata Matrisi Tablosu

		Gerçekleşen	
		Gerçek Pozitif	Gerçek Negatif
Tahmin	Pozitif Değerlendirilen	Önerildi / Satın Alındı	Önerildi / Satın alınmadı
	Negatif Değerlendirilen	Önerilmedi / Satın Alındı	Önerilmedi / Satın Alınmadı

Matris üzerinden elde edilen değerler ile kesinlik ve hassasiyet değerleri temelde şu şekilde formüle edilir (Gunawardana ve Shani, 2009: 2945):

$$hassasiyet = \frac{DP}{DP + YN} = \frac{\text{Önerildi ve Satın Alındı}}{\text{Önerildi ve Satın Alındı} + \text{Önerilmedi ve Satın Alındı}}$$

$$kesinlik = \frac{DP}{DP + YP} = \frac{\text{Önerildi ve Satın Alındı}}{\text{Önerildi ve Satın Alındı} + \text{Önerildi ve Satın Alınmadı}}$$

Araştırmada, eğitim veri seti içerisinde referans günü içerisinde satın alma işlemini gerçekleştiren ve her müşteri için hedef ürünlerle birlikte değerlendirilmiştir. Belirlenen referans günü ve zaman aralığı doğrultusunda çalıştırılan modellerden test veri seti üzerinde çalıştırılması ile kesinlik, hassasiyet ve F1 skorlar elde edilerek sonuçlara ulaşılmıştır. Hesaplamalar her referans günü için tekrarlanarak ilgili skorların elde edimi tekrarlanmıştır.

Tablo 3.2 ve formüllerden anlaşılacağı üzere kesinlik değeri gerçekleşen sonuçların tüm sonuçlara olan oranı ise algoritma tarafından doğru sınıflandırılan sonuçların oranını belirtmektedir.

Modellerde, belirlenen müşteri m ve referans gün d için N adet ürünün kesinlik, hassasiyet ve F1 skoru şu şekilde belirlenmiştir (Pradel vd., 2011: 380);

$$kes_{m,d}@N = \frac{\text{card}(R_{m,d}@N \cap T_{u,d})}{N}$$

$$has_{m,d}@N = \frac{\text{card}(R_{m,d}@N \cap T_{u,d})}{\text{card}(T_{u,d})}$$

Formüllerde $T_{m,d}$ hedef veri setini ve $R_{m,d}@N$ ise önerilen N ürünü simgelemektedir. Bu doğrultuda F1 skoru;

$$F1_{m,d}@N = \frac{2rec_{m,d}@N \cdot prec_{u,d}@N}{rec_{n,d}@N + prec_{n,d}@N}$$

olarak belirlenecektir. Öneri algoritmalarının üreteceği ürün sayısı veri kısıtlılığı ve karar vericilerin tercihi doğrultusunda 5 olarak tespit edilmiştir. Formüllerde $N = 5$ sabiti geçerli olacaktır.

Araştırma için kullanılacak değerlendirme skorunu ise her referans günü için kullanıcıların ortalaması alınıp aşağıdaki hesaplama ile elde edilecektir (Pradel vd., 2011: 380);

$$kes_d@N = \frac{\sum_{k_{test,d}} prec_{k,d}@N}{N_d}$$

$$has_d@N = \frac{\sum_{k_{test,d}} rec_{k,d}@N}{N_d}$$

$$per f@N = \frac{\sum_{d, ref.} perf_d@N}{d \text{ ref. gün sayısı}} \quad f \in \{kes, has, f1\}$$

Belirtilen formüller ile yola çıkılarak bahse konu performans ölçüm kriterleri ile bir skor elde edilmesi sağlanmıştır. Böylelikle test veri seti yardımıyla öneri modellerinin kıyaslaması yapılmıştır.

3.8. Araştırmada Kullanılan Yöntemler

Belirtildiği üzere öneri sistemleri için model oluştururken, işbirlikçi filtreleme yöntemlerinin uygulanmasına karar verilmiştir. Bu doğrultuda belirlenen zaman dilimleri içerisinde işbirlikçi filtreleme modellerinde sıklıkla kullanılan üç yöntem tespit edilerek veri seti için tahmin oluşturma modellerinin ortaya çıkarılması amaçlanmıştır. Bu seçim yapılırken farklı veri tipi kombinasyonları ile oluşturulan ve genelde değerlendirme matrisleri ile sonuca ulaşılan algoritma ve yöntemlerden faydalanılmıştır. Bu nedenle hafıza temelli, model temelli ve birliktelik kuralları başlığı altından birer yöntem seçimi tercih edilmiştir. Söz konusu kriterler çerçevesinde çevrimdışı hesaplama yöntemlerine uygun olarak ürün (hafıza) temelli komşuluk modeli kapsamında benzerlik/uzaklık ölçümü, kullanıcı (model) temelli modellerde

kullanılan boyut indirgeme yöntemleri ile birliktelik kuralları belirlenen kriterlerle ele alınarak farklı modellerin oluşturulmasına ve tahmin performansının ortaya çıkarılmasına karar verilmiştir.

Ürün temelli işbirlikçi filtreleme modellerinden sık kullanılan yöntemlerden biri benzerlik/uzaklık ölçümüdür (Aggarwal, 2016: 37). Benzerlik ve benzeşmezlik (uzaklık) ölçümü sınıflandırma, kümeleme gibi belirli bir desen/düzen keşfi için gerçekleştirilen analizlerde önemli bir rol üstlenebilmektedir. Problemin çözümü ya da modelin oluşturulması aşamasında yapılacak tercihler performans üzerinde önemli bir etki farkını yaratabilmesi bu tür bir yöntemin kullanıldığı çalışmalarda önemli karar noktalarından biri haline gelmektedir. Araştırmada literatürde sıklıkla kullanılan ve çoğu benzer çalışmada da yer verilen kosinüs (cosine) benzerliği ile daha çok belirlenmiş veri tipi için kullanımı gözlemlenen ve spesifik durumlar için tercih edilen Jaccard benzerliği üzerinden analizlerin gerçekleştirilmesi amaçlanmıştır.

Kullanıcı temelli işbirlikçi filtreleme modelinde ise ürün temelli yaklaşımın aksine tüm matrisin hafızada tutularak işlemlerin bu veri üzerinden yapılmasına ihtiyaç duymayacak modeller tercih edilmiştir. Kümeleme, sınıflandırma, örtük sınıf analizi, Markov karar süreci ve matris ayrıştırma yöntemleri gibi metotlar bu model içerisinde değerlendirilmektedir (Aggarwal, 2016: 73). Araştırmada yine literatürde yaygın olarak kullanılan ve öneri sistemleri içerisinde verimi sonuçlar elde edilebilen boyutsal indirgeme tekniklerinden TDA (Tekel Değer Ayrışımı), veri tipine özgü analizlerde kullanımı uygun olan NMF (Negatif Olmayan Matris Ayrışımı) modeline yer verilmiştir.

Öneri sistemlerinde kullanılan başka bir yöntem ise birliktelik kuralı öğrenmesidir. Bu yöntem özellikle büyük veri tabanlarında ilişkisel örüntülerin çıkarımı konusunda başarılı sonuçlar vermektedir. Temel algoritmalar arasında yer alan FP-Growth ve veri tipine uygun PSO modeli analize dahil edilmiştir. FP-Growth, apriori algoritmasını temel alması ve büyük veri setleri için uygun olması nedeni ile tercih edilirken, PSO özelleştirilerek ikili veriler için kullanılabilmesi seçim aşamasında önemli bir kriter olmuştur.

Belirtilen metotlar veri seti üzerinde uygulaması yapılırken veri setindeki satın alma işlemlerinin zamansal dağılımı da karşılaştırmaya konu olmuştur. Uygulamada belirli bir zaman döngü içerisinde hesaplamalar gerçekleştirilecektir. Bu döngü belirtilen tüm metotlar için uygulanarak elde edilen skorların ortalaması hesaplanması planlanmaktadır. Döngü şu şekilde sağlanmıştır:

- Veri seti içerisinde yer alan en az bir satın alma işleminin gerçekleştiği günler referans gün olarak tespit edilmiştir.

- Analizin ilk aşamasında kısa orta ve uzun dönem performansların karşılaştırılması için referans noktaları belirlenmiştir. Bunlar 45, 90 ve 365 olarak seçilmiştir.
- Belirlenen referans günlerine uygulanacak referans gün noktaları kadar gün eklenerek alt veri setleri oluşturulmuştur. Bu veri setleri üzerinde modeller oluşturularak model performansları ölçülmüştür.
- Belirlenen referans noktaları üzerinden elde edilen skorların devamlılığının tespit edilebilmesi için 30 ile 540 gün arasındaki her gün yeni referans noktaları olarak analize dahil edilmiş ve skor performanslarının grafikleştirilmesi sağlanmıştır.

Belirtilen yöntemlerde, çıktı olarak belirli sayıda (N) bir ürün listesini oluşturarak (Araştırmada bu sayının 5 olarak kararlaştırıldığı belirtilmişti) satın alma geçmişinin, ikili sütun vektörü olarak $G \in [0,1]^{1 \times M}$ ile gösterimi sağlanmıştır. Örnek olarak k 'nıncı satırdaki 1 değeri k ürünün M müşterisi tarafından satın alındığını göstermektedir. G tek bir müşteri için oluşturulmuş satın alma geçmişi matrisidir. Bahsi geçtiği üzere satın alma geçmişleri belirlenen her zaman aralığı için yeniden oluşturulmuş ve analizlerde bu şekilde yer verilmiştir.

Benzerlik Ölçümü Uygulaması

İkinci bölümde bahsedildiği üzere ürün temelli komşuluk modeli, kullanıcıların bir ürün havuzu içerisindeki tercih ve davranışlarının, bir değerlendirme matrisi üzerinden modelleştirilmesi üzerine kurgulanmıştır. Bu tercih ve davranışlar puanlama, yorum vb. gibi doğrudan görüşler olabileceği gibi dolaylı olarak ürün hakkındaki tutumunu belirleyebileceğimiz satın alma eylemi, ürün sayfalarında gezinme süreleri vb. izler bu modelde kullanılabilir. Örtülü faktör modellemelerinde verinin çok taraflı olarak işlenebilmesi ve bu farklı açıların değerlendirilebilmesinden ötürü çoğu öneri sistemlerinde sıklıkla kullanılan bir metot olması, araştırmada da öncelikli olarak bu modelin denenmesinde öncülük etmiştir.

Uygulama Modeli

Bu modelin uygulaması esnasında farklı benzerlik ölçüm araçları üzerinde durulmuş ve literatürdeki uygulamalar incelenmiştir. Pearson, Kosinüs gibi literatürde sık kullanılan benzerlik ölçüm metotlarının yanı sıra özel durumlarda daha sık kullanımına rastlanan PIP, HSNM, Jaccard gibi ölçümlerin veri yapısına uygunluğu değerlendirilmiştir.

Bunların arasından ikili ve tekli veri tipleri için sıklıkla kullanılan Jaccard benzerlik ölçümü, satın alma matrisi üzerinden oluşturulması planlanan modele uygun olduğu kabul edilerek araştırmada yer verilmiştir. Jaccard benzerlik ölçümü sınırlı örneklem setleri arasında

örneklem boyutlarının birleşim büyüklüğünün kesişim büyüklüğüne bölümü ile elde edilen bir değer ölçüm aracıdır (Jaccard, 1908: 223). İkili yapıya sahip veri setlerinde değerlerin karakteristik olarak birbiri ile örtüşmelerinin ölçümünü yaparak işlem gerçekleştirilmektedir (Choi vd., 2010: 43). Araştırmada satın alma matrisleri üzerinden gerçekleştirilen analizlerde, eğer ürün iki müşteri tarafından da satın alındığı değerler kesişim kümesi olarak kabul edilmiş, diğer olasılıklar ise birleşim kümesi içerisinde bırakılarak analize dahil edilir.

$$uz_j(k, l) = 1 - J(k, l) = \frac{P_{.kP.l}^T}{P_{.k}^T P_{.k} - P_{.l}^T P_{.l}}$$

Jaccard benzerlik uzaklık ölçümünde k ve l ürün davranışlarının paylaştığı örüntüler kapsamında yukarıdaki formül oluşturulmuştur. Denklemden $P.k$ P 'nin k 'nıncı sütunu, P^T ise t matrisinin transpozesidir. Eğitim süresinin çıktısı (k, l) 'inci girdinin sim($k'l$) ye eşit olduğu G ürünlerinin arasındaki benzerlik matrisi olarak tanımlanır.

Tablo 3.3 Benzerlik Ölçüm Model Algoritması

```

Fonksiyon BenzerlikÖlçümü(){
  Her  $j \rightarrow l$ 'den  $m$  değerine
    Her  $i \rightarrow l$ 'den  $n$  değerine
      Eğer  $i \neq j$ 
         $M_{i,j} \rightarrow \text{benzerlik}(R_{*i}, R_{*j})$ 
      Değilse
         $M_{i,j} \rightarrow 0$ 
    Her  $i \rightarrow l$ 'den  $n$  değerine
      Eğer  $M_{i,j} \neq \text{belirlenen } N \text{ eni değer listesinde}(M_{*}, j)$ 
         $M_{i,j} \rightarrow 0$ 
  Sonuç ( $M$ )
}
```

Öte yandan ikili veri setleri üzerinde uygulanan kosinüs (cosine) benzerliği de popüler ölçüm araçları arasında yer alır. Araştırmada performans karşılaştırması yapmak üzere bu yöntemin de analizlere dahil edilmesi uygun görülmüştür. Kosinüs benzerlik ölçümü ile, iki sıfırdan farklı vektörün aralarındaki açısal değerinin kosinüsü hesaplanarak vektörlerin benzerliği bulunur.

Bahsedilen satın alma geçmiş matrisleri üzerinden aşağıdaki adımlar takip edilerek hafıza temelli için öneri modelleri oluşturulmuştur. Hesaplama için temel mekanizma Tablo 3.3'te gösterilmiştir.

Belirlenen döngü ilk aşamada belirlenen referans gün noktalarına uygulanmış, daha sonra da elde edilen değerlerin tutarlı bir sonuç gösterip göstermediğini anlamak ve grafiksel olarak ortaya çıkarmak için 30 - 570 arasındaki her günü referans noktası olarak tanımlanmış ve üç nokta yerine 540 nokta ile analizler tekrarlanmıştır. Bu doğrultuda eğitim veri seti üzerinde belirlenen modelin test veri seti üzerinde uygulanabilir sonuçlar elde edilmiştir.

Metrik Değerlendirme

Ürün temelli öneri sistemlerinin değerlendirmesinin yapılabilmesi için daha önce bahsedilen kesinlik, hassasiyet ve F1 skor metrik ölçüm yöntemlerinden yararlanılmıştır. Belirtildiği gibi uygulama, kullanıcıların belirli dönemler içerisinde kullanıcıların satın alma davranışlarını dikkate alarak oluşturulacak modeller üzerine kurgulanmıştır. Benzerlik ölçümün gerçekleştirildiği bu çalışmada literatürde yer alan standart yöntemler kullanılmış veri seti üzerinde başka bir değişikliğe gidilerek farklılaştırma çalışmaları yapılmamıştır. Oluşturulan algoritmalar sonucu hata matrisi ortalamaları EK-1’de gösterilmiştir. Belirtilen dönemler içerisinde uygulanan ürün temelli öneri sistemi metodu skorları şu şekilde elde edilmiştir;

Tablo 3.4 Jaccard ve Kosinüs Benzerlik Ölçümü 45 Günlük Metrik Değerlendirmesi

45 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
Jaccard Benzerlik Ölçümü	11,15	12,89	11,96
Kosinüs Benzerlik Ölçümü	13,02	14,30	13,63

Tablo 3.5 Jaccard ve Kosinüs Benzerlik Ölçümü 90 Günlük Metrik Değerlendirmesi

90 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
Jaccard Benzerlik Ölçümü	10,82	12,51	11,60
Kosinüs Benzerlik Ölçümü	12,70	13,72	13,19

Tablo 3.6 Jaccard ve Kosinüs Benzerlik Ölçümü 365 Günlük Metrik Değerlendirmesi

365 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
Jaccard Benzerlik Ölçümü	8,96	10,21	9,54
Kosinüs Benzerlik Ölçümü	9,48	11,72	10,48

Tablo 3.7 Jaccard ve Kosinüs Benzerlik Ölçümü Metrik Değerlendirmesi

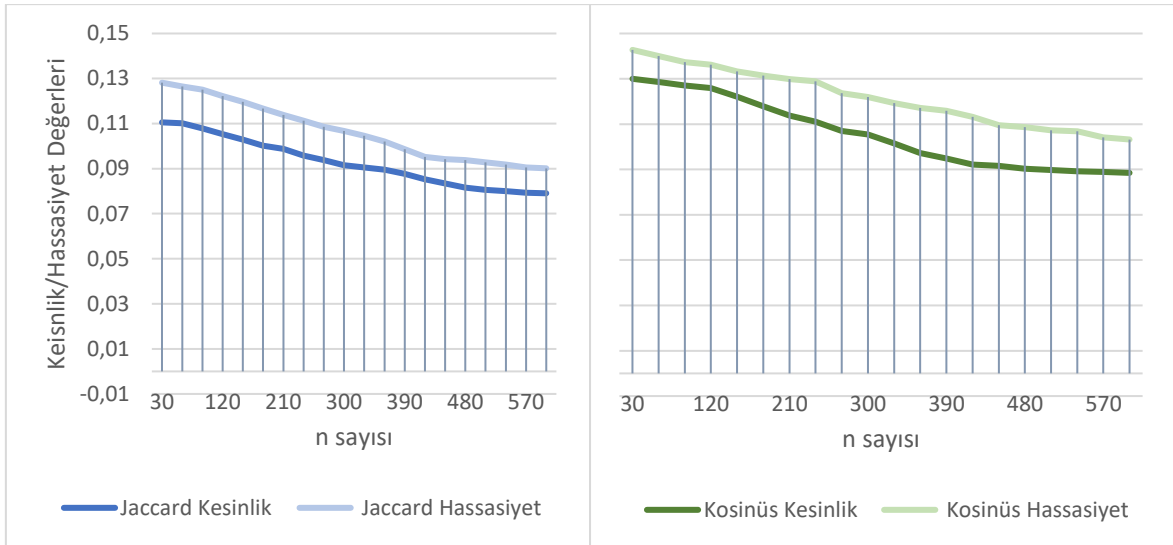
Tüm Veri Seti Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
Jaccard Benzerlik Ölçümü	7,90	9,01	8,42
Kosinüs Benzerlik Ölçümü	8,84	10,32	9,52

Araştırma için kullanılan veri seti, belirtilen şartlar ve kıstaslar çerçevesinde 0,0884 ile 0,1430 arasında tahmin öneri sistemi performans skoru elde edebilmektedir. Hem kısa dönem olarak kabul edilen 45 (Tablo 3.4) ve 90 (Tablo 3.5) günlük zaman dilimlerinde hem de 365

(Tablo 3.6) gün ve tüm veri setinde (Tablo 3.7) Kosinüs benzerlik ölçümünün Jaccard benzerlik ölçümüne kıyasla daha iyi sonuçlar verdiği gözlemlenmektedir. Kosinüs, 45 günlük zaman dilimi için kesinlik ve hassasiyet değerleri sırasıyla 0,1302 ve 0,1430 iken, Jaccard benzerlik ölçümü 0,1115 ile 0,1289 olarak belirlenmiştir. Bu durum 90 günlük zaman dilimi için de benzer bir durum ortaya koymaktadır. Bu zaman dilimi için kosinüs kesinlik ve hassasiyet verileri 0,1270 ile 0,1372 olarak elde edilmiş, Jaccard benzerlik ölçümünde ise 0,1082 ve 0,1251 olmuştur. Uzun dönem olarak kabul edilen 365 gün ve tüm veri seti ile yapılan uygulamada değerler 0,0896'ya kadar düşmüştür. Kosinüs benzerlik ölçümü 365 gün için kesinlik ve hassasiyet değerleri 0,0948 ve 0,1172; tüm veri seti kullanılarak yapılan çalışmada ise 0,0884 ile 0,1032 olarak sonuçlar ortaya çıkmıştır. Jaccard yönteminde ise değerler daha da düşmektedir. 365 gün için kesinlik ve hassasiyet değerleri 0,0896 ile 0,1021 olurken, tüm veri seti üzerinde yapılan çalışmada 0,0790 ile 0,0901 gibi oldukça düşük değerlere gelmiştir. Bu değerlere bakılarak kosinüs benzerlik ölçümünün dört zaman dilimi için gerçekleştirilen veri setleri için daha başarılı bir performans sergilediği söylenebilmektedir.

Elde edilen değerlerdeki değişimlerin belirli bir dönem için mi yoksa genele yayılabilecek bir çıkarım mı olduğunu tespit edebilmek adına 30 günlük zaman diliminden başlayarak 10'ar gün artırımla 30 ile 540 gün arasında yeni zaman dilimi noktaları belirlenerek çalışma genişletilmiştir. Şekil 3.1'de de gözlemlendiği gibi zaman dilimlerinin artması ile birlikte her iki yöntemde de kesinlik ve hassasiyet değerlerinde ciddi azalışlar görülmektedir. Kosinüs benzerlik ölçümü için kesinlik ve hassasiyet değerleri açısından 30 günlük zaman dilimi ile tüm zaman dilimi arasındaki azalma yüzdesel olarak %32,10 ve %27,83 gibi değerlere ulaşmakta, bu değer Jaccard benzerlik ölçümünde ise %29,14 ve %30,01 olarak gözlemlenmiştir.

Kesinlik değerinin tüm sonuçlar içerisindeki, hassasiyet değerinin ise yapılan tahminler arasındaki başarı oranını gösterdiğini düşünerek yorumlandığı zaman her iki skorda da iyi sonuçlarla karşılaşılmamaktadır. Ancak bu skorlar tek başına bir anlam ifade etmemekte diğer yöntemlerle yapılan kıyaslamalarla yorumlanabilir hale gelmektedir. Araştırma öncesi herhangi bir referans değere göre kıyaslama yapılmadığı vurgulanarak yöntemin tek başına başarılı olup olmayacağı konusunda bir yoruma başvurulmamaktadır.



Şekil 3.4 Jaccard ve Kosinüs Benzerlik Ölçümleri Kesinlik/Hassasiyet Değerlerinin Grafiksel Gösterimi

Matris Çarpanlara Ayırma Metodu

Matris Ayırıştırma metodu, kullanıcı puanlama ve değerlendirme yöntemleriyle oluşturulan matrisler üzerinden elde edilen öneri oluşturma modelleme yöntemi oldukça yaygın bir biçimde kullanılmaktadır (Koren vd., 2009: 43). Bu metodun satın alma değerler matrisinde uygulanması için iki farklı boyut indirgeme tekniği kullanılması planlanmaktadır.

Araştırma Modeli

Bu yöntemlerde veri seti boyutunda bir matris olduğu varsayıldığı ve kullanıcıların satın alma davranışlarına uygun olarak 0 ve 1 rakamları ile gösterilen değerlerden oluştuğu algoritmalar çalışmadan ön tanımlanmıştır. Bu nedenle boyutsal indirgeme teknikleri kullanıcı ve ürün verileri üzerinden analizin sonucunu etkileyecek bilgi kaybını önleyecek en iyi optimizasyonların gerçekleştirilmesi ve denemelerin yapılarak sonuç üzerinden yorum yapılmasına dikkat edilmiştir.

Çarpanlara ayırma işleminde TDA metodunun kullanımı uygun bulunmuştur. Ürün için öncelikli olarak TDA yardımıyla elde edilen Frobenius norm doğrultusunda P için en iyi sıralama yaklaşımı ile ayırıştırma işlemi gerçekleştirilecektir (Sarwar, B. vd., 2000: 8).

Bunun yanında yine popüler bir matris ayırıştırma yöntemi olan NMF yöntemi de araştırma kapsamında değerlendirilmiştir. Bu yöntem indirgenen boyuttaki değerlerin negatif değeri kapsamaması nedeniyle diğer metotlardan ayrılmaktadır. Yüksek boyuttan düşük boyuta indirgenen matriste boyutlarda negatif elemanlara yer verilmez ve tüm elemanların pozitif değere sahip olması sağlanır.

$$U * V = \underset{*}{\operatorname{argmin}} \|P - UV\|_F^2$$

$$U \in \mathbb{R}_+^{R \times r} \quad V \in \mathbb{R}_+^{R \times c}$$

Vektör matrisi V^* üzerinden satın alma geçmişleri G için uygun bir parametre vektörü elde edimi ile sonuca ulaşılması planlanmaktadır. Yukarıda belirtilen her iki yöntemde de V^* vektör matrisi olarak kabul edilebilir. Bu matris $\mathbb{R}^r$ içerisindeki her k ürününü yansıtabilir. Böylece V^* içerisinde k sütünü vektörel alanda gösterimini sağlayacak U^* satın alma geçmişleri değeri matrisinin V^* satırlarından oluşan lineer birleşimini simgeleyebilecektir. Bu yöntemle eğitimi seti üzerinden V^* matrisi sayesinde test veri seti üzerinde tahmin modeli uygulanabilmektedir.

TDA ve NMF ile öneri sistemleri ile performans skorları elde edebilmek için iki farklı yol kullanılabilir. Kullanıcı ürün arasındaki örtülü ilişkinin tespiti için tek boyuta indirgenerek ortaya çıkarılabilir ya da düşük-boyut gösterimi oluşturularak benzerlik ölçümleri ya da ANN metodu ile sonuca gidilebilmektedir. Araştırmada tek boyut üzerinden elde edilecek bilgiler kullanılmış sonuç olarak aşağıdaki işlemler her iki boyut indirgeme tekniği için de kullanılarak işlemler gerçekleştirilmiştir;

- S matrisini k boyutuna indirgemek için $\mathbb{R}^{R \times r}$ ve $\mathbb{R}^{R \times c}$ faktörize ederek U, S ve V boyutları elde edildi.
- S_k indirgenmiş matrisinin karekökü olarak $S_k^{1/2}$ elde edildi
- Meydana gelen iki matris hesaplaması: $U_k S_k^{1/2}$ ve $S_k^{1/2} V_k$

Oluşan matrisler öneri skorun oluşturulmasında kullanılabilir durumdadır. Kullanıcı c için : $U_k S_k^{1/2}$ ile ürün p için $S_k^{1/2} V_k$ matrislerinin çarpımına kullanıcı ortalaması eklenerek bir tahmin değeri oluşturulmuştur.

$$C_{tahm} = U_k S_k^{1/2} \cdot S_k^{1/2} V_k$$

Öneri sistemi modeline adapte edilmesi durumunda hesaplama şu şekilde gerçekleştirilecektir:

- Her referans gününden başlayarak referans noktası gün sayısını ekleyerek analizin gerçekleştirileceği tarih aralığını bul:
 - Veri seti içerisinde elde edilen tarih aralığı için verileri seç ve her tarih için;
 - Belirtilen boyutsal indirgeme yöntemiyle öneri değerini oluştur. r değeri ön tanımlı olarak belirtilmeden kullanılan sistem tarafından her veri seti için optimum değerin otomatik belirlenmesi sağlanmıştır.
 - Elde edilen indirgenmiş boyutta veri seti benzerlik ölçümü yöntemi ile yeniden model oluşumu sağlanır. Model üzerinden oluşturulan skorların ortalamasını alarak referans gün noktası için ortalama değeri ortaya çıkar.

Elde edilen değerler test veri seti üzerinde değerlendirilerek öneri skorları oluşturulmuştur. Araştırmada ön tanımlı olarak herhangi bir boyut sayısı değeri tanımlanmamış, algoritmanın en optimum değeri her veri seti için kendi belirlemesi sağlanmıştır.

Metrik Değerlendirme

Bahse konu matris ayırıştırma yöntemleri ile eğitim veri seti üzerinde gerçekleştirilen model oluşturma işlemi sonucunda test verileri üzerinde daha önce belirlenen zaman dilimleri için oluşturulmuş algoritmalar çalıştırılarak ortalama kesinlik hassasiyet ve F1 skor değerleri elde edilmiştir. Bu skorların elde ediminde kullanılan hata matrislerinin ortalama değerleri EK-2’de gösterilmiştir.

Tablo 3.8 TDA ve NMF Metotlarının 45 Günlük Metrik Değerlendirmesi

45 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
TDA	17,07	18,79	17,89
NMF	15,47	17,05	16,22

Tablo 3.9 TDA ve NMF Metotlarının 90 Günlük Metrik Değerlendirmesi

90 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
TDA	16,71	18,28	17,46
NMF	14,93	16,00	15,45

Tablo 3.10 TDA ve NMF Metotlarının 365 Günlük Metrik Değerlendirmesi

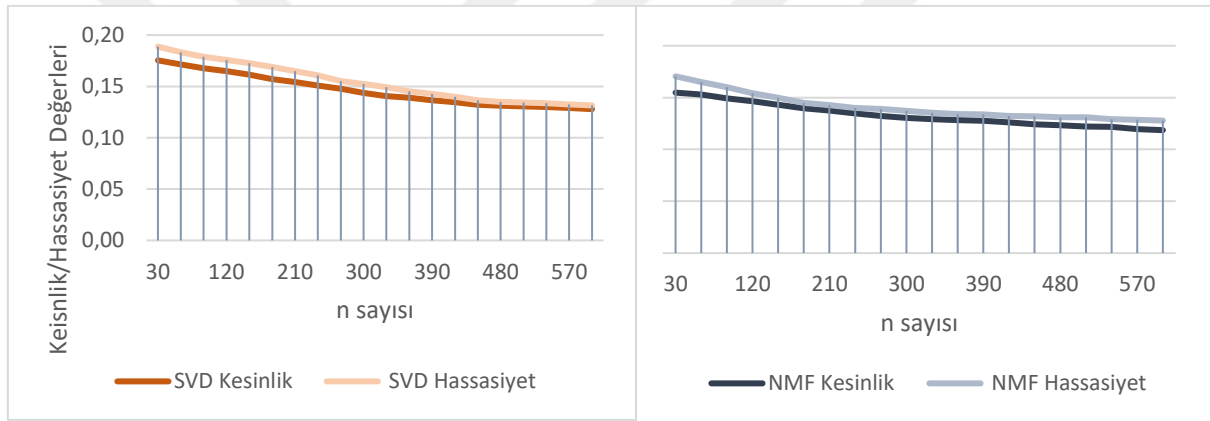
365 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
TDA	13,89	14,49	14,18
NMF	12,81	13,58	13,18

Tablo 3.11 TDA ve NMF Metotlarının Tüm Test Veri Seti Üzerindeki Metrik Değerlendirmesi

Tüm Veri Seti Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
TDA	12,78	13,15	12,96
NMF	11,85	12,48	12,16

Daha önceden tanımlanmış olan 45, 90, ve 365 günlük zaman dilimleri için boyutsal ayırıştırma teknikleri gerçekleştirilmiş ve vektör tabanlı benzerlik ölçüm yöntemlerinden kosinüs benzerliği uygulanarak Tablo 3.8, Tablo 3.9, Tablo 3.10 ve Tablo 3.11 ile sonuçlar ortaya konulmuştur. Yapılan çalışmada 0,1828 ile 0,1185 arasında gösterge değerler elde edilmiştir.

TDA boyutsal ayrıştırma tekniği kullanılan durumlarda kısa dönem olarak değerlendirilen 45 günlük dönemde kesinlik ve hassasiyet değerleri açısından 0,1707 ve 0,1879 değerleri elde edilmiş, bir değer kısa dönem değeri olan 90 günlük zaman dilimi içinse 0,1671 ve 0,1828 olarak belirlenmiştir. Ancak uzun dönem verileri olan 365 günlük zaman dilimi ve tüm veri seti üzerinde yapılan çalışmada kesinlik ve hassasiyet değerleri sırasıyla 0,1389-0,1278 ile 0,1281-0,1358 sonuçlarını ortaya koymaktadır. Değerlerden de anlaşılacağı üzere kısa dönemde hem kesinlik hem de hassasiyet değerleri için daha iyi performans skorları elde edilirken bu değerler arasındaki azalma kesinlik değeri için %25,13; hassasiyet değeri içinse %26,28'dir. Kısa ve uzun dönem arasındaki fark, TDA yöntemine göre nispeten daha azdır. Ancak TDA hem kısa dönemde hem de uzun dönemde daha iyi performans skorları elde etmektedir.



Şekil 3.5 TDA ve NMF Metotları Kesinlik/Hassasiyet Değerlerinin Grafiksel Gösterimi

TDA ve NMF kesinlik ve hassasiyet sonuçlarındaki değişimin daha detaylı gözlemlenebilmesi için 10'ar günlük zaman dilimleri için tekrar hesaplaması yapılmış ve bu dilimler arasındaki değişimin nasıl seyrettiği konusunda Şekil 3.5'de gösterildiği üzere gözlemlenmiştir. Her iki yöntem de dar zaman dilimleri ile geniş zaman dilimleri arasında kesinlik veya hassasiyet ölçümlerinde keskin değişimler gözlemlenmemiştir. Referans gün sayısı 30 olarak belirlenen değerler ile tüm veri seti içinde yapılan değerler arasında F1 skorlarındaki azalama %32 ve %23 oranında değişim olduğu anlaşılmıştır. Bu değerlerden de anlaşılacağı üzere TDA kısa dönemde daha iyi bir performans skoru elde etmekte ancak aradaki fark uzun dönemde oldukça azalmaktadır.

Kesinlik ve hassasiyet değerleri açısından değerlendirildiğinde ise herhangi bir boyut ayrıştırma yöntemi kullanılmadan elde edilen sonuçlardaki birbirleri arasındaki fark burada gözlemlenmemiştir. Kesinlik değeri gerçekleşen sonuçlar üzerindeki başarıyı, hassasiyet değeri

ise tahmin edilen sonuçlar üzerindeki başarıyı göstermesi nedeniyle bu iki değerin birbirine yakın sonuçlar vermesi istenilen bir durumdur.

Birliktelik Kuralları

Araştırmada birliktelik kuralı öğrenme metotlarından ikili veri yapısına uygun olarak iki tanesi seçilerek uygulamadaki sonuçları karşılaştırılmıştır. Bunlardan bir tanesi Öneri sistemlerinde sıklıkla kullanılan Fp-Growth metodu olurken, bir diğeri parçacık yığın optimizasyonu (Particle swarm optimization) olarak adlandırılan yöntem kullanılmıştır.

Araştırma Modeli

Fp-Growth diğer apriori, eclat gibi birliktelik kuralları gibi sık tekrarlanan nesne veri setleri üzerinden hesaplamaları gerçekleştirmektedir. Uygulamada bu algoritma üzerinden analizleri gerçekleştirmek için şu ön işlemler uygulanmıştır:

- Nesnelerin tekrarlama oranlarının anlık olarak veri seti üzerinde tespit edilmesi
- Sıklık tespit edilmeyen tüm nesnelerin veri setinden çıkartılması
- Nesnelerin belirlenen sıklık adedine göre azalan şekilde sıralanması

Bu işlemler sonucu elde edilen satın alma işlemleri üzerinden algoritmanın temel uygulama yolu olan fp-tree ağaç yapısı oluşturularak kurallar havuzu elde edilmiştir. Uygulamada minimum destek (min. support) 0,03, minimum güven (min. confidence) değeri 0,4 olarak tespit edilmiştir. Nesne başına oluşturulacak maksimum kural sayısı ise 5 olarak belirlenmiştir. Bu değerlerin tespitinde uygulama sonucu elde edilen skorların karşılaştırmasında eşit ve birbiri ile uyumlu yorumlamalar yapılabilmesi hususu dikkate alınmış, elde edilebilecek fazla sayıda kuralın diğer yöntemlerle yapılacak değerlendirmelerde hatalı sonuçlara yol açabileceği için kısıtlanmıştır.

Literatürde FP-Growth algoritması ile ilgili oldukça fazla varyasyon çalışması mevcuttur. Ancak bu uygulamada bu varyasyonlar üzerinde bir çalışma gerçekleştirilmemiş ve orijinal algoritmaya sadık kalınmıştır.

Algoritmanın çalışma mantığı Tablo 3.12’de özetlenmiştir. S değeri; referans günü ile referans nokta gün sayısının eklenmesiyle oluşan referans bitiş günü arasındaki satış kayıtlarının oluşturduğu veri setidir.

PSO algoritması Sarath ve Ravi (2013: 1832) tarafından gerçekleştirilen çalışmadaki adımlar takip edilerek analizlerin gerçekleştirilmesi düşünüldüğü için Michigan yaklaşımı olarak adlandırılan her parçacık üzerinden oluşturulan kuralların bağımsız değerlendirildiği yöntem kullanılmıştır.

Araştırmada ek olarak her parçacığın öneminin belirlenmesi için uygunluk değeri (fitness value) hesaplaması yapılmıştır. $k \Rightarrow l$ ürünü için uygunluk değeri şu şekilde açıklanmıştır;

$$Uygunluk = destek(k \rightarrow l) * güven(k \rightarrow l)$$

Destek değeri k ve l ürünlerinin dahil olduğu işlemleri tüm işlemler içindeki yüzdesel oranını, güven değerini kuralın sağ tarafını kapsayan kuralların aynı zamanda sol tarafın da dahil olduğu kuralların olasılık değeridir. Bu doğrultuda adımlar şu şekilde sıralanmaktadır:

- Tüm parçacıklar için uygunluk denkleminin uygulanarak hesaplamaların yapılması
- Yerel en iyi değerler ve global en iyi değerlerin tespiti

Değerler arasında en iyi sonuç elde edilene kadar tekrarlı işlemlerin gerçekleştirilmesi.

Tablo 3.12 Fp-Growth Model Oluşturma Algoritması

Fonksiyon FPGrowth(S){ <i>F-List oluştur: F[l];</i> <i>S veri seti içerisindeki her T_i işlemi için:</i> <i>T_i işlemi içerisindeki her n_i nesnesi için:</i> <i>$F[n_i]++$;</i> <i>işlem sonu</i> <i>işlem sonu</i> <i>F[] sırala;</i> <i>FP-tree oluştur: A;</i> <i>S veri seti içerisindeki her T_i işlemi için:</i> <i>F'ye göre T_i içindeki nesneleri sırala:</i> <i>AğaçOluştur()</i> <i>İşlem sonu</i> }
--

Bu işlem sonucunda en iyi kural bulunarak bu kural doğrultusunda eğitim veri seti içerisinde alınan sonuçlarla test verileri analiz edilebilmektedir. Tablo 3.11’de kullanılan temel algorithma optimizasyon minimizasyon kurallarına göre, sayı popülasyonu değeri ise 20 olarak belirlenmiştir. Maksimum iterasyon sayısı 500’dür.

Fp-growth ve ikili PSO gösterimde aynı ürünün aynı oturumda birden fazla satın alımı için kullanımı uygun olmamaktadır. $k \Rightarrow k$ güven kuralının oluşturulabilmesi için işlenebilir veri seti içerisinde şu şekilde hesaplanması uygun bulunmuştur;

$$b_{k,k} = \frac{k \text{ ürünün en az iki kez satın alındığı oturum sayısı}}{k \text{ ürünün satın alındığı oturum sayısı}}$$

Uygulamada, Birliktelik kuralları en fazla iki ürünün kıyaslamasını gerçekleştirilmesi amaçlanmıştır. Ancak kurallar sonucunda iki ürün sayısı yeterli olmaması durumunda birden fazla ürün içeren oluşumlarda listeye eklenmiştir. Daha önce de belirtildiği gibi liste uzunluğu 5 olarak kararlaştırılmıştır. Veri seti için her ürün aynı oturum içerisindeki ürünlerle tekrarlanarak hesaplamaların gerçekleştirilmesi sağlanmış ve bu hesaplar üzerinden ortalama değerle kıyaslama işlemi gerçekleştirilmiştir.

Araştırma analizinin gerçekleştirilmesinde Sarwar ve Karypis (2000: 161) tarafından gerçekleştirilen adımlar uygulanmış ve ürün k için kullanıcı geçmiş değer matrisleri hesaba katılarak $l \rightarrow k$ kuralı ile maksimum güven değerine göre sıralama gerçekleştirilmiştir. Uygulamada gerçekleştirilen adımlar Tablo 3.13 ile gösterilmiştir. Oluşturulan algoritma ile farklı zaman dilimleri için oluşturulmuş veri setlerinin tekrarlanan bir süreç içerisinde tek tek hesaplanarak ortalama bir sonuç elde etmeye yönelik bir düzenleme gerçekleştirilmiştir.

Tablo 3.13 PSO Model Oluşturma Algoritması

<i>FONKSİYON PSO(S)</i>
Her i parçacığı için:
Her d boyutu için:
İzin verilen alan içerisinde x_{id} pozisyonunu rasgele belirle
İzin verilen alan içerisinde v_{id} hızını rasgele belirle
İŞLEM SONU
İŞLEM SONU
DÖNGÜ $k:1$:
Her i parçacığı için:
Uygunluk değeri hesapla
EĞER uyg. değeri geçmiş verilerden elde edilen P_best_{id} 'den iyiyse:
Uygunluk değerini p_best_{id} olarak yeniden belirle
İŞLEM SONU
İŞLEM SONU
G_best_{id} içerisindeki en iyi uygun değerine sahip parçacığı seç
Her i parçacığı için:
Her d boyutu için:
Denkleme göre hızı hesapla:
$v_{id}(k+1) = wv_{id}(k) + c_1rand_1(p_{id} + x_{id}) + c_2rand_2(p_{gd} - x_{id})$
Parçacık pozisyonunu hesapla:
$x_{id}(k+1) = x_{id}(k) + v_{id}(k+1)$
İŞLEM SONU
İŞLEM SONU
$k=k+1$
DÖNGÜYE DEVAM ET:maksimum iterasyon sayısına veya minimum hata oranına erişim

Metrik Değerlendirme

Eğitim veri seti üzerinde yapılan çalışma ile elde edilen çıkarımlar test seti üzerinde uygulanarak aşağıdaki sonuçlar elde edilmiştir. Veri setinin oluşturulmasında referans gün mantığı uygulanarak 45, 90 ve 365 gün zaman noktaları ile herhangi bir zaman dilimi olmadan elde edilen değerlerin ortalaması alınarak aşağıdaki tablolar oluşturulmuştur. Tablolarda kesinlik ve hassasiyet değerlerinin yanı sıra bu değerlerin harmonik ortalaması olan F1 skorlarına da yer verilmiştir. Bu skorların elde ediminde kullanılan hata matrislerinin ortalama değerleri EK-3’de gösterilmiştir.

Tablo 3.14 PSO ve FpGrowth Metotlarının 45 Günlük Metrik Değerlendirmesi

45 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
PSO	19,83	21,57	20,66
FpGrowth	17,67	19,81	18,68

Tablo 3.15 PSO ve FpGrowth Metotlarının 90 Günlük Metrik Değerlendirmesi

90 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
PSO	19,77	20,84	20,29
FpGrowth	17,51	19,28	18,35

Tablo 3.16 PSO ve FpGrowth Metotlarının 365 Günlük Metrik Değerlendirmesi

365 Günlük Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
PSO	17,85	18,75	19,29
FpGrowth	15,91	17,32	16,59

Tablo 3.17 PSO ve FpGrowth Metotlarının Tüm Test Veri Seti Üzerindeki Metrik Değerlendirmesi

Tüm Veri Seti Değerlendirme (,00)	Kesinlik	Hassasiyet	F1 Skoru
PSO	15,97	17,88	16,87
FpGrowth	14,22	16,52	15,52

Test veri seti üzerinden gerçekleştirilen hesaplamalarda iki yöntemin kıyaslanmasında zaman dilimleri açısından farklılıklar göze çarpmaktadır. Kesinlik ve hassasiyet değerleri kıyaslandığında PSO ölçümleri arasındaki fark, fp-Growth yöntemine göre daha fazla olmakla birlikte PSO ve fp-Growth yöntemleri ile elde edilen sonuçlarda PSO yönteminin ölçümlenen tüm zaman dilimleri içinde Fp-Growth yöntemine göre kesinlik ve hassasiyet değerleri açısından daha iyi sonuçlar elde edilmiştir.

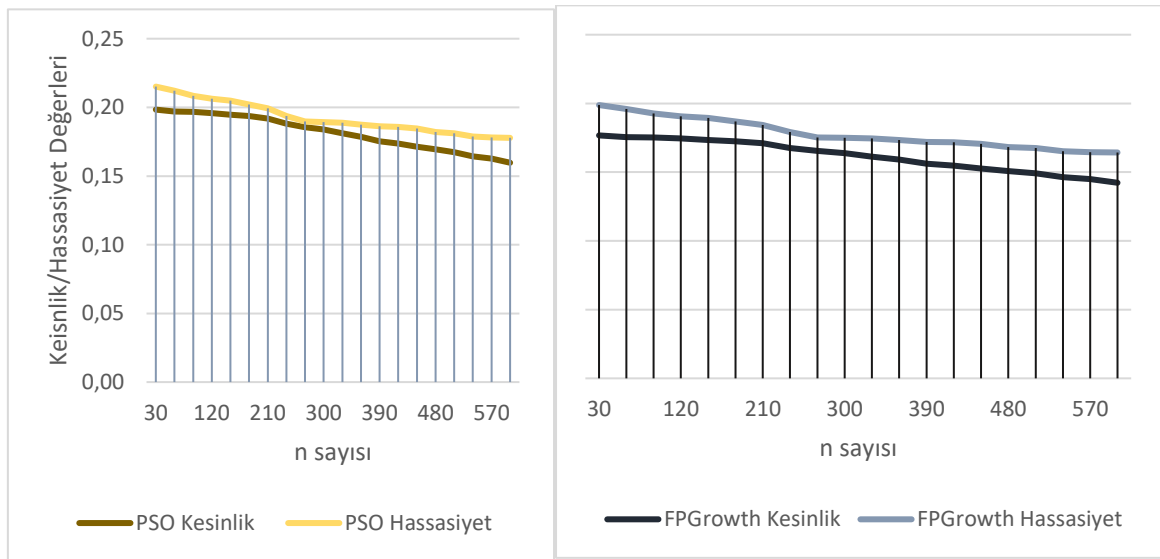
PSO yöntemi kullanılan durumlarda kısa dönem olarak değerlendirilen 45 günlük (Tablo 3.14) dönemde kesinlik ve hassasiyet değerleri açısından 0,1983 ve 0,2157 değerleri elde edilmiş, bir diğer kısa dönem değeri olan 90 günlük (Tablo 3.15) zaman dilimi içinse 0,1977 ve 0,2084 olarak belirlenmiştir. Bu dönemler arasında veri setleri açısından ciddi bir skor farklılığı gözlemlenmemiştir. Uzun dönem verileri olan 365 günlük (Tablo 3.16) zaman dilimi ve tüm veri seti (Tablo 3.17) üzerinde yapılan çalışmada kesinlik ve hassasiyet değerleri sırasıyla 0,1785-0,1875 ile 0,1597-0,1788 sonuçlarını ortaya koymaktadır. Değerlerden de anlaşılacağı üzere kısa dönemde hem kesinlik hem de hassasiyet değerleri için daha iyi performans skorları elde edilirken en uzun döneme kıyasla bu değerler arasındaki azalma kesinlik değeri için %19,46; hassasiyet değeri içinse %17,10'dur.

Araştırmada kullanılan bir başka birliktelik kural yöntemi olan Fp-Growth'un kullanıldığı durumlarda kısa dönem olarak değerlendirilen 45 günlük dönemde kesinlik ve hassasiyet değerleri açısından 0,1767 ve 0,1981 değerleri elde edilmiş, bir diğer kısa dönem değeri olan 90 günlük zaman dilimi içinse 0,1751 ve 0,1928 olarak belirlenmiştir. Uzun dönem verileri olan 365 günlük zaman dilimi ve tüm veri seti üzerinde yapılan çalışmada kesinlik ve hassasiyet değerleri sırasıyla 0,1591-0,1732 ile 0,1422-0,1652 sonuçlarını ortaya koymaktadır. Elde edilen performans skorları kısa dönemde hem kesinlik hem de hassasiyet değerleri için daha iyi sonuçlar elde edilirken, en uzun döneme kıyasla bu değerler arasındaki azalma kesinlik değeri için %19,54; hassasiyet değeri içinse %16,60 olmuştur.

Belirtilen dönemler arasında PSO yöntemi ile elde edilen model sonuçlar hem kısa dönemde hem de uzun dönemde daha iyi performans skoru oluşturmuştur.

Zaman aralıkları açısından daha detaylı bir araştırma yapıldığında ise geniş zaman aralıkları için gözlemlenen değer değişimlerinin, veri seti üzerinde genel bir trend olduğu gözlemlenmiştir. Daha önce de bahsedildiği gibi 10'ar günlük referans değerleri üzerinden yapılan incelemede her iki yöntem için de daha dar zaman dilimleri içerisinde daha iyi tahmin sonuçları elde edilebildiği rahatlıkla söylenebilmektedir.

30 ve 570 gün arasında her 10'ar gün arttırılarak hesaplanan yeni veri setleri üzerinde değişim daha net gözlemlenmektedir. PSO ve Fp-Growth algoritmalarının test veri setindeki kesinlik ve hassasiyet değerleri üzerinden elde edilen sonuçların gösterildiği Şekil 3.6'te kısa dönemdeki performans skorları zaman dilimlerinin genişliğinin arttırılmasıyla azalmakta ve algoritmanın başarısının düştüğü gözlemlenmiştir. Bu durum birliktelik kurallarının kısa zaman aralıkları ile oluşturulan veri setlerinde daha başarılı olduğunu söyleyebilmemize olanak sağlamaktadır. Bu iki algoritma açısından değerlendirildiğinde aynı veri setleri ile yapılan analiz PSO için bir nebze daha iyi sonuçları ortaya çıkarmaktadır.



Şekil 3.6 PSO ve FpGrowth Metotları Kesinlik/Hassasiyet Değerlerinin Grafiksel Gösterimi

Elde edilen veriler ışığında yapılan incelemede hassasiyet değerlerinin kesinlik değerlerinden daha yüksek olduğu gözlemlenmekte bu durum tahmin yapılan değerler üzerinde başarı oranının, gerçekleşen tüm değerler üzerinde başarılı oranının daha yüksek olduğunun göstergesidir. Hassasiyet değerleri her iki birliktelik kuralı uygulamasında orta dönemde yakın sonuçlar oluşturmuş ancak kısa zaman aralıkları ve uzun zaman aralıklarında birbirinden sonuçlar elde edilmiştir.

3.9. Yöntemlerin Karşılaştırılması

Araştırmada öneri sistemleri için oluşturulan söz konusu modeller neticesinde oluşturulan çıktılar üzerinden elde edilen skorlarla bir kıyaslama yapılmıştır. Bu doğrultuda belirlenen üç modelleme tekniği için ikişer alt yöntem uygulanarak hesaplamalar gerçekleştirilmiştir. Tablo 3.18 ile gerçekleştirilen modellerin test veri seti üzerindeki skorları özet olarak sunulmuştur. Söz konusu tablo her modelde kullanılan yöntemlerin kesinlik, hassasiyet ve bunların harmonik ortalamasını temsil eden F1 skorunu göstermektedir. Günlük zaman aralıkları olarak da değerlendirmeye alınan metot performansları incelendiği zaman her bir uygulamadan farklı çıkarımlar elde edilmesi söz konusudur. Bu durum aynı aileye altından gelmelerine rağmen modellerin farklı kriter ve mantıksal yaklaşımlardan oluşmaları nedeniyle araştırma öncesi de beklenen bir olguydu. Araştırmanın içinde bulunduğu kısıtlar dikkate alındığında performans skorundan tek başına değerlendirilmesinden çok birbiri ile farklılaşan performans karşılaştırmaları üzerinde durulması, araştırmanın sonuçlarının yorumlanması açısından daha sağlıklı olacaktır.

Tablo 3.18 Araştırmada Kullanılan Metotların, Kesinlik, Hassasiyet ve F1 Skorlarının Toplu Gösterimi

(,00)	Ürün Temelli Komşuluk Modeli						Matris Ayrıştırma						Birliktelik Kuralı					
	Jaccard			Kosinüs			TDA			NMF			PSO			FpGrowth		
	Kes.	Has.	F1	Kes.	Has.	F1	Kes.	Has.	F1	Kes.	Has.	F1	Kes.	Has.	F1	Ke	Has.	F1
45 Günlük	11,15	12,89	11,96	13,02	14,30	13,63	17,07	18,79	17,89	15,47	17,05	16,22	19,83	21,57	20,66	15,67	13,12	18,68
90 Günlük	10,82	12,51	11,60	12,70	13,72	13,19	16,71	18,28	17,46	14,93	16,00	15,45	19,77	20,84	20,29	14,90	12,98	18,35
365 Günlük	8,96	10,21	9,54	9,48	11,72	10,48	13,89	14,49	14,18	12,81	13,58	13,18	17,85	18,75	19,29	12,11	10,50	16,59
Tüm Veri Seti	7,90	9,01	8,42	8,84	10,32	9,52	12,78	13,15	12,96	11,85	12,48	12,16	15,97	17,88	16,87	10,71	7,54	15,52

Elde edilen skorlar ışığında bahse konu metotların performans skorları ile birlikte zamansal değişimleri yorumlanmıştır. Benzerlik ölçümleri ile elde edilen skorlara bakıldığında hem boyutsal ayrıştırma hem de birliktelik kuralları ile elde edilen skorların gerisinde kaldığı gözlemlenmektedir. Benzerlik ölçümü yöntemleri içinde yapılan karşılaştırmada Kosinüs benzerlik ölçümünün, Jaccard benzerlik ölçümüne kıyasla hem kısa zaman dilimleri hem de orta ve uzun zaman dilimlerinde daha iyi bir performans sergilediği vurgulanmıştır. F1 skoru bazında bakıldığı zaman Kosinüs benzerlik ölçümü en kısa dönemden uzun döneme 11,96 ile 8,42 performans skorları arasında değerler aldığı gözlemlenmektedir.

Bu değerler kosinüs benzerlik ölçümünde 14,30 ile 10,32 arasında değişmektedir. Görüleceği üzere her iki yöntem kısa zaman aralıklarında daha iyi bir performans çıkarmış ancak kosinüs değeri daha iyi bir tahmin skoru oluşturmuştur. Ancak bu değerler diğer yöntemlerden elde edilen skorlarla kıyaslandığı zaman düşük ve sistem yöneticilerin kullanımına uygun oranlar değildir. Bunun temel sebeplerinden biri veri setleri içerisindeki seyreklik sorunu en büyük etkidir. Bu durum benzerlik ölçülerinde görülen problemlerden olmakla birlikte literatürde bununla ilgili çalışmalar gerçekleştirilmiştir (Anand ve Bharadwaj, 2011: 5102).

Matris ayrıştırma teknikleri vasıtasıyla oluşturulan modelde, veri setleri öncelikli olarak boyut indirgeme tekniğine tabi tutulmuş ve vektör bazlı benzerlik ölçümü araçlarıyla boyut indirgeme tekniğinin sonuca nasıl etki ettiği gözlemlenmiştir. Bu yöntem kullanılırken iki farkı metot üzerinde durulmuştur. TDA ile test verileri üzerinde yapılan uygulamada elde edilen F1 skorları 17,89 ile 12,96 arasında değişmektedir. NMF F1 skorları ise 16,22 ile 12,16 arasında değerler almaktadır. Kısa zaman aralıklarında başlayarak uzun zaman aralıklarına kadar sabit bir çizgi ile sonuçların ortaya çıkmadığı zaman dilimi uzadıkça performansın azaldığı vurgulanmıştır. Bir diğer husus ise kısa zaman aralıkları ile uzun zaman aralıkları arasındaki değişim ile ilgilidir. Benzerlik ölçümü ile elde edilen sonuçlar arasındaki fark boyutsal indirgeme ile yapılan modelleme sonucuna göre daha fazladır. Boyutsal ayrıştırma metotları arasındaki farkı da düşürerek daha tutarlı bir sonuç elde edilmesini sağlamıştır.

Görüleceği üzere TDA yöntemi ile elde edilen boyut indirgenmiş veri setleri NMF yöntemine göre daha başarılı sonuçlar vermiş ve yüksek bir performans skoru elde edilmiştir. Safi benzerlik ölçümü kullanılarak yapılan ölçümlere kıyasla veri setleri içerisinde 45 günlük zaman diliminde TDA %31,25; NMF ise %19,00 daha iyi bir skor üretmiştir. 365 günlük dönemi dikkate aldığımızda ise TDA %35,30; NMF ise %25,76 daha iyi bir performans sergilemiştir. NMF kısa dönemde daha düşük bir performans sergilemiş uzun dönemde ise bu

farkı azaltmıştır. Ancak TDA ile her hâlükârda daha iyi bir performans yakalanmıştır. Boyutsal indirgeme teknikleri ile veri setleri içerisindeki seyreklik sorunları bir nebze azaltılmış ve daha başarılı sonuçlara ulaşılmıştır. Bu durum hem TDA hem de NMF için de geçerli bir ölçümdür. Benzer veri yapılarına sahip veri setleri için bu ve benzeri tekniklerin uygulanması sonucu olumlu etkileyebileceği ortaya çıkarılmıştır.

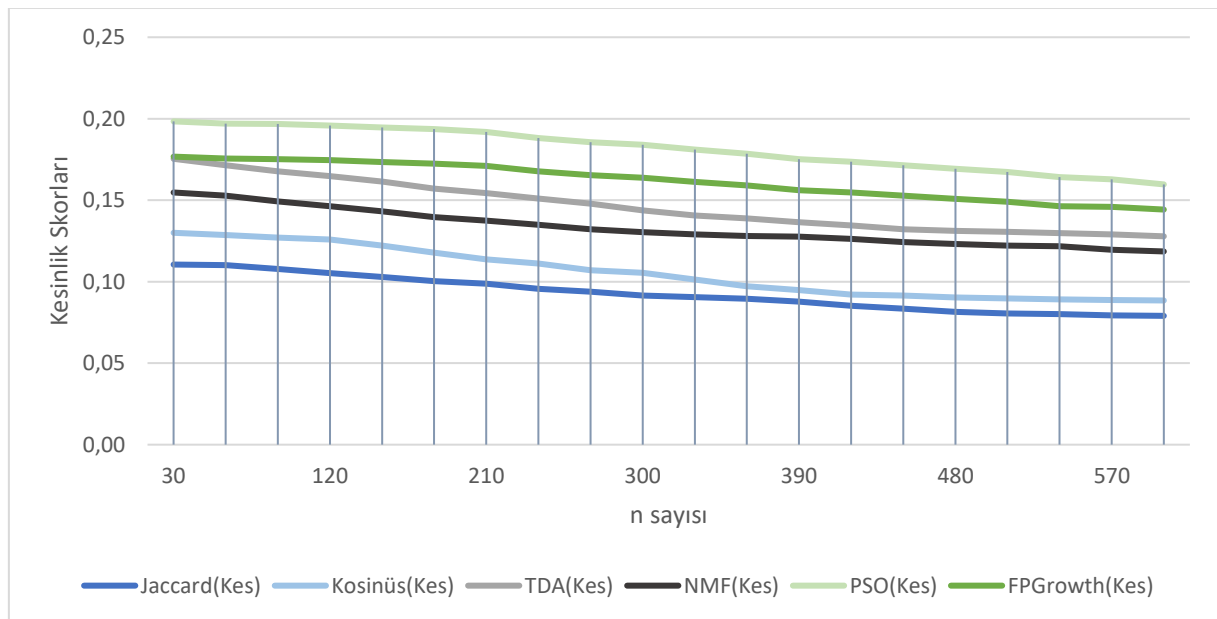
Bahsedildiği üzere birliktelik kuralları hem benzerlik ölçüm hem de boyutsal indirgeme yöntemleri ile elde edilen sonuçlardan daha iyi bir skor elde etmektedir. F1 skorları ile yapılan kıyaslamada 45 günlük zaman dilimi için PSO 20,66; Fp- Growth ise 18,68; 90 günlük zaman diliminde ise PSO 20,29; Fp-Growth 18,35 olarak performans skoru oluşturmuştur. Bu iki zaman dilimi için diğer metotlarla yapılan kıyaslamada benzerlik ölçümüne göre 45 günlük dönemde PSO yaklaşık %51 daha iyi bir skor üretirken, Fp-Growth ise yaklaşık %37 daha iyi bir skor üretmektedir. 90 günlük zaman diliminde bu durum PSO yaklaşık %32 daha iyi bir skor üretirken, Fp-Growth ise %26 daha iyi bir skor olarak gözlemlenmektedir. PSO, Matris ayrıştırma ile oluşturulan model ile kıyaslandığında (TDA) ise %15,48 daha başarılı aynı karşılaştırma Fp-Growth'da ise %4,41 olarak gözlemlenmektedir. 90 günlük sonuçlar ise sırasıyla %16,20 ve %5,09'dur. PSO tekniği boyutsal indirgeme tekniğine göre daha belirgin bir performans skoru üretebilmektedir.

Daha uzun dönemi temsil eden 365 günlük dönemde ise PSO F1 skoru 19,29; Fp-Growth F1 skoru ise 16,59'dur. Benzerlik ölçümü modellerinde daha başarılı sonuç elde edilen kosinüs benzerlik ölçümü ile aradaki fark kıyaslandığında PSO ve FP-Growth değerlerinin %84,06 ile %64,03 ; TDA matris ayrıştırma yöntemi ile kıyaslandığında ise sırasıyla %36,03 ve %16,99 daha iyi bir performans skoru elde edildiği gözlemlenmektedir. Diğer iki yöntemde göre daha iyi tahmin sonuçları üretebilen birliktelik kuralları hem kısa zaman hem de daha uzun zaman aralıkları içerisinde tatmin edici skorlar üretebilmiştir. Birliktelik kuralları içerisinde gerçekleştirilen bu iki metot arasında ise PSO, FP-Growth'a kıyasla iyi bir skor elde etmiştir.

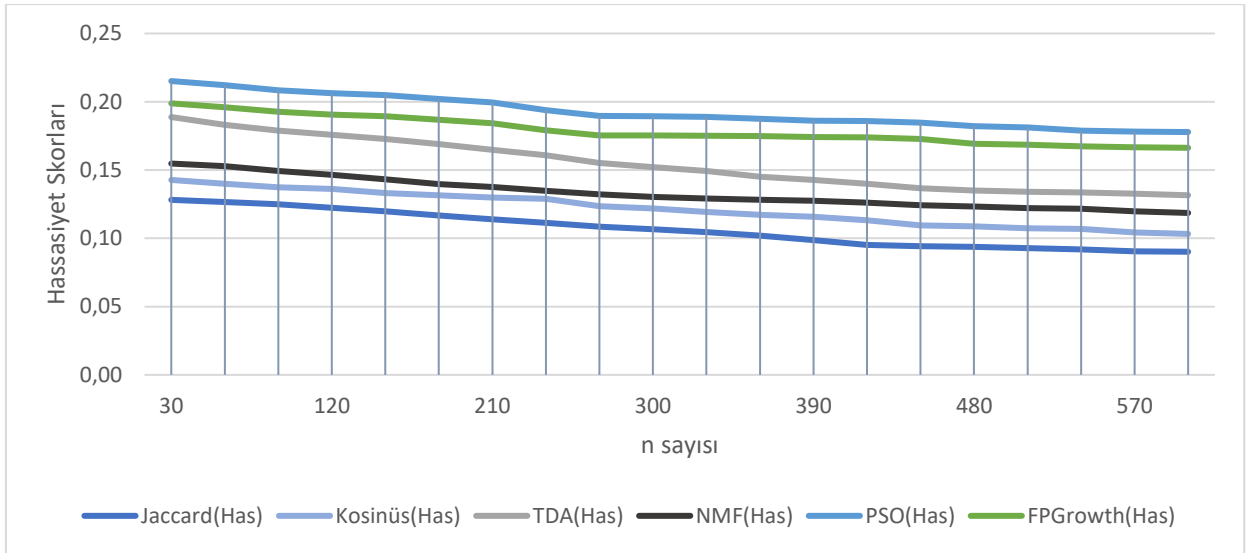
Kesinlik ve hassasiyet değerleri tüm metotlar özelinde hesaplanarak yorumlanabilir sonuçlar üretilmesine yardımcı bir araç olması sağlanmıştır. Kesinlik ve hassasiyet değerleri açısından bakıldığında uygulanan tüm yöntemlerde hassasiyet değerinin kesinlik değerine göre daha yüksek bir skor ürettiği gözlemlenmektedir. Bu durum hassasiyet değerinin tahmin edilen sonuçlar üzerinden geçerli tahmin oranını vermesinden kaynaklanmaktadır. Kesinlik değerinde tüm gerçekleşen sonuçlar ile bu kıyaslama yapılmaktadır. İlk etapta incelenen benzerlik ölçümünde bu fark diğer yöntemlere göre fazla olduğu gözlemlenmiştir. Bu durum daha önce de belirtildiği gibi veri setleri içindeki seyreklik sorunun bu metodun uygulamasında verimliliği azaltmasından kaynaklanmaktadır. Öte yandan boyut indirgeme teknikleri kullanılarak yapılan

çalışmada bu farkın kapandığı ve benzerlik ölçümü yöntemi ile elde edilen model sonucuna göre daha düşük bir farkın ortaya çıktığı gözlemlenmiştir. Yine bu durum boyut indirgeme tekniği ile seyreklik sorunun kısmen azaltılmasından kaynaklanmaktadır. Birlikte kurallarında ise bu fark kabul edilebilir bir alan içerisinde gerçekleşmekte orta süreli zaman aralıklarında minimuma indiği gözlemlenmektedir. Zaman dilimleri açısından belirli noktalarda veri zenginliğinin bu aralıklarda sağlanmış olması bunun nedenleri arasında sayılabilmektedir. Birlikte kuralları veri seti içerisindeki bağlantıların sayısal dağılımı üzerinden bir çıkarım oluşturması ikili veri tipine sahip bir veri setinde daha iyi bir sonuç oluşturmaya imkan sağlamıştır. Bu nedenle hem performans skoru hem de kesinlik hassasiyet değer ölçümü daha tutarlı bir sonuç oluşturmıştır.

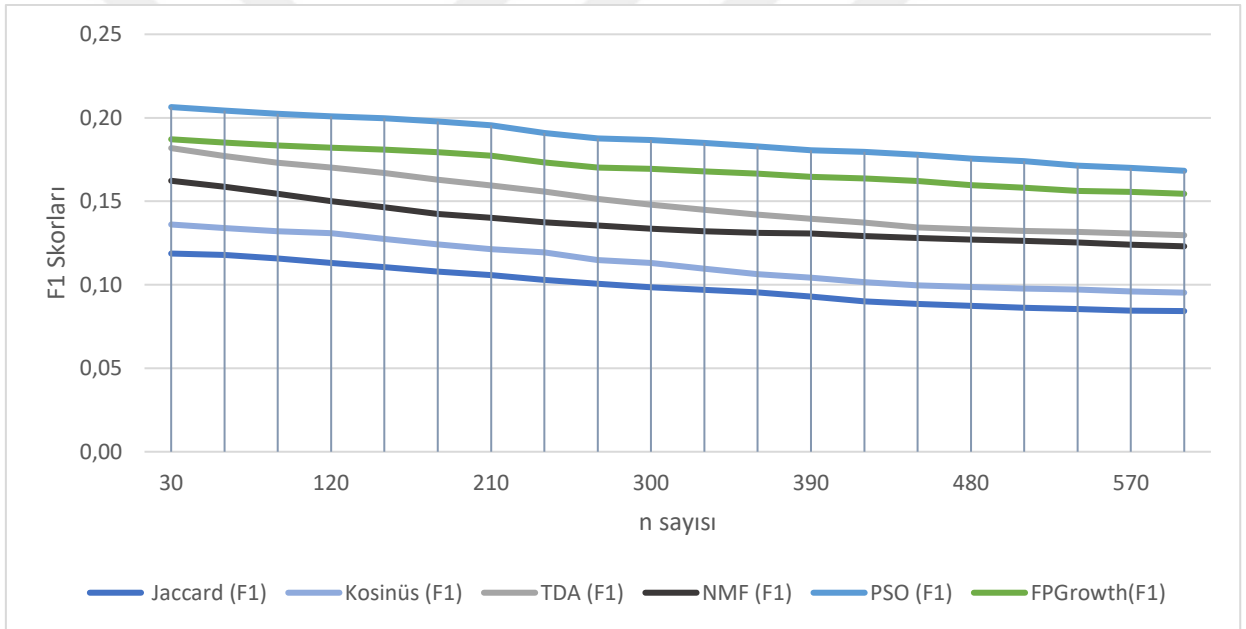
Araştırmada gerçekleştirilen modellemelerin genişletilmiş referans noktaları ile oluşturulan kesinlik, hassasiyet ve F1 skorları karşılaştırmalı olarak Şekil 3.7, Şekil 3.8 ve Şekil 3.9'da da gösterilmiştir. Hem Kesinlik hem hassasiyet hem de dolaylı olarak F1 skorunda zaman aralığının genişlemesi ile modellerin performansında düşüş olduğu daha net gözlemlenmektedir. Bu düşüş benzerlik ölçümünün yapıldığı modellerde daha keskin matris ayrıştırma ve birlikte kurallarında daha az gerçekleşmiştir. Yine tüm zaman dilimleri için matris ayrıştırma ve birlikte kuralları benzerlik ölçümü modellerine kıyasla daha yüksek skor elde etmektedir. Dar zaman aralıklarında fp-Growth, NMF ve PSO belirgin bir şekilde daha iyi bir skor üretirken, jaccard ve kosinüs değerlerinin düşük skorlarda kaldığı gözlemlenmiştir. Zaman aralığı genişletildikçe tüm metotlarda düşüş gözlemlenmesine rağmen birlikte kuralları olan TDA, NMF, PSO ve Fp-Growth'da bu düşüş diğerlerine göre daha az olduğu fark edilmektedir.



Şekil 3.7 Araştırmada Kullanılan Metotların Kesinlik Değeri Karşılaştırması

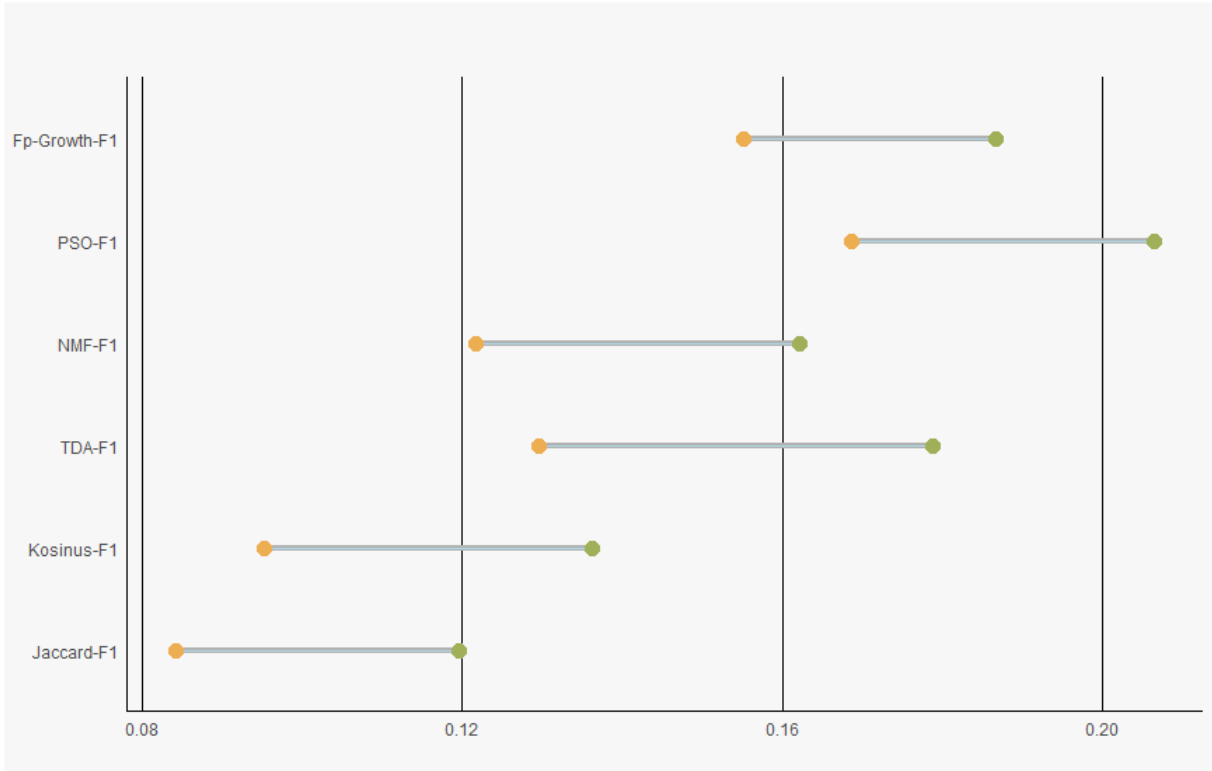


Şekil 3.8 Araştırmada Kullanılan Metotların Hassasiyet Değeri Karşılaştırması



Şekil 3.9 Araştırmada Kullanılan Metotların F1 Değeri Karşılaştırması

Şekil 3.10 ile üretilen modellerin performans değerlendirmesi zamansal olarak daha net gösterilmiştir. Turuncu nokta herhangi bir zaman dilimi sınırlaması yapılmadan gerçekleştirilen analizleri gösterirken yeşil nokta en dar zaman dilimini simgelemektedir. Skor değerlerinin karşılaştırılması F1 skoru üzerinden oluşturulmuştur. Bahsedildiği üzere standart benzerlik ölçümü metotları ile elde edilen skorların diğer metotlara göre daha düşük olduğu bu grafikten de gözlemlenebilmektedir.



Şekil 3.10 Araştırmada Kullanılan Metotların Sürelere Göre F1 Skor Karşılaştırması

Hem kosinüs hem de jaccard benzerlik ölçümü zaman dilimleri arasındaki değişkenlik benzerlik göstermekle beraber kosinüs ölçümünde başlangıç ve bitiş değerleri daha yüksek bir skorla başlamaktadır. Matris ayrıştırması yöntemi kullanılarak elde edilen sonuçlar ise yine bahsedildiği üzere standart benzerlik ölçümü metotlarından daha verimli olduğu net bir şekilde gözlemlenebilmektedir. Bununla birlikte zamansal dilimler arası değişimler daha az ve kısa ve uzun vade değişimlerinin daha sınırlı olduğu görülmektedir. En iyi performansı sergileyen birliktelik kurallarında zaman dilimleri arasındaki farkın diğer yöntemlere göre daha geniş bir ölçekte olduğu belirlenmiştir. Ancak her hâlikârda yüksek bir skor elde etmesi nedeniyle hem kısa dönemde hem de uzun dönemde daha iyi bir sonuç çıktısı elde etmiştir.

SONUÇ

Araştırma, literatürde makine öğrenme tekniklerinin oldukça yaygın bir kullanım alanı olan öneri sistemleri üzerine kurgulanmıştır. Bu kurgu, sıklıkla karşılaşılan kullanıcı ürün ilişkisinden yola çıkarak kullanıcı davranışlarının tahmin edilebilirliği üzerine birtakım çıkarımlar yapmak üzere inşa edilmiş ve farklı metotların sonucu nasıl değiştirebileceği sorgulanmıştır. Araştırma kullanıcı ve ürünlerden elde edilen verilerin kullanıldığı genel uygulama alanından sıyrılarak daha özelleştirilmiş model varyasyonlarının ortaya çıkarılmasını amaçlamıştır.

Araştırma, öneri sistemlerinde kullanılan farklı metotlarının performansını incelerken, bu işlemi belirli zaman aralıklarını dikkate alarak gerçekleştirmektedir. Belirtilen metotlar öncelikli olarak tüm veri seti ile birlikte 45, 90 ve 365 gün için oluşturulan zaman aralıkları için oluşturulan satın alma geçmiş değerler üzerinden analiz edilmiştir. Böylece metotların genel performansının yanında zaman aralıklarının değiştirilmesi ile performansın nasıl değiştiği ölçülmüştür. Belirlenen zaman aralıklarının değişiminin daha detaylı gözlemlenebilmesi için onar günlük dilimlerin sonuçları da araştırmaya dahil edilerek trend değişimleri ortaya koyulmuştur.

Ürün temelli komşuluk modeli ile gerçekleştirilen hesaplamalarda kullanılan ölçüm yöntemlerinden Jaccard ve Kosinüs benzerlik ölçüm yöntemleri veri yapısına uygun olduğu düşünüldükçe öneri sistem modeline dahil edilmiştir. Gerçekleştirilen işlemler bahsedildiği üzere belirli zaman aralıklarında test veri setinin şekillendirilmesi ile oluşturulmuş ve karşılaştırma zaman dilimleri ve tüm veri seti üzerinde gerçekleştirilmiştir. Yapılan karşılaştırma ile kısa zaman dilimlerinde her iki yöntemin daha iyi tahmin skorları ürettiği gözlemlenirken zaman dilimlerinin aralığı arttıkça tahmin skorlarının düştüğü gözlemlenmiştir. İki yöntem arasında ise kosinüs ölçüm metodu için daha iyi bir performans ölçümü sağlanmıştır. Ancak bu yöntemlerden elde edilen çıktıların geneli incelendiğinde diğer yöntemlere göre performans skorlarının düşük olduğu gözlemlenmiştir.

Kullanılan bir diğer yöntem olan matris ayrıştırma yönteminde ise TDA ve NMF metotları değerlendirilmiştir. Gerçekleştirilen analizler sonucu benzerlik ölçümü ile oluşturulan modellerde olduğu gibi, zaman dilimlerinin genişlemesi ile elde edilen skorda düşüş yaşanmakta, TDA'nın daha dar zaman aralıklarında NMF'ye göre daha iyi bir skor üretme potansiyeli gözlemlenmektedir. Elde edilen sonuçlar her hâlükârda veri seti üzerinde hiçbir ek model uygulaması gerçekleştirilmeyen benzerlik kurallarından daha iyi bir değere sahip olduğu

gözlemlenmiştir. Benzerlik ölçümüne göre 45 ve 90 günlük zaman dilimlerinde yaklaşık %30 daha iyi bir performans skoru elde edilmiştir. 365 gün ve üzeri zaman dilimlerinde ise bu fark %35'in üzerine çıkmaktadır. Boyut ayrıştırma yönteminin daha iyi sonuçlar oluşturması veri yapısındaki seyreklik sorunu ile bağdaştırılmıştır. Boyutsal ayrıştırma yöntemleri veri seti içerisindeki bu tür sorunların ortadan kaldırılmasına yardımcı olmakta ve daha tutarlı sonuçlar elde edilmesini sağlamaktadır. İkili veri tipi üzerinde bu tür uygulamaların gerçekleştirilmesi sonucu olumlu etkilediği ortaya çıkarılmıştır.

Son olarak birliktelik kurallarının da öneri sistemler içerisindeki kullanım alanları gözetilerek, bu kurallar içerisinde veri setine uygun olarak belirlenen Fp-Growth ve PSO metotları üzerine çalışılmıştır. Fp-Growth ve PSO hem kısa dönemde hem de uzun dönemde benzerlik ölçümü ve matris ayrıştırma yöntemlerine göre daha yüksek performans skorları üretmiştir. Benzerlik ölçümleri ile arasındaki fark tüm zaman dilimlerinde %50'nin üzerinde gözlemlenmekte boyutsal ayrıştırma yönteminin kullanıldığı durumlarla karşılaştırıldığında ise %30 daha iyi bir başarı elde etmektedir. Bu metotlar büyük veri tabanlarında iyi bir performans göstermesi ve nesne özelliklerinin karşılaştırılması yerine her satış kaydı için nesneler arası etkileşim üzerinden çıkarım gerçekleştirmesi ikili veri yapılarına kolay adapte edilmesine imkan tanımaktadır (Tyagi ve Bharadwaj, 2012). Böylelikle metotların işlemsel kayıt altında ikili bir veri yapısına sahip olması birliktelik kurallarının adapte edilerek daha iyi sonuçlar elde edilmesine olanak sağlamaktadır.

Uygulanan tüm modeller arasında birliktelik kurallarının daha iyi sonuç vermesinin veri yapısı ve adapte edilebilirliğindeki kolaylıkla açıklanabilmesine rağmen çalışma süresi ve işlem gücü açısından diğer yöntemlere göre dezavantajları bulunmaktadır. Sistem karar vericileri için oldukça tatminkâr performans skoru üretmeleri, olumlu bir değerlendirme olarak da bu yöntemler diğer yöntemlere göre daha yavaş ve yüksek kapasiteye ihtiyaç duymaktadırlar. Eğitim veri seti üzerinden oluşturulan model yapısı diğer yöntemlerle elde edilen modellere göre çok daha yavaş bir hızda oluşturulmuştur. Araştırmanın, eğitim ve test veri seti üzerinden gerçekleştirilmesi çalışmada süre ve işlem gücü açısından bir performans kıyaslamasına gidilmesine olanak tanınmamıştır. Ancak bu modellerin gerçek zamanlı bir öneri sistemine dönüştürülmesi aşamasında bu araştırmaların yapılarak karar vericilere sunulması maliyet-süre kriterlerine göre skorların değerlendirilmesine yardımcı olacaktır. Gerçek zamanlı bir uygulamada doğru sonuçların elde edilebilmesi için modelin kendini güncelleyerek bir döngü için öneri motorunun maksimum verimlilikte çalışmasının gerekliliği kabul edilen bir gerçektir. Sonraki çalışmalarda süre ve işlem gücü kriterinin de çalışmaya dahil edilerek gerçek zamanlı

öneri sunumu ile skor üretimi konusunda geliştirmelerin yapılması sonuçların iyileştirilmesine yardımcı olacaktır.

İkili veri tipi üzerinden kullanıcıların satın alma geçmişi değerlendirmeye alınarak tahmin modelleri oluşturulması oldukça güç ve geniş bir veri havuzuna ihtiyaç duymaktadır. Kullanıcı ve ürün bilgilerinin minimize edilerek yalın bir analiz sistemi kurulması elde edilen çıktıların karar vericileri tatmin etmekte zorluklar yaşatabilmektedir. Buna rağmen araştırmada bu tip veri yapılarının analizlere dahil edilerek çıktılar oluşturulmasında hangi metotların daha verimli bir sonuç elde edilebileceği göz önüne serilmiştir. Çıktıların verimliliği sistem karar vericileri tarafından değerlendirilmelidir.

Öneri sistemleri içinde satın alma değerleri üzerinden gerçekleştirilen değerlendirmelere literatürde gözlendiği daha önce belirtmişti. Bodapati, Anand (2008) tarafından kullanıcıların satın alma geçmişi ile elde edilen ikili veri yapısının öneri sistemlerinde kullanım alanı ilgili bir çalışma yapılmış ve uygulama yöntemleri üzerine detaylı bir şekilde durulmuştur. Liu ve Shih (2005) yine benzer bir yapı ile farklı modellerin birleşimi ile alternatif yollar üzerine çalışma gerçekleştirilmiştir. Pradel vd. (2011) gerçekleştirilen çalışma tez konusu araştırma ile benzer bir yol izleyerek çeşitli metotların kıyaslamasını yaparak satın alma davranışlarının zamansal boyutta performans değerlemesini yapmıştır. Bu tür veri yapılarından oluşan veri setlerinde genellikle seyreklik sorunu ile karşı karşıya kalmaktadır. Kullanıcı-ürün yapısına sahip matrislerde ürün sütunu satın alma tercihlerini yansıtmaması ve satın alma tercihlerinin almama tercihlerine kıyasla azınlık sayılabilecek bir sayıda olması modellerin oluşturulmasında zorluklara sebep olmaktadır. Bu araştırmada da benzer yapıya sahip kullanıcı-ürün matrisleri ile modeller oluşturulmuştur. Bu seyrekliğin azaltılması için model oluşturma işlemleri öncesi tedbirler alınarak minimize etmeye yönelik çalışmalar gerçekleştirilmiştir. Bu doğrultuda zamansal dilimler kullanılmış ve bu zaman dilimleri içerisinde birden fazla alışveriş yapmış kullanıcılar tercih edilmiştir. Tüm bu önlemlere rağmen seçilen veri türünün kaderi olarak seyreklik sorunu ile karşılaşmıştır. Daha tutarlı sonuçlar elde edebilmek amacıyla veri seçimi geniş tutulması ile birlikte mümkün olduğu kadar daha yoğun bir satın alma tercihi matrisi oluşturulmalıdır.

Özellikle e-ticarette karar vericilerin kullanıcı ile olan etkileşimini arttırma amacıyla öneri sistemine gün be gün ilgiyle yaklaşımları literatürde de modellerin farklılaştırma çabalarına hız katmaktadır. Bu çalışma temel bazda bir yöntem karşılaştırması olsa da kullanılan tekil veya hibrit çalışmalar için yol gösterici nitelikte olması amaçlanmıştır. Karar vericiler tarafından eldeki imkân ve veri niteliği açısından değerlendirilerek öneri sistem mekanizmalarının geliştirilmesi için farklı bir bakış açısı oluşturulmaya çalışılmıştır.

KAYNAKÇA

- Aamir, M. ve Bhusry, M. (2015) "Recommendation system: State of the art approach". *International Journal of Computer Applications*; 120 (12): 25-32.
- Abdi, H. ve Williams, L. J. (2010) "Principal component analysis". *Wiley interdisciplinary reviews: computational statistics*; 2 (4): 433-459.
- Adamic, L. A. ve Adar, E. (2003) "Friends and neighbors on the web". *Social networks*; 25 (3): 211-230.
- Aditya, P. H., Budi, I. ve Munajat, Q. (2016). "A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for e-commerce in indonesia: A case study". *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. 2016.303-308.
- Adomavicius, G. ve Tuzhilin, A. (2005) "Personalization technologies: A process-oriented perspective". *Communications of the ACM*; 48 (10): 83-90.
- Adomavicius, G. ve Tuzhilin, A. (2005) "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions". *IEEE transactions on knowledge and data engineering*; 17 (6): 734-749.
- Agarwal, R. ve Srikant, R. (1994). "Fast algorithms for mining association rules". *Proc. of the 20th VLDB Conference*. 1994.487-499.
- Aggarwal, C. C. (2016). *Ensemble-based and hybrid recommender systems*. In *Recommender systems* (pp. 199-224): Springer.
- Aggarwal, C. C. (2016). *Recommender systems*. Springer, Amerika Birleşik Devletleri.
- Aggarwal, C. C. ve Parthasarathy, S. (2001). "Mining massively incomplete data sets by conceptual reconstruction". *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. 2001.227-232.
- Aggelis, V. ve Christodoulakis, D. (2005). "Customer clustering using rfm analysis". *Proceedings of the 9th WSEAS International Conference on Computers*. 2005.2.
- Alyari, F. ve Navimipour, N. J. (2018) "Recommender systems: A systematic review of the state of the art literature and suggestions for future research". *Kybernetes*; 47 (5): 985-1017.
- Amershi, S. ve Conati, C. (2009) "Combining unsupervised and supervised classification to build user models for exploratory". *Journal of Educational Data Mining*; 1 (1): 18-71.

- Anand, D. ve Bharadwaj, K. K. (2011) "Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities". *Expert systems with applications*; 38 (5): 5101-5109.
- Arazy, O., Kumar, N. ve Shapira, B. (2009) "Improving social recommender systems". *IT professional*; 11 (4).
- Beck, J. R., Shultz, E. K. J. A. o. p. (1986) "The use of relative operating characteristic (roc) curves in test performance evaluation". 110 (1): 13-20.
- Bell, R. M. ve Koren, Y. (2007) "Lessons from the netflix prize challenge". *SIGKDD Explor. Newsl.*; 9 (2): 75-79.
- Bennett, J. ve Lanning, S. (2007). "The netflix prize". *Proceedings of KDD cup and workshop*. 2007.35.
- Blattberg, R. C., Kim, B.-D. ve Neslin, S. A. (2008). Why database marketing? *Database marketing* (pp. 13-46): Springer.
- Bobadilla, J., Bojorque, R., Esteban, A. H. (2017) "Recommender systems clustering using bayesian non negative matrix factorization". *IEEE Access*; 6: 3549-3564.
- Bodapati, A. (2008) "Recommendation systems with purchase data". *Journal of marketing research*; 45 (1): 77-93.
- Bodapati, A. (2008) "Recommendation systems with purchase data". *Journal of marketing research*; 45 (1): 77-93.
- Boutsidis, C. ve Gallopoulos, E. (2008) "Svd based initialization: A head start for nonnegative matrix factorization". *Pattern recognition*; 41 (4): 1350-1362.
- Bouza, A., Reif, G., Bernstein, A. (2008) "Semtree: Ontology-based decision tree algorithm for recommender systems".
- Breese, J. S., Heckerman, D. ve Kadie, C. (1998). "Empirical analysis of predictive algorithms for collaborative filtering". *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. 1998.43-52.
- Brusilovski, P., Kobsa, A. ve Nejdl, W. (2007). *The adaptive web: Methods and strategies of web personalization*. Springer Science & Business Media.
- Buckland, M. ve Gey, F. (1994) "The relationship between recall and precision". *Journal of the American society for information science*; 45 (1): 12-19.
- Bult, J. R. ve Wansbeek, T. (1995) "Optimal selection for direct mail". *Marketing Science*; 14 (4): 378-394.
- Burke, R. (2002) "Hybrid recommender systems: Survey and experiments". *User modeling and user-adapted interaction*; 12 (4): 331-370.

- Burke, R. (2007). *The adaptive web. Chapter hybrid web recommender systems*.
- Burke, R., Felfernig, A. ve Göker, M. H. (2011) "Recommender systems: An overview". *Ai Magazine*; 32 (3): 13-18.
- Campbell, B. M. (1978) "Similarity coefficients for classifying relevés". *Vegetatio*; 37 (2): 101-109.
- Cheng, R. ve Jin, Y. (2015) "A social learning particle swarm optimization algorithm for scalable optimization". *Information Sciences*; 291: 43-60.
- Cho, Y. H., Kim, J. K. ve Kim, S. H. (2002) "A personalized recommender system based on web usage mining and decision tree induction". *Expert systems with Applications*; 23 (3): 329-342.
- Choi, S.-S., Cha, S.-H. ve Tappert, C. (2010) "A survey of binary similarity and distance measures". *Journal of Systemics, Cybernetics and Informatics*; 8 (1): 43-48.
- Chou, S., Chen, C.-W. ve Lin, J.-Y. (2015) "Female online shoppers: Examining the mediating roles of e-satisfaction and e-trust on e-loyalty development". *Journal of retailing and Consumer services*; 25 (4): 542-561.
- Cremonesi, P., Koren, Y. ve Turrin, R. (2010). "Performance of recommender algorithms on top-n recommendation tasks". *Proceedings of the fourth ACM conference on Recommender systems*. 2010.39-46.
- Çano, E. ve Morisio, M. (2017) "Hybrid recommender systems: A systematic literature review". *Intelligent Data Analysis*; 21 (6): 1487-1524.
- Danielsson, P.-E. (1980) "Euclidean distance mapping". *Computer Graphics and image processing*; 14 (3): 227-248.
- de Mello Neto, W. L. ve Nowé, A. (2012). "Insights on social recommender system". *Proceedings of the Workshop on Recommendation Utility Evaluation: Beyond RMSE, at ACM RecSys12*. 2012.33-38.
- DeMers, D. ve Cottrell, G. W. (1993). "Non-linear dimensionality reduction". *Advances in neural information processing systems*. 1993.580-587.
- Demiriz, A. (2001). "An association mining-based product recommender". *NFORMS Miami 2001 Annual Meeting Cluster: Data Mining*. 2001
- Denning, P. J. (1982) "Acm president's letter: Electronic junk". *Communications of the ACM*; 25 (3): 163-165.
- Dwivedi, S., Kasliwal, P. ve Soni, S. (2016). "Comprehensive study of data analytics tools (rapidminer, weka, r tool, knime)". *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*. 2016.1-8.

- Eberhart, R. ve Kennedy, J. (1995). "A new optimizer using particle swarm theory". *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*. 1995.39-43.
- Eirinaki, M., Gao, J., Varlamis, I. (2018). Recommender systems for large-scale social networks: A review of challenges and solutions. *Elsevier*.
- Ekstrand, M. D., Riedl, J. T. ve Konstan, J. A. (2011) "Collaborative filtering recommender systems". *Foundations Trends in Human-Computer Interaction*; 4 (2): 81-173.
- Famili, A., Shen, W.-M., Weber, R. (1997) "Data preprocessing and intelligent data analysis". *Intelligent data analysis*; 1 (1): 3-23.
- Felfernig, A., Friedrich, G. ve Schmidt-Thieme, L. (2007) "Guest editors' introduction: Recommender systems". *IEEE Intelligent systems*; 22 (3): 18-21.
- Fischer, G. (2001) "User modeling in human-computer interaction". *User modeling and user-adapted interaction*; 11 (1): 65-86.
- Fleder, D. M. ve Hosanagar, K. (2007). "Recommender systems and their impact on sales diversity". *Proceedings of the 8th ACM conference on Electronic commerce*. 2007.192-199.
- Foltz, P. W. ve Dumais, S. T. (1992) "Personalized information delivery: An analysis of information filtering methods". *Communications of the ACM*; 35 (12): 51-60.
- Garbarino, E. ve Strahilevitz, M. (2004) "Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation". *Journal of Business Research*; 57 (7): 768-775.
- Ge, M., Delgado-Battenfeld, C. ve Jannach, D. (2010). "Beyond accuracy: Evaluating recommender systems by coverage and serendipity". *Proceedings of the fourth ACM conference on Recommender systems*. 257-260.
- Gerard Salton ve McGill, M. (1983). *Introduction to modern information retrieval*. Pergamon, New York.
- Gershman, A., Meisels, A., Lüke, K.-H. (2010) "A decision tree based recommender system". *10th International Conference on Innovative Internet Community Systems* 12.
- Ghani, R. ve Fano, A. (2002). "Building recommender systems using a knowledge base of product semantics". *Proceedings of the Workshop on Recommendation and Personalization in ECommerce at the 2nd International Conference on Adaptive Hypermedia and Adaptive Web based Systems*. 27-29.
- Ghazanfar, M. ve Prugel-Bennett, A. (2010) "An improved switching hybrid recommender system using naive bayes classifier and collaborative filtering".

- Ghazarian, S. ve Nematbakhsh, M. A. (2015) "Enhancing memory-based collaborative filtering for group recommender systems". *Expert systems with applications*; 42 (7): 3801-3812.
- Glance, N. S. (2005) "Recommender system and method for generating implicit ratings based on user interactions with handheld devices".
- Golbeck, J. (2006) "Generating predictive movie recommendations from trust in social networks". *Trust Management*: 93-104.
- Goldberg, D., Nichols, D., Oki, B. M. (1992) "Using collaborative filtering to weave an information tapestry". *Communications of the ACM*; 35 (12): 61-70.
- Gong, S., Ye, H. ve Dai, Y. (2009). "Combining singular value decomposition and item-based recommender in collaborative filtering". *Second International Workshop on Knowledge Discovery and Data Mining*. 2009.769-772.
- Good, N., Schafer, J. B., Konstan, J. A. (1999). "Combining collaborative filtering with personal agents for better recommendations". *AAAI/IAAI*. 1999.439-446.
- Gorakala, S. K. ve Usielli, M. (2015). *Building a recommendation system with r*. Packt Publishing Ltd, Amerika Birleşik Devletleri.
- Grčar, M., Mladenich, D., Fortuna, B. (2005). "Data sparsity issues in the collaborative filtering framework". *International Workshop on Knowledge Discovery on the Web*. 58-76.
- Groh, G. ve Ehmig, C. (2007). "Recommendations in taste related domains: Collaborative filtering vs. Social filtering". *Proceedings of the 2007 international ACM conference on Supporting group work*. 127-136.
- Gunawardana, A. ve Shani, G. (2009) "A survey of accuracy evaluation metrics of recommendation tasks". *Journal of Machine Learning Research*; 10 (Dec): 2935-2962.
- Hájek, P., Havel, I. ve Chytil, M. (1966) "The guha method of automatic hypotheses determination". *Computing*; 1 (4): 293-308.
- Hájek, P. ve Havránek, T. (1977) "On generation of inductive hypotheses". *International Journal of Man-Machine Studies*; 9 (4): 415-438.
- Hallinan, B. ve Striphos, T. (2016) "Recommended for you: The netflix prize and the production of algorithmic culture". *New media society*; 18 (1): 117-137.
- Hamers, L. (1989) "Similarity measures in scientometric research: The jaccard index versus salton's cosine formula". 25 (3): 315-318.
- Han, J., Pei, J. ve Yin, Y. (2000). "Mining frequent patterns without candidate generation". *ACM sigmod record*. 2000.1-12.
- Han, P., Xie, B., Yang, F. (2004) "A scalable p2p recommender system based on distributed collaborative filtering". *Expert systems with applications*; 27 (2): 203-210.

- Hastie, T., Tibshirani, R. ve Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.
- Hawkins, D. (2004) "The problem of overfitting". *Journal of chemical information computer sciences*; 44 (1): 1-12.
- Hay, A. (1988) "The derivation of global estimates from a confusion matrix". *International Journal of Remote Sensing*; 9 (8): 1395-1398.
- Herlocker, J., Konstan, J. ve Riedl, J. (2002) "An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms". *Information retrieval*; 5 (4): 287-310.
- Herlocker, J., Konstan, J., Terveen, L. (2004) "Evaluating collaborative filtering recommender systems". *ACM Transactions on Information Systems*; 22 (1): 5-53.
- Herlocker, J. L., Konstan, J. A., Borchers, A. (1999). "An algorithmic framework for performing collaborative filtering". *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 230-237.
- Hershey, J. R. ve Olsen, P. A. (2007). "Approximating the kullback leibler divergence between gaussian mixture models". *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2007.IV. 317-320.
- Hill, W., Stead, L., Rosenstein, M. (1995). *Recommending and evaluating choices in a virtual community of use*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA.
- Hong, T.-P., Lin, J.-W. ve Wu, Y.-L. (2006). "A fast updated frequent pattern tree". *2006 IEEE International Conference on Systems, Man and Cybernetics*. 2167-2172.
- Hoyer, P. O. (2004) "Non-negative matrix factorization with sparseness constraints". *Journal of machine learning research*; 5 (Nov): 1457-1469.
- Hoyer, P. O. J. (2004) "Non-negative matrix factorization with sparseness constraints". *Journal of machine learning research*; 5 (Nov): 1457-1469.
- Hu, Y., Koren, Y. ve Volinsky, C. (2008). "Collaborative filtering for implicit feedback datasets". *Eighth IEEE International Conference on Data Mining*. 263-272.
- Huang, A. (2008). "Similarity measures for text document clustering". *Proceedings of the sixth new zealand computer science research student conference*. 9-56.
- Huang, Z., Chen, H. ve Zeng, D. (2004) "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering". *ACM Transactions on Information Systems*; 22 (1): 116-142.

- Huang, Z., Zeng, D. D. ve Chen, H. (2007) "Analyzing consumer-product graphs: Empirical findings and applications in recommender systems". *Management science*; 53 (7): 1146-1164.
- Hyvärinen, A. ve Oja, E. (2000) "Independent component analysis: Algorithms and applications". *Neural networks*; 13 (4-5): 411-430.
- Isinkaye, F., Folajimi, Y. ve Ojokoh, B. (2015) "Recommendation systems: Principles, methods and evaluation". *Egyptian Informatics Journal*; 16 (3): 261-273.
- Jaccard, P. (1908) "Nouvelles recherches sur la distribution florale". *Bull. Soc. Vaud. Sci. Nat.*; 44: 223-270.
- Jahrer, M., Töschner, A. ve Legenstein, R. (2010). "Combining predictions for accurate recommender systems". *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 693-702.
- Jameson, A. (2004). "More than the sum of its members: Challenges for group recommender systems". *Proceedings of the working conference on Advanced visual interfaces*. 2004. 48-54.
- Jamiy, F. E., Daif, A., Azouazi, M. (2015) "The potential and challenges of big data-recommendation systems next level application". *arXiv preprint arXiv*; 3424.
- Kaminskas, M. ve Bridge, D. (2016) "Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems". *ACM Transactions on Interactive Intelligent Systems*; 7 (1): 1-42.
- Katz, L. (1953) "A new status index derived from sociometric analysis". *Psychometrika*; 18 (1): 39-43.
- Kelly, D. (2009) "Methods for evaluating interactive information retrieval systems with users". *Foundations Trends in Information Retrieval*; 3 (1-2): 1-224.
- Kim, C. ve Kim, J. (2003). "A recommendation algorithm using multi-level association rules". *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*. 2003.524-527.
- Koren, Y., Bell, R. ve Volinsky, C. (2009) "Matrix factorization techniques for recommender systems". *Computer*; 42 (8): 30-37.
- Koren, Y. ve Sill, J. (2011). "Ordrec: An ordinal model for predicting personalized item rating distributions". *Proceedings of the fifth ACM conference on Recommender systems*. 2011, 117-124.
- Kumar, B. S. ve Rukmani, K. (2010) "Implementation of web usage mining using apriori and fp growth algorithms". *Int. J. of Advanced networking and Applications*; 1(06): 400-404.

- Lam, X. N., Vu, T., Le, T. D. (2008). "Addressing cold-start problem in recommendation systems". *Proceedings of the 2nd international conference on Ubiquitous information management and communication*. 2008, .208-211.
- Lee, C.-H., Gutierrez, F. ve Dou, D. (2011). "Calculating feature weights in naive bayes with kullback-leibler measure". *2011 IEEE 11th International Conference on data mining*. ,1146-1151.
- Lee, D. D. ve Seung, H. S. (1999) "Learning the parts of objects by non-negative matrix factorization". *Nature*; 401 (6755): 788.
- Lee, S. K., Cho, Y. H. ve Kim, S. H. (2010) "Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations". *Information Sciences*; 180 (11): 2142-2155.
- Leydesdorff, L. (2008) "On the normalization and visualization of author co-citation data: Salton's cosine versus the jaccard index". 59 (1): 77-85.
- Li, H., Wang, Y., Zhang, D. (2008). "Pfp: Parallel fp-growth for query recommendation". *Proceedings of the 2008 ACM conference on Recommender systems*.107-114.
- Lika, B., Kolomvatsos, K. ve Hadjiefthymiades, S. (2014) "Facing the cold start problem in recommender systems". *Expert Systems with Applications*; 41 (4): 2065-2073.
- Linden, G., Smith, B. ve York, J. (2003) "Amazon. Com recommendations: Item-to-item collaborative filtering". *IEEE Internet computing*; 7 (1): 76-80.
- Liu, D.-R. ve Shih, Y.-Y. (2005) "Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences". *Journal of Systems Software*; 77 (2): 181-191.
- Loeb, S. ve Terry, D. (1992) "Information filtering". *Commun. ACM*; 35 (12): 26–28.
- Lu, J., Wu, D., Mao, M. (2015) "Recommender system application developments: A survey". *Decision Support Systems*; 74: 12-32.
- Mahmood, T. ve Ricci, F. (2009). *Improving recommender systems with adaptive conversational strategies*. Proceedings of the 20th ACM conference on Hypertext and hypermedia, Torino, Italy.
- Manning, C., Raghavan, P. ve Schütze, H. (2010) "Introduction to information retrieval". *Natural Language Engineering*; 16 (1): 100-103.
- Marlin, B. M., Zemel, R. S., Roweis, S. T. (2011). "Recommender systems: Missing data and statistical model estimation". *Twenty-Second International Joint Conference on Artificial Intelligence*.

- Martín-Vicente, M. I., Gil-Solla, A., Ramos-Cabrera, M. (2014) "A semantic approach to improve neighborhood formation in collaborative recommender systems". *Expert Systems with Applications*; 41 (17): 7776-7788.
- Massa, P. ve Avesani, P. (2004). "Trust-aware collaborative filtering for recommender systems". *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*.492-508.
- Melville, P. ve Sindhvani, V. (2010) "Recommender systems". *Encyclopedia of machine learning*; 1: 829-838.
- Miyahara, K. ve Pazzani, M. J. (2000). "Collaborative filtering with the simple bayesian classifier". *Pacific Rim International conference on artificial intelligence*. 679-689.
- Montaner, M., López, B. ve De La Rosa, J. L. (2003) "A taxonomy of recommender agents on the internet". *Artificial intelligence review*; 19 (4): 285-330.
- Nathanson, T., Bitton, E. ve Goldberg, K. (2007). "Eigentaste 5.0: Constant-time adaptability in a recommender system using item clustering". *Proceedings of the 2007 ACM conference on Recommender systems*. 149-152.
- O'mahony, M. P., Hurley, N. J. ve Silvestre, G. C. (2004) "An evaluation of neighbourhood formation on the performance of collaborative filtering". *Artificial Intelligence Review*; 21 (3-4): 215-228.
- Omondi, A. O. ve Mbugua, A. W. (2017) "An application of association rule learning in recommender systems for e-commerce and its effect on marketing". *Pan African Conference on Science, Computing and Telecommunications*; 20.
- Ozgun, C., Colliau, T., Rogers, G. (2017) "Matlab vs. Python vs. R". *Journal of Data Science*; 15 (3): 355-372.
- Park, D. H., Kim, H. K., Choi, I. Y. (2012) "A literature review and classification of recommender systems research". *Expert Systems with Applications*; 39 (11): 10059-10072.
- Park, Y.-J. ve Tuzhilin, A. (2008). "The long tail of recommender systems and how to leverage it". *Proceedings of the 2008 ACM conference on Recommender systems*. 2008.11-18.
- Pazzani, M. J. ve Billsus, D. (2007). *Content-based recommendation systems*. In *The adaptive web* (pp. 325-341): Springer.
- Peters, T. A. (1993) "The history and development of transaction log analysis". *Library hi tech*; 11 (2).

- Portugal, I., Alencar, P. ve Cowan, D. (2018) "The use of machine learning algorithms in recommender systems: A systematic review". *Expert Systems with Applications*; 97: 205-227.
- Pradel, B., Sean, S., Delporte, J. (2011). "A case study in a recommender system based on purchase data". *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 377-385.
- Qiu, J., Lin, Z. ve Li, Y. (2015) "Predicting customer purchase behavior in the e-commerce context". *Electronic commerce research*; 15 (4): 427-452.
- Raghavan, V., Bollmann, P. ve Jung, G. S. (1989) "A critical investigation of recall and precision as measures of retrieval system performance". *ACM Transactions on Information Systems*; 7 (3): 205-229.
- Resnick, P., Iacovou, N., Suchak, M. (1994). *Grouplens: An open architecture for collaborative filtering of netnews*. Chapel Hill, North Carolina, USA.
- Resnick, P. ve Varian, H. R. (1997) "Recommender systems". *Commun. ACM*; 40 (3): 56-58.
- Ricci, F., Rokach, L. ve Shapira, B. (2011). *Introduction to recommender systems handbook*. In *Recommender systems handbook* (pp. 1-35): Springer.
- Ricci, F., Rokach, L., Shapira, B. (2015). *Recommender systems handbook*. Springer.
- Robles, V., Larranaga, P., Menasalvas, E. (2003). "Improvement of naive bayes collaborative filtering using interval estimation". *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*. 2003.168-174.
- Rocchio, J. J. (1971) "Relevance feedback in information retrieval". *The Smart retrieval system-experiments in automatic document processing*.
- Rodgers, S. ve Harris, M. A. (2003) "Gender and e-commerce: An exploratory study". *Journal of advertising research*; 43 (3): 322-329.
- Rokach, L. ve Maimon, O. (2005). *Decision trees*. In *Data mining and knowledge discovery handbook* (pp. 165-192): Springer.
- Rossetti, M., Stella, F. ve Zanker, M. (2013). "Towards explaining latent factors with topic models in collaborative recommender systems". *2013 24th International Workshop on Database and Expert Systems Applications*. 162-167.
- Salton, G., Wong, A. ve Yang, C.-S. (1975) "A vector space model for automatic indexing". *Communications of the ACM*; 18 (11): 613-620.
- Sarath, K. ve Ravi, V. (2013) "Association rule mining using binary particle swarm optimization". *Engineering Applications of Artificial Intelligence*; 26 (8): 1832-1840.

- Sarwar, B., Karypis, G., Konstan, J. (2000). "Analysis of recommendation algorithms for e-commerce". *EC*. 2000.158-167.
- Sarwar, B., Karypis, G., Konstan, J. (2000). *Application of dimensionality reduction in recommender system-a case study*. Retrieved from
- Sarwar, B., Karypis, G., Konstan, J. (2001). "Item-based collaborative filtering recommendation algorithms". *Proceedings of the 10th international conference on World Wide Web*. 285-295.
- Sarwar, B. M., Karypis, G., Konstan, J. (2002). "Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering". *Proceedings of the fifth international conference on computer and information technology*. 291-324.
- Sarwar, B. M., Karypis, G., Konstan, J. A. (2001) "Item-based collaborative filtering recommendation algorithms". *Www*; 1: 285-295.
- Schafer, J. B., Konstan, J. ve Riedl, J. (1999). "Recommender systems in e-commerce". *Proceedings of the 1st ACM conference on Electronic commerce*. 1999.158-166.
- Schafer, J. B., Konstan, J. A. ve Riedl, J. (2001). *E-commerce recommendation applications*. In *Applications of data mining to electronic commerce* (pp. 115-153): Springer.
- Schein, A. I., Popescul, A., Ungar, L. H. (2002). "Methods and metrics for cold-start recommendations". *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. 253-260.
- Schneider, D. (2018) "Machine learning predicts home prices". *IEEE Spectrum*; 56 (1): 42-43.
- Schröder, G., Thiele, M. ve Lehner, W. (2011). "Setting goals and choosing metrics for recommender system evaluations". *UCERSTI2 workshop at the 5th ACM conference on recommender systems, Chicago, USA*. 50-53.
- Shardanand, U. ve Maes, P. (1995). "Social information filtering: Algorithms for automating "word of mouth"". *Proceedings of the SIGCHI conference on Human factors in computing systems*. 210-217.
- Sidhu, S., Meena, U. K., Nawani, A. (2014) "Fp growth algorithm implementation". *International Journal of Computer Applications*; 93 (8).
- Smith, B. ve Linden, G. (2017) "Two decades of recommender systems at amazon. Com". *Ieee internet computing*; 21 (3): 12-18.
- Smyth, B. (2007) "Case-based recommendation". *The adaptive web*: 342-376.
- Suryavanshi, B. S., Shin, N. ve Mudur, S. P. (2005). "Improving the effectiveness of model based recommender systems for highly sparse and noisy web usage data". *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*. 618-621.

- Tanbeer, S. K., Ahmed, C. F., Jeong, B.-S. (2008). "Cp-tree: A tree structure for single-pass frequent pattern mining". *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. 1022-1027.
- Townsend, J. T. (1971) "Theoretical analysis of an alphabetic confusion matrix". *Perception Psychophysics*; 9 (1): 40-50.
- Tsuge, S., Shishibori, M., Kuroiwa, S. (2001). "Dimensionality reduction using non-negative matrix factorization for information retrieval". *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236)*. 960-965.
- Tyagi, S. ve Bharadwaj, K. K. (2012) "Enhanced new user recommendations based on quantitative association rule mining". *Procedia Computer Science*; 10: 102-109.
- Ujjiin, S. ve Bentley, P. J. (2003). "Particle swarm optimization recommender system". *Proceedings of the 2003 IEEE Swarm Intelligence Symposium. SIS'03 (Cat. No. 03EX706)*. 124-131.
- Van Der Maaten, L., Postma, E. ve Van den Herik, J. J. J. (2009) "Dimensionality reduction: A comparative". *Mach Learn Res*; 10 (66-71): 13.
- Van Dongen, S. ve Enright, A. J. (2012) "Metric distances derived from cosine similarity and pearson and spearman correlations". *arXiv preprint arXiv*.
- Van Meteren, R. ve Van Someren, M. (2000). "Using content-based filtering for recommendation". *Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop*. 47-56.
- Vargas, S. ve Castells, P. (2011). "Rank and relevance in novelty and diversity metrics for recommender systems". *Proceedings of the fifth ACM conference on Recommender systems*. 109-116.
- Vozalis, M. G. ve Margaritis, K. G. (2007) "Using svd and demographic data for the enhancement of generalized collaborative filtering". *Information Sciences*; 177 (15): 3017-3037.
- Wang, J., De Vries, A. P. ve Reinders, M. J. (2006). "Unifying user-based and item-based collaborative filtering approaches by similarity fusion". *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 501-508.
- Wei, K., Huang, J. ve Fu, S. (2007). "A survey of e-commerce recommender systems". *2007 international conference on service systems and service management*. 1-5.

- Wu, J. ve Lin, Z. (2005). "Research on customer segmentation model by clustering". *Proceedings of the 7th international conference on Electronic commerce*. 316-318.
- Xiang, L. ve Nan, W. (2016). "The impact of online information to the internet reservations of hotels in guangzhou". *International workshop on Mathematics and Decision Science*. 404-413.
- Yeh, I.-C., Yang, K.-J. ve Ting, T.-M. (2009) "Knowledge discovery on rfm model using bernoulli sequence". *Expert Systems with Applications*; 36 (3): 5866-5871.
- Yin, H., Cui, B., Li, J. (2012) "Challenging the long tail recommendation". *arXiv preprint arXiv*.
- Zhang, J., Craciun, G. ve Shin, D. (2010) "When does electronic word-of-mouth matter? A study of consumer product reviews". *Journal of Business Research*; 63 (12): 1336-1341.
- Zhang, L., Yu, H. ve Hu, S. (2003). "A new approach to improve particle swarm optimization". *Genetic and Evolutionary Computation Conference*. 134-139.
- Zhou, W., Wen, J., Qu, Q. (2018) "Shilling attack detection for recommender systems based on credibility of group users and rating time series". *PloS one*; 13 (5): e0196533.

EK-1 BENZERLİK ÖLÇÜMÜ MODELİ HATA MATRİSLERİ ORTALAMA DEĞERLERİ

45 Gün / Jaccard		Tahmin	
		1	0
Gerçekleşen	1	295,36	1995,25
	0	2353,12	998,14
90 Gün / Jaccard		Tahmin	
		1	0
Gerçekleşen	1	865,35	6051,92
	0	7129,21	3296,61
365 Gün / Jaccard		Tahmin	
		1	0
Gerçekleşen	1	1768,14	15542,88
	0	17946,77	3217,21
45 Gün / Kosinüs		Tahmin	
		1	0
Gerçekleşen	1	342,54	2051,54
	0	2287,36	1102,91
90 Gün / Kosinüs		Tahmin	
		1	0
Gerçekleşen	1	989,61	6223,17
	0	6798,65	3395,11
365 Gün / Kosinüs		Tahmin	
		1	0
Gerçekleşen	1	1985,67	14945,41
	0	18946,77	5217,21

EK-2 MATRİS AYRIŞTIRMA MODELİ HATA MATRİSLERİ ORTALAMA DEĞERLERİ

45 Gün / TDA		Tahmin	
		1	0
Gerçek leşen	1	454,85	1965,44
	0	2209,35	1094,93
90Gün / TDA		Tahmin	
		1	0
Gerçekl eşen	1	1254,54	5607,02
	0	6251,33	3724,4
365 Gün / TDA		Tahmin	
		1	0
Gerçek leşen	1	2345,83	13836,21
	0	14541,73	6324,04
45 Gün / NMF		Tahmin	
		1	0
Gerçek leşen	1	431,52	2099,34
	0	2356,21	1195,38
90Gün / NMF		Tahmin	
		1	0
Gerçek leşen	1	1126,12	5911,14
	0	6411,83	4098,55
365 Gün/ NMF		Tahmin	
		1	0
Gerçek leşen	1	2223,012	14142,08
	0	15124,32	6592,22

EK-3 BİRLİKTELİK KURALLARI MODELİ HATA MATRİSLERİ ORTALAMA DEĞERLERİ

45 Gün / PSO		Tahmin	
		1	0
Gerçek leşen	1	522,18	1898,25
	0	2111,01	1315,22
90Gün / PSO		Tahmin	
		1	0
Gerçek leşen	1	1487,3	5648,36
	0	6033,59	4251,23
365 Gün / PSO		Tahmin	
		1	0
Gerçekle şen	1	2499,35	11554,87
	0	12692,23	7001,21

45 Gün / FP-GROWTH		Tahmin	
		1	0
Gerçek leşen	1	492,57	1993,21
	0	2293,51	1190,24
90Gün / FP-GROWTH		Tahmin	
		1	0
Gerçek leşen	1	1350,98	5654,2
	0	6541,02	4620,02
365 / FP-GROWTH		Tahmin	
		1	0
Gerçekl eşen	1	2519,22	13012,21
	0	13246,21	7982,21

ÖZGEÇMİŞ

Adı ve SOYADI	Ömer UÇAN
EĞİTİM DURUMU	
Mezun Olduğu Lise	Adem Tolunay Fen Lisesi, Burdur, 2001
Lisans Diploması	İstanbul Üniversitesi, İşletme Fakültesi, İngilizce İşletme Bölümü, İstanbul, 2007.
Yüksek Lisans Diploması	Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, İşletme Bölümü, Antalya, 2011.
Tez Konusu	Dijital kütüphanelerde veri madenciliği uygulamaları: Akdeniz Üniversitesi Merkez Kütüphanesi örneği
Yabancı Dil / Diller	İngilizce
İŞ DENEYİMİ	
Çalıştığı Kurum	İller Bankası A.Ş. Antalya Bölge Müdürlüğü, Uzman
E-Posta	omerucan@gmail.com