

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

**ÖNERİ SİSTEMLERİ VE E-TİCARETTE ÖNERİ  
SİSTEMLERİNİN KULLANIMI**

**YÜKSEK LİSANS TEZİ**

**Şahiner GÜLER**

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM  
MÜHENDİSLİĞİ**  
**Tez Danışmanı : Doç. Dr. Kürşat AYAN**

**Eylül 2019**

T.C.  
SAKARYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

ÖNERİ SİSTEMLERİ VE E-TİCARETTE ÖNERİ  
SİSTEMLERİNİN KULLANIMI

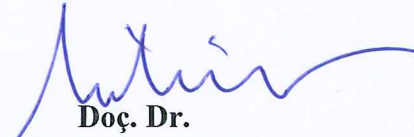
YÜKSEK LİSANS TEZİ


Şahiner GÜLER

Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM  
MÜHENDİSLİĞİ

Bu tez 26.09.2019 tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.

  
Doç. Dr.  
Kürşat AYAN  
Jüri Başkanı

  
Doç. Dr.  
Nilüfer YURTAY  
Üye

  
Dr. Öğr. Üyesi  
Ali GÜLBAĞ  
Üye

## **BEYAN**

Tez içindeki tüm verilerin ve dokümantasyonun akademik ve insanî etik bakımından tarafımda yapıldığını, kullanılan verilerde tahrifat yapmadan, başka eserlerden yararlanılan durumlarda bilimsel etik ve kurallar çerçevesinde atıfta bulunulduğunu beyan ederim.

Şahiner GÜLER

26.09.2019

## **TEŐEKKÜR**

Yüksek lisans ders ve tez dönemlerimde tecrübesini, bilgisini ve ilgisini esirgemeyen arařtırmalarımnda, tezimde ve planlanmasında beni yönlendiren, teşvik eden değerli danışman hocam sayın Doç. Dr. Kürşat AYAN'a teşekkürlerimi sunarım.

Gerek lisans gerekse yüksek lisans eğitimim ve tezimin hazırlanması süreçlerinde benden desteğini hiçbir zaman esirgemeyen değerli dostum Arş. Gör. Hüseyin ESKİ'ye, yine tezimin hazırlanması sürecinde desteklerini her zaman hissettiğim dostlarım Dr. Öğr. Üyesi Sabahattin YEŐİLÇINAR'a ve Dr. Öğr. Üyesi Kenan BULDURUN'a teşekkür ederim.

## İÇİNDEKİLER

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

### BÖLÜM 1.

GİRİŞ .....	1
1.1. Öneri Sistemlerinin Kullanım Amaçları ve Faydaları .....	1
1.1.1. Trafik çekmek .....	2
1.1.2. İlgili içerik sunmak.....	2
1.1.3. Müşterileri bağlamak.....	2
1.1.4. Satış miktarını artırmak.....	2
1.1.5. Stokta birikmiş ürünlerin eritilmesini sağlamak .....	3
1.1.6. İş yükünü azaltmak.....	3
1.1.7. Raporlama yapmak.....	3

### BÖLÜM 2.

ÖNERİ SİSTEMLERİNDE KULLANILAN YÖNTEMLER .....	4
2.1. İçerik Bazlı Filtreleme Yöntemi .....	5
2.1.1. İçerik bazlı filtreleme yöntemlerinin artıları ve eksileri.....	7
2.1.2. İçerik bazlı filtreleme yöntemi örnekleri.....	7
2.2. İşbirlikçi Filtreleme Yöntemleri .....	8
2.2.1. Hafıza bazlı işbirlikçi filtreleme yöntemleri .....	9

2.2.1.1. Kullanıcı bazlı filtreleme yöntemi .....	9
2.2.1.2. Ürün bazlı filtreleme yöntemi.....	9
2.2.2. Model bazlı işbirlikçi filtreleme yöntemleri.....	12
2.2.2.1. Tekil değer ayrıştırma.....	12
2.2.2.2. Boyutsallık indirgeme tekniği.....	13
2.2.2.3. Saklı anlamsal analiz .....	14
2.2.2.4. Model bazlı işbirlikçi filtreleme yöntemlerinde kullanılan algoritmalar .....	15
2.2.3. İşbirlikçi filtreleme yöntemlerinin artı ve eksileri.....	18
2.3. Hibrit Filtreleme Yöntemleri .....	18
2.3.1. Değişmeli hibrit yöntemi.....	19
2.3.2. Basamaklı hibrit yöntemi .....	19
2.3.3. Karışık hibrit yöntemi .....	19
2.3.4. Özellik kombinasyonu.....	20
2.3.5. Özellik büyütme yöntemi .....	20
2.3.6. Meta düzeyi yöntemi.....	20

### BÖLÜM 3.

ÖNERİ SİSTEMLERİNDE KARŞILAŞILAN PROBLEMLER.....	23
3.1. Veri Seyrekliği.....	23
3.2. Ölçeklenebilirlik .....	23
3.3. Soğuk Başlangıç .....	23
3.3.1. Sistemlere etkileri.....	24
3.3.2. Sorunu azaltma yöntemleri.....	25
3.3.3. Profil tamamlama .....	25
3.3.4. Özellik haritalaması.....	27
3.3.5. Hibrit özellik ağırlıklandırma.....	27
3.3.6. Çözümler .....	28
3.4. Gri Koyun Kullanıcılar .....	29
3.5. Şilin Saldırıları.....	30
3.6. Veri Eksikliği.....	30
3.7. Tahmin Edilemeyen Ürünler .....	30

#### BÖLÜM 4.

ÖNERİ SİSTEMLERİ UYGULAMALARI VE KARŞILAŞTIRILMALARI.....	31
4.1. Kullanılan Kütüphane ve Sınıflar .....	31
4.2. İçerik Bazlı Öneri Sistemi Uygulaması .....	32
4.3. İşbirlikçi Öneri Sistemi Uygulaması .....	33
4.4. Hibrit Öneri Sistemi Uygulaması .....	33
4.5. Öneri Sistemlerinin Karşılaştırılması .....	33

#### BÖLÜM 5.

SONUÇ ve ÖNERİLER.....	37
KAYNAKLAR .....	38
ÖZGEÇMİŞ .....	46

## **SİMGELER VE KISALTMALAR LİSTESİ**

<b>BFM</b>	: Beş Faktör Modeli
<b>HBİFY</b>	: Hafıza Bazlı İşbirlikçi Filtreleme Yöntemi
<b>İBFY</b>	: İçerik Bazlı Filtreleme Yöntemi
<b>İFY</b>	: İşbirlikçi Filtreleme Yöntemi
<b>KBFY</b>	: Kullanıcı Bazlı Filtreleme Yöntemi
<b>MBFY</b>	: Model Bazlı İşbirlikçi Filtreleme Yöntemi
<b>SAA</b>	: Saklı Anlamsal Analiz
<b>TDA</b>	: Tekil Değer Ayrıştırma
<b>ÜBFY</b>	: Ürün Bazlı Filtreleme Yöntemi



## ŞEKİLLER LİSTESİ

Şekil 2.1. Öneri Sistemleri Yöntemleri.....	4
Şekil 2.2. Öneri Sistemi Çalışma Modeli.....	22

## TABLolar LİSTESİ

Tablo 2.1. Ürünler Arası Benzerlik Tablosu.....	7
Tablo 2.2. Kullanıcı – Ürün Beğenme Değerleri Tablosu .....	7
Tablo 2.3. İşbirlikçi Filtreleme Yöntemlerine Genel Bakış.....	21
Tablo 4.1. İçerik Bazlı Öneri İçin En Çok Etkileşime Giren Kullanıcıların Listesi	36
Tablo 4.2. İşbirlikçi Öneri İçin En Çok Etkileşime Giren Kullanıcıların Listesi ...	36
Tablo 4.3. Hibrit Öneri İçin En Çok Etkileşime Giren Kullanıcıların Listesi .....	36
Tablo 5.1. Yöntemlerin Karşılaştırılması.....	37

## ÖZET

Anahtar kelimeler: Öneri sistemleri, içerik bazlı öneri sistemleri, işbirlikçi öneri sistemleri, hibrit içerik sistemleri, e-ticaret.

İnternet üzerinden alışveriş, artık hayatımızda alışveriş alışkanlıklarımızı büyük bir ölçüde değiştirdi ve değiştirmeye de devam ediyor. Günün her saatinde her kategoride ürüne ulaşmamız için ihtiyacımız olan sadece internetle birlikte bir bilgisayar, akıllı telefon veya bir tablet. Bu sayede, ihtiyaç duyduğumuz veya almayı düşündüğümüz herhangi bir ürün için farklı alternatiflere, farklı kalite ve fiyatlara ulaşmamız artık çok kolay. Yaşadığımız ülkenin hatta dünyanın herhangi bir noktasından kapımıza kadar ürünün teslimini sağlayabiliyoruz. Öneri sistemleri, alınması düşünülen ürünler için aynı veya benzer ürünleri önceden almış başka müşterilerin yorum, değerlendirme ve oylarının da yardımıyla alternatif seçenekler veya tamamlayıcı başka ürünler önererek yapılacak alışverişi çok kolaylaştırmaktadır. Bu sayede zamandan tasarruf edilmesi, ihtiyaca daha yakın ürünlerin incelenmesi sağlanmış olur.

# **RECOMMENDATION SYSTEMS AND THE USAGE OF RECOMMENDATION SYSTEMS IN E-COMMERCE**

## **SUMMARY**

Keywords: Recommendation systems, content-based recommendation systems, collaborative recommendation systems, hybrid recommendation systems, e-commerce

Online shopping has changed our shopping habits to a great extent now and continues to change. In order to reach a product in any categories at any time of the day, we just need to have a computer, a smartphone or a tablet which has Internet connection. Thus, it becomes easy to reach different alternatives, different quality and prices for any product that we need or want to buy. Thanks to Internet, we can deliver the product from any place of the world. Suggestion systems make shopping much easier for the products to be bought by offering alternative options or complementary products with the help of the comments, evaluations and votes of other customers who have already bought the same or similar products for the intended products. Therefore, they help save time and examine the products that are most needed.

## **BÖLÜM 1. GİRİŞ**

Öneri Sistemleri, kullanıcıların internet üzerinden alışveriş yaparken, haber okurken veya film izlerken farklı haber ve ürünlerden haberdar olmalarını ve böylece yeni ürünler keşfetmelerini sağlayan sistemlerdir.

Herhangi bir alışveriş sitesinden ürün bakılmaya başlanmasından itibaren, incelenen ürünlere benzer ürünler hem alışveriş yaptığımız sitede hem de internette dolaşırken örneğin bir haber sitesinin reklamlar kısmında karşımıza çıkabiliyor.

İnternet üzerindeki alışveriş, film, oyun sitelerinde dolaşırken sayfaların alt veya yan kısımlarında “en çok satılan ürün”, “bu ürün grubundaki diğer ürünler”, “en çok satan kitaplar”, “kullanıcılar şu ürünleri de inceledi”, “beğenebileceğiniz oyun veya filmler”, “ilginizi çekebilecek haberler” gibi başlıklar altında web siteleri tarafından ilgili oldukları ürün veya mesajlarla bize tavsiyede bulunulur.

Günlük yaşantımızda uyku ve çalışma hayatımız dışındaki zamanımızın büyük bir bölümünü internette harcadığımızı düşünürsek, internetin ticarî anlamda çok büyük bir pazar olması kaçınılmaz olmaktadır. Bu pazarda, web sitelerinin türüne göre ziyaretçileri için sunulan ürün ve hizmet tavsiyeleri hem ziyaretçilerin işini kolaylaştıracaktır hem de daha çok ürün satışını sağlayacaktır. Bir öneri sistemi oluşturmak için matematiksel modeller kullanan tekniklere ihtiyaç vardır.

### **1.1. Öneri Sistemlerinin Kullanım Amaçları ve Faydaları**

Hayatımızın neredeyse vazgeçilmezi noktasına gelen internet artık alışveriş alışkanlıklarımızı da tümüyle değiştirmek üzeredir. Öneri sistemleri, internette

alışveriş yapmak için harcadığımız zamanı hem daha verimli bir şekilde kullanmamıza yardımcı olur hem de aradığımız ürünlere yakın ürünler önererek alternatifler sunar.

#### **1.1.1. Trafik çekmek**

Bir tavsiye motoru, kişiselleştirilmiş e-postalarla ve reklamlarla siteye trafik getirebilir.

#### **1.1.2. İlgili içerik sunmak**

Öneri motoru tarafından yapılan tavsiyeler sayesinde müşterinin araştırdığı, gezdiği ürünler, göz atma geçmişi analiz edilerek, ilgilenebileceği farklı ürünleri keşfetmesi sağlanır. Veriler gerçek zamanlı olarak toplanıp, işlendiği için daha isabetli tavsiyeler yapılması sağlanır.

#### **1.1.3. Müşterileri bağlamak**

Kişiselleştirilmiş ürün tavsiyeleri yapılması kişinin siteye ilgisini daha da artıracak ve kişinin amacına daha kolay ulaşmasında etkili olacaktır. Kişinin dolaştığı sitede kendisine özel ürünlerin tavsiye edilmesi sayesinde site tarafından kendisine değer verildiği, önemsendiği duygusu artacaktır ve bu da müşterinin siteye olan sadakatini artırıp, öneri motoru sayesinde sitenin kendi müşterisini oluşturmasına yardımcı olacaktır.

#### **1.1.4. Satış miktarını artırmak**

Öneri motorunun kişiselleştirilmiş ürün önerileri sayesinde müşterilerin ilgi alanına göre ürün tavsiye edilmesi müşterilerin sipariş miktarlarının ve her siparişteki ürün miktarının artması kaçınılmazdır.

### **1.1.5. Stokta birikmiş ürünlerin eritilmesini sağlamak**

Öneri sistemlerinde hangi ürünlerin ön plana çıkarılıp, vurgulanabileceğinin kontrol altında olması sayesinde yeni veya fazla stoklanmış ürünlerin satışı artırılabilir. Müşterilerin promosyonlu ürünlere daha çabuk ve kolay ulaşması sağlanabilir.

### **1.1.6. İş yükünü azaltmak**

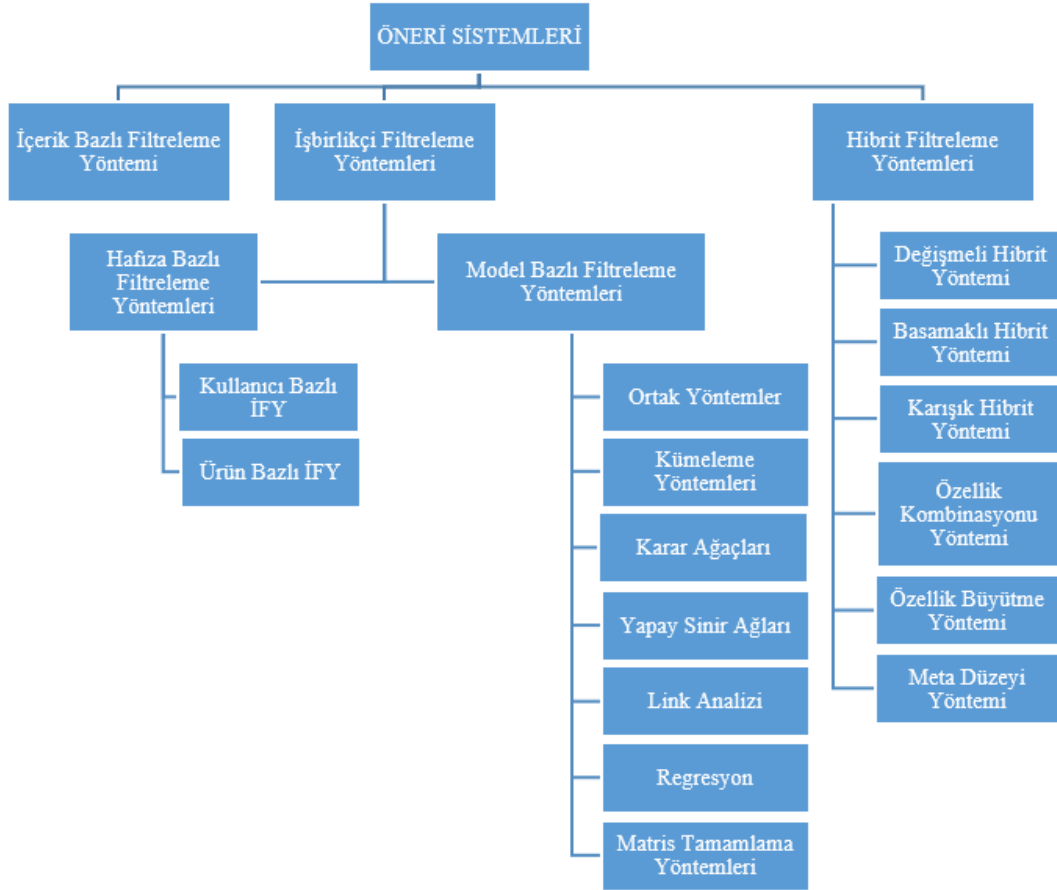
Her bir müşteri için kişiselleştirilmiş bir alışveriş sunmak gerek veri hacminin büyüklüğü gerekse müşterinin ilgilendiği ürünler arasında bir ilişki kurup, kendisine uygun ürünler tavsiye etmek, yönetilmesi çok zor olan bir iş yükü gerektirecektir. Bu işlemlerin bir yazılım sayesinde otomatik olarak yapılması hem personelin iş yükü hem de bu işlemler için harcanması gereken bütçeye çok büyük bir tasarruf sağlayacaktır.

### **1.1.7. Raporlama yapmak**

Raporlama yapılması, bir sistemde kişiselleştirilmiş veriler sunulması açısından olmazsa olmazdır. Müşterilere doğru ve anlık raporlar sunulması müşterilerin kendilerine daha çok hitap eden ürünlerle ilgilenmesini sağlayacaktır. Öneri motoru doğru ve yerinde önerilerle müşterilere bir satış danışmanı gibi yardımcı olup, satışların ve müşteri memnuniyetinin artmasını sağlayacaktır [1-3].

## BÖLÜM 2. ÖNERİ SİSTEMLERİNDE KULLANILAN YÖNTEMLER

Şekil 2.1.'de öneri sistemlerinde kullanılan yöntemler ve her yöntemin altında farklı tekniklere göre sınıflandırılmış alt yöntemler gösterilmektedir. Şekildeki her bir yöntem hakkında detaylı bilgi ilerleyen bölümlerde verilecektir.



Şekil 2.1. Öneri Sistemleri Yöntemleri



## 2.1. İçerik Bazlı Filtreleme Yöntemi

Bu yöntem, ürün açıklamaları ve kullanıcı profillerine dayanır [6, 7]. İçerik bazlı bir öneri sisteminde anahtar kelimeler, ürünleri tanımlamak için kullanılır ve kullanıcılar tarafından beğenilen ürünlerin türlerinin belirlenmesi için kullanıcı profilleri oluşturulur. Yani kullanıcının önceki beğenileri ve şu anda incelemekte olduğu ürünlere benzer ürünler önerir.

Önerilecek ürünler özellikle kullanıcı tarafından değerlendirilmiş ürünler ile karşılaştırılıp, en uygun ürünler önerilir. Bu yöntemin temeli, bilgi edinme ve bilgi filtreleme araştırmaları üzerinedir.

Sistemdeki ürünlerin özelliklerinin ayıklanması için bir sunum algoritması uygulanır. Yaygın olarak kullanılan algoritmalarından bir tanesi de Terim Sıklığı / Ters Doküman Sıklığı (TF / IDF)' dir [4, 5].

Bir kullanıcı profili oluşturulabilmesi için iki tür bilgi gereklidir.

1. Kullanıcıların tercihleri
2. Öneri sistemindeki kullanıcıların önceki etkileşimleri

Bu yöntem, temel olarak ürünleri sistem içinde tanımlamak için nitelik ve özelliklerine göre bir ürün profili oluşturur. Kullanıcı için ağırlıklı ürün özellikleri vektörü yardımıyla bir profil oluşturulur. Bu vektördeki ağırlıklar ürünün her bir özelliğinin kullanıcı için önemini gösterir ve çeşitli yöntemlerle ayrı ayrı değerlendirilmiş içerik vektörleri kullanılarak hesaplanır.

Karmaşık olmayan basit yaklaşımlar derecelendirilmiş ürün vektörlerinin ortalama değerlerini kullanırken, daha karmaşık yaklaşımlar kullanıcının ürünü beğenme ihtimali üzerinde de durarak Bayesian Sınıflandırıcıları, kümeleme analizi, karar ağaçları ve yapay sinir ağları gibi makine öğrenme tekniklerini kullanır [8].

Genelde ürün üzerinde beğenme ya da beğenmeme seçeneği gibi doğrudan bir geri bildirim, belirli özelliklere göre bazen daha yüksek ya da daha düşük ağırlıklar vermek için kullanılabilir

Bu yöntem yardımıyla aynı zamanda bir kullanıcının bir içerikle ilgili davranışlarından yararlanılarak başka bir içerik hakkında değerlendirme yapılması sağlanabilir. Kişiyeye göre özelleştirme ve ilgili olabilecek ürün, film ve haber tavsiyelerinin daha yararlı olması için sezgisel, daha kullanışlı bir yaklaşıma ihtiyaç duyulmuştur. Bu teknikte tahmin ve tavsiyelerin üretilmesi için ürünlerin içerikleri yani özellikleri daha fazla önem arz etmektedir. Bu yüzden yayın, haber gibi önerilerde içerik bazlı filtreleme yöntemi en başarılı olan yöntemdir.

Kullanıcıların önceki değerlendirilen içeriklerinden daha çok olumlu olan özellikler kullanılarak öneriler yapılır.

İBFY, anlamlı öneriler üretmek amacıyla büyük veri kümelerindeki belgeler arasında benzerlik ve bağlantı bulabilmek için Terim Sıklığı / Ters Doküman Sıklığı (TF / IDF) gibi Vektör Uzay Modelleri ya da Naive Bayes Sınıflandırıcısı, Karar Ağaçları veya Sinir Ağları gibi Olasılık Modelleri kullanır [9 – 12].

Yapılan bu öneriler ya istatistiksel analizler ya da makine öğrenme teknikleriyle elde edilirler. İBFY tarafından yapılan önerilerde kullanıcı profillerinin bir etkisi yoktur çünkü bu teknikte önemli olan ürünün içeriği yani özellikleridir. Bu yüzden kullanıcı profili değişse de önerilen ürünler etkilenmez. Bu teknikte temel ihtiyaç, ürün özelliklerinin tam olarak belirtilmesidir.

Aynı ürün grubuna ait ürünler arasındaki benzerlik değerlerini gösteren Tablo 2.1.'e göre; değer, 10'a yaklaştıkça ortak özellikler artmakta, 10'dan uzaklaştıkça ortak özellikler azalmaktadır. Örneğin; Ürün 1 ile Ürün 3, Ürün 4 ve Ürün 6 arasında, Ürün 2 ile Ürün 5 arasında oldukça fazla ortak özellik vardır.

Tablo 2.1. Ürünler Arası Benzerlik Tablosu

	Ürün 1	Ürün 2	Ürün 3	Ürün 4	Ürün 5	Ürün 6
Ürün 1	10	2	8	9	4	8
Ürün 2	2	10	1	4	8	3
Ürün 3	8	1	10	2	3	4
Ürün 4	9	4	2	10	2	5
Ürün 5	4	8	3	2	10	2
Ürün 6	8	3	4	5	2	10

Tablo 2.2.'de 1. kullanıcı Ürün 3 ve Ürün 4'ü, 2. kullanıcı Ürün 1'i ve 3. kullanıcı da Ürün 3'ü oldukça beğenmiştir. Tablo 2.1.'deki ürün benzerlikleri değerlerinden hareketle 1. kullanıcıya Ürün 6, 2. kullanıcıya Ürün 3, Ürün 6 ve Ürün 4, 3. kullanıcıya da Ürün 1 tavsiye edilebilir.

Tablo 2.2. Kullanıcı – Ürün Beğenme Değerleri Tablosu

	Beğeniler (1 – 5 Arası)				
	1. Kullanıcı	2. Kullanıcı	3. Kullanıcı	...	n. Kullanıcı
Ürün 1	–	5	1	...	–
Ürün 2	–	3	1	...	–
Ürün 3	5	–	5	...	–
Ürün 4	4	–	2	...	–
Ürün 5	–	1	–	...	–
Ürün 6	–	–	–	...	–

### 2.1.1. İçerik bazlı filtreleme yöntemlerinin artıları ve eksileri

İBFY ile kullanıcılar tarafından bir değerlendirme olmadan da yeni ürünler önerilebilir. Veritabanında kullanıcı tercihleri olmasa da önerilerin doğruluğu etkilenmez. Kullanıcılar kendi profillerini paylaşmadan da öneriler alabilirler. Bu yöntemle, kullanıcılara önerilerin nasıl üretildiğine dair açıklamalar da sağlanabilir. İBFY, ürünlerin meta verilerine dayanır [13 – 15].

### 2.1.2. İçerik bazlı filtreleme yöntemi örnekleri

News Dude, okuyucularına haberleri okumak için sentezlenmiş konuşmayı sunan bir haber sistemidir. Terim Sıklığı / Ters Doküman Sıklığı (TF / IDF) modelini kullanarak haber hikayelerini tanımlar. Kosinüs Benzerlik Ölçümü (Cosine Similarity Measure) ile karşılaştırma yapılan haberler bir öğrenme algoritmasına sunularak kısa öneriler belirlenir.

CiteSeer, çeşitli sezgisel tarama ve makine öğrenme algoritmaları kullanarak dokümanları işleyen otomatik bir alıntı indeksleme sistemidir. Bu sayede internet üzerinden kullanılan en geniş ve en çok tercih edilen araştırma makalesi depolarından biri olmuştur [16].

LIBRA, kitaplarla ilgili bilgileri internetten toplayan içerik bazlı bir kitap öneri sistemidir. Naive Bayes Sınıflandırıcısı kullanarak sıralı bir başlık listesi oluşturacak kullanıcı profilini öğrenir. En yüksek puanlamaya katkıda bulunan özellikler listelenir. Bu sayede kullanıcıların sisteme güvenmeleri sağlanmış olur [17].

## 2.2. İşbirlikçi Filtreleme Yöntemleri

İFY, daha geniş bir kullanım alanına sahiptir. Bu yaklaşımla, kullanıcıların tercihleri ve ilgi alanları gibi bilgiler toplanıp, analiz edilerek kullanıcı özelliklerinin birbirleriyle olan benzerliklerine göre neyi beğenebilecekleri tahmin edilmeye çalışılarak tavsiyede bulunulur. Kullanıcı ve ürün benzerliklerini ölçmek ve değerlendirmek için K – En Yakın Komşu Yaklaşımı ve ilk olarak Allen tarafından kullanılan Pearson Korelasyonu gibi çok sayıda algoritma kullanılır. [20]

İFY, kişilerin önceden beğendiklerinin benzerlerini sonraki zamanlarda da beğenecekleri varsayımına dayanır. Açık veri toplama örnekleri şu şekillerde toplanmaktadır:

- Kullanıcıların bir ürünü derecelendirmeleriyle
- Kullanıcıların yaptıkları aramalarla,
- Kullanıcıların favori ürünlerini beğenilme derecelerine göre sıralamalarıyla,
- Kullanıcılara iki ürün sunulup, bunlardan daha iyi olanını seçmeleriyle,
- Kullanıcıların beğendikleri bir ürün listesi oluşturmalarıyla,
- Kullanıcıların ürünleri görüntüleme süreleriyle [21]

İFY, kullanıcılar tarafından ürünler hakkında yapılan tercihlere göre kullanıcı – ürün matrisi ile elde edilen bir veritabanı oluşturur. Daha sonra kullanıcı profilleri

arasındaki benzerliklere göre ilgi ve tercihlere göre eşleştirmeler yapar. Kullanıcı, kendisiyle aynı bölgedeki kullanıcılar tarafından beğenilen ancak kendisinin henüz değerlendirmedeği ürünler için öneriler alır [21, 22].

İşbirlikçi yöntemleri, Hafıza Bazlı ve Model Bazlı Filtreleme Yöntemleri olarak sınıflandırılmıştır. HBİFY'nin en iyi bilinen yaklaşımlarından bir tanesi Kullanıcı Bazlı Filtreleme Yöntemi, MBİFY'nin en iyi bilinen yaklaşımlarından bir tanesi de Çekirdek Haritalama Yöntemi'dir [22, 23].

### **2.2.1. Hafıza bazlı işbirlikçi filtreleme yöntemleri**

Daha önceden kullanıcı tarafından değerlendirilen ürünler, kullanıcıyla beğenilerini paylaşan bir komşunun aranmasında uygun bir rol oynar. Bir kullanıcı için bir komşu bulunduğunda, yeni öneriler üretmek için komşuların tercihlerini birleştirecek farklı algoritmalar kullanılabilir. Bu tekniklerin etkinliği sayesinde kullanılan uygulamalarda büyük oranda başarı elde edilmiştir. HBİFY, Kullanıcı Bazlı ve Ürün Bazlı olmak üzere iki farklı şekilde elde edilebilir [26, 27].

#### **2.2.1.1. Kullanıcı bazlı filtreleme yöntemi**

KBFY, öncelikle kullanıcıların aynı ürün hakkındaki derecelendirmelerini karşılaştırarak kullanıcılar arasındaki benzerliği ve daha sonra da bir ürün için tahmin edilecek derecelendirmeyi hesaplar.

#### **2.2.1.2. Ürün bazlı filtreleme yöntemi**

ÜBFY, kullanıcılar arasındaki benzerlikle ilgilenmez, ürünler arasındaki benzerlikleri kullanarak tahminleri hesaplar. Aktif kullanıcı tarafından derecelendirilen tüm öğeleri kullanıcı – ürün matrisinden alarak bir ürün benzerlikleri modeli oluşturur. Matristen alınan bu ürünlerin hedef ürüne benzerlik derecelerini daha sonra da en çok benzeyen ürünleri belirler. Benzer ürünlerde aktif kullanıcının derecelendirmesinin ağırlıklı ortalaması alınarak tahmin yapılır.

Kullanıcı – ürün benzerliğinin hesaplanabilmesi için birkaç farklı benzerlik ölçüm türü vardır. En çok bilinen iki tanesi korelasyon temelli ve kosinüs temelli olanlarıdır.

$$s(a, u) = \frac{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i=1}^n (r_{u,i} - \bar{r}_u)^2}} \quad (2.1)$$

$s(a, u)$ , yukarıdaki formülden  $a$  ve  $u$  kullanıcıları arasındaki benzerliği gösterir.  $r_{-}(a, i)$ ,  $a$  kullanıcısının  $i$  ürünü için verdiği puandır.  $(r_{-}a)$ ,  $a$  kullanıcısı tarafından verilen ortalama puan,  $n$  ise kullanıcı – ürün listesindeki toplam ürün sayısıdır. Ayrıca bir ürün için yapılacak tahmin, seçilmiş komşuların oranlarının ağırlıklı kombinasyonundan oluşturulur. Bu komşuların oranları, komşuların ortalamasından ağırlıklı sapma olarak hesaplanarak bulunur.

Genel tahmin formülü:

$$p(a, i) = \bar{r}_a + \frac{\sum_{i=1}^n (r_{u,i} - \bar{r}_u) \times s(a, u)}{\sum_{i=1}^n s(a, u)} \quad (2.2)$$

Kosinüs benzerliği, Pearson tabanlı ölçümden farklıdır. Şöyle ki; kosinüs benzerliği, istatistiksel yaklaşım yerine lineer cebire dayalı bir vektör uzay modelidir. Aralarındaki açıya göre iki tane  $n$  boyutlu vektör arasındaki benzerliği ölçer.

Kosinüs tabanlı ölçüm, iki metin dosyasını karşılaştırmak için bilgi edinme, metin madenciliği alanlarında yaygın olarak kullanılır, bu karşılaştırmada dokümanlar, terimlerin vektörleri olarak gösterilir.  $u$  ve  $v$  ürünleri arasındaki benzerlik aşağıdaki gibi tanımlanır [28 – 30].

$$\text{benzerlik} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}} \quad (2.3)$$

Örneğin;

- A cümlesi: Öneri sistemleri alışveriş alışkanlıklarımızı etkiler.
- B cümlesi: Alışveriş merkezleri alışkanlıklarımızı değiştirir.

İki cümledeki kelimeler bir kümede toplanır, aynı kelimedenden birden fazla adette varsa bir tanesi yazılır.

Ortak Küme: {öneri, sistemleri, alışveriş, alışkanlıklarımızı, etkiler, merkezleri, değiştirir}

Ortak kümenin elemanlarının A ve B cümlelerinde olup olmadığı kontrol edilir, eğer varsa oluşturulacak matriste 1 yazılır yoksa 0 yazılır. Öneri kelimesi A cümlesinde vardır, B cümlesinde yoktur, alışveriş kelimesi hem A cümlesinde hem de B cümlesinde vardır.

$$A = [1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0]$$

$$B = [0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1]$$

$$A.B = 1.0 + 1.0 + 1.1 + 1.1 + 1.0 + 0.1 + 0.1 = 2 \quad (2.4)$$

$$||A|| = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2} = \sqrt{5} \quad (2.5)$$

$$||B|| = \sqrt{0^2 + 0^2 + 1^2 + 1^2 + 0^2 + 0^2 + 1^2} = \sqrt{3} \quad (2.6)$$

$$\text{benzerlik} = \cos(\theta) = \frac{A.B}{||A|| ||B||} = \frac{2}{\sqrt{5} \cdot \sqrt{3}} = \frac{2}{\sqrt{15}} = \frac{2}{3,87} = 0,5167 \quad (2.7)$$

İki cümlelerin benzerlik oranı %51,67'dir.

### 2.2.2. Model bazlı işbirlikçi filtreleme yöntemleri

Bu yöntem, İFY'nin performansını artıracak yeni bir model öğrenmek için önceki ürün değerlendirmelerini kullanılır. Bu modelin oluşturulma işlemleri, makine öğrenmesi ya da veri madenciliği teknikleriyle yapılabilir. Bu teknikler, önceden hesaplanan bir model sayesinde bir ürün setini hızlı bir şekilde önerebilirler. Ayrıca komşuluk tabanlı tavsiye sistemlerine benzer sonuçlar sağladıkları da kanıtlanmıştır.

Bu teknikler arasında Tekil Değer Ayrıştırma (Singular Value Decomposition – SVD) gibi Boyutsallık Azaltma Tekniği (Dimensionality Reduction Technique), Matris Tamamlama Tekniği (Matrix Completion Technique), Saklı Anlamsal Yöntemler (Latent Semantic Methods) and Regresyon (Regression) and Kümeleme (Clustering) gibi teknikler vardır.

#### 2.2.2.1. Tekil değer ayrıştırma

Tekil Değer Ayrıştırma (TDA), karmaşık bir matrisi daha basit matrislerin çarpımı şeklinde gösterilmesini sağlayan bir işlemdir. Bu işlemlerde en çok kullanılan algoritmalar, SVD, QR, LU, CS Ayrışımı, QLP Ayrışımı, Pivoted QR ve Pivoted QLP Ayrışımı'dır [31].

Özellikle matrisin köşegenleştirilmesi ile sıkı bir ilişkisi olan TDA, tüm matris ayrışım algoritmalarının başında gelerek, özellikle regresyon analizinde önemli bir yer tutmaktadır [32].

Örneğin, 
$$X = \begin{bmatrix} 3.01 & 0.01 & -2.99 \\ 2.99 & -0.01 & -3.01 \\ 2.00 & -4.00 & 2.00 \end{bmatrix}$$
 olsun

$X$  matrisi aşağıdaki gibi ayrıştırılabilir [32]:

$$X = \begin{bmatrix} 3 & 0 & -3 \\ 3 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & -4 & 2 \end{bmatrix} = \begin{bmatrix} 0.01 & 0.01 & 0.01 \\ -0.01 & -0.01 & -0.01 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.8)$$



$$= 6 \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} + 2\sqrt{6} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} + \frac{\sqrt{6}}{100} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & 0 & 0 \end{bmatrix} \quad (2.9)$$

$$6 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} + 2\sqrt{6} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} + \frac{\sqrt{6}}{10} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \quad (2.10)$$

$$= \theta_1 U_1 V_1 + \theta_2 U_2 V_2 + \theta_3 U_3 V_3 = \sum_{i=1}^3 \theta_i U_i V_i^T \quad (2.11)$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2\sqrt{6} & 0 \\ 0 & 0 & \frac{\sqrt{6}}{100} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} = \theta UV^T \quad (2.12)$$

Görülüyor ki, denklem (4)'teki  $U_i$ , denklem (5)'teki  $U$  orthonominal matrisinin sütunudur. Denklem (4)'teki  $V_i^T$  ise, denklem (5)'teki  $V^T$  orthonominal matrisinin satırdır.  $\theta_i$  ise  $\theta$  köşegen matrisinin köşegen öğesidir. Denklem (4) ve denklem (5),  $X$  matrisinin tekil değer ayrışımının açıklamasıdır [32].

#### 2.2.2.2. Boyutsallık indirgeme tekniği

Günlük artan veri miktarları, bu verilerin takibini zorlaştırmaktadır. Çok büyük oranlarda artan bu verilerle aynı zamanda verilerin görselleştirilmesi de çok zor olmaya başlıyor. Değişken sayısının çok olduğu durumlarda bu veri setini görselleştirmek ve içinden çıkarımlar yapmak oldukça zordur. Bu teknikler sayesinde büyük miktardaki veri setinden alt kümeler çıkarılarak yorum ve çıkarımların daha kolay yapılabilirdiği normal boyutta bir veri seti elde edilebilir.  $X$  boyutunda bir veri seti  $Y$  boyutunda bir veri setinin alt kümesine dönüştürülebilir. Buna Boyutsallık İndirgeme denir [90].

### 2.1.1.1. Saklı anlamsal analiz

Saklı Anlamsal Analiz, kelime ve kelime gruplarının metinlerde kullanımlarını analiz etme yöntemidir. Aşağıdaki sorulara cevap vermek için kullanılır.

- Kullanılan metnin temel anlamı nedir?
- Kelimelerin, kullanıldıkları bölüme etkisi nedir?
- Bir bölümdeki kelimelerin yaklaşık anlamlarının, bölümün genel anlamıyla ilgisi nedir?

Bir dilde birden fazla anlam taşıyan (eş sesli) kelimeler, dilin anlaşılmasını karmaşılaştırır. Kelimelerin anlamları, içinde kullanıldıkları içeriğe bağlı olarak değişir. Birden fazla anlama sahip olan bir kelimenin kullanılacağı cümlede kullanılmak istenen anlamından tamamen farklı bir anlamda kullanılması, anlatılmak istenen düşüncenin çok farklı bir boyuta çekilmesine sebep olabilir. Dolayısıyla da bir cümlenin birden fazla anlamı ortaya çıkabilmektedir.

Metnin gerçek anlamına ulaşılabilmesi için de sözcükler ve bağlı oldukları kavramlar karşılaştırılır. Saklı Anlamsal Analiz (SAA – Latent Semantic Analysis), aynı zamanda Saklı Anlamsal İndeksleme (Latent Semantic Indexing) olarak da adlandırılır.

SAA, matrisleri çarpanlarına ayırmak için Tekil Değer Ayrıştırma – TDA adı verilen gelişmiş bir matris cebir yöntemi kullanır.

Tekil Değer Ayrıştırma yöntemi, karmaşık algoritmaları kullanmak için geliştirildiği için küçük boyuttaki metinler için pek pratik değildir.

Temel olarak;

- Metin, küçük matrislere dönüştürülür. Matristeki her bir hücre, bir kelimenin metin içindeki bölümde kaç defa kullanıldığını gösterir.

- Matris, her bölümün bir vektör olarak temsil edilmesini sağlayan çarpanlarına ayrılır.
- Hedef ürünler, kosinüs ve benzer ölçümler bölümler arasındaki benzerlikleri göstermek için kullanılır [33].

Genel tahmin formülü:

$$p(a, i) = \bar{r}_a + \frac{\sum_{i=1}^n (r_{u,i} - \bar{r}_u) \times s(a, u)}{\sum_{i=1}^n s(a, u)} \quad (2.13)$$

Kosinüs benzerliği, Pearson tabanlı ölçümden farklıdır. Şöyle ki; kosinüs benzerliği, istatistiksel yaklaşım yerine lineer cebire dayalı bir vektör uzay modelidir. Aralarındaki açığa göre iki tane  $n$  boyutlu vektör arasındaki benzerliği ölçer.

Kosinüs tabanlı ölçüm, iki metin dosyasını karşılaştırmak için bilgi edinme metin madenciliği alanlarında yaygın olarak kullanılır, bu karşılaştırmada dokümanlar, terimlerin vektörleri olarak gösterilir.  $u$  ve  $v$  ürünleri arasındaki benzerlik aşağıdaki gibi tanımlanır [28 – 30].

$$s(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}| * |\vec{v}|} = \frac{\sum_i r_{u,i} r_{v,i}}{\sqrt{\sum_i r_{u,i}^2} \times \sqrt{\sum_i r_{v,i}^2}} \quad (2.14)$$

### 2.1.1.2. Model bazlı işbirlikçi filtreleme yöntemlerinde kullanılan algoritmalar

Model bazlı teknikler, ürünler arasındaki ilişkileri tanımlamak için kullanıcı – ürün matrisini analiz eder ve bu ilişkileri de en çok önerilen ürünler listelerini karşılaştırmak için kullanırlar. Ayrıca öğrenme algoritmaları, kullanıcıların ilgilendikleri ürünleri ne zaman tüketmeleri gerektiği konusunda da tavsiyelerde bulunmaktadır. Bu nedenle bu tekniklerde kullanıcıların algoritmaların incelenmeleri önem arz etmektedir.

Kümeleme: Kümeleme teknikleri, desen tanımlama, görüntü işleme, istatistiksel veri analizi gibi farklı alanlara uygulanmıştır. Kümeleme algoritması, kullandığı veri kümesi içinde anlamlı gruplar tespit etmek için verileri alt kümelere ayırmaya çalışır.

Kümeler oluşturulduktan sonra bir kümedeki kullanıcıların fikirlerinin ortalaması alınabilir ve kullanıcılar için önerilerde bulunmakta kullanılabilir.

Küme içi yüksek benzerlik ve kümeler arası düşük benzerlik içeren gruplar iyi bir kümeleme yöntemiyle yüksek kalitede olacaktır. Bazı kümeleme yaklaşımlarında bir kullanıcının farklı kümelerde kısmî katkısı olabilir ve tavsiyeler bu kümelerin ortalaması üzerine kurulur. K – Ortalama (K – Means) ve Kendini Düzenleyen Harita, farklı kümeleme metodları arasında genelde en çok kullanılanlardır. K – Ortalama (K – Means), bir giriş parametresi alır ve daha sonra n elemanlı K kümelerine ayırır.

Kendini Düzenleyen Harita, yapay nöronları kümeleme tekniği üzerine kurulu bir denetimsiz öğrenme metodudur. Kümeleme teknikleri, işbirlikçi tabanlı algoritmalarındaki küme adaylarını azaltmak için kullanılır [34 – 38].

Karar Ağaçları: Karar ağacı ile amaç, eldeki verilerin ağaç dalları şeklinde benzer gruplara ayrılmasıdır. Daha sonra önceden görülmemiş örnekleri sınıflandırmak için uygulanır.

Kullanılan verinin kalitesi, yapılacak tahminlerin doğruluğunu doğrudan etkiler [39]. Karar ağaçları, Destek Vektör Makinesi ve Sinir Ağları gibi sınıflandırıcılardan daha anlaşılabilirler ve daha iyi yorumlanabilirler. Ayrıca karar ağaçları, gerçek verilerle kategorik özelliklerin karışık olduğu verilerin ve bazı eksik özelliklerin işlenmesinde daha esneklerdir.

Yapay Sinir Ağları: Bu algoritma türünde, birbirlerine sistematik bir şekilde bağlı katmanlara yerleştirilmiş çok sayıda nöronun (düğüm) oluşmaktadır. Nöronlar arasında her birinin diğeri üzerindeki etki miktarına bağlı olarak birbirleriyle ilişkili ağırlıkları vardır. Bazı özel problemlerin çözümünde yapay sinir ağlarının kullanımında bazı avantajlar vardır.

YSA, birçok nöron ve bu nöronların her birinin arasında bağlantılar içerir. Buna rağmen hatalı veri kümelerine oranla daha sağlamdır. Doğrusal olmayan fonksiyonları

tahmin edebilir ve veri kümelerindeki karmaşık ilişkileri tespit edebilir. Ayrıca ağın bir kısmı başarısız olsa bile ağ verimli bir şekilde çalışabilir [40].

**Link Analizi:** Link Analizi, birbirlerine bağlı nesnelere arasında desen ve eğilimleri keşfetmek için gerekli ağ kurma sürecidir. Link Analizi sayesinde web aramalarının başarısında büyük ilerleme sağlanmıştır. Link Analizi, Sayfa Sıralaması (Page Rank) ve HITS Algoritmalarını içerir. Çoğu link analizi algoritması web grafiğinde bir web sayfasını tek bir düğüm olarak ele alır [41, 42].

**Regresyon Analizi:** Burada amaç, iki veya daha fazla değişken arasındaki ilişkiyi belirlemektir. Bu değişkenler arasında sebep – sonuç ilişkisi olmalıdır ki bu ilişki kullanılarak konu ile ilgili tahminler yapılabilir [43].

Örneğin;

– Sebep	Sonuç
– Emek	Kazanç
– Kitap okuma	Zihinsel gelişim
– Örnek sayısı	Konuyu öğrenme
– Uyku süresi	Dinlenme miktarı

**Matris Tamamlama Tekniği:** Matris tamamlama tekniğinin amacı, ürün – kullanıcı matrisinde bilinmeyen ya da eksik kalan elemleri öngörmektir. Korelasyon tabanlı K – en yakın komşu, işbirlikçi filtreleme yöntemlerinde kullanılan en önemli tekniktir. Büyük ölçüde kullanıcı verilerinin ürünler üzerindeki tarihe göre yapılan derecelendirmelerine dayanır.

Derecelendirme matrisi, kullanıcıların matristeki bütün elemanları derecelendirmemesinden dolayı çoğu zaman seyrek ve büyük olur. Büyük çoğunlukla kullanıcıların ürünler hakkındaki tarihsel derecelendirmelerine bağlıdır. Çoğu zaman kullanıcı matrisi, çok büyük ve kullanıcıların ürünlerin çoğunu değerlendirmemesi nedeniyle de seyrek olur. Bu problem, kullanıcılara güvenilir ve doğru öneriler verme noktasında sistemin yetersiz kalmasına sebep olur.

Düşük maliyetli modellerin farklı türleri pratikte özellikle işbirlikçi filtrelemede uygulamaya yönelik olarak matris tamamlamak için kullanılmıştır. Biçimsel olarak matris tamamlama tekniğinde amaç, matris girdilerini tahmin etmektir [45 – 47].

### **2.1.2. İşbirlikçi filtreleme yöntemlerinin artı ve eksileri**

İşbirlikçi Filtreleme'nin İçerik Bazlı Filtreleme'ye göre önemli avantajları vardır. Çünkü ürünlerle ilgili çok fazla içeriğin olmadığı ve bir bilgisayar sisteminin analiz etmesinin zor olduğu içeriğin (görüş ve öneriler gibi) olduğu durumlar mümkün olabiliyor [5]. Ayrıca İFY, kullanıcının profilinde olmasa bile kullanıcıyla alakalı ürünler gibi çeşitli öneriler sunabilir [48].

İFY'nin başarılı yönlerine rağmen Bölüm 4'te değinilecek sorunlar mümkün olabiliyor.

## **2.2. Hibrit Filtreleme Yöntemleri**

İçerik bazlı ve işbirlikçi filtreleme yöntemlerinin kendilerine göre güçlü ve zayıf oldukları durumlar vardır [49].

En başarılı öneri tekniklerinden biri olan hibrit filtreleme yöntemi yeni kullanıcılara ürün önermek için mevcut kullanıcıların bilinen tercihlerini kullanır. Bu filtreleme yöntemi amazon.com gibi ticarî sistemlerde popüler olarak kullanılmaktadır. En son Netflix film önerileri ödülü HFY araştırmalarına olan ilgiyi artırdı [50 – 54].

Hafıza bazlı ve model bazlı hibrit filtreleme yöntemleri öneri değerlerini yalnızca derecelendirme matrisine dayanarak tahmin etmeye çalışırlar. İBFY, ürünün içeriğinde bulunan bilgileri kullanır. HFY ise hem derecelendirme matrisini hem de ürünün içeriğinde bulunan bilgileri kullanır. İBFY ile İFY'nin dezavantajlarını diğer yaklaşımın avantajlarıyla örtterek daha verimli sonuçlar alınmasını sağlar.

Etkili bir sonuç için henüz değerlendirilmemiş ürünlerin verilerinin azlığı olan veri seyrekliği ve çok büyük sayıda kullanıcı ve ürün bilgilerinin bulunduğu ölçeklenebilirlik arasında bir ortak bir filtre uygulanmalıdır [52].

### **2.2.1. Değişmeli hibrit yöntemi**

Bu yöntem, güven değeri ve dış ölçüt gibi kriterleri mevcut olan ve bileşenleri farklı durumlarda farklı performanslar gösteren durumlar arasından önerilerde bulunur [62]. İşbirlikçi yöntem gibi davranarak içerik bazlı yönetimdeki yeni kullanıcı sorununu önleme özelliğine sahiptir.

Öneri sistemlerinin güçlü ve ve zayıf yönlerine karşı duyarlı olması bu sistemin çok önemli bir özelliğidir. Bu sistemin dezavantajı ise parametre sayısını artırdığı için öneri sistemine karmaşıklık getirmesidir. DailyLearner, içerik tabanlı yöntemin yeterli önerilerde bulunamadığı durumlarda işbirlikçi yöntemin kullanıldığı sistemlere bir örnektir [5].

### **2.2.2. Basamaklı hibrit yöntemi**

Bu yöntemle farklı ürünler arasında bir tercih listesi oluşturulurken tekrarlı bir işlem uygulanır. Bu tekniğin önerileri başka bir öneri tekniğiyle işlenir. EntreeC, işbirlikçi öneri sistemi olarak kullanılan Basamaklı Hibrit Yöntemi'ne bir örnektir. [5, 62].

### **2.2.3. Karışık hibrit yöntemi**

Bu yöntemle her bir üründen sadece bir öneri almak yerine farklı öneri sistemlerinin sonuçlarını birleştirilir. Bireysel performanslar, yerel bir bölgenin genel performansını etkilemez [5]. Farklı öneri sistemlerinin sonuçlarının birleştirilmesi, çoklu sıralama listelerinin bir araya getirilmesine dayanır [62].

#### **2.2.4. Özellik kombinasyonu**

Belirli bir öneri tekniğiyle üretilen veriler başka bir öneri tekniği için kullanılabilir. Örneğin; Papper, işbirlikçi yöntemin içerik bazlı yöntemdeki derecelendirmelerini film önerisi olarak kullanan bir özellik kombinasyonu örneğidir [62].

#### **2.2.5. Özellik büyütme yöntemi**

Bu teknik, önceki öneri sistemleri tarafından üretilen derecelendirmelerden ve diğer bilgilerden faydalanır. Örneğin Libra Sistemi [17], amazon.com'da bulunan veriler hakkında bir Bayes Metin Sınıflandırıcısı kullanarak içerik temelli önerilerde bulunur [5].

#### **2.2.6. Meta düzeyi yöntemi**

Bu öneri yöntemiyle oluşturulan dahili model başka bir yöntem için girdi verisi olarak kullanılabilir ve bu model tek bir derecelendirmeye karşılaştırıldığında bilgi bakımından her zaman zengindir. Meta düzeyi hibritleri, İBFY'nin seyreklik problemini ilk tekniği ikinci tekniğin girdisi olarak öğrendiği tüm modeli kullanarak çözebilir. Bu yöntem için birinci teknik sonucundaki veri, ikinci teknikte kullanılacak verinin bir kısmının değil, tamamının yerini alır [5, 62].

Tablo 2.3.'te işbirlikçi yöntemlerin kullandıkları tekniklerle birlikte birbirlerine göre avantaj ve dezavantajları gösterilmiştir.



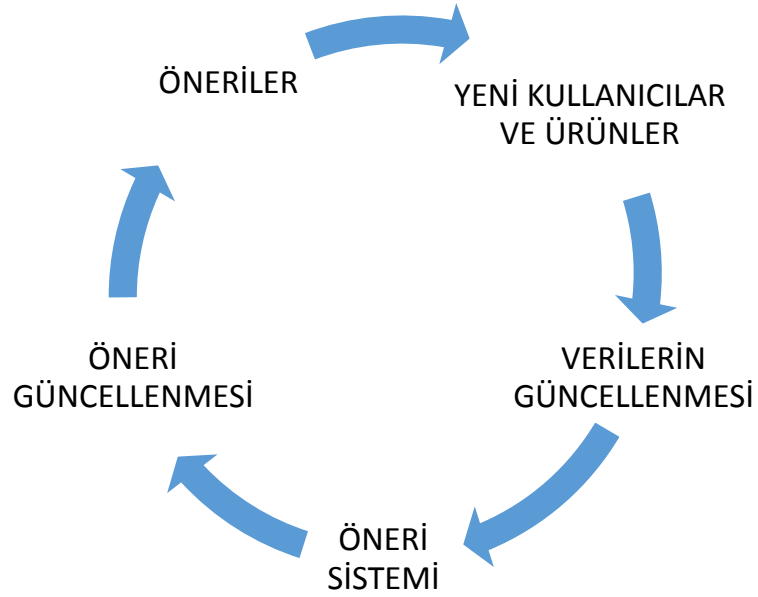
Tablo 2.3: İşbirlikçi Filtreleme Yöntemlerine Genel Bakış

İşbirlikçi Filtreleme Türü	Örnek Teknikler	Temel Avantajlar	Temel Dezavantajlar
Hafıza Bazlı İşbirlikçi Filtreleme	<ul style="list-style-type: none"> <li>- Komşu tabanlı İFY (Pearson / vektör kosinüs korelasyonu ile ürün bazlı / kullanıcı bazlı İFY algoritmaları)</li> <li>- Ürün bazlı / kullanıcı bazlı ilk N adet öneri</li> </ul>	<ul style="list-style-type: none"> <li>- Uygulaması kolay</li> <li>- Yeni veriler kolayca ve aşamalı olarak eklenebilir</li> <li>- Önerilen ürünlerin içeriklerinin dikkate alınmasına gerek yok</li> <li>- Ortak puanlama maddeleriyle iyi bir şekilde ölçeklendirilebilir</li> </ul>	<ul style="list-style-type: none"> <li>- İnsanların derecelendirmelerine bağlı</li> <li>- Verinin seyrek olduğu durumlarda performansın düşmesi</li> <li>- Yeni kullanıcı ve ürünler için tavsiyede bulunamaz</li> <li>- Büyük veri kümeleri için sınırlı ölçeklenebilirlik</li> </ul>
Model Bazlı İşbirlikçi Filtreleme	<ul style="list-style-type: none"> <li>- Bayesian İFY</li> <li>- Kümeleme İFY</li> <li>- Saklı Anlamsal İFY</li> <li>- Seyrek Faktör Analizi</li> <li>- Boyutsallık azaltma tekniklerini kullanan İFY</li> </ul>	<ul style="list-style-type: none"> <li>- Seyreklik, ölçeklenebilirlik ve diğer sorunlar daha kolay ele alınabilir</li> <li>- Tahmin performansını artırır</li> <li>- Öneriler için sezgisel bir açıklama verir</li> </ul>	<ul style="list-style-type: none"> <li>- Model oluşturmak pahalı</li> <li>- Tahmin performans ve ölçeklenebilirlik arasında denge kurulması</li> <li>- Boyutsallık azaltma teknikleri için faydalı bilgilerin kaybedilmesi</li> </ul>
Hibrit Önericiler	<ul style="list-style-type: none"> <li>- İFY, Örneğin Fab</li> <li>- İçeriği artırılmış İF</li> <li>- Hafıza bazlı ve Model bazlı öneri sistemlerini bir araya getiren yaklaşım. Örneğin kişilik Teşhisi</li> </ul>	<ul style="list-style-type: none"> <li>- Önerici sistemlerin sınırlandırmalarının üstesinden gelir.</li> <li>- Tahmin performansını artırır.</li> <li>- Seyreklik ve Gri Koyun (Gray Sheep) gibi problemleri çözer</li> </ul>	<ul style="list-style-type: none"> <li>- Yüksek maliyet ve karmaşık olması</li> <li>- Genellikle uygun olmayan harici bilgilere ihtiyaç duyması</li> </ul>

Şekil 2.2.'de görüldüğü gibi gerek kullanıcılardan gelen değerlendirme ve derecelendirmelerle gerekse yeni gelen ürün veya mevcut ürünlerin içeriklerinde yapılması muhtemel güncellemeler sonucu verilerde güncelleme işlemi olabilir.

Öneri sistemi tarafından mevcut veriler işlenerek bir öneri sunulur veya önceki öneriler güncellenebilir. Bu öneriler kullanıcılara sunulurken kullanıcının farklı ürünlerden haberdar edilmesi ya da daha uygun ürünlere yönlendirilmesi sağlanır.

Bir kullanıcının kendisine önerilen önerilerden sonra yapacağı değerlendirme ve derecelendirmeler de öneri sistemi için yeni bir giriş veya güncelleme olacak ve öneri sisteminin geliştirilmesi için katkı olacaktır.



Şekil 2.2. Öneri Sistemi Çalışma Modeli

## **BÖLÜM 3. ÖNERİ SİSTEMLERİNDE KARŞILAŞILAN PROBLEMLER**

Öneri sistemlerinin güvenilirliği ve uygulanabilirliği, sistemin devamı için çok önemlidir. Kullanıcının ihtiyaçlarına göre öneriler yapılabilmesi, kullanılan öneri sistemlerinin güvenilirliğini ve uygulanabilirliğini artıracaktır. Kullanıcının değerlendirmeleri, yeni ürünlerin sisteme dahil olması, verilerin giderek çoğalması vs. ilgilenilmesi gereken bazı problemleri de beraberinde getirecektir.

### **3.1. Veri Seyrekliği**

Veri seyrekliği problemi, uygun verilerin benzer kullanıcıları tanımlamasında yetersiz olduğu durumlarda ortaya çıkar. Önerilerin kalitesini ve işbirlikçi filtrelemeyi etkileyen önemli bir durumdur [91].

### **3.2. Ölçeklenebilirlik**

Öneri sistemlerinde öneri doğruluğunu artırmak için kullanıcılar arasındaki benzerliği hesaplarken her özelliğe farklı bir ağırlık verilir. Özelliklerin ağırlık değerlerinin belirlenmesi için genetik algoritma daha uygundur. Önerilen metodla öneri doğruluğu ve ölçeklenebilirliği bir miktar iyileştirilir. Böylece veri yoğunluğunu artırarak veri seyrekliği problemine sebep olan ürün derecelendirmeleri ve demografik bilgiler bir araya getirilip, birleştirilir [18].

### **3.3. Soğuk Başlangıç**

Soğuk Başlangıç (Cold Start), bilgisayar tabanlı bilgi sistemlerinde otomatik veri modellemesini içeren potansiyel bir problemdir. Temel sorun, haklarında henüz yeterli

bilgiye sahip olunamayan kullanıcılar ve ürünler için herhangi bir çıkarımda bulunulamaması durumudur [66].

### 3.3.1. Sistemlere etkileri

Soğuk Başlangıç (Cold Start), en yaygın öneri sistemleri problemidir. Normalde bir öneri sistemi, kullanıcı profillerini bazı referans özelliklerine göre karşılaştırır. Bu özellikler, İçerik Tabanlı Filtreleme Yöntemi veya İşbirlikçi Filtreleme Yöntemi tarafından kullanılan ürün oylamaları, yer işaretleri, alımlar, sayfa ziyaret sayıları gibi bilgilerden elde edilebilir [64].

- İçerik Bazlı Filtreleme Yöntemi (İBFY)'nde sistem tarafından bir ürünün özellikleri, kullanıcının profilindeki ilgili ürünün özellikleriyle eşleştirebilmelidir. Bunun yapılabilmesi için öncelikle kullanıcının tercihleri hakkında ayrıntılı bir bilgi edinilmelidir. Bilgi edinme, kullanıcının sorgulanmasıyla veya kullanıcının davranışlarının takip edilmesiyle mümkündür. Soğuk Başlangıç Problemi, hem kullanıcının sorgulanma sürecinde hem de kullanıcının davranışlarının takip sürecinde kullanıcıya henüz akıllı bir öneri sunmaya başlamadan önceki durumda olan sistemi kullanmasında biraz iş düşeceği anlamına gelmektedir.
- İşbirlikçi Filtreleme Yöntemi (İFY)'nde ise öneri sistemi, aktif kullanıcılarla aynı tercihleri paylaşan kullanıcıları tespit eder ve benzer fikirli kullanıcıların tercih ettiği ama aktif kullanıcıların henüz görmediği ürünleri tavsiye eder. Soğuk Başlangıç Problemi'nden dolayı İFY, daha önceden kullanıcılar tarafından oylanmamış ürünleri dikkate almaz.
- Kullanıcının tercihleri, kullanıcı davranışları gözlemlenerek oluşturulduğu için tavsiyelerin kullanıcıya göre kişiselleştirilmesi zaman almaktadır.

Üç durumda soğuk başlangıç olur [66].

1. Yeni Ortak Nokta: Ürünle ilgili bilgiler olmasına rağmen henüz kullanıcının ya da kullanıcı tercihlerinin olmadığı durumlarda güvenilir öneriler pek mümkün olmayabilir.
2. Yeni Ürün: İçerik bilgisine sahip olmasına rağmen etkileşim olmadığı için sisteme yeni eklenen ürünler için öneriler sağlıklı ve güvenilir olmayabilir.
3. Yeni Kullanıcı: Henüz etkileşime girmediği ve herhangi bir öneri ya da işlem yapmadığı için yeni bir kullanıcının kişiselleştirilmiş herhangi bir öneride bulunması mümkün değil.

### 3.3.2. Sorunu azaltma yöntemleri

Sistem tipi ve karakteristikleriyle birlikte çok sayıda öneri algoritması olması nedeniyle soğuk başlangıç problemini hafifletmek için birçok strateji geliştirilmiştir. En önemli strateji, hibrit önerilere güvenmektir. Bu da bir kategori veya modelin diğeriyle birleştirilmek suretiyle dezavantajlarını azaltarak gerçekleştirilir [67 – 69].

Soğuk başlangıcın her üç durumunun da (yeni ürün ve yeni kullanıcı) ortak kullanıcı etkileşimi (yani aynı ürün hakkında farklı kullanıcıların yorum ve değerlendirmelerinin) eksikliği vardır. Yeni ürünler için İFY (İşbirlikçi Filtreleme Yöntemi), mevcut ürünler için ise İBFY (İçerik Bazlı Filtreleme Yöntemi) kullanmak, öneri sistemlerinin performansını artırmak için ortak bir strateji olarak değerlendirilebilir.

### 3.3.3. Profil tamamlama

Yeni ürün veya kullanıcılar söz konusu olduğunda uygun seçeneklerden bir tanesi de referans olabilecek bazı verileri hızlıca elde etmektir. Gerekli bilginin büyüklüğüne göre bu verilerin elde edilmesinin çeşitli yolları vardır. Bu tekniklere ortaya çıkarma stratejileri denir. Bu stratejiler, kullanıcıya yöneltilecek sorularla açıkça yapılır veya kullanıcı davranışları gözlemlenerek habersiz yapılır.

Her iki durumda da soğuk başlangıç problemi, kullanıcının sistemi herhangi bir öneri sunmaya başlamadan önceki durumu olan “aptal” durumunda (kullanıcıların kendi profillerini oluşturmalarına katkıda bulunma durumunda) kullanabilmesi için bir miktar çaba sarfetmesi anlamına gelir.

Örneğin; filmler için bir öneri sistemi olan MovieLens, yeni kullanıcı kayıtlarında kullanıcılardan bazı filmleri değerlendirmelerini ister. Tercihleri ortaya çıkarma stratejisi yeni kullanıcılarla başa çıkmak için etkili bir yöntem olsa da yeni kayıt sırasındaki ek bilgiler, kullanıcılardan daha fazla zaman alacaktır. Bununla birlikte, kullanıcıların uzun zaman önce gördükleri, kullandıkları ürünleri değerlendirmeleri, tercihlerin kalitesinin ideal seviyede olmasını engelleyebilir ya da kayıt işleminin uzaması söz konusu olduğunda kullanıcının kayıt işlemini dikkatsizce yapmasına bu da derecelendirmelerin rastgele olmasına sebebiyet verebilir.

Kullanıcı profilleri, kullanıcılardan gelen sosyal medya veya internette ziyaret ettikleri sayfa bilgileriyle de otomatik olarak oluşturulabilir. Örneğin bir kullanıcı bir bilgisayar oyunu hakkında bir haber okuyorsa, kullanıcı oyun mağazasına ziyaret ettiğinde öneri sistemi, önerilen oyunlar içerisinde bu oyunu da önerir [70 – 73].

Bir kullanıcının kişisel karakteristiğine dayalı olarak başlangıç profilini oluşturmak ve bu profili kişiselleştirilmiş öneri oluşturmak için kullanmak da mümkündür. Kullanıcının karakteristik özellikleri, Beş Faktör Modeli (Five Factor Model – FFM) kullanılarak tanımlanabilir.

Bir diğer teknik ise makine öğrenmesidir. Makine öğrenmesinin en önemli amacı, en bilgilendirici olabilecek ürünleri değerlendirip, puanlama yapmasını istemek için tercih çıkarma sürecine yönlendirmektir. Bu yönlendime işlemi, uygun verilerin analiz edilmesi ve verilerin (puanlama ve etkileşim verileri) yararlarının ne kadar olduğunun tahmin edilmesiyle yapılır [74, 75].

### 3.3.4. Özellik haritalaması

Son dönemlerde, makine öğrenmesine dayanan, içerik ile ortak bilgileri tek bir modelde birleştirmeye çalışan daha gelişmiş stratejiler ortaya çıkarıldı. Bu yaklaşımların bir örneği de matris faktoringi algoritmalarına uyarlanmış özellik haritalama tekniğidir [76, 77]. Bu tekniğe göre:

Bir faktoring matrisi, kullanıcı – ürün etkileşimlerini iki dikdörtgen matris olarak temsil eder. Her kullanıcı ilk matrisin bir satırıyla, her ürün de ikinci matrisin bir sütunuyla ilişkilendirilir. Bir kullanıcı veya ürünle ilişkilendirilmiş satır veya sütuna gizli faktörler denir. Yeni bir ürün eklendiğinde gizli faktörlerle henüz ilişkilendirilmediği için ortaya çıkan etkileşim eksiklikleri gizli faktörlerin öğrenilmesine izin vermez [78].

Her bir ürün bazı özellikleriyle (marka, yazar, yayınevi, yapımcı, aktör vs.) ilişkilendirilirse bu özelliklere karşılık gelen gizli faktörlerin tahmin edilebilmesini sağlayacak yerleşik bir işlev tanımlanabilir. Mevcut ürün verileriyle eğitilmiş olan bu yerleşik işlev, birçok şekilde tasarlanabilir.

### 3.3.5. Hibrit özellik ağırlıklandırma

Özellik haritalamasına benzer bir diğer yaklaşım ise, ürün ve kullanıcıların özelliklerinin kullanıcı tarafından önem durumuna göre ağırlıklandırıldığı “Hibrit Özellik Ağırlıklandırma”dır. Kullanıcılara önerilecek bir ürünün her bir özelliğinin (marka, ürün grubu, kullanılan malzeme vs.) önemi farklı olacaktır. Dolayısıyla ağırlıkları da farklı olacaktır

Hibrit özellik ağırlıklandırma teknikleri öneri sistemleri için uyarlanmıştır. Ayrıca hibrit yöntemlerinin çoğu faktoring makinelerinin özel durumları olarak düşünülebilirler [79, 80].

### 3.3.6. Çözümler

Soğuk Başlangıç problemi için birkaç çözüm vardır. Etkin çözümlerden bir tanesi, mevcut veri kümesi içerisinde daha fazla veri seçerek ve elde ederek öneri sistemlerinin performansını artırabilen makine öğrenmesi tekniklerini uygulamaktır. Bu çözümde, mevcut verilerin analiz edilmesi ile veri derecelendirmelerinin ne kadar yararlı olduğu tahmin edilir. İşbirlikçi filtreleme yöntemlerinde bu tekniklere derecelendirme çıkarma stratejileri denir.

Arayüz ajanlarının olduğu durumlarda farklı kullanıcılar için çalışan ajanlar arasında işbirliği yapılarak soğuk başlangıç problemi çözüme kavuşturulabilir. Bu sayede ajanların kendi kullanıcılarından gelecek değerlendirmeler kullanılarak yeni ürünler için derecelendirmeler yapılabilir. Öneri sistemlerinde İçerik Bazlı ve İşbirlikçi Filtreleme yöntemleri arasında hibrit bir yaklaşım benimsenerek soğuk başlangıç problemi azaltılır.

Henüz değerlendirme ve derecelendirme yapılmamış yeni ürünlerin puanları kullanıcıların diğer ürünlere yaptıkları derecelendirmelere göre otomatik olarak atanır. Ürün benzerlikleri içerik özelliklerine göre belirlenir. Kullanıcı profilleri, gözetme geçmişi gibi kullanıcı bilgilerinden elde edilen verilerle otomatikleştirilebilir [81 – 85].

Örneğin, bir müzik portalından bir sanatçı hakkında bilgiler okuyan bir kullanıcıya, müzik mağazasını ziyaret ettiğinde de ilgili sanatçı hakkında önerilerde bulunulur [86]. Ayrıca bir kullanıcının kişilik özelliklerine dayanarak, bu kişiye kişiselleştirilmiş önerilerde kendi özelliklerine göre ürünler tavsiye edilebilir. Kullanıcının kişilik özellikleri Beş Faktör Modeli (Five Factor Model – FFM) gibi bir kişilik modeli kullanılarak tanımlanabilir [87, 88].

Beş Faktör Modeli: Genellikle “Büyük Beş” (Big Five) olarak adlandırılana beş geniş özellik kümesidir.



- Dışadönürlük: Dışa dönük bireyler, sessiz ve çekingen olmanın aksine iddialı ve girişkendirler.
- Anlaşılabilirlik: Anlaşılabilir bireyler, sert ve kaba olmaktan çok yardıma hazır ve kibardırlar.
- Vicdanlılık: Vicdanlı bireyler, dikkat dağıtıcı ve düzensiz değil, görev odaklı ve düzenlidirler.
- Nörotizm (Duygusal İstikrar): Nörotik bireyler, duygusal olarak esnek ve sabırlı olmaktan çok endişeli ve sürekli olumsuz düşüncelere eğilimlidirler.
- Deneyime Açıklık: Bu gruptaki bireyler, geniş bir ilgi alanına sahip olup, çevrelerine duyarlı, rutinden daha çok yeniliklere açıktırlar.

Beş Faktör Modeli (BFM), küçük bir özellik kümesi kullanarak, bireyin kişiliğindeki değişkenliğin mümkün olduğu kadar fazlasını tespit etmek için geliştirilmiştir. Pek çok psikolog, bu beş alanın kişilik özelliklerinde en önemli temel farklılıkları yakaladığını düşünmektedirler [61].

### 3.4. Gri Koyun Kullanıcılar

Bir öneri sistemi için en önemli görevlerden bir tanesi, önerilerin kalitesini iyileştirmektir. Bir kullanıcı güvendiği bir öneri sisteminde aradığını bulamadığını farkederse bu sistemi tekrar kullanması beklenmez. Öneri sistemlerinde sıradışı tercihlere sahip kullanıcılar olabilir.

İşbirlikçi Filtreleme Yöntemleri bu kullanıcılar için yüksek kalitede öneri üretemeyecektir. Bu kullanıcılar, diğer kullanıcılardan ayırt edilebilmeleri için gri koyun (gray sheep) olarak adlandırılmışlardır. Diğer kullanıcılarla aynı fikirde olmadıkları ya da kısmen katıldıkları için korelasyon katsayıları düşüktür. Bu kullanıcıların değerlendirmeleri iki soruna sebep olabilir.

1. Kullanıcılar doğru öneriler alamayabilirler.
2. Kullanıcıların önerilerini olumsuz yönde etkileyebilirler [61, 65].

### 3.5. Şilin Saldırıları

Şilin Saldırıları, sahte kullanıcı hesaplarıyla beraber sahte derecelendirme ve değerlendirmeler gibi içerikler oluşturularak ürünlerin öneri sıralamalarını ve değerlendirmelerini değiştirmek için yapılan saldırılardır. Bu saldırılarla, ürünlerin daha üst ya da daha alt sıralarda yer almaları sağlanabilir. Bu saldırıların tespit edilmesi, kullanılan öneri sisteminin güvenilirliğini devam ettirebilmek için çok önemlidir.

### 3.6. Veri Eksikliği

Öneri sistemlerindeki en büyük sorun belki de etkili öneriler için çok sayıda veriye ihtiyaç duyulmasıdır. Google, Amazon, Netflix gibi öneri sistemlerinde en başarılı firmaların sahip oldukları tüketici verilerinin büyüklüğü, başarılarının ana kaynağıdır. Çünkü bir öneri sistemi ne kadar çok ürün ve kullanıcı bilgisi barındırıyorsa o kadar iyi tavsiyelerde bulunur.

### 3.7. Tahmin Edilemeyen Ürünler

Kullanım amacı veya ihtiyacı bakımından birbirlerine yakın ürünlerin tahmin edilip, önerilmesi zaten öneri sistemlerinden beklenen sonuçlardır. Buna karşın birbirlerine kullanım amacı, kullanım ihtiyacı vb. durumlardan dolayı uzak, hatta zıt denebilecek ürünlerin tahmin edilmeleri beklenen sonuçlar değildir.

Örneğin bir kişi hem arabesk müzik hem sanat müziği ya da hem klasik müzik hem de halk müziği sevebilir. Hem korku filmleri hem romantik filmler hoşuna gidebilir ya da aynı anda hem dağ sporlarına hem de su sporlarına ilgili duyuyor olabilir. Böyle durumlarda ilk ilgi alanındaki ürün grubundan yapacağı alışverişler, diğer ilgi alanındaki ürün gruplarından öneriler için pek yardımcı olamayacaktır.

## **BÖLÜM 4. ÖNERİ SİSTEMLERİ UYGULAMALARI VE KARŞILAŞTIRILMALARI**

İçerik bazlı, işbirlikçi ve hibrit öneri sistemlerinin ayrı ayrı uygulamaları ve karşılaştırmaları için Python dili ve bu dilin bazı kütüphane ve sınıfları kullanılmıştır. Bu dilin seçilmesinin sebebi kullanımın kolay ve anlaşılır olması, etkili çözümler sunması ile birlikte zengin ve yine kullanımı çok kolay kütüphanelere sahip olmasıdır.

### **4.1. Kullanılan Kütüphane ve Sınıflar**

NumPy, Pandas, Sci-kit Learn kütüphaneleri ve Sci-kit Learn kütüphanesinden:

- cosine\_similarity,
- SVD,
- TfidfVectorizer,

sınıfları kullanılmıştır.

NumPy: NumPy kütüphanesi, yapısal olarak Python listelerine benzeyen ama işlev ve hız bakımından daha kullanışlı olan, işlemlerin hızlı bir şekilde yapılmasını sağlayan, bilimsel hesaplamalar için kullanılan bir matematik kütüphanesidir. NumPy kütüphanesinde diziler homojen yapıda olmalıdır yani dizilerin elemanları aynı veri tipinde olmalıdır.

Pandas: Pandas kütüphanesi, homojen olmayan dizilerle yapılacak işlemlerde etkili olarak kullanılan bir kütüphanedir. Csv ve txt dosyalarındaki verilerle işlemlerin hızlı ve etkin bir şekilde yapılmasını sağlar.

Sci-kit Learn: Sci-kit Learn kütüphanesi, makine öğrenmesi için hazırlanmış sınıflandırma, regresyon, karar ağaçları, vektör ve matris vb. işlemler için kullanılan çok faydalı ve kullanması kolay bir kütüphanedir.

Cosine\_similarity: Kosinüs benzerliği ile iki metin arasındaki benzerlik, trigonometrik fonksiyon olan kosinüs fonksiyonu aracılığıyla formülize edilerek hesaplanır.

SVD: Singular Value Decomposition (Tekil Değer Ayrıştırma), karmaşık bir matrisle yapılması gereken işlemlerin daha basit matrislerin çarpımı şeklinde işlem yapılmasını sağlayan bir sınıftır. Özellikle matrisin köşegenleştirilmesi ile sıkı bir ilişkisi olan SVD, tüm matris ayrışım algoritmalarının başında gelerek, özellikle regresyon analizinde önemli bir yer tutmaktadır [32].

TfidfVectorizer: Bu sınıf sayesinde işleme alınacak kelimeler TF / IDF matrisine dönüştürülür. Bu teknikle yapılandırılmamış bir metin, her kelimenin vektör içinde bir konumla temsil edildiği bir vektör yapısına dönüştürülür. Bu sayede ürün özelliklerine göre ürünlerin birbirleriyle ne kadar alakalı olduğu ölçülür.

#### **4.2. İçerik Bazlı Öneri Sistemi Uygulaması**

Bu uygulama için kullanılan veri setinde 14.400 filme ait film adı, türü, imdb\_id, filmin özeti gibi bilgilerin bulunduğu 24 bilgi bulunmaktadır. Filmlerin özetlerinin birbirlerine benzerliklerinden faydalanılarak bir tavsiye listesi oluşturulacaktır. TfidfVectorizer sınıfı yardımıyla yapılandırılmamış bir metin, her kelimenin matris içinde bir konumla temsil edildiği bir matris yapısına dönüştürülür.

Kosinüs benzerliği kullanılarak bu matristen elde edilen verilerle her bir filme içerik bakımından en yakın olan filmler sıralanır. Filmin adının girilmesiyle bu film adına en yakın içerikteki filmlerin sıralanması sağlanır.

### 4.3. İşbirlikçi Öneri Sistemi Uygulaması

İşbirlikçi öneri sisteminde 2.500 adet film bilgisi ve 264.000 adet kullanıcı değerlendirmesinin olduğu veri setleri kullanılmıştır.

Kosinüs benzerliği sınıfıyla her bir kullanıcı için en yakın 20 kişi belirlenmiştir. Daha sonra kullanıcı ve film parametlerini kullanan bir fonksiyon tanımlanmıştır. Bu fonksiyon çıktısından elde edilen değerle kullanıcıya tavsiye edilecek filmler belirlenir.

### 4.4. Hibrit Öneri Sistemi Uygulaması

Hibrit öneri sistemi için 9.219 adet film bilgilerinin olduğu veri setleri kullanılmıştır. Aynı zamanda bu filmlerin birbirleriyle benzerliklerinin olduğu bir kosinüs benzerlik veri seti de sistemin eğitilmesi amacıyla kullanılmıştır. Tanımlanan dataframe, split metoduyla 5'e bölünüp, bölünen kısım, SVD sınıfı sayesinde eğitim seti olarak kullanılmıştır.

Oluşturulan fonksiyonda kullanıcı ID'si ve filmin adı parametre olarak alınır. ID'si girilen kullanıcının incelediği ya da izlediği filmin adına göre kendisine istenen sayıda film önerisi yapılır.

### 4.5. Öneri Sistemlerinin Karşılaştırılması

Kullanılan veri setinde kullanılan makalelerin tarih, gerçek web adres, başlık, metin içeriği, hangi dilde yazıldığı ve kim tarafından paylaşıldığı (yazarı) bilgileri mevcuttur.

Bu örnekte makalelerin görüntülenme, beğenilme, yorum yapılma, takip edilme, sık kullanılanlara eklenmesi gibi bilgileri farklı puanlarla ağırlıklandırılıp, değerlendirilmiştir.

Yeterli verinin olmadığı Soğuk Başlangıç probleminin önüne geçilebilmesi için en az 5 etkileşimi olan kullanıcılar değerlendirmeye alınmıştır.

Veri setinde toplam 2.215 kullanıcı olmasına karşın değerlendirmeye alınan en az 5 etkileşimi olan kullanıcıların sayısı 1.260'tır. 93.814 etkileşime karşılık, en az 5 etkileşimi olan kullanıcı tarafından yapılan etkileşim sayısı 74.212'dir.

Kullanıcı / makale tekil etkileşim sayısı 42.2. Bu etkileşimler içinden eğitim verisi olarak veri setinin %80'i olarak 33.792 adet, test verisi olarak da %20'si olarak 8.448 adet alınmıştır.

Bu çalışmada Terim Sıklığı / Ters Doküman Sıklığı (TF / IDF) tekniği kullanılmıştır. Bu teknikle yapılandırılmamış bir metin, her kelimenin vektör içinde bir konumla temsil edildiği bir vektör yapısına dönüştürülür. Bu sayede her kelimenin makale ile ne kadar alakalı olduğu ölçülür. Kullanıcının etkileşimde bulunduğu her bir ürün için:

- Kullanıcının hiç etkileşime girmediği 100 ürün örneklenir.
- Kullanıcının bu ürünlerle ilgisi olmadığı için bu ürünlerden haberdar değildir.
- Öneri sisteminden 1 etkileşimli, 100 etkileşimsiz üründen oluşan bir liste istenir.
- Bu listeden kullanıcı ve etkileşimde olduğu ürün için "İlk N Tane" doğruluk ölçümü hesaplanır.
- Bütün "İlk N Tane" doğruluk ölçümleri bir araya getirilir.

İçerik Bazlı Öneri için:

Tablo 4.1.'e göre birinci kullanıcının toplam etkileşim sayısı 202, rastgele seçilen 20 etkileşimi yani yaklaşık %10'u ilk 5'e, 30 etkileşimi yani yaklaşık %15'i ise ilk 10'a girmiştir.

İkinci kullanıcının toplam etkileşim sayısı 162, rastgele seçilen 20 etkileşimi yani yaklaşık %12'si ilk 5'e, 38 etkileşimi yani yaklaşık %23'ü ise ilk 10'a girmiştir.

Tablo 4.4. İçerik Bazlı Öneri İçin En Çok Etkileşime Giren Kullanıcıların Listesi

kisi_id	İlk 10'daki sayı	İlk 5'teki sayı	Etkileşim sayısı	İlk 10'daki oran	İlk 5'teki oran
user022935	30	20	202	0,148515	0,108911
user346735	38	20	162	0,234568	0,123457
user936696	36	26	140	0,257143	0,185714
user664872	60	40	124	0,483871	0,322581
user812977	18	9	91	0,197802	0,098901

İşbirlikçi Öneri için:

Tablo 4.2.'de birinci kullanıcının toplam etkileşim sayısı 202, rastgele seçilen 24 etkileşimi yani yaklaşık %12'si ilk 5'e, 48 etkileşimi yani yaklaşık %24'ü ise ilk 10'a girmiştir.

İkinci kullanıcının toplam etkileşim sayısı 162, rastgele seçilen 28 etkileşimi yani yaklaşık %16'sı ilk 5'e, 58 etkileşimi yani yaklaşık %46'sı ise ilk 10'a girmiştir.

Tablo 4.5. İşbirlikçi Öneri İçin En Çok Etkileşime Giren Kullanıcıların Listesi

kisi_id	İlk 10'daki sayı	İlk 5'teki sayı	Etkileşim sayısı	İlk 10'daki oran	İlk 5'teki oran
user022935	48	24	202	0,237624	0,118812
user346735	58	28	162	0,358024	0,160494
user936696	40	19	140	0,285714	0,135714
user664872	50	40	124	0,403226	0,322581
user812977	52	37	91	0,571429	0,406593

Hibrit Öneri için:

Tablo 4.3.'te birinci kullanıcının toplam etkileşim sayısı 202, rastgele seçilen 30 etkileşimi yani yaklaşık %15'i ilk 5'e, 42 etkileşimi yani yaklaşık %21'i ise ilk 10'a girmiştir.

İkinci kullanıcının toplam etkileşim sayısı 162, rastgele seçilen 40 etkileşimi yani yaklaşık %25'i ilk 5'e, 54 etkileşimi yani yaklaşık %33'ü ise ilk 10'a girmiştir.

Tablo 4.6. Hibrit Öneri İçin En Çok Etkileşime Giren Kullanıcıların Listesi

kisi_id	İlk 10'daki sayı	İlk 5'teki sayı	Etkileşim sayısı	İlk 10'daki oran	İlk 5'teki oran
user022935	42	30	202	0,207921	0,148515
user346735	54	40	162	0,333333	0,246914
user936696	38	30	140	0,271429	0,214286
user664872	50	38	124	0,403226	0,306452
user812977	33	30	91	0,362637	0,340659



## BÖLÜM 5. SONUÇ VE ÖNERİLER

Tablo 5.1.'de görüldüğü gibi genel olarak rastgele seçilen etkileşimlerin ilk 10'a girme oranının işbirlikçi yöntemde yaklaşık %45, içerik bazlı yöntemde yaklaşık %52, hibrit yöntemde ise yaklaşık %53, ilk 5'e girme oranının ise işbirlikçi yöntemde yaklaşık %33, içerik bazlı yöntemde yaklaşık %41, hibrit yöntemde ise yaklaşık %43 olduğu ve hibrit yöntemin daha başarılı sonuçlar çıkardığı görülmüştür.

Tablo 5.1.Yöntemlerin Karşılaştırılması

Kullanılan Yöntem	İlk 10'daki oran	İlk 5'teki oran
İşbirlikçi Filtreleme	0,453133	0,332458
İçerik Bazlı Filtreleme	0,515578	0,409334
Hibrit Filtreleme	0,527425	0,426157

Öneri sistemlerinin tahminlerinin doğruluğu, kullanılacak öneri sistemi türünün hangi internet sitesinde kullanılacağı ile doğrudan ilgilidir.

Örneğin, haber siteleri için; sadece içeriğin daha önemli olduğu, kullanıcı yorum ve derecelendirmelerinin çok etkisinin olmadığı İçerik Bazlı Filtreleme Yöntemi, kullanıcı derecelendirmeleriyle beraber ürün içeriğinin de önemli olduğu siteler için İşbirlikçi Filtreleme Yöntemi daha uygun olacaktır.

Öneri sisteminin kullanılacağı sektör, internet sitesi türü vb. yerler için diğer yöntemlerin birbirlerine göre daha iyi sonuçlar verebilmesi de söz konusudur. Hazırlanan uygulama ile bu durumlar da göz önünde bulundurularak elde edilecek sonuçlara göre, hangi yöntemin daha iyi sonuçlar verdiğine bakılıp daha uygun olan öneri sistemi kullanılmalıdır.

## KAYNAKLAR

- [1] <http://www.certona.com/article/benefit-of-recommendation-engines/>, Eriřim Tarihi: 06.04.2019.
- [2] <http://www.geo-viz.com/blog/advantages-of-a-recommendation-system>, Eriřim Tarihi: 06.04.2019.
- [3] <https://www.quora.com/What-are-the-benefits-of-using-product-recommendations-on-my-e-commerce-site>, Eriřim Tarihi: 06.04.2019.
- [4] Xiaoyuan Su, Taghi M. Khoshgoftaar, A Survey of Collaborative Filtering Techniques, October 2009 *Advances in Artificial Intelligence* 2009(12) DOI: 10.1155/2009/421425.
- [5] Su, X., Khoshgoftaar, T. M., A Survey of Collaborative Filtering Techniques, *Advances in Artificial Intelligence*, Article ID 421425, 2009(12) DOI: 10.1155/2009/421425.
- [6] Aggarwal, Charu C. (2016). *Recommender Systems: The Textbook*. Springer. ISBN 9783319296579.
- [7] Peter Brusilovsky (2007). *The Adaptive Web*. p. 325. ISBN 978-3-540-72078-2.
- [8] Blanda, Stephanie (May 25, 2015). "Online Recommender Systems – How Does a Website Know What I Want?". *American Mathematical Society*. Retrieved October 31, 2016.
- [9] <http://bilgisayarkavramlari.sadievrenseker.com/2012/10/22/tf-idf/>, Eriřim Tarihi: 01.02.2019.
- [10] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach Learn* 1997;29 (2 –3): 131 – 63.
- [11] Duda RO, Hart PE, Stork DG. *Pattern classification*. John Wiley & Sons; 2012.
- [12] Bishop CM. *Pattern recognition and machine learning*, vol. 4, no.4. Springer, New York; 2006.

- [13] Shyong K, Frankowski D, Riedl J. Do you trust your recommendations? An exploration of security and privacy issues in recommender systems. In: *Emerging trends in information and communication security*. Berlin, Heidelberg: Springer; 2006. p. 14–29.
- [14] Adomavicius G, Tuzhilin A. Toward the next generation of recommender system. A survey of the state-of-the-art and possible extensions. *IEEE Trans Knowl Data Eng* 2005;17(6):734–49.
- [15] Zhang T, Vijay SI. Recommender systems using linear classifiers. *J Mach Learn Res* 2002; 2:313–34.
- [16] Billsus D, Pazzani MJ. User modeling for adaptive news access. *User Model User-adapted Interact* 2000;10(2–3):147–80.
- [17] Mooney RJ, Roy L. Content-based book recommending using learning for text categorization. In: *Proceedings of the fifth ACM conference on digital libraries*. ACM; 2000. p. 195–204.
- [18] Qian Wang, Xianhu Yuan, Min Sun “Collaborative Filtering Recommendation Algorithm Based On Hybrid User Model”, *FSKD*, 2010.
- [19] Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. (2000). "Application of Dimensionality Reduction in Recommender System A Case Study",.
- [20] Allen, R.B. (1990). "User Models: Theory, Method, Practice". *International J. Man-Machine Studies*.
- [21] Parsons, J.; Ralph, P.; Gallagher, K. (July 2004). "Using viewing time to infer user preference in recommender systems". *AAAI Workshop in Semantic Web Personalization*, San Jose, California.
- [22] Breese, John S.; Heckerman, David; Kadie, Carl (1998). *Empirical Analysis of Predictive Algorithms for Collaborative Filtering (PDF) (Report)*. Microsoft Research.
- [23] "Kernel-Mapping Recommender system algorithms". *Information Sciences*. 208: 81–104. doi:10.1016/j.ins.2012.04.012. Retrieved 1 June 2015.
- [24] Bobadilla J, Ortega F, Hernando A, Gutie´rrez A. Recommendersystems survey. *Knowl-Based Syst* 2013; 46:109–32.
- [25] Breese J, Heckerma D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: *Proceedings of the 14th conference on uncertainty in artificial intelligence (UAI-98)*; 1998. p. 43–52.

- [26] Zhao ZD, Shang MS. User-based collaborative filtering recommendation algorithms on Hadoop. In: Proceedings of 3rd international conference on knowledge discovering and data mining, (WKDD 2010), IEEE Computer Society, Washington DC, USA; 2010. p. 478–81. doi: 10.1109/WKDD.2010.54.
- [27] Zhu X, Ye HW, Gong S. A personalized recommendation system combining case-based reasoning and user-based collaborative filtering. In: Control and decision conference (CCDC 2009), Chinese; 2009. p. 4026–28.
- [28] Adomavicius G, Tuzhilin A. Toward the next generation of recommender system. A survey of the state-of-the-art and possible extensions. *IEEE Trans Knowl Data Eng* 2005;17(6):734–49.
- [29] Herlocker JL, Konstan JA, Terveen LG, Riedl JT. Evaluating collaborative filtering recommender systems. *ACM Trans Inform Syst* 2004;22(1):5–53.
- [30] Jannach D, Zanker M, Felfernig A, Friedrich G. Recommender systems – an introduction. Cambridge University Press; 2010.
- [31] Ansley and Graig F., Hern ve Thomas, 1993; Stewart G. W., 1993; Watkins ve David. S., 1982.
- [32] Li, Z. 2006. Matris Ayrışımı. İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü. Ekonometri Anabilim Dalı, Yüksek Lisans Tezi.
- [33] <http://www.statisticshowto.com/latent-semantic-analysis/>, Erişim Tarihi: 03.03.2019.
- [34] Kuzelewska U. Advantages of information granulation in clustering algorithms. In: Agents and artificial intelligence. NY: Springer; 2013. p. 131–45.
- [35] McSherry D. Explaining the pros and cons of conclusions in CBR. In: Calero PAG, Funk P, editors. Proceedings of the European conference on case-based reasoning (ECCBR-04). Madrid (Spain): Springer; 2004. p. 317–30.
- [36] Linden G, Smith B, York J. Amazon.com recommendation: item-to-item collaborative filtering. *IEEE Internet Comput* 2003;7(1): 76–80.
- [37] Michael JA, Berry A, Gordon S, Linoff L. Data mining techniques. 2nd ed. Wiley Publishing Inc.; 2004.
- [38] Hosseini-Pozveh M, Nemartbakhsh M, Movahhedinia N. A multimedia approach for context-aware recommendation in mobile commerce. *Int J Comput Sci Inform Secur* 2009;3(1).

- [39] Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. In: Cohen W. Moore AW, editors. Machine Learning, Proceedings of the twenty-third international conference, ACM, New York; 2003. p. 161–8.
- [40] Larose TD. Discovering knowledge in data. Hoboken (New Jersey): John Wiley; 2005.
- [41] Berry MJA, Linoff G. Data mining techniques: for marketing, sales, and customer support. New York: Wiley Computer Publishing; 1997.
- [42] Deng C, Xiaofe H, Ji-Rong W, Wei-Ying M. Block-level link analysis. In: Proceedings of the 27th annual international ACM SIGIR conference on research and development in information retrieval; 2004. p. 440–7.
- [43] <http://w3.balikesir.edu.tr/~bsentuna/wp-content/uploads/2013/03/Regresyon-Analizi.pdf>, Erişim Tarihi: 01.04.2019.
- [44] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Mach Learn 1997;29(2–3):131–63.
- [45] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. IEEE Comput 2009; 8:30–7.
- [46] Bojnordi E, Moradi P. A novel collaborative filtering model based on combination of correlation method with matrix completion technique. In: 16th CSI international symposium on artificial intelligence and signal processing (AISP), IEEE; 2012.
- [47] Taka'cs G, Istva'n P, Bottya'n N, Tikk D. Investigation of various matrix factorization methods for large recommender systems. In: IEEE international conference on data mining workshops. ICDMW'08. IEEE; 2008. p. 553–62.
- [48] Schafer JB, Frankowski D, Herlocker J, Sen S. Collaborative filtering recommender systems. In: Brusilovsky P, Kobsa A, Nejdl W, editors. The Adaptive Web, LNCS 4321. Berlin Heidelberg (Germany): Springer; 2007. p. 291–324. [http://dx.doi.org/10.1007/978-3-540-72079-9\\_9](http://dx.doi.org/10.1007/978-3-540-72079-9_9).
- [49] Su, X., Greiner, R., Khoshgoftaar, M. T., Zhu, Xingquan., Hybrid Collaborative Filtering Algorithms Using a Mixture of Experts, Web Intelligence, IEEE/WIC/ACM International Conference on, 2007. DOI: 10.1109/WI.2007.10.
- [50] Netflix prize, <http://www.netflixprize.com>, Erişim Tarihi: 20.04.2019.
- [51] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A Constant Time Collaborative Filtering Algorithm, Information Retrieval, 4(2), pp. 133 – 151, 2001.

- [52] B.M. Sarwar, G. Karypis, J.A. Konstan, and J. Riedl. Item-based Collaborative Filtering Recommendation Algorithms, 10th International World Wide Web Conference, pp. 285 – 295, 2001.
- [53] K. Miyahara, and M.J. Pazzani, Improvement of Collaborative Filtering with the Simple Bayesian Classifier, Information Processing Society of Japan, 43(11), 2002.
- [54] P. Melville, R.J. Mooney, and R. Nagarajan. ContentBoosted Collaborative Filtering for Improved Recommendations, Proceedings of the 18th National Conference of Artificial Intelligence, 2002.
- [55] R. Greiner, X. Su, B. Shen, and W. Zhou. Structural Extension to Logistic Regression: Discriminative Parameter Learning of Belief Net Classifiers, Machine Learning, 59(3), pp. 297 – 322, 2005.
- [56] X. Su, and T.M. Khoshgoftaar. Collaborative Filtering for Multi-class Data Using Belief Net Algorithms, the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp. 497-504, 2006.
- [57] GroupLens. <http://movielens.umn.edu>, GroupLens Research group, Department of Computer Science and Engineering, University of Minnesota, 2006.
- [58] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A Constant Time Collaborative Filtering Algorithm, Information Retrieval, 4(2), pp. 133-151, 2001.
- [59] Suriati, Meisyarah, D., Tulus, 2017. Weighted hybrid technique for recommender system. IOP Conf. Series: Journal of Physics: Conf. Series 930: 012050. doi :10.1088/1742-6596/930/1/012050.
- [60] Mr. Avadhut D.Wagavkar, Prof. Mrs.S.S.Vairagar, Weighted Hybrid Approach in Recommendation Method, International Journal of Computer Science Trends and Technology (IJCST) –5 (2), 2017.
- [61] <http://www.oxfordbibliographies.com/view/document/obo-9780199828340/obo-9780199828340-0120.xml>, Erişim Tarihi: 20.04.2019.
- [62] <http://www.pitt.edu/~peterb/3954-061/hybrid%20summary.doc>, Erişim Tarihi: 01.05.2019.
- [63] Ghazanfar, M. Ali., Bennett, A. Prugel., Building Switching Hybrid Recommender System Using Machine Learning Classifiers and Collaborative Filtering, IAENG International Journal of Computer Science, 37:3, IJCS\_37\_3\_09.

- [64] [https://readwrite.com/2009/01/28/5\\_problems\\_of\\_recommender\\_systems/](https://readwrite.com/2009/01/28/5_problems_of_recommender_systems/), Erişim Tarihi: 01.02.2019.
- [65] Robin Burke, Hybrid recommender systems: Survey and experiments, *User Modeling and User-Adapted Interaction* 12 (2002), no. 4, 331– 370.
- [66] Bobadilla, Jesús; Ortega, Fernando; Hernando, Antonio; Bernal, Jesús (February 2012). "A collaborative filtering approach to mitigate the new user cold start problem". *Knowledge-Based Systems*. **26**: 225 – 238. doi:10.1016/j.knosys.2011.07.021.
- [67] Huang, Zan; Chen, Hsinchun; Zeng, Daniel (1 January 2004). "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering". *ACM Transactions on Information Systems*. **22** (1): 116–142. doi:10.1145/963770.963775.
- [68] Salter, J.; Antonopoulos, N. (January 2006). "CinemaScreen Recommender Agent: Combining Collaborative and Content-Based Filtering". *IEEE Intelligent Systems*. **21** (1): 35 – 41. doi:10.1109/MIS.2006.4.
- [69] Burke, Robin (2007). "Hybrid Web Recommender Systems". *The Adaptive Web*. Springer Berlin Heidelberg: 377 – 408. doi:10.1007/978-3-540-72079-9\_12.
- [70] Elahi, Mehdi; Ricci, Francesco; Rubens, Neil. *Active Learning in Collaborative Filtering Recommender Systems*. Springer International Publishing. pp. 113–124. ISBN 978-3-319-10491-1.
- [71] Elahi, Mehdi; Ricci, Francesco; Rubens, Neil (2016). A survey of active learning in collaborative filtering recommender systems. *Computer Science Review – via Elsevier*.
- [72] Andrew I. Schein; Alexandrin Popescul; Lyle H. Ungar; David M. Pennock (2002). *Methods and Metrics for Cold-Start Recommendations*. Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002). New York City, New York: ACM. pp. 253–260. ISBN 1-58113-561-0. Retrieved 2008-02-02.
- [73] "Vendor attempts to crack 'cold start' problem in content recommendations"(PDF). *Mobile Media*. United Kingdom: Informa Telecoms & Media: 18. 2007-06-29. Archived from the original (PDF) on 2008-08-01. Retrieved 2008-02-02.
- [74] Tkalcic, Marko; Chen, Li (2016). "Personality and Recommender Systems". In Ricci, Francesco; Rokach, Lior; Shapira, Bracha. *Recommender Systems Handbook* (2nd ed.). Springer US. ISBN 978-1-4899-7637-6.

- [75] Rubens, Neil; Elahi, Mehdi; Sugiyama, Masashi; Kaplan, Dain (2016). "Active Learning in Recommender Systems". In Ricci, Francesco; Rokach, Lior; Shapira, Bracha. *Recommender Systems Handbook* (2nd ed.). Springer US. doi:10.1007/978-1-4899-7637-6\_24. ISBN 978-1-4899-7637-6.
- [76] Koren, Yehuda; Bell, Robert; Volinsky, Chris (August 2009). "Matrix Factorization Techniques for Recommender Systems". *Computer*. **42** (8): 30 – 37. CiteSeerX 10.1.1.147.8295. doi:10.1109/MC.2009.263.
- [77] Gantner, Zeno; Drumond, Lucas; Freudenthaler, Cristoph (20 January 2011). *2010 IEEE International Conference on Data Mining*. pp. 176 – 185. CiteSeerX 10.1.1.187.5933. doi:10.1109/ICDM.2010.129. ISBN 978-1-4244-9131-5.
- [78] Agarwal, Deepak; Chen, Bee-Chung (28 June 2009). *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining – KDD '09*. ACM. pp. 19 – 28 doi:10.1145/1557019.1557029. ISBN 9781605584959.
- [79] Ferrari Dacrema, Maurizio; Gasparin, Alberto; Cremonesi, Paolo. "Deriving item features relevance from collaborative domain knowledge" (PDF). *Proceedings of Knowledge-aware and Conversational Recommender Systems (KaRS) Workshop 2018 (co-located with RecSys 2018)*.
- [80] Bernardis, Cesare; Ferrari Dacrema, Maurizio; Cremonesi, Paolo (2018). "A novel graph-based model for hybrid recommendations in cold-start scenarios". *Proceedings of the Late-Breaking Results Track Part of the Twelfth ACM Conference on Recommender Systems*. arXiv:1808.10664.
- [81] Lika, Blerina; Kolomvatsos, Kostas; Hadjiefthymiades, Stathes (March 2014). "Facing the cold start problem in recommender systems". *Expert Systems with Applications*. **41** (4): 2065–2073. doi:10.1016/j.eswa.2013.09.005.
- [82] Hou, Lei; Pan, Xue; Liu, Kecheng (7 March 2018). "Balancing the popularity bias of object similarities for personalised recommendation". *The European Physical Journal B*. **91** (3). doi:10.1140/epjb/e2018-80374-8.
- [83] Abdollahpouri, Himan; Burke, Robin; Mobasher, Bamshad (27 August 2017). *Proceedings of the Eleventh ACM Conference on Recommender Systems – Rec Sys '17*. ACM. pp. 42 – 46. doi:10.1145/3109859.3109912. ISBN 9781450346528.
- [84] Rashid, Al Mamunur; Karypis, George; Riedl, John (20 December 2008). "Learning preferences of new users in recommender systems". *ACM SIGKDD Explorations Newsletter*. **10** (2): 90. doi:10.1145/1540276.1540302.



- [85] Bobadilla, Jesús; Ortega, Fernando; Hernando, Antonio; Bernal, Jesús (February 2012). "A collaborative filtering approach to mitigate the new user cold start problem". *Knowledge-Based Systems*. **26**: 225 – 238. doi:10.1016/j.knosys.2011.07.021.
- [86] Park, Yoon-Joo; Tuzhilin, Alexander (23 October 2008). *Proceedings of the 2008 ACM conference on Recommender systems – Rec Sys '08*. ACM. pp. 11–18. CiteSeerX 10.1.1.421.1833. doi:10.1145/1454008.1454012. ISBN 9781605580937.
- [87] Pazzani, Michael J.; Billsus, Daniel (2007). *Content-Based Recommendation Systems. The Adaptive Web*. Lecture Notes in Computer Science. **4321**. pp. 325 – 341. CiteSeerX 10.1.1.130.8327. doi:10.1007/978-3-540-72079-9\_10. ISBN 978-3-540-72078-2.
- [88] Chen, Li; Chen, Guanliang; Wang, Feng (22 January 2015). "Recommender systems based on user reviews: the state of the art". *User Modeling and User-Adapted Interaction*. **25** (2): 99 –154. doi:10.1007/s11257-015-9155-5.
- [89] Ghazanfar, M. Ali., Bennett, A. Prugel., Fulfilling the Needs of Gray-Sheep Users in Recommender Systems, A Clustering Solution, International Conference on Information Systems and Computational Intelligence, China, 2011-01-18 - 2011-01-20, 2011.
- [90] <https://medium.com/datadriveninvestor/dimensionality-reduction-techniques-27049b5a4c55>, Erişim Tarihi: 23.04.2019.
- [91] Manos Papagelis, Dimitris Plexousakis and Themistoklis Kutsuras "Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences". iTrust'05 Proceeding of 3rd International on Trust Management. ACM DL(2005).
- [92] Qian Wang, Xianhu Yuan, Min Sun "Collaborative Filtering Recommendation Algorithm Based On Hybrid User Model", FSKD, 2010.

## **ÖZGEÇMİŞ**

Şahiner Güler, 15.11.1980 yılında İzmir’de doğdu. İlk, orta ve lise eğitimini İstanbul’da yaptı. 1997 yılında Orhan Cemal Fersoy Lisesi’nden mezun olup, 1998 yılında başladığı Sakarya Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü’nü 2003 yılında bitirdi. 2011 yılında Sakarya Üniversitesi Bilgisayar ve Bilişim Mühendisliği Bölümü’nde yüksek lisans eğitimine başladı. 2013 yılında Muş Alparslan Üniversitesi’nde öğretim görevlisi olarak başladığı görevine devam etmekte olup, evli ve 3 çocuk babasıdır.