



**TAVSİYE SİSTEMLERİ İLE KULLANICI DAVRANIŞLARINA UYGUN  
REKLAM SEÇİMİ**

**Mehmet Türkey YOLDAR**

**DOKTORA TEZİ**

**YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ  
BİLİŞİM ENSTİTÜSÜ**

**HAZİRAN 2019**

Mehmet Türkay YOLDAR tarafından hazırlanan “TAVSİYE SİSTEMLERİ İLE KULLANICI DAVRANIŞLARINA UYGUN REKLAM SEÇİMİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi YÖNETİM BİLİŞİM SİSTEMLERİ Anabilim Dalında DOKTORA TEZİ olarak kabul edilmiştir.

**Danışman:** Doç. Dr. Uğur ÖZCAN

Endüstri Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ....

**Başkan:** Doç. Dr. Metehan TOLON

Pazarlama Anabilim Dalı, Hacı Bayram Veli Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ....

**Üye:** Doç. Dr. Diyar AKAY

Endüstri Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ....

**Üye:** Doç. Dr. Fatih Emre BORAN

Enerji Sistemleri Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ....

**Üye:** Dr. Öğr. Üyesi Evrencan ÖZCAN

Endüstri Mühendisliği Anabilim Dalı, Kırıkkale Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ....

Tez Savunma Tarihi: 19/06/2019

Jüri tarafından kabul edilen bu tezin Doktora Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....  
Doç. Dr. Aslıhan TÜFEKÇİ

Bilişim Enstitüsü Müdürü

## ETİK BEYAN

Gazi Üniversitesi Bilişim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Mehmet Türkay YOLDAR

19/06/2019



TAVSİYE SİSTEMLERİ İLE KULLANICI DAVRANIŞLARINA  
UYGUN REKLAM SEÇİMİ  
(Doktora Tezi)

Mehmet Türkay YOLDAR

GAZİ ÜNİVERSİTESİ  
BİLİŞİM ENSTİTÜSÜ

Haziran 2019

ÖZET

Çevrimiçi reklamcılık alanında, uygun reklamı seçerek hedef kullanıcıya göstermek oldukça önemli fakat bir o kadar da zor bir görevdir. Geleneksel hedefleme yöntemlerinin başarılı sonuçlar vermesine rağmen, bu yöntemler de tüm kullanıcılar için farklı reklamları etkili bir şekilde seçememektedir. Bu çalışmada, çevrimiçi reklam veri kümeleri üzerinde tavsiye sistemlerinde kullanılan işbirlikçi filtreleme yöntemleri uygulanmaktadır. Aynı zamanda kullanıcılara farklı ve etkili reklamlar önermek için yeni bir yöntem önerilmektedir. İkili kümeleme yönteminin ve sıralı ağırlıklı ortalama birleştirme operatörünün kullanımına dayanan bu yöntem ile kullanıcıların reklamlarla ilgili bir geçmişi olmasa bile gezinti geçmişleri yardımıyla işbirlikçi bir şekilde reklam tavsiyesinde bulunulabilmektedir. İşbirlikçi filtreleme yöntemlerinin ve de önerilen yöntemin farklı deneysel kurulumlar ile elde edilmiş çevrimdışı değerlendirme sonuçları paylaşılmaktadır. Bu çalışmada elde edilen sonuçlar, işbirlikçi filtreleme yöntemlerinin reklam tavsiyesinde etkili bir araç olarak kullanılabilmesi tekrarlamakta ve önerilen yöntemin özellikle gezinme geçmişi gibi kullanıcı davranışlarının iyi bilindiği sistemlerde daha başarılı sonuçlar vereceğini göstermektedir.

Bilim Kodu : 114603  
Anahtar Kelimeler : Çevrimiçi Reklamcılık, Davranışsal Hedefleme, Tavsiye Sistemleri, İşbirlikçi Filtreleme, İkili kümeleme, OWA  
Sayfa Adedi : 97  
Danışman : Doç. Dr. Uğur ÖZCAN

ADVERTISEMENT SELECTION BASED ON USER BEHAVIOR  
BY RECOMMENDER SYSTEMS

(Ph. D. Thesis)

Mehmet Türkay YOLDAR

GAZİ UNIVERSITY  
INFORMATICS INSTITUTE

June 2019

ABSTRACT

In online advertising, it is crucial to show relevant advertisements to target users. However, this is a challenging process. Although conventional targeting approaches yield useful results, they cannot effectively select different advertisements for all users. In this study, collaborative filtering techniques are explored on online advertisement data sets. Also, a new method is proposed to recommend different and compelling advertisements to users. The proposed method, which is based on biclustering and ordered weighted average aggregation operators, can address circumstances such as the lack of implicit feedback on advertisements. Results of the offline analysis of the proposed method and other collaborative filtering methods are presented. It is shown that collaborative filtering methods are beneficial and that the proposed method provides better results, especially in systems where users' behaviors, like their navigation histories, are well known.

Science Code : 114603  
Keywords : Online Advertising, Behavioral Targeting, Recommender Systems, Collaborative Filtering, Biclustering, OWA  
Page Number : 97  
Supervisor : Assoc. Prof. Dr. Uğur ÖZCAN

## TEŐEKKÜR

Çalıőmalarım boyunca deęerli yardım ve katkılarıyla beni yönlendiren, kıymetli tecrübelerinden faydalandığım danışmanım Doç. Dr. Uęur Özcın'a teőekkürlerimi sunmayı borç bilirim.

Çalıőmanın hazırlanmasının her aşamasında bilgi ve deneyimleriyle çalıőmaya yön veren, ilgi ve desteklerini esirgemeyen tez izleme komitesi hocalarım Doç. Dr. Diyar Akay'a ve Doç. Dr. Metehan Tolon'a içtenlikle teőekkür ederim.

Ayrıca, uzun ve zorlu tez süreci boyunca manevi desteęiyle beni hiçbir zaman yalnız bırakmayan çok deęerli eőim Zeynep Akkutay Yoldar'a tüm kalbimle çok teőekkür ederim.



## İÇİNDEKİLER

	<b>Sayfa</b>
ÖZET .....	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER .....	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ .....	x
SİMGELER ve KISALTMALAR .....	xi
1. GİRİŞ .....	1
2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR .....	5
2.1. Çevrimiçi Reklamcılık.....	5
2.1.1. İçeriğe dayalı hedefleme .....	7
2.1.2. Demografik hedefleme.....	8
2.1.3. Konum tabanlı hedefleme .....	8
2.1.4. Davranışsal hedefleme .....	9
2.2. Kullanıcı Profilleri ve Web Madenciliği .....	10
2.2.1. Web madenciliği .....	13
2.2.2. Web kullanım madenciliği .....	14
2.3. Kümeleme Yöntemleri ve İkili Kümeleme .....	17
2.3.1. Kümeleme yöntemleri.....	18
2.3.2. İkili kümeleme .....	19
2.4. Tavsiye Sistemleri .....	25
2.4.1. Tavsiye süreci .....	26
2.4.2. Tavsiye yöntemleri.....	27
2.4.3. Tavsiye sistemlerinin karşılaştığı zorluklar .....	30
2.5. Birleştirme İşlemleri ve OWA Birleştirme Operatörleri .....	33
2.5.1. Sıralı ağırlıklı ortalama (OWA) birleştirme operatörleri .....	34
2.6. Bölüm Sonucu .....	39
3. YÖNTEM .....	41
3.1. Veri Kümelerinin Tanıtımı .....	41
3.1.1. Gerçek dünya veri kümesi .....	41

	<b>Sayfa</b>
3.1.2. Veri hazırlama.....	42
3.1.3. Sentetik veri kümesi.....	45
3.2. Genel Bakış .....	46
3.3. Kullanıcı Profillerinin Oluşturulması .....	47
3.3.1. Kullanıcı ziyaret matrisi.....	48
3.3.2. Kullanıcı ilgi matrisi .....	48
3.3.3. Kullanıcı demografi matrisi .....	50
3.3.4. Kullanıcı önemi.....	51
3.3.5. Kullanıcı reklam matrisi .....	53
3.4. Kullanıcı ve İlgi Kümelerinin Bulunması .....	53
3.5. Tavsiye .....	54
3.5.1. Kullanıcının ait olduğu ikili kümelerin bulunması .....	55
3.5.2. Reklam seçiminin yapılması .....	56
3.6. Bölüm Sonucu .....	61
4. BULGULAR VE YORUM.....	63
4.1. Değerlendirme Ölçütleri.....	63
4.2. Parametre Optimizasyonu .....	64
4.3. Deneysel Kurulum.....	64
4.4. Deney Sonuçları ve Sonuçların Yorumlanması .....	66
4.4.1. Deney 1: Tüm veri kümesi (gerçek veri kümesi).....	67
4.4.2. Deney 2: En çok ziyarette bulunan 100 kullanıcı (gerçek veri kümesi) ..	69
4.4.3. Deney 3: En az ziyarette bulunan 100 kullanıcı (gerçek veri kümesi) ....	71
4.4.4. Deney 4: Tüm veri kümesi (sentetik veri kümesi).....	73
4.4.5. Deney 5: Uzun kuyruk (sentetik veri kümesi) .....	75
4.5. Bölüm Sonucu .....	76
5. GENEL SONUÇLAR VE ÖNERİLER.....	79
5.1. Öneriler.....	81
KAYNAKLAR .....	85
ÖZGEÇMİŞ .....	97

## ÇİZELGELERİN LİSTESİ

<b>Çizelge</b>	<b>Sayfa</b>
Çizelge 2.1. İşbirlikçi filtreleme ve içerik tabanlı yöntemlerinin kıyaslanması .....	33
Çizelge 3.1. Veri türleri .....	42
Çizelge 3.2. Dış kaynaklardan temin edilen veri türleri .....	43
Çizelge 3.3. Reklam tıklanma verilerine ait bazı istatistikî değerler .....	45
Çizelge 3.4. Gezinti verilerine ait bazı istatistikî değerler .....	45
Çizelge 3.5. Kullanıcı ziyaret matrisi, V .....	52
Çizelge 3.6. Yerel kullanıcı ilgi matrisi, L .....	52
Çizelge 3.7. Küresel ilgi vektörü, g .....	52
Çizelge 3.8. Kullanıcı ilgi matrisi, I .....	53
Çizelge 3.9. Kullanıcı önem vektörü, imp .....	53
Çizelge 3.10. Kullanıcı reklam matrisi, A .....	53
Çizelge 3.11. İkili kümedeki kullanıcıların ilgisi .....	56
Çizelge 3.12. Hedef kullanıcının ilgisi .....	56
Çizelge 3.13. OWA için yeniden sıralama adımının gösterimi .....	59
Çizelge 3.14. İki farklı dilsel niceleyici (az ve çoğu) için elde edilen ağırlık vektörleri	60
Çizelge 3.15. Her reklam için hesaplanan ve sıralanan OWA ve IOWA değerleri. ....	61
Çizelge 4.1. Reklamların sınıflandırılması .....	63
Çizelge 4.2. Parametreler ve değerleri .....	65
Çizelge 4.3. Deney 1 için karşılaştırılan yöntemlerin değerlendirme ölçütleri .....	67
Çizelge 4.4. Deney 2 için karşılaştırılan yöntemlerin değerlendirme ölçütleri .....	69
Çizelge 4.5. Deney 3 için karşılaştırılan yöntemlerin değerlendirme ölçütleri .....	71
Çizelge 4.6. Deney 4 için karşılaştırılan yöntemlerin değerlendirme ölçütleri .....	73
Çizelge 4.7. Deney 5 için karşılaştırılan yöntemlerin değerlendirme ölçütleri .....	75
Çizelge 4.8. Tüm deneylerde en iyi başarımları gösteren yöntemlerin karşılaştırılması ....	77

## ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Web madenciliği süreci .....	14
Şekil 2.2. İkili küme çeşitleri .....	21
Şekil 2.3. Tavsiye süreci .....	26
Şekil 2.4. Dilsel niceleyiciler; birkaç, en az yarısı, mümkün olduğunca çok ve çoğu. .	38
Şekil 3.1. Veri zenginleştirme işlemlerinin özetlenmiş gösterimi .....	44
Şekil 3.2. Önerilen yöntemin ana hatlarıyla gösterimi .....	47
Şekil 3.3. İkili kümeleme öncesi (solda) ve sonrası (sağda) kullanıcı ilgi matrisi .....	54
Şekil 3.4. Tavsiye edilecek reklamların seçilmesinde izlenen akış şeması. ....	58
Şekil 4.1. Önerilen yönteme uygulanan k-kat çapraz doğrulama yöntemi .....	66
Şekil 4.2. Deney 1 için hassasiyet, hatırlama ve $F_1$ değerleri .....	68
Şekil 4.3. Deney 2 için hassasiyet, hatırlama ve $F_1$ değerleri .....	70
Şekil 4.4. Deney 3 için hassasiyet, hatırlama ve $F_1$ değerleri .....	72
Şekil 4.5. Deney 4 için hassasiyet, hatırlama ve $F_1$ değerleri .....	74
Şekil 4.6. Deney 5 için hassasiyet, hatırlama ve $F_1$ değerleri .....	76

## SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

<b>Simgeler</b>	<b>Açıklama</b>
$\alpha$	Veyalık derecesi
$a_i$	i-inci argüman
$b_j$	$a_i$ değerlerinin j-inci büyük olanı
$\delta$	Kullanıcı demografisi
$d_a$	Demografik bilgi: yaş aralığında bulunma olasılığı, %
$d_{female}$	Demografik bilgi: kadın olma olasılığı, %
$d_{male}$	Demografik bilgi: erkek olma olasılığı, %
$g$	Küresel ilgi vektörü
$i$	Kullanıcı ilgisi
$imp$	Kullanıcı önemi
$l$	Yerel ilgi
$p$	Kullanıcı ilgi vektörü
$t$	Eşik değer
$u_j$	$b_j$ kriterine ait önem derecesi (Eş. 2.13)
$u_j$	Uyarılmış sıralama bileşeni (Eş. 2.8)
$v$	Ziyaret sayısı
$w_j$	Ağırlık
$A$	Kullanıcı reklam matrisi
$B$	Sıralanmış argümanlar vektörü
$D$	Kullanıcı demografi matrisi
$F_1$	$F_1$ değeri
$H$	Dağınım derecesi
$I$	Kullanıcı ilgi matrisi
$J$	Jaccard indeksi
$L$	Yerel ilgi matrisi
$P$	Hassasiyet

$Q$	Üyelik fonksiyonu
$R$	Hatırlama
$V$	Kullanıcı ziyaret matrisi
$W$	Ağırlık vektörü
$W^T$	Ağırlık vektörünün devriği
@N	Tavsiye sayısı

### Kısaltmalar

### Açıklama

<b>ABD</b>	Amerika Birleşik Devletleri
<b>AUC</b>	ROC eğrisi altındaki alan
<b>BPR</b>	Bayes kişileştirilmiş sıralama
<b>BPRKNN</b>	En yakın komşu tabanlı Bayes kişileştirilmiş sıralama
<b>BPRMF</b>	Matris ayrıştırma tabanlı Bayes kişileştirilmiş sıralama
<b>CPA</b>	Etkileşim başına ödeme (Cost Per Action)
<b>CPC</b>	Tıklama başına ödeme (Cost Per Click)
<b>CPL</b>	Fırsat başına ödeme (Cost Per Lead)
<b>CPM</b>	Bin gösterim başına ödeme (Cost Per Mille)
<b>CPV</b>	Gösterim başına ödeme (Cost Per View)
<b>CTR</b>	Tıklama oranı (Click Through Rate)
<b>GPS</b>	Küresel Konumlandırma Sistemi
<b>IBCF</b>	Öge tabanlı işbirlikçi filtreleme
<b>IOWA</b>	Uyarılmış sıralı ağırlıklı ortalama
<b>LOWA</b>	Dilsel sıralı ağırlıklı ortalama
<b>MF</b>	Matris ayrıştırma
<b>OPSM</b>	Sırası korunmuş alt matris
<b>OWA</b>	Sıralı ağırlıklı ortalama
<b>OWGA</b>	Sıralı ağırlıklı geometrik ortalama
<b>RMSE</b>	Kare ortalamalarının karekökü hatası
<b>ROC</b>	Alıcı işletim karakteristiği
<b>SAMBA</b>	İkili küme analizi için istatistiksel-algoritmik yöntem
<b>SLIM</b>	Seyrek Doğrusal Yöntemler
<b>SVD</b>	Tekil değer ayrışımı
<b>UBCF</b>	Kullanıcı tabanlı işbirlikçi filtreleme
<b>WRMF</b>	Ağırlıklandırılmış düzenli matris ayrıştırma

## 1. GİRİŞ

Çevrimiçi reklamcılık tüm işletmelerin yeni müşteriler bulmaları, yeni gelir kaynakları yaratmaları ve tanınırlıklarını arttırmaları için kullanabileceği en etkili yollardan birisidir. Pazardaki payı 2017 yılında 83 milyar ABD dolarına (IAB, 2018a) kadar yükselen çevrimiçi reklamcılık, görüntülü reklamlardan ücretli sosyal medya tanıtımlarına kadar geniş bir alanı kapsamaktadır.

Başlangıçta ana paydaşları kullanıcılar, reklam verenler ve yayıncılar olan çevrimiçi reklamcılık alanı, reklam verenlerin ve yayıncıların sayısının artmasıyla, üçüncü taraf bir hizmet sağlayıcıya, reklam dağıtımıcılara da ihtiyaç duyar hale gelmiştir (Chen, Cox, Uluagac ve Copeland, 2016).

Bir reklamın her bir görüntülenmesine izlenim, kullanıcının reklam ile etkileşmesine tıklama, her bir izlenime karşın gelen tıklama sayısına tıklama oranı denilmektedir (Rosenkrans, 2007). Tıklama oranını artırmak için kullanılan yöntemler hesaplamalı reklamcılık alanında değerlendirilebilir. Hesaplamalı reklamcılığın temel sorunu, belirli bir bağlamda belirli bir kullanıcı için en iyi reklamları bulmaktır (Broder, 2008). Hedefleme, reklam gösterimlerinin anlamlı ve değerli olmasını sağlamak için kullanılan ana yöntem haline gelmiştir. En yaygın hedefleme yöntemlerinden birisi davranışsal hedeflemedir. Günümüzde, hedefleme tekniklerinin etkinliği büyük veri kullanımı sayesinde önemli ölçüde artmıştır. Fakat hala bazı zorluklar vardır. Bir reklamın içeriği, kullanıcıların geçmişi veya web sitesinin içeriği ile ilişkilendirilmedikçe, bu reklam etkili bir şekilde gösterilememektedir. Yaş, konum, cinsiyet, gelir gibi demografik verilere dayalı reklam gösterimi, diğer yöntemlerle birleştirilmediğinde ise yüzeysel kalmaktadır.

Bu çalışmaya başlanılmasına sebep olan temel soru, *“İçeriğe dayalı ve davranışsal hedeflemenin ötesinde, kullanıcılara nasıl farklı reklamlar gösterilebilir?”* idi. Bu soruyu yanıtlama amacıyla çevrimiçi reklamcılık alanında tavsiye sistemlerinin, özellikle işbirlikçi filtreleme tekniklerinin kullanımı araştırılmıştır. Bu çalışmada ek bir araştırma sorusu daha incelenmektedir: *Seyrek bir veri kümesinde yeni kullanıcılara reklamlar etkin bir şekilde nasıl önerilebilir?* Kullanıcılar çoğunlukla reklamlarla etkileşime girmekten kaçınırlar ve bu da dolaylı geri bildirimleri toplamayı zorlaştırır. Soğuk başlangıç problemi, tavsiye

sistemleri için bilinen sorunlardan biridir, aynı zamanda reklam veri kümesi doğası gereği çok seyrekler. Seyreklik de tavsiye sistemlerinde çözülmeye çalışılan sorunlardan bir diğeridir. Önerilen yöntem ile kullanıcı daha önceden hiçbir reklama tıklamamış olsa bile tavsiyelerde bulunulabilmektedir.

Hazırlanan bu çalışma ile hesaplamalı reklamcılık yöntemlerinden davranışsal hedefleme ve tavsiye sistemlerinde kullanılan yöntemlerden biri olan işbirlikçi filtreleme çevrimiçi reklamcılık veri kümeleri üzerinde uygulanmaktadır. İhtiyaç duyulan kullanıcı profilleri web kullanım madenciliği ile elde edilmekte, kullanıcılar ve ilgi alanları ikili kümeleme ile kümelenebilmektedir. Web kullanım madenciliğine girdi olan veriler, diğer benzer çalışmaların aksine kullanıcıların tek bir web sitesi ya da uygulama üzerindeki hareketlerinden değil, tüm gezinti geçmişlerinden ve davranışlarından oluşmaktadır. Reklam önerisi ise ikili kümede bulunan kullanıcılarının ortak karar vermesi üzerine sıralı ağırlıklı ortalama birleştirme operatörü yardımıyla gerçekleştirilmektedir. Önerilen bu yeni yöntem alanlar arası tavsiye sistemleri sınıfında değerlendirilebilir. Aynı zamanda kullanılan verilerde ve oluşturulan kullanıcı profillerinde kullanıcıya ait tanımlayıcı bilgiler kullanılmaması, günümüzde yükselen gizlilik ile ilgili endişeler için de önerilen yöntemin bir alternatif olabileceğini düşündürmektedir.

Bu çalışma kapsamına çevrimiçi reklamcılık türlerinden biri olan görüntülü reklamlar dahil edilmiş, diğer reklam türleri göz ardı edilmiştir. Bu çalışmada oluşturulan yöntemin ve sonuçlarının doğruluğu çeşitli kaynaklardan temin edilerek kullanılan verilerin doğruluğuna bağlıdır. Temin edilen veriler anonimdir; bu nedenle, kişiler doğrudan veya veri kümesindeki diğer meta tanımlayıcılar aracılığıyla tanımlanamaz. Özet veriler, bilgilerin, kişilerin kimlikleriyle bağlantılı olamayacağı şekilde temin edilmiştir. Çalışmanın deneysel bölümlerinde R projesi (R Core Team, 2013) ve kütüphanelerinden yararlanılmıştır.

Bu çalışma beş ana bölümden oluşmaktadır;

- *Bölüm 2: Kavramsal Çerçeve ve İlgili Araştırmalar* altında çalışmanın temelini oluşturan kavramlardan çevrimiçi reklamcılık, kullanıcı profilleri, web madenciliği, kümeleme yöntemleri, ikili kümeleme, tavsiye sistemleri ve tavsiye sistemlerinde



birleřtirme operatörlerine deęinilmekte ve bu kavramlarla ilgili yapılan temel ve güncel arařtırmaları içeren literatür taraması yer almaktadır.

- *Bölüm 3: Yöntemde* tez çalışması kapsamında kullanılan veri kümeleri tanıtılmakta ve ardından önerilen yöntem oluştururken izlenen adımlar olan kullanıcı profillerinin oluşturulması, kullanıcı ve ilgi kümelerinin bulunması ve tavsiye adımları detaylı olarak açıklanmaktadır.
- *Bölüm 4: Bulgular ve Yorumda* işbirlikçi filtreleme yöntemleri ve önerilen yöntemin başarımı k katlamalı çapraz doğrulama yöntemi ile hesaplanmakta, elde edilen bulgular paylaşılmakta ve bu bulgular yorumlanmaktadır.
- Son olarak *Bölüm 5: Genel Sonuçlar ve Öneriler* altında ise çalışmanın sonuçları değerlendirilmekte ve gelecek çalışmalar için öneriler paylaşılmaktadır.



## 2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde çalışmanın temelini oluşturan kavramlara değinilmekte ve bu kavramlardan çevrimiçi reklamcılık, kullanıcı profilleri, web madenciliği, kümeleme yöntemleri, ikili kümeleme, tavsiye sistemleri ve tavsiye sistemlerinde birleştirme işlemleri üzerine yapılan temel ve güncel araştırmaları içeren literatür taramasına yer verilmektedir.

### 2.1. Çevrimiçi Reklamcılık

Çevrimiçi reklamcılık (ya da internet reklamcılığı) her ölçekteki işletmenin erişimlerini genişletmek, yeni müşteriler bulmak, yeni gelir kaynakları yaratmak ve de tanınırlıklarını arttırmak için kullandığı en etkili yollardan birisidir.

Görüntülü reklamlardan (İng. Display Ads), ücretli sosyal medya tanıtımlarına, uygulama içi reklamlardan, elektronik posta tanıtımlarına kadar geniş bir alanı kapsayan çevrimiçi reklamcılık gelirleri özellikle 2010 yılında kazandığı ivme ile 2016 yılında 72,5 milyar ABD Doları'na, 2017 yılında ise yaklaşık %14 artarak 83 milyar ABD Doları'na kadar yükselmiştir (IAB, 2018a). Türkiye'de ise 2017 yılı için dijital reklam yatırımları 2,2 milyar TL'ye ulaşarak, bir önceki yıla göre yaklaşık %15 artmıştır (IAB, 2018b).

Çevrimiçi reklamcılıktaki ana paydaşlar kullanıcılar, reklam verenler, yayıncılar ve de reklam dağıtımçılarıdır. Reklam veren tanıtımını yapmak istediği mala ya da hizmete ait reklamı, yayıncıların sahip olduğu web siteleri ya da mobil uygulamalar üzerinden kullanıcılara ulaştırmak ister. Reklam verenlerin ve de yayıncıların sayısı ve çeşitliliği düşünüldüğünde bu işbirliğini sağlamak için bir üçüncü parti hizmet sağlayıcıya yani reklam dağıtımçılarına ihtiyaç ortaya çıkmıştır (Chen ve diğerleri, 2016). Bir reklamın her bir görüntülenmesine izlenim ya da gösterim (İng. impression), kullanıcının reklam ile etkileşmesine tıklama (İng. click), her bir izlenime karşın gelen tıklama sayısına tıklama oranı (İng. CTR: Click Through Rate) denilmektedir (Rosenkrans, 2007).

Mangani (2004) çevrimiçi reklamcılığın ilk dönemlerinde var olan iki fiyatlandırma modelini kıyaslamıştır. CPV (İng. Cost Per View); her gösterim başına reklam veren tarafından ödeme yapıldığı, CPC (İng. Cost Per Click) ise kullanıcının reklama tıklaması

halinde reklam verenin ödeme yaptığı fiyatlandırma modelleridir.

Yakın zamanlı bir çalışmada ise Chen ve diğerleri (2016) çevrimiçi reklam türlerini görüntülü reklamcılık ve arama reklamcılığı olarak iki, fiyatlandırma modellerini ise dört ana başlık altında toplamıştır. Bir web sayfasında yer alan afiş, açılır reklam, video ve benzeri tüm reklamlar görüntülü reklam olarak adlandırılmaktadır. Arama reklamları bir anahtar kelime ya da metin ile arama motorunda arama yapıldığında gösterilen reklamlardır. Görüntülü reklamcılıkta yaygın olarak kullanılan modellerden CPM (İng. Cost Per Mille); her bin gösterim başına reklam veren tarafından ödeme yapıldığı bir fiyatlandırma modelidir. Kullanıcının reklama tıklaması halinde ödeme yapılan model olan CPC ise hem görüntülü hem de arama reklamcılığında tercih edilmektedir. CPA (İng. Cost Per Action) ve CPL (İng. Cost Per Lead) modellerinde ise kullanıcının reklamı görmesi ya da tıklamasından ziyade, tıklamanın ardından bir işlemi gerçekleştirmesi (örneğin malı satın alması ya da iletişim bilgilerini bırakması) durumunda reklam veren tarafından ödeme yapılır. Görüntülü reklamlarda tüm ödeme yöntemleri kullanılırken, arama reklamlarında ise CPM tercih edilen bir yöntem değildir (Asdemir, Kumar ve Jacob, 2012; Chen ve diğerleri, 2016; Hoffman ve Novak, 2000).

İlk örneklerine 1993'de rastlanan afiş reklamlar (İng. banner), kullanılan en eski standart format olarak düşünülmektedir (Chen ve diğerleri, 2016). O günden günümüze çevrimiçi reklamcılığın muazzam ölçüde büyümesine rağmen, 2000'li yılların başında %2 ile %4 arasında olan tıklama oranları artan reklam gösterim sayıları nedeniyle hızla düşerek (Kazienko ve Adamski, 2007) bugün ortalama %0,05 seviyesine gerilemiştir (SmartInsights, 2018).

Tıklama oranının artırılması için kullanılan yöntemler hesaplamalı reklamcılık (İng. Computational Advertising) alanı altında değerlendirilebilir. Tıklama oranı ile ilgili yöntemler bu alanın sadece çok küçük bir bölümünü oluşturmaktadır. Karmaşık bir ekosisteme sahip çevrimiçi reklamcılıkta, bir reklamın izleniminden yola çıkarak muhtemelen bir satın alma sürecine kadar olan yolculuğunu daha iyi bir şekilde modellemeye çalışan (Rajan, 2017) hesaplamalı reklamcılık çoğu zaman çevrimiçi reklamcılık ya da dijital reklamcılık olarak adlandırılrsa da aslında bilgi toplama, istatistiksel modelleme, makine öğrenimi, optimizasyon, büyük ölçekli arama ve metin analizinin

kesiřtiđi noktada yer alan, çevrimiçi reklamcılıđın bilimsel bir alt disiplindir (Dave ve Varma, 2014). Hesaplamalı reklamcılıkta ele alınan temel sorun, belirli bir bağlam için en uygun reklamları bulmaktır (Broder, 2008).

Reklam dağıtımıcıları en uygun reklamın seçimini, reklam veren tarafından verilen gösterim tekliflerini büyükten küçüđe sıralayarak yapmaktaydı. Sezgisel olarak geliri en üst düzeye çıkarmanın en belirgin yolu olan bu yöntem, reklam dağıtımıcılarının en yüksek teklifli reklamların içerikle alakalı olmadığını fark etmeleri ile sona ermiştir. CPM dışındaki fiyatlandırma modellerinin reklam verenler tarafından daha fazla tercih edilir olmasıyla ve de alakasız reklamlar kullanıcı memnuniyetsizliđi ile sonuçlanabileceđi için (Wang, Zhang, Choi ve D'Eredita, 2002); reklam dağıtımıcıları reklamları alaka düzeyi ve beklenen gelirin bir fonksiyonu olarak sıralayarak göstermeye başlamışlardır (Richardson, Dominowska ve Ragno, 2007). İçeriđe göre reklamları gösterebilmek genellikle iki adımlı bir yol izlenerek mümkün olmuştur. İlk adımda, reklam veri tabanından içerikle örtüşükleri derecede reklamlar sıralanarak çekilir; ardından seçilen en iyi reklamlar, beklenen geliri en üst seviyeye çıkaracak şekilde tekrar sıralanır. Böylelikle kullanıcı, reklam veren, yayıncı ve de reklam dağıtımıcısı olmak üzere tüm paydaşların gereksinimleri de karşılanmış oluyordu. Daha sonraları içeriđe dayalı hedefleme olarak adlandırılacak olan bu basit iki adımlı yöntemin evrimleşmiş birçok çeşidi günümüzde kullanılmaya devam edilmektedir.

Hedefleme, reklam gösterimlerinin anlam kazanması ve deđerli hale gelmesi için kullanılan ana adımlardan birisidir. Hesaplamalı reklamcılıkta kullanılan hedefleme yöntemlerinden en yaygın olanları içeriđe dayalı hedefleme, demografik hedefleme, konum tabanlı hedefleme ve davranışsal hedefleme olarak sıralanabilir. Goldfarb (2014) çevrimiçi ve çevrimdışı reklamcılık arasındaki en temel ekonomik farkı, hedefleme maliyetindeki azalış olarak belirtmektedir.

### **2.1.1. İçeriđe dayalı hedefleme**

İçeriđe dayalı hedefleme, tüketicinin görmekte olduđu içeriđe göre reklam dağıtımını yapılmasını ifade eder (Goldfarb, 2014; Zhang ve Katona, 2012). Zhang ve Katona (2012) içeriđe dayalı hedeflemenin çevrimiçi ortamda bir çok farklı hedef kitleye kolaylıkla hedeflenmiş reklamlar sunmak için kullanıldığını, ayrıca gelişmiş içerik analizi

algoritmaları ile de çok verimli hale geldiğini belirtmiştir.

Arama reklamcılığı, reklam verenlerin reklamlarını arama içeriğiyle eşleştirdiği bir içeriğe dayalı hedefleme biçimidir. Görüntülü reklamcılıkta da içeriğe dayalı hedefleme yaygındır: bir teknoloji sitesinde teknolojik mal veya hizmetlere ait reklamlar, tatil sitesinde ise otel reklamları gösterilmektedir.

### **2.1.2. Demografik hedefleme**

Geleneksel yöntemlerden birisi olan ve birçok medya aracında yaygınca kullanılan demografikler, görüntülü ya da metin reklamı dağıtımı yapan herhangi bir çevrimiçi reklam dağıtımıcısı için de olmazsa olmaz veri kaynaklarından birisidir.

Demografik hedefleme; reklam verenlerin belirttikleri demografik özelliklerle eşleşen kullanıcılara, reklam dağıtımıcısı tarafından ilgili reklamların gösterilmesi olarak tarif edilmektedir (Goldfarb, 2014).

Yaş, cinsiyet, konum, ilgi alanları, gelir seviyesi gibi birçok boyuttan oluşabilecek demografik verilerin bir kullanıcı için toplanması ise oldukça zorlayıcı olabilir. Yasal yükümlülüklerin yanı sıra, kullanıcıların internet üzerinde gizliliklerine verdikleri önemin artması ile bu alandaki temel zorluk demografik bilgilerin edinilmesi olmuştur. Kullanıcı tarafından sağlanan demografik bilgiler daha güvenilirdir. Genellikle kayıt esnasında ya da anketler aracılığı ile toplanan bu veriler yine de çoğu zaman kullanıcıların gizlilik endişeleri nedeniyle gerçeği yansıtmamaktadır. Bir diğer yöntem ise kullanıcıların gezinti geçmişlerinden ve davranışlarından demografik bilgilerin tahmin edilmesidir (Hu, Zeng, Li, Niu ve Chen, 2007).

### **2.1.3. Konum tabanlı hedefleme**

Kullanıcının konumunu tespit ederek bölgesel reklamları göstermek için kullanılan yöntemler konum tabanlı hedefleme içerisinde yer alır. Kullanıcı tarafından sağlanmış adres bilgileri kullanılabilirliği gibi, IP adresinden konum tespiti, GPS bilgisi, kullandığı tarayıcının dil ayarları bile hedefleme için kullanılabilir (Plummer, Rappaport, Hall ve

Barocci, 2007). Mobil cihazların yaygınlaşmasıyla kullanıcıya doğru zamanda, doğru yerde, doğru mesaj iletilebilir. Elbette bu tür uygulamalar gizlilik ile ilgili endişeleri arttırabileceği gibi kullanıcının tepkisini de çekebilir (Shankar ve Balasubramanian, 2009).

#### **2.1.4. Davranışsal hedefleme**

Davranışsal hedefleme, kullanıcının ziyaret ettiği sayfalar, yaptığı aramalar gibi gezinti davranışları üzerinden toplanan bilgilerin değerlendirilmesi ile etkili bir şekilde reklam gösterimi yapılması olarak tanımlanabilir (Levene, 2011; Yan ve diğerleri, 2009).

Davranışsal hedeflemenin bir türü olarak görülebilecek olan yeniden hedeflemeyi, Goldfarb (2014) daha önceden aradığı ya da gördüğü bir içerik üzerinden kullanıcıya reklam gösterimi olarak tanımlamıştır. Örneğin bir alışveriş sitesinde bir cep telefonunu inceleyen bir kişi, gezdiği diğer sitelerde aynı siteye ait cep telefonu reklamlarıyla karşılaşabilir. Benzer bir şekilde arama motorunda cep telefonu arayan bir kişi, herhangi bir alışveriş sitesini gezmeseyse bile, gezdiği diğer sitelerde cep telefonu reklamlarıyla karşılaşabilir.

Goldfarb ve Tucker (2011a) hedeflemenin kullanıcıların satın alma niyetini nasıl etkileyebileceğini incelemiş ve iki önemli sonuca ulaşmıştır. İlki, reklamların web sitesi içeriğiyle eşleşmesinin ya da reklamların kendisini fazlasıyla hissettirmesinin (İng. obtrusive) satın alma niyetini arttırdığı; ikincisi ise bu iki yöntemin beraber kullanıldığında rahatsız edici olduğu ve satın alma niyetini azalttığıdır. Bu muhtemelen kullanıcıların gizlilikle ilgili kaygılarından kaynaklanmaktadır.

#### **Gizlilik**

Çevrimiçi reklamcılık alanındaki tüm paydaşlar, kullanıcılar hakkında çok sayıda bilgi toplar ve saklar. Bu verileri toplayan paydaşlar, hizmetlerini geliştirmek için araştırma yapmak amacıyla bunları kullanır. Kullanıcının geçmiş gezinti davranışları, davranışsal hedefleme yöntemlerinin uygulanması için de önemlidir. Bu veriler, kullanıcıların cinsiyetleri ve belirli sporlar gibi satın alma eğilimlerini veya bir tatil planlayıp planlamadıkları gibi genel özelliklerini ortaya çıkarmak için kullanılabilir (Evans, 2009). Bu verilerin toplanması ve kullanılması, tüketici mahremiyeti ile ilgili endişelere neden

olmuştur. Özellikle, İnternet'in hızla yaygınlaşması ile birlikte gizlilik kaygılarının arttığı görülmektedir (Goldfarb, 2014).

Turow, King, Hoofnagle, Bleakley ve Hennessy (2009) ve Malheiros, Jennett, Patel, Brostoff ve Sasse (2012) yaptığı benzer çalışmalarda da kullanıcılar benzer fikirlere sahiptirler. Katılımcıların %66'sı kendi ilgi alanlarına yönelik reklamların gösterilmesini uygun bulmazken, aynı durum haberler için de (%58) geçerlidir (Turow ve diğerleri, 2009). Tucker (2014) çalışmasında sosyal ağ kullanıcılarına daha şeffaf politikalar sağlandığında, kullanıcıların kişileştirilmiş içerik ve reklamlara iki kat daha pozitif tepki gösterdiğini belirtmektedir.

Zhu ve Chang (2016) gizlilik kaygısı ve kişileştirilmiş reklamlar arasındaki ilişkinin anlaşılması üzerine çalışmışlar, son derece alakalı kişiselleştirilmiş reklamların kullanıcıların gizlilikle ilgili kaygılarını hafifletme konusunda elverişli olduğunu göstermişlerdir. Özellikle, kullanıcının ilgi alanlarına ve zevklerine uyan içeriğin doğru bir şekilde sağlanmasıyla, gizlilik ile ilgili endişeler azalmaktadır.

Goldfarb ve Tucker (2011b) gizlilik ile ilgili yasal düzenlemelerin çevrimiçi reklamcılık alanına etkilerini incelemişlerdir. Sonuçlar, Avrupa'da gizlilik ile ilgili yasaların yürürlüğe girmesi ile afiş reklamlar üzerinden satın alma niyetinin ortalama %65 oranında azaldığını göstermektedir. Aynı dönemde yasanın uygulanmadığı diğer ülkelerde benzer bir değişim olmamıştır.

## **2.2. Kullanıcı Profilleri ve Web Madenciliği**

Kullanıcıya uyarlanan herhangi bir sistemin nihai amacı, kullanıcılara ihtiyaç duymadan ihtiyaç duydukları şeyleri açıkça sağlamaktır (Mulvenna, Anand ve Büchner, 2000). Çevrimiçi reklamcılık alanında hedeflemenin en önemli adımı kullanıcı profili oluşturmaktır. Mevcut kişileştirme yöntemleri kullanıcı ilgisinin doğrudan ya da dolaylı olarak ifade edilmesi üzerine kurgulanmaktadır (Mobasher, Dai, Luo ve Nakagawa, 2001). Kullanıcıdan ihtiyaç duyulan veriler geçmiş eylemleri üzerinden çıkartılabileceği gibi doğrudan da talep edilebilir. Kullanıcı profili oluşturmak için kullanıcının sıralı eylemlerinden olasılık tabanlı modeller üretmek, zaman, etkinlik türü ve içeriğine dayalı



ağırlıklandırmalar yapmak, belirli kategorilerdeki olayların sayısını dikkate almak ya da kümeleme ve diğer denetlenmeyen teknikler kullanmak tercih edilebilir (Mobasher, Dai, Luo ve Nakagawa, 2002; Schiaffino ve Amandi, 2009).

Kullanıcı profili oluşturmada kullanılacak gerekli verileri elde etmek için sayfa görüntüleme, satın alma gibi dolaylı ifadelerden faydalanılabileceği gibi beğeni, kayıt formları ya da mal veya hizmet derecelendirmeleri gibi açık ifadelerden de yararlanılabilir (Eirinaki ve Vazirgiannis, 2003).

Mobasher, Cooley ve Srivastava (2000a) internet kullanıcıların davranışlarına göre kişileştirilmiş sonuçlar oluşturmak üzere yaptıkları çalışmada, sürecin kullanıcı profili oluşturmada insan katılımının aktif bir şekilde yer alarak yürüdüğünü ve bunun öznel kullanıcı profilleri oluşmasına neden olduğu belirtilmişler, aynı zamanda zaman içinde değişen kullanıcı tercihlerinin oluşturulan profillerin geçerliliğini kaybettiğinden söz etmişlerdir. Veri madenciliği yöntemleri uygulanarak bu sürecin nesnel, güncel ve otomatik bir şekilde gelmesi için genel bir mimari önerilmiştir.

Nasraoui, Frigui, Krishnapuram ve Joshi (2000) bulanık kümeleme yöntemiyle otomatik olarak internet kullanıcı profilleri çıkartmaya çalışmışlar, bunun için de kullanıcı oturumlarını ve gezilen web sitelerinin yapılarını (içerik, konu) veri olarak kullanmışlardır. Oluşturulan algoritma ile yeni eklenen kullanıcı tercihleri sonucu, tüm verinin tekrar analizine gerek kalmadan, kullanıcı profillerin güncellenmesi sağlanarak dikkate değer bir zaman kazancı elde edilmiştir.

Bu ve benzeri çalışmalardan da (Golemati, Katifori, Vassilakis, Lepouras ve Halatsis, 2007; Hoppe, Nicolle ve Roxin, 2013; Qiu ve Cho, 2006; Raad, Chbeir ve Dipanda, 2010; Soltysiak ve Crabtree, 1998; Sugiyama, Hatano ve Yoshikawa, 2004) görüldüğü üzere, kullanıcı profilleri oluşturulmasında insan faktörünün azaltılarak yerini bilgisayarlara bırakması hem hızlı hem de objektif sonuçların elde edilmesini sağlamaktadır. Kullanıcı davranışları üzerinden toplanan bilgiler veri madenciliği ve makine öğrenimi yöntemleriyle analiz edilerek, kullanıcı profilleri oluşturulabilir.

Kullanıcıların gezinti geçmişleri üzerinden profil oluşturulması için Hafri, Djeraba, Stanchev ve Bachimont (2003) tarafından Markov Modeline göre geliştirilen yaklaşım ile kümeleme yöntemi uygulanarak tahminler yapılmış ve üç önemli nokta üzerinde durulmuştur: kullanıcıların keşfetme şekli, ulaşılan içerik ve kullanıcı davranışları.

Eirinaki ve Vazirgiannis (2003) kişileştirilmiş sonuçlar için istatistiksel ve veri madenciliği yöntemleri uygulamış, gezinti geçmişi ile içerik, kullanıcı profili gibi verileri girdi olarak kullanmışlardır. Kullanıcı grupları ile web sayfaları arasındaki olası ilişkileri çıkarmışlardır.

Kullanıcının herhangi bir katkısı olmadan kullanıcı profili oluşturmak için Sugiyama ve diğerleri (2004) arama sonuçlarına dayalı bir yöntem geliştirmişlerdir. Kullanıcının bir günlük gezinti geçmişi üzerinden detaylı bir analiz ile işbirlikçi filtreleme yapılarak kullanıcı profilleri oluşturulmuştur.

Jin, Zhou ve Mobasher (2004) gizli anlamsal analize dayalı veri madenciliği teknikleri kullanarak kullanıcıların gezinti davranışları üzerine desenler keşfetmeye çalışmışlardır. Kullanıcı tercih ve ilgisini anlayabilmek için, kullanıcıları, kullanıcıların hareketlerini ve web nesnelere üzerinden semantik ilişkileri incelenmişlerdir.

Tüm kişileştirme yaklaşımları, özellikle veri madenciliğine dayalı olan yöntemler, kullanıcıların ilgilerini yansıtacak verilerin doğru bir şekilde toplanmasını gerektirir. Kişiselleştirilmiş sistemler sadece öneriler üretmek veya tahmin yapmak için kullanılan yöntemlerle değil, aynı zamanda kullanıcı profillerinin oluşturulma şekliyle de farklılık gösterir (Mobasher, 2007).

Kural tabanlı ve içerik tabanlı kişiselleştirme sistemleri genellikle kullanıcı ilgisi üzerinden bireysel bir kullanıcı profili oluşturur ve bu profili yalnızca o kullanıcının gelecek etkileşimlerini uyarlamak için kullanır. Bireysel profillere dayanan yaklaşımların önemli bir dezavantajı, tavsiyelerin kullanıcının önceki ilgi alanlarına çok odaklı olması nedeniyle sınırlı kalması; yani kullanıcının belirgin bir şekilde ilgisi olduğu bilinmese bile, ilgileneceği tavsiyelerde bulunabilmesinin yetersizliğidir. Kural tabanlı sistemlerde ise kullanıcı profilleri kişisel ve demografik verilere dayalı olsa da öneri sürecinde demografik verilerin kullanımı çok yaygın değildir. Bunun nedeni, bu verilerin internette toplanmasının daha zor

olması, toplanan verilerin ise kalitesiz olmaya meyilli olmasıdır. Ayrıca, yalnızca demografik verilere dayanan önerilerin, içerik tabanlı veya kullanıcı davranışına göre yapılan önerilere göre daha az doğru olduğu gösterilmiştir (Pazzani, 1999). İşbirlikçi filtrelemede ise sistem sadece etkin kullanıcı için kullanıcının profilini kullanmakla kalmaz, aynı zamanda diğer kullanıcıların profillerini de dikkate alır. Profiller genellikle kullanıcının bir öge alt kümesi üzerinde tercihlerini gösteren bir derecelendirme kümesi olarak temsil edilir. Etkin bir kullanıcının profili, ona benzer tercihlere sahip diğer kullanıcıları, yani komşularını, bulmak için kullanılır (Mobasher, 2007).

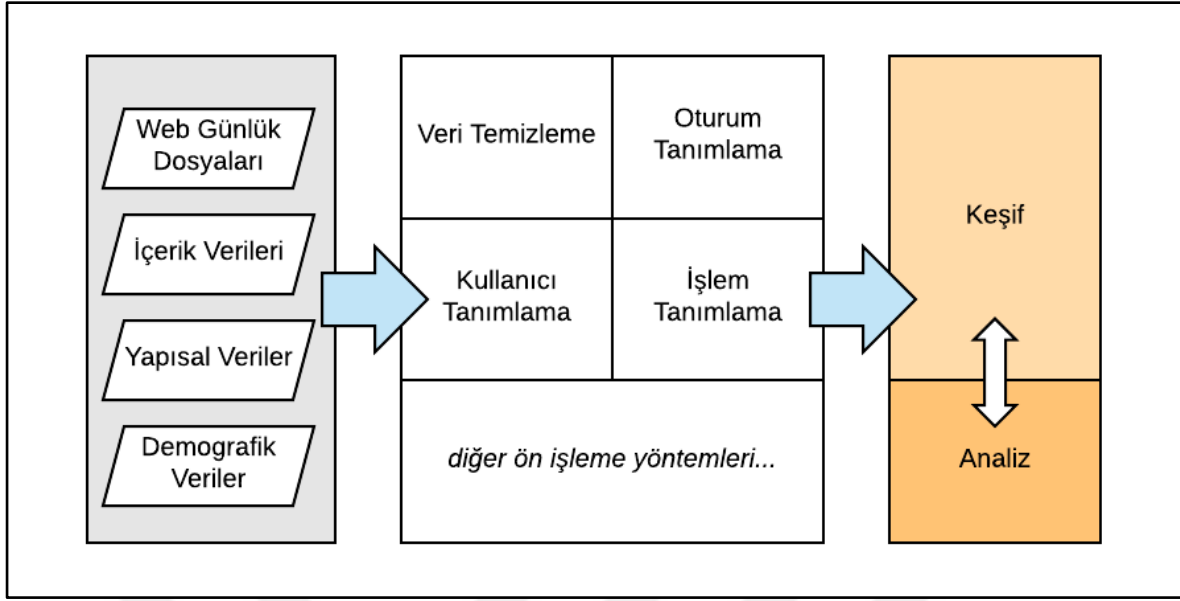
Bu çalışmada ana veri kaynağı olarak kullanıcılardan dolaylı olarak toplanan kayıtları barındıran günlük dosyalarının kullanılması nedeniyle kullanıcı profillerinin oluşturulmasında web madenciliğinin bir alt alanı olan web kullanım madenciliğinden faydalanılmaktadır.

### **2.2.1. Web madenciliği**

Web madenciliği, veri madenciliği yöntemlerinin, tekniklerinin ve de modellerinin internet tabanlı veriler, yapılar ya da kullanım kalıpları üzerinde kullanılması olarak tanımlanır (Markov ve Larose, 2007). Şekil 2.1'de basitleştirilmiş haliyle web madenciliği süreci gösterilmektedir.

Markov ve Larose (2007) web madenciliğini üç alt başlık altında toplamaktadır.

- Yapısal web madenciliği; bilgilerin toplanması, dokümanların indekslenmesi, metin içeriklerine ve bağlantı (köprü) yapılarına göre web sayfalarının sıralanması ve aranması da dahil olmak üzere bilgi çıkarma işlemlerini kapsamaktadır.
- Web içeriği madenciliği; kümeleme ve sınıflandırma gibi makine öğrenimi ve veri madenciliği yaklaşımları ile webde bulunan verinin bilgiye çevrilmesine odaklanmaktadır.
- Web kullanım madenciliği; web verilerinin kullanım kalıplarını açığa çıkarmak için veri madenciliği yöntemlerinin uygulanmasını ifade eder. Takip eden bölümde web kullanım madenciliğine değinilmektedir.



Şekil 2.1. Web madenciliği süreci

### 2.2.2. Web kullanım madenciliği

Web kullanım madenciliği, internet uygulamalarının ihtiyaçlarını daha iyi anlayabilmek ve kullanıcılara daha iyi bir şekilde sunmak için, kullanım kalıplarını keşfetmeye yönelik veri madenciliği tekniklerinin uygulanmasıdır (Srivastava, Cooley, Deshpande ve Tan, 2000).

Web kullanım madenciliğini, yapısal web madenciliği ve web içeriği madenciliğinden ayıran en büyük fark insanların davranışlarını yansıtmasıdır. Kullanıcı davranışının analizi, kullanıcı deneyiminin özelleştirilmesine ve kişiselleştirilmesine yol açabilecek bilgiler sağlayabilir (Markov ve Larose, 2007).

Kendine has özel bir algoritması olmayan web kullanım madenciliği, tipik veri madenciliği döngüsünü izlemektedir. Mobasher (2007) bu özelliği sayesinde web kullanım madenciliğinin farklı veri kaynaklarını kapsamlı bir şekilde kullanabilmek için büyük bir esneklik sağlandığını ve kişiselleştirme görevlerinin diğer mevcut uygulamalarla daha iyi entegre olabileceğini belirtmiştir. Web kullanım madenciliği döngüsü verilerin toplanması, ön işleme, desenlerin keşfi ve analiz aşamalarından oluşmaktadır (Kosala ve Blockeel, 2000; Varnagar, Madhak, Kodinariya ve Rathod, 2013).

Ham verilerin elde edilebileceği kaynaklar genellikle istemci günlük dosyaları, vekil sunucu günlük dosyaları ve de web sunucusu günlük dosyalarıdır (Varnagar ve diğerleri, 2013). Bunlar arasından en yaygın kullanılanı ve belki de en işlevseli web sunucusu günlükleridir. Sunucu tarafından otomatik olarak yaratılan bu günlükler, kullanıcının sunucu ile olan her etkileşiminde oluşturulur. Bu zengin bir içerik sunarken, oldukça büyük olması nedeniyle de işlenmesini zorlaştırabilir. Bir diğer veri kaynağı ise istemcide meydana gelen işlemlerin kayıtlarından oluşan istemci günlük dosyalarıdır. Bu işlemler fare hareketlerinden, içerik seçimine kadar farklılıklar gösterebilir. Bu verilerin toplanarak elde edilmesi için JavaScript dosyaları hazırlanabileceği gibi tarayıcı eklentileri de kullanıcıya sunulabilir (Choi ve Lee, 2009; Varnagar ve diğerleri, 2013). Vekil sunucu günlük dosyaları ise kullanımına bağlı olarak anonim kullanıcılar için gezinti davranışlarını tutabileceği gibi, çok katmanlı olarak çalışan web uygulamalarında kayıtların ortak bir noktada tutulmasına yardımcı olur (Varnagar ve diğerleri, 2013).

Kosala ve Blockeel (2000) web kullanım madenciliği sürecini kaynak bulma, bilgi seçimi ile ön işleme, genelleştirme ve analiz olarak dört ana basamağa ayırmaktadır. Ana veri kaynakları sunucu ve tarayıcı günlükleri gibi dosyalar olan web kullanım madenciliğinde genellikle istatistiksel ya da makine öğrenimi yöntemleri uygulanmakta, kullanım alanları ise öncelikli olarak pazarlama ve kullanıcı modelleme olarak öne çıkmaktadır.

Benzer şekilde Spiliopoulou (2000) web günlükleri üzerinden kullanıcı davranışlarını çıkarmak için üç adımlı bir yoldan bahsetmiştir. Bu adımlar sırasıyla, günlüklerin hazırlanması, gezinti desenlerinin çıkarılması ve sonuçların analizidir. Srivastava ve diğerleri (2000) çalışmasında yine üç adımlı bir süreci tanımlamıştır; ön işleme, desen keşfi ve desenlerin analizi.

Mobasher ve diğerleri (2000a) web kullanımını tabanlı bir kişileştirme mimarisi önermiştir. İki parçalı bu mimaride; çevrim dışı olarak önce veriler işlenmekte, kümeleme ve birliktelik kuralları gibi madencilik işlemleri yapılmaktadır. Ardından elde edilen sonuçlar kullanılarak kullanıcının mevcut gezinti aktivitelerine göre öneriler çevrimiçi olarak sunulmaktadır. Web kullanım madenciliği ile web içeriği madenciliği beraber kullanıldığında, kısıtlı kullanım verileri olduğu durumlara göre daha etkili bir kişileştirme elde edilmektedir (Mobasher, Dai, Luo, Sun ve Zhu, 2000b).

Web kullanım madenciliği, işbirlikçi filtrelemenin anonim kullanıcılar üzerinde etkili bir şekilde kullanılmasında da yardımcı olmaktadır (Mobasher ve diğerleri, 2001). Mobasher ve diğerleri (2001) yaptıkları çalışmada uygun veri dönüşümleri ve ön işlemlerle web kullanım madenciliğinin uygulanmasının işbirlikçi filtrelemenin etkinliğini önemli ölçüde artırabildiğini göstermiştir.

İşbirlikçi filtreleme gibi kişiselleştirmeye yönelik yaklaşımlarla birlikte kullanılan web kullanım madenciliği, öznel kullanıcı derecelendirmelerine, ölçeklenebilirlik, seyreklik ya da yüksek boyutluluk gibi tavsiye sistemlerinin yetersiz kaldığı durumlarda yardımcı olabilir. Web günlüklerinden işlemlerin kümelenmesi ve sayfa gösterimlerinin kümelenmesi ile oluşturulan bütünleştirilmiş kullanıcı profilleri, kullanım tabanlı tavsiye sistemlerinin oluşturulmasında kullanılabilmesi gibi, işbirlikçi filtreleme gibi yöntemlerin de doğruluk ve ölçeklenebilirliğini arttırmak için kullanılabilir (Cho, Kim ve Kim, 2002; Mobasher ve diğerleri, 2002).

Facca ve Lanzi (2005) hazırladıkları derlemede web kullanım madenciliğinin genel kullanım alanlarını kişileştirme, ön okuma (İng. pre-fetching) ve ön belleğe alma (İng. caching), tasarımı destekleme ve e-ticaret olarak dört kategori altında toplamıştır. Kişileştirmede, tavsiye sistemleri yaygın olarak kullanılmakta, kullanıcının ilgilenebileceği bağlantılar önerilmektedir.

Kişileştirme için kullanılan veri madenciliği yaklaşımlarının çoğu, işbirlikçi filtrelemenin uzantıları olarak görülebilir; bu yaklaşımlarda girdi olarak kullanıcılara ait geçmiş derecelendirme ya da gezinti profilleri dikkate alınarak, farklı algoritmalar yardımıyla kullanıcı profilleri oluşturulur. Bu kullanıcı profilleri, hedef kullanıcının profiliyle birlikte dikkate alınarak öneriler üretmek ya da kullanıcı davranışlarını tahmin etmek için kullanılabilir (Mobasher, 2007).

Nasraoui, Soliman, Saka, Badia ve Germain (2008) oluşturdukları web kullanım madenciliği kullanarak profil oluşturma yaklaşımını beş adımla özetlemektedir. Kullanıcı oturumları web günlük dosyalarından ayıklamak için ön işlemeye tabi tutulur. Ardından kullanıcı oturumları H-UNC (İng. Hierarchical Unsupervised Niche Clustering) algoritması ile kümelenir. Bu kümelerden kullanıcı profilleri oluşturulur ve ek kaynaklar yardımıyla bu

profiller zenginleştirilir. Farklı zaman aralıklarında farklı profil etkinliklerini izlenerek, profiller sürekli geliştirilir.

Mele (2013)'nin arama motoru başarımını arttırmak ve web sayfası önerilerini iyileştirmek için yaptığı çalışma; web sayfası tavsiyesi için işbirlikçi filtrelemeye dayanmakta, benzer ilgi alanlarına ve zevklere sahip kullanıcıları tanımlamak için tıklama bilgileri gibi içeriğe sahip kullanıcı gezinti günlüklerini kullanmaktadır.

Adeniyi, Wei ve Yongquan (2016) k en yakın komşu yöntemi kullanarak oluşturdukları tavsiye sisteminde, var olan kullanıcıların tıklama verilerini web kullanım madenciliği ile işleyip, hedef kullanıcıya daha hızlı bir şekilde bir öneri seti oluşturmak için bunu belirli kullanıcı gruplarıyla eşleştirmiştir. Bu sayede ölçeklendirme sorununu aşmayı hedeflemiştir.

Web kullanım madenciliği süreçleri, yöntemleri ve teknikleriyle ilgili daha detaylı bilgi için (Varnagar ve diğerleri, 2013) ve (Markov ve Larose, 2007) kaynaklarına başvurulabilir.

### **2.3. Kümeleme Yöntemleri ve İkili Kümeleme**

Veri madenciliğinin temel alanlarından biri olan kümeleme, bir dizi nesne göz önüne alındığında, bu nesnelerin benzerlik ölçülerine göre gruplandırılması olarak tanımlanabilir. Böylece kümedeki nesneler birbirleriyle benzer özelliklere sahip iken, diğer kümedeki nesnelere göre benzer değildir. Nesnelere daha az sayıda küme ile temsil etmek, bazı özelliklerinin kaybedilmesine neden olsa da basitleştirmeyi sağlar. Nesnelere ait oldukları kümeler ile modellenir. Kümeleme analizinin insanlar tarafından değil de kümeleme algoritması tarafından yapılması, nesnelere arasında daha önce bilinmeyen grupların keşfedilmesine yol açabilmesi nedeniyle yararlıdır (Han, Pei ve Kamber, 2011). Kümeleme, biyoloji, güvenlik, iş zekâsı, pazarlama, satış ve benzeri birçok alanda bir veri madenciliği aracı olarak kullanılmaktadır (Berkhin, 2006).

### 2.3.1. Kümeleme yöntemleri

Kümeleme algoritmaları, yöntemlerine göre kategorilere ayrılabilir. Bir algoritma birden fazla kategoriye ait özelliği gösterebilir.

Bölümlemeli yöntemler; bir nesnelere seti üzerinden, her bölümün bir kümeyi temsil ettiği nesne bölümleri oluşturur. Her bir grupta en az bir tane nesne bulunur. Nesnelere ve kümeler arasında herhangi bir hiyerarşik bağ bulunmaz. Bu yöntemlere dayalı algoritmaların çoğu mesafeye dayalı algoritmalarlardır. Verilen  $k$  bölüm sayısına göre, bir ilk bölümleme oluşturulur. Ardından nesnelere bir bölümden diğerine taşınarak yinelemeli bir yer değiştirme tekniği kullanılır. Aynı kümedeki nesnelere birbirlerine yakın ve diğer kümelerdekilere de uzak olması sağlandığında yineleme işlemi sonlanır (Han ve diğerleri, 2011; Xu ve Wunsch, 2005).

Hiyerarşik yöntemler; nesnelere seti üzerinde hiyerarşik bir ayrışma yaratır. Hiyerarşik yapılaşmanın nasıl oluştuğuna bağlı olarak yığınsal veya bölücü olarak bir sınıflandırma daha yapılabilir. Yığınsal yaklaşım her nesne ayrı bir grup oluşturarak gruplamaya başlar ve tüm gruplar birleştirilene kadar nesnelere veya grupları birbirine yaklaştırır. Bu yaklaşım alttan üste olarak da adlandırılmaktadır. Üstten alta, yani bölücü yaklaşım ise aynı kümedeki tüm nesnelere, ardışık yinelemelerle daha küçük kümelere bölünmesi, ardından her kümede yalnızca bir nesne kalması temeline dayanır (Han ve diğerleri, 2011; Murtagh, 1983).

Yoğunluğa dayalı yöntemler: Bölümlemeli yöntemlerinin çoğu nesnelere arası mesafeye bakmaları nedeniyle sadece küresel şekilli kümeler bulabilir ve rasgele şekilli kümeleri keşfetmekte güçlük çekerler. Yoğunluk kavramına dayalı olarak geliştirilen yöntemler komşuluklara ait yoğunluk bir miktar eşiği aştığı sürece kümeyi büyütür. Bu tür yöntemler gürültüyü ve uç değerleri filtrelemek için de kullanılabilir (Ester, Kriegel, Sander ve Xu, 1996).

Izgara bazlı yöntemler, adından da anlaşılacağı üzere nesne uzayını sınırlı sayıda hücrelere bölerek bir ızgara yapısı oluşturur. Tüm kümeleme işlemleri bu ızgara yapısı üzerinde gerçekleştirilir. Bu tür yöntemler, başarımları nesne sayısından ziyade ızgara yapısına bağlı



olduğundan yüksek boyutlu veriler üzerinde etkin bir çözüm sağlamaktadır (Hinneburg ve Keim, 1999).

Bu çok yaygın kullanılan yöntemlere ek olarak bazı ileri teknikler de söz konusudur.

Birçok kümeleme algoritması geliştirilmiş ve günümüz ihtiyaçları doğrultusunda geliştirilmeye devam edilmektedir. Bu algoritmaları bölümlene kriterleri, kullanılan benzerlik yöntemleri, nesne özellikleri gibi farklı açılardan daha birçok kategoriye ayırmak mümkündür (Xu ve Wunsch, 2005).

Geleneksel kümeleme yöntemleri bir nesneyi yalnızca bir kümeye atamaktadır, fakat bazı durumlarda nesnenin bir ya da birden fazla kümeye atanması istenebilir. Bulanık kümeleme ve olasılık tabanlı kümeleme yöntemleri buna olanak sağlamaktadır. Bir nesnenin farklı kümelere aitlik dereceleri hesaplanarak ileri işlemler yapılabilir.

Yüksek boyutlu veriler hem kümelemenin nasıl modelleneceği hem de böyle kümelerin nasıl aranacağı konusunda çeşitli zorluklar oluşturmaktadır. Alt uzay kümeleme yöntemleri, orijinal alanın alt uzaylarında kümeler arar (Agrawal, Gehrke, Gunopulos ve Raghavan, 1998). Boyut azaltma yöntemleri, daha düşük boyutlu yeni bir alan yaratır ve orada kümeler arar (Kaski, 1998).

Bu çalışmada hem yüksek boyutlu verilerle çalışmayı sağlayan hem de nesnelerin birden fazla kümede bulunmasına izin veren ikili kümeleme yöntemlerinden faydalanılmıştır.

### **2.3.2. İkili kümeleme**

İkili kümeleme diğer kümeleme yöntemlerinden farklı olarak nesnelere ve de niteliklerini eş bir şekilde göz önüne alarak veri matrisleri içinden farklı alt matrisler oluşturur. Geleneksel yöntemler nesnelerin nitelik değerlerine göre kümelerken, ikili kümeleme nesnelere kümelediği gibi niteliklerini de kümeleyerek, diğer yöntemlerin bulamayacağı benzersiz desenleri ortaya çıkarabilir. Bu tarza olan kümeleme algoritmaları ikili kümeleme yöntemleri altında değerlendirilebilir.

İkili kümeleme, gen anlatım analizinden (Cheng ve Church, 2000), işbirlikçi filtrelemeye (Fabricio, Ferreira ve Von Zuben, 2007) ve hatta çevrimiçi reklamcılık alanında (Ignatov, Kuznetsov ve Poelmans, 2012a) bile farklı uygulamalarda kendine yer bulmuş bir yöntem haline gelmiştir.

Bu bölümde öncelikle ikili küme çeşitleri, ardından ise ikili kümeleme yöntemleri üzerine değinilecektir.

### İkili küme çeşitleri

İkili kümelerin modellenmesinde ve bulunmasında ikili kümenin çeşidi önem kazanır.

$O = \{o_1, \dots, o_n\}$  nesnelar dizisi ve de  $F = \{f_1, \dots, f_m\}$  nitelikler dizisi olsun.  $E = [e_{ij}]$  nesnelar ve nitelikler dizilerinden oluşan bir veri matrisi olsun. Bu matriste  $i$  ve  $j$  değeri  $1 \leq i \leq n$  ve  $1 \leq j \leq m$  şartlarını sağlamaktadır. Nesnelar dizisinin elemanlarından oluşan  $I \subseteq N$  alt dizisi ile nitelikler dizisinin elemanlarından oluşan  $J \subseteq F$  alt dizisi üzerinden bir alt matris  $I \times J$  tanımlanmış olsun. Bu tanımlamalar çerçevesinde farklı ikili küme çeşitleri tanımlanmıştır.

- Sabit değeri sahip ikili küme: Bu tarz ikili kümelerdeki tüm değeri sahip bir değeri eşittir.  $I \times J$  alt matrisinde yer alan her  $e_{ij} = c$  olmalıdır, burada  $i \in I$  ve  $j \in J$ 'dir.
- Satırlarda sabit değeri sahip ikili küme: Her bir satırdaki değeri sahip sabit değeri sahip iken, farklı satırlar farklı değeri sahiptir.  $I \times J$  alt matrisinde yer alan her  $e_{ij} = c + \alpha_i$  (ya da  $e_{ij} = c \cdot \alpha_i$ ) olup, burada  $i \in I, j \in J$  ve  $\alpha_i$   $i$  satırı için düzeltme değeri dir.
- Sütunlarda sabit değeri sahip ikili küme: Her sütundaki değeri sahip sabit değeri sahip iken, farklı sütunlar farklı değeri sahiptir.  $I \times J$  alt matrisinde yer alan her  $e_{ij} = c + \beta_j$  (ya da  $e_{ij} = c \cdot \beta_j$ ) olup, burada  $i \in I, j \in J$  ve  $\beta_j$   $j$  sütunu için düzeltme değeri dir.
- Tutarlı değeri sahip ikili küme: Satırlardaki değeri sahip sütunlara göre tutarlı bir şekilde değeri sahipir, ya da tersi şekilde sütunlardaki değeri sahipir satırlara göre tutarlı bir

şekilde değişir.  $I \times J$  alt matrisinde yer alan her  $e_{ij} = c + \alpha_i + \beta_j$  (ya da  $e_{ij} = c \cdot \alpha_i \cdot \beta_j$ ) olup, burada  $i \in I, j \in J, \alpha_i$  i satırı için düzeltme değeri ve  $\beta_j$  j sütunu için düzeltme değeridir. Benzer bir tanım satırlar yerine sütunlardaki değerler için de yapılabilir.

- Tutarlı değişimlere sahip ikili küme: Nesnelere ya da niteliklerin değerleri arasında tutarlı bir artış ya da azalış vardır.  $I \times J$  alt matrisinde yer alan her değer için  $(e_{i_1 j_1} - e_{i_1 j_2})(e_{i_2 j_1} - e_{i_2 j_2}) \geq 0$  olup, burada  $i_1, i_2 \in I$  ve  $j_1, j_2 \in J$ 'dir. Benzer bir tanım satırlar yerine sütunlardaki değişim için de yapılabilir.

Şekil 2.2'de farklı ikili küme çeşitlerine ait örnekler gösterilmektedir.

<table border="1" style="border-collapse: collapse;"> <tr><td>2</td><td>2</td><td>2</td></tr> <tr><td>2</td><td>2</td><td>2</td></tr> <tr><td>2</td><td>2</td><td>2</td></tr> </table>	2	2	2	2	2	2	2	2	2	<table border="1" style="border-collapse: collapse;"> <tr><td>2</td><td>2</td><td>2</td></tr> <tr><td>3</td><td>3</td><td>3</td></tr> <tr><td>4</td><td>4</td><td>4</td></tr> </table>	2	2	2	3	3	3	4	4	4	<table border="1" style="border-collapse: collapse;"> <tr><td>2</td><td>3</td><td>4</td></tr> <tr><td>2</td><td>3</td><td>4</td></tr> <tr><td>2</td><td>3</td><td>4</td></tr> </table>	2	3	4	2	3	4	2	3	4	<table border="1" style="border-collapse: collapse;"> <tr><td>2</td><td>4</td><td>6</td></tr> <tr><td>3</td><td>6</td><td>9</td></tr> <tr><td>4</td><td>8</td><td>12</td></tr> </table>	2	4	6	3	6	9	4	8	12	<table border="1" style="border-collapse: collapse;"> <tr><td>6</td><td>4</td><td>5</td></tr> <tr><td>2</td><td>1</td><td>4</td></tr> <tr><td>4</td><td>3</td><td>6</td></tr> </table>	6	4	5	2	1	4	4	3	6
2	2	2																																															
2	2	2																																															
2	2	2																																															
2	2	2																																															
3	3	3																																															
4	4	4																																															
2	3	4																																															
2	3	4																																															
2	3	4																																															
2	4	6																																															
3	6	9																																															
4	8	12																																															
6	4	5																																															
2	1	4																																															
4	3	6																																															
Sabit değerlere sahip	Satırlarda sabit değerlere sahip	Sütunlarda sabit değerlere sahip	Tutarlı değerlere sahip	Tutarlı değişimlere sahip																																													

Şekil 2.2. İkili küme çeşitleri

### İkili kümeleme yöntemleri

İkili kümeleme yöntemlerini, buldukları ikili küme çeşidine, sayısına (Madeira ve Oliveira, 2004) ya da algoritmaların uyguladığı yöntemlere göre (Bozdağ, Kumar ve Catalyurek, 2010; Busygin, Prokopyev ve Pardalos, 2008; Tanay, Sharan ve Shamir, 2005) sınıflandıran çeşitli çalışmalar yapılmıştır. Burada yalnızca yaygın olarak kullanılan ve de alanında öncü olan algoritmalara yer verilmiştir.

Cheng ve Church tarafından genlerin bir alt kümesinin (nesnelere) ve koşulların bir alt kümesinin (nitelikler) tutarlılığını yakalamak için, ikili küme kavramı ortaya atılmıştır (Cheng ve Church, 2000). Fakat ilk ikili kümeleme algoritması olan doğrudan kümeleme (Hartigan, 1972), simültane olarak satırları ve sütunları kümelemek için bir model olarak ortaya çıkmıştır. Hartigan (1972), o anda mevcut olan her ikili kümeyi, her yinelemede iki

parçaya bölerek iki ikili küme tanımlar. Bu yaklaşım aynı anda bir grup ikili kümeyi keşfeder. Veri içindeki ikili kümeler en düşük tutarsızlık ile bulunur.

$\delta$ -bicluster algoritması (Cheng ve Church, 2000) bir seferde bir ikili küme keşfeder. İkili kümeyi rastgele sayılarla saklayarak, bir sonraki yinelemede yeni bir ikili küme keşfeder. Sabit değerlere sahip, satırlarda ya da sütunlarda sabit değerlere sahip ve de tutarlı değerlere sahip ikili kümeleri bulmada başarılıdır. Cheng ve Church algoritması olarak yazarlarının adıyla da isimlendirilen algoritma, ikili küme algoritmaları üzerine olan çalışmalarını tekrar canlandırmış, özellikle biyoteknoloji alanında kendine önemli bir yer edinmiştir.

Çift iki yönlü kümeleme (İng. Coupled two way clustering) (Getz, Levine ve Domany, 2000) bir ikili küme elde etmek için hem satırlarda hem sütunlarda ayrı ayrı bir kümeleme yapar. Her yinelemede istikrarlı bir sıra ve sütun küme oluşturan hiyerarşik bir kümeleme algoritması kullanır ve sonuç olarak bir kerede bir grup ikili küme keşfeder.

Örme model (İng. Plaid model) de her yinelemede tek bir ikili küme keşfeden algoritmalarından birisidir (Lazzeroni ve Owen, 2002). Tutarlı değerlere sahip ikili kümeler bulan bu algoritma, veri matrisindeki bir nesnenin değerini katman adı verdiği terimlerin toplamı olarak görür. Bu modele göre veri matrisi ikili kümelere karşılık gelen katmanların doğrusal bir fonksiyonu olarak tanımlanır.

SAMBA (İng. Statistical-Algorithmic Method for Bicluster Analysis) algoritması (Tanay, Sharan ve Shamir, 2002) veri matrisini nesnelere ve niteliklerin toplu olarak köşe kümesini oluşturduğu iki parçalı bir çizgeye (İng. graph) dönüştürür. İkili kümeler arasında keskin tutarsızlıkların olduğu, tutarlı değişimlere sahip ikili kümeleri eş zamanlı olarak bulur.

Yinelemeli imza algoritması (İng. Iterative Signature Algorithm, ISA) (Bergmann, Ihmels ve Barkai, 2003) bir başlangıç gen grubundan yola çıkılarak, tüm örnekleri bu gen grubuna göre puanlar ve puanı önceden tanımlanmış bir eşığı aştığı örnekleri seçer. Aynı şekilde, seçilen örneklerle ilgili olarak tüm genler puanlanır ve başka bir kullanıcı tanımlı eşığe dayanan yeni bir gen seti seçilir. Tüm süreç, gen seti ve örnek seti birleşinceye kadar tekrarlanır. Bu algoritma birden fazla ikili küme bulur.

Sırası korunmuş bir alt matris (İng. Order Preserved Submatrix, OPSM), değerleri sütunların bir alt kümesi boyunca lineer bir sıraya değişen satırlardan oluşmaktadır. OPSM algoritması bu tarz alt matrisleri bulmaktadır. Algoritma tek seferde tutarlı değişimlere sahip bir tane ikili kümeyi bulur (Ben-Dor, Chor, Karp ve Yakhini, 2003).

Tutarlı değişimlere sahip ikili kümeler xMotif algoritması ile de bulunabilir. Bu algoritma, korunmuş gen anlatım motifleri (ikili kümeler, xMotif) bulmayı hedefler, bunun için sütunlar seti üzerinden sabit değerlere sahip satırları arar. Bir gen durumunun bir dizi ifade değeri olduğunu düşünülür ve belirli bir sayıda durum bulunduğunu varsayılır. Bu durumlar, yukarı ve aşağı yönlü olarak iki tane olabilir. Veri matrisinin birden çok ikili küme içerebileceği düşünüldüğünden en yüksek satır sayısını içeren ikili kümenin bulunması hedeflenir (Murali ve Kasif, 2003).

Tutarlı değerlere sahip ikili kümeleri eş zamanlı bulan bir diğer algoritma ise spektral ikili kümelemedir (İng. Spectral biclustering). Veri matrisinden ikili kümeleri bulmak için, SVD (İng. Singular Value Decomposition) yöntemini kullanarak elde edilen öz değerleri (İng. Eigenvalue) ve öz vektörleri (İng. Eigenvector) kullanır. Bu normalleştirme işlemi sonrası veri matrisi dama tahtası yapısına sahip olur (Kluger, Basri, Chang ve Gerstein, 2003; Madeira ve Oliveira, 2004).

BiMax ikili kümeleme algoritması (İng. Binary inclusion-maximum biclustering) ikili bir veri matrisi içinden en büyük ikili kümeyi bulmak için kullanılır. Bu algoritma ikili veri matrisi içinde, “0” ve “1”leri ayırarak, “1” değerine sahip en kapsamlı alt matrisleri bulur ve bu alt matrisler, kendilerinden daha büyük bir başka alt matrisin parçası olamaz (Prelić ve diğerleri, 2006).

İkili kümeleme yöntemleri ile tavsiye sistemleri (Ignatov, Poelmans ve Zaharchuk, 2012b; Symeonidis, Nanopoulos ve Manolopoulos, 2008a; 2009), işbirlikçi filtreleme (Alqadah, Reddy, Hu ve Alqadah, 2015; Fabricio ve diğerleri, 2007; Symeonidis, Nanopoulos, Papadopoulos ve Manolopoulos, 2008b) ve çevrimiçi reklamcılık (Ignatov ve diğerleri, 2012a) alanlarında da yapılmış çalışmalar mevcuttur.

Symeonidis ve diğeri (2008b) işbirlikçi filtrelemede kullanıcıları ve öğeleri aynı anda gruplamak için en yakın ikili küme yöntemi kullanmıştır. BiMax ve xMotif algoritmaları ile geliştirdikleri en yakın ikili kümeleme yöntemini birlikte kullanarak etkin sonuçlar elde etmişlerdir.

Fabricio ve diğeri (2007) sadece kullanıcılar ya da öğeler arasındaki benzerlikleri göz önünde bulundurmak yerine, kullanıcılar ile öğeler arasındaki mevcut ikiliği hesaba katacak şekilde bir ikili kümelemeli işbirlikçi filtreleme yöntemi geliştirmiş ve literatürdeki diğer işbirlikçi filtreleme yöntemlerine göre daha iyi sonuçlar elde edildiğini bildirmişlerdir.

Alqadah ve diğeri (2015) hassas ve bölgesel sonuçlar elde etmek için ikili küme komşuluğu tabanlı işbirlikçi filtreleme yöntemi kullanarak tavsiye sistemi geliştirmişlerdir. Geliştirilen yöntem özellikle dağınık veri üzerinde, genel kabul görmüş algoritmalara göre daha iyi sonuç vermiştir.

Symeonidis ve diğeri (2009) açıklamalı tavsiye sistemleri üzerine yaptıkları çalışmada kullanıcıları gruplamak için ikili kümeleme yöntemi uygulamıştır. Bireysel kullanıcılar yerine grupların kullanılması, çıkartılan özelliklerin bütün toplulukların tercihini yansıtmasına ve işbirlikçi özelliklerin daha geniş kullanıcı tercihlerini kapsaması nedeniyle daha iyi açıklamalar oluşturulmasına neden olmaktadır.

Lise öğrencilerinin lisans eğitimi yönelimi için geliştirilen bir tavsiye sisteminde de (Ignatov ve diğeri, 2012b) ikili kümeleme yöntemi uygulanmış ve eğitim alanında da ikili kümelemenin başarılı bir şekilde uygulanabileceği gösterilmiştir.

Çevrimiçi reklamcılık için kavram temelli ikili kümeleme üzerine yapılan bir çalışmada (Ignatov ve diğeri, 2012a) ikili veri üzerinden yoğun-yaklaşık (İng. Dense approximate) ikili kümeler çıkaran bir algoritma geliştirmişlerdir. Sonuçlar karşılaştırdıkları klasik yöntemlere göre daha yüksek başarımlar göstermiştir.

## 2.4. Tavsiye Sistemleri

Bir tavsiye sistemi, girdi olarak kullandığı verileri, kullanıcılarının gelecekte beğeneceği ve ilgi duyacağı öğeleri tahmin etmek için kullanır (Lü ve diğerleri, 2012). Kullanıcı geçmişleri kullanılan verilerin önemli bir kısmını oluşturmaktadır.

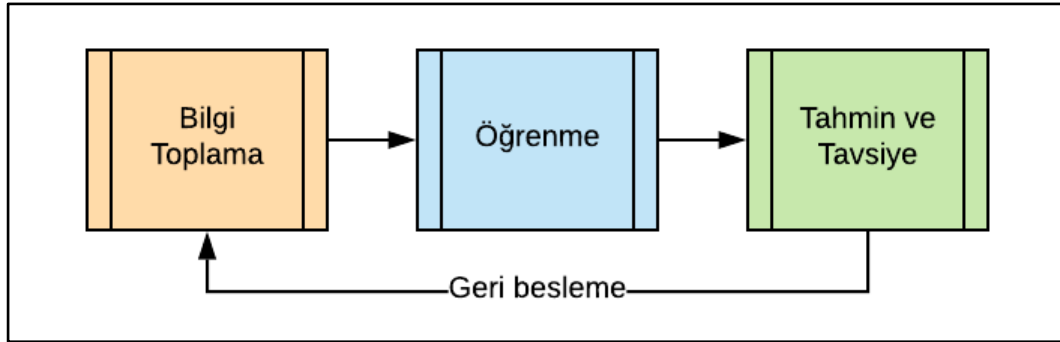
Bir tavsiye sisteminin görevi, kişiselleştirilmiş, önerilen öğeler listesini kullanıcılarına ulaştırmaktır. Bu iki yolla yapılabilir:

- *Tahmin (Derecelendirme Tahmini)*: Belirli bir öğe için öngörülen derecelendirme puanını hesaplanır. Öğeleri önermek için, tavsiye sisteminin birkaç alternatif sunmaya da hazır olması gerekir. Derecelendirme tahmininin zorluğu, nadiren derecelendirilmiş olsalar bile istenen herhangi bir öğe hakkında bir derecelendirmede bulunması gerekliliğidir (Schafer, Frankowski, Herlocker ve Sen, 2007). Bu tür görevlerin başarımı literatürde yaygınca kullanılan kare ortalamalarının karekökü hatası (RMSE) ile değerlendirilmektedir (Steck, 2013).
- *Tavsiye (Sıralama)*: Hedef kullanıcıya, ilgisini çekebilecek bir öğe listesi gösterir. Genellikle bu işlem, kullanıcının öğeyi nasıl derecelendireceğini tahmin etmek ve ardından öğeleri bu öngörülen dereceye göre sıralamak olarak tanımlanır. Ancak, bazı başarılı tavsiye algoritmaları, öngörülen derecelendirme değerlerini hiç hesaplamaz (Schafer ve diğerleri, 2007). Sıralama, her kullanıcı için, öğe kümesindeki tüm öğeler arasından öğelerin küçük bir sayısını (N) seçmek gerektiğinde uygulanabilecek yararlı bir yaklaşımdır. Bu tür görevlerin başarımı ise hassasiyet, hatırlama ve alıcı işletim karakteristiği (ROC, İng. Receiver Operating Characteristics) eğrisi altındaki alan (AUC, İng. Area under the ROC curve) gibi ölçütlerle değerlendirilmektedir (Steck, 2013).

Daha sonra hedef kullanıcıya sunulacak olan tavsiye listesi, sıralamada en yüksek yere ya da öngörülen en yüksek derecelendirme puanlarına sahip olan öğeler seçilerek oluşturulur. (Lü ve diğerleri, 2012; Steck, 2013).

### 2.4.1. Tavsiye süreci

Isinkaye, Folajimi ve Ojokoh (2015) tavsiye sürecini üç aşamalı bir döngü olarak tanımlamaktadır. Şekil 2.3 bilgi toplama, öğrenme ve son olarak tahmin ve tavsiye evrelerinden oluşan basit bir tavsiye sürecini göstermektedir.



Şekil 2.3. Tavsiye süreci

#### Bilgi toplama evresi

Bu aşamada tahmin görevinde kullanılmak üzere bir kullanıcı profili ya da modeli oluşturmak için gerekli bilgiler toplanır. Bu bilgiler kullanıcının özellikleri, davranışları ve de eriştiği kaynakların içerikleri olabilir. Kullanıcı profili ya da modeli iyi yapılandırılmadan bir tavsiye sistemi düzgünce çalışmaz. Sistem başlangıçtan itibaren makul bir öneride bulunmak için kullanıcıdan mümkün olduğunca bilgi edinmelidir (Isinkaye ve diğerleri, 2015).

Herhangi bir tavsiye sisteminin başarısı büyük oranda kullanıcının mevcut ilgilerini temsil etme kabiliyetine bağlıdır. Tavsiye sisteminin herhangi bir tahmin yöntemi ile doğru tavsiyelerde bulunabilmesi, kullanıcı profilleri ya da modelleri olmadan oldukça zordur (Isinkaye ve diğerleri, 2015).

Kullanıcı profilleri ve modelleri doğrudan, dolaylı ya da bu ikisinin karışımı olan melez geri besleme yöntemleri ile elde edilen bilgiler ile hazırlanabilir. *Kullanıcı Profilleri ve Web Madenciliği* bölümü altında bu konu hakkında detaylı bilgi verilmektedir.



### Öğrenme evresi

Bilgi toplama aşamasında toplanan bilgilerden ve geri bildirimlerden kullanıcının özelliklerini filtrelemek ve ortaya çıkarmak için bir öğrenme algoritması uygulanır (Isinkaye ve diğerleri, 2015).

### Tahmin ve tavsiye evresi

Bu son aşamada ise kullanıcının hangi öğeleri beğeneceği tahmin edilir ya da kullanıcıya tavsiyede bulunulur. Tahmin görevi bilgi toplama aşamasında elde edilmiş veri kümesi üzerine yapılabilirken, tavsiye görevi ise kullanıcının hareketleri gözlemlenerek tetiklenebilir (Isinkaye ve diğerleri, 2015).

#### **2.4.2. Tavsiye yöntemleri**

Tavsiye sistemlerinin sınıflandırılmasında farklı yaklaşımlar bulunsa da, genel olarak üç ana yöntem üzerinden sınıflandırma yapılmaktadır: İçerik tabanlı yöntemler, işbirlikçi filtreleme ve melez yöntemler (Adomavicius ve Tuzhilin, 2005; Lü ve diğerleri, 2012). Bir diğer genel kabul görmüş sınıflandırma ise bellek tabanlı ve model tabanlı yöntemlerdir (Adomavicius ve Kwon, 2007). Bellek tabanlı yöntemler doğrudan önceki kullanıcı faaliyetlerini dikkate alarak, tavsiye talebi esnasında önerileri hesaplamaktadır. Buna karşılık model tabanlı yöntemler ise kullanıcı faaliyetlerinden topladığı verileri kullanarak önceden bir tahmin modelini oluşturmakta, ardından bu modeli tavsiyede bulunmak için kullanmaktadır (Adomavicius ve Kwon, 2007).

#### İçerik tabanlı yöntemler

İçerik tabanlı yöntemler, bir hedef kullanıcının önceden tercih etmiş olduğu öğelerin içeriğine benzer içerikte olan diğer öğeleri bulur ve önerir (Lü ve diğerleri, 2012). Öğelerin benzerliği, karşılaştırılan öğelerle ilişkili özelliklere dayalı olarak hesaplanır. Örneğin, bir kullanıcı komedi tarzına ait bir filmi pozitif olarak derecelendirdiyse, sistem bu türden diğer filmleri önermeyi öğrenebilir.

Vakaya dayalı mantık yürütme, bir kullanıcı belirli bir öğeyi severse, muhtemelen benzer öğeleri de seveceğini varsayar. Yeni ama benzer öğeler önerir. Öznitelik tabanlı teknikler ise öğe özelliklerinin kullanıcı profiliyle eşleşmesine durumunda öğeleri önerir (Ricci, Rokach ve Shapira, 2011).

### İşbirlikçi filtreleme

Resnick, Iacovou, Suchak, Bergstrom ve Riedl (1994) işbirlikçi filtrelemeyi “İnsanların diğer insanların görüşlerine dayalı seçimler yapmasına yardımcı olur” şeklinde tanımlamıştır. Günümüzde halen işbirlikçi filtreleme, tavsiye sistemlerinde kullanılan en yaygın tavsiye yöntemlerinden birisidir.

İşbirlikçi filtreleme, kullanıcıların bir öğe kümesi hakkında derecelendirme yapmalarını sağlar ve sistemde yeterli veri toplandığında; sistem her kullanıcıya önerilerde bulunabilir (Bobadilla, Ortega, Hernando ve Gutiérrez, 2013). Kullanıcı derecelendirmeleri, kullanıcıların faaliyet ve davranışları üzerinden veri madenciliği teknikleri kullanılmasıyla dolaylı olarak toplanabileceği gibi (Anand ve Mobasher, 2003; Herlocker, Konstan ve Riedl, 2000; Nasraoui ve diğerleri, 2008), diğer sistemlerden veya alanlardan alınabilir (Burke, 2002; Li, 2011).

İşbirlikçi filtreleme algoritmaları iki ana kategoriye ayrılabilir: bellek tabanlı ve model tabanlı (Sarwar, Karypis, Konstan ve Riedl, 2001). Kullanıcı ve öğe tabanlı filtrelemenin dahil olduğu bellek tabanlı işbirlikçi filtreleme yöntemlerinde, derecelendirme kümesinin tamamı önerilerde bulunmak veya öğelerin sıralamasını tahmin etmek için kullanılır. Öte yandan, model tabanlı işbirlikçi filtreleme yöntemlerinde, önce matris ayrıştırma gibi çeşitli makine öğrenme algoritmaları tarafından gerçekleştirilen bir model oluşturularak önerilerde bulunulur. Hem bellek tabanlı hem de model tabanlı algoritmalar kullanan melez algoritmalar da vardır (Pennock, Horvitz, Lawrence ve Giles, 2000). İşbirlikçi filtreleme algoritmaları, önerilerin etkinliğini arttırmak için diğer tavsiye teknikleriyle de birlikte kullanılabilir (Burke, 2002).

Kullanıcı tabanlı işbirlikçi filtreleme (UBCF, İng. User Based Collaborative Filtering) algoritmaları, önerilerde bulunmak için benzer kullanıcıların öğe derecelendirmelerini toplar

(Schafer ve diğeri, 2007). Aynı öğeyi benzer şekilde derecelendiren kullanıcılar muhtemelen aynı tada (İng. taste) sahiptir. Bu varsayımına dayanarak, bu teknik, benzer kullanıcılar tarafından zaten derecelendirilmiş hedef kullanıcı tarafından henüz görünmemiş öğeleri önerir (Ricci ve diğeri, 2011).

Öğeye tabanlı işbirlikçi filtreleme (IBCF, İng. User Based Collaborative Filtering) algoritmaları benzer öğeleri bulur ve daha sonra hedef kullanıcıya bu öğeleri önerir (Sarwar ve diğeri, 2001; Schafer ve diğeri, 2007). Bu teknik, benzer şekilde derecelendirilen öğelerin muhtemelen benzer olduğunu farz eder. En yüksek bağıntıya sahip öğeleri önerir.

Ağırlıklandırılmış düzenli matris ayrıştırma (WRMF, İng. Weighted Regularized Matrix Factorization) (Hu, Koren ve Volinsky, 2008; Pan ve diğeri, 2008) yöntemi örtük geri besleme veri kümeleri için sıkça kullanılan bir matris ayrıştırma yöntemidir. Üst-N tavsiyeyi bir sıralama problemi olarak ele alan Bayes kişileştirilmiş sıralama (BPR, İng. Bayesian personalized ranking) (Rendle, Freudenthaler, Gantner ve Schmidt-Thieme, 2009) yöntemi Bayes analizinden elde edilen maksimum postör tahmincisidir. En yakın komşu (BPRKNN) ya da matris ayrıştırma (BPRMF) gibi diğer yöntemlerle birleştirilebilir. Seyrek Doğrusal Yöntemler (SLIM, İng. Sparse Linear Methods) ise Üst-N tavsiye sistemleri için düzenli bir  $l_1$ -norm ve  $l_2$ -norm optimizasyonuna dayanan matris ayrıştırma tabanlı tavsiye yaklaşımlarıdır (Ning ve Karypis, 2011).

### Melez ve diğer yöntemler

Melez (İng. Hybrid) tavsiye yöntemleri iki ya da daha fazla tavsiye yönetimini birleştirerek daha yüksek başarımlar elde etmeye yönelik olarak hazırlanan yöntemlerdir. Böylelikle her yöntemin zayıflıkları diğer yöntemlerin üstünlükleri ile en düşük seviyeye çekilmeye çalışılır. Melez yöntemlerin en önemli uygulaması, işbirlikçi filtreleme ve içerik tabanlı yöntemleri birleştirerek soğuk başlangıç sorununu çözmeye çalışmasıdır (Lü ve diğeri, 2012).

Burke (2002) ve Isinkaye ve diğeri (2015) melez tavsiye yöntemlerini ağırlıklı, değişimli, karışık, kademeli, özellik birleştirmeli, özellik aktarmalı ve meta katmanlı olarak sınıflandırmıştır.

- Ağırlıklı melez tavsiye yöntemi, tüm tavsiye yöntemleriyle hesaplanmış tavsiyeleri barındırır. Daha sonra her birine farklı ağırlıklar verilerek genel bir tahmin derecesi elde edilir. Bu türdeki sistemler belirsizliğin çok olduğu durumlarda kolayca sonuçlara göre uyarlanarak yeniden ayarlanabilir
- Değişimli melez tavsiye yöntemi ise mevcut duruma göre yöntemler arasından bir seçim yaparak en uygun tavsiyeyi bulmaktadır. Bu yöntemin dezavantajı ise değişim mekanizmasının varlığının sistemi daha karışık bir hale getirebilmesidir.
- Karışık melez tavsiye yöntemi çok sayıda önerinin verilebileceği durumlarda ön plana çıkar. Her tavsiye yönteminden elde edilen sonuçlar karışık bir şekilde sunulmaktadır.
- Kademeli melez tavsiye yöntemi, farklı öğeler arasında bir tercih sırası oluşturulmasında yinelemeli bir arttırma işlemi uygular. Bir tekniğin önerileri, başka bir tavsiye tekniğiyle iyileştirilmiştir.
- Diğer melez yöntemlerde ise kullanılan yöntemler arası oluşan özellikler birbirlerine birleştirilmekte, aktarılmakta ya da ara katmanlar oluşturularak diğer modellere girdi olarak sağlanmaktadır.

Bir diğer tavsiye yöntemi ise alanlar arası (İng. cross domain) tavsiye sistemleridir. Melez yöntemlerden farklı olarak, alanlar arası tavsiye sistemleri farklı alanlardaki verileri ya da tavsiyeleri birlikte ya da ayrı ayrı değerlendirerek çalışmaktadırlar. Alanlar arasında içerik tabanlı bir benzerliğin ya da işbirlikçi filtrelemeye yönelik bir uygulamanın var olması durumunda çeşitli yaklaşımlarda bulunulabilir (Fernández-Tobías, Cantador, Kaminskas ve Ricci, 2012). Bu yaklaşımlardan en yaygın olanları; kullanıcı derecelendirmesini tek birçok alanlı derecelendirme matrisine toplama, birden fazla alana ait kullanıcı tercihleri için ortak bir gösterimi kullanma, kullanıcı tercihlerini çoklu alan grafiği ile bağlama ve kullanıcı tercihlerini alan bağımsız özelliklere eşlemedir (Fernández-Tobías ve diğerleri, 2012).

### 2.4.3. Tavsiye sistemlerinin karşılaştığı zorluklar

Tavsiye sistemleri kullanılan yöntemler birbirlerine karşı bazı üstünlüklere sahip iken, hemen hemen tüm yöntemler farklı zorluklar ile de karşılaşmaktadırlar. Bu zorluklar, kullanılan yöntemin başarımını düşürebileceği gibi, kullanılmasını bile engelleyebilir. Lü ve

diğerleri (2012) tavsiye sistemlerinin karşılaştığı ana zorlukları sekiz başlık altında sıralamaktadır.

- *Seyreklik*, tavsiye sistemlerinin başarımını düşüren en büyük zorluklardan biridir. Bir veri kümesinde çoğu zaman çok sayıda öge bulunur, ancak kullanıcıların derecelendirmeleri yetersizdir, bu nedenle benzer kullanıcıları bulma olasılığı düşüktür (Lü ve diğerleri, 2012). İşbirlikçi filtreleme için transfer öğrenmesini<sup>1</sup> (Li, Yang ve Xue, 2009) veya alanları arası tavsiyeleri (Fernández-Tobías ve diğerleri, 2012; Gao ve diğerleri, 2013) kullanarak seyrekliği azaltmak için yapılan farklı araştırmalar bulunmaktadır.
- *Ölçeklenebilirlik*: Veri kümeleri genellikle seyrek olmakla birlikte, büyük siteler için milyonlarca kullanıcı ve öge içerirler. Dolayısıyla, yöntemlerin paralel çalışabilir, az kaynak tüketen ve de düşük maliyetli olması önemlidir (Lü ve diğerleri, 2012).
- Bir diğer zorluk olan *soğuk başlangıç* problemi, tavsiye sisteminin *yeni kullanıcılar* için etkili bir şekilde tahminde ya da öneride bulunamamasına yol açar. *Yeni öğeler* ve hatta *yeni alanlar* yetersiz verilerden benzer şekilde etkilenir (Schafer ve diğerleri, 2007). Bu sorunun çözümü melez yöntemler oluşturmak (Adomavicius ve Tuzhilin, 2005), kullanıcıları farklı sitelerde takip etmek (Lü ve diğerleri, 2012) ve alanlar arası tavsiye sistemlerinin kullanılmasıdır (Fernández-Tobías ve diğerleri, 2012).
- *Doğruluk*, bir kullanıcıya verilen tavsiyelerin kabul görmesidir ve bu en popüler ve yüksek derecelendirilen öğelerin önerilmesiyle etkin bir şekilde sağlanabilir. Bu tür öğeler kullanıcılar gözünde çok değerli değildir (*Popüler Tat*) ve kullanıcılar aslında sistemin *Farklılık* yaratarak kendilerinin kolayca bulamayacağı öğeleri önermesini beklerler (Lü ve diğerleri, 2012).
- *Zaman etkisi*, çoğu tavsiye yönteminin zaman ölçeğini göz ardı etmesinden kaynaklanır. Kullanıcı profillerinin, kullanıcının kısa ve uzun dönem ilgileri üzerinden çıkaran yaklaşımlar (Ahmed, Low, Aly, Josifovski ve Smola, 2011; Koşir, Kononenko ve Bosnić, 2014) zaman etkisini de göz önüne bu sorunu çözmeye yöneliktir.

---

<sup>1</sup> Transfer öğrenmesi, bir problemi çözerken elde edilen bilgiyi saklayarak, daha sonra farklı ama ilgili bir probleme uygulamak üzerine odaklanan bir makine öğrenimi yöntemidir.

- *Kullanıcı arayüzü* ve de *şeffaflık* tavsiyelerin kullanıcılar tarafından dikkate alınmamasına neden olan zorluklardan birisidir. Kullanıcılar öğelerin neden kendilerine önerildiği açıklandığında tavsiyeleri kabul etmektedirler (Lü ve diğerleri, 2012). Bu problemler, e-ticaret sitelerinde, “Bu ürüne bakanlar, bunlara da baktılar” ya da “Bu ürünü satın alanlar yanında şunları da aldı” gibi açıklamalar ile çözülmeye çalışılmaktadır.
- E-ticaret sitelerinde yer alan tavsiye sistemlerinin karşılaştığı bir diğer sorun ise *Saldırılarına karşı açıklıktır*. Bu, tavsiye sisteminin farklı saldırı yöntemleri aracılığıyla (Ricci ve diğerleri, 2011) bir öğeyi daha yüksek ya da düşük derecelendirilmesine neden olmaktadır.
- *Tavsiyelerin değerlendirilmesi* farklı değerlendirme ölçütleri kullanılarak yapılmakta olsa da hangisinin en uygun olduğu farklı durumlarda ve görevlerde farklılık göstermektedir. Tavsiye algoritmalarının farklı algoritmalar ile kıyaslanması da her algoritmanın aynı görev için tasarlanmaması nedeniyle yine sıkıntılıdır. Bir diğer sorun ise değerlendirmelerin genellikle çevrimdışı olarak yapılması ve kullanıcıların gerçek davranışlarının sadece çevrimiçi değerlendirmeler ile ortaya çıkabileceğidir (Lü ve diğerleri, 2012).

Bu çalışmada kullanılan veri kümeleri çok seyrek olduğu için önerilen yöntem, alanlar arası tavsiye yaklaşımı ile güçlendirilmektedir. Alanlar arası tavsiye sisteminin bir görevi, hedef alandaki öğelere ilişkin önerilerin kalitesini artırmak için kaynak alandaki kullanıcılar ve öğeler hakkında bilgi edinmektir (Fernández-Tobías ve diğerleri, 2012). Benzer şekilde, bu çalışmada önerilen yöntem, kullanıcı profillerini kaynak alanda oluştururken, önerileri hedef alanda yapmaktadır.

Çizelge 2.1 işbirlikçi filtreleme ve içerik tabanlı yöntemlerin artı ve eksi yönlerini kıyaslamakta, aynı zamanda karşılaştıkları zorlukları da listelemektedir.

Çizelge 2.1. İşbirlikçi filtreleme ve içerik tabanlı yöntemlerinin kıyaslanması (Manouselis, Drachler, Vuorikari, Hummel ve Koper, 2011)

Yöntem	Artıları	Eksileri
<b>İşbirlikçi filtreleme yöntemleri</b>		
Kullanıcı tabanlı işbirlikçi filtreleme	<ul style="list-style-type: none"> <li>✓ Farklılık</li> <li>✓ İçerik analizine ihtiyaç duymaz</li> <li>✓ Alandan bağımsızdır</li> <li>✓ Zaman geçtikçe kalitesi yükselir</li> </ul>	<ul style="list-style-type: none"> <li>- Yeni kullanıcı sorunu</li> <li>- Yeni öge sorunu</li> <li>- Popüler tat</li> <li>- Ölçeklenebilirlik</li> <li>- Seyreklik</li> <li>- Soğuk başlangıç</li> </ul>
Öge tabanlı işbirlikçi filtreleme	<ul style="list-style-type: none"> <li>✓ Farklılık</li> <li>✓ İçerik analizine ihtiyaç duymaz</li> <li>✓ Alandan bağımsızdır</li> <li>✓ Zaman geçtikçe kalitesi yükselir</li> </ul>	<ul style="list-style-type: none"> <li>- Yeni öge sorunu</li> <li>- Popüler tat</li> <li>- Seyreklik</li> <li>- Soğuk başlangıç</li> </ul>
<b>İçerik tabanlı yöntemler</b>		
Vakaya mantık yürütme	<ul style="list-style-type: none"> <li>✓ İçerik analizine ihtiyaç duymaz</li> <li>✓ Alandan bağımsızdır</li> <li>✓ Zaman geçtikçe kalitesi yükselir</li> </ul>	<ul style="list-style-type: none"> <li>- Yeni kullanıcı sorunu</li> <li>- Seyreklik</li> <li>- Soğuk başlangıç</li> </ul>
Öznitelik teknikler	<ul style="list-style-type: none"> <li>✓ Soğuk başlangıç sorunu yoktur</li> <li>✓ Yeni kullanıcı ve öge sorunu yoktur</li> <li>✓ Tercih değişikliklerine karşı duyarlıdır</li> <li>✓ Öğeden bağımsız özniteliklerle uyumludur</li> <li>✓ Kullanıcıdan öğelere eşleme yapabilir</li> </ul>	<ul style="list-style-type: none"> <li>- Aşırı uzmanlaşma</li> <li>- Öğrenmez</li> <li>- Sadece kategoriler üzerinden çalışmaktadır</li> <li>- Aşırı uzmanlaşma</li> <li>- Sürekli bakım gerektirir</li> </ul>

## 2.5. Birleştirme İşlemleri ve OWA Birleştirme Operatörleri

Birçok numerik değeri birleştirerek tek bir tane ile temsil etme işlemine birleştirme ve bu işlemi yapan fonksiyona ise birleştirme fonksiyonu adı verilmektedir. Uygulamalı matematikten bilgisayar bilimlerine, ekonomiden çok kriterli karar vermeye kadar birçok alanda kendine yer edinen bu işlem değişik yöntemlerle yapılabilmektedir (Grabisch, Marichal, Mesiar ve Pap, 2009).

Tercihlerin, kriterlerin veya benzerliklerin bir araya getirilerek birleştirilmesi, tavsiye sistemlerinde çeşitli aşamalarda gerçekleşir. Genellikle birleştirme işlemleri toplama, aritmetik ortalama veya maksimum, minimum operatörleri kullanılarak yapılır. Daha uygun tavsiyelerin bulunmasına yol açabilecek birçok diğer birleştirme operatörü genellikle göz ardı edilmektedir (Beliakov, Calvo ve James, 2011).

Beliakov ve diğeri (2011) kullanılan tavsiye sisteminin özelliklerine göre farklı birleştirme operatörleri kullanılması tavsiye sonuçlarının iyileşmesine yol açabileceğini, fakat daha karmaşık operatörlerin kullanılmasının daha doğru tavsiyelere neden olmayacağını belirtmektedir.

Bu çalışmada birleştirme işlemlerini yapmak için sıralı ağırlıklı ortalama (İng. Ordered weighted averaging, OWA) birleştirme operatörleri kullanılmıştır.

### 2.5.1. Sıralı ağırlıklı ortalama (OWA) birleştirme operatörleri

Yager (1988) çalışmasında çoklu kriterlerin bir araya getirildiği genel bir karar fonksiyonu oluşturmuştur. Sıralı ağırlıklı ortalama (OWA) operatörü olarak adlandırılan bir birleştirme operatörü tanıtmış ve bu operatörün özelliklerini incelemiştir. OWA operatörünün başarımının, tüm ölçütleri yerine getirmesini gerektiren VE operatörünü kullanarak elde edilen başarımla en az bir kriterin karşılanması gerektiği VEYA operatörü arasında bulmuştur.

$n$  boyutlu bir OWA operatörünün ( $F: R^n \rightarrow R^n$ ) ilişkili  $n$  boyutlu ağırlık vektörü,  $W$ ;

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad (2.1)$$

olmak üzere

$$w_j \in [0,1] \quad (2.2)$$

ve

$$\sum_{j=1}^n w_j = 1 \quad (2.3)$$

şartlarını sağlar. OWA operatörü



$$F(a_1, \dots, a_n) = \sum_{j=1}^n w_j b_j \quad (2.4)$$

olarak tanımlanır. Burada  $b_j$ ,  $a_i$  değerlerinin  $j$ -inci büyük olanıdır.

Bu operatörün önemi, argümanların değerlerine dayalı olarak yeniden sıralanmasıdır. Ağırlıklar, belli bir argümana bağlı olmadan sıralamada belli bir konumla ilişkilendirilir.

B vektörü argümanların sıralanmış haline karşılık geliyorsa, OWA operatörü basitçe

$$F(a_1, \dots, a_n) = W^T B \quad (2.5)$$

şeklinde tanımlanabilir (Yager ve Filev, 1999). Burada  $W^T$ , ağırlık vektörünün devriğidir.

Yager (1988) ağırlık vektörü ile ilgili iki önemli ölçüt tanımlamıştır. Bunlardan birisi operatörün VE ile VEYA arasındaki derecesini gösteren, veyalık (İng. orness) ya da bir diğer adıyla en büyüklük (İng. maxness) derecesi,  $\alpha$ 'dır.

$$\alpha(W) = \frac{1}{n-1} \sum_{j=1}^n w_j (n-j) \quad (2.6)$$

olarak tanımlanır.  $\alpha(W) = 1$  durumunda, ağırlık vektörü  $W$  en büyük (İng. max),  $\alpha(W) = 0,5$  durumunda ortalama ve de  $\alpha(W) = 0$  olduğu durumda ise en küçük (İng. min) operatörü gibi davranır. Yager (1988)'in tanımladığı diğer ölçüt ise dağınımdır (İng. dispersion). Dağınım,  $H$ , ağırlık vektörü  $W$ 'nin birleştirme işlemi esnasında tüm bilgileri dikkate alma derecesini gösterir. Ağırlık vektörünün düzenli bir şekilde dağılması, dağınım derecesini de yükseltir. Dağınım derecesi

$$H(W) = - \sum_{j=1}^n w_j \ln(w_j) \quad (2.7)$$

eşitliği ile hesaplanır.

Fodor, Marichal ve Roubens (1995) çalışmasında OWA operatörlerini nitelendirmiş ve farklı ağırlık vektörlerine göre davranışlarını sıralamıştır.

Dilsel sıralı ağırlıklı ortalama (İng. Linguistic ordered weighted averaging, LOWA) operatörleri dilsel niceleyicilerin kullanıldığı özel bir OWA türüdür (Herrera, Herrera-Viedma ve Verdegay, 1996). Bir dilsel değişken, sayısal değerlerden farklı olarak doğal dilde yazılmış sözcükler veya cümlelerdir. Genellikle, problem alanına bağlı olarak, belirsiz veya kesin olmayan bilgiyi açıklamak için uygun bir dilsel terim seti seçilir ve kullanılır. Belirlenen setteki öğeler farklı belirsizlik seviyeleri arasındaki ayrım düzeyini belirler (Delgado, Verdegay ve Vila, 1993; Herrera ve Herrera-Viedma, 1997).

Uyarılmış sıralı ağırlıklı ortalama (İng. Induced ordered weighted averaging, IOWA) operatörleri Yager ve Filev (1999) tarafından genişletilmiştir. Bu operatörler, girdi olarak OWA çiftlerini kullanır. OWA çifti,  $\langle u_j | a_j \rangle$ , argüman ve argümanın niteliklerinden uyarılmış sıralama bileşeninden oluşmaktadır. Argümanlar sıralama bileşenine göre sıralanır.

$$\text{IOWA}(\langle u_1 | a_1 \rangle, \dots, \langle u_n | a_n \rangle) = W^T B \quad (2.8)$$

Burada B vektörü argümanların u bileşenlerine göre sıralanmış haline karşılık gelmektedir. Chiclana, Herrera-Viedma, Herrera ve Alonso (2007) çalışmasında bazı IOWA operatörlerini karşılaştırmış ve karar verme süreçlerindeki kullanımlarını açıklamıştır.

Sıralı ağırlıklı geometrik ortalama (İng. Ordered weighted geometric averaging, OWGA) operatörleri Xu ve Da (2002) tarafından tanımlanmıştır.

$$\text{OWGA}(a_1, \dots, a_n) = \prod_{j=1}^n w_j^{b_j} \quad (2.9)$$

Burada  $b_j$   $a_i$  değerlerinin j-inci büyük olanıdır.

### OWA operatörlerinin ağırlıklarının belirlenmesi

Sıralı ağırlıklı ortalama operatörlerinin uygulanmasında ağırlıklarının belirlenmesi çok önemlidir.

Ağırlıkların belirlenmesinde uygulanan ilk yöntemlerden birisi, O'Hagan (1988) tarafından öne sürülmüştür. Belirli bir veyalık derecesi için en büyük dağınım değerine sahip ağırlıkları belirlemek için bir çözüm önermiştir.

Yager (1988; 1993) OWA operatörünün ağırlıklarını dilsel niceleyicileri kullanarak hesaplamak için bir yol önermiştir. Oransal nicelikler birçoğu, en azından yarısı gibi göreceli miktarları temsil etmek için kullanılır. Bir oransal nicelik,  $[0,1]$  birim aralığındaki bulanık bir alt kümeyle ( $Q$ ) temsil edilebilir.  $r \in [0,1]$  olmak üzere, üyelik fonksiyonu  $Q(r)$ ,  $r$  oranının temsil ettiği niceleyicinin anlamıyla ne kadar uyumlu olduğunu gösterir.

Bir orantılı niceleyici,  $Q: [0,1] \rightarrow [0,1]$ , aşağıdaki şartı sağlar:

$$Q(0) = 0 \text{ ve } \exists r \in [0,1] \text{ için } Q(r) = 1 \quad (2.10)$$

Bir azalmayan niceleyici ise

$$\forall a, b \text{ için eğer } a > b \text{ ise } Q(a) \geq Q(b) \quad (2.11)$$

şartını sağlar. Bir azalmayan orantılı niceleyici için üyelik fonksiyonu,  $a, b, r \in [0,1]$  olmak üzere

$$Q(r) = \begin{cases} 0 & \text{eğer } r < a \\ \frac{r-a}{r-b} & \text{eğer } a \leq r \leq b \\ 1 & \text{eğer } r > b \end{cases} \quad (2.12)$$

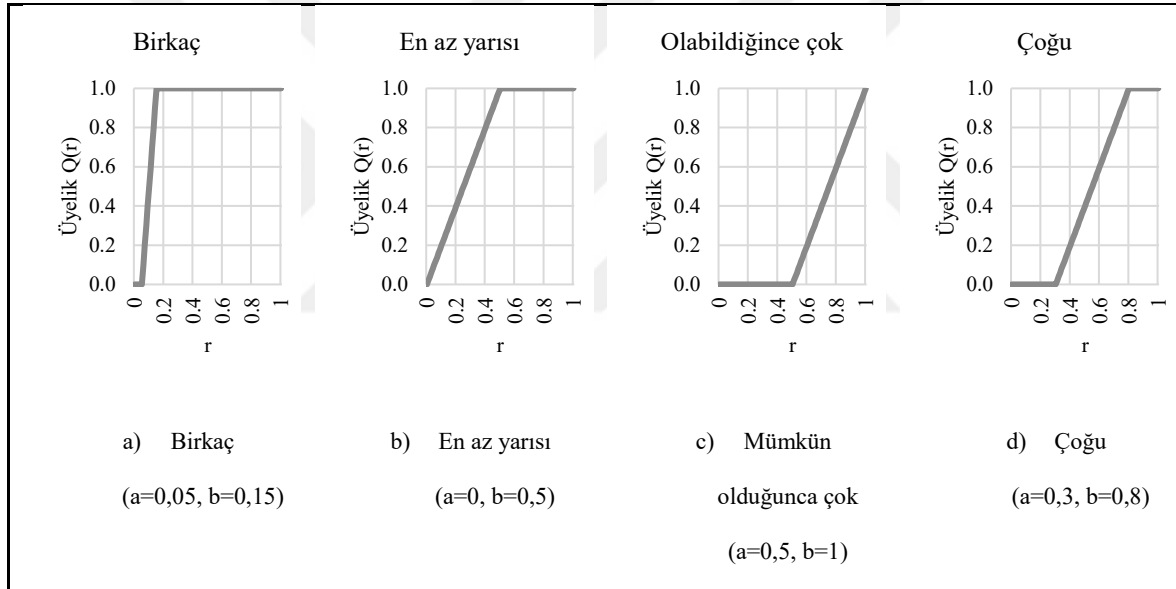
olarak tanımlanabilir.

$j$ . en büyük  $a_i$  değeri  $b_j$ , ve bu kritere ait önem derecesi  $u_j$ ,  $S_j = \sum_{k=1}^j u_k$  ve  $T = \sum_{k=1}^n u_k$  olmak üzere ağırlıklar

$$w_j = Q\left(\frac{S_j}{T}\right) - Q\left(\frac{S_{j-1}}{T}\right) \quad (2.13)$$

olarak hesaplanabilir (Herrera ve Herrera-Viedma, 1997; Herrera ve diğerleri, 1996; Yager, 1988; 1993).

Bazı dilsel niceleyiciler Şekil 2.4'de gösterilmektedir.



Şekil 2.4. Dilsel niceleyiciler; birkaç, en az yarısı, mümkün olduğunca çok ve çoğu.

Fullér ve Majlender (2001) dağılım derecesini en büyük olacağı ağırlık vektörleri elde etmek için analitik bir yaklaşım önermiştir. Fullér ve Majlender (2003) OWA operatörü ağırlıklarını elde etmek için yaptığı başka bir çalışmada ise en küçük değişimli (İng. minimal variability) ağırlıkları bulacak bir yaklaşım önermiştir.

Wang ve Parkan (2005) OWA operatörü ağırlıkları için en küçük – en büyük farklılığı (İng. minimax disparity) yaklaşımı öne sürmüşlerdir. Önerilen yaklaşım herhangi bitişik iki ağırlık arasındaki en büyük farkı küçülterek ağırlıklar belirlemektedir.

Liu (2007) en küçük deęişim ve en küçük – en büyük farklılık yaklaşımlarını ile elde edilen ağırlık vektörlerinin tamamen aynı olması üzerine çalışmış, eş farklı (İng. equidifferent) bir OWA ağırlık vektörü biçimi tanıtmış ve doğrudan ağırlık vektörü üretme yöntemi önermiştir. Bu yöntem ikinci dereceden veya doğrusal programlama problemlerini çözmeyi gerektirmemektedir.

Xu (2005) OWA operatörü için ağırlıkların elde edilmesi üzerine yapılan çalışmaları incelemiş, normal dağılım tabanlı yeni ve pratik bir yöntem önermiştir.

OWA operatörleri ve ağırlıklılarının belirlenmesi konusunda birçok çalışma yapılmış ve halen de yeni çalışmalar yapılmaya devam etmektedir (Carlsson ve Fullér, 2018; Liu, 2007; Sang ve Liu, 2014; Wang, Luo ve Liu, 2007; Wang ve Parkan, 2005; Yager, 2017; Yager ve Alajlan, 2016).

OWA operatörleri ve uygulamaları hakkında daha detaylı bilgi için Yager, Kacprzyk ve Beliakov (2011) ve Yager ve Kacprzyk (2012) kaynaklarına başvurulabilir.

## 2.6. Bölüm Sonucu

Bu bölümde kavramsal çerçeve ve ilgili araştırmalara yer verilmiştir.

Çevrimiçi reklamcılık hem Türkiye’de hem de Dünya’da büyük bir pazar payına ulaşmıştır. Çevrimiçi reklamcılık alanının ana ölçütlerinden biri olan tıklama oranı ise hızla düşmekte, reklamların etkinliği azalmaktadır. Tıklama oranının artırılması için kullanılan yöntemler hesaplamalı reklamcılık alanı altında değerlendirilmektedir. Hedefleme, reklam gösterimlerinin anlam kazanması ve değerli hale gelmesi için kullanılan ana adımlardan birisidir ve en yaygın olanları içeriğe dayalı hedefleme, demografik hedefleme, konum tabanlı hedefleme ve davranışsal hedefleme olarak sıralanmaktadır.

Kullanıcı profili oluşturmak çevrimiçi reklamcılık alanında hedeflemenin en önemli adımıdır. Kullanıcı profilleri oluşturabilmek için gerekli veriler dolaylı yoldan (ihtiyaç duyulan veriler kullanıcının geçmiş eylemleri üzerinden çıkartılır) ya da doğrudan (anket, kayıt esnasında verilen bilgiler vb.) talep edilebilir. Dolaylı yoldan kullanıcı profili

oluşturmanın bir yolu da web kullanım madenciliğinden elde edilen verileri kullanmaktır. Web kullanım madenciliğini, diğer yöntemlerden ayıran en büyük fark insanların davranışlarını yansıtmasıdır.

İkili kümeleme diğer kümeleme yöntemlerinden farklı olarak nesnelere ve de niteliklerini eş bir şekilde göz önüne alan bir kümeleme yöntemidir, böylelikle nesnelere kümelediği gibi niteliklerini de kümeleyerek, diğer yöntemlerin bulamayacağı benzersiz desenleri ortaya çıkarabilmektedir.

Bir tavsiye sistemi, girdi olarak kullandığı verileri, kullanıcılarının gelecekte beğeneceği ve ilgi duyacağı öğeleri tahmin etmek için kullanır. Tavsiye sisteminin görevi, kişiselleştirilmiş, önerilen öğeler listesini kullanıcılarına ulaştırmaktır. Bu iki yolla yapılabilir: *Tahmin* ve *Tavsiye*. Tavsiye süreci bilgi toplama, öğrenme ile tahmin ve tavsiye evrelerinden oluşmaktadır. İşbirlikçi filtreleme kullanıcıların diğer kullanıcıların görüşlerine dayalı seçimler yapmasına olanak sağlayan bir tavsiye yöntemidir. Alanlar arası tavsiye sistemleri ise farklı alanlardaki verileri ya da tavsiyeleri birlikte ya da ayrı ayrı değerlendirerek kullanıcılar zenginleştirilmiş tavsiyeler sunmaktadır. Tercihlerin, kriterlerin veya benzerliklerin bir araya getirilerek birleştirilmesi, tavsiye sistemlerinde çeşitli aşamalarda, farklı yöntemlerle gerçekleştirilmektedir.

OWA birleştirme operatörleri genel bir karar fonksiyonu oluşturmak için kullanılan bir araç olarak tanıtılmıştır. Bu operatörü diğer operatörlerden farklı kılan özelliği, argümanlar değerlerine dayalı olarak yeniden sıralanırken, ağırlıklar ise belli bir argümana bağlı olmadan sıralamada belli bir konuyla ilişkilendirilmesidir. Bu ağırlıklar dilsel niceleyiciler kullanarak da hesaplanabilmektedir.

Sonraki bölümde bu çalışmada kullanılan veri kümeleri tanıtılmakta ve çalışmada uygulanan yöntemler detaylı bir şekilde anlatılmaktadır.

### 3. YÖNTEM

Bu çalışma kapsamında önerilen yöntem bu bölümde detaylandırılmaktadır. Öncelikle çalışmada kullanılan veri kümeleri tanıtılmaktadır. Ardından önerilen yöntemin ana hatları paylaşılmaktadır. Son olarak yöntemin ana adımları olan kullanıcı profillerinin oluşturulması, kullanıcı ve ilgi kümelerinin bulunması ve tavsiye ana adımları detaylı olarak açıklanmaktadır.

#### 3.1. Veri Kümelerinin Tanıtımı

Bu çalışmada, çevrimiçi reklam alanına özgü bir gerçek dünya veri kümesi ve bir de bu kümeyi baz alan sentetik veri kümesi kullanılmıştır.

##### 3.1.1. Gerçek dünya veri kümesi

Gerçek dünya veri kümesi kısaca reklam görüntüleme günlükleri, reklam tıklanma günlükleri ve kullanıcı gezinti günlükleri olarak adlandırılacak üç farklı günlük dosyasından oluşmaktadır. Veriler 1 Kasım 2015 ile 31 Aralık 2015 tarihleri arasında reklam dağıtım ağı tarafından toplanmıştır. Bu verilerin toplanmasının amacı, reklam dağıtım ağının müşterilerine alakalı reklamları sunabilmektir. Veriler halka açık değildir. Çalışmada kullanılan tüm veriler, en az bir reklama tıklamış ve ABD bölgesinde bulunan, rastgele seçilen 250 isimsiz kullanıcı için 2 aylık bir aktiviteye karşılık gelmektedir. Bu çalışma amacıyla temin edilen verilerin tamamı anonimleştirilmiştir.

Veri kümesi üç farklı kayıt türünü tutan günlük dosyalarından oluşmaktadır (Çizelge 3.1). Bunlar reklam görüntüleme, reklam tıklanma ve kullanıcı gezinti kayıtlarını tutan günlük dosyalarıdır.

Reklam görüntüleme kayıtlarını tutan günlük dosyaları, gösterilen her reklam için tutulan ham verileri içermektedir. Her bir kayıt 72 özellik barındırmakta, bu özellikler; reklama ait bilgiler, reklamın gösterim şekline ait bilgiler, reklamı gören kullanıcıya ait bilgiler, reklamın bağlantılı olduğu reklam kampanyasının detayları ve zaman damgası olarak sınıflandırılabilir. Reklam kampanyası, aynı malı veya hizmeti tanıtan reklam grubudur ve

birden fazla reklam içerebilir.

Reklam tıklanma kayıtlarını tutan günlük dosyaları, tıklanan her reklam için tutulan ham verileri içermektedir. Her bir kayıt, reklam görüntüleme kayıtları ile aynı özellikleri içermektedir.

Kullanıcı gezinti kayıtlarını içeren günlük dosyaları, kullanıcının hareketlerini takip etmek amacıyla tutulan kayıtlardan oluşmaktadır. Her bir kayıt kullanıcıya ait bilgiler, gezinti adresi ve zaman damgasını içeren altı özellikten oluşmaktadır.

Çizelge 3.1. Veri türleri

Veri Türü	Miktar	Alanlar
Reklam Görüntülenme Kayıtları	1 254 961	72
Reklam Tıklanma Kayıtları	5 109	72
Kullanıcı Gezinti Kayıtları	1 321 757	6

Çevrimiçi reklamcılık alanının doğası gereği, reklam tıklanma verilerinin seyrekliği %99'dur.

### 3.1.2. Veri hazırlama

Verilerden genel bazı istatistiki bilgiler elde edilerek, üzerlerinde yapılabilecek olan işlemler araştırılmıştır. Başlıca uygulanabilecek işlemler, veri madenciliğinde de sıklıkla kullanılan, temizleme, indirgeme ve zenginleştirme işlemleridir.

Elde edilen veriler üzerinde ek bir temizleme işlemi gerçekleştirilmemiştir. Bunun nedeni, tavsiye sisteminin kullanılması planlanan sistem üzerinde bu işlemlerin gerçek zamanlı yapılmasının zaman ve maliyet bakımından uygulanabilir olmamasıdır.

Verilerde bazı alanların tüm kayıtlar için aynı olması nedeniyle (kullanıcıların konum, dil bilgisi gibi) bu alanlar ilerleyen işlemlerde göz ardı edilmiştir. Reklam önerisinin, reklamın bağlı olduğu kampanya üzerinden yapılmak istenmesi nedeniyle de reklama ait alanlar



(örneğin reklama ait boyut bilgileri gibi) göz ardı edilerek, bunun yerine reklamın bağlı olduğu kampanyaya ait alanlar dikkate alınmıştır.

Veri zenginleştirme amacıyla iki servis hazırlanmıştır. Bu servisler kullanıcılar için anlamlı bir profil oluşturabilmek için üçüncü parti servisleri ya da veri kaynaklarını kullanarak, kullanıcının yapmış olduğu aktiviteler üzerinden yeni veriler çıkarmaktadır. Bu veriler bir veri tabanında tutulmuş, bu aşamada üzerlerinde herhangi bir işlem yapılmamıştır. Çizelge 3.2’de dış kaynaklardan temin edilen veri türleri gösterilmektedir.

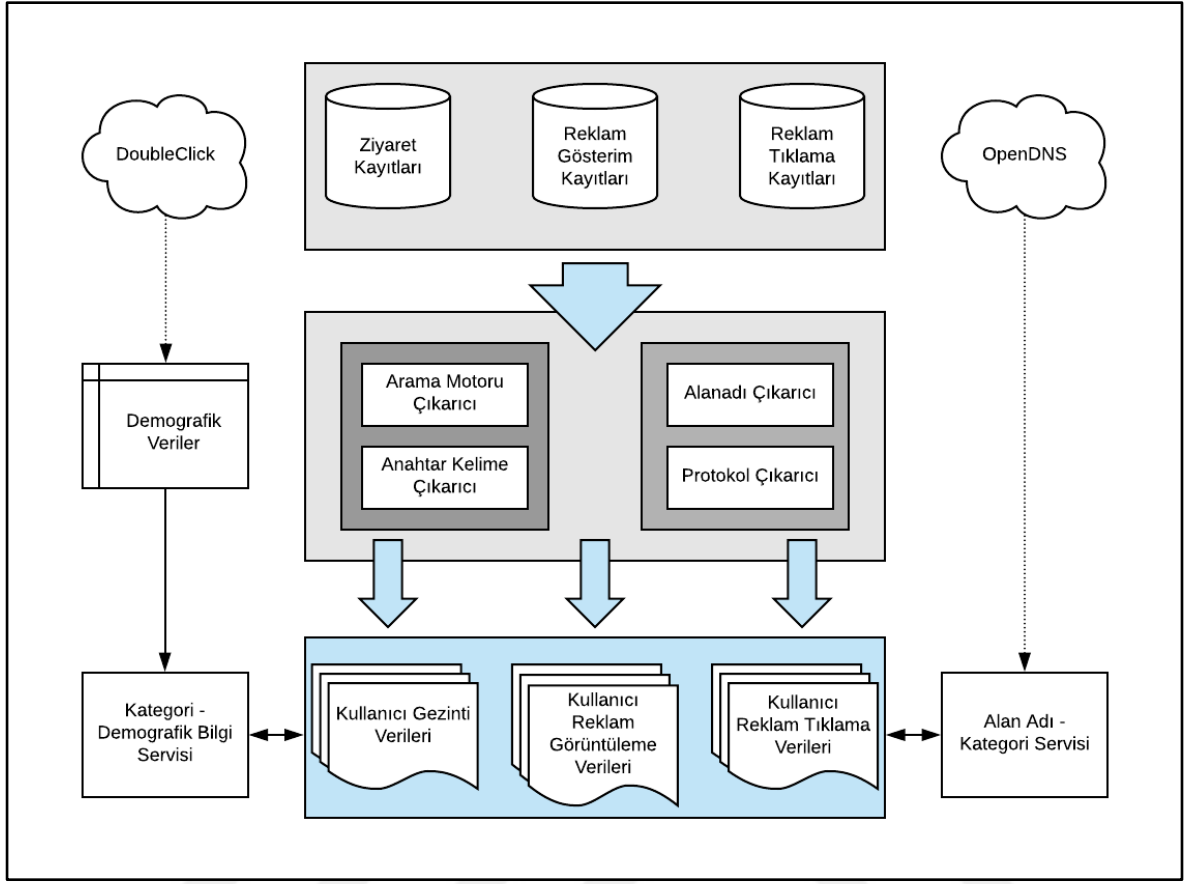
Çizelge 3.2. Dış kaynaklardan temin edilen veri türleri

Veri Türü	Miktar	Açıklama
Kategoriler	61	Alan Adı – Kategori servisinden alınan eşsiz kategoriler
Alan Adı – Kategori Verileri	35 816	Alan adlarının ilgili olduğu kategoriler listesi
Kategori – Demografik Bilgi Verileri	61	Kategori için yaş aralığı ve cinsiyet dağılımı bilgileri

Şekil 3.1’de veri kümesinde bulunan ziyaret, reklam gösterim ve reklam tıklanma kayıtlarının, bahsedilen ek servisler ile nasıl zenginleştirilerek kullanıcı gezinti, kullanıcı reklam görüntüleme ve kullanıcı reklam tıklama verilerine dönüştürüldüğü gösterilmektedir.

*Alan Adı – Kategori Servisi*, verilen web sitesinin ait olduğu kategorileri dönmektedir. Bu servis, web sitelerinin folksonomi yöntemiyle sınıflandırılmasıyla oluşturulan OpenDNS Domain Tagging (OpenDNS, 2016) veri tabanından yararlanmaktadır. Oyunlar, Seyahat ve Filmler gibi 61 ana kategori bulunmaktadır.

*Kategori – Demografik Bilgi Servisi*, verilen bir kategori için yaş aralığı ve cinsiyet dağılımı bilgilerini dönmektedir. Bu servis ise DoubleClick (DoubleClick, 2016) veri tabanından yararlanmaktadır.



Şekil 3.1. Veri zenginleştirme işlemlerinin özetlenmiş gösterimi

Çizelge 3.3 ve Çizelge 3.4 sırasıyla reklam tıklanma ve de gezinti verilerine ait bazı istatistikî bilgilere yer vermektedir.

Çizelge 3.3. Reklam tıklanma verilerine ait bazı istatistikî değerler

İstatistik	Değer (Gerçek)	Değer (Sentetik)
Reklam kampanyalarının sayısı	356	1427
Kullanıcıların sayısı	250	7921
Kampanyaların tekil tıklanma sayısı	1 146	89924
Kullanıcı başına tekil tıklama sayısı	4,58	11,35
Kampanya başına tekil tıklama sayısı	3,22	63,01
En çok reklama tıklayan kullanıcının tıklama sayısı	20	38
En az reklama tıklayan kullanıcının tıklama sayısı	2	4
En üst %10 kullanıcının toplam tıklamalardaki oranı	%27	%21
En üst %10 kampanyanın toplam tıklamalardaki oranı	%87	%38
Seyreklik	0,987	0,992

Çizelge 3.4. Gezinti verilerine ait bazı istatistikî değerler

İstatistik	Değer (Gerçek)	Değer (Sentetik)
Kategori sayısı	61	61
Ortalama	%1,69	%1,69
Medyan	%0,57	%1,42
En çok ziyaret edilen kategorinin oranı	%12,05	%3,04
En az ziyaret edilen kategorinin oranı	<%0,01	%0,09
En üst %10 kategorinin toplam ziyaretlerdeki oranı	%47	%19

### 3.1.3. Sentetik veri kümesi

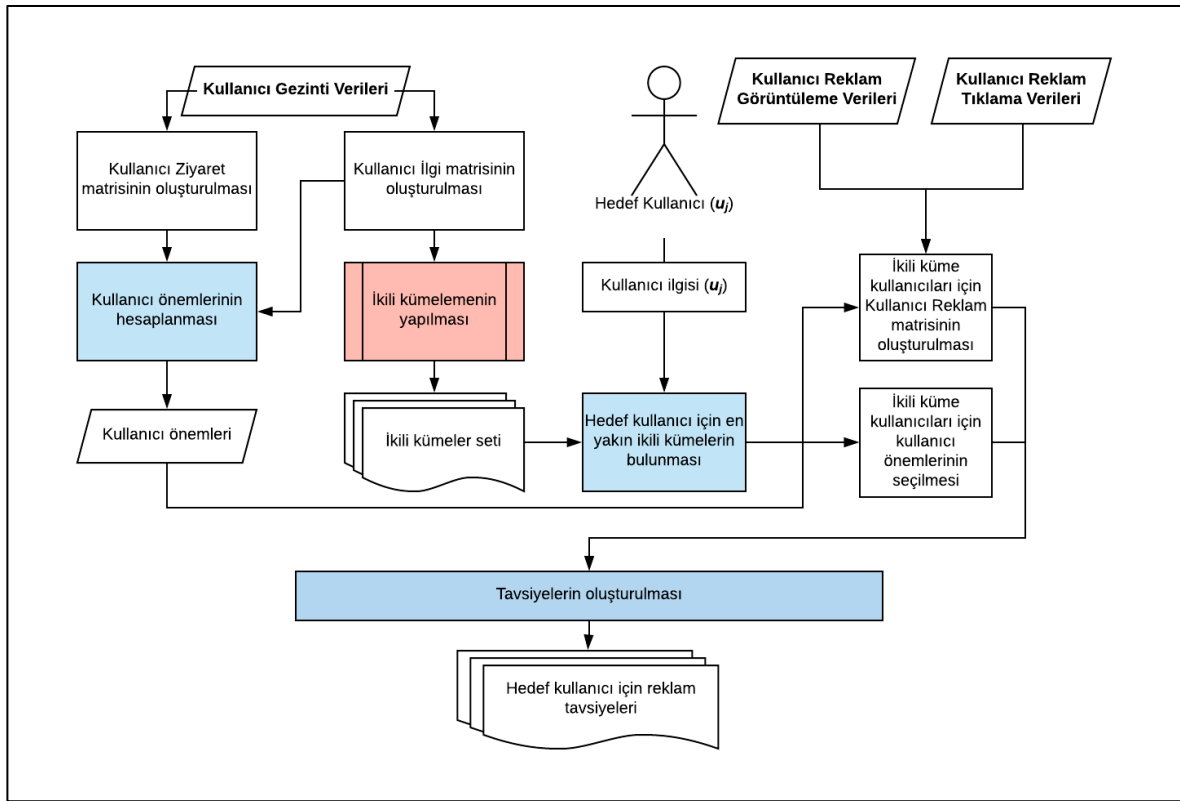
Gerçek dünya veri kümesi nispeten küçük olduğundan, özel ihtiyaçları karşılamak için sentetik bir veri kümesi oluşturulmuştur. Sentetik bir veri kümesi oluşturma işlemi, bazı istatistiksel dağılımlara dayanan küçük bir kod yardımıyla yapılmıştır. Sentetik veri kümesi aşağıdaki ihtiyaçlara göre oluşturulmuş ve sınırlandırılmıştır:

- Gezinti verileri, gerçek dünya veri kümesine göre rastgele dağıtılmıştır.
- Kullanıcılar bazı ortak ilgi alanlarını paylaşmak için gruplandırılmıştır. Grup büyüklüğü toplam nüfusun en fazla %3'ü ile sınırlıdır ve bir grubun ortak ilgi kategorisi büyüklüğü 4 ile 16 arasındadır.
- En popüler reklamlar, tüm reklamların %7,5'i ile sınırlıdır.
- Ortalama tıklama oranı %0,75 ile sınırlıdır.
- Reklamlara yapılan tıklamalar, ters güç yasası dağılımına ( $1/x^n$ ) dayanılarak istatistiksel olarak dağıtılmaktadır. n reklamlar arasında daha yumuşak tıklama dağılımı için 0,25 olarak seçilmiştir.
- Kullanıcı başına tıklama sayısı rastgele olacak şekilde 4 ile 40 arasında dağıtılmıştır.

Çizelge 3.3 ve Çizelge 3.4 ayrıca sırasıyla sentetik veri kümesinin tıklanma verileri ve gezinme verileri için bazı istatistikleri göstermektedir. Veri hazırlama aşamaları sentetik veri kümesine uygulanmamıştır.

### 3.2. Genel Bakış

Önerilen yöntem üç alt bölüm üzerinden incelenebilir. Öncelikle *Kullanıcı Profillerinin Oluşturulması* bölümü incelenecektir. Bu bölümde kullanıcılara ait gezinti verilerinden yararlanarak, kullanıcı ziyaret matrisi ve kullanıcı ilgi matrisi oluşturulur. Kullanıcı önemi değerleri de bu bölümde hesaplanır. Aynı zamanda reklam görüntüleme ve reklam tıklama verileri vasıtasıyla kullanıcı reklam matrislerinin oluşturulması bu bölümde açıklanmaktadır. İkinci bölüm ise *Kullanıcı ve İlgi Kümelerinin Bulunması* aşamasını içerir. Bu bölümde kullanıcı ilgi matrisi üzerinden ikili kümeleme işlemi yapılarak ikili kümeler oluşturulur. Son bölüm olan *Tavsiye* bölümü ise hedef kullanıcının ait olduğu ikili kümelerin nasıl bulunduğunu ve tavsiye edilecek reklamların nasıl seçildiğini içermektedir.



Şekil 3.2. Önerilen yöntemin ana hatlarıyla gösterimi

Önerilen yöntem ana hatlarıyla Şekil 3.2’de gösterilmektedir. Burada bir hedef kullanıcı için reklam önerisinin nasıl yapıldığının akışı gösterilmiştir: Kullanıcı ziyaret ve kullanıcı ilgi matrisi gezinti verilerinden oluşturulur. Kullanıcı önemi ve ikili kümeler bu matrislerden üretilir. Bir hedef kullanıcı reklam önerisi istediğinde, en yakın ikili kümeler hedef kullanıcının ilgisine göre bulunur. Görüntüleme ve tıklama verilerinden, seçilen ikili küme kullanıcıları için bir kullanıcı reklam matrisi oluşturulur. Son olarak, hedef kullanıcı için reklam önerisi oluşturulur.

### 3.3. Kullanıcı Profillerinin Oluşturulması

Belirlenmiş kategorilere karşı kullanıcının ilgisini içeren kullanıcı profilleri elde etmek için kullanıcı ziyaret matrisi ve kullanıcı ilgi matrisi hazırlanması gerekmektedir. Bu matrislerin hazırlanma adımları bu alt bölümde ayrıntılı olarak açıklanmaktadır. Başka bir matris, kullanıcı reklam matris de reklam tavsiyesinin yapılması adımıyla kullanılmak üzere hazırlanmaktadır. Reklam tavsiyesinde anahtar rol oynayacak olan kullanıcı önemi kavramı da bu alt bölümde ortaya konmaktadır.

### 3.3.1. Kullanıcı ziyaret matrisi

Tüm veri setlerinde kullanıcılara ait gezinti verileri işlenerek, kullanıcı ziyaret matrisi hazırlanmıştır. Bu matris bir kullanıcının bir kategoriye kaç defa ziyaret ettiği bilgilerinden oluşmaktadır.

$$V = v_{u,c} \quad (3.1)$$

burada  $V$ , kullanıcı ziyaret matrisidir ve  $v$ ,  $u$  kullanıcısının  $c$  kategorisi için gerçekleştirdiği ziyaret sayısını ifade eder.

### 3.3.2. Kullanıcı ilgi matrisi

Kullanıcı ziyaret matrisi, bir kullanıcının bir kategoriye kaç kez ziyaret ettiği hakkında bilgiler içerir. Fakat, iki farklı kullanıcının (Ayşe ve Burak olarak adlandırılсын) aynı kategoriye eşit olarak ziyaret etmiş olması, tek başına anlam ifade etmez. Her iki kullanıcının da Spor kategorisine ait web sitelerine 300 ziyarette bulunduğunu varsayalım. Ayşe'nin toplam ziyaret sayısı 1000 ve Burak'ın toplam ziyaret sayısı 50000 ise, bu iki kullanıcının Spor kategorisine ilgisi aynı sayılamaz. Bu yüzden kullanıcının ziyaret sayısından çok, ziyaret oranına dikkat etmek gerekir.

Benzer şekilde, bazı kategoriler tüm kullanıcılar tarafından diğer kategorilerden daha sık ziyaret edilmektedir. Örneğin, Spor kategorisindeki web siteleri, tüm kullanıcı ziyaretlerinin ortalama %10'u oranında ziyaret edilirken, Oyuncak kategorisindeki web siteleri yalnızca %0,5'i oranında ziyaret ediliyor olsun. Aynı zamanda Ayşe, ziyaretlerinin %5'ini Spor, %2'sini Oyuncak kategorisine yapmış olsun. Ayşe, yalnızca kendi içinde değerlendirildiğinde Spor kategorisiyle daha ilgili olacaktır, ancak diğer kullanıcılarla karşılaştırıldığında Oyuncak kategorisine daha fazla ilgi duyduğu ortaya çıkacaktır.

Bu iki koşulu ele almak için yerel ilgi ve küresel ilgi olarak adlandırılacak iki ilgi türü tanımlanmaktadır. Daha sonra bu iki ilgi türünü kullanıcı ilgi matrisi ile birleştirilerek ikili kümeleme işlemleri için kullanılacak olan matris elde edilmektedir.

Kullanıcının bir kategoriye olan yerel ilgisi kullanıcının o kategoriye ait gezintilerinin toplam sayısının, tüm kategorilere ait gezintilerine oranı şeklinde hesaplanmıştır. Dolayısıyla bu 0 ile 1 arasında bir değerdir.

$$l_{u,c} = \frac{v_{u,c}}{\sum_{c^*} v_{u,c^*}} \quad (3.2)$$

burada  $l$ ,  $u$  kullanıcısının  $c$  kategorisi için yerel ilgisidir. Kullanıcı ilgi vektörü, kullanıcının tüm kategoriler için ilgisini barındıran bir vektör olarak tanımlanmıştır.

$$p_u = \begin{bmatrix} l_{c1} \\ l_{c2} \\ \vdots \\ l_{cn} \end{bmatrix} \quad (3.3)$$

burada  $p$ ,  $u$  kullanıcısı için kullanıcı ilgi vektörü ve  $n$ , kategori sayısıdır. Yerel ilgi matrisi, kullanıcı ziyaret matrisi üzerinden hesaplanabilir.

$$L = \frac{V}{V_j} \quad (3.4)$$

burada  $L$  yerel ilgi matrisidir ve  $V_j$ , kullanıcı ziyaret matrisinin satır toplamlarının bir vektörüdür<sup>2</sup>.

Her bir kullanıcı için ilgilerinin yoğun olduğu kategorileri bulabilmek amacıyla küresel ilgi vektörü tanımlamasına ihtiyaç duyulmaktadır.

$$g = \frac{j^T L}{\sum j^T L} \quad (3.5)$$

burada  $g$ , küresel ilgi vektörüdür ve  $j^T L$ , yerel ilgi matrisinin sütun toplamlarının bir vektörüdür. Artık kullanıcı ilgi matrisi,  $I$ , hesaplanabilir.

---

<sup>2</sup>  $j = (1, 1, \dots, 1)$  sütun vektörünün,  $V$  matrisi ile çarpımından, matrisin satır toplamlarını içeren bir vektör elde edilir. Bu sayede  $l_{u,c}$  değerleri ve  $p_u$  vektörleri kullanıcı ziyaret matrisi üzerinden hesaplanmış olur.

$$I = \frac{L}{g} \quad (3.6)$$

Belirlenen eşik değere göre matrisin her elemanı Eş. 3.7'ye göre değiştirilerek kullanıcı ilgi matrisi ikili hale dönüştürülür.

$$i_{u,c} = \begin{cases} 1, & \text{eğer } i_{u,c} \geq t \\ 0, & \text{eğer } i_{u,c} < t \end{cases} \quad (3.7)$$

burada  $t$ , eşik değerdir. Yüksek eşik değeri ( $>1$ ) seçimi ortak ilgi alanlarına sahip kullanıcıları daha hassas bir şekilde bulmayı sağlayacak iken, bu değerin aşırı yüksek seçilmesi ( $>3$ ) hiçbir kullanıcının birbiriyle eşleşmemesine neden olabilir.

Kullanıcı ilgi matrisi,  $I$ , satırlarında kullanıcılar,  $u$ , ve sütunlarında kategoriler,  $c$ , olan ve kullanıcı ilgisini,  $i$ , kategori bazında tutan bir matristir. Eğer  $u$  kullanıcısının etkinlikleri,  $c$  kategorisi ile ilgili herhangi bir bağ oluşturmuyorsa,  $I(u, c) = 0$ 'dır.  $u$  kullanıcısının yaptığı etkinlikler kullanıcının  $c$  kategorisine ilgisi olduğu gösteriyorsa ise,  $I(u, c) = 1$ 'dir.

$$I = \begin{bmatrix} i_{u1,c1} & \cdots & i_{u1,cn} \\ \vdots & \ddots & \vdots \\ i_{um,c1} & \cdots & i_{um,cn} \end{bmatrix} \quad (3.8)$$

burada  $m$ , kullanıcı sayısı ve  $n$ , kategori sayısıdır.

### 3.3.3. Kullanıcı demografi matrisi

Kullanıcının demografik profili,  $\delta$ , demografik özelliklere ait olasılık değerlerini barındıran bir vektördür. Demografik bilgilerin,  $d$ , yer aldığı bu vektör kullanıcı ilgi vektörü,  $\rho$ , üzerinden kategori-demografik bilgi verileri kullanılarak oluşturulmuştur.

$$\delta_u = \begin{bmatrix} d_{a[18-25]} \\ d_{a[26-40]} \\ \vdots \\ d_{female} \\ d_{male} \end{bmatrix} \quad (3.9)$$



Oluşturulan bir diğer matris ise, demografi matrisidir (kullanıcı – demografi matrisi),  $D$ . Satırlarında kullanıcılar ve sütunlarında demografik kategoriler olan bu matris, kullanıcının ilgili demografik kategoride bulunma olasılığını tutmaktadır.

$$D = \begin{bmatrix} d_{u1,da[18*25]} & \cdots & i_{u1,dmale} \\ \vdots & \ddots & \vdots \\ d_{um,da[18*25]} & \cdots & d_{um,dmale} \end{bmatrix} \quad (3.10)$$

### 3.3.4. Kullanıcı önemi

Her bir kullanıcının reklam önerisinde farklı ağırlıkları olması istenmektedir. Eğer bir kullanıcı bir başka kullanıcıya göre daha yoğunlaşmış ilgiye sahipse, daha önemlidir. İki kullanıcı kıyaslandığında, çok ziyareti bulunup, sınırlı sayıda kategoriye ilgisi olan bir kullanıcının, daha az ziyareti veya daha çok kategoriye ilgisi bulunan kullanıcılara göre daha önemli olması gerektiği düşünülmüştür.

Kullanıcı önemi

$$imp = \left( \frac{\widehat{V}}{\widehat{I}} \right)^{1/e} \quad (3.11)$$

olarak tanımlanmıştır. Burada  $imp$ , kullanıcı önemi,  $\widehat{V} = \widehat{V}_j$ , kullanıcı ziyaret matrisinin satır toplamlarından oluşan normalize edilmiş ziyaret vektörü,  $\widehat{I} = \widehat{I}_j$  kullanıcı ilgi matrisinin satır toplamlarından oluşan normalize edilmiş ilgi vektörüdür. Kullanıcı önemi 0 ile 1 arasında değerler almakta ve 1'e ne kadar yakınsa kullanıcının önemi artmaktadır.

**Örnek 3.1:** Üç kullanıcı (Ayşe, Barış ve Cenk olsun) ve beş kategoriden oluşan bir mikro-sisteme ait kullanıcı ziyaret matrisi Çizelge 3.5'de verilmiştir. Bu örnek kullanıcı ilgi matrisinin ve kullanıcı öneminin nasıl hesaplandığını göstermektedir. Çizelge 3.5'de verilen  $V$  matrisine Eş. 3.4 uygulanarak yerel kullanıcı ilgi matrisi elde edilir (Çizelge 3.6). Küresel ilgi vektörü ise Eş. 3.5'in uygulanmasıyla elde edilir ve Çizelge 3.7'da gösterilmektedir. Eş. 3.6 ve Eş. 3.7 birlikte uygulanarak kullanıcı ilgi matrisi elde edilir (Çizelge 3.8). Son olarak, kullanıcı önemi Eş 3.11 yardımıyla hesaplanır (Çizelge 3.9).

Küresel ilginin etkisi, Cenk'in Kategori 1 ve Kategori 2'ye olan ilgili üzerinde rahatlıkla görülebilir. Cenk, Kategori 1'i Kategori 2'den neredeyse iki kat daha fazla ziyaret etmiştir. Fakat, bu durum diğer kullanıcılar içinde geçerlidir. Bu yüzden Cenk'in Kategori 1'e özel bir ilgisi olmadığı, daha az ziyaret ettiği Kategori 2'ye ise ilgisi olduğu söylenebilir. Bu örnekte göze çarpan bir diğer nokta ise, Ayşe ve Barış'ın eşit sayıda toplam ziyarete sahip olmaları fakat Ayşe'nin üç kategoriye ilgisi varken, Barış'ın yalnızca bir kategoriye ilgisi olmasıdır. Bu durumda Barış'ın Ayşe'ye göre daha yoğunlaşmış bir ilgiye sahip olduğundan bahsedilebilir. Diğer yandan, Ayşe ve Cenk eşit sayıda kategoriye ilgi göstermektedir, fakat Ayşe'nin daha fazla ziyareti vardır. Bu durum da Ayşe'nin Cenk'ten daha önemli olmasına yol açmıştır. Örnekteki kullanıcıların önemi, kullanıcı önem vektörüyle aynı yöndedir.

Çizelge 3.5. Kullanıcı ziyaret matrisi, V

Kullanıcı	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5	Toplam
Ayşe	30	5	17	66	32	150
Barış	25	5	60	47	13	150
Cenk	5	2	18	3	7	35
<b>Toplam (Kategori)</b>	60	12	95	116	52	335

Çizelge 3.6. Yerel kullanıcı ilgi matrisi, L

Kullanıcı	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5
Ayşe	0,20	0,03	0,11	0,44	0,21
Barış	0,17	0,03	0,40	0,31	0,09
Cenk	0,14	0,06	0,51	0,09	0,20

Çizelge 3.7. Küresel ilgi vektörü, g

	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5
<b>Küresel ilgi</b>	0,18	0,04	0,28	0,35	0,16

Çizelge 3.8. Kullanıcı ilgi matrisi, I

Kullanıcı	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5	Toplam
Ayşe	1	0	0	1	1	3
Barış	0	0	1	0	0	1
Cenk	0	1	1	0	1	3

Çizelge 3.9. Kullanıcı önem vektörü, imp

Kullanıcı	Önem
Ayşe	0,528
Barış	0,791
Cenk	0,309

### 3.3.5. Kullanıcı reklam matrisi

Kullanıcılara ait reklam görüntüleme ve reklam tıklama verileri işlenerek, kullanıcının reklam kampanyaları ile olan ilişkisi, kullanıcının bir reklam kampanyasına ait reklamları kaç defa görüntüleyip, bunların kaçına tıkladığının oranı olan CTR (İng. click-through rate) değerleri üzerinden oluşturulmuştur (Çizelge 3.10).

Çizelge 3.10. Kullanıcı reklam matrisi, A

	ad1	ad2	ad3	ad4	ad5	ad6	ad7	...	ad356
1	0	0	0	0	0	0	0,1575	...	0
2	0	0	0	0	0	0	0	...	0
3	0	0	0,5173	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...
250	0,1202	0	0,1172	0	0,2321	0	0	...	0

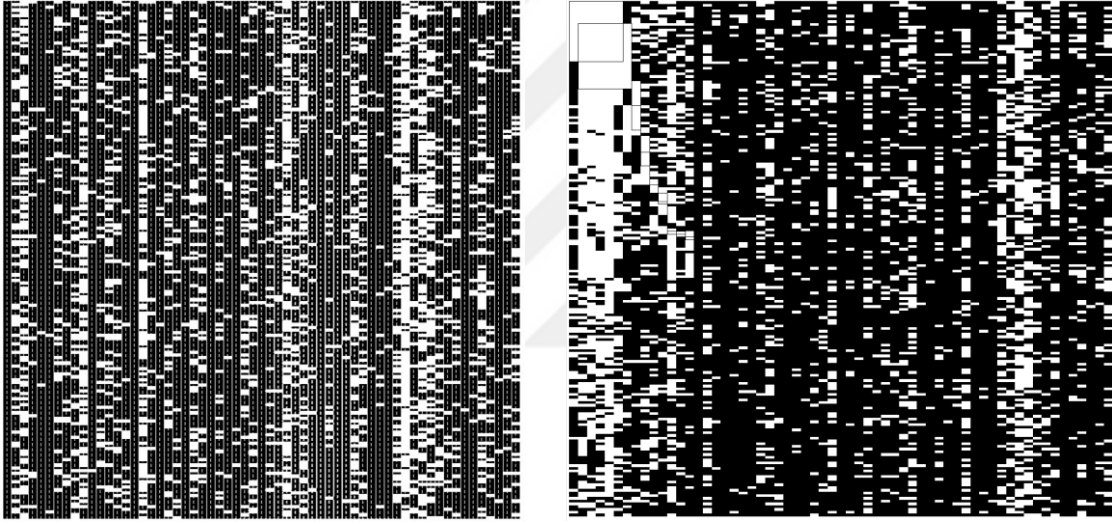
### 3.4. Kullanıcı ve İlgi Kümelerinin Bulunması

İkili kümeleme işlemi, kullanıcı ilgi matrisi üzerinden yapılmaktadır. Kullanıcı ilgi matrisi ikili bir matris olduğu için ikili kümeleme yöntemlerinden olan BiMax algoritmasının kullanılması uygun görülmüştür. BiMax algoritması azami katılım (İng. inclusion-maximal) gösteren tüm kümeleri bulduğu için, bu davranış, kümelerin kısmen üst üste gelmesine neden

olur; bu nedenle, hedef kullanıcı birden fazla ikili kümeye ait olabilir.

İkili kümeleme işlemini yapabilmek için biclust (Kaiser ve diğerleri, 2015) adlı R paketi kullanılmıştır. Bu paket, iki boyutlu verilerde ikili kümelerin bulunması için BiMax'da dahil birçok algoritmanın kullanımına olanak sağlamaktadır. İkili kümeleme işlemi yapılırken, bir ikili kümenin en düşük satır ve sütun sayılarının kaç olacağı ve bulunacak en fazla ikili küme sayısı gibi parametreler de belirtilebilmektedir.

Şekil 3.3'de kullanıcı ilgi matrisinin ikili kümeleme öncesi ve sonrası gösterilmektedir.



Şekil 3.3. İkili kümeleme öncesi (solda) ve sonrası (sağda) kullanıcı ilgi matrisi.

### 3.5. Tavsiye

Reklam tavsiyesi iki ana adımda oluşturulur. Öncelikle, hedef kullanıcının ait olduğu ikili kümeler belirlenir ve ardından ikili kümedeki kullanıcılarının ortak kararına dayanarak Üst- $N^3$  tavsiyeler seçilir.

---

<sup>3</sup> Üst-N (İng. Top-N), en yüksek dereceli N adet tavsiyenin seçildiği bir yöntemdir.

### 3.5.1. Kullanıcının ait olduğu ikili kümelerin bulunması

Hedef kullanıcının hangi ikili kümeye ya da kümelere ait olduğunu (ya da ilgi alanlarının en benzer olduğunu) bulmak için ikili veriler üzerinde kullanımı tercih edilen Jaccard indeksi hesaplanmıştır. Her bir ikili küme için, kullanıcının küme kategorilerine olan ilgisi üzerinden Jaccard indeksi

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (3.12)$$

olarak hesaplanır. Burada  $X$ , hedef kullanıcının ilgi kümesidir ve  $Y$ , ikili kümenin ilgi kümesidir. Her bir ikili küme için Jaccard indeksi hesaplanarak hedef kullanıcının bu ikili kümelere verdiği değerler (ikili küme skoru olarak adlandırılacaktır) kıyaslanmak üzere toplanır.

En yüksek skorlu ikili kümeden başlayarak, eğer sırasıyla diğer ikili kümelerin skoru belirlenen bir eşik değerin<sup>4</sup> içinde ise, o ikili kümelere dahil olmak üzere, küme kullanıcıları elde edilir. Eğer dahil edilecek ikili küme limiti 1 ise, sadece ilk ikili küme dikkate alınır. Bu işlem için Algoritma 1 kullanılmıştır.

---

**Algoritma 1:** Hedef kullanıcının ait olduğu ikili küme kullanıcılarını getir

---

**Girdi:** Hedef kullanıcı, İkili kümeler, Eşik değer, İkili küme limiti

**Çıktı:** Kullanıcı listesi

**Hedef kullanıcı** ile eşleşen tüm **ikili kümelerin skorlarını** getir

**while** *ikili küme skoru Eşik değer aralığındaysa* **ve** *İkili küme limiti geçilmediyse* **do**

Mevcut **ikili küme kullanıcılarını** getir;

**if** *Kullanıcı listesi tanımlıysa* **then**

Yeni ve tekil **kullanıcıları Kullanıcı listesine** ekle;

**else**

Yeni bir **Kullanıcı listesi** oluştur;

**Kullanıcıları Kullanıcı listesine** ekle;

**end**

**end**

**return** **Kullanıcı listesi**;

---

<sup>4</sup> Eşik değeri 0 ile 1 arasındadır. Eşik değeri 0 ise, bir sonraki küme, mevcut küme ile aynı ikili küme skoruna sahip olmalıdır. Eşik değeri 1 ise, ikili küme limiti sınırı aşılmadıkça tüm ikili kümeler dikkate alınır.

**Örnek 3.2:** Çizelge 3.11 ve Çizelge 3.12’de sırasıyla bir ikili kümedeki kullanıcılar ile ilgileri ve hedef kullanıcının o ilgi alanları için olan ilgisi gösterilmektedir. İkili küme skoru hesaplanırken hedef kullanıcının diğer ilgi alanlarına olan ilgisi göz ardı edilir. Bu örnek için ikili küme skoru Eş. 3.12 yardımıyla 0,667 olarak hesaplanmıştır.

Çizelge 3.11. İkili kümedeki kullanıcıların ilgisi

Kullanıcı	Kategori 11	Kategori 12	Kategori 23	Kategori 25	Kategori 32	Kategori 60
2	1	1	1	1	1	1
13	1	1	1	1	1	1
57	1	1	1	1	1	1
128	1	1	1	1	1	1

Çizelge 3.12. Hedef kullanıcının ilgisi

	Kategori 11	Kategori 12	Kategori 23	Kategori 25	Kategori 32	Kategori 60
Hedef	0	1	1	1	0	1
Kullanıcı						

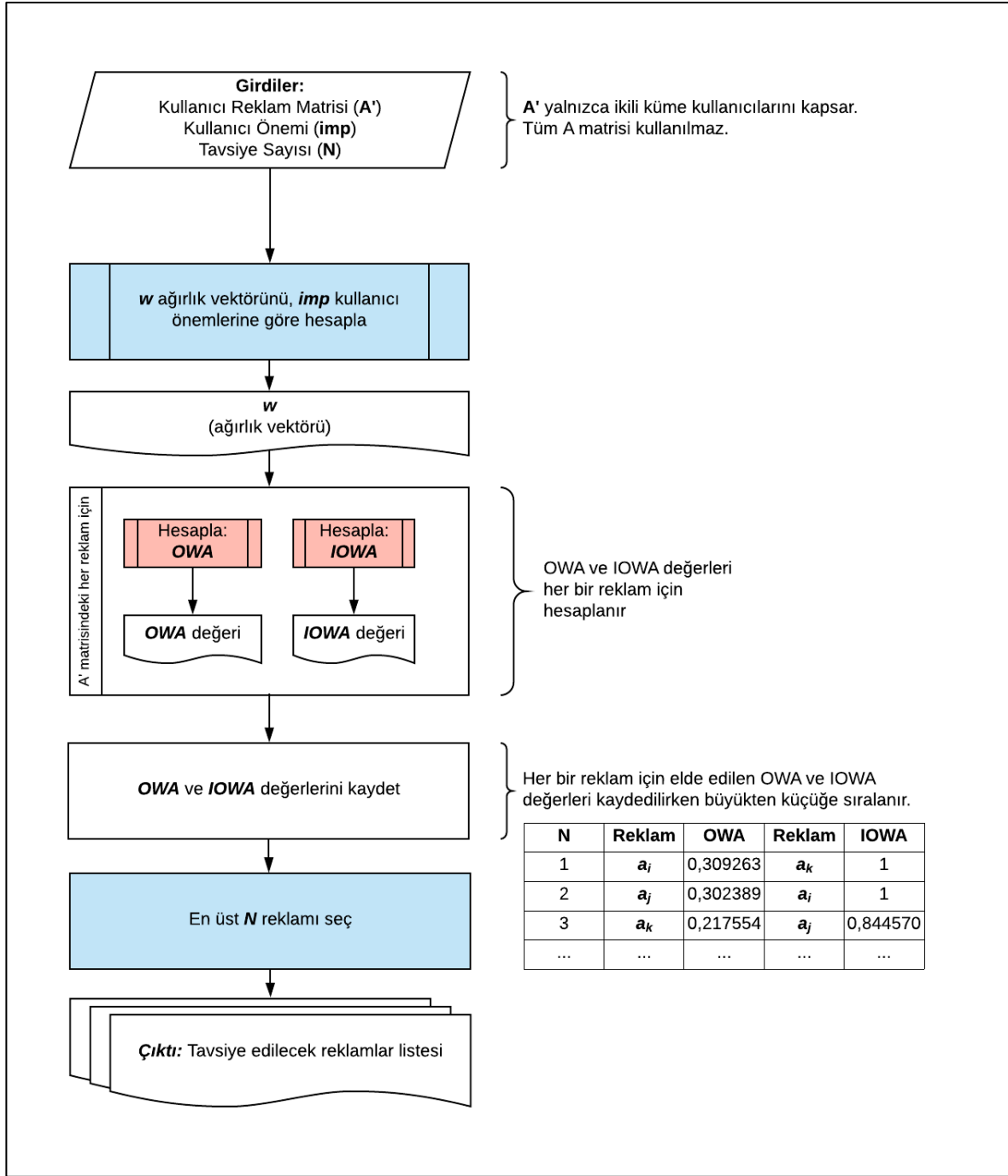
### 3.5.2. Reklam seçiminin yapılması

Reklam seçimi, hedef kullanıcının ait olduğu ikili küme kullanıcılarının ortak karar vermesi üzerine yapılır.

Bunun için OWA ve IOWA operatörleri kullanılmıştır. Kullanıcı reklam matrisi üzerinden bir reklam için kullanıcıların CTR değerleri Eş. 2.4’teki  $a_i$  vektörünü oluşturmaktadır. Bu değerlerin büyükten küçüğe sıralanmış hali ise Eş. 2.4’teki  $b_j$  vektörünü oluşturur. Her bir kullanıcı bir kriter olarak kabul edildiğinde; ağırlıklar Eş. 2.13 kullanılarak hesaplanır. Dilsel niteleyici için Eş. 2.12’de belirtilen a ve b parametreleri sırasıyla 0,05 ve 0,25 olarak seçilmiştir. Bu değerler "en az bir" ve "az" dilsel niteleyicileri arasına denk gelmektedir. Ağırlıkların hesaplanmasıyla her bir reklam için ayrı ayrı OWA değerleri hesaplanmaktadır. Aynı zamanda IOWA değerleri de Eş. 2.8 yardımıyla hesaplanmıştır. IOWA için uyarılmış sıralama bileşeni  $u_i$  olarak  $a_i$  değeri alınmıştır.  $a_i$  değerleri eğer  $a_i > 0$  ise 1,  $a_i = 0$  ise 0 olacak şekilde değiştirilmiştir. Bu değişiklik ile kullanıcının reklam tıklayıp, tıklamadığı

dikkate alınırken, uyarılmış sıralama bileşeni olarak CTR değerleri kullanılmıştır. OWA ve IOWA değerlerini hesaplamak için agop (Gagolewski ve Cena, 2014) adlı R paketi kullanılmıştır.

Her bir reklam için elde edilen OWA ve IOWA değerleri bir listede toplanır. Bu değerler büyükten küçüğe sıralanarak belirtilen sayıda reklam önerisi hedef kullanıcıya sunulmaktadır. Yanlış pozitif önerilerin sayısını sınırlamak için 0 ile 1 arasında bir eşik değeri seçilebilir. Eşik değer 1'e yaklaştıkça, tavsiye sayısı azaltılacak, ancak gerçek pozitif oranı daha iyi yüksek olacaktır. Eşik değer kullanımı diğer reklam hedefleme seçeneklerinin mevcut olması durumunda faydalı olacaktır. Şekil 3.4'de reklam önerisinin oluşturulma adımları bir akış şeklinde gösterilmektedir.



Şekil 3.4. Tavsiye edilecek reklamların seçilmesinde izlenen akış şeması.



**Örnek 3.3:** Bu örnekte öncelikle 20 kullanıcıyı bir ikili küme üzerinden tek bir reklam için Şekil 3.4’de gösterilen OWA değerleri hesaplanmakta, ardından aynı yöntemle tüm reklamlar için hesaplanan OWA değerleri üzerinden Üst-N reklam tavsiyesinin oluşturulması açıklanmaktadır. Çizelge 3.13 yeniden sıralama adımını göstermektedir. Bu ikili kümenin yalnızca beş kullanıcısı reklamı tıklamıştır ( $a_i$  değeri 0’dan büyük olanlar). Ağırlıklar, kullanıcının önemi ( $u$  sütunu) dikkate alınarak hesaplanır.

Çizelge 3.13. OWA için yeniden sıralama adımının gösterimi

i	Kullanıcı	$a_i$	$u$		j	Kullanıcı	$b_j$	$u$
1	2	0	0,055125		1	58	0,315536	0,015847
2	15	0	0,124215		2	68	0,315536	0,074681
3	17	0	0,057386		3	43	0,29546	0,062233
4	22	0	0,0785		4	93	0,243545	0,067826
5	30	0,2321	0,096331		5	30	0,2321	0,096331
6	43	0,29546	0,062233		6	2	0	0,055125
7	48	0	0,064936		7	15	0	0,124215
8	52	0	0,054962		8	17	0	0,057386
9	55	0	0,127911		9	22	0	0,0785
10	58	0,315536	0,015847	→	10	48	0	0,064936
11	68	0,315536	0,074681	(Sıralama)	11	52	0	0,054962
12	71	0	0,06863		12	55	0	0,127911
13	74	0	0,105373		13	71	0	0,06863
14	75	0	0,06179		14	74	0	0,105373
15	76	0	0,067028		15	75	0	0,06179
16	89	0	0,059802		16	76	0	0,067028
17	93	0,243545	0,067826		17	89	0	0,059802
18	95	0	0,067422		18	95	0	0,067422
19	102	0	0,069605		19	102	0	0,069605
20	109	0	0,068092		20	109	0	0,068092

Çizelge 3.14 iki farklı dilsel niceleyici için elde edilen ağırlık vektörlerini göstermektedir. Ağırlık hesaplaması için “Çoğu” dilsel nicelik belirteci seçildiyse, en az yedi kullanıcının reklamı tıklamış olması gerektiği için bu reklam tavsiye edilmeyecektir. Burada dikkat edilmesi gereken konu, yediden fazla kullanıcının reklamı tıklamış olması reklamın tavsiye edileceğini de garanti etmemektedir. “Az” dilsel nicelik belirteci seçildiyse, reklamın en az iki kullanıcı tarafından tıklanmasını gerektirdiğinden, bu reklamın önerilme şansı olacaktır.

Ağırlıklar da hesaplandıktan sonra OWA ve IOWA değerleri hesaplanır. Bu işlem tüm reklamlar için tekrar edilir ve elde edilen OWA ve IOWA değerleri büyükten küçüğe sıralanarak Çizelge 3.15 oluşturulur.

Çizelge 3.14. İki farklı dilsel niceleyici (az ve çoğu) için elde edilen ağırlık vektörleri

Kullanıcı	$b_j$	u	w	
			a= 0,05 ve b=0,25 (Az)	a=0,3 ve b=0,8 (Çoğu)
58	0,31554	0,01585	0	0
68	0,31554	0,07468	0,06266	0
43	0,29546	0,06223	0,21494	0
93	0,24355	0,06783	0,23426	0
30	0,2321	0,09633	0,33271	0
2	0	0,05513	0,15544	0
15	0	0,12422	0	0,08558
17	0	0,05739	0	0,07928
22	0	0,0785	0	0,10845
48	0	0,06494	0	0,08971
52	0	0,05496	0	0,07593
55	0	0,12791	0	0,17671
71	0	0,06863	0	0,09481
74	0	0,10537	0	0,14557
75	0	0,06179	0	0,08536
76	0	0,06703	0	0,05859
89	0	0,0598	0	0
95	0	0,06742	0	0
102	0	0,06961	0	0
109	0	0,06809	0	0
<b>OWA</b>			<b>0,21755</b>	<b>0</b>
<b>IOWA</b>			<b>0,84457</b>	<b>0</b>

Son olarak, Üst-N reklam tavsiyesi hedef kullanıcıya önerilir. OWA ve IOWA değerleri farklı olduğu için, tavsiye listeleri de kullanılan birleştirme operatörüne göre farklılık göstermektedir.

Çizelge 3.15. Her reklam için hesaplanan ve sıralanan OWA ve IOWA değerleri (Sıralama operatöre göre değişmektedir.)

N	Reklam	OWA	Reklam	IOWA
1	ad28	0,309263	ad7	1
2	ad44	0,302389	ad28	1
3	ad7	0,287252	ad44	1
4	ad46	0,217554	ad46	0,844570
5	ad24	0,192411	ad1	0,786439
6	ad8	0,151433	ad5	0,727088
7	ad25	0,140774	ad8	0,679124
8	ad47	0,134341	ad24	0,665121
9	ad5	0,104363	ad47	0,521154
10	ad34	0,089903	ad6	0,50607
11	ad1	0,077375	ad25	0,417876
12	ad6	0,07233	ad34	0,344299
13	ad18	0,046889	ad18	0,15763
14	ad23	0,016631	ad21	0,061186
15	ad11	0,015206	ad23	0,061186
16	ad21	0,012095	ad11	0,060665
17	ad20	0,010739	ad20	0,04094
18	ad30	0,000504	ad30	0,001544
...	...	0	...	0

### 3.6. Bölüm Sonucu

Bu bölümde çalışmada kullanılan veriler detaylı bir şekilde tanıtılmış ve de izlenen yöntemler açıklanmıştır.

Veriler; gezinti, reklam görüntüleme ve reklam tıklama kayıtlarının işlenmesi ve dış kaynaklı veriler ile de zenginleştirilerek çalışmada kullanılabilir hale getirilmiştir.

Kullanıcı profillerinin oluşturulması için takip edilen işlemler açıklanmış, yerel ve küresel ilgi ile kullanıcı önemi kavramlarından bahsedilmiştir. Ardından kullanıcı ve ilgi kümelerinin ikili kümeleme yardımı ile bulunması aşaması detaylandırılmıştır. Son olarak hedef kullanıcının ait olduğu ikili kümelerin bulunması ve ardından reklam seçiminin yapılmasında izlenen yöntemler adım adım gösterilmiştir.

Bir sonraki bölümde çalışmada elde edilen bulgulara yer verilmekte ve sonuçlar yorumlanmaktadır.



## 4. BULGULAR VE YORUM

Bu bölümde, öncelikle, tavsiye sistemlerinin değerlendirilmesinde kullanılan değerlendirme ölçütleri tanıtılmaktadır. Daha sonra önerilen yöntemin verimliliğini artırmak için yöntemin parametreleri optimize edilmekte ve bu parametreler paylaşılmaktadır. Son olarak, önerilen yöntem işbirlikçi filtreleme yöntemlerinden IBCF, UBCF, WRMF, BPRMF ve SLIM yöntemleri ile karşılaştırılmaktadır.

### 4.1. Değerlendirme Ölçütleri

Hassasiyet ve hatırlama, tavsiye sistemleri için yaygın olarak kullanılan değerlendirme ölçütlerinden ikisidir (Herlocker, Konstan, Terveen ve Riedl, 2004). Bobadilla ve diğerleri (2013) ve Steck (2013) tarafından yapılan çalışmalara göre çapraz doğrulama, hassasiyet, hatırlama ve ROC değerleri Üst-N tavsiye sistemlerini değerlendirmek için yaygın olarak kullanılmaktadır. Ayrıca, hassasiyet ve hatırlama yaygın olarak kullanılan ROC eğrilerine göre özellikle dengesiz dağılmış veri setlerinde yapılan ikili sınıflandırmalarda daha fazla bilgi sağlamaktadır (Saito ve Rehmsmeier, 2015).

Hassasiyeti ve hatırlamayı hesaplamak için, öğelerin ilgili ve ilgisiz sınıflara ayrılması gerekir. Hassasiyet, seçilen bir öğenin alakalı olma olasılığıdır; seçilen ilgili öğelerin seçilen öğelerin toplam sayısına oranı olarak hesaplanır. Hatırlama, ilgili bir öğenin seçilme olasılığıdır; seçilen ilgili öğelerin, ilgili öğelerin toplam sayısına oranıdır (Herlocker ve diğerleri, 2004). Çizelge 4.1 bu çalışmada reklamların nasıl sınıflandırıldığını göstermektedir.

Çizelge 4.1. Reklamların sınıflandırılması

	Seçilen	Seçilmeyen	Toplam
<b>İlgili</b>	Önerilen ve hedef kullanıcı tarafından tıklanan	Hedef kullanıcı tarafından tıklanmış fakat önerilmeyen	Hedef kullanıcı tarafından tıklanan tüm reklamlar
<b>İlgisiz</b>	Önerilen ama tıklanmamış	Önerilmeyen ve tıklanmamış	Hedef kullanıcı tarafından tıklanmayan tüm reklamlar

Çizelge 4.1'deki tanımlara göre, hassasiyet ve hatırlama;

$$P = \frac{\text{tıklanan} \cap \text{önerilen}}{\text{önerilen}} \quad (4.1)$$

$$R = \frac{\text{tıklanan} \cap \text{önerilen}}{\text{tıklanan}} \quad (4.2)$$

olarak tanımlanabilir; burada *tıklanan* hedef kullanıcı tarafından tıklanan tüm reklamlar, diğer bir deyişle, alakalı öğeler, *önerilen* tavsiye edilen tüm reklamlar, yani seçilen öğeler, *P* hassasiyet ve *R* hatırlamadır.

Farklı algoritmaları karşılaştırmak için hassasiyet ve hatırlama kullanmanın ana zorluklarından biri, bir algoritmanın başarımını iyice değerlendirmek için hassasiyet ve hatırlamanın birlikte ele alınması gerektiğidir. Hassasiyet ve hatırlamanın ters yönde bağlantılı olduğu görülmüştür (Herlocker ve diğerleri, 2004).  $F_1$  değeri, hassasiyetle hatırlamayı birleştirerek tek bir değerde tanımlar:

$$F_1 = \frac{2PR}{P + R} \quad (4.3)$$

## 4.2. Parametre Optimizasyonu

Önerilen yöntemin parametrelerini önceki bölümlerde sunulmuştur. Her parametre önerilen yöntemin başarımını etkiler. Zhen, Li ve Yeung (2009) ve Symeonidis ve diğerleri (2008b) uyguladıkları yönteme benzer bir şekilde, önerilen yöntemin her bir parametresi, diğer parametrelerin sabit tutulmasıyla tek tek optimize edilmiştir. Parametreler ve kullanılan veri kümeleri için optimize edilmiş değerleri Çizelge 4.2’de gösterilmektedir.

## 4.3. Deneysel Kurulum

Önerilen yöntem, beş farklı işbirlikçi filtreleme yöntemi ile karşılaştırılmıştır: IBCF, UBCF, WRMF, BPRMF ve SLIM. Bu yöntemler, Üst-N tavsiye görevi için iyi başarımlar gösteren güncel teknik yöntemlerini içerir. Bu yöntemlere ek olarak, popüler öneriler için elde edilen tavsiye sonuçları (Popüler) da bir alt sınır olarak paylaşılmaktadır. Bu yöntemlerden Popüler,

IBCF ve UBCF ile çalışmak için recomenderlab (Hahsler, 2017) adlı R paketi kullanılmıştır. WRMF, BPRMF ve de SLIM yöntemleri için ise bir Java kütüphanesi olan Librec (Guo, Zhang, Sun, & Yorke-Smith, 2015) kullanılmıştır.

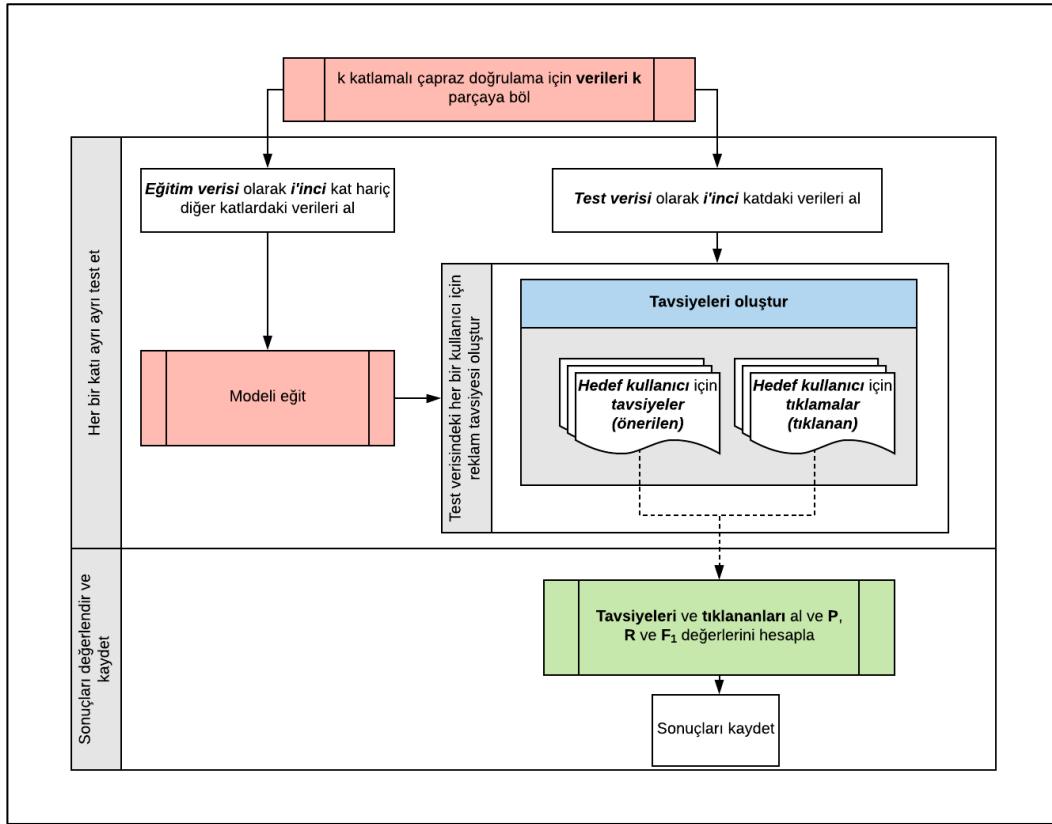
Çizelge 4.2. Parametreler ve değerleri

Parametre	Referans	Değer (Gerçek)	Değer (Sentetik)
Eşik değer (1)	3.3.2 Kullanıcı ilgi matrisi > Eş. 3.7	1	1,5
İkili küme sayısı	3.4 Kullanıcı ve İlgi Kümelerinin Bulunması	25	50
İkili kümedeki en az satır sayısı	3.4 Kullanıcı ve İlgi Kümelerinin Bulunması	10	32
İkili kümedeki en az sütun sayısı	3.4 Kullanıcı ve İlgi Kümelerinin Bulunması	6	8
Eşik değer (2)	3.5.1 Kullanıcının ait olduğu ikili kümelerin bulunması > Algoritma 1	0,02	0,05
İkili küme limiti	3.5.1 Kullanıcının ait olduğu ikili kümelerin bulunması > Algoritma 1	3	3
Eşik değer (3)	3.5.2 Reklam seçiminin yapılması	0,1	0,05
Q(r)	3.5.2 Reklam seçiminin yapılması	a=0,05, b=0,25	

Tavsiye yöntemlerinin başarımını ölçmek için k katlamalı çapraz doğrulama yöntemi (Kohavi, 1995) uygulanmıştır. Bu yöntemde göre veri kümesi belirtilen sayıda parçaya bölünmekte, bu parçaların bir bölümü test verisi olarak ayrılmakta, geri kalanları ise eğitim verisi olarak kullanılmaktadır. Yöntem eğitim verisi üzerinden eğitilmekte, ardından test verileri üzerinden sınanmaktadır. Bu işlem her parça için tekrarlanır. Tavsiye yöntemlerini k-kat çapraz doğrulama ile doğrulamak için caret (Kuhn, 2017) adlı R paketi kullanılmıştır. Yöntemin detayları, Şekil 4.1'de gösterilmiş olup öneriler, hedef kullanıcı için önerilen reklamları ve tıklamalar, hedef kullanıcı tarafından tıklanan reklamları belirtmektedir. Şekil 4.1 aynı zamanda eğitim ve test verilerinin nasıl kullanıldığını da göstermektedir.

k değeri (kat sayısı) 11 olarak seçilmiştir. Kohavi (1995)'ye göre; 10 ila 20 arasındaki ılımlı k değerleri, daha küçük k değerlerinin tersine varyansı azaltır. k değerinin artması ise hesaplamayı zorlaştırır. Deneyler farklı N değerlerine (tavsiye sayısı) karşı koşulmuştur, bu değerler 1, 3, 5, 7, 10, 15 ve 20'dir. Her bir deney için, hassasiyet, hatırlama ve  $F_1$  değerleri hesaplanmıştır.

Önerilen yöntemi istisnai ve sınırlayıcı durumlar için test etmek amacıyla, gerçek dünya veri kümesinden iki alt küme daha seçilmiştir. Bu alt kümelerden biri, en çok ziyaret eden 100 kişiyi (Deney 2'de kullanılan), diğeri ise en az ziyaret eden 100 kullanıcıyı (Deney 3'te kullanılan) içerir. Ek olarak, sentetik veri kümesi, daha büyük bir veri kümesinde bir değerlendirme yapılmasını sağlayan Deney 4'te kullanılmıştır. Son olarak, en popüler reklamların kaldırıldığı sentetik veri kümesinden bir alt küme, uzun kuyruk deneyi için kullanılmıştır (Deney 5).



Şekil 4.1. Önerilen yönteme uygulanan k-kat çapraz doğrulama yöntemi

#### 4.4. Deney Sonuçları ve Sonuçların Yorumlanması

Hem gerçek dünya veri kümesinin hem de sentetik veri kümesinin tamamı ve de farklı alt kümeleri ile beş deney yapılmıştır. Sunulan sonuçlar 11 kat üzerinden alınmış ortalamalardır. Burada not düşülmesi gereken bir önemli nokta diğer yöntemlere hedef kullanıcıları değerlendirme sırasında en az bir pozitif kullanıcı-reklam ilişkisi sunulurken ( $verilen \geq 1$ ), önerilen yönteme kullanıcı-reklam ilişkisi hakkında hiçbir bilgi verilmiyor olmasıdır ( $verilen = 0$ ).



#### 4.4.1. Deney 1: Tüm veri kümesi (gerçek dünya veri kümesi)

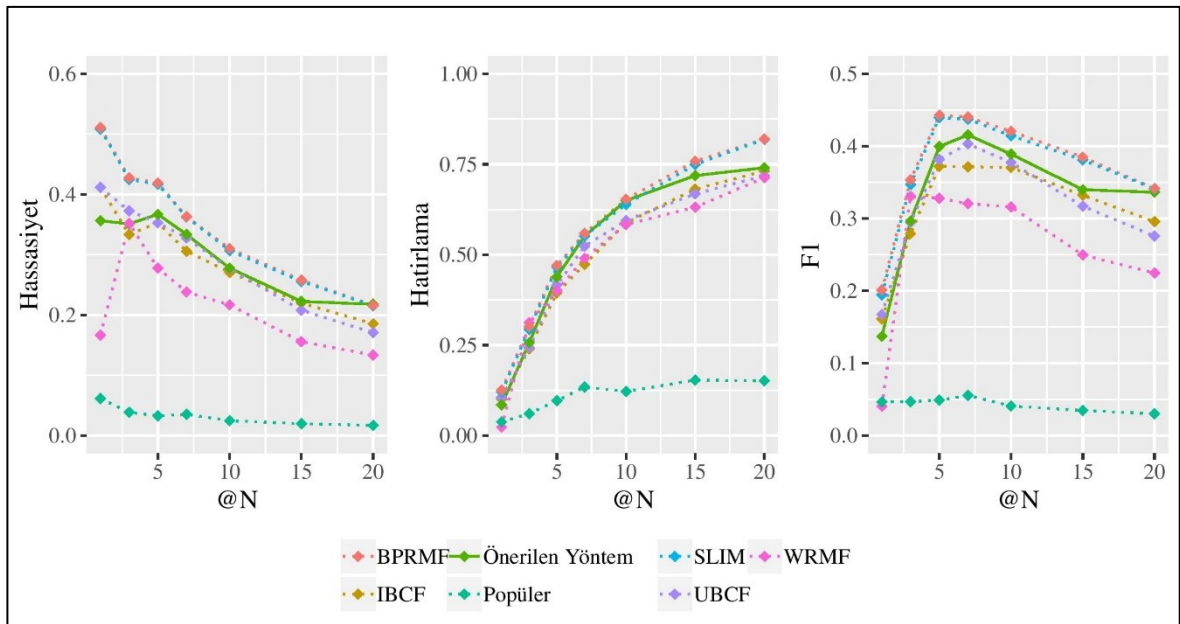
İlk deneyde, gerçek dünya veri kümesinin tamamı kullanılmıştır. Deneyle sonuçları, karşılaştırılan yöntemlerin değerlendirme ölçütlerini içeren Çizelge 4.3’de gösterilmektedir. Şekil 4.2 farklı @N değerlerinde sırasıyla hassasiyet, hatırlama ve  $F_1$  değerlerini göstermektedir. İdeal olarak, değerlerin 1'e yakın olması arzu edilir.

Çizelge 4.3. Deney 1 için karşılaştırılan yöntemlerin değerlendirme ölçütleri

@N	Ölçüt	Popüler	IBCF	UBCF	WRMF	BPRMF	SLIM	Önerilen Yöntem
1	P	0,069	0,412	0,412	0,167	<b>0,511</b>	<u>0,509</u>	0,356
	R	0,041	0,100	0,105	0,023	<b>0,125</b>	<u>0,120</u>	0,085
	$F_1$	0,051	0,161	0,167	0,040	<b>0,201</b>	<u>0,195</u>	0,137
3	P	0,042	0,333	0,373	0,352	<b>0,427</b>	<u>0,425</u>	0,351
	R	0,075	0,240	0,243	<b>0,312</b>	<u>0,301</u>	0,293	0,256
	$F_1$	0,054	0,279	0,294	0,331	<b>0,353</b>	<u>0,347</u>	0,296
5	P	0,039	0,353	0,353	0,278	<b>0,419</b>	<u>0,417</u>	0,367
	R	0,109	0,394	0,416	0,401	<b>0,470</b>	<u>0,465</u>	0,439
	$F_1$	0,057	0,372	0,382	0,328	<b>0,443</b>	<u>0,440</u>	0,400
7	P	0,040	0,306	0,328	0,238	<b>0,363</b>	<u>0,362</u>	0,334
	R	0,146	0,473	0,524	0,490	<b>0,559</b>	<u>0,552</u>	0,551
	$F_1$	0,062	0,372	0,403	0,320	<b>0,440</b>	<u>0,437</u>	0,416
10	P	0,031	0,271	0,276	0,217	<b>0,310</b>	<u>0,307</u>	0,278
	R	0,152	0,588	0,595	0,583	<b>0,653</b>	0,639	<u>0,650</u>
	$F_1$	0,051	0,371	0,377	0,316	<b>0,421</b>	<u>0,414</u>	0,389
15	P	0,023	0,220	0,208	0,156	<b>0,258</b>	<u>0,255</u>	0,223
	R	0,174	0,681	0,668	0,631	<b>0,759</b>	<u>0,749</u>	0,719
	$F_1$	0,041	0,332	0,317	0,250	<b>0,385</b>	<u>0,381</u>	0,340
20	P	0,018	0,185	0,171	0,133	<u>0,216</u>	0,215	<b>0,218</b>
	R	0,180	0,731	0,718	0,713	<b>0,820</b>	<u>0,819</u>	0,740
	$F_1$	0,033	0,296	0,276	0,225	<b>0,342</b>	<u>0,341</u>	0,336

@N'ye karşılık gelen sütun, tavsiye sayısını göstermektedir. P, R ve  $F_1$ 'e karşılık gelen satırlar sırasıyla hassasiyet, hatırlama ve  $F_1$  değerini gösterir. Her bir @N değeri için verilen ölçütler bakımından, **koyu** sayılar en iyi başarıyı ve altu çizili sayılar en iyi ikinci başarıyı göstermektedir.

BPRMF ve SLIM, birbirine yakın olan ve en iyi sonuçları veren iki yöntem olarak görülmektedir. IBCF ve UBCF de yakın sonuçlar vermektedir, ancak başarımları daha düşüktür. Sahip olduğu düşük hassasiyet başarımı nedeniyle WRMF göz ardı edilirse, önerilen yöntemin başarımı en yakın komşu tabanlı yaklaşımlar ve matris ayrıştırma modelleri arasında gözükmemektedir. Bu deney için yöntemler arasındaki farklar istatistiksel olarak anlamlı<sup>5</sup> değildir ( $\alpha=0.05$ , %99 güvenle). Önerilen yöntemin  $verilen=0$  olarak değerlendirilmesine rağmen, başarımının diğer yöntemlerin başarımlarına eş olduğu söylenebilir.



Şekil 4.2. Deney 1 için hassasiyet, hatırlama ve  $F_1$  değerleri

<sup>5</sup> ANOVA kullanarak Tukey Pairwise Comparisons yöntemiyle anlamlılık testi yapılmıştır.

#### 4.4.2. Deney 2: En çok ziyarette bulunan 100 kullanıcı (gerçek dünya veri kümesi)

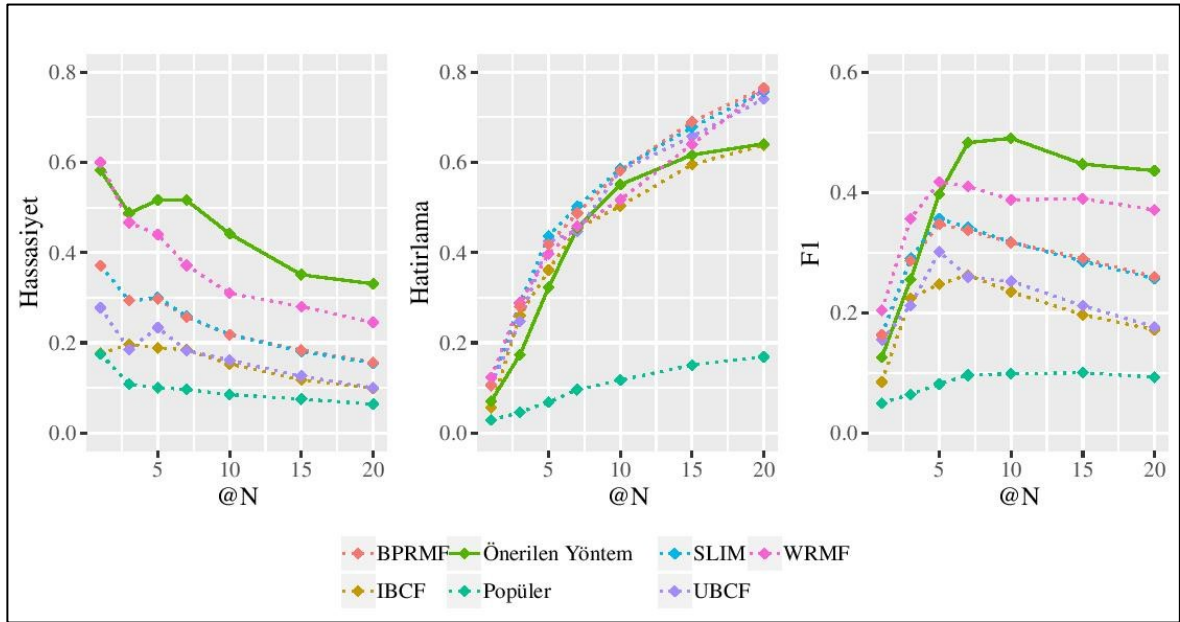
Bu deneyde kullanılmak üzere tüm veri kümesi yerine en çok ziyaret eden 100 kullanıcıyı içeren bir alt küme seçilmiştir. Deneylerin sonuçları Çizelge 4.4’de gösterilmektedir. Şekil 4.3 farklı @N değerlerinde sırasıyla hassasiyet, hatırlama ve  $F_1$  değerlerini göstermektedir.

Çizelge 4.4. Deney 2 için karşılaştırılan yöntemlerin değerlendirme ölçütleri

@N	Ölçüt	Popüler	IBCF	UBCF	WRMF	BPRMF	SLIM	Önerilen Yöntem
<b>1</b>	<b>P</b>	0,175	0,176	0,278	<b>0,600</b>	0,371	0,371	<u>0,582</u>
	<b>R</b>	0,029	0,056	<u>0,107</u>	<b>0,123</b>	0,104	0,104	0,071
	<b>F<sub>1</sub></b>	0,049	0,085	0,155	<b>0,204</b>	<u>0,163</u>	<u>0,163</u>	0,126
<b>3</b>	<b>P</b>	0,108	0,196	0,185	<u>0,467</u>	0,294	0,294	<b>0,488</b>
	<b>R</b>	0,046	0,260	0,247	<b>0,288</b>	0,280	<u>0,288</u>	0,173
	<b>F<sub>1</sub></b>	0,064	0,224	0,212	<b>0,356</b>	0,286	<u>0,291</u>	0,255
<b>5</b>	<b>P</b>	0,100	0,188	0,233	<u>0,440</u>	0,297	0,300	<b>0,516</b>
	<b>R</b>	0,068	0,361	<u>0,425</u>	0,397	0,418	<b>0,437</b>	0,323
	<b>F<sub>1</sub></b>	0,081	0,248	0,301	<b>0,417</b>	0,348	0,356	<u>0,397</u>
<b>7</b>	<b>P</b>	0,096	0,185	0,183	<u>0,371</u>	0,257	0,259	<b>0,516</b>
	<b>R</b>	0,096	0,454	0,447	0,458	<u>0,488</u>	<b>0,503</b>	0,454
	<b>F<sub>1</sub></b>	0,096	0,263	0,259	<u>0,410</u>	0,337	0,342	<b>0,483</b>
<b>10</b>	<b>P</b>	0,085	0,153	0,161	<u>0,310</u>	0,218	0,218	<b>0,442</b>
	<b>R</b>	0,117	0,503	<u>0,582</u>	0,517	0,581	<b>0,585</b>	0,551
	<b>F<sub>1</sub></b>	0,098	0,235	0,252	<u>0,388</u>	0,317	0,317	<b>0,490</b>
<b>15</b>	<b>P</b>	0,075	0,118	0,126	<u>0,280</u>	0,184	0,181	<b>0,351</b>
	<b>R</b>	0,150	0,595	0,658	0,640	<b>0,690</b>	<u>0,678</u>	0,617
	<b>F<sub>1</sub></b>	0,100	0,196	0,211	<u>0,390</u>	0,290	0,286	<b>0,447</b>
<b>20</b>	<b>P</b>	0,064	0,099	0,100	<u>0,245</u>	0,156	0,154	<b>0,331</b>
	<b>R</b>	0,169	0,638	0,741	<u>0,762</u>	<b>0,765</b>	0,758	0,641
	<b>F<sub>1</sub></b>	0,093	0,171	0,176	<u>0,371</u>	0,259	0,256	<b>0,436</b>

@N'ye karşılık gelen sütun, tavsiye sayısını göstermektedir. P, R ve  $F_1$ 'e karşılık gelen satırlar sırasıyla hassasiyet, hatırlama ve  $F_1$  değerini gösterir. Her bir @N değeri için verilen ölçütler bakımından, **koyu** sayılar en iyi başarıyı ve altu çizili sayılar en iyi ikinci başarıyı göstermektedir.

Önceki deneyin sonuçlarından farklı olarak, WRMF düşük @N değerlerinde daha iyi başarımlar göstermekte ve önerilen yöntem daha yüksek @N değerlerinde en iyi sonuçları vermektedir. Özellikle, önerilen yöntemin güçlü hassasiyet değerleri daha yüksek  $F_1$  değerlerine sahip olmasıyla sonuçlanmıştır. Yüksek hassasiyet ve yüksek hatırlama değerleri, uygun reklamların başarılı bir şekilde bulunduğunu ve uygun olmayan reklamların da filtrelenebileceğini göstermektedir. SLIM ve BPRMF birbirine yakın başarımlar göstermiştir. Önerilen yöntemin başarımları @N = 10'da WRMF'den %26, SLIM ve BPRMF'den ise %55 daha iyidir. Önceki deneyde olduğu gibi, UBCF ve IBCF benzer başarımlar göstermiştir ve başarımları diğer yöntemlerden daha düşüktür. Önerilen yöntem daha iyi başarımlar gösterse de istatistiksel olarak incelendiğinde UBCF ve IBCF yöntemlerinden anlamlı şekilde farklı çıkmakta, ancak WRMF, SLIM ve BPRMF ile bir farklılık göstermemektedir ( $\alpha=0.05$ , %99 güvenle). Önerilen yöntem daha yüksek bir başarımlar göstermesi bu deney için beklenen bir sonuçtu, çünkü kullanıcıların gezinme kalıpları hakkındaki bilgiler zengin olduğunda ve diğer yöntemler bu bilgilerden olumlu veya olumsuz yönde etkilenmediğinden dolayı, diğer yöntemlere karşı bir avantaja sahip olmaktadır.



Şekil 4.3. Deney 2 için hassasiyet, hatırlama ve  $F_1$  değerleri

#### 4.4.3. Deney 3: En az ziyarette bulunan 100 kullanıcı (gerçek dünya veri kümesi)

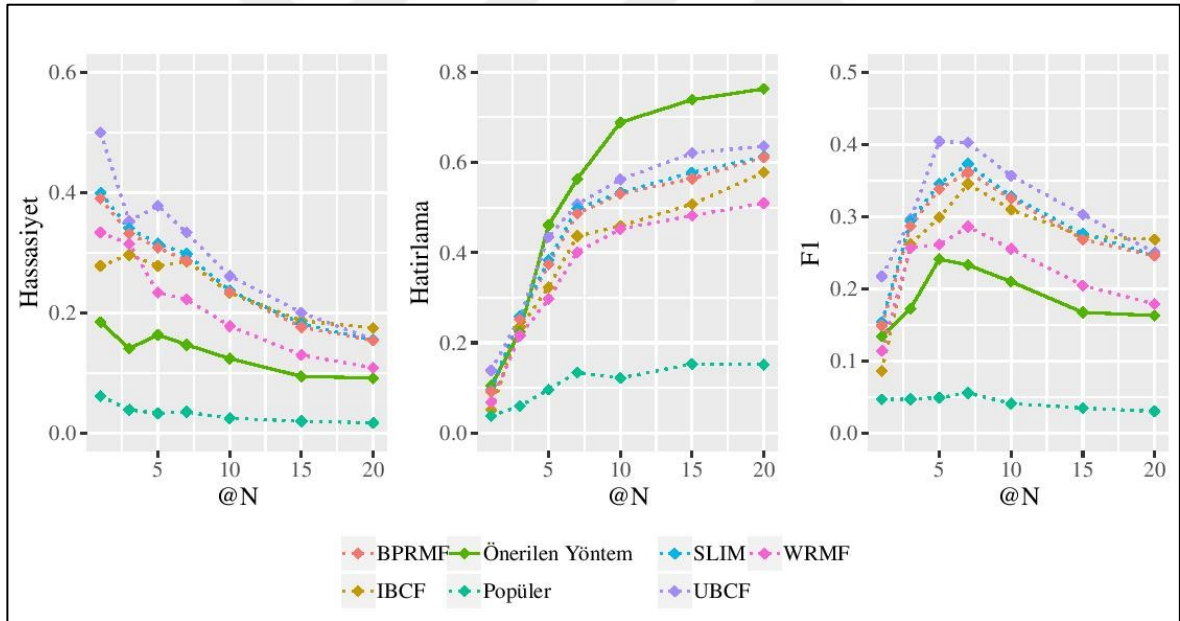
Gerçek dünya veri kümesi ile yapılan bu son deneyde ise en az ziyarete sahip 100 kullanıcıyı içeren bir alt küme seçilmiştir. Deneylerin sonuçları Çizelge 4.5’de gösterilmektedir. Şekil 4.4 farklı @N değerlerinde sırasıyla hassasiyet, hatırlama ve  $F_1$  değerlerini göstermektedir.

Çizelge 4.5. Deney 3 için karşılaştırılan yöntemlerin değerlendirme ölçütleri

@N	Ölçüt	Popüler	IBCF	UBCF	WRMF	BPRMF	SLIM	Önerilen Yöntem
<b>1</b>	<b>P</b>	0,061	0,278	<b>0,500</b>	0,333	0,390	<u>0,399</u>	0,184
	<b>R</b>	0,037	0,051	<b>0,139</b>	0,069	0,092	0,095	<u>0,105</u>
	<b>F<sub>1</sub></b>	0,046	0,086	<b>0,217</b>	0,114	0,149	<u>0,154</u>	0,134
<b>3</b>	<b>P</b>	0,038	0,296	<b>0,352</b>	0,315	0,332	<u>0,340</u>	0,140
	<b>R</b>	0,060	0,235	<u>0,256</u>	0,216	0,252	<b>0,259</b>	0,224
	<b>F<sub>1</sub></b>	0,047	0,262	<b>0,297</b>	0,257	0,286	<u>0,294</u>	0,172
<b>5</b>	<b>P</b>	0,033	0,278	<b>0,378</b>	0,233	0,308	<u>0,315</u>	0,163
	<b>R</b>	0,096	0,323	<u>0,435</u>	0,297	0,374	0,383	<b>0,461</b>
	<b>F<sub>1</sub></b>	0,049	0,299	<b>0,404</b>	0,261	0,338	<u>0,346</u>	0,241
<b>7</b>	<b>P</b>	0,035	0,286	<b>0,333</b>	0,222	0,287	<u>0,298</u>	0,147
	<b>R</b>	0,134	0,436	<u>0,507</u>	0,401	0,487	0,498	<b>0,564</b>
	<b>F<sub>1</sub></b>	0,055	0,345	<b>0,402</b>	0,286	0,361	<u>0,373</u>	0,233
<b>10</b>	<b>P</b>	0,024	0,233	<b>0,261</b>	0,178	0,235	<u>0,237</u>	0,124
	<b>R</b>	0,122	0,459	<u>0,562</u>	0,452	0,531	0,533	<b>0,689</b>
	<b>F<sub>1</sub></b>	0,041	0,309	<b>0,357</b>	0,255	0,325	<u>0,328</u>	0,210
<b>15</b>	<b>P</b>	0,019	<u>0,187</u>	<b>0,200</b>	0,130	0,176	0,182	0,094
	<b>R</b>	0,153	0,508	<u>0,621</u>	0,482	0,564	0,577	<b>0,739</b>
	<b>F<sub>1</sub></b>	0,035	0,273	<b>0,303</b>	0,204	0,268	<u>0,276</u>	0,167
<b>20</b>	<b>P</b>	0,017	<b>0,174</b>	<u>0,156</u>	0,108	0,154	0,154	0,091
	<b>R</b>	0,151	0,578	<u>0,636</u>	0,510	0,611	0,614	<b>0,763</b>
	<b>F<sub>1</sub></b>	0,030	<b>0,268</b>	<u>0,250</u>	0,179	0,246	0,247	0,163

@N'ye karşılık gelen sütun, tavsiye sayısını göstermektedir. P, R ve  $F_1$ 'e karşılık gelen satırlar sırasıyla hassasiyet, hatırlama ve  $F_1$  değerini gösterir. Her bir @N değeri için verilen ölçütler bakımından, **koyu** sayılar en iyi başarıyı ve altu çizili sayılar en iyi ikinci başarıyı göstermektedir.

UBCF genel olarak diğer yöntemlerden daha iyi başarımlar göstermiştir. Önerilen yöntem yüksek @N değerlerinde daha iyi bir hatırlamaya sahip olmasına rağmen; genel olarak en düşük hassasiyete sahiptir. Yüksek hatırlama ve düşük hassasiyetli değerler, uygun reklamların başarılı bir şekilde bulunduğunu ancak uygun olmayan reklamların verimli bir şekilde filtrelenemediğini, bu nedenle hedef kullanıcıya da önerildiğini gösterir. Bu, önerilen yöntemin diğer yöntemlerden daha kötü başarımlar göstermesini neden olmuştur. Bu sorunun arkasındaki kök neden, önerilen yöntemin kullanıcıların gezinme kalıpları hakkındaki bilgiler normalden daha az bilindiğinde diğer yöntemlere karşı olan avantajını yitmesidir. Bu deneyde kullanılan veri kümesi için önerilen yöntem istatistiksel olarak UBCF'den anlamlı bir şekilde farklı, ancak SLIM, IBCF, WRMF ve BPRMF ile farklı değildir; öte yandan, diğer tüm yöntemlerinde başarımları kendi aralarında anlamlı bir şekilde farklı değildir ( $\alpha=0.05$ , %99 güvenle).



Şekil 4.4. Deney 3 için hassasiyet, hatırlama ve F<sub>1</sub> değerleri

#### 4.4.4. Deney 4: Tüm veri kümesi (sentetik veri kümesi)

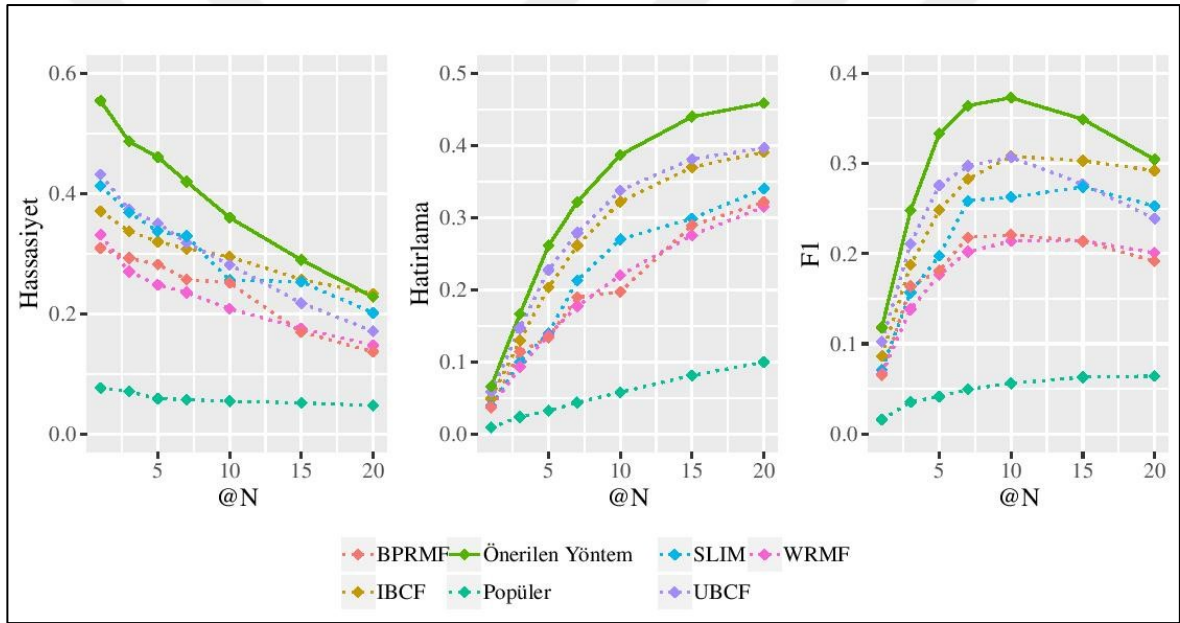
Bu deneyde ise sentetik veri kümesinin tamamı kullanılarak yöntemlerin başarımı ölçülmüştür. Deneylerin sonuçları Çizelge 4.6'da gösterilmektedir. Şekil 4.5 farklı @N değerlerinde sırasıyla hassasiyet, hatırlama ve  $F_1$  değerlerini göstermektedir.

Çizelge 4.6. Deney 4 için karşılaştırılan yöntemlerin değerlendirme ölçütleri

@N	Ölçüt	Popüler	IBCF	UBCF	WRMF	BPRMF	SLIM	Önerilen Yöntem
<b>1</b>	<b>P</b>	0,076	0,370	<u>0,431</u>	0,331	0,309	0,413	<b>0,554</b>
	<b>R</b>	0,009	0,049	<u>0,058</u>	0,037	0,036	0,039	<b>0,066</b>
	<b>F<sub>1</sub></b>	0,016	0,086	<u>0,102</u>	0,067	0,065	0,071	<b>0,118</b>
<b>3</b>	<b>P</b>	0,070	0,337	<u>0,373</u>	0,270	0,292	0,369	<b>0,486</b>
	<b>R</b>	0,024	0,130	<u>0,147</u>	0,093	0,114	0,100	<b>0,166</b>
	<b>F<sub>1</sub></b>	0,035	0,187	<u>0,211</u>	0,138	0,164	0,157	<b>0,248</b>
<b>5</b>	<b>P</b>	0,059	0,320	<u>0,350</u>	0,248	0,282	0,337	<b>0,460</b>
	<b>R</b>	0,032	0,203	<u>0,227</u>	0,137	0,133	0,140	<b>0,261</b>
	<b>F<sub>1</sub></b>	0,041	0,248	<u>0,276</u>	0,176	0,181	0,198	<b>0,333</b>
<b>7</b>	<b>P</b>	0,057	0,308	0,318	0,235	0,257	<u>0,329</u>	<b>0,420</b>
	<b>R</b>	0,044	0,261	<u>0,279</u>	0,177	0,189	0,213	<b>0,322</b>
	<b>F<sub>1</sub></b>	0,049	0,283	<u>0,297</u>	0,202	0,218	0,259	<b>0,364</b>
<b>10</b>	<b>P</b>	0,054	<u>0,295</u>	0,281	0,208	0,251	0,256	<b>0,360</b>
	<b>R</b>	0,058	0,322	<u>0,337</u>	0,220	0,197	0,270	<b>0,387</b>
	<b>F<sub>1</sub></b>	0,056	<u>0,308</u>	0,307	0,214	0,221	0,263	<b>0,373</b>
<b>15</b>	<b>P</b>	0,051	<u>0,257</u>	0,218	0,175	0,170	0,253	<b>0,289</b>
	<b>R</b>	0,081	0,370	<u>0,381</u>	0,276	0,289	0,299	<b>0,440</b>
	<b>F<sub>1</sub></b>	0,063	<u>0,303</u>	0,277	0,214	0,214	0,274	<b>0,349</b>
<b>20</b>	<b>P</b>	0,047	<b>0,233</b>	0,171	0,147	0,137	0,201	<u>0,228</u>
	<b>R</b>	0,099	0,391	<u>0,397</u>	0,315	0,321	0,340	<b>0,459</b>
	<b>F<sub>1</sub></b>	0,064	<u>0,292</u>	0,239	0,201	0,192	0,253	<b>0,304</b>

@N'ye karşılık gelen sütun, tavsiye sayısını göstermektedir. P, R ve  $F_1$ 'e karşılık gelen satırlar sırasıyla hassasiyet, hatırlama ve  $F_1$  değerini gösterir. Her bir @N değeri için verilen ölçütler bakımından, **koyu** sayılar en iyi başarımı ve altu çizili sayılar en iyi ikinci başarımı göstermektedir.

Önerilen yöntem,  $F_1$  değerinin yanı sıra, hassasiyet ve de hatırlama açısından diğer yöntemlerden daha iyi başarımlar göstermektedir. UBCF, ikinci en iyi başarımları gösteren yöntemdir. Önerilen yöntemin başarımları, @N=10'da UBCF'den %24 daha iyidir. Daha iyi başarımların bir nedeni sentetik verilerin üretilmesinde gerçek dünya veri kümesi göz önünde bulundurulmuş fakat kullanıcıların ilgilerinin dağılımı daha yumuşak bir şekilde dağıtılmıştır. Aslında, bu deneyin sonuçları *Deney 2* ile de tutarlıdır: her iki veri kümesinde de zengin gezinme geçmişi olan kullanıcılar vardır. Önerilen yöntem daha iyi başarımlar gösterse de, yöntemler arasındaki farklar istatistiksel olarak anlamlı değildir ( $\alpha=0.05$ , %99 güvenle).



Şekil 4.5. Deney 4 için hassasiyet, hatırlama ve  $F_1$  değerleri



#### 4.4.5. Deney 5: Uzun kuyruk (sentetik veri kümesi)

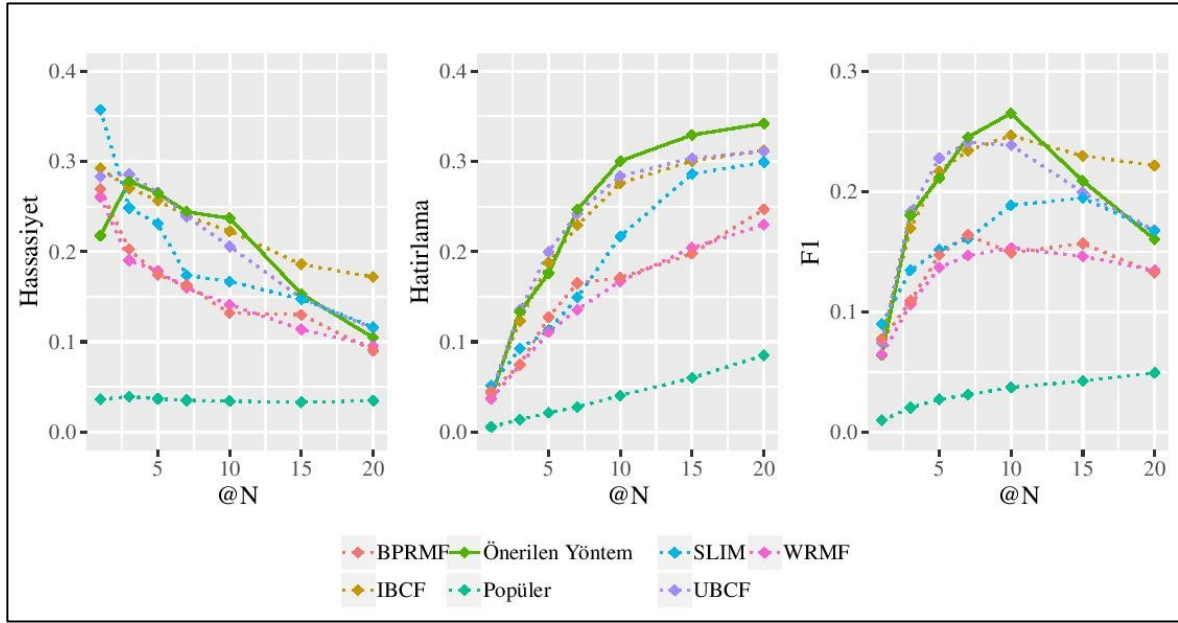
Bu son deneyde sentetik veri kümesinden en popüler olan reklamların çıkartılmasıyla elde edilmiş bir alt küme seçilmiştir. Deneylerin sonuçları Çizelge 4.7'de gösterilmektedir. Şekil 4.6 farklı @N değerlerinde sırasıyla hassasiyet, hatırlama ve  $F_1$  değerlerini göstermektedir.

Çizelge 4.7. Deney 5 için karşılaştırılan yöntemlerin değerlendirme ölçütleri

@N	Ölçüt	Popüler	IBCF	UBCF	WRMF	BPRMF	SLIM	Önerilen Yöntem
1	P	0,036	<u>0,293</u>	0,283	0,261	0,269	<b>0,357</b>	0,218
	R	0,006	0,044	0,043	0,037	<u>0,045</u>	<b>0,051</b>	0,038
	$F_1$	0,010	<u>0,077</u>	0,074	0,064	0,077	<b>0,090</b>	0,064
3	P	0,039	0,270	<b>0,286</b>	0,190	0,203	0,248	<u>0,278</u>
	R	0,014	0,123	<b>0,135</b>	0,074	0,074	0,092	<u>0,133</u>
	$F_1$	0,020	0,169	<b>0,183</b>	0,106	0,109	0,135	<u>0,180</u>
5	P	0,037	0,257	<b>0,265</b>	0,178	0,174	0,231	<u>0,265</u>
	R	0,021	<u>0,188</u>	<b>0,200</b>	0,111	0,127	0,113	0,176
	$F_1$	0,027	<u>0,217</u>	<b>0,228</b>	0,137	0,147	0,151	0,211
7	P	0,035	<u>0,240</u>	0,239	0,160	0,163	0,174	<b>0,244</b>
	R	0,028	0,229	<u>0,242</u>	0,136	0,165	0,149	<b>0,246</b>
	$F_1$	0,031	0,234	<u>0,241</u>	0,147	0,164	0,160	<b>0,245</b>
10	P	0,034	<u>0,223</u>	0,206	0,141	0,132	0,167	<b>0,237</b>
	R	0,040	0,276	<u>0,284</u>	0,167	0,171	0,217	<b>0,342</b>
	$F_1$	0,037	<u>0,246</u>	0,239	0,153	0,149	0,189	<b>0,280</b>
15	P	0,033	<b>0,186</b>	0,148	0,114	0,130	0,147	<u>0,153</u>
	R	0,060	0,300	<u>0,304</u>	0,204	0,198	0,286	<b>0,329</b>
	$F_1$	0,042	<b>0,230</b>	0,199	0,146	0,157	0,194	<u>0,209</u>
20	P	0,035	<b>0,172</b>	0,115	0,095	0,090	<u>0,116</u>	0,105
	R	0,085	<b>0,312</b>	<u>0,311</u>	0,230	0,247	0,299	0,300
	$F_1$	0,049	<b>0,222</b>	<u>0,168</u>	0,134	0,132	0,167	0,155

@N'ye karşılık gelen sütun, tavsiye sayısını göstermektedir. P, R ve  $F_1$ 'e karşılık gelen satırlar sırasıyla hassasiyet, hatırlama ve  $F_1$  değerini gösterir. Her bir @N değeri için verilen ölçütler bakımından, **koyu** sayılar en iyi başarıyı ve altu çizili sayılar en iyi ikinci başarıyı göstermektedir.

$F_1$  değeri bakımından en iyi başarımları sırasıyla IBCF ve önerilen yöntem göstermektedir. Genel olarak deneyin sonuçları incelendiğinde her bir @N değeri için farklı bir yöntem daha iyi başarımlar gösterdiği göze batmaktadır. @N=1 için, SLIM'in daha yüksek hassasiyet ve hatırlama değerleri vardır ama @N=3 ve 5 için bu durum UBCF için geçerlidir. Tavsiye sayısı arttıkça, önerilen yöntem ve IBCF daha iyi başarımlar göstermeye başlamaktadır. Sonuç olarak, yöntemler arasında istatistiksel olarak anlamlı bir fark yoktur ( $\alpha=0.05$ , %99 güvenle).



Şekil 4.6. Deney 5 için hassasiyet, hatırlama ve  $F_1$  değerleri

#### 4.5. Bölüm Sonucu

Bu bölümde öncelikle değerlendirmede kullanılan ölçütler tanıtılmış ve önerilen yöntemin başarımlarını ölçmek için kurgulanan deneysel kurulumun detaylarına yer verilmiştir.

Önerilen yöntem de dahil olmak üzere toplam yedi yöntemin başarımları, iki farklı veri kümesi üzerinden beş farklı deney ile ölçülmüştür.

Hem gerçek dünya hem de sentetik veri kümelerinin kullanıldığı deneylere ek olarak, sentetik veri kümesinden popüler reklamlar arındırılarak uzun kuyruk veri alt kümesi, gerçek dünya veri kümesinden ise en çok ve en az ziyarette bulunan 100 kullanıcıyı içeren alt veri kümeleri hazırlanmış ve ardından önerilen yöntem işbirlikçi filtreleme yöntemleri ile

kıyaslanmıştır.

Çizelge 4.8, en iyi ve ikinci en iyi başarımları gösteren yöntemler ve önerilen yöntem ile en iyi başarımları gösteren yöntem arasındaki başarımları farkı açısından deney sonuçlarını özetlemektedir. Önerilen yöntem en iyi başarımları sergiledi ise, ikinci en iyi yöntemle göre karşılaştırılmıştır.

Çizelge 4.8. Tüm deneylerde en iyi başarımları gösteren yöntemlerin karşılaştırılması

Deney	En iyi yöntem (Başarımları)	İkinci en iyi yöntem (Başarımları)	Önerilen Yöntemin Başarımları	Başarımları Farkı (%)
<b>D1 – Tüm veri kümesi (G<sup>6</sup>)</b>	BPRMF (0,3692)	SLIM (0,3650)	0,3306	<b>-10</b>
<b>D2 – En çok ziyarette bulunan 100 kullanıcı (G)</b>	Önerilen Yöntem (0,3765)	WRMF (0,3622)	0,3765	<b>+4</b>
<b>D3 – En az ziyarette bulunan 100 kullanıcı (G)</b>	UBCF (0,3185)	SLIM (0,2881)	0,1885	<b>-41</b>
<b>D4 – Tüm veri kümesi (S<sup>7</sup>)</b>	Önerilen Yöntem (0,2985)	UBCF (0,2440)	0,2985	<b>+22</b>
<b>D5 – Uzun kuyruk (S)</b>	IBCF (0,1993)	Önerilen Yöntem (0,1921)	0,1921	<b>-4</b>

Gerçek dünya veri kümesinin tümünün kullanıldığı ilk deneyde önerilen yöntemin başarımları en yakın komşu tabanlı yaklaşımlar ve matris ayrıştırma modelleri arasında gözükmemektedir.

En çok ziyarette bulunan 100 kullanıcı ile yapılan deneyde ise önerilen yöntemin başarımları WRMF'den %4 daha iyidir. Önerilen yöntem daha iyi başarımları sahip olsa da WRMF, SLIM ve BPRMF ile istatistiksel olarak anlamlı bir farklılık göstermemektedir.

<sup>6</sup> Gerçek dünya veri kümesi

<sup>7</sup> Sentetik veri kümesi

En az ziyarette bulunan 100 kullanıcı ile yapılan deneyde ise UBCF en iyi başarımı gösterirken, önerilen yöntem yüksek @N değerlerinde daha iyi bir hatırlamaya sahip olmasına rağmen; genel olarak en düşük hassasiyete sahiptir. Bu da kötü bir başarı sergilemesine neden olmuştur.

Sentetik veri kümesinin tamamının kullanıldığı deneyde ise önerilen yöntem, hassasiyet ve de hatırlama açısından diğer yöntemlerden daha iyi başarı göstermiş ve en iyi ikinci başarıyı gösteren UBCF'den yaklaşık %22 daha yüksek bir başarı sergilemiştir.

Uzun kuyruk deneyinde ise her bir @N değeri için farklı bir yöntem daha iyi başarı göstermiş fakat yöntemler arasında istatistiksel olarak anlamlı bir fark görülemedi.

Tüm bu deneyler, önerilen yöntemin, kullanıcıların gezinme kalıpları hakkındaki bilgilerin zengin olduğu veri kümelerinde bir avantaja sahip olduğunu ve diğer yöntemlere göre daha üstün bir başarı gösterdiğine işaret etmektedir. Aksi durumda ise önerilen yöntem bu avantajını kaybetmekte ve başarımı düşmektedir. Yine de önerilen yöntemin *verilen=0* olarak değerlendirilmesine rağmen diğer yöntemlere eş başarı sergilemesi dikkat çekmektedir.

Son bölümde genel sonuçlar ve öneriler yer almaktadır.

## 5. GENEL SONUÇLAR VE ÖNERİLER

Bu çalışmada geleneksel hedefleme yöntemlerinin ötesinde etkili bir şekilde reklam gösterimini sağlamak amacıyla bir yöntem önerilmiş ve bu yöntemin kullanıldığı bir tavsiye sistemi tanımlanmıştır. İşbirlikçi filtreleme yöntemleri ve önerilen yöntem çevrimiçi reklamcılık veri kümeleri üzerinde uygulanarak, başarımları hesaplanmış, uygulanan yöntemler birbirleri ile kıyaslanmış ve sonuçlar yorumlanmıştır.

Çalışmanın ilk bölümünde kavramsal çerçeve ve ilgili araştırmalara yer verilmiştir. Hem Türkiye’de hem de Dünya’da büyük bir pazara ulaşan çevrimiçi reklamcılıkta hedefleme, reklam gösterimlerinin etkin bir şekilde yapılabilmesi ve değerli hale gelmesi için uygulanan yöntemlerin başında gelmektedir. Hedeflemenin en önemli adımı ise kullanıcı profilleri oluşturmaktır. Kullanıcı profilleri doğrudan ya da dolaylı yollarla oluşturulabilir. Bu çalışmada kullanılan verilerin web kullanım madenciliği teknikleriyle işlenmesi ile, kullanıcı profilleri dolaylı yollarla oluşturulabilmiştir. Web kullanım madenciliğinin diğer yöntemlere göre üstünlüğü, kullanıcı davranışlarını yansıtmasıdır. İkili kümeleme diğer kümeleme yöntemlerinin belirlemesinin zor olduğu benzersiz desenleri bulabilmektedir. Bu kümelenecek nesnelere (kullanıcıların) ve de niteliklerinin (ilgi kategorileri) eş bir şekilde göz önüne alınarak kümeleme işlemlerinin gerçekleştirilmesi sayesinde olmaktadır. Bir tavsiye sistemi, bilgi toplama, öğrenme ve ardından tavsiye ya da tahmin evrelerini gerçekleştirerek, kullanıcıya kişileştirilmiş önerilen öğeler listesi sunmaktadır. İşbirlikçi filtreleme yöntemleri tavsiyelerin diğer kullanıcıların görüşlerine dayalı olarak yapılmasına imkân sağlar. Tavsiye değerinin zenginleştirilmesinde kullanılan yöntemlerden biri olan alanlar arası tavsiye sistemleri ise farklı alanlardaki veri ya da tavsiyeleri göz önüne almaktadır. Tavsiye sistemlerinde uygulama alanına ve amaca göre farklı birleştirme operatörleri kullanılabilir. Genel bir karar fonksiyonu oluşturmak için kullanılan OWA birleştirme operatörlerini diğer operatörlerden farklı kılan özelliği ise yeniden sıralama adımıdır. Sıralamadan bağımsız olan ağırlıklar ise dilsel niceleyiciler kullanılarak hesaplanabilmektedir.

Çevrimiçi reklamcılık veri kümesi kullanıcıların gezinti kayıtlarını ve de kullanıcıların reklamlarla ilişkilerini içeren reklam görüntüleme ile reklam tıklama kayıtlarını tutan günlük dosyalarından oluşmaktadır. Çalışmanın uygulama bölümünde kullanılan veriler,

çevrimiçi reklamcılık veri kümesindeki kayıtların işlenmesi ve üçüncü parti servisler yardımıyla zenginleştirilmesinin ardından elde edilmektedir. Bu servisler kullanıcılar için anlamlı bir profil oluşturabilmek kullanıcının yapmış olduğu aktiviteler yardımıyla zenginleştirme işlemini yapmaktadır.

Önerilen yöntem üç ana aşamadan oluşmaktadır. İlk aşama kullanıcılara ait gezinti verilerinden yararlanarak kullanıcı profillerinin oluşturulmasıdır. İkinci aşama olan, kullanıcı profilleri yardımıyla ikili kümeleme işlemlerinin yapıldığı, kullanıcı ve ilgi kümelerinin bulunması aşamasını, son olarak hedef kullanıcıya reklam önerilerinin sunulduğu tavsiye aşaması takip etmektedir.

İşbirlikçi filtreleme yöntemleri ve de önerilen yöntem, kullanıcıların bireysel geçmişlerini dikkate alan içeriğe dayalı hedefleme ya da davranışsal hedeflemenin aksine, farklı kullanıcıların tıklamış olduğu reklamları birbirlerine etkin bir şekilde önermesini hedeflemektedir. Çalışmanın bir diğer amacı ise yeni kullanıcılara da reklam önerebilmektir. İşbirlikçi filtreleme yöntemlerinde kullanıcıların birbirlerine olan benzerlikleri etkileştikleri reklamlar üzerinden hesaplanmaktadır ya da öge tabanlı sistemlerde birbirlerine benzer reklamlar dikkate alınmaktadır. Çalışmada önerilen yöntem ise diğerlerinden farklı olarak, kullanıcıların ve ilgi alanlarının (ilgi kategorileri) ikili kümelenmesine odaklanmaktadır. Böylelikle hiçbir reklam ile etkileşmemiş (ya da çok az etkileşime girmiş) kullanıcılara bile reklam önerisinde bulunabilmektedir.

Son olarak, önerilen yöntem diğer işbirlikçi filtreleme yöntemleri ile karşılaştırılmaktadır. Yöntemlerin başarımlarını gösteren hassasiyet, hatırlama ve bu iki değeri birleştirerek tek bir değerde tanımlayan  $F_1$  ölçütleri, farklı deneysel kurulumlar için hesaplanmıştır. Her bir deney farklı  $N$  değerlerine (tavsiye sayısı) karşı  $k$  katlamalı çapraz doğrulama yöntemi kullanılarak koşulmuştur.

Çalışmada elde edilen bulgular ışığında;

- İşbirlikçi filtreleme yöntemlerinin reklam tavsiyesinde etkili bir araç olarak kullanılabilceği tekrar gözlemlenmiş,

- Kullanıcıların gezinti geçmişlerine ait kayıtların zengin olması durumunda, önerilen yöntemin diğer yöntemlere göre ciddi bir üstünlük gösterdiği,
- Önerilen yöntemin hiçbir reklam ile etkileşmemiş kullanıcılar için bile diğer yöntemlerin reklamlarla etkileşen kullanıcılar için oluşturabildiği tavsiyelere eş tavsiyeler oluşturabileceği

sonuçlarına ulaşılmıştır.

Özetle, bu çalışma kullanıcıların gezinti geçmişleri bilindiğinde, reklamları işbirlikçi bir şekilde tavsiye etmenin mümkün olduğu fikrini tazelemektedir. Bu çalışmanın bulguları, önerilen yöntemin gezinti geçmişi bakımından zengin veri kümelerine sahip sistemlerde yüksek başarımlar göstereceği, diğer sistemlerde işbirlikçi filtreleme yöntemleri kadar etkili olacağını göstermektedir.

## 5.1. Öneriler

Çevrimiçi reklamcılık da dahil olmak üzere birçok pazarlama alanında da kullanılabilecek tavsiye sistemleri olası fırsatları satışa çevirmek için güçlü bir araç olarak değerlendirilmelidir. Makine öğrenmesi temelli olmaları nedeniyle tavsiye sistemlerinin

- Daha net kişileştirmeye imkân tanınması sayesinde, müşteriye iletilmek istenen mesajları (reklam, promosyon vb.) basit yeniden hedeflemenin ötesinde daha yüksek bir hassasiyetle sunabilmesi,
- İlgili bir mal veya hizmet hakkında gerçek zamanlı olarak reklam ögesi sunabilmesi, böylelikle tüketici ilgisi yüksekken yatırımın geri dönüşümünü sağlayabilmesi,
- Geleneksel yöntemlerden farklı olarak gizli kalmış ilgi alanlarına yönelik reklamları belirleyebilmesi,
- Reklam önerilerini zaman içinde geliştiren daha akıllı bir yapıya sahip olması durumunda, pazarlama kampanyalarının verimliliğini arttırabilmesi

özellikleri ön plana çıkmaktadır. Bu özellikler de dikkate alındığında, önerilen yöntemin gelişmiş bir reklam dağıtım sisteminin bir parçası olarak kullanılması durumunda

sağlayacağı faydanın daha yüksek olması beklenmektedir. Çevrimiçi reklamcılık ekosisteminin paydaşlarına göre tahmin edilen bu faydalardan bazıları

- *Reklam verenler için:*
  - Ön görülemeyen kullanıcı kümelerine erişimin sağlanması
  - Etkisiz reklam gösterimlerinin azaltılarak yatırım getirisinin yükseltilmesi
  - Başarılı ve etkili reklam kampanyaları sayesinde işletme itibarının artması
- *Yayıncılar için:*
  - Etkili reklam gösterimleri sayesinde kısa süre gelir elde edebilme
  - Etkili reklamlar sayesinde yayın ortamının (web sitesi, mobil uygulama vb.) itibarının yükselmesi
  - Kullanıcıların yayın ortamını reklam engelleyici eklentilerin listesine eklemesini engelleme
- *Kullanıcılar için:*
  - Gerçekten ilgilendikleri reklamlarla karşılaşma
  - Farkında olmadıkları fakat ihtiyaç duyabilecekleri mal ve hizmetler hakkında bilgi sahibi olmaları
- *Reklam dağıtımıcılar için:*
  - Daha etkili hedefleme ve reklam dağıtımı sayesinde rekabet gücü
  - Pazar payını artırma olanağı
  - Kullanıcıların reklam dağıtımıcıya güven kazanması ve böylelikle reklam gösterimlerinden ayrılmaması

olarak sıralanabilir.

Bu çalışma yalnızca sınırlı bir veri kümesi üzerinde gerçekleştirilebilmiştir. Çalışmanın sonuçları çevrimdışı olarak doğrulanmıştır. Gelecekteki çalışmalarda, daha büyük bir veri kümesinin çevrimiçi bir değerlendirmesini yapmak mümkündür. Çevrimiçi bir değerlendirme yapılması durumunda reklam kampanyalarının etkinliğine ek olarak yukarıda maddeler halinde verilmiş faydalar da detaylı bir şekilde incelenebilir.



Bir diđer nokta ise, önerilen yöntem, kullanıcı profillerini oluştururken demografik özellikleri dikkate almamaktadır. Bu çalışmada kullanılan veri kümesinin boyutu kullanıcıların demografik özellikleri arasında bir fark oluşmasına imkân vermemiştir. Dolayısıyla daha kapsamlı bir veri kümesi ile çalışılması durumunda bu da mümkün olacaktır.

Ayrıca, kullanıcı profilleri oluşturulurken zaman etkisinin de göz önüne alınması daha güçlü kullanıcı profilleri elde edilmesine olanak sağlayacaktır. Benzer şekilde bu çalışmada yine veri kümesinin zaman aralığı nedeniyle bu etki dikkate alınamamıştır.

Son olarak, önerilen yöntemin farklı dilsel niceleyicilerle kullanılmasının, tavsiye sistemlerinin karşılaştığı uzun kuyruk soruna farklı bir çözüm olup olmayacağı ilerleyen çalışmalar kapsamında araştırılabilir.



## KAYNAKLAR

- Adeniyi, D., Wei, Z. ve Yongquan, Y., (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method, *Applied Computing and Informatics*, 12 (1), 90-108.
- Adomavicius, G. ve Kwon, Y., (2007). New recommendation techniques for multicriteria rating systems, *IEEE Intelligent Systems*, 22 (3).
- 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.
- Agrawal, R., Gehrke, J., Gunopulos, D. ve Raghavan, P., (1998). *Automatic subspace clustering of high dimensional data for data mining applications*: ACM.
- Ahmed, A., Low, Y., Aly, M., Josifovski, V. ve Smola, A.J., (2011). Scalable distributed inference of dynamic user interests for behavioral targeting. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 114-122. ACM.
- Alqadah, F., Reddy, C.K., Hu, J. ve Alqadah, H.F., (2015). Biclustering neighborhood-based collaborative filtering method for top-n recommender systems, *Knowledge and Information Systems*, 44 (2), 475-491.
- Anand, S.S. ve Mobasher, B., (2003). Intelligent techniques for web personalization. *Proceedings of the 2003 international conference on Intelligent Techniques for Web Personalization*, 1-36. Springer-Verlag.
- Asdemir, K., Kumar, N. ve Jacob, V.S., (2012). Pricing models for online advertising: CPM vs. CPC, *Information Systems Research*, 23 (3-part-1), 804-822.
- Beliakov, G., Calvo, T. ve James, S., (2011). *Aggregation of Preferences in Recommender Systems*. In Ricci, F., Rokach, L. ve Shapira, B. (Eds.), *Recommender systems handbook*. Boston, MA: Springer, 705-734.
- Ben-Dor, A., Chor, B., Karp, R. ve Yakhini, Z., (2003). Discovering local structure in gene expression data: the order-preserving submatrix problem, *Journal of computational biology*, 10 (3-4), 373-384.
- Bergmann, S., Ihmels, J. ve Barkai, N., (2003). Iterative signature algorithm for the analysis of large-scale gene expression data, *Physical review E*, 67 (3), 031902.
- Berkhin, P., (2006). A survey of clustering data mining techniques, *Grouping multidimensional data*, 25, 71.
- Bobadilla, J., Ortega, F., Hernando, A. ve Gutiérrez, A., (2013). Recommender systems survey, *Knowledge-based systems*, 46, 109-132.

- Bozdağ, D., Kumar, A.S. ve Catalyurek, U.V., (2010). Comparative analysis of biclustering algorithms. *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology*, 265-274. ACM.
- Broder, A.Z., (2008). Computational advertising and recommender systems. *Proceedings of the 2008 ACM conference on Recommender systems*, 1-2. ACM.
- Burke, R., (2002). Hybrid recommender systems: Survey and experiments, *User modeling and user-adapted interaction*, 12 (4), 331-370.
- Busygin, S., Prokopyev, O. ve Pardalos, P.M., (2008). Biclustering in data mining, *Computers & Operations Research*, 35 (9), 2964-2987.
- Carlsson, C. ve Fullér, R., (2018). *Maximal Entropy and Minimal Variability OWA Operator Weights: A Short Survey of Recent Developments*. In Collan, M. ve Kacprzyk, J. (Eds.), *Soft Computing Applications for Group Decision-making and Consensus Modeling*: Springer, 187-199.
- Chen, G., Cox, J.H., Uluagac, A.S. ve Copeland, J.A., (2016). In-Depth Survey of Digital Advertising Technologies, *IEEE Communications Surveys & Tutorials*, 18 (3), 2124-2148.
- Cheng, Y. ve Church, G.M., (2000). Biclustering of expression data. *Ismb*, 93-103.
- Chiclana, F., Herrera-Viedma, E., Herrera, F. ve Alonso, S., (2007). Some induced ordered weighted averaging operators and their use for solving group decision-making problems based on fuzzy preference relations, *European Journal of Operational Research*, 182 (1), 383-399.
- 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, J. ve Lee, G., (2009). New techniques for data preprocessing based on usage logs for efficient web user profiling at client side. *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on*, 54-57. IEEE.
- Dave, K. ve Varma, V., (2014). Computational advertising: Techniques for targeting relevant ads, *Foundations and Trends® in Information Retrieval*, 8 (4-5), 263-418.
- Delgado, M., Verdegay, J.L. ve Vila, M.A., (1993). On aggregation operations of linguistic labels, *International journal of intelligent systems*, 8 (3), 351-370.
- İnternet: DoubleClick, (2016). DoubleClick by Google. URL: <https://www.doubleclickbygoogle.com/>, Son Erişim Tarihi: 2016/06/01.
- Eirinaki, M. ve Vazirgiannis, M., (2003). Web mining for web personalization, *ACM Transactions on Internet Technology (TOIT)*, 3 (1), 1-27.

- Ester, M., Kriegel, H.-P., Sander, J. ve Xu, X., (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, 226-231.
- Evans, D.S., (2009). The online advertising industry: Economics, evolution, and privacy, *Journal of Economic Perspectives*, 23 (3), 37-60.
- Fabricio, O., Ferreira, H.M. ve Von Zuben, F.J., (2007). Applying biclustering to perform collaborative filtering. *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*, 421-426. IEEE.
- Facca, F.M. ve Lanzi, P.L., (2005). Mining interesting knowledge from weblogs: a survey, *Data & Knowledge Engineering*, 53 (3), 225-241.
- Fernández-Tobías, I., Cantador, I., Kaminskas, M. ve Ricci, F., (2012). Cross-domain recommender systems: A survey of the state of the art. *Spanish Conference on Information Retrieval*, 24. sn.
- Fodor, J., Marichal, J.-L. ve Roubens, M., (1995). Characterization of the ordered weighted averaging operators, *IEEE Transactions on Fuzzy Systems*, 3 (2), 236-240.
- Fullér, R. ve Majlender, P., (2001). An analytic approach for obtaining maximal entropy OWA operator weights, *Fuzzy Sets and Systems*, 124 (1), 53-57.
- Fullér, R. ve Majlender, P., (2003). On obtaining minimal variability OWA operator weights, *Fuzzy Sets and Systems*, 136 (2), 203-215.
- İnternet: Gagolewski, M. ve Cena, A., (2014). agop: Aggregation Operators package for R. URL: <http://agop.rexamine.com/>, Son Erişim Tarihi: 2018/12/17.
- Gao, S., Luo, H., Chen, D., Li, S., Gallinari, P. ve Guo, J., (2013). Cross-domain recommendation via cluster-level latent factor model. *Joint European conference on machine learning and knowledge discovery in databases*, 161-176. Springer.
- Getz, G., Levine, E. ve Domany, E., (2000). Coupled two-way clustering analysis of gene microarray data, *Proceedings of the National Academy of Sciences*, 97 (22), 12079-12084.
- Goldfarb, A., (2014). What is different about online advertising?, *Review of Industrial Organization*, 44 (2), 115-129.
- Goldfarb, A. ve Tucker, C., (2011a). Online display advertising: Targeting and obtrusiveness, *Marketing Science*, 30 (3), 389-404.
- Goldfarb, A. ve Tucker, C.E., (2011b). Privacy regulation and online advertising, *Management science*, 57 (1), 57-71.
- Golemati, M., Katifori, A., Vassilakis, C., Lepouras, G. ve Halatsis, C., (2007). Creating an ontology for the user profile: Method and applications. *Proceedings of the first RCIS conference*, 407-412.
- Grabisch, M., Marichal, J., Mesiar, R. ve Pap, E., (2009). *Aggregation Functions*.

- Hafri, Y., Djeraba, C., Stanchev, P. ve Bachimont, B., (2003). A markovian approach for web user profiling and clustering. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 191-202. Springer.
- İnternet: Hahsler, M., (2017). recommenderlab: Lab for Developing and Testing Recommender Algorithms. URL: <https://CRAN.R-project.org/package=recommenderlab>, Son Erişim Tarihi: 2018/12/17.
- Han, J., Pei, J. ve Kamber, M., (2011). *Data mining: concepts and techniques*: Elsevier.
- Hartigan, J.A., (1972). Direct clustering of a data matrix, *Journal of the american statistical association*, 67 (337), 123-129.
- Herlocker, J.L., Konstan, J.A. ve Riedl, J., (2000). Explaining collaborative filtering recommendations. *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 241-250. ACM.
- Herlocker, J.L., Konstan, J.A., Terveen, L.G. ve Riedl, J.T., (2004). Evaluating collaborative filtering recommender systems, *ACM Transactions on Information Systems (TOIS)*, 22 (1), 5-53.
- Herrera, F. ve Herrera-Viedma, E., (1997). Aggregation operators for linguistic weighted information, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27 (5), 646-656.
- Herrera, F., Herrera-Viedma, E. ve Verdegay, J., (1996). Direct approach processes in group decision making using linguistic OWA operators, *Fuzzy Sets and systems*, 79 (2), 175-190.
- Hinneburg, A. ve Keim, D.A., (1999). Optimal grid-clustering: Towards breaking the curse of dimensionality in high-dimensional clustering. *25th International Conference on Very Large Data Bases*.
- Hoffman, D.L. ve Novak, T.P., (2000). Advertising pricing models for the world wide web, *Internet publishing and beyond: The economics of digital information and intellectual property*, 5, 2.
- Hoppe, A., Nicolle, C. ve Roxin, A., (2013). Automatic ontology-based user profile learning from heterogeneous web resources in a big data context, *Proceedings of the VLDB Endowment*, 6 (12), 1428-1433.
- Hu, J., Zeng, H.-J., Li, H., Niu, C. ve Chen, Z., (2007). Demographic prediction based on user's browsing behavior. *Proceedings of the 16th international conference on World Wide Web*, 151-160. ACM.
- Hu, Y., Koren, Y. ve Volinsky, C., (2008). Collaborative Filtering for Implicit Feedback Datasets. *ICDM*, 263-272. Citeseer.
- İnternet: IAB, (2018a). IAB internet advertising revenue report. URL: [https://www.iab.com/wp-content/uploads/2018/05/IAB-2017-Full-Year-Internet-Advertising-Revenue-Report.REV\\_.pdf](https://www.iab.com/wp-content/uploads/2018/05/IAB-2017-Full-Year-Internet-Advertising-Revenue-Report.REV_.pdf), Son Erişim Tarihi: 2018/08/01.

- İnternet: IAB, (2018b). IAB Türkiye AdEx 2017. URL: <https://www.iabturkiye.org/iab-turkiye-2017-yili-dijital-reklam-yatirimlarini-acikladi>, Son Erişim Tarihi: 2018/08/01.
- Ignatov, D.I., Kuznetsov, S.O. ve Poelmans, J., (2012a). Concept-based biclustering for internet advertisement. *IEEE 12th International Conference on Data Mining Workshops (ICDMW)*, 123-130. IEEE.
- Ignatov, D.I., Poelmans, J. ve Zaharchuk, V., (2012b). Recommender system based on algorithm of bicluster analysis RecBi, *arXiv preprint arXiv:1202.2892*.
- Isinkaye, F., Folajimi, Y. ve Ojokoh, B., (2015). Recommendation systems: Principles, methods and evaluation, *Egyptian Informatics Journal*, 16 (3), 261-273.
- Jin, X., Zhou, Y. ve Mobasher, B., (2004). Web usage mining based on probabilistic latent semantic analysis. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 197-205. ACM.
- İnternet: Kaiser, S., Santamaria, R., Khamiakova, T., Sill, M., Theron, R., Quintales, L., Leisch, F. ve De Troyer, E., (2015). biclust: BiCluster Algorithms. R package version 1.2. 0. Son Erişim Tarihi: 2018/12/17.
- Kaski, S., (1998). Dimensionality reduction by random mapping: Fast similarity computation for clustering. *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*, 413-418. IEEE.
- Kazienko, P. ve Adamski, M., (2007). AdROSA—Adaptive personalization of web advertising, *Information Sciences*, 177 (11), 2269-2295.
- Kluger, Y., Basri, R., Chang, J.T. ve Gerstein, M., (2003). Spectral biclustering of microarray data: coclustering genes and conditions, *Genome research*, 13 (4), 703-716.
- Kohavi, R., (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, 1137-1145. Montreal, Canada.
- Kosala, R. ve Blockeel, H., (2000). Web mining research: A survey, *ACM Sigkdd Explorations Newsletter*, 2 (1), 1-15.
- Košir, D., Kononenko, I. ve Bosnić, Z., (2014). Web user profiles with time-decay and prototyping, *Applied intelligence*, 41 (4), 1081-1096.
- İnternet: Kuhn, M., (2017). caret: Classification and Regression Training. URL: <https://CRAN.R-project.org/package=caret>, Son Erişim Tarihi: 2018/12/17.
- Lazzeroni, L. ve Owen, A., (2002). Plaid models for gene expression data, *Statistica sinica*, 61-86.
- Levene, M., (2011). *An introduction to search engines and web navigation*: John Wiley & Sons.

- Li, B., (2011). Cross-domain collaborative filtering: A brief survey. *2011 23rd IEEE International Conference on Tools with Artificial Intelligence*, 1085-1086. IEEE.
- Li, B., Yang, Q. ve Xue, X., (2009). Transfer learning for collaborative filtering via a rating-matrix generative model. *Proceedings of the 26th annual international conference on machine learning*, 617-624. ACM.
- Liu, X., (2007). The solution equivalence of minimax disparity and minimum variance problems for OWA operators, *International Journal of Approximate Reasoning*, 45 (1), 68-81.
- Lü, L., Medo, M., Yeung, C.H., Zhang, Y.-C., Zhang, Z.-K. ve Zhou, T., (2012). Recommender systems, *Physics Reports*, 519 (1), 1-49.
- Madeira, S.C. ve Oliveira, A.L., (2004). Biclustering algorithms for biological data analysis: a survey, *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 1 (1), 24-45.
- Malheiros, M., Jennett, C., Patel, S., Brostoff, S. ve Sasse, M.A., (2012). Too close for comfort: A study of the effectiveness and acceptability of rich-media personalized advertising. *Proceedings of the SIGCHI conference on human factors in computing systems*, 579-588. ACM.
- Mangani, A., (2004). Online advertising: Pay-per-view versus pay-per-click, *Journal of Revenue and Pricing Management*, 2 (4), 295-302.
- Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H. ve Koper, R., (2011). *Recommender systems in technology enhanced learning*. In Ricci, F., Rokach, L. ve Shapira, B. (Eds.), *Recommender systems handbook*: Springer, 387-415.
- Markov, Z. ve Larose, D.T., (2007). *Data mining the Web: uncovering patterns in Web content, structure, and usage*: John Wiley & Sons.
- Mele, I., (2013). Web usage mining for enhancing search-result delivery and helping users to find interesting web content. *Proceedings of the sixth ACM international conference on Web search and data mining*, 765-770. ACM.
- Mobasher, B., (2007). *Data mining for web personalization*. In Brusilovski, P., Alfred, K. ve Wolfgang, N. (Eds.), *The adaptive web*: Springer, 90-135.
- Mobasher, B., Cooley, R. ve Srivastava, J., (2000a). Automatic personalization based on web usage mining, *Communications of the ACM*, 43 (8), 142-151.
- Mobasher, B., Dai, H., Luo, T. ve Nakagawa, M., (2001). Improving the effectiveness of collaborative filtering on anonymous web usage data. *Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWP01)*, 53-61.
- Mobasher, B., Dai, H., Luo, T. ve Nakagawa, M., (2002). Discovery and evaluation of aggregate usage profiles for web personalization, *Data mining and knowledge discovery*, 6 (1), 61-82.



- Mobasher, B., Dai, H., Luo, T., Sun, Y. ve Zhu, J., (2000b). Integrating web usage and content mining for more effective personalization, *Electronic commerce and web technologies*, 165-176.
- Mulvenna, M.D., Anand, S.S. ve Büchner, A.G., (2000). Personalization on the Net using Web mining: introduction, *Communications of the ACM*, 43 (8), 122-125.
- Murali, T. ve Kasif, S., (2003). Extracting conserved gene expression motifs from gene expression data. *Pacific symposium on biocomputing*, 77-88.
- Murtagh, F., (1983). A survey of recent advances in hierarchical clustering algorithms, *The Computer Journal*, 26 (4), 354-359.
- Nasraoui, O., Frigui, H., Krishnapuram, R. ve Joshi, A., (2000). Extracting web user profiles using relational competitive fuzzy clustering, *International journal on artificial intelligence tools*, 9 (04), 509-526.
- Nasraoui, O., Soliman, M., Saka, E., Badia, A. ve Germain, R., (2008). A web usage mining framework for mining evolving user profiles in dynamic web sites, *IEEE transactions on knowledge and data engineering*, 20 (2), 202-215.
- Ning, X. ve Karypis, G., (2011). Slim: Sparse linear methods for top-n recommender systems. *2011 IEEE 11th International Conference on Data Mining*, 497-506. IEEE.
- O'Hagan, M., (1988). Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic. *Twenty-Second Asilomar Conference on Signals, Systems and Computers*, 681-689. IEEE.
- İnternet: OpenDNS, (2016). OpenDNS Domain Tagging. URL: <https://community.opendns.com/domaintagging/>, Son Erişim Tarihi: 2016/06/01.
- Pan, R., Zhou, Y., Cao, B., Liu, N.N., Lukose, R., Scholz, M. ve Yang, Q., (2008). One-class collaborative filtering. *2008 Eighth IEEE International Conference on Data Mining*, 502-511. IEEE.
- Pazzani, M.J., (1999). A framework for collaborative, content-based and demographic filtering, *Artificial intelligence review*, 13 (5-6), 393-408.
- Pennock, D.M., Horvitz, E., Lawrence, S. ve Giles, C.L., (2000). Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach. *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, 473-480. Morgan Kaufmann Publishers Inc.
- Plummer, J., Rappaport, S.D., Hall, T. ve Barocci, R., (2007). *The online advertising playbook: Proven strategies and tested tactics from the advertising research foundation*: John Wiley & Sons.
- Prelić, A., Bleuler, S., Zimmermann, P., Wille, A., Bühlmann, P., Gruissem, W., Hennig, L., Thiele, L. ve Zitzler, E., (2006). A systematic comparison and evaluation of biclustering methods for gene expression data, *Bioinformatics*, 22 (9), 1122-1129.

- Qiu, F. ve Cho, J., (2006). Automatic identification of user interest for personalized search. *Proceedings of the 15th international conference on World Wide Web*, 727-736. ACM.
- İnternet: R Core Team, (2013). R: A language and environment for statistical computing. URL: <https://www.r-project.org/>, Son Erişim Tarihi: 2019/2/25.
- Raad, E., Chbeir, R. ve Dipanda, A., (2010). User profile matching in social networks. *13th International Conference on Network-Based Information Systems (NBIS)*, 297-304. IEEE.
- Rajan, S., (2017). The Evolution of Computational Advertising. *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, 99-99. ACM.
- Rendle, S., Freudenthaler, C., Gantner, Z. ve Schmidt-Thieme, L., (2009). BPR: Bayesian personalized ranking from implicit feedback. *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, 452-461. AUAI Press.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. ve Riedl, J., (1994). GroupLens: an open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 175-186. ACM.
- Ricci, F., Rokach, L. ve Shapira, B., (2011). *Introduction to recommender systems handbook*. In Ricci, F., Rokach, L. ve Shapira, B. (Eds.), *Recommender systems handbook*: Springer, 1-35.
- Richardson, M., Dominowska, E. ve Ragno, R., (2007). Predicting clicks: estimating the click-through rate for new ads. *Proceedings of the 16th international conference on World Wide Web*, 521-530. ACM.
- Rosenkrans, G., (2007). *Online advertising metrics*: Idea Group Reference Hershey, London, PA.
- Saito, T. ve Rehmsmeier, M., (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, *PloS one*, 10 (3), e0118432.
- Sang, X. ve Liu, X., (2014). An analytic approach to obtain the least square deviation OWA operator weights, *Fuzzy Sets and Systems*, 240, 103-116.
- Sarwar, B., Karypis, G., Konstan, J. ve Riedl, J., (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th international conference on World Wide Web*, 285-295. ACM.
- Schafer, J.B., Frankowski, D., Herlocker, J. ve Sen, S., (2007). *Collaborative filtering recommender systems*. In Brusilovski, P., Kobsa, A. ve Nejdl, W. (Eds.), *The adaptive web*: Springer, 291-324.
- Schiaffino, S. ve Amandi, A., (2009). *Intelligent user profiling*. In Bramer, M. (Eds.), *Artificial Intelligence An International Perspective*: Springer, 193-216.

- Shankar, V. ve Balasubramanian, S., (2009). Mobile marketing: a synthesis and prognosis, *Journal of interactive marketing*, 23 (2), 118-129.
- İnternet: SmartInsights, (2018). Average display advertising clickthrough rates. URL: <https://www.smartinsights.com/internet-advertising/internet-advertising-analytics/display-advertising-clickthrough-rates/>, Son Erişim Tarihi: 2018/10/11.
- Soltysiak, S. ve Crabtree, I., (1998). Automatic learning of user profiles—towards the personalisation of agent services, *BT Technology Journal*, 16 (3), 110-117.
- Spiliopoulou, M., (2000). Web usage mining for web site evaluation, *Communications of the ACM*, 43 (8), 127-134.
- Srivastava, J., Cooley, R., Deshpande, M. ve Tan, P.-N., (2000). Web usage mining: Discovery and applications of usage patterns from web data, *Acm Sigkdd Explorations Newsletter*, 1 (2), 12-23.
- Steck, H., (2013). Evaluation of recommendations: rating-prediction and ranking. *Proceedings of the 7th ACM conference on Recommender systems*, 213-220. ACM.
- Sugiyama, K., Hatano, K. ve Yoshikawa, M., (2004). Adaptive web search based on user profile constructed without any effort from users. *Proceedings of the 13th international conference on World Wide Web*, 675-684. ACM.
- Symeonidis, P., Nanopoulos, A. ve Manolopoulos, Y., (2008a). Providing justifications in recommender systems, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38 (6), 1262-1272.
- Symeonidis, P., Nanopoulos, A. ve Manolopoulos, Y., (2009). MoviExplain: a recommender system with explanations. *Proceedings of the third ACM conference on Recommender systems*, 317-320. ACM.
- Symeonidis, P., Nanopoulos, A., Papadopoulos, A.N. ve Manolopoulos, Y., (2008b). Nearest-biclusters collaborative filtering based on constant and coherent values, *Information Retrieval*, 11 (1), 51-75.
- Tanay, A., Sharan, R. ve Shamir, R., (2002). Discovering statistically significant biclusters in gene expression data, *Bioinformatics*, 18 (suppl\_1), S136-S144.
- Tanay, A., Sharan, R. ve Shamir, R., (2005). Biclustering algorithms: A survey, *Handbook of computational molecular biology*, 9 (1-20), 122-124.
- Tucker, C.E., (2014). Social networks, personalized advertising, and privacy controls, *Journal of Marketing Research*, 51 (5), 546-562.
- Turow, J., King, J., Hoofnagle, C.J., Bleakley, A. ve Hennessy, M., (2009). Americans reject tailored advertising and three activities that enable it, *Available at SSRN 1478214*.
- Varnagar, C.R., Madhak, N.N., Kodinariya, T.M. ve Rathod, J.N., (2013). Web usage mining: a review on process, methods and techniques. *International Conference on Information Communication and Embedded Systems (ICICES)*, 40-46. IEEE.

- Wang, C., Zhang, P., Choi, R. ve D'Eredita, M., (2002). Understanding consumers attitude toward advertising, *AMCIS 2002 Proceedings*, 158.
- Wang, Y.-M., Luo, Y. ve Liu, X., (2007). Two new models for determining OWA operator weights, *Computers & Industrial Engineering*, 52 (2), 203-209.
- Wang, Y.-M. ve Parkan, C., (2005). A minimax disparity approach for obtaining OWA operator weights, *Information Sciences*, 175 (1), 20-29.
- Xu, R. ve Wunsch, D., (2005). Survey of clustering algorithms, *IEEE Transactions on neural networks*, 16 (3), 645-678.
- Xu, Z., (2005). An overview of methods for determining OWA weights, *International journal of intelligent systems*, 20 (8), 843-865.
- Xu, Z. ve Da, Q., (2002). The ordered weighted geometric averaging operators, *International Journal of Intelligent Systems*, 17 (7), 709-716.
- Yager, R.R., (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking, *IEEE Transactions on systems, Man, and Cybernetics*, 18 (1), 183-190.
- Yager, R.R., (1993). Families of OWA operators, *Fuzzy sets and systems*, 59 (2), 125-148.
- Yager, R.R., (2017). OWA aggregation of multi-criteria with mixed uncertain satisfactions, *Information Sciences*, 417, 88-95.
- Yager, R.R. ve Alajlan, N., (2016). Some issues on the OWA aggregation with importance weighted arguments, *Knowledge-Based Systems*, 100, 89-96.
- Yager, R.R. ve Filev, D.P., (1999). Induced ordered weighted averaging operators, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29 (2), 141-150.
- Yager, R.R. ve Kacprzyk, J., (2012). *The ordered weighted averaging operators: theory and applications*: Springer Science & Business Media.
- Yager, R.R., Kacprzyk, J. ve Beliakov, G., (2011). *Recent developments in the ordered weighted averaging operators: theory and practice*: Springer Science & Business Media.
- Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y. ve Chen, Z., (2009). How much can behavioral targeting help online advertising? *Proceedings of the 18th international conference on World wide web*, 261-270. ACM.
- Zhang, K. ve Katona, Z., (2012). Contextual advertising, *Marketing Science*, 31 (6), 980-994.
- Zhen, Y., Li, W.-J. ve Yeung, D.-Y., (2009). TagiCoFi: tag informed collaborative filtering. *Proceedings of the third ACM conference on Recommender systems*, 69-76. ACM.

Zhu, Y.-Q. ve Chang, J.-H., (2016). The key role of relevance in personalized advertisement: Examining its impact on perceptions of privacy invasion, self-awareness, and continuous use intentions, *Computers in Human Behavior*, 65, 442-447.





## ÖZGEÇMİŞ

### Kişisel Bilgiler

Soyadı, adı : YOLDAR, Mehmet Türkay  
Uyruğu : T.C.  
Telefon : 0 (505) 700 25 00  
e-posta : mehmetturkay.yoldar@gazi.edu.tr

Eğitim Derecesi	Okul/Program	Mezuniyet Yılı
Doktora	Gazi Üniversitesi Yönetim Bilişim Sistemler	2019
Yüksek lisans	Hacettepe Üniversitesi Temiz Tükenmez Enerjiler	2010
Lisans	Anadolu Üniversitesi İşletme	2012
Lisans	Hacettepe Üniversitesi Fizik Mühendisliği	2007
Lise	Çankaya Ömer Seyfettin Lisesi	2000

### Yabancı Dil

İngilizce



**GAZİLİ OLMAK AYRICALIKTIR**