

# **İÇERİK BAZLI MEDYA TAKİP VE HABER TAVSİYE SİSTEMİ**

## **CONTENT BASED MEDIA TRACKING AND NEWS RECOMMENDATION SYSTEM**

**Servet TAŞCI**

**PROF. DR. İLYAS ÇİÇEKLI**

**Tez Danışmanı**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin Bilgisayar Mühendisliği  
Anabilim Dalı için öngördüğü YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2015

## KABUL VE ONAY

Servet TAŞCI tarafından hazırlanan "**İçerik Bazlı Medya Takip ve Haber Tavsiye Sistemi**" başlıklı bu çalışma aşağıdaki jüri tarafından **Bilgisayar Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** başarılı bulunarak jürimiz tarafından yüksek lisans tezi olarak kabul edilmiştir.

Prof. Dr. Ferda Nur ALPASLAN

Başkan

Prof. Dr. İlyas ÇİÇEKLİ

Danışman

Prof. Dr. Hayri SEVER

Üye

Yrd. Doç. Dr. Ayça TARHAN

Üye

Yrd. Doç. Dr. Gönenç ERCAN

Üye

Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **YÜKSEK LİSANS TEZİ** olarak onaylanmıştır.

Prof. Dr. Fatma Sevin DÜZ

Fen Bilimleri Enstitüsü Müdürü

## ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi kullanılan verilerde herhangi bir tahrifat yapmadığımı
- Ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

11/06/2015

SERVET TAŞCI

## ÖZET

### İÇERİK BAZLI MEDYA TAKİP VE HABER TAVSİYE SİSTEMİ

**Servet TAŞCI**

**Yüksek Lisans, Bilgisayar Mühendisliği**

**Tez Danışmanı: Prof. Dr. İlyas ÇİÇEKLI**

**Haziran 2015, 70 sayfa**

İnternetin hayatımıza yoğun bir şekilde girmesiyle birlikte yapısal olmayan verilerin özellikle metinsel verilerin miktarı çok fazla artmıştır. Kullanıcıların bu verilere erişim noktasının internet kaynakları olduğunu düşünürsek, bu kaynakların güvenilirliği, doğruluğu gibi konular da ön plana çıkmaktadır. Kaynakların çokluğunun yanında farklı kaynaklarda benzer içeriklerle bulunduğunu da göz önünde bulundurursak, bu kaynaklar arasından kısa zamanda sadece ihtiyaç duyulan metnin okunması son derece zor olmaktadır. Erişilen kaynağın da geçekten ihtiyaç duyulan bilgiye barındırıyor olması ve bunun kullanıcı gözüyle teyit ediliyor olması gerekmektedir. Tavsiye sistemleri bu anda devreye girerek kullanıcıların farklı özelliklerini göz önünde bulundurarak erişilen içerik ile kullanıcı arasında ilişki kurmakta ve içeriği belirli kriterlere göre değerlendirerek, içerik kullanıcıya özgü ise kullanıcıya tavsiye etmektedir. Tavsiye sistemleri ilk kullanılmaya başladığında basit anlamda içerik filtreleme özellikleri kullanılmaktaydı fakat günümüz itibarıyla bu sistemler çok daha karmaşık hesaplamalar ve algoritmalar kullanarak kullanıcıların birçok özellikleri ile veri arasında ilişki kurmaya çalışmaktadır. Bu da tavsiye sistemlerinin karar destek sistemleri olarak kullanılması haline dönüşmesini sağlamıştır.

Bu tez metin halinde bulunan haber kaynaklarından verilerin elde edilmesi, verinin sınıflandırılması, özetlenmesi ve haberlerin kullanıcının özellikleri ile ilişkilendirilerek tavsiye etmeyi amaçlamıştır. Tavsiye sistemleri tavsiye ederken başlıca üç metot olan içerik-bazlı filtreleme, işbirlikçi filtreleme ve karma filtrelemeyi kullanmaktadır. Bizim sistemimizde içerik-bazlı filtreleme kullanılmıştır.

**Anahtar Kelimeler:** Tavsiye sistemi, işbirlikçi filtreleme, içerik bazlı filtreleme, sınıflandırma algoritmaları, metin özetleme

## **ABSTRACT**

# **CONTENT BASED MEDIA TRACKING AND NEWS RECOMMENDATION SYSTEM**

**Servet TAŞCI**

**Graduate School, Computer Engineering**

**Supervisor: Prof. Dr. İlyas ÇİÇEKLİ**

**June 2015, 70 Pages**

With the increasing use of Internet in our life, amount of unstructured data, and particularly amount of textual data, has increased dramatically. Thinking that the access point of users to this data is Internet, reliability and accuracy of these resources stands out as a concern. Besides multitude of resources, most resources have similar content and it is quite challenging to read only the needed news among these resources in a short time. It is also needed that accessed resource really includes the required information and that it is confirmed by the user. Recommender systems assess different characteristics of the users and correlate the accessed content and user and then evaluate the content according to the specific criteria and recommends to the user. First recommender systems were using simple content filtering features, but current systems use much more complicated calculations and algorithms and try to correlate many characteristics of users and the data. These improvements allowed usage of recommender systems as decision support systems.

This thesis aims at getting data from textual news resources, classification of data, summarization, and recommend the news by correlating the news with the characteristics of users. Basically, recommender systems mainly use three

methods: content-based filtering, cooperative filtering, and mixed filtering. In our system, content-based filtering is used.

**Keywords:** Recommendation Systems, Content-based Filtering, Collaborative Filtering, Text Summarization, Text Classification, Classification Algorithms, Keyphrase Extraction

## TEŐEKKÜR

Tez alıőmamın her aőamasında deęerli katkı ve eleőtirileriyle yol gosteren, sonsuz sabırla beni her zaman alıőmaya teővik eden ve guven veren Sayın Prof. Dr. İlyas İEKLİ' ye, jüri üyelerine ve her zaman yanımda olan eőime ve kızlarıma içtenlikle teőekkür ederim.



# İÇİNDEKİLER

## Sayfa

ÖZET .....	i
ABSTRACT .....	iii
TEŞEKKÜR .....	v
İÇİNDEKİLER .....	vi
ŞEKİLLER .....	viii
TABLolar .....	ix
SİMGELER VE KISALTMALAR .....	x
<b>1. GİRİŞ .....</b>	<b>1</b>
1.1. Tezin Amacı .....	1
1.2. Tezin Kapsamı .....	2
1.3. Tez Taslağı .....	3
<b>2. TAVSİYE SİSTEMLERİ .....</b>	<b>5</b>
2.1. Tavsiye Sistemleri Nedir ve Nasıl Çalışır .....	5
2.2. Tavsiye Sistemlerinin Amacı .....	7
2.3. Tavsiye Sistemleri .....	8
2.3.1. Tavsiye Sistemlerinde Kullanılan Metotlar .....	8
2.3.2. Tavsiye Sistemi Sınıflandırma Yaklaşımları .....	16
2.4. Haber Tavsiye Sistemleri .....	20
2.5. Doküman Bazlı Tavsiye Sistemleri .....	20
2.5.1. Dokümanların Yapısallaştırılması .....	20
2.5.2. Benzerlik Ölçme Yöntemleri ve Sınıflandırma Algoritmaları .....	21
2.5.3. Dokümanların Tavsiye Sistemlerinde Etkin Kullanımı .....	27
<b>3. BENZER ÇALIŞMALAR .....</b>	<b>28</b>
3.1. Doküman Tavsiye Sistemleri .....	28
3.2. Haber Tavsiye Sistemleri .....	31
<b>4. HABER ANALİZ SİSTEMİ .....</b>	<b>33</b>
4.1. Genel Anlatım .....	33
4.2. Zaman Bazlı (Time-based) Tavsiye Sistemi .....	36
4.2.1. Haber Sitelerinden Haberin Çekilmesi .....	36
4.2.2. Sınıflandırılma İşlemleri ve Haber Sınıflandırması .....	37
4.2.3. Özetleme Sistemleri ve Haber Özetlenmesi .....	40

4.2.4.	Profil Oluřturma .....	43
4.2.5.	Zaman-bazlı Tavsiye Sistemi Algoritması .....	45
4.3.	<i>Anahtar Kelime Bazlı Tavsiye Sistemi</i> .....	47
<b>5.</b>	<b>Tavsiye Sistemi Deęerlendirme Yöntemleri</b> .....	<b>52</b>
5.1.	<i>Deęerlendirme Metrikleri</i> .....	52
5.1.1.	Test Ortamı Oluřturulması.....	52
5.1.2.	Deęerlendirme Metrikleri .....	53
5.1.3.	Problem Sahaları.....	54
5.2.	<i>Tavsiye Sistemi Deęerlendirmesi</i> .....	56
5.2.1.	Zaman-bazlı tavsiye Sisteminde Optimum Haber Sayısının Tespiti .....	57
5.3.	<i>İki Tavsiye Sisteminin Karşılaştırılması</i> .....	59
<b>6.</b>	<b>SONUÇ VE GELECEK ÇALIřMALAR</b> .....	<b>60</b>
6.1.	<i>Sonuç</i> .....	60
6.2.	<i>Gelecek Çalıřmalar</i> .....	61
<b>KAYNAKLAR</b>	.....	<b>62</b>
<b>EKLER</b>	.....	<b>68</b>
<b>ÖZGEÇMİř</b>	.....	<b>70</b>

## ŞEKİLLER

	<u>Sayfa</u>
Şekil 2.1 Tuzhilin'in Sınıflandırması.....	17
Şekil 2.2 Burke'in Sınıflandırması.....	17
Şekil 2.3 Asanov'un Sınıflandırması.....	19
Şekil 2.4 Örnek Pearson Korelasyonu.....	23
Şekil 2.5 Support Vector Machine .....	26
Şekil 4.1 Haberanalizi.org Sitesinde Haberlerin Gösterilmesi.....	34
Şekil 4.2 Program Çalışması Görünümü .....	35
Şekil 4.3 Örneklem Sınıflandırma Sonuçları.....	40
Şekil 4.4 Bir Haber ve Özeti Görünümü .....	42
Şekil 4.5 Örnek Profil Ekranı .....	44
Şekil 4.6 Örnek Bir Tavsiye Profili .....	45
Şekil 4.7 Haber Tavsiye Sistemi Çalışma Şeması.....	46
Şekil 4.8 Anahtar Kelime Tavsiye Sistemi Akış Şeması .....	49
Şekil 4.9 Haber Benzerlik sonuçları.....	50
Şekil 5.1 Haber Sayısı İçin ROC Analizi .....	58

## TABLULAR

	<b><u>Sayfa</u></b>
Tablo 2.1 İşbirlikçi Filtreleme Teknikleri.....	10
Tablo 4.1 İş Akışı Şematik Gösterimi .....	33
Tablo 4.2 Örnek Haber Siteleri XPATH İfadeleri .....	37
Tablo 4.3 Sınıflandırma Algoritması Doğruluk (Accuracy) Sonuçları.....	39
Tablo 5.1 Precision-Recall.....	53
Tablo 5.2 Tavsiye Değerlendirme Sonuçları (20000 haber için).....	56
Tablo 5.3 Optimum Haber Sayısı Tespiti.....	58
Tablo 5.4 Zaman Bazlı Tavsiye Sistemi Değerlendirmesi .....	59
Tablo 5.5 Keyphrase Bazlı Tavsiye Sistemi Değerlendirmesi .....	59

## SİMGELER VE KISALTMALAR

### Kısaltmalar

CF	Collaborative Filtering (İşbirlikçi Filtreleme)
CBF	Content Based Filtering (İçerik Bazlı Filtreleme)
IR	Information Retrieval (Bilgi Erişim)
TF-IDF	Term Frequency- Inverse Document Frequency (Dönem Frekans-Ters Belge Frekansı)
ROC	Receiver Operatin Characteristics (Alıcı Çalışma Özellikleri)
MAE	Mean Absolute Error (Ortalama Mutlak Hata)
RMSE	Root Mean Absolute Error (Köksel Ortalama Mutlak Hata)
SVD	Single Value Decomposition (Tek Değer Ayrışma)
VSM	Vector Space Model (Vektör Uzay Modeli)

# 1. GİRİŞ

## 1.1. Tezin Amacı

İnternetin yaygınlaşması ile birlikte kullanıcıların aradığı konuyu çok fazla kaynak arasından seçmek zorunda olması ve alternatiflerinin artması problemin temelini oluşturmaktadır. Bu noktada devreye tavsiye sistemleri girmektedir. Tavsiye sistemleri kullanıcılara tavsiyelerde bulunurken aynı zamanda kullanıcıları belli kaynaklara yönlendirme işlemi de gerçekleştirmektedir ki bu da onun ticari önemini arttırmıştır. Daha detaylı olarak belirtmek gerekirse, tavsiye sistemlerinin amacının kullanıcıların aradığı ve ihtiyaç duyduğu bilgilerin bir an önce kullanıcıya doğru bir şekilde getirmesi olarak özetlenebilir [1].

Tavsiye sistemleri, kullanıcıların ilgilendikleri konu ve takip ettikleri olaylara dayanarak bilgi elde eder ve kullanıcıya özgü yeni ilgili konuları kullanıcıya sunmaya çalışır. İnternet bu kaynakların tespitinde kullanılacak en etkili araçtır. Bu kaynaklar her geçen gün statik ve dinamik içeriğiyle birlikte katlanarak artmaktadır. Sonuç olarak kullanıcılara istediği bilgileri filtreleyerek ulaştırmak ve gereksiz diyebileceğimiz veriden kurtarmak, tavsiye sistemlerinin başlıca amacıdır.

Tavsiye sistemleri, internet sitelerini semantik olarak indeksleme ve içeriklerini ayrıştırma gibi çeşitli yöntemler kullanarak bu sorunu çözmeye çalışmışlardır. En temel seviyede, filtreleme özellikleri kullanılsa da günümüzde bu çok daha karmaşık hesaplamaları beraberinde getirmiştir ve günümüzde veri madenciliği gibi birçok farklı teknolojiyi de içinde barındırır [2].

İlk tavsiye sistemleri genel olarak sorgu temelli çalışmaktaydı ve içerik-bazlı filtreleme basit seviyede kullanılmaktaydı [3]. Günümüzde, ihtiyaçlar arttığından daha karmaşık sistemlere ihtiyaç duyulmaktadır. Kullanıcı modelleme ve kişiselleştirilmiş bilgiyi kullanıcının hizmetine sunma son derece önem kazanmıştır. Çoğu arama motorları da bu soruna çözüm olabilmek adına indekslenmiş veriyi kullanıcıya sunmakla birlikte kullanıcıya tavsiye etme yoluna da gitmişlerdir. Bu konuda başarılı çalışmalar yapılmasına karşın ciddi denebilecek bir başarı da tam olarak sağlanamamıştır. Her ne kadar arama motorları bilginin filtrelenmesi ve sıralanması için kullanılsa da [4] araştırma sonuçları arama motorlarında yapılan

aramaların neticesinde kullanıcıların çoğu zaman ilk birkaç sayfada aramayı sonlandırdığını göstermiştir. Kullanıcılara hızlı kararlar alıracak seçenekler sunulması tavsiye sistemleri açısından önemlidir.

Haber tavsiye sistemleri de kullanıcılara farklı kaynaklardan haber sunmaya çalışır. Sunulan haberler arasından kullanıcı kendisine hitap eden haberleri öncelikle okuyacaktır. Hatta kendisine sunulan haberlerin hepsini kullanıcının okuması, haber tavsiye sisteminin başarısını göstermektedir. Haber analiz sistemi de bu kapsamda yapılmış olan Türkçe bir haber tavsiye sistemidir. Kullanıcılara kendisine hitap edecek haberler sunulmaya çalışılmış ve haber tavsiyesi kullanıcının dinamik olarak oluşan profiline bağlanmıştır.

## 1.2. Tezin Kapsamı

Bu tezin kapsamını belirtilirken tavsiye sistemlerinin genel çalışma esasları ve sistemimiz ile arasındaki ilişki belirtilecektir. Genel olarak sistemimiz, farklı haber sitelerinden haberlerin çekilmesi, haberlerin özetlenmesi, kullanıcıların tercih ve arayışlarının belirlenmesi, profillerinin oluşturulması ve güncel olarak tutulması, kullanıcılardan gelen geri beslemelere bağlı olarak uygun haber tavsiyelerinin yapılabilmesini kapsamaktadır. Benzer haberlerin tavsiye edilmemesi, tavsiye sisteminde anahtar kelimelerin daha etkin olarak kullanılması için de çalışmalar yapılmıştır.

Tavsiye sistemlerinin uygulanması üzerine birçok çalışma yapılmış ve bu çalışmalar iki paradigma altında toplanmıştır. İçerik tabanlı filtreleme<sup>1</sup>; kullanıcının geçmişte ilgilendiği nesnelere dayanarak, o nesnelere en yakın nesnelere tavsiye edilmesi esasına dayanır. İşbirlikçi filtreleme<sup>2</sup> ise kullanıcıya en yakın / benzer kullanıcıların incelediği nesnelere dayanarak kullanıcıya tavsiyede bulunmayı esas alır. İki paradigmanın olumlu ve olumsuz özellikleri bulunmaktadır. Karma<sup>3</sup> tavsiye sistemi bu esnada devreye girerek iki paradigmayı birleştirmiş ve olumlu yönlerini kullanarak tavsiyelerde bulunmayı esas almıştır. Bu iki paradigmanın uygulanması esnasında veri madenciliği, doğal dil işleme, sınıflandırma teknikleri gibi birçok

---

<sup>1</sup> İng : Content-based Filtering

<sup>2</sup> İng : Collaborative Filtering

<sup>3</sup> İng : Hybrid

yöntem de kullanılmaktadır. Bu tez çalışmasında içerik tabanlı filtreleme kullanılmıştır.

### **1.3. Tez Taslağı**

Tez altı ana kısımdan oluşmaktadır. İlk kısım giriş kısmıdır ve giriş kısmında tavsiye sistemlerinin çıkış sebepleri, neden ihtiyaç duyulduğu ve hangi aşamalardan geçerek bugünlere ulaştığı anlatılmıştır. Tavsiye sistemlerinde hangi yöntemlerin kullanıldığı bu kısımda verilmiştir ve bu yöntemler detaylı olarak sonraki kısımlarda anlatılmıştır. Daha sonra da bu tez kapsamında geliştirilen haber tavsiye sisteminin hangi yöntemle ele alındığı işlenmiştir.

İkinci kısımda tavsiye sistemleri daha detaylı olarak ele alınmış, tavsiye sistemlerinin amacı irdelenmiş, kullanılan yöntemler detaylı olarak tartışılmıştır. Genel olarak kullanılan içerik tabanlı tavsiye sistemleri ile işbirlikçi filtreleme tabanlı tavsiye sistemleri yanında tavsiye sistemlerinde kullanılan diğer yöntemler de burada anlatılmıştır. Ayrıca tavsiye sistemleri haber tavsiye sistemleri ve doküman tabanlı tavsiye sistemleri olarak iki grupta incelenmiş ve bu iki grup arasındaki bağımlılıklar ve ayrılık teşkil eden hususlar da yine bu kısımda anlatılmıştır. Tavsiye sistemlerine ilişkin farklı gruplamalar da bu kısımda şematik olarak belirtilmiştir.

Üçüncü kısımda, gerçekleştirilen benzer çalışmalar, çalışmaların olumlu ve olumsuz tarafları ele alınmıştır. Benzer çalışmalar haber tavsiye sistemleri ve diğer tavsiye sistemleri olarak iki grup altına irdelenmiş ve yine bu iki grup altında incelenmesi gerekliliği de bu kısımda anlatılmıştır. İncelenen benzer çalışmalar ile tavsiye sistemimiz aralarındaki ilişkiler de bu kısımda vurgulanmış olumlu ve olumsuz yönleri anlatılmaya çalışılmıştır.

Dördüncü kısımda, tez kapsamında geliştirilen haber tavsiye sistemine kısaca bir giriş yapılmış ve daha sonra üzerinde çalışılmış olan zaman bazlı tavsiye sistemi ile anahtar kelime kullanılarak uygulanan tavsiye sistemi anlatılmıştır. Zaman bazlı tavsiye sistemi safhalara ayrılarak her bir safhası detaylı olarak incelenmiştir. Bu safhaları kısaca şu şekilde belirtebiliriz; haber sitelerinden haberlerin çekilmesi, haberlerin ortak bir sınıflandırma algoritması ile sınıflandırılması, özetleme sistemi, kullanıcıların dinamik ve statik profillerini oluşturan tavsiye sistemi algoritması,



anahtar kelimelerin bulunması, anahtar kelime bazlı tavsiye sisteminin anlatılması ve zaman bazlı tavsiye sistemi ile anahtar kelime bazlı tavsiye sisteminin karşılaştırılması.

Beşinci kısımda tavsiye sistemlerinin değerlendirilmesini sağlayan metrikler anlatılmış, bu metrikler ışığında tavsiye sistemimiz değerlendirilmiş ve çıkarılan sonuçlar bu kısımda ele alınmıştır. Karşılaşılan sorunlar ve problem sahaları da bu kısımda belirtilmiştir.

Altıncı kısımda ise çıkarılan sonuçlar ve beklenen gelişmeler neticesinde tavsiyeler ortaya konulmuştur.

## 2. TAVSİYE SİSTEMLERİ

### 2.1. Tavsiye Sistemleri Nedir ve Nasıl Çalışır

Günümüzde pek çok ürünle ilgili o kadar çok kaynak vardır ki (bu kaynakları sadece internet veya basılı yayınlar olarak değil her çeşit ürünü kapsayacak geniş bir yelpazede düşünmek gerekir) bu kaynaklar arasında tüketiciler veya kullanıcılar ihtiyacına uygun olan tercihleri yapmakta zorlanmaktadır. Tam bu noktada, karar vermeyi kolaylaştıracak karar destek sistemleri devreye girmektedir. Bu karar destek sistemleri tavsiye sistemleri olarak ta adlandırılabilir. Başka bir tanım da *Paul Resnick* ve *Hal R. Varian* tarafından 1997' de yapılmıştır. "Tipik bir tavsiye sisteminde insanlar tavsiyeleri girdi olarak sunar ve sistem yorumlayarak alıcılara ulaştırır." [5]

Bu kısa tanımdan sonra, tavsiye sistemleri, daha dar anlamda internet kaynaklarının oluşturduğu veri setleri içinde uygun tavsiyelerin yapıldığı ve veri setlerinin daraltıldığı sistem olarak değerlendirilecektir. Bir tavsiye sistemi genel olarak üç ana öğeden oluşmaktadır. Bunlar; veri seti, tavsiye motoru ve kullanıcı arayüzü olarak düşünülebilir. Veri setleri istenilen özellikteki metinleri ve kullanıcının belirlenen eğilim ve tercihlerini kapsamaktadır. Tavsiye motoru, veri seti içinden kullanıcıya hitap eden en önemli makalelerin kullanıcıya tavsiye edilmesini sağlayan sistem olarak düşünülebilir ve tavsiye motoru tavsiye sisteminin beynidir. Arayüz ise tavsiyelerin kullanıcılara gösterildiği ve geri beslemelerin alındığı programdır.

Tavsiye sistemlerini fonksiyonel olarak detaylandırırsak genel olarak üç kısımdan oluştuğunu söylemek mümkündür;

***İçerik Analizi*<sup>4</sup>** ; İçerik analizi anlamlandırılmamış ve özellikleri üzerinden yapılandırılmamış dokümanı anlamlandırarak sonuç çıkarılabilecek bir hale getirme işlemi olarak tanımlanabilir. Doküman tabanlı sistemler olarak düşünürsek metinlerin işlenmesi ve yapısal hale getirilmesi olarak özetlenebilir.

***Profil Oluşturucu*<sup>5</sup>**; Profil oluşturucu sistemin kalbi konumundadır. Kullanıcılardan alınan olumlu ve olumsuz geri dönüşleri kullanarak kullanıcının

---

<sup>4</sup> Ing : Content Analyzer

<sup>5</sup> Ing : Profile Learner

profilinin oluşmasını sağlar. Geri beslemeler net veya bulanık olabileceği gibi ender olarak otomatik olarak ta alınabilir. Lee ve Park [6] 'ın tasarladığı ticari sistemde olduğu gibi satın alım süresi ve ürün üzerinde geçirilen süre göz önüne alınarak ta tespit edilebilir. Kullanıcıdan tavsiye edilen nesneye oylama, beğenme veya yorum yaparak geri dönüş elde etme şeklinde net olabildiği gibi, nesneye incelediği süreyi değerlendirme gibi muğlak ta olabilir.

***Nesneleri Filtreleme***<sup>6</sup>; Kullanıcıların profillerine dayanarak tavsiyelerde bulunmayı kapsamaktadır.

Yukarıda belirtilen tanımlamalar ile birlikte her tavsiye sisteminin de tavsiye ettiği nesne türlerine dayanarak bazı özellikler arz ettiğini belirtmek gerekir. Bizim çalışmamız haber tavsiye sistemi olduğundan haber özelliklerini belirtmemizde fayda vardır. Haber tavsiye sistemlerinin bazı temel özellikleri [7];

- Haber tavsiye sistemi çok fazla hesaplama gerektirdiği gibi haberlerin kısa ömürlü olması da bir özellik olarak değerlendirilebilir.
- Her haber sitesinin haber sunuşu farklı olmakla birlikte haberlerin belli bir formatı bulunmamaktadır.
- Haberler çok kısa zaman içinde önemini yitirebileceğinden ömürleri de kısa olmaktadır.
- Çoğu haber makaleleri, özel bir olayı kim, nerede, ne zaman yaptı gibi net sorular sorarak tanımlar.
- Kullanıcıların haber türlerine olan ilgisi çok kısa zamanda değişebilmektedir.
- Çok fazla haber makalesi bulunduğundan bu haberleri ayırmak için veri madenciliği ve makine öğrenmesi gibi karmaşık algoritmalar ve hesaplamalar gerekmektedir.

şeklinde belirtilebilir.

Bunlar arasında kuşkusuz en önemli olanın haberlerin kısa süre zarfında önemini yitiriyor olması ve dolayısıyla bu kısa zamanda kullanıcıya haberlerin tavsiye

---

<sup>6</sup> Ing : Filtering Component

ediliyor olması gelmektedir. Bir film, kitap ya da müzik tavsiye sistemlerinde olduğu gibi uzun süre kullanıcıya aynı kaynağı sunmak çok anlamlı değildir. Bizim sistemimizde de haberler günlük tavsiye edilmektedir.

*Pazzani ve D. Billsus'un* [8] de belirttiği gibi tavsiye edilecek nesnelere restorandaki menü örneğinde olduğu gibi belirgin özelliklere sahip olabilir ve bu özellikler yapısal olarak da gösterilebilir. Bunun yanında, haber makalelerinde olduğu gibi özellikleri net olarak çıkartılmayacak şekilde de olabilir. Özelliklerin net olarak karşılaştırılmadığı durumlarda, yapısal olmayan özelliklerin yapısal hale getirilmesi söz konusudur. Haber sistemleri için karmaşık ve uzun süren hesaplamalar başarılı sonuçlar üretse de haberin güncelliğini yitirmemesi önem arz etmektedir. Bu yüzden hesaplama ve kullanıcıya hızlı cevap dönme arasında bir denge oluşturulmalıdır. Farklı tavsiye sistemlerinin kendilerine has özellikleri sonraki kısımlarda anlatılacaktır.

Tavsiye metotları kişiselleştirilmiş veya kişiselleştirilmemiş metotlar olmak üzere ikiye ayrılabilir [13]. Kişiselleştirilmiş metotlarda kullanıcının profili önem kazanırken, kişiselleştirilmemiş metotlarda kullanıcının tavsiye edilen makalelere yaptığı oylamalar ve gösterdiği ilgiden oluşur [7].

## **2.2. Tavsiye Sistemlerinin Amacı**

Tavsiye sistemlerinin karar destek sistemleri olduğunu yukarıda vurgulamıştık. Karar destek sistemleri olarak kullanıcıların karar vermesinde etkili olabileceği gibi kullanıcıları yanlış yönlendirme özelliğini de göz önünde bulundurmaları gerekir. Ticari alanda daha çok bu amaca hizmet ettiğini söyleyebiliriz. Diğer sistemler için kullanıcıların aradıkları kaynaklara daha kolay ulaşmalarını sağlayacak sistemler olarak belirtilebilir. Daha resmi ifadelerle tavsiye sistemlerinin amaçlarını şu şekilde sayabiliriz:

- Gereğinden fazla veri ile baş edebilmek için veri miktarını ihtiyaçlarımız seviyesine indirmek.
- Müşterilere ve kullanıcılara karar vermede yardımcı olarak karar destek sistemleri olarak davranmak.

- Müşteri veya kullanıcı sayısını arttırmak.
- Müşteri ve kullanıcıların sistemi tekrar kullanmasını sağlamak.
- Ticareti, satışları ve kullanıcı sayısını arttırmak [9].
- Yüksek kalitede arama yapmak [10].

Bu maddeleri düşündüğümüzde bir kısmının kullanıcılara hitap ederken, bir kısmının ise sistem sahiplerine hitap ettiğini görürüz. Örneğin, ticari yönü, satışların artırılması yönü sistem sahiplerine bakarken, çok miktarda bilgi ile baş edebilmek, sisteme tekrar girmek, karar vermeyi kolaylaştırmak yönü ise ağırlıklı olarak müşteri/ kullanıcıya hitap eder.

### **2.3. Tavsiye Sistemleri**

Tavsiye sistemleri tanımı, amacı ve işleyişi genel hatlarıyla buraya kadar olan kısımlarda anlatılmıştı. Bu kısımda tavsiye sistemlerinde kullanılan metotlar ve daha sonra haber tavsiye sistemleri ve doküman tabanlı sistemler olarak tavsiye sistemleri ayrıştırılarak anlatılacaktır.

#### **2.3.1. Tavsiye Sistemlerinde Kullanılan Metotlar**

##### **2.3.1.1. İşbirlikçi Filtreleme**

En çok kullanılan ve en olgun metot olduğu kabul edilir. Bu sistemde esas olarak kullanıcılar veya ürünler gruplanarak, gruplar arasında benzerlikler bulunur. Bulunan benzerlikler yardımıyla kullanıcılara yeni tavsiyelerde bulunulur. Tipik bir işbirlikçi filtreleme senaryosunda *Xiaoyuan Su ve Taghi Khoshgoftaar'ın* [11] belirttiği gibi belirli sayıda kullanıcı ve nesne vardır. Eğer kullanıcılar benzer nesnelere oyladılar veya kullandırlarsa muhtemelen bir sonraki sefer de benzer nesnelere beğenecek veya oylayacaktır. Bu senaryo çoğu zaman tutsa da haber tavsiye sistemlerinde kullanıcıların benzer haberleri defalarca okumaktan sıkılacağını düşünmek gerekir. Fakat spor haberleri için bunun tersi de söz konusu olabilir ki, taraftarlar tuttıkları takımın elde ettiği zaferleri farklı ağızlardan farklı yorumlarla dinlemekten keyif alabilirler. Kullanıcılar benzer özelliklerine göre

gruplanacağı gibi nesnelere de benzer özelliklerine göre gruplanabilir. Kullanıcı ve nesnelere arasında matris oluşturmak suretiyle tavsiye sistemi oluşturulur [11].

İşbirlikçi filtrelemede genellikle iki farklı teknik kullanılır;

**Hafıza-bazlı işbirlikçi filtreleme**, kullanıcı filtreleme tekniğinde tüm kullanıcılar veya belli gruptaki kullanıcılar bir küme kabul edilerek benzer kullanıcılara benzer öğeler tavsiye edilir. Model öğrenme olmamakla birlikte temel algoritmalar kullanılır.

**Model-bazlı işbirlikçi filtreleme**, kullanıcı filtreleme tekniğinde kullanılan tekniğin kendini geliştirmesini ve öğrenmesini sağlayacak model kurgulanır. Bu modelin uygulanması zor olmakla birlikte performans sorunlarına da daha açıktır. Büyük miktarda veri için düşünüldüğünde, model-bazlı sistemlerde benzerlik algoritmaları boyut azaltma teknikleriyle birlikte kullanılarak hız ve ölçeklenebilirlik açısından fayda elde edilebilir. Temel farklılıklarını Tablo 2.1 deki gibi gösterebiliriz.

<b>İşbirlikçi filtreleme</b>	
<b>Hafıza-Bazlı</b>	<b>Model-Bazlı</b>
<ul style="list-style-type: none"><li>• Kullanıcı özellik ve oylarına bağlıdır.</li><li>• Yoğun veri kullanıldığında performans sorunları oluşabilir.</li><li>• Yeni kullanıcılar için tavsiye mekanizması çalışmaz. Çünkü öğrenme modeli yoktur.</li><li>• Diğer metoda göre daha az ölçeklenebilir.</li><li>• Pearson, cosine ve jaccard gibi daha temel benzerlik algoritmaları tercih edilir.</li></ul>	<ul style="list-style-type: none"><li>• Model üretmek başlangıçta zor ve karmaşıktır.</li><li>• Model-bazlı tavsiye sistemlerinde özellikle veri çok büyükse ve modellenmesi doğru yapılmışsa hafıza-bazlı sistemlere göre daha hızlıdır. Çünkü veri setinin tamamı değil, bir kısmı üzerinde sorgulama yapılacaktır. Veri setini daraltmak için boyut azaltma teknikleri kullanılmalıdır.</li><li>• Tavsiye esnasında veri setinin tamamı kullanılmadığından tavsiyeler</li></ul>

	<p>hafıza-bazlı sistemlerdeki kadar net olmayabilir.</p> <ul style="list-style-type: none"> <li>• Ölçeklenebilirlik ve tavsiye algoritması arasında seçim yapmak gerekebilir.</li> </ul>
--	--

**Tablo 2.1 İşbirlikçi Filtreleme Teknikleri**

İşbirlikçi filtrelemeyi genel olarak anlattıktan sonra kullanılan temel iki tekniğinden (model-bazlı ve hafıza-bazlı ) bahsettik. Daha farklı gruplamalar da mevcuttur [12]. İşbirlikçi filtrelemeyi daha iyi anlamak için bazı karşılaşılan problemlerden de bu kısımda bahsetmek yerinde olacaktır. Sayılan problemler, işbirlikçi filtrelemeyle birlikte, içerik bazlı filtreleme tekniğinde de karşımıza çıkabilecek problemlerdir;

- **Veri dağınıklığı<sup>7</sup>**: Çok fazla nesne olması ve bu nesnelere oylayan kullanıcıların arasında bağ kurmanın zorlaştığı zaman ortaya çıkan durumdur. Bu durum çoğu zaman yeni bir ürün geldiğinde veya yeni kullanıcı sisteme üye olduğunda meydana gelir ki buna *soğuk başlangıç<sup>8</sup>* problemi de denir. *Soğuk başlangıç* problemi hem yeni ürünler hem de yeni kullanıcılar için kullanılır. Yeni ürünler henüz oylanmadığı için nasıl tavsiye edileceği ile ilgili bir problemdir. Farklı sistemlerde çözüm olarak bazı yöntemler denenmiştir. Bu yöntemler ağırlıklı olarak ürünlerin diğer ürünlere benzerliğinin bulunması esasına dayanır. Ürünlerin geniş bir özellik yelpazesine sahip olması gerektiğini unutmamak gerekir. Yeni kullanıcılar için de aynı problem söz konusudur. Bu durumda da kullanıcıların diğer kullanıcılara benzerliği bulunmaya çalışılır. Farklı bir yöntem de yeni haberler üzerinde güven ilişkisi kurularak oluşturulmaya çalışılmıştır [13].
- **Ölçeklendirilebilme<sup>9</sup>** : Nesne ve kullanıcı sayısı çok fazla arttığında hesaplamalar zorlaşacaktır. Bu durumda boyut azaltma teknikleri vb. kullanılır. Haber tavsiye sistemleri için hesaplama süresinin uzamasının olumsuz etkilerini yukarıda bahsetmiştik.

<sup>7</sup> İng : Data Sparsity

<sup>8</sup> İng : Cold-Start

<sup>9</sup> İng : Scalability

- **Benzerlik<sup>10</sup>** : Aynı veya birbirine çok benzeyen nesnelere benzer özellikleri taşıdığı için aynı kullanıcıya tekrar tekrar tavsiye edilir. Bu durumda genellikle kullanıcılar aynı nesnenin tavsiye edilmesinden sıkılmaktadır. Çözüm olarak çok benzeyen veya aynı nesnelere bulan algoritmalar kullanılmalıdır. Bu durum farklı tavsiye sistemleri için farklı yorumlanabilir. Film veya müzik tavsiye sistemleri için benzer tavsiyeler uygunken, haber tavsiye sistemlerinde kullanıcıların benzer haberleri tekrar okuması sıkıcı olabilir.
- **Kararsız Kullanıcılar<sup>11</sup>**: Kararsız kullanıcılar sistemin istikrarını zedeleyebilir. Bu kullanıcılar kendilerine benzeyen kullanıcılar için de olumsuz örnek teşkil etmektedir. Bu durumlar için güvene dayalı sistemler denenebilir.
- **Yanlış Yönlendirme<sup>12</sup>**: Kişiler, sahibi oldukları bazı ürünleri özellikle tavsiye ederek kullanıcılar üzerinde yanlış etki bırakabilir. Benzer durum ticari ürünler satma maksatlı kurulan sitelerde söz konusudur. Bazı ürünler özellikle ön plana çıkartılabilir. Belli bir zümre bu ürünleri kullanarak bu ürünlerin değerlerini arttırabilirler.

İşbirlikçi filtreleme en çok kullanılan metottur. Bazı önemli avantaj ve dezavantajları aşağıda belirtilmiştir:

#### **Avantajları;**

- En eski metot olmasına rağmen etkinliğini hala korumaktadır.
- Sistem tavsiyelerini kullanıcının aldığı ürünlere dayandırdığından güncelliğini korumaktadır.

#### **Dezavantajları;**

- Yeni kullanıcılara ilk tavsiyelerde bulunmak zordur. Bu problem *early-rater problem* olarak tanımlanabilir.

---

<sup>10</sup> Ing : *Synonymy*

<sup>11</sup> Ing : *Gray Sheep*

<sup>12</sup> Ing : *Shilling Attacks*



- Kullanıcı profili dar olduğunda sınıflandırma yapmak zordur (*Data sparseness*). Yeni ürünlerin alıcısı başlangıçta oluşmadığından tavsiye etmekte zor olmaktadır.
- Karmaşık hesaplamalara ihtiyaç duymaktadır. Ayrıca ürün ve müşteri yelpazesi genişlediğinde hesaplamalar daha da karmaşıklaşmaktadır.

### 2.3.1.2. İçerik Bazlı Filtreleme

İçerik bazlı filtreleme işbirlikçi filtrelemenin aksine sadece müşterinin kendisinin geçmişine ve tercihlerine dayanarak tavsiyede bulunur. Daha genel ifadeyle kullanıcıların başlangıçta oluşturdukları profillere veya zaman içinde oluşturdukları dinamik profillere dayandığı söylenebilir [4]. Profilin başlangıçta oluşturulması yeni kullanıcı probleminin bir çözümü olarak değerlendirilebilir. Temel çalışma prensibi kullanıcıların tercihlerinin belirlenmesi ve daha sonra tavsiye edilebilecek nesnelere kullanıcı ile ilişkisinin kurulması ve nesne kullanıcı matrisi kurularak tavsiye sisteminin oluşturulmasıdır. Fab sistemi [14] kullanıcılara internet sayfası tavsiye etmektedir ve her sayfayı yüz kelime üzerinden değerlendirmektedir. İçerik bazlı filtreleme sistemlerinde de VSM<sup>13</sup> yoğun olarak kullanılmaktadır. Pazzani'de verdiği restoran örneğinde [8] *TF-IDF* ve VSM'nin öneminden bahsetmiştir. Kullanıcı profillerinin de iki şekilde oluşturulabileceğinden bahsetmiştir; Kullanıcının tercihleri ve kullanıcının sistem ile olan geçmiş etkileşimi. Bunlarla birlikte kullanıcı profili ne kadar etkili olursa olsun eğer tavsiye edilebilecek kaynaklar diğer kaynaklara göre ayırt edici özelliklere sahip değilse iyi bir tavsiye sisteminin oluşturulmayacağı ve bunun kısıtlayıcı bir faktör olduğu [8] söylenebilir.

Adomavicious ve Tuzhilin'de içerik bazlı filtrelemenin köklerinin bilgi erişimi<sup>14</sup> ve bilgi filtrelemede<sup>15</sup> olduğunu söyler [15]. Bununla birlikte içerik tabanlı filtreleme sistemlerinin veri filtrelemenin devamı olduğu düşünülebilir.

---

<sup>13</sup> Vector Space Model (Vektör Uzay Modeli)

<sup>14</sup> Ing : information retrieval

<sup>15</sup> Ing : information filtering

### **Avantajları;**

- Sadece kullanıcının tercihlerine dayandığından işbirlikçi filtreleme kadar karmaşık hesaplamalara ihtiyaç duymamaktadır.
- Benzer kullanıcıları bulma esnasında karşılaşılabilecek yanlış sonuçlardan etkilenmez.

### **Dezavantajları;**

- Kullanıcılara her seferinde benzer ürünler sunulduğundan kullanıcı bir süre sonra bıkmaya başlayabilir.
- Farklılaşmamış ürünleri tavsiye etmek kolay iken, farklılaşan daha önceki ürünlere benzer özellikleri taşımayan ürün piyasaya girdiğinde sınıflandırma yapmak zorlaşır. İşbirlikçi filtrelemede ürün kullanılmaya başladıkça özellikler dinamik olarak oluşmaktadır.
- Yeni kullanıcı problemi; kullanıcı başlangıçta yeterince ürün kullanmadığından tavsiyede bulunmak zordur.

#### **2.3.1.3. Karma Yaklaşımlar**

Genel olarak konuşacak olursak, tavsiye sistemleri sınıflandırmaları ve uygulamaları ağırlıklı olarak içerik-bazlı ve işbirlikçi filtreleme yöntemleri üzerine yoğunlaşmıştır. *Burke, Rubin* [16] bu sistemleri incelemiş ve bazı karşılaştırmalar yapmıştır. *Belkin* [17] bilgi filtrelemeyi kullanıcının doküman ile etkileşimi olarak tanımlamıştır. *Burke* ise, araştırmasında, içerik-bazlı filtrelemeyi bilgi filtrelemenin gelişimi ve devamı olarak görmektedir. Bu iki sistemin de olumlu ve olumsuz yönleri bulunmaktadır. Daha performanslı bir sistem oluşturmak için bu iki sistemin etkileşimi ve iyi yönlerini birleştirmek düşünülebilir. Bu tür sistemlere karma sistemler denilmektedir.

Karma<sup>16</sup> yaklaşımlar içerik-bazlı ve işbirlikçi filtrelemeyi birlikte ele alarak ikisinin de olumlu yanlarını kullanır. *Konstan'da* tekniklerin zorluklarını ve kolaylıklarını incelemiş [18], yeni kullanıcı ve yeni metinlerle ilgili sıkıntıları belirtmiştir. Yeni

---

<sup>16</sup> Ing : Hybrid

kullanıcıların yeni dokümanları oylaması zaman alacağından oylama yapan otomatik robotlar geliştirmişlerdir. İşbirlikçi filtreleme ve içerik-bazlı filtreleme de en büyük sorunların başında hala yeni kullanıcılar ve yeni dokümanlara nasıl davranılacağı ile ilgili sorunlar gelmektedir.

İçerik-bazlı filtreleme ile işbirlikçi filtreleme kıyaslaması birçok yerde yapılmaktadır. İçerik-bazlı filtrelemede kullanıcılara tavsiye edilen dokümanlar tanımlanan özellikler üzerinden yapılırken, işbirlikçi filtrelemede yeni dokümanlar da tavsiye tekniği açısından başarılı sonuçlar çıkarabilir. Fakat haber tavsiye sistemlerinde sık aralıklarla haberlerin yenilenmesi nedeniyle tavsiye sisteminin günlük olarak çalışması gerekmektedir. Bu da zaman açısından diğer kullanıcıların oylamalarını dikkate almayı zorlaştırmaktadır.

İki tekniğin de zorlandığı yönlerden birisi de, benzer haberlerin aynı özellikleri taşıması nedeniyle tekrar tekrar tavsiye edilmesi olarak ifade edilebilir. Bu etkiye "*portfolio effect*" de denir. Bunu önlemek için *Pazzani* [19] geliştirdikleri *DailyLearner* sisteminde kullanıcılardan ilgili, ilgili değil ve biliniyor şeklinde geri besleme almışlar ve biliniyor özelliğine göre tavsiye edilen dokümanları tavsiye sisteminden çıkarmaya çalışmışlardır. *Burke, Robin* çalışmasının devamında teknikleri birbirleri ile karşılaştırmıştır[16].

Karma modeller işbirlikçi filtrelemenin ve içerik-bazlı filtrelemenin olumlu yönlerini farklı algoritmalar kullanılarak birleştirilmesiyle elde edilir. Genele işbirlikçi filtrelemeye daha yakındır, fakat işbirlikçi filtrelemede yer alan ilk oylayıcı<sup>17</sup> problemini çoğunlukla saf dışı bırakır. *Wesomender* sistemi [20] iki metodu birleştirme işini algoritmayla çözmüşlerdir. İki yöntemin sonuçları bir ağırlık katsayısı ile ( $r=W.CB+(1-W)CF$ ) formülü ile birleştirilir. Bu formülde CB içerik temelli filtreleme sonucu, CF işbirlikçi filtreleme sonucu ve W içerik temelli filtrelemenin ağırlık katsayısıdır. Karma sistemlerinin uygulanması aşamasında farklı teknikler kullanılmaktadır. Bu teknikleri kısaca açıklayacak olursak;

**Ağırlıklandırılmış Sistemler:** Bu sistemlerde başlangıçta işbirlikçi ve içerik tabanlı tavsiyeye eşit ağırlık verilmiştir. Sistem geliştikçe ağırlıklandırma artmaktadır. Teorik olarak sistem başarılı gözükse de, ilk profiller oluştuktan sonra,

---

<sup>17</sup> First-rater problem

ağırlık içerik tabanlı tavsiye motoruna ve daha sonra, tavsiyeler arttıkça, işbirlikçi filtrelemeye kaydığı aşıkardır.

**Dönüştürme:** Bu sistemde tavsiye sonucunda oluşacak durumlara göre farklı sistemler uygulanmaktadır. *DailyLearner* [19] sisteminde başlangıçta içerik tabanlı sistem kullanılmışken daha sonra işbirlikçi filtrelemeye geçişler olmaktadır.

Diğer örnekleri de karışım ki farklı durumlar için yöntemlerden birisinin denenmesi şeklinde uygulanmıştır. Özellik birleştirme ve üst üste bindirme gibi yöntemler [16] sayılabilir.

#### **2.3.1.4. Basitleştirilmiş Yaklaşımlar**

İstatistiksel analiz ve elle seçim kişiselleştirilmemiş ve karmaşık hesaplamalar gerektirmeyen tekniklerdir. İstatistiksel metotlar müşterilerden alınan geri dönüşlere dayanarak belirlenir. Elle seçime nazaran bazı hesaplamalar yapılır. Elle yapılan seçimler ise uzmanların tercihlerine dayanır. Bu sistemler günümüz açısından çok basit kalmakla birlikte hızlı çözümler bulma açısından denenebilecek sistemlerdir. Anlık çözümler bulma açısından değerlendirilebilir.

“Karmaşık hesaplamalar gerektirmedığından güncelliği sağlamak kolaydır.” avantaj olarak sayılabilecekken, “Bütün kullanıcılar için aynı tavsiyeleri barındırır. Bu yüzden kişiselleştirilmemiştir” fikri dezavantaj olarak belirtilebilir.

#### **2.3.1.5. Demografik Filtreleme**

Demografik filtreleme yakın davranış gösteren kişileri araştıran işbirlikçi filtreleme ile temelde benzerlik göstermektedir. İki teknik arasındaki fark ise işbirlikçi filtrelemenin kişi özellikleri yerine kullanıcıların önceki davranışları (oylama, sıralama) arasındaki benzerliği değerlendirmesidir. Demografik filtreleme tekniği kullanan tavsiye sistemlerinin dikkate aldıkları kullanıcı özellikleri, kullanıcıların cinsiyet, yaş, eğitim, meslek ve yaşadıkları şehir gibi özellikle sabit olan demografik bilgileri olabilirken; kullanıcıların sistemle etkileşimleri geçmişte aldıkları, inceledikleri ürünler, gezindikleri web siteleri gibi dinamik veriler de olabilir. Bu metotta, kullanıcılar demografik özelliklerine göre sınıflandırılır. Kullanıcılar veya müşteriler, profillerine göre gruplanarak sistem tarafından henüz

denenmemiş, kullanılmamış ürünler kullanıcılara tavsiye edilir. İyi bilinen ilk örneği kitap tavsiye sistemi olan *Grundy*'dir [21], *Prizm* ise tipik başka bir örnektir [9].

Literatürde son yıllarda demografik tavsiye sistemleri üzerinde çok fazla çalışılmamaktadır. Kişilerin demografik bilgilerini toplarken karşılaşılan zorluklar, demografik bilgilerin sabit kalıp, insanların zevk ve tercihlerinin zaman içerisinde değişmesi dolayısıyla çok anlamlı olmamaktadır. Bunun gibi nedenler araştırmacıların demografik tavsiye sistemlerine olan ilgisini azaltmıştır.

#### **Avantajları;**

- Karmaşık bilgiye ihtiyaç duymaz. Temel sınıflandırmadan sonra tavsiyede bulunmak daha kolaydır.

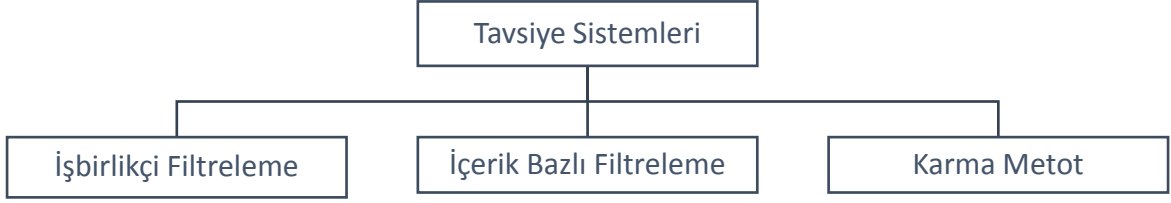
#### **Dezavantajları;**

- Sınıflandırma çok genel olduğundan başarı ihtimali biraz düşüktür.
- Sınıflandırmaya esas girdi verisi kullanıcılardan doğrudan alındığından eksik veya yanlış olma ihtimali vardır.
- Kullanıcı/müşteri ihtiyaçları zamanla değişebileceğinden güncelliğini yitirebilmektedir.

### **2.3.2. Tavsiye Sistemi Sınıflandırma Yaklaşımları**

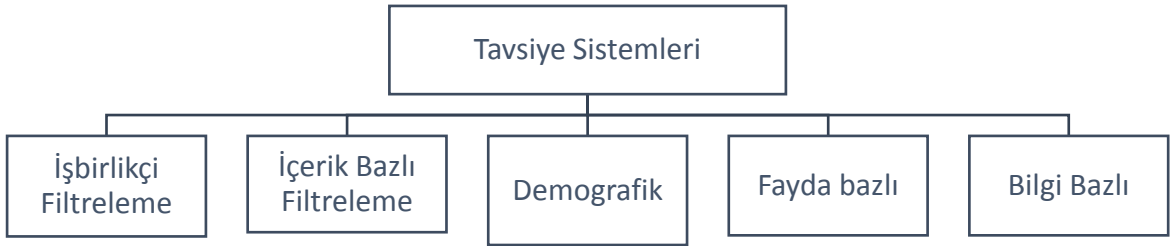
Literatüre geçmiş birçok tavsiye sistemi sınıflandırması mevcuttur. Bunların bir kısmı kişiselleştirilmiş iken, bir kısmı kişiselleştirilmemiş sistemlerdir. Kişiselleştirilmiş sistemin en belirgin örnekleri İşbirlikçi filtreleme, içerik-bazlı filtreleme sayılırken, kişiselleştirilmemiş sisteme örnek olarak istatistiksel analizler verilebilir.

Tavsiye sistemlerini kısaca anlattıktan sonra tavsiye sistemleri sınıflandırmaları incelenebilir. Net olarak belirlenmiş bir standart sınıflandırma olmamakla birlikte değişik sınıflandırmalar yapılmıştır. *Adomavicious* ve *Tuzhilin* en popüler sınıflandırmanın tavsiye sistemlerini üçe ayırmak olduğunu bu ayrımın kısıtlarını belirterek söylemişlerdir [15] (Şekil 2.1). Bu sınıflandırma gerçekten de en kabul gören sınıflandırmadır. Bu tez için de kabul edilen sınıflandırma budur. Bazı diğer yöntemler ağırlıklı olarak işbirlikçi filtrelemenin altında kabul edilir.



**Şekil 2.1 Tuzhilin'in Sınıflandırması**

*Robin Burke*, tavsiye sistemlerini beş kategoriye ayırır. Ayrımı yaparken dikkate aldığı durum girdi verisidir [16] (Şekil 2.2). Bu sınıflandırma içinde en çok kullanılanın ise işbirlikçi filtreleme olduğunu belirtmiştir.



**Şekil 2.2 Burke'in Sınıflandırması**

Fayda bazlı<sup>18</sup> ve bilgi bazlı<sup>19</sup> işbirlikçi filtreleme tekniklerinden de bu kısımda kısaca bahsetmek yerinde olacaktır. Bu iki teknik birbirlerine benzemekle birlikte aralarındaki ilişkiyi *Burke* [16] oldukça güzel anlatmıştır. Fayda bazlı sistemlerde kullanıcıya genel bir tavsiyeden ziyade her bir nesnenin kullanıcı açısından önemini belirten bir fayda matrisi oluşturur. Bu tekniğin diğerlerinden farkı ele alınmamış bazı durumların fonksiyon içinde ele alınabiliyor olmasıdır. Örneğin; bir bilgiye kullanıcının o an için çok fazla ihtiyaç duyuyor olması gibi. Bilgi bazlı sistemler de benzer şekilde her bir nesne ile kullanıcı arasındaki ilişkinin fonksiyonel olarak ortaya çıkarılabiliyor olmasıdır. Bu teknikte kullanıcı hakkında bilgi nettir ve bu bilgi sistem açısından kullanılabilir olmalıdır.

<sup>18</sup> Ing : Utility-based

<sup>19</sup> Ing : Knowledge-based



**Şekil 2.3 Schafer'in Sınıflandırması**

*Schafer et al* sadece tavsiye metotlarını değerlendirmez. Aynı zamanda, müşterilerden gelen oylama ve geri beslemeleri de kullanır. Ona göre sınıflandırmalar biraz daha çeşitlidir ve farklı açılardan değerlendirilebilir. *Schafer* tavsiye sistemlerinin ticari etkilerini ele almıştır [22][23] (Şekil 2.3). Tavsiye sistemlerini *Schafer* beşe ayırmıştır.

Kişiselleştirilmemiş Tavsiye<sup>20</sup> temel olarak işbirlikçi filtrelemeyle aynıdır. Kişinin tercihlerinden ziyade diğer kullanıcıların tercihlerine dayanır. Özellik bazlı<sup>21</sup> sistemlerde ise kullanıcının açık olarak bir istekte bulunması gerekir. Örneğin belli bir özelliğe bir kitap araması gibi. Nesne-nesne ilişkili<sup>22</sup> sistemler ise ürün satan siteler için konuşursak, kullanıcının belli ürünleri sepetine eklemesi ve ardından sistemin bu nesnelere içinden tavsiye edilecekleri sıralama yaparak belirlemesine dayanır. Kullanıcı-Kullanıcı ilişkili<sup>23</sup> sistemlerde tavsiyelerin benzer ürünü alan diğer kullanıcıların tercihlerine bakılarak yapılıyor olmasıdır ki, bu aslında bir nevi işbirlikçi filtreleme olarak tanımlanabilir. Kişiselleştirilmemiş Tavsiye tekniğinden farkı ise nesnelere arasında tercih yapabilecek şekilde verinin oluşmuş olmasıdır. Son teknikte kullanıcı girişi<sup>24</sup> ki, bu teknikte kullanıcıların girdiği veriler, yorumlar, oylamalar dikkate alınmaktadır.

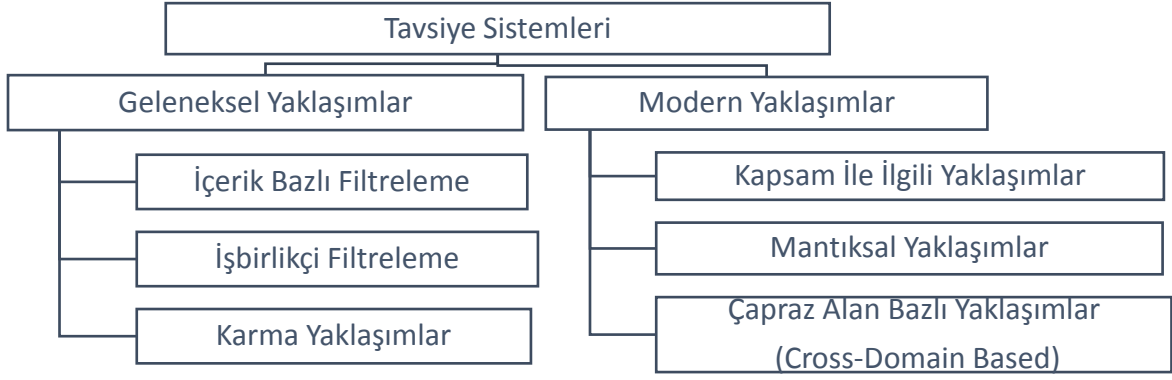
<sup>20</sup> Ing : Non-personalized Recommendation

<sup>21</sup> Ing : Attribute-based Recommendation

<sup>22</sup> Ing : Item-to-item Correlation

<sup>23</sup> Ing : People-to-people Correlation

<sup>24</sup> Ing : User Input



**Şekil 2.3 Asanov'un Sınıflandırması**

*Asanov* [4] ise sınıflandırma sistemlerini Şekil 2.4 te görüldüğü gibi üst seviyede iki dala ayırmış ve daha sonra bu dalların altında sınıflandırma yaparak biraz daha farklı bir yol izlemiştir. Geleneksel yaklaşımlar güncelliğini korumakla birlikte tavsiye sistemleri açısından da hala geçerliliğini korumaktadır. *Asanov* ise tercihlerin çok hızlı değiştiği varsayımıyla modern yaklaşımları farklı kategori olarak değerlendirmiştir. Modern yaklaşımları inceleyecek olursak, Kapsam ile ilgili yaklaşımlardan<sup>25</sup> kasıt kullanıcının bulunduğu çevre ve şartlarla ilgili tavsiye sisteminin oluşmasıdır. GPS üzerinden kullanıcının konum bilgisinin alınarak alışveriş veya gezi yerlerinin tavsiye edilmesi gibi. Bu, aslında demografik filtrelemeyle benzerlik göstermektedir. Anlamsal yaklaşımlarda<sup>26</sup> aramalarda kullanılan kelimelerin bağımsız kelimeler olmaktan ziyade semantik olarak kullanılabilir olmasıdır ki bu ender olarak yapılabilir. Çapraz alan bazlı yaklaşımlarda<sup>27</sup> klasik tek bir domain yerine birden fazla domainin etkileşiminin kullanılmasıdır. Modern yaklaşımlar bütün olarak düşünüldüğünde uygulanması zor fakat yakın gelecek için bir zorunluluk olarak karşımıza çıkmaktadır. Özellikle kullanıcıların tercihlerinin, konum ve çevre gibi bilgilerinin çok hızlı değiştiği günümüzde bu durumların ele alınması önemlidir.

Bunlara ek olarak birçok sınıflandırma mevcuttur. Sınıflandırmaların hepsi düşünüldüğünde genel olarak klasik İşbirlikçi filtreleme, içerik-bazlı filtreleme, karma sistemler ve bunun yanında tavsiye sistemlerinin niteliğine göre farklı

<sup>25</sup> Ing : Context-aware Approaches

<sup>26</sup> Ing : Semantic-based Approaches

<sup>27</sup> Ing : Cross-domain based Approaches



durumlara öncelik verecek disiplinleri uygulamak yerinde olacaktır: Konum bilgisi, çevre bilgisi, tavsiyenin saat kaçta kullanıcıya gösterildiği gibi.

#### **2.4. Haber Tavsiye Sistemleri**

Tavsiye sistemleri, temel sınıflandırmaların dışında uygulamalarda görülen farklılıklara göre de farklı bir değerlendirmeye tabi tutulabilir. Film, müzik, kitap, araba, film, haber, ticari ürünler gibi. Her bir sistemin kendine özgü özellikleri bulunmaktadır. Haber tavsiye sistemleri anlık değişebilirken, film ve müzik tavsiye sistemlerinde anlık tavsiyeler çok önem arz etmemektedir. Haber tavsiye sistemlerini doküman bazlı sistemlerin altında değerlendirmek de yanlış olmaz. Doküman bazlı sistemleri değerlendirmede en önemli husus ise yapısal ve işlenmemiş durumda olan dokümanlardan anlamlı sonuçlar ve bilgiler çıkartmaktır. Daha detaylı olarak doküman tabanlı sistemler aşağıda incelenmiştir.

#### **2.5. Doküman Bazlı Tavsiye Sistemleri**

Dokümanlara dayalı tavsiye sistemlerinde dokümanları sınıflandırmak, kümelemek son derece zordur. Zor olmasının sebebi dokümanların yapısal halde bulunmaması ve dolayısıyla işlenecek durumda olmamasıdır. Dokümanların matematiksel formülleri yoktur. Bu nedenle yapılacak ilk işlem, yapısal olmayan dokümanın yapısal ve anlamlı hale getirilmesidir. Yapısal hale geldikten sonra benzerlik ölçme yöntemlerinden biri kullanılarak dokümanlar sınıflandırılmalıdır. Bu işlemler bittikten sonra dokümanların tavsiye sistemlerinde nasıl değerlendirileceği belirlenmelidir. Bu aşamaların detayları bu bölümde tartışılmıştır.

##### **2.5.1. Dokümanların Yapısallaştırılması**

Yapısal olmayan dokümanların yapısal hale getirilmesi için birçok çalışma yapılmaktadır. Bir doküman incelendiğinde dokümanların belli sıklıktaki kelimelerden oluştuğu aşikardır. Bu özellik kullanılarak dokümanlar işlenebilir. Bu bağlamda kullanılan en temel model *vektör uzayı modelidir* (Rocchio [24]). Bu modelde her haber veya her metin belgesi, belirli sıklıkta kelimelerden oluşan bir vektör olarak düşünülebilir. Bu vektörün değişkenlerini ise kelimeler ve kelime ağırlıkları oluşturmaktadır. Bir dokümanı kelimelerden oluşan bir vektör olarak tanımlayacak;

$$d_i = \{W_{i1}, W_{i2}, W_{i3}, \dots, W_{iN}\}.$$

Doküman içerisindeki bir kelimenin skoru şu şekilde tanımlanır:

$$W_{ij} = tf_{ij} * \log \left( 1 + \frac{N}{df_i} \right)$$

- $tf_{ij}$ : i'nci dokümanın j'ninci kelimesinin doküman içerisindeki sıklığı.
- N: Örneklem üzerindeki toplam doküman sayısı.
- $df_i$ :  $W_{ij}$  kelimesini içinde bulduran doküman sayısı.

Vektör uzay modelini dokümana uyguladığımızda elde edeceğimiz sonuç ise;

$d_i = \{W_{i1}, W_{i2}, W_{i3}, W_{i4}, \dots\}$  ve örneklem de  $N = \{d_1, d_2, \dots, d_N\}$  şeklinde ifade edilir. Bu formül neticesinde doküman genel olarak yapısal hale getirilmiş olur. Logaritma kullanımı ağırlığın ani artışlarını kontrol altında tutmak için yapılmış bir normalizasyon işlemidir. Dokümanların bu şekilde formüleleştirilmiş gösterimine *vektör uzayı modeli*<sup>28</sup> denmektedir. VSM ilk olarak **SMART** [25] sisteminde kullanılmak üzere geliştirilmiştir. VSM her dokümanı noktalardan oluşan bir bütün olarak ele almaktadır. Her dokümanın da birbirine olan mesafelerini matematiksel olarak hesaplamak mümkündür. *Peter Turney, Patrick Pantell'in* [26] yaptığı araştırmada vektör uzayı modelin dokümanların benzerliklerini bulmada çok etkili olduğunu gözler önüne sermiştir.

### 2.5.2. Benzerlik Ölçme Yöntemleri ve Sınıflandırma Algoritmaları

Dokümanlar yapısal hale getirildikten sonra, dokümanların birbirleriyle ilişkileri benzerlik ölçme yöntemleri ve sınıflandırma yöntemleri kullanılarak yapılmaktadır.

**Benzerlik Ölçme Yöntemleri;** Benzerliği ölçülecek iki nesne x ve y olarak tanımlanacak olursa, bu iki nesne arasındaki yakınlık, benzerlik fonksiyonlarıyla ifade edilebilir. Bazı benzerlik algoritmaları üzerinde inceleyecek olursak;

- **Minkowsky Distance;** *Minkowsky, Manhattan ve Euclidean* [27] metrikleri temel olarak iki nesne arasındaki benzerliği koordinat düzlemi üzerinden ilişkilendirerek hesaplar. x ve y benzerliği hesaplanacak olan iki nesne olarak düşünülürse, fonksiyon şu şekilde çıkarılabilir;

---

<sup>28</sup> Ing : Vector Space Model

$D = \sqrt{\sum_{i=0}^k (x_i - y_i)^n}$  ;  $n, 2$  değerini aldığında fonksiyon euclidean denklemi olarak adlandırılır.  $k$  kelimenin iki dokümanda  $x_i$  ağırlığı ve  $y_i$  ağırlığı dikkate alınarak hesaplaması yapılmıştır.

- **Manhattan Distance;** *Minkowsky* benzerliğinin genelleşmiş biçimi olarak tarif edilebilir. Formülü şu şekildedir;  $D = \sum_{i=0}^k |x_i - y_i|$  . Blok metrik olarak ta isimlendirilir. Grid yapısına dönüştürüldüğünde yollar arasında yatay veya dikey birer birer gidilerek benzerlik ölçülebilir. Euclidean metriğinde ise yollar arasındaki yolculuk kare ve karekök ilişkisiyle kurulduğundan, birer birer yatay veya dikey gidilemez.
- **Euclidean Distance;**  $\sqrt{\sum_{i=0}^k (x_i - y_i)^2}$  formülüyle ifade edilebildiğini minkowski metriğini belirtirken söylemiştik. Euclidean metriği iki nesne arasındaki benzerliği nesnelerin kare farklarının kareköklerini alarak hesaplar.
- **Rocchio Algorithm [24];** İlk sınıflandırma metotlarından. Bu metotta her doküman bir vektör olarak ifade edilir ve vektör oluşturulurken, tf-idf [28] kullanılır. *Vektör Space Model* 'e dayalı olarak geliştirilmiştir. İlk kullanımı ve kaynağı **SMART** sistemidir. Formülü şu şekildedir:

$$q_m = \alpha q_0 + \beta \frac{1}{|D_r|} \sum_{d_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{d_j \in D_{nr}} \vec{d}_j$$

$q_m$  : Düzenlenmiş Vektör

$q_0$ : Orijinal Vektör

$|D_r|$  : Bilinen İlgili Dokümanlar

$|D_{nr}|$  : Bilinen İlgisiz Dokümanlar

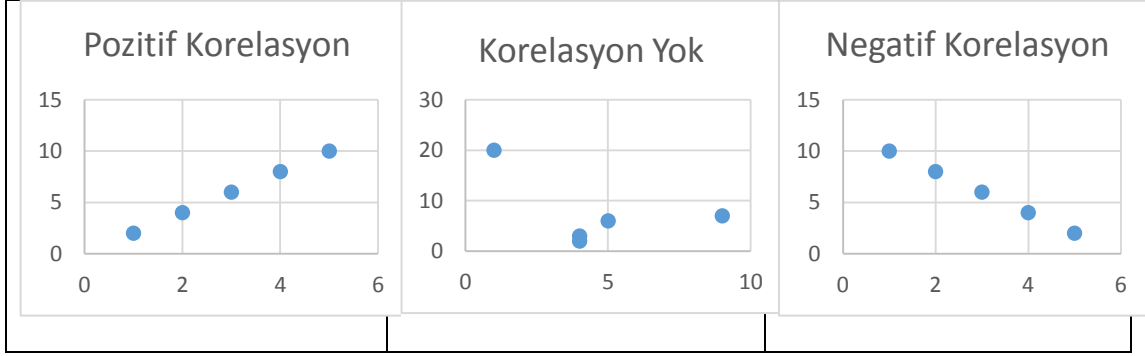
$\alpha, \beta, \gamma$  : Ağırlıkları oluşturmaktadır.

Ağırlık katsayıları değiştirilerek nesnenin sınıfı değiştirilebilir.

- **Pearson Similarity [11];** Pearson benzerliği iki değişkenin veya iki dokümanın vektörel olarak birbirleriyle ne kadar ilişkili olduğunu belirler. -1 ile +1 arasında değerler alır. İki doküman arasındaki benzerlik +1 çıktığında dokümanların tamamen benzer olduğunu, -1 çıktığında iki doküman arasında negatif benzerlik olduğu, 0 çıktığında ise ilişki olmadığı sonucu çıkarılabilir. Dokümanların tamamen benzer olması karşılaştırılan özelliklerin aynı olması demektir. Temel formülizasyonu;

$$Pearson(x, y) = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{N}\right) \cdot \left(\sum y^2 - \frac{(\sum y)^2}{N}\right)}}$$

Koordinat düzleminde örnek Pearson korelasyonu Şekil 2.5' teki gibi gösterilebilir.



Şekil 2.4 Örnek Pearson Korelasyonu

- **Cosine Similarity [29]**; Dokümanlar vektörel olarak ifade edildikten sonra, aralarındaki benzerlik vektörler arasındaki korelasyon ile açıklanabilir. Cosine benzerliği de dokümanlar arasındaki benzerliği, iki vektör arasındaki açıya dayanarak bulmaktadır. Popüler metriklerden olup formülü;

$$\cos(x, y) = \frac{(x \bullet y)}{\|x\| \|y\|}. \text{ Negatif değer almamakla birlikte aldığı değer } [0,1]$$

aralığındadır. Sınıflandırma ve kümeleme algoritmalarında da sıkça kullanılmaktadır.

- **Jaccard Similarity [29]**; Tanimoto benzerliği olarak ta bilinir.

$$\text{Jaccard benzerliği; } J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \quad M_{11}; \text{ Kesişen nesne sayıları.}$$

$$\text{Jaccard Uzaklığı; } J = \frac{M_{10} + M_{01}}{M_{01} + M_{10} + M_{11}} \quad M_{10}, M_{01}; \text{ Farklı nesne sayısı.}$$

Dokümanlar açısından bakacak olursak, jaccard benzerliği iki doküman arasındaki benzerliği, ortak kelimelerin doküman içerisindeki ağırlığını karşılaştırarak bulmaktadır. [0,1] aralığında değer alır. 1 değeri, iki dokümanın benzer olduğunu, 0 ise, iki doküman arasında korelasyon olmadığını gösterir. Doküman özellikleri kelime olarak tanımlandıysa ve kelime sayısı kısıtlandıysa başarılı bir şekilde kullanılabilir.

### **Sınıflandırma Yöntemleri;**

Modern bilgi teknolojilerinin gelişmesiyle birlikte kullanışlı bilginin yoğun bilgi arasından nasıl elde edileceği ile ilgili çalışmalar yapılmaktadır. Bu çalışmalar ile birlikte işlenen dokümanların sınıflandırması da son yıllarda önem kazanmıştır. Sınıflandırma işlemleri makine öğrenmesi, veri madenciliği, bilgi erişim (IR) ve daha birçok alanı kapsamaktadır. Farklı sınıflandırma ayrımları mevcuttur. Bunlardan bir tanesine göre sınıflandırma algoritmaları kural-bazlı, doğrusal sınıflandırma algoritmaları ve örnek-bazlı olduğu gibi, başka bir ayrıma göre de başka bir ayrımda da daha farklı ele alınmış olabilir. Sınıflandırma işlemleri yapılabileceği gibi dokümanların kümelenmesi işlemleri de yapılabilir. Kümeleme işleminde sınıflar belli olmayıp dokümanlar benzerliklerine göre gruplanmaktadır. Kümeleme işlemleri daha bulanık yapıda değerlendirilebilir. Kısaca kümeleme işlemlerinden bahsedip üzerinde durduğumuz sınıflandırma algoritmalarına geri dönebiliriz.

Mevcut metin kümeleme teknikleri başlıca beş kategoriye ayrılabilir [30];

- Kısmi metin kümeleme,
- Hiyerarşik metin kümeleme,
- Kural-bazlı metin kümeleme,
- Makine öğrenmesi ile metin kümeleme,
- Üst boyutlu metin kümeleme.

Kısmi kümeleme basit olmakla birlikte karmaşık kümeler oluşturmaktadır. Hiyerarşik kümelemede alt kümeler oluşarak daha yapısal kümeleme oluşturulabilir. Kural-bazlı kümeleme basittir, aynı zamanda belli kurallara dayandığından kısıtlı kümeleme yapabilir. Bunlarla birlikte genetik algoritmalar kullanılarak da kümeleme ve sınıflandırmalar yapılabilir. *Yonghong* ve *Wenyang* genetik algoritma kullanmış ve terimlerin ağırlıklarını değiştirerek etkili bir

kümeleme yapmaya çalışmışlardır [31]. Metinlerin kümelenmesinin yanında eğer metinleri belli kategorilere bölmeye çalışıyorsak onları sınıflandırma işlemine tabi tutabiliriz. Bu kısımdan sonra da belli başlı sınıflandırma algoritmalarını ve kullanım alanlarını inceleyebiliriz.

Kural-bazlı, doğrusal ve akıllı sınıflandırma türleri vardır. Kural-bazlı sınıflandırma algoritmaları belli başlı kurallara bağlıdır ve bu kurallara göre sınıflandırma yaparlar.

Konsept olarak her sınıflandırma algoritması eğitim, uygulama ve değerlendirme fazından oluşur. Eğitim fazında test verileri üzerinde algoritma çalıştırılır. Sınıflandırma neticesinde dokümanlar tek bir sınıfa girebileceği gibi birden fazla sınıfa da dahil olabilir. İdeal olan her bir dokümanın tek bir sınıfta yer almasıdır. Son olarak değerlendirme işlemleri yapılır ve uygulamaya geçilir.

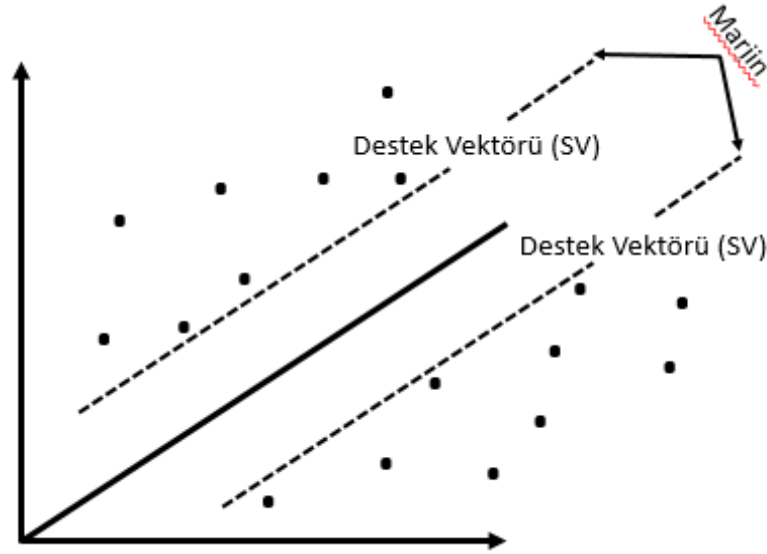
Tavsiye sistemleri kullanıcılar ve nesnelere arasındaki benzerliği bulmada genel olarak sorun yaşarlar. Ve bu sorunu çözmek için de benzerlik metrikleri, sınıflandırma algoritmaları veya nesne kümeleme gibi teknikler kullanılır. Bazı sınıflandırma teknikleri;

- **Simple Bayesian Classification (SBC) [32];**

Simple Bayesian sınıflandırıcı, çoğu sınıflandırma işlemlerinde etkin olarak kullanılan bir algoritmadır. Bir dokümanın  $c_i$  sınıfına ait olma ihtimali  $P(c_i|X)$  olarak ifade edilebilir. Bayes teorisi kullanılarak  $P(c_i|X)$  değeri  $P(X|c_i) \cdot P(c_i) / P(X)$  şeklinde ifade edilebilir. Burada en yüksek değeri verecek  $c_i$  sınıfı arandığından  $p(X)$  değerinin hesaplanmasına gerek yoktur.  $P(X|c_i)$  değeri  $c_i$  sınıfında  $X$  dokümanının oluşma olasılığını gösterir ve eğer  $X$  dokümanı  $(a_1, a_2, a_3, \dots, a_n)$  kelimelerinden oluşuyorsa  $P(X|c_i)$  değerini  $P(a_1|c_i) \cdot \dots \cdot P(a_n|c_i)$  şeklinde ifade edebiliriz. Başka bir şekilde ifade edecek olursak; dokümanın kategorisi, doküman içindeki tüm kelimelerin kategorilerinin çarpımları olarak belirtilebilir. Formül çarpma işlemi gerektirdiğinden Tüm özelliklerin sıfırdan büyük olması gerekmektedir. Tüm dokümanlar 0'dan büyük ihtimalle tüm sınıflara yakındır sonucunu bu formülden çıkarmak mümkündür. Dokümanı da en yüksek ihtimalli olduğu sınıfa veya kategoriye dahil etmek yerinde olur. Bununla birlikte, bayesian sınıflandırıcı çok fazla özellik karşılaştırılacaksa işlemler karmaşık

hesaplamalar gerektireceğinden özellikleri budama, eleme yaparak bu tekniği kullanmak gerekir [33].

- **SVM (Support Vector Machine) [34];** Makine öğrenmesi metotları arasında son dönemde popüler doğrusal olmayan sınıflandırma teknikleri arasında geçer. Regresyon analizlerinde de ağırlıklı olarak kullanılır. İki sınıf arasında, iki sınıfa da yakın olan noktaların oluşturduğu vektörlere destek vektörleri denir. Sınıflar birbirinden bu vektörler vasıtasıyla ayrılır. Bir örneğin hangi sınıfa girdiği veya hangi sınıfa daha yakın olduğunu bulma SVM algoritması içinde ele alınır. Kernel tabanlı metotlardandır. Sınıflandırma ve regresyon analizlerinde kullanılabilir. İkili karşılaştırmalarda eşik değerleri de tanımlanarak kullanılır. Doğrusal sınıflandırmalarda da kullanılabilir. Sınıflar arasında sınıfları birbirinden ayıran sınırlar bulunmaktadır. Sınırlar arası marjin olarak tanımlanır. Sınırlar üzerindeki noktalar destek vektörleri (Support Vectors) olarak tanımlanır. Şekil 2.6 'da örnek şekil gösterilmektedir.



Şekil 2.5 Support Vector Machine

- **Nearest Neighbour Classifier (KNN);** KNN, belli sınıflara dahil olmuş nesnelere içine, sınıflandırılması belli olmayan yeni bir nesne geldiğinde, yeni nesneyi de, bu yeni nesneye en çok benzeyen sınıflandırılmış nesnelere dayanarak sınıflandırma işlemidir. KNN, dokümanları sınıflandırmak için kullanılabilir gibi, kişileri sınıflandırmak için de tavsiye sistemlerinde sıkça kullanılmaktadır. Algoritma şu şekilde çalışır [35];

- Sınıflandırılmamış yeni örneklerle birlikte K sabit sayısı belirlenir.
- Yeni örneğe en çok benzeyen k nesne seçilir.
- Nesnelerin çoğunun dahil olduğu sınıf seçilir.
- Yeni örneğe bu sınıf verilir.

Yeni bir haber dokümanı üzerinden örnek verecek olursak: Bu habere en çok benzeyen on haberi seçtiğimizi düşünelim ( K on olarak belirlendi). Benzerlik metriklerini kullanarak haberimize en yakın on haberin sınıflandırmasının da iki magazin, dört politika, iki ekonomi ve iki güncel olduğunu var sayalım. Bu durumda haberimiz politika haberi olarak sınıflandırılmış olur. KNN, birçok benzerlik metriğiyle birlikte kullanılabilir.

Sınıflandırma algoritmalarının da kendilerine özgü bazı problemleri vardır. Bunlardan en önemlileri;

- Çok özellik barındırma (**High Dimensionality**) ; Özellik sayısının fazla olması hesaplama maliyetini getirmektedir. Farklı boyut azaltma teknikleri kullanılmaktadır. Bunlara genel olarak az kullanılan özelliklerin kaldırılması ve dikkate alınmaması olarak kullanılabilir. *Principal Concept Analysis(PCA)* ve *Single Value Decomposition (SVD)* [36] uygulanabilir.
- Çoklu sınıflandırma (**Multi-classification** ) ; Bir dokümanın birden fazla sınıfa dahil olmasıdır ki, genellikle özellik sayısının az olmasından kaynaklanmaktadır.

### 2.5.3. Dokümanların Tavsiye Sistemlerinde Etkin Kullanımı

Dokümanlar yapısal hale getirildikten sonra cosine, euclidean, jaccard, KNN ve bayesian gibi benzerlik algoritmaları kullanılarak dokümanlar belli sınıflara ayrılmalıdır. Çünkü birçok kaynak doküman kaynakları anahtar kelimeleri çıkartmamakta ve hatta dokümanın kategorilendirmesini dahi yapmamaktadır. Sınıflandırma işlemi tavsiye sisteminin temelini oluşturmaktadır. Haber kaynakları için düşünüldüğünde, kullanıcılara okuduğu haberlere yakın haberleri tavsiye etmek için; okunan haberlere benzer ve aynı sınıfta haberlerin tespiti gerekmektedir. Spor haberleri okuyan bir kullanıcıya ağırlıklı olarak spor haberleri tavsiye edilmeli, politika ile ilgilenen kişiye daha çok siyasi haber tavsiye edilmelidir. Bu bağlamda, yeni haberleri, eski haberler ile yakınlık ilişkisi kurarak



sınıflandırmak gerekmektedir. Yakınlık ilişkisinin de benzerlik algoritmaları kullanarak tespit edilebileceğini belirtmiştik. Eğer kaynak miktarı çoksa veya sınıflandırma işlemi sağlıklı bir şekilde yapılamıyorsa kümeleme işlemi de yapılabilir.

Tavsiye sisteminin temeli, dokümanların ve kullanıcıların sınıflandırılması esasına dayanmaktadır. Dokümanların sınıfı olduğu gibi kullanıcıların da sınıfları bulunmaktadır. Kullanıcı sınıflandırma işlemi profillere dayanır. Kullanıcı profilleri statik olarak kullanıcıdan alınabileceği gibi, kullanıcının sistem ile etkileşiminden de elde edilebilir. Bu etkileşim neticesinde elde edilen profile dinamik profil denmektedir. Doküman tavsiye sisteminin temeli, kullanıcıların profilleri ile dokümanların içeriği arasında benzerliğin bulunması esasına dayanır [37]. Doküman sınıflandırmaları, daha ileri seviyede uygulamaya göre farklı şekillerde yapılabilir [38].

Doküman ve kullanıcı analiz işlemleri tamamlandıktan sonra son aşama, kullanıcı ve doküman verilerinin tavsiye motorunda işletilmesi ve tavsiyenin gerçekleşmesidir. Tavsiye motoru da kullanıcıların geçmişte sistem ile etkileşimi ve yeni dokümanların eski dokümanlar ile benzerliğinin irdelenmesi neticesinde kullanıcılara tavsiyeler sunmaktadır.

### **3. BENZER ÇALIŞMALAR**

Bugüne kadar içerik tabanlı filtreleme, işbirlikçi filtreleme ve karma yaklaşımlarda birçok çalışma yapılmıştır. Yapılan çalışmaların bir kısmı *Netflix, google, Yahoo* ve *Amazon* gibi ürünlere dönüşmüş ve tavsiye sistemleri her geçen gün daha da fazla günlük hayatımıza girmektedir. Haber analiz sistemini de bazı sistemlerle karşılaştırırken haber tavsiye sistemleriyle ve diğer tavsiye sistemleriyle ilişkisini ayrı ayrı inceleyerek karşılaştırmaya çalıştık.

#### **3.1. Doküman Tavsiye Sistemleri**

*Liu ve Dolan*, zaman-tabanlı bir yaklaşım uyguladılar [39]. Kullanıcıların makale okuma sürelerini, makalenin uzunluğunu ve kelimelerin makale üzerindeki yerlerinin de dikkate alarak bir sistem kurguladılar. Kullanıcının makale üzerinde

geçirdiği zamana bakarak hükümler vermek ve tavsiyelere bulunmak bazen yanlış sonuçlar doğurabilmektedir. Sistemimizde bu fikrin yanlış yorumlamalara açık olduğu düşünülerek kullanılmamıştır.

*Tan ve Teo*, PIN [40] ismini verdikleri bir sistem kurdular. Bu sistemde kullanıcılar başlangıçta profillerini ilgilendikleri kelime ve tamlamaları sisteme girerler. Sistem de daha sonra bu terimlere dayanarak en uygun tavsiyelerde bulunmaya çalışır. Kullanıcının tercihlerinin zaman içinde değişeceği bu sistemde çok dikkate alınmamıştır.

*Jiahui Liu, Peter Dolan ve elin Ronby Peterson* [39] tavsiye sistemlerinin başarısının kullanıcıların hali hazırdaki ilgilerinin bilinmesiyle ortaya çıktığını söylerler. Onlar, sistemlerini içerik-tabanlı filtreleme tekniği kullanarak geliştirdiler. Çünkü onlara göre en önemli sorun first-rater problemiydi. Kullanıcıların sistemi kullanmalarını ve yeterince oylama yapmalarını beklemek zaman alıcıydı. Profiller kullanıcıların sistemi aktif olarak kullanmaları neticesinde oluşuyordu. Bu da sistem açısından maliyetliydi. Oylama ve geri beslemeyi gizlilik nedeniyle kullanmıyorlardı. Kullanıcıların tek tek tıklaması haricinde genel eğilimi de takip ettiler. Kullanıcıların oylamalarını almamaları haricinde sayfada kalma sürelerini de hesaplamamaları sayfaların ne kadar okunduğu bilgisini muğlak kılar.

*Saranyak. G. ve G. Sudha* [7] 'nın önerdiği sistemde ise veriler otomatik olarak haber ajanslarından belli aralıklarla alınarak kendi veritabanlarına yazılmaktadır. Sınıflandırma haber ajansları tarafından yapılmaktadır. Standart bir algoritma kullanılması farklı sınıflandırmalara yol açabilmektedir. Bizim sistemimizde ise haberler farklı kaynaktan alınmakta ve standart bir algoritmayla sınıflandırma işlemi yapılmaktadır. Sistemlerinde kullanıcıların profillerine önem vermektedirler. Onlara göre kullanıcıların statik ve dinamik profilleri vardır. Dinamik profilleri her interaktif durumda güncellenmektedir. Bu da haberlerin sıkça değiştiği bir ortamda maliyetli olabilmektedir.

*Lei li, Daniel Knox* [41], tavsiye sistemlerini dört iş akışından oluşan bir süreç olarak tanımlarlar. Bu süreçler; haber seçimi, haber gösterimi, haber işleme ve kullanıcı profili oluşturmaktır. Sistem üç ana öğeden oluşmaktadır. Bunlar sırasıyla haber kümeleme, kullanıcı profili oluşturma, haber sınıflandırmadır. Haberleri kümelemek için LSH ve hiyerarşik kümeleme kullanırlar. Kullanıcı profillerini

oluştururken üç hususu göz önünde bulundururlar; haber içeriği, benzer erişim paternleri ve tercih edilen haberler. Kullanıcılar  $U=<Topic\ Vector,\ list\ of\ users,entities>$  olarak tanımlanırken, haberler de benzer şekilde profiller oluşturur:  $F=<T,P,E>$  . Bir sonraki adımda ise kullanıcı ve haber benzerlikleri bulunarak matris oluşturulur. Sistemlerini test etmek maksadıyla beş gün veri toplamış için ve sonraki on günde de haber tavsiyesi yapmışlardır. Değerlendirme aşamasında tatmin edici sonuç elde ettiklerini belirtmişlerdir. On beş gün yerine daha uzun süre tercih edilse belki daha farklı sonuçlar alınabilirdi. Bizim sistemimizde ise bu süre hayli uzun tutulmuştur.

*Pazzani* [42] ise karma bir model kullanmıştır. Bu modelde kullanıcıların okuduğu haberlere göre kısa dönemli ve uzun dönemli iki farklı algoritmayı birleştirerek kullanmıştır. Kısa dönemli algoritmayla birlikte KNN kullanmış, uzun dönemli algoritmayla birlikte Naive-Bayes kullanmıştır. Bu ikisini birleştirerek kullanıcılara kısa dönemde daha hızlı tepki verilmesini sağlamıştır. Kısa dönemde büyük çaplı eğilimlerden ziyade genel anlık eğilimlere göre haber tavsiye edilmesini sağlamış, uzun dönemde ise naive-bayes kullanarak büyük çaplı eğilimlere göre haber tavsiyesinde bulunulmuştur. Benzer bir yaklaşım bizim sistemimizde de mevcuttur. Kullanıcıların son okuduğu haberlerin ağırlığı arttırılarak önem dereceleri yükseltilmiştir.

*Streamrec* sistemi [43] ise gerçek zamanlı bir sistem olması sebebiyle bizim sistemimize benzer bir sistem olarak tasarlanmıştır. Bu sistemde haberler gruplanmış ve kullanıcının haberi ilk olarak mı okuduğu yoksa yeniden mi değerlendirdiği bilgisi üzerinde durulmuştur. Kullanıcı yeniden değerlendirdi ise yeni oylaması zaman ile birlikte tutularak o zaman içinde aynı habere benzer oylamayı verenler üzerinden tekrar gruplandırma yapılmıştır. Ve kullanıcıya yeni haber tavsiyelerinde bulunulmuştur. Bu sistemin düzgün ve verimli çalışması aynı haberi çok kişinin okuduğu ve farklı zamanlarda tekrar gözden geçirdiği sistemlerde daha anlamlı sonuçlar üretirken az verinin ve az okuyucunun bulunduğu sistemlerde sistemin değerlendirmesi son derece zor gözükmektedir.

Çağrılan bir web sayfası üzerinde durulan süreyi göz önüne alarak oluşturulan bir sistemde de [44] mevcuttur. Google news sisteminde haber sayfası üzerinde durulan süreye bakılarak bir sistem oluşturulmuştur. Bu sistemde haberin uzunluğu dikkate yeterince alınmamış ayrıca haberin okunup okunmadığı bilgisinin

tespitini yapmanın da zorluğu çok göz önünde bulunulmamıştır. Kullanıcıların haber üzerinde durduğu süre ile haber kelime sayısı arasında oran ilişkisi kurarak değerlendiren zaman tabanlı tavsiye sistemlerinden birisi de *Liang ve Lai* tarafından tasarlanmıştır [45].

### 3.2. Haber Tavsiye Sistemleri

*News-Dude* [42] tavsiye sisteminde, kullanıcının araştırmasına göre altı farklı haber kaynağından haber sunulmaktadır. Daha sonra kullanıcı haberi ilgili veya ilgisiz şeklinde onaylamaktadır. Tavsiyeler bu sistemde kısa ve uzun vadeli olmak üzere ikiye ayrılmış, kısa vadelide *KNN*, uzun vadelide ise *naive-bayes* kullanılmıştır. Bu sistem karma bir sistem olarak değerlendirilebilir. Sunulan haberlerin oylanması bizim sistemimizde de mevcuttur. Sistemimizde beş farklı kategori kullanılmış, ayrıca kullanıcıların oylama yapmaması durumunda da kullanıcıların okuduğu haberlere dayanarak da profil güncellemesi yapılmaktadır.

*Gong'ta* [46] işbirlikçi filtreleme kullanmış, haberleri ve kullanıcıları kümeleyerek bu kümeler arasındaki ilişkilere dayanan bir sistem tasarlamıştır. Bu sistemdeki büyük problemlerin başında kullanıcı kümelerinin oluşturulması yeni haberlerin kümelenmesi ve algoritmanın fazla karmaşık olması gelmektedir. Haber sistemleri için bu işin biraz karmaşık olduğu düşünülebilir.

Çok ilginç bir haber tavsiye sistemi de *hot-news-recommendation* sistemidir [47]. Bu sistemi kısaca şöyle özetleyebiliriz: Sıcak haber tavsiye sisteminde tavsiyeler, o gün yayımlanan birçok haber sitesinde dikkat çeken en sıcak ve önemli haberleri kapsamaktadır. Bu haberler kişiye özel değil, kişilerin tıklamasıyla oluşmuyor ve kişilerin profillerine de ihtiyaç duymuyor. Haber başlıklarında ki önemli dört kelime seçilerek haberin diğer haberlerle olan benzerliği bayesian sınıflandırma algoritması kullanılarak bulunuyor. Daha sonra ise en sıcak ve önemli 20-50 haber kullanıcılara sunuluyor. Sistemden anlaşılacağı gibi sistem kişiye özel değil, grup müşterilerine özel olarak sunuluyor. *Hot-news-recommendation* sistemi genel hatlarıyla bizim sistemimize benziyor. Bizim sistemimizde okuyucuya sunulacak en iyi haber sayısını bulmak için optimum haber sayısı algoritması geliştirdik ve değerlendirmesini *ROC Curve* kullanarak bulduk. Benzer şekilde de bu sistemde belli haber sayısı bulunmuş ve bu sayının 100 olduğuna karar verilmiş. Bu sistem,

pratik ve gruba hitap edebilecek güzel haberleri sunarken değerlendirme kısmı tam yapılamamaktadır. Bunun yanında, sistemin benzeri olmadığından değerlendirme yaparken hem karşılaştırma yapılmamış, hem de değerlendirme aşamasında 4-5 uzman personele tüm haberler arasından en dikkat çekici olan haberleri belirlemeleri istenmiş ve bu haberlerin sistemin belirlediği haberler ile örtüşüp örtüşmediği üzerinden değerlendirme yapılmış.

İlginç olan bir diğer tavsiye sistemi de karma tavsiye sistemi olarak geliştirilen *Vertical News Recommendation System* (Dikey haber tavsiye sistemi) [48]. Dikey haber tavsiye sisteminin geleneksel haber tavsiye sisteminden farkları şu şekilde çıkarılmıştır. Sistem kampüs ortamında denenmiş, bu nedenle kullanıcı tipleri öğretmen, öğrenci gibi belirli profillere sahiptir. Haber çeşitleri kısıtlı sadece kampüsü ilgilendiren haberler. Kullanıcıların isteği de ilginç olan haber değil, kendisini ilgilendirecek haberin seçilmesi olmuştur. Ayrıca kullanıcılar üye olduğu grupların profillerini de otomatik olarak almaktadır. Bunun yanında okudukları haberler de dinamik profillerini oluşturmada etkilidir. Karma bir sistemde dikkat edilecek üç soru vardır: Hangi metotları birleştireceksiniz, neden birleştireceksiniz ve nasıl birleştireceksiniz [49]. Dikey haber tavsiye sisteminde de bizim sistemimizde olduğu gibi *Jaccard* sınıflandırması kullanılmıştır. Kullanıcıya haber tavsiye edilirken kullanıcı profili oluşturulmuş, haber profili oluşturulmuş, daha sonra aralarındaki ilişki *jaccard* benzerliği, ilişki ve kümeleme gibi üç metotla belirlenerek çerçeve daraltılmış ve neticesinde bir skor oluşturulmuş, daha sonra da en yüksek skorlu haberler kullanıcıya tavsiye edilmiştir. Dikey haber tavsiye sistemi ismini her aşamada, kısıtlamalar neticesinde çerçevenin daralmasından almıştır.

Bir başka haber tavsiye sistemi de anahtar kelimelere dayalı tavsiye sistemi olarak kullanılmıştır [50]. Bu sistemde de komşuluk ilişkisi bulunmuş, doküman seviyesinde anahtar kelimeler bulunmuş ve filtreleme yapılarak belli sayıya indirgenmiştir.

## 4. HABER ANALİZ SİSTEMİ

### 4.1. Genel Anlatım

Bu kısım, tavsiye sistemimizin kısaca anlatımı, şekil olarak gösterimi, kullanılan teknikler ve sistemin safhalarını kapsamaktadır. İnternet sitelerine dayandırılarak oluşturulan tavsiye sistemleri, esas verinin hazırlanması ve örneklemin oluşturulmasına dayanmaktadır [51]. Bizim sistemimizde, kaynaklar internet siteleri olmakla birlikte uygun verinin elde edilmesi ve verinin çekilmesi ilk aşamadır.

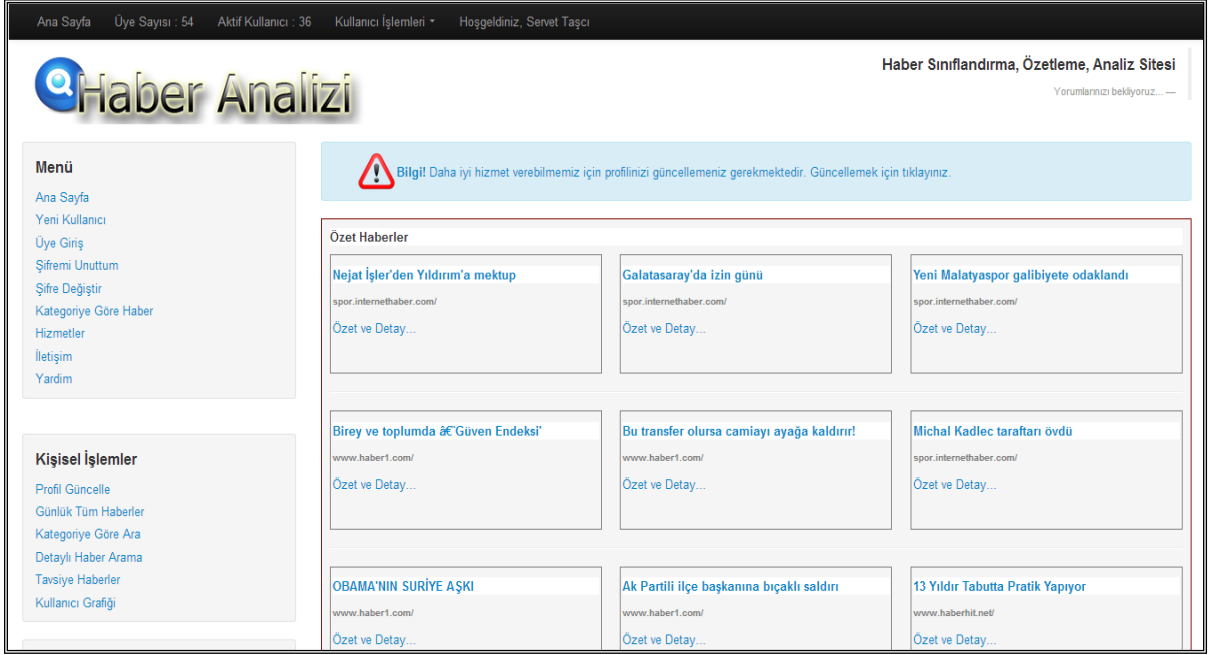
Sistemimiz temel olarak kullanıcılara sunumun ve sonucun gösterildiği bir haber internet sitesinden ([www.haberalizi.org](http://www.haberalizi.org)) ve tavsiye sisteminin motorunu oluşturan bir masaüstü uygulamasından oluşmaktadır. Masaüstü uygulaması, haberlerin çekilmesi, sınıflandırılması, özet çıkarılması, anahtar kelimelerin bulunması ve tavsiye motoru gibi birçok modülü barındırmaktadır. Tavsiye sistemlerinin safhaları Tablo 4.1 de gösterilmektedir.

Kaynaklar	Tavsiye Motoru	Sunum
Çeşitli İnternet Siteleri	Haberlerin Çekilmesi Anahtar Kelime Çıkarımı	Tavsiye Değerlendirilmesi Tavsiye Gösterimi
	Haberlerin Sınıflandırılması	
	Haberlerin Özetlenmesi	
	Haber Tavsiyesi	
	Geri Besleme	

Tablo 4.1 İş Akışı Şematik Gösterimi

Kısaca sistemi anlatacak olursak: İlk olarak, belirlediğimiz ortalama yirmi haber sitesinden günlük olarak yaklaşık bin beş yüz civarında haberi çekiyor ve ardından haberlerin anahtar kelimelerini tespit ediyoruz. Daha sonra işlenebilir hale gelmiş haberleri, sınıflandırma modülü vasıtasıyla sınıflandırmasını yapıyor ve haberlerin özetlerini çıkarıyoruz. Haberlerin özetlerinin çıkarımı, tavsiye sistemiyle doğrudan

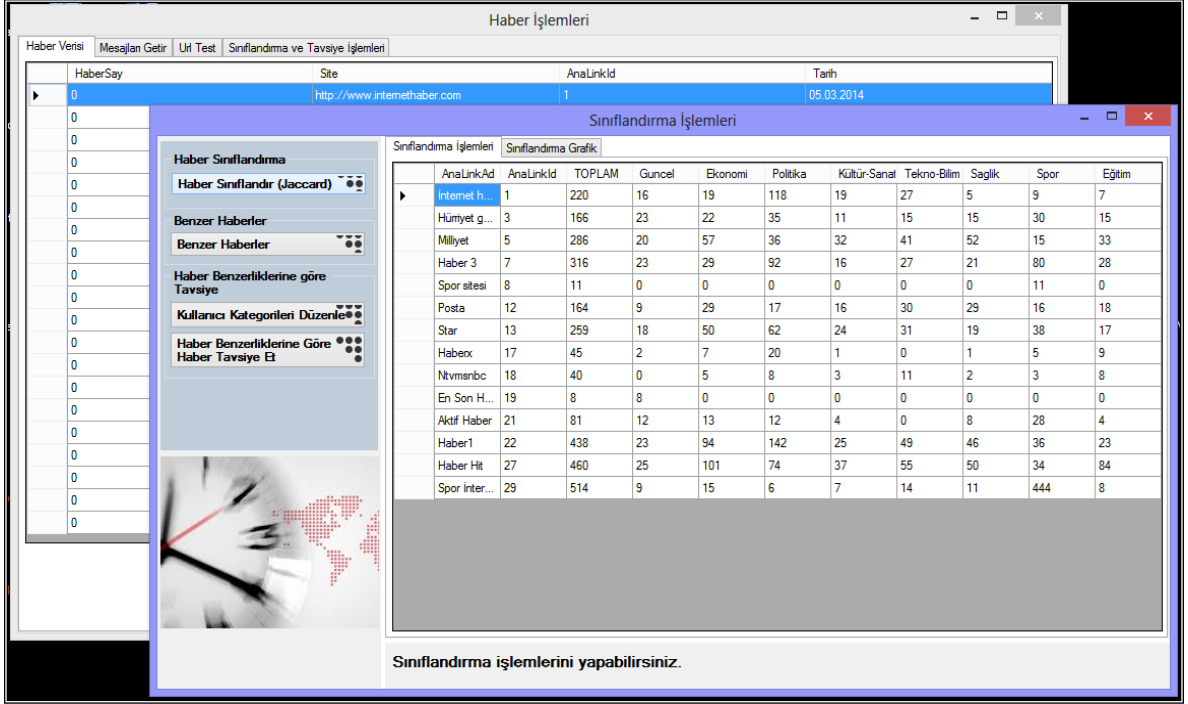
ilişkili olmasa da kullanıcılara haber seçiminde kolaylık sağlamaktadır. Sonraki aşamada ise, internet sitemize üye olan ve statik profillerini ilk aşamada oluşturan kullanıcılara haber tavsiyesi yapıyoruz. Okunan ve oylanan haberlere göre de kullanıcı profillerini güncelliyor ve sistemimizin başarısını kullanıcılardan aldığımız geri beslemeler vasıtasıyla ölçüyoruz.



The screenshot shows the Haber Analizi website interface. At the top, there is a navigation bar with links for 'Ana Sayfa', 'Üye Sayısı : 54', 'Aktif Kullanıcı : 36', 'Kullanıcı İşlemleri', and 'Hoşgeldiniz, Servet Taşcı'. The main header features the 'Haber Analizi' logo and the text 'Haber Sınıflandırma, Özetleme, Analiz Sitesi'. A blue banner with a warning icon and text reads: 'Bilgi! Daha iyi hizmet verebilmemiz için profilinizi güncellemeniz gerekmektedir. Güncellemek için tıklayınız.' Below this, the 'Özet Haberler' section displays a grid of news articles. Each article card includes a title, a source URL, and a link to 'Özet ve Detay...'. The articles shown are: 'Nejat İşler'den Yıldırım'a mektup' (spor.internethaber.com/), 'Galatasaray'da izin günü' (spor.internethaber.com/), 'Yeni Malatyaspor galibiyete odaklandı' (spor.internethaber.com/), 'Birey ve toplumda 'Güven Endeksi'' (www.haber1.com/), 'Bu transfer olursa camiayı ayağa kaldırr!' (www.haber1.com/), 'Michal Kadlec taraftarı övdü' (spor.internethaber.com/), 'OBAMA'NIN SURIYE AŞKI' (www.haber1.com/), 'Ak Partili ilçe başkanına bıçaklı saldırı' (www.haber1.com/), and '13 Yıldır Tabutta Pratik Yapıyor' (www.haberhit.net/). On the left side, there is a 'Menü' section with links for 'Ana Sayfa', 'Yeni Kullanıcı', 'Üye Giriş', 'Şifremi Unuttum', 'Şifre Değiştir', 'Kategoriye Göre Haber', 'Hizmetler', 'İletişim', and 'Yardım'. Below the menu is a 'Kişisel İşlemler' section with links for 'Profil Güncelle', 'Günlük Tüm Haberler', 'Kategoriye Göre Ara', 'Detaylı Haber Arama', 'Tavsiye Haberler', and 'Kullanıcı Grafiği'.

**Şekil 4.1 Haberalanizi.org Sitesinde Haberlerin Gösterilmesi**

Sistemimizi görsel olarak anlatacak olursak, şekil 4.1 **haberalanizi.org** alan adlı sistemi uyguladığımız siteden bir görünümüdür. Kullanıcı siteye kullanıcı adıyla giriş yaptığında günlük haberler karşısına çıkmaktadır. Sol menüde, kullanıcıya kategoriye göre haber arama, detaylı haber arama ve Tavsiye Haberler gibi menüler yer almaktadır. Ayrıca, kullanıcılar profillerini güncelleme işlemlerini de site üzerinden yapabilmektedir. Okunan haberler ile ilgili oylama işlemleri de yapılabilmektedir. Oylama neticesinde dinamik profiller oluşturmakta ve güncellemektedir.



Şekil 4.2 Program Çalışması Görünümü

Şekil 4.2' de haber sitelerinin taranması, özetlenmesi, sınıflandırılması gibi tavsiye sisteminin motoru olarak çalışan masaüstü uygulaması yer almaktadır. Masaüstü uygulaması sayesinde bağlantı ve zaman aşımı gibi sorunlar minimum seviyeye indirilmiştir. Program geliştirildikçe pek çok optimizasyon işlemleri de yapılmıştır. Günlük olarak oluşturulan tavsiyeler, kullanıcılara **habernalizi.org** alan adlı site üzerinden gösterilmektedir.

Tezimizi birbirine benzeyen iki farklı tavsiye sistemi üzerine kurguladık. Bu iki sistem;

- Anahtar Kelimelere Dayalı<sup>29</sup> Tavsiye Sistemi
- Zaman Bazlı<sup>30</sup> Tavsiye Sistemi

Anahtar kelime bazlı sistemi ile zaman bazlı tavsiye sistemi karşılaştırarak değerlendirmesini de son aşamada yaptık. Değerlendirmeler neticesinde zaman bazlı tavsiye sistemini ana sistemimiz olarak kullandık. Karşılaştırma sonuçları tezin beşinci kısmında anlatılmaktadır. Öncelik olarak şunu belirtmek gerekir ki, iki sistemin de anahtar kelimelere dayalı olması sebebiyle çok ciddi benzerlikleri

<sup>29</sup> Ing : Keyphrase-based

<sup>30</sup> Ing : Time-basec



vardır. İki sistem de temel olarak kelime ve kelime benzerliklerini kullanmaktadır. Haberler arasındaki benzerliği zaman bazlı sistem, tüm kelimeler üzerinden yaparken diğer sistem belirlenen anahtar kelimeler üzerinden yapmaktadır. Öncelikle zaman-bazlı tavsiye sistemi, daha sonra anahtar kelime bazlı sistem anlatılacak ve iki sistemin karşılaştırması yapılacaktır

## **4.2. Zaman Bazlı (Time-based) Tavsiye Sistemi**

Haber tavsiye sistemi beş aşamadan oluşmaktadır. Bu aşamalar;

1. Haber Sitelerinden Haberin Çekilmesi
2. Sınıflandırma İşlemleri ve Haber Sınıflandırması
3. Özetleme Sistemleri ve Haber Özetlenmesi
4. Profil Oluşturma
5. Haber Tavsiye Edilmesi

### **4.2.1. Haber Sitelerinden Haberin Çekilmesi**

İnternet sitelerinin her birinin tasarımın standartlarının ve içerik düzenlemesinin birbirinden farklı olması sebebiyle içerik analizi yapmak son derece güçtür. Güç olmasının sebeplerinden birisi de HTML dilinin esnek, eski ve kuralsız olmasından kaynaklanmaktadır. Örneğin bir internet sayfasında haber başlığı h1..h6 tagları arasında yer alırken başka bir sitede başlıklar sillerle düzenlenmiştir. Bir sitede haber içeriğine ulaşmak çok kolayken başka bir sitede haber haricinde içerikte reklam ve görsel öğelerin çokluğundan dolayı haber içeriğine ulaşmak son derece zor olmaktadır. Dolayısıyla haber içeriğini alırken her sitenin kendisine özgü özelliklerini incelemek ve ona göre çözümler bulmak gerekmektedir.

Belli başlı haber sitelerini incelediğimizde, haber içeriklerinin yaklaşık olarak dört ile yirmi cümle arasında sıkıştırıldığını gördük. Genel olarak, sitelerin haber içerik standartlarının olduğunu ve her sitenin kendi standartlarını devam ettirdiğini tespit ettik. Haberlerin sadece ilgili bölümlerini çıkarmak için de XPATH ifadeleri [52] kullandık. XPATH ifadeleri bize xml, html, xhtml biçimindeki dokümanları parçalara ayırarak belli taglara ulaşmamızı sağlamaktadır.

Bazı örnek XPATH ifadeleri ve açıklamaları;

- //h1, dokümandaki h1 tagını bulur.

- //h1[@class='title'] h1 tagı ve title sınıfını bulur.

Bazı haber sitelerinin XPATH ifadeleri Tablo 4.2 dedir. Bu ifadelerin mümkün olduğunca değişmemesi gerekir. İdeal bir sistem, haber sitelerinin haberleri belli formatlarla yayınlaması, rss kullanması veya servisler şeklinde açması şeklinde olabilir.

http://www.internethaber.com	//h1[@class='title']	//span[@id='contextual']
http://www.hurriyet.com.tr	//h1	//div[@class='txt']
http://www.milliyet.com.tr	//h1	//div[@class='text']
http://www.haberturk.com	//h1[@class='header']	//span[@id='news-detail-body']/p
http://www.haber3.com	//h1	//span[@id='contextual']
http://www.spor3.com	//h1[@class='title']	//span[@id='contextual']/p
http://www.taraf.com.tr	//span[@class='yazi_baslik']	//div[@class='yazi']
http://www.aksam.com.tr	//h1	//div[@class='double-wide']/p

**Tablo 4.2 Örnek Haber Siteleri XPATH İfadeleri**

Haber sitelerinden içeriği elde ettikten sonra başlık ve içerikle birlikte sistemimize kaydetmekteyiz. Sonraki aşamada elde edilen içeriğin sınıflandırma işlemini yapmaktayız.

#### **4.2.2. Sınıflandırılma İşlemleri ve Haber Sınıflandırması**

Haberlerin hangi kategori içerisinde değerlendirileceği doğrudan bildirilmemişse, kategorisinin tespit edilmesi son derece zordur. Örneğin, futbol maçı esnasında meydana gelen olayların anlatıldığı bir haber spor kategorisinde mi değerlendirilmeli, politik olarak mı, yoksa magazin olarak mı? Spor olarak değerlendirilmesi gereken bir haber magazin kategorisinde değerlendirildiğinde okuyan kişi bu haber ile ne kadar ilgilenir, bu haberi okur mu, okumaz mı? Haberi okumak isteyecek mi, yoksa istemeyecek mi? Bir kişi sadece spor haberlerini okuyorsa böyle bir haber tavsiye edilmeli mi? Edilecekse kaçınıcı öncelik olarak

tavsiye edilmeli. İşte bu ve benzeri soruların cevabını bulmak için öncelikle haberin sınıflandırılması gerekmektedir. Bazı haber siteleri bu sınıflandırmayı otomatik olarak yaparken, bazı haber siteleri ise bu sınıflandırmaya hiç değinmeden doğrudan baş sayfa haberi olarak vermektedir. Haberi sınıflandıran siteler de haberi farklı kategorilere ayırabilmektedir. Farklı sitelerdeki farklı kategorilendirme sorunlarının çözümü için biz haber sitelerini kendi belirlediğimiz standartlarda sınıflandırmaktayız.

Haber makalelerini, haber sitelerini inceleyerek genel eğilime uygun olacak şekilde sekiz farklı kategoriye ayırdık. Bu kategoriler; *Magazin, Ekonomi, Politika, Kültür-Sanat, Tekno-Bilim, Sağlık, Spor ve Eğitimdir*. Sonraki aşamada, kategorilere ait kelimeleri belirledik. Kategorilere ait kelimelerin neler olacağını ve hangi ağırlıklarla olacağı kısmını belirlemek için sınıflandırması yapılmış haberleri ve kelimeleri kullanarak bir havuz oluşturduk. Yaklaşık olarak her bir kategori, 400 ila 700 kelimedenden oluşmaktadır ve her kelimenin belirli bir ağırlığı vardır. Haber dokümanlarını *vektör uzay modeli* kullanarak inceledik. Doküman içinde geçen kelimelerin ağırlığını o doküman içinde geçen kelimeler içindeki ağırlığına ve tüm dokümanlarda kaç kere geçtiği bilgisini de kullanarak **TF\*IDF** değerini belirledik.

İçerik yapısal formda ise, yanı özellikler net ise *euclidean distance, jaccard distance*, değil ise *cosine similarity* kullanılabilir. Sınıflandırma algoritmalarından hangisinin daha iyi sonuç vereceğini tespit edebilmek amacıyla yayımladığı haberlerin sınıflandırmasını yapmış olan birkaç haber sitesinden çekmiş olduğumuz yaklaşık dört bin civarında haberin sınıflandırmasını belirlemiş olduğumuz benzerlik metrikleriyle test ettik. Sınıflandırması yapılmış olan derlemin metriklerle yaptığımız ölçümleri ve sonuçları Tablo 4.3' te verilmiştir. Derlem üzerinde yaptığımız sınıflandırma sonucunda ulaştığımız doğruluk sonuçları kullanılan metrikler için Tablo 4.3' te her sınıf için ayrı ayrı verilmiştir.

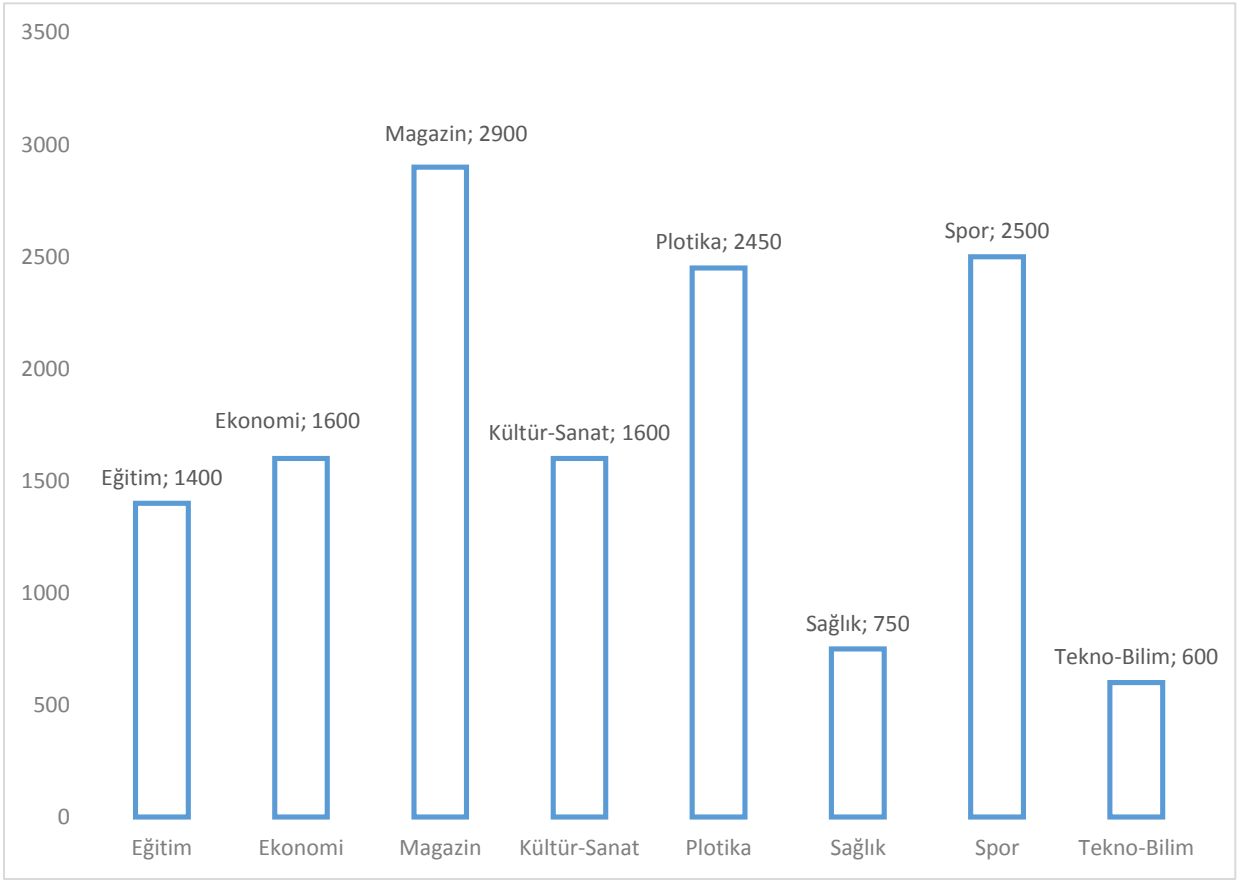
	Toplam	Jaccard	Cosine	Euclidean	SVM
Eğitim	240	0,65	0,52	0,4	0,6
Ekonomi	360	0,75	0,47	0,35	0,7
Magazin	370	0,65	0,5	0,42	0,6

<b>Kültür-Sanat</b>	250	0,7	0,45	0,4	0,7
<b>Politika</b>	490	0,67	0,4	0,3	0,72
<b>Sağlık</b>	700	0,82	0,45	0,6	0,75
<b>Spor</b>	920	0,8	0,7	0,65	0,7
<b>Tekno-Bilim</b>	370	0,75	0,55	0,35	0,7
<b>Toplam</b>	3700	0,72	0,51	0,43	0,7

**Tablo 4.3 Sınıflandırma Algoritması Doğruluk (Accuracy) Sonuçları**

Yaklaşık olarak dört bin haber üzerinde yaptığımız sınıflandırma çalışmasında en başarılı sonuçların *Jaccard* sınıflandırmasında olduğunu tespit ettik. Bu sonucun etkili olmasında doküman sınıflandırma işlemi kelimeye dayandırılmış olmamızın da payı büyüktür.

Sınıflandırma yapmadan önce şunu belirtmek gerekir ki; sınıflandırma kelimelerini isim ve özel isimleri kullanarak yaptık. Ayrıca kelimelerin köklerini tespit ederek benzerlik algoritmasını bu kökler üzerinde değerlendirerek kullandık. Örneğin "**Arabaya**" ve "**Arabada**" aynı köke sahiptir. Köküne gidecek olursak kökünün **araba** olduğunu tespit edebiliyoruz. Kelime köklerini *zemberek* [53] vasıtasıyla bulduk. Ayrıca yaklaşık olarak 150 kelimedenden oluşan *stopword*'leri (EK-A' da tam liste gösterilmiştir) değerlendirme dışında tuttuk. On günlük bir örneklemeden seçilen sınıflandırma sonucu, haberlerin kategori dağılımlarının incelenmesi açısından örnek olarak Şekil 4.3 de verilmiştir.



**Şekil 4.3 Örneklem Sınıflandırma Sonuçları**

Sınıflandırma neticesinde, düzgün dağılım olduğu görülebilmektedir. Spor haberlerinin fazla olması, eğitim, sağlık ve tekno-bilim haberlerinin az olması tahmin edilebilecek bir sonuçtur.

### 4.2.3. Özetleme Sistemleri ve Haber Özetlenmesi

#### 4.2.3.1. Özetleme Sistemleri

Geçmişte, bir konu hakkında bilgi bulma son derece zorken, günümüzde kaynakların artmasıyla birlikte veri miktarı da artmıştır. Bilgiye erişim süresinin kısalması için de bu bilgilerin derlenip kısa notlar halinde son kullanıcıya ulaştırılması ve kullanıcının tüm veriyi incelemek yerine özetlere bakarak karar vermesi ve veriyi bulması da önem kazanmıştır [54]. Bununla birlikte aranan içerik birçok doküman içerisinde yer alacağından bu dokümanlar arasından hangisinin daha faydalı ve etkin olduğunu bulmak ta son derece güçtür.

Bir doküman genellikle birçok konuyu içerisinde barındırır ve bazı konuları yüzeysel olarak ele alırken bazı konulara daha derince temas eder. Bu bağlamda özetleme metotları devreye girerek kullanıcıları gereksiz veriden kurtarabilir. Otomatik metin özetleme işlemleri devreye girerek bu sorunu çözmeye yardımcı olmaktadır.

Bir makale veya haber özeti ana konuları mutlaka ele almalı ve bunun yanında gereksiz özet barındırmamalıdır [55]. Bunun yanında genel olarak bir makalede bulunması gereken belli başlı özellikler vardır [56]. Bunlar; Girdinin karakteristiği, çıktının (özetin) metin olarak karakteristiği, özetin amacı vb.

Özetleme Metotları ne özetlediğiniz ve nasıl özetlediğinize göre kategoriye ayrılır [55] ve tek bir dokümandan çıkarılabileceği gibi birçok dokümandan da çıkarılabilir [56]. Birçok dokümandan çıkarıldığı takdirde bütünlük bozulmadan birleştirilebilmeli ve anlamlı bir şekilde kullanıcının hizmetine sunulmalıdır.

Doküman özetleme iki farklı şekilde yapılabilir; Bunlar cümle ayıklama ve özet çıkarmadır [57]. Özet çıkarma, NLP işleme, gramer ve lexicon çıkarma, ayıklama gibi karmaşık hesaplamaları barındırdığından uygulanması zordur. Buna karşılık, cümle ayıklama daha kolay uygulanabilir durumdadır. Biz bu metotlardan cümle ayıklamayı kullandık.

Özet çıkarılacak makale incelenirken özete esas teşkil edeceğinden şu konulara özel ağırlık verilmesi uygun olur;

- Makalenin başlığı: Makalenin başlığı, kısaca makalenin özeti olabileceği gibi, içerdiği kelimeler özete esas önem derecesi yüksek kelimelerdir.
- Kelime sıklığı: Çok kullanılan kelimeler (isimler) önemli olabilmektedir.
- Cümleler içinde geçen kelime sıklığını da dikkate alarak otomatik puan verilebilir. Puanlara normalleştirme işlemi uygulanmalıdır.
- Anahtar kelimeler<sup>31</sup> ana konuları ele alabilmektedir. Bu kelimeler tespit edilirse özet daha anlamlı olabilir.
- Cümlenin paragraf ve makale içindeki yeri cümlenin önemini belirtebilir.
- Alıntılar ya bölünmemeli ya da özetle hiç yer almamalıdır.

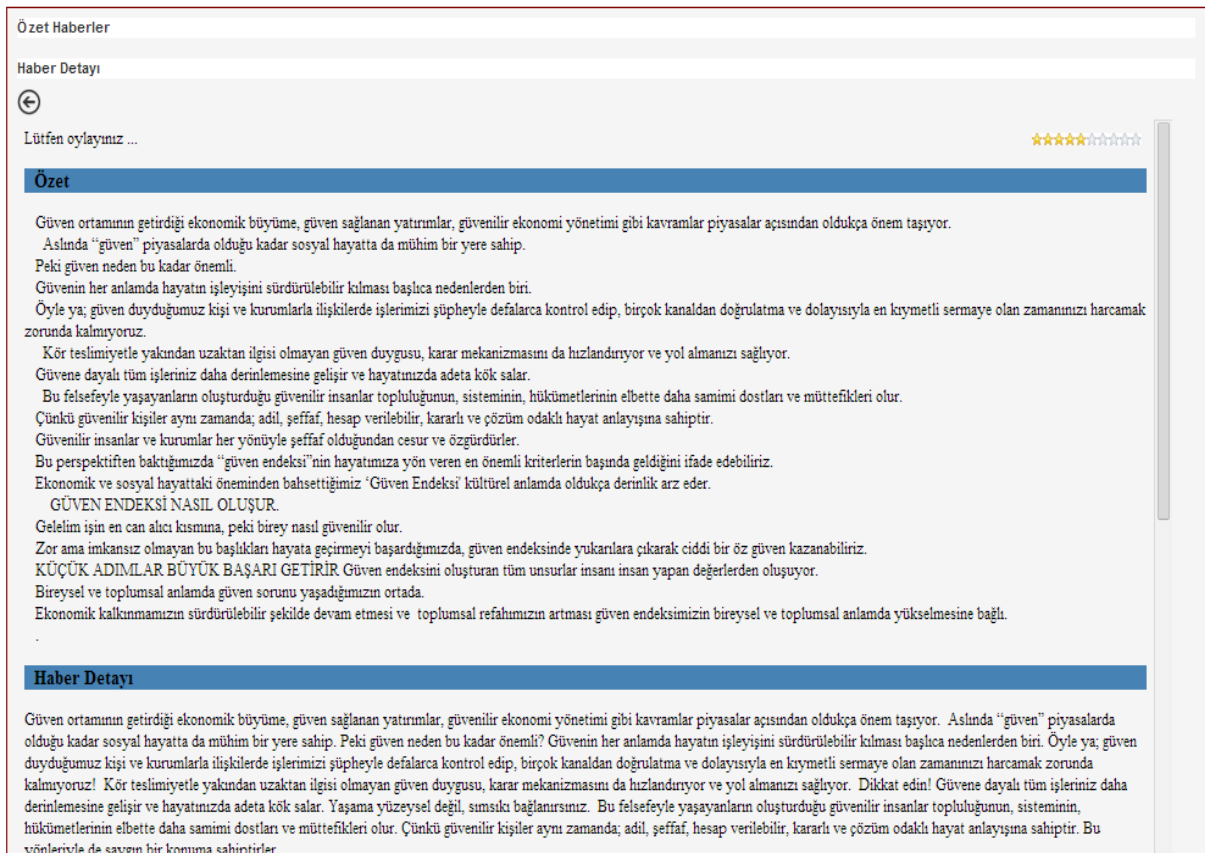
---

<sup>31</sup> Ing : Keyphrase

Yukarıda belirtilen ifadeler dışında bazı hususlar da uygulanan algoritmalarla göre önem kazanmaktadır.

#### 4.2.3.2. Haber Özetlenmesi

Haber analizi sisteminde cümle ayıklama metodunu kullanarak özetleme yaptığımızı belirtmiştik. Uyguladığımız sistem Cıgır, Kutlu ve Çiçekli [54] 'nin özetleme sistemidir. Ortalama olarak haber özetleri dört ila on iki cümle arasında olmaktadır. Özetleme algoritması var olan bir sistemin uygulanması şeklinde olduğundan detaylara girmeyip sadece sonuçlarıyla ilgileniyoruz.



The screenshot shows a news article interface. At the top, there is a section titled "Özet Haberler". Below it, the "Haber Detayı" section is visible. The article text is as follows:

Güven ortamının getirdiği ekonomik büyüme, güven sağlanan yatırımlar, güvenilir ekonomi yönetimi gibi kavramlar piyasalar açısından oldukça önem taşıyor. Aslında "güven" piyasalarda olduğu kadar sosyal hayatta da mühim bir yere sahip. Peki güven neden bu kadar önemli? Güvenin her anlamda hayatın işleyişini sürdürülebilir kılması başlıca nedenlerden biri. Öyle ya, güven duyduğumuz kişi ve kurumlarla ilişkilerde işlerimizi şüphesiz defalarca kontrol edip, birçok kanaldan doğrulama ve dolayısıyla en kıymetli sermaye olan zamanımızı harcamak zorunda kalmıyoruz. Kör teslimiyetle yakından uzaktan ilgisi olmayan güven duygusu, karar mekanizmasını da hızlandırıyor ve yol almanızı sağlıyor. Güvene dayalı tüm işleriniz daha derinlemesine gelişir ve hayatınızda adeta kök salar. Bu felsefeyle yaşayanların oluşturduğu güvenilir insanlar topluluğunun, sisteminin, hükümetlerinin elbette daha samimi dostları ve müttefikleri olur. Çünkü güvenilir kişiler aynı zamanda; adil, şeffaf, hesap verilebilir, kararlı ve çözüm odaklı hayat anlayışına sahiptir. Güvenilir insanlar ve kurumlar her yönüyle şeffaf olduğundan cesur ve özgürdürler. Bu perspektiften baktığımızda "güven endeksi"nin hayatımıza yön veren en önemli kriterlerin başında geldiğini ifade edebiliriz. Ekonomik ve sosyal hayattaki öneminden bahsettiğimiz "Güven Endeksi" kültürel anlamda oldukça derinlik arz eder.

**GÜVEN ENDEKSİ NASIL OLUŞUR.**

Gelelim işin en can alıcı kısmına, peki birey nasıl güvenilir olur. Zor ama imkansız olmayan bu başlıkları hayata geçirmeyi başardığımızda, güven endeksinde yukarılara çıkarak ciddi bir öz güven kazanabiliriz. **KÜÇÜK ADIMLAR BÜYÜK BAŞARI GETİRİR.** Güven endeksinin oluşturulan tüm unsurlar insanı insan yapan değerlerden oluşuyor. Bireysel ve toplumsal anlamda güven sorunu yaşadığımız ortada. Ekonomik kalkınmamızın sürdürülebilir şekilde devam etmesi ve toplumsal refahımızın artması güven endeksimizin bireysel ve toplumsal anlamda yükselmesine bağlı.

**Haber Detayı**

Güven ortamının getirdiği ekonomik büyüme, güven sağlanan yatırımlar, güvenilir ekonomi yönetimi gibi kavramlar piyasalar açısından oldukça önem taşıyor. Aslında "güven" piyasalarda olduğu kadar sosyal hayatta da mühim bir yere sahip. Peki güven neden bu kadar önemli? Güvenin her anlamda hayatın işleyişini sürdürülebilir kılması başlıca nedenlerden biri. Öyle ya, güven duyduğumuz kişi ve kurumlarla ilişkilerde işlerimizi şüphesiz defalarca kontrol edip, birçok kanaldan doğrulama ve dolayısıyla en kıymetli sermaye olan zamanımızı harcamak zorunda kalmıyoruz! Kör teslimiyetle yakından uzaktan ilgisi olmayan güven duygusu, karar mekanizmasını da hızlandırıyor ve yol almanızı sağlıyor. Dikkat edin! Güvene dayalı tüm işleriniz daha derinlemesine gelişir ve hayatınızda adeta kök salar. Yaşama yüzeysel değil, sınıksız bağlanırsınız. Bu felsefeyle yaşayanların oluşturduğu güvenilir insanlar topluluğunun, sisteminin, hükümetlerinin elbette daha samimi dostları ve müttefikleri olur. Çünkü güvenilir kişiler aynı zamanda; adil, şeffaf, hesap verilebilir, kararlı ve çözüm odaklı hayat anlayışına sahiptir. Bu yönleriyle de sizin bir konuma sahiptirler.

Şekil 4.4 Bir Haber ve Özeti Görünümü

Şekil 4.4 hakkında kısaca açıklama yapacak olursak; site üzerinde herhangi bir haberin detayını okumak için seçtiğinizde öncelikle haberin özetini ve hemen altında haberin detayını görebilirsiniz. Bu sayede kullanıcılar özete kısaca göz gezdirerek ilgilenmediği takdirde başka bir habere göz gezdirebilmektedir.

#### 4.2.4. Profil Oluřturma

Tavsiye sistemleri kullanıcılara haber tavsiye edebilmek için onların karakteristiklerine ve ilgi alanlarına ihtiyaç duymaktadır. Bu nedenle kullanıcı profillerini oluşturmak önemlidir, özellikle içerik tabanlı filtreleme kullanan sistemlerde profilin düzgün oluşturulması ve güncel tutulması daha da önemlidir. Profil oluşturulmadığında *ilk oylayıcı*<sup>32</sup> problemiyle karşılaşılacağını belirtmiřtik. Profil oluşturmak için birçok yöntem mevcuttur. Bu yöntemlerin birçoęu [58] makalede ele alınmıřtır.

Bireysel kullanıcıların profilleri kullanıcıdan istenerek doldurulabileceęi gibi, daha önce yapmıř olduęu tercihlere veya aktivitelere dayanarak otomatik olarak ta oluşturulabilir [58]. Deęiřtirilebilen profillere dinamik profil, zaman içinde aynı kalan profillere ise statik profiller denir. Pek çok tavsiye sistemi hem statik, hem de dinamik profili kullanarak tavsiyede bulunur. İki profil oluřturma metodunun da olumlu ve olumsuz yönleri bulunmaktadır.

Statik profilin faydası profilin bařlangıçta oluşturulabilmesi ve bu profile yönelik olarak tavsiye edilecek haberlerin kolayca bulunabilmesidir. Bunun yanında, sadece statik profili kullanarak tavsiye sistemi kurduęunuzda aynı kullanıcıya her zaman benzer haberler tavsiye edilecek ve kullanıcının zaman içindeki tercih deęiřiklikleri sisteme yansıtılamayacaktır.

Dinamik profil oluřturmada ise en büyük sorun bařlangıçta kullanıcının profili oluřmadıęından kullanıcıya tavsiye yapılamamasıdır. Fakat zaman içinde kullanıcının tercihleri deęiřtikçe profil otomatik olarak güncellenebilmektedir. Dinamik profiller kısa ve uzun dönemli olarak ikiye ayrılırlar [59][60]. Bununla birlikte zaman kısıtı konulmadan oluşturulmuř sistemler de mevcuttur.

İki metodun da artılar ve eksilerini ortadan kaldıracak řekilde profil oluřturma yöntemi belirlemek ideal olandır. Bizim sistemimizde iki metot da göz önünde bulundurulmaktadır. Kullanıcı sisteme ilk üye olduęunda statik profilini oluřturmasını bekliyoruz ve ilk tavsiyeleri de bu profile göre gerçekleřtiriyoruz.

---

<sup>32</sup> Ing : First-rater



Ana Sayfa Üye Sayısı : 54 Aktif Kullanıcı : 35
Kullanıcı İşlemleri · Hoşgeldiniz, Servet Taşcı

<p><b>Menü</b></p> <ul style="list-style-type: none"> <li><a href="#">Ana Sayfa</a></li> <li><a href="#">Yeni Kullanıcı</a></li> <li><a href="#">Üye Giriş</a></li> <li><a href="#">Şifremi Unuttum</a></li> <li><a href="#">Şifre Değiştir</a></li> <li><a href="#">Kategoriye Göre Haber</a></li> <li><a href="#">Hizmetler</a></li> <li><a href="#">İletişim</a></li> <li><a href="#">Yardım</a></li> </ul>	<div style="border: 1px solid #ccc; padding: 5px;"> <p><b>Kişisel Profil</b></p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; border-bottom: 1px solid #ccc;">E-Mail servet.boos@example.com</td> <td style="width: 50%; border-bottom: 1px solid #ccc;">Doğum Tarihi 12.12.1982</td> </tr> <tr> <td style="border-bottom: 1px solid #ccc;">Ad Servet</td> <td style="border-bottom: 1px solid #ccc;">Soyad Taşcı</td> </tr> <tr> <td style="border-bottom: 1px solid #ccc;">Güncel çatışma, kaza, trafik</td> <td style="border-bottom: 1px solid #ccc;">Ekonomi Borsa,imkb</td> </tr> <tr> <td style="background-color: #e0f0ff; padding: 2px;">Bilgi! Güncel haberlerle ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.</td> <td style="background-color: #e0f0ff; padding: 2px;">Bilgi! Ekonomi haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.</td> </tr> <tr> <td style="border-bottom: 1px solid #ccc;">Politika hükümet,siyaset,darbe</td> <td style="border-bottom: 1px solid #ccc;">Kültür-Sanat belgesel, sanat,tarih</td> </tr> <tr> <td style="border-bottom: 1px solid #ccc;">Tekno-Bilim microprocessor, computer,bilgisayar</td> <td style="border-bottom: 1px solid #ccc;">Sağlık kanser, kansızlık</td> </tr> <tr> <td style="border-bottom: 1px solid #ccc;">Spor galatasaray,fatih,terim</td> <td style="border-bottom: 1px solid #ccc;">Eğitim hacettepe, üniversite, derslane</td> </tr> <tr> <td style="background-color: #e0f0ff; padding: 2px;">Bilgi! Tekno-Bilim haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.</td> <td style="background-color: #e0f0ff; padding: 2px;">Bilgi!Sağlık haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.</td> </tr> <tr> <td style="background-color: #e0f0ff; padding: 2px;">Bilgi! Politika haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.</td> <td style="background-color: #e0f0ff; padding: 2px;">Bilgi! Kültür-Sanat haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.</td> </tr> <tr> <td style="background-color: #e0f0ff; padding: 2px;">Bilgi! Spor haberleri ile ilgili ilgilendiğiniz terimleri aralana "" koyarak</td> <td style="background-color: #e0f0ff; padding: 2px;">Bilgi! Fâtım haberleri ile ilgili ilgilendiğiniz terimleri aralana "" koyarak</td> </tr> </table> </div>	E-Mail servet.boos@example.com	Doğum Tarihi 12.12.1982	Ad Servet	Soyad Taşcı	Güncel çatışma, kaza, trafik	Ekonomi Borsa,imkb	Bilgi! Güncel haberlerle ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Bilgi! Ekonomi haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Politika hükümet,siyaset,darbe	Kültür-Sanat belgesel, sanat,tarih	Tekno-Bilim microprocessor, computer,bilgisayar	Sağlık kanser, kansızlık	Spor galatasaray,fatih,terim	Eğitim hacettepe, üniversite, derslane	Bilgi! Tekno-Bilim haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Bilgi!Sağlık haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Bilgi! Politika haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Bilgi! Kültür-Sanat haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Bilgi! Spor haberleri ile ilgili ilgilendiğiniz terimleri aralana "" koyarak	Bilgi! Fâtım haberleri ile ilgili ilgilendiğiniz terimleri aralana "" koyarak
E-Mail servet.boos@example.com	Doğum Tarihi 12.12.1982																				
Ad Servet	Soyad Taşcı																				
Güncel çatışma, kaza, trafik	Ekonomi Borsa,imkb																				
Bilgi! Güncel haberlerle ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Bilgi! Ekonomi haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.																				
Politika hükümet,siyaset,darbe	Kültür-Sanat belgesel, sanat,tarih																				
Tekno-Bilim microprocessor, computer,bilgisayar	Sağlık kanser, kansızlık																				
Spor galatasaray,fatih,terim	Eğitim hacettepe, üniversite, derslane																				
Bilgi! Tekno-Bilim haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Bilgi!Sağlık haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.																				
Bilgi! Politika haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.	Bilgi! Kültür-Sanat haberleri ile ilgili ilgilendiğiniz terimleri aralana "", koyarak yazınız.																				
Bilgi! Spor haberleri ile ilgili ilgilendiğiniz terimleri aralana "" koyarak	Bilgi! Fâtım haberleri ile ilgili ilgilendiğiniz terimleri aralana "" koyarak																				

**Şekil 4.5 Örnek Profil Ekranı**

Şekil 4.5 haberanalizi.org sayfamızdan alınmış profil görüntüleme ekranıdır. Bu ekranda kişiler statik profillerini görebilmekte ve kategorilere göre güncelleme yapabilmektedir. Statik profilin güncellenmesi ilk oylama problemini kısmen çözmek için kullanılmaktadır.

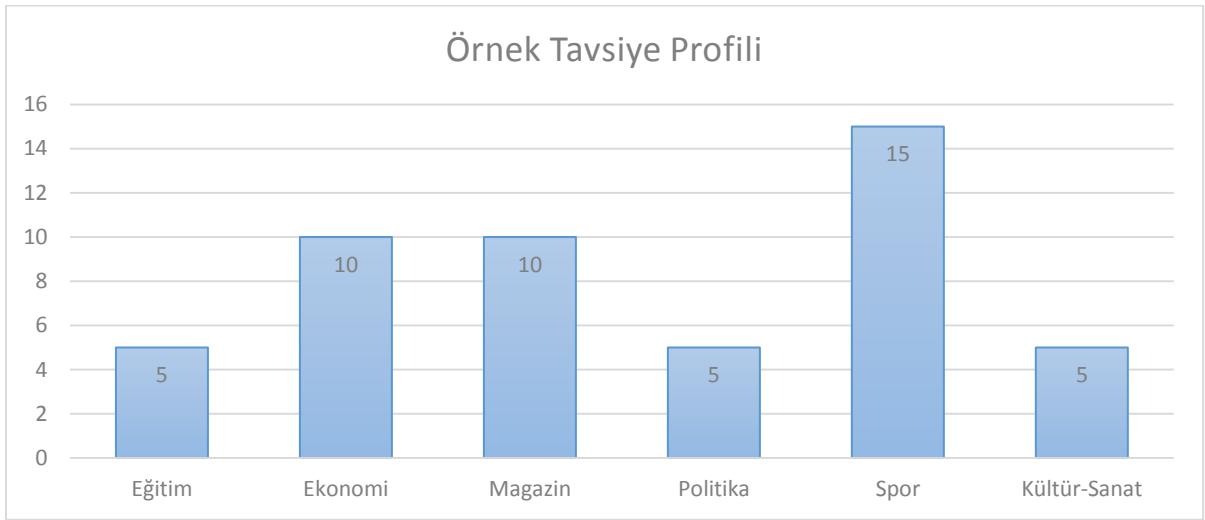
Statik ve dinamik profiller, her bir kategori için kategori ağırlığı ve ağırlıklandırılmış kategori kelimeleri tutan bir listeden oluşur. Kullanıcıların kategori ağırlık toplamları bire eşittir. Kategori ağırlıkları, kullanıcıya tavsiye edilecek haber sayısını belirler. Her bir kategorideki ilgili kelimeler, kelime ve kelimenin ağırlığından oluşan bir çift olarak tutulmaktadır. Kelimelerin ağırlığı bu kelimenin geçtiği dokümanlarla ilgilenen kullanıcı sayısını yansıtmaktadır ve normalize edilmiştir. Örnek bir profil;

Profile = {

Magazine	: 0.20, Magazin kelimeleri,
Economy	: 0.15, Ekonomi kelimeleri,
Politics	: 0.25, Politika ve siyaset kelimeleri,
Culture-Art	: 0, Kültür kelimeleri,
Techno-Science	: 0.12 ,Tekno-bilim kelimeleri
Health	: 0.08, Sağlık kelimeleri,
Sports	: 0.16, Spor kelimeleri,

Education : 0.04, Eğitim kelimeleri  
}

Zaman içinde, kullanıcı haber okudukça, haber oyladıkça, dinamik profil oluşmakta ve dinamik profil her gün güncellenmektedir. Tavsiyeler de benzer şekilde günde bir kez güncellenmektedir. Dinamik profil ağırlıklandırmasına göre son iki gün okunan haber kelimelerinin ağırlıkları 1,5 ile, sonraki iki günün ağırlığı 1,25 ile, sonraki 3 günün ise 1 ile çarpılmaktadır. Bu da kullanıcının son okudukları haberlerin önemini arttırmaktadır. **Zaman-bazlı tavsiye sistemi tanımı da buradan kaynaklanmaktadır.** Kullanıcılara haber tavsiye edilirken profilin nasıl kullanıldığı bir sonraki kısımda anlatılmıştır.



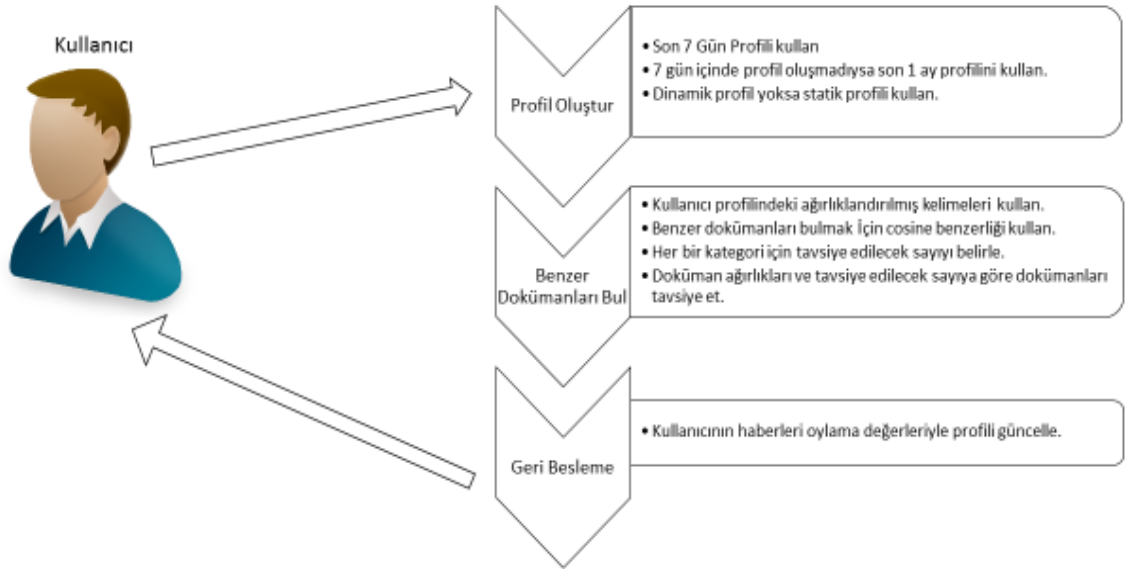
Şekil 4.6 Örnek Bir Tavsiye Profili

Şekil 4.6' daki örnek profili inceleyecek olursak, kullanıcının daha çok spor, ekonomi ve magazin haberlerini okuduğunu söylemek yanlış olmaz. Örnekteki kullanıcıya göre dağılım; Eğitim: 5, Ekonomi: 10, Magazin: 10, Politika: 5, Spor: 15, Kültür-Sanat: 5 şeklindedir.

#### 4.2.5. Zaman-bazlı Tavsiye Sistemi Algoritması

Algoritmayı açıklamadan önce şunu belirtmek gerekir ki kullanıcılara haber tavsiyesi yapıldığında, kullanıcılar benzer haberlerin de tavsiye edilmesini isteyebilir. Bu kullanıcının benzer haberleri okuyarak karşılaştırma yapmasını sağlar veya benzer haberleri hiç okumama tercihi de kullanıcının önüne sunulmuş

olur. Benzer bir sistem Yuanhua Lv [61] tarafından uygulanmıştır. Bu sistemde benzer haberler tespit edilerek kullanıcının sonra okuması için önüne sunulmuştur. Sistemimizin algoritmasını kısaca açıklayacak olursak, kullanıcıya haber tavsiye edilmesi için öncelikle kullanıcının profilinin oluşması gerekmektedir. Kullanıcının son yedi gün içinde oluşan dinamik profili (**PDlastsevendays**), son yedi günde haber okumadıysa kullanıcının bir önceki ayının dinamik profili (**PDlastmonth**), eğer o da oluşmadıysa kullanıcının statik profili (**PS**) kullanılır. Kullanıcının profili seçildikten sonra tavsiye bu profile dayanılarak yapılır. Her bir kategori için (**Ci**) kullanıcının kategori değeri ile tavsiye edilecek toplam haber sayısı çarpılarak o kategoride tavsiye edilecek haber sayısı (**NCi**) bulunur. Aynı kategorideki günlük haberler ile kullanıcının profilindeki kelimeler arasında benzerlik bulunarak kullanıcıya en yakın haberler tavsiye edilir. Bu tavsiye işlemi için *cosine similarity* kullanılmıştır. Karşılaştırma yapılmadan önce stopwords elenerek işlem yapılır. Haber başlıkları benzerlik hesaplanırken haber detayına göre daha fazla ağırlığa sahiptir. Zaman bazlı haber tavsiye sisteminin şematik olarak anlatımı Şekil 4.7’ de yapılmıştır.



**Şekil 4.7 Haber Tavsiye Sistemi Çalışma Şeması**

- Profil oluřturma safhasında kullanıcının yedi gn, bir ay ve yoksa statik profili kullanılarak profil oluřturulmuřtur. Statik profil de mevcut deęilse (Soęuk bařlangıç problemi) popler haberler tavsiye edilmiřtir.
- Benzer haberleri bulma esnasında cosine benzerlięi kullanılmıř, dokmanlar ile kullanıcı profili arasındaki benzerlik, kullanıcının okuduęu haberlerdeki kelimelerin aęırlıklarına gre bulunmuřtur. Kategorilere gre tavsiye edilecek sayı belirlenerek tavsiye sistemi kullanılmıřtır. Her bir kategori iin tavsiye edilecek sayı belirlenmiř ve benzer haberler aęırlıklarına gre sıralanarak kullanıcıya tavsiye edilmiřtir.
- Geri besleme vasıtasıyla kullanıcının dinamik profili gncellenmekte ve tavsiye motoru her gn tekrar alıřtırılmaktadır.

#### 4.3. Anahtar Kelime Bazlı<sup>33</sup> Tavsiye Sistemi

Kullanıcılar benzer ierikli haberleri okumaya meyillidirler. Haber tavsiyeleri de bu fikirden yola ıkılarak yapılabilir. oęu haber makaleleri aynı kaynaklı olma gibi benzer karakteristiklere sahiptir. Benzer haberler de genel olarak benzer anahtar kelime ve yapıları barındırır. Bu ıkarım, anahtar kelimelere dayalı tavsiye sistemlerinin temel fikrini oluřturur.

Anahtar kelimeler, dokmanın kısa bir zeti olduęu gibi, aynı zamanda dokman hakkında aydınlatıcı bilgiler de veren ifadelerdir [62]. Anahtar kelimeler dokmanın nemli noktalarına vurgu yapmaktadır. Akademik ve bazı dięer yayınlarda, anahtar kelimeler dokman hakkında bilgi vermeyi saęlamak iin yazarlar tarafından belirlenmiřtir. Dokman zetlemede anahtar kelimeler kullanılabileceęi gibi zetin ierięini belirlemede de etkili olmaktadır. Anahtar szckler dokmandan ıkarılacaęı gibi bazı kelime ve kelime grupları anahtar szckler olarak etiketlenebilir.

Anahtar kelime ıkartmadaki ama, dokman veya makale hakkında metadata oluřturma dır. Metadata ise dokmanın bir nevi zeti olarak deęerlendirilebilir. *KEA* [60] sistemi de dokman ierisinden anahtar kelimeleri ıkartmayı amalayan bir sistemdir.

---

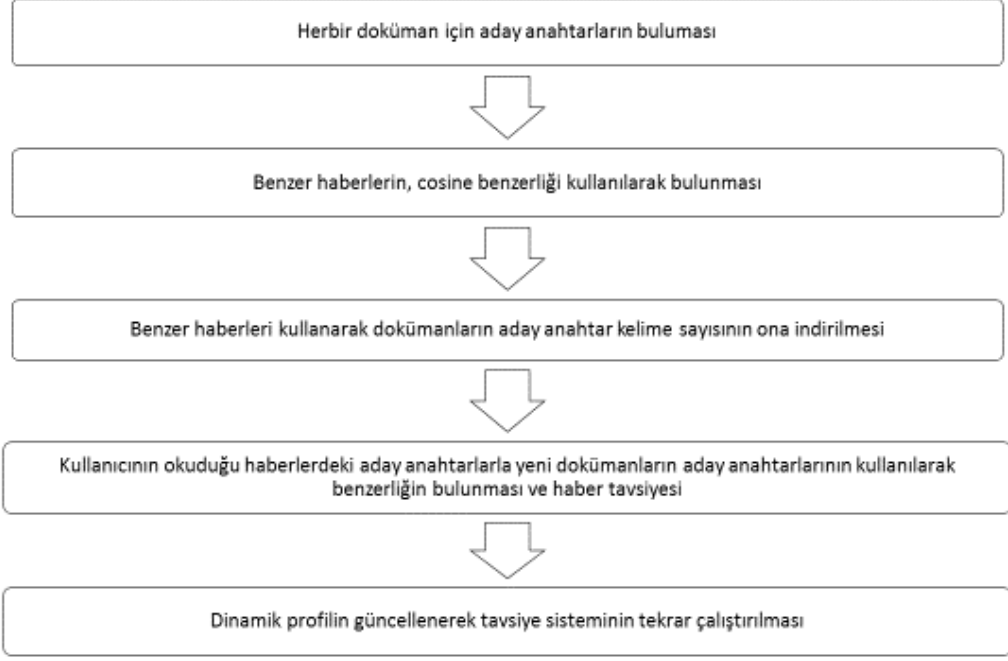
<sup>33</sup> Ing : Keyphrase-Based

Pala ve Cicekli'nin geliřtirmiş olduđu sistemde [64] dokümanlardan anahtar kelimelerin belirlenmesini kapsamaktadır. Bu sistem anahtar kelimeleri belirlerken Türkçe'ye özgü yapıları da dikkate alarak anahtar kelime çıkartmayı amaçlamıştır. Anahtar kelimelerin belli başlı amaçlarını sıraladıktan sonra sistemimizde neden kullanıldığından da kısaca bahsedebiliriz;

Anahtar kelimeler farklı amaçlar için kullanılabilir [65];

- Özetleme, anahtar kelimeler dokümanın bir özeti veya doküman hakkında fikir veren ifadeler olarak değerlendirilebilir.
- İndeksleme, dokümanları daha kolay bulmayı sağlar.
- Arama, anahtar sözcük kullanılarak yapılan aramalar daha etkili sonuçlar döndürebilir.
- Doküman sınıflandırmada da kullanılabilir [65].

Sistemimizde anahtar kelimeler dokümanın özetini çıkarmada ve tavsiye etme esnasında kullanılmaktadır. Anahtar kelime belirleme esnasına kadar yapılan işlemler iki sistem için de aynıdır. Haberlerin çekilmesi, sınıflandırılması, özetlenmesi, kelimelerin belirlenmesi işlemleri beraber olmakta, daha sonra anahtar kelime tavsiye sisteminde aday anahtar kelimeler çıkartılmakta ve filtreleme yapılarak muhtemel anahtar kelimeler belirlenmektedir. Muhtemel anahtar kelimeler belirlendikten sonra tavsiye sistemi çalışmaktadır. Haber makalelerinde anahtar kelimeler etkin olarak kullanıldığından haber analizi sisteminde de bu kelimelere dayalı bir sistem kullanmak yerinde olur. Anahtar kelime-bazlı sistemin şematik gösterimi Şekil 4.8 'de mevcuttur.



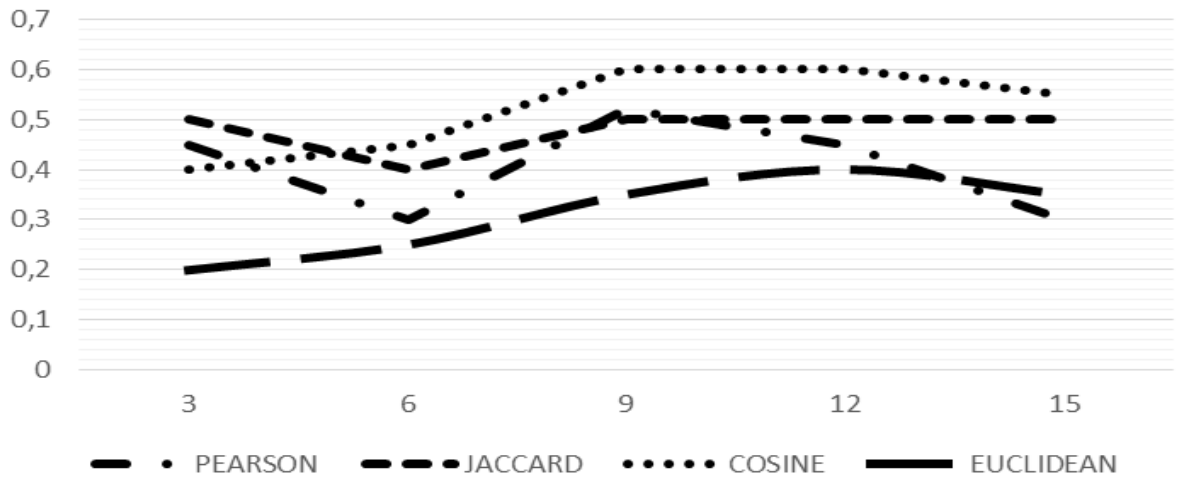
**Şekil 4.8 Anahtar Kelime Tavsiye Sistemi Akış Şeması**

Şekil 4.8 teki şema üzerinden giderek sistemi safhalandıracak olursak;

ilk aşama her bir doküman için aday anahtarların bulunmasıdır. Haber makalelerindeki anahtar kelimeleri bulmak çok da kolay değildir. Daha doğrusu, anahtar kelimeler yazar tarafından açık olarak belirtilmedikçe anahtar kelimeleri belirlemek için bazı metotlar kullanmak gerekir. Bunun için öncelikle aday anahtar kelimeler belirlenmeli ve ardından aday anahtar kelimeler mümkün olduğunca sınırlı bir sayıya indirgenmelidir. Aday anahtarları bulabilmek için sistemimizde KEA [62] uygulanmıştır. Bu anahtar sözcüklerin kaçının doğru olduğu ve ne kadar doğrulukla sonuçlar ürettiği değerlendirilmemiştir. Çünkü incelediğimiz haber sitelerinin hemen hemen hiç biri anahtar kelimeleri belirtmediği gibi her gün haberlerin yenilenmesi de bu anahtar sözcüklerin doğruluğunun bulunmasında zorluklar oluşturmaktadır.

İkinci aşama, benzer haberlerin bulunmasıdır. Benzer dokümanları bulmak için *vektör uzay modelini* ve *Rocchio* algoritmasını kullandık. Öncelikle dokümanları yapısal hale getirdik. Bu safha, zaten zaman bazlı haber tavsiye sisteminde uygulanmıştı. Benzer haberleri bulmak için farklı algoritmalar kullandık. Hangi algoritmanın daha başarılı olduğunu test etmek üzere benzerliği belirlenmiş iki makale dizisinden oluşan bir veri seti oluşturduk. Bu iki makale seti arasındaki

benzerliđi ve benzer haberleri bulmak için bazı benzerlik algoritmaları kullandık. Bunlar; *Cosine*, *Jaccard*, *Euclidean* ve *Pearson Coefficient* 'tir. Çıkan sonuçlar incelendiđinde (Şekil 4.9), *Cosine* benzerliđinin en iyi sonucu ürettiđi görölmektedir. Dolayısıyla *cosine* benzerliđi kullanılmıř ve benzer haberler her bir doküman için on haber ile kısıtlanmıřtır. Bu sayının artması durumunda aday anahtarların tespitinde sorunlar yařanmaktadır. Zaten, farklı kaynaklarda haberin birer kere yer aldığını ve yaklaşık olarak on farklı kaynaktan haber çekildiđini düşünürsek on sayısının uygun bir sayı olduđu görülecektir.



Şekil 4.9 Haber Benzerlik sonuçları

Üçüncü aşamada, aday anahtar sözcükler arasından gerçek aday anahtar sözcüklerin belirlenebilmesi için bazı çalışmalar yaptık. Bu çalışmamızın kaynađını tıbbi dokümanlardaki anahtar kelimeleri bulmak için yapılmıř olan sistem [66] oluşturmaktadır. Tıbbi dokümanları incelemek için yapılan çalışmadaki algoritma özetle şöyledir;

$X = PF * \log(N/DF)$  kullanılmıřtır.

**PF**, aday anahtarın dokümandaki sıklığı,

**N**, Doküman sayısı,

**DF**, aday anahtarın dokümanların kaçında bulunduđu ile ilgilidir.

Bu algoritma, temelde vektör uzay modelinden farklı deđildir. Aynı algoritmayı sistemimize uygun bazı deđişiklikler yaparak anahtar kelime bazlı tavsiye sisteminde kullandık. **N deđeri sabit ve on deđerini aldı.** DF ve PF deđerlerini de

bu deęerlere gre hesaplayabiliriz. DF deęeri sıfır ile on arasında olacaktır. Daha sonra bu algoritmayı kullanarak her bir doküman için aday anahtar kelimeler arasından on tanesini muhtemel anahtar kelime olarak kabul ettik.

Dördüncü aşama ise on adet anahtar kelime belirlenmiş haberlere dayanarak tavsiye sisteminin uygulanmasıdır. Bu aşamada kullanıcıların okuduęu haberler ile okumadığı haberler arasında anahtar sözcükler üzerinden ilişki kurarak tavsiye sistemimizi geliştirdik. Yeni haberler tavsiye edilirken son bir hafta içerisinde kullanıcının okumuş olduęu haberlerde geçen anahtar kelimelerle yeni haberler arasında benzerlik esasına dayanmaktadır. Bu benzerlik, anahtar kelimelerin kesişim kümesine göre tespit edilmiştir. Tavsiye edilen haber miktarı, zaman-bazlı tavsiye sisteminde tespit edildięi gibi en fazla otuzdur.

Son aşamada da her gün tavsiye sistemi çalıştırılmış ve kullanıcıdan dönen geri beslemeler ile kullanıcının dinamik profili güncellenmiştir. Tavsiye edilen haberlerle önceki haberler arasındaki benzerlik, kısmen de olsa benzer haberlerin tavsiye edilmesi sonucunu doğurmaktadır. Zaman-bazlı haber tavsiye sistemi ile anahtar kelime bazlı tavsiye sisteminin karşılaştırması değerlendirme metrikleri içerisinde ele alınmıştır.



## 5. Tavsiye Sistemi Deęerlendirme Yöntemleri

### 5.1. Deęerlendirme Metrikleri

Tavsiye sistemleri artık birçok modern uygulamada mevcuttur. Bu uygulamalarla birlikte tavsiye sisteminin deęerlendirilmesi de önem kazanmıştır. Kullanıcılara karar destek sistemleri olarak yardımcı olan tavsiye istemlerinin kullanıldığı alana göre farklı deęerlendirme metrikleri eşlik etmektedir. Genel geçer tek bir deęerlendirme sistemi olmadığı gibi her uygulama için farklı deęerlendirme sistemleri daha gerçekçi çözümler sunabilir. Bununla birlikte deęerlendirme sistemlerinin de olumsuz özellikleri bulunmaktadır. Bu kısımlar da aşağıda anlatılacaktır.

#### 5.1.1. Test Ortamı Oluşturulması

Öncelikle, uygulamada olmayan test sürecinde yaşanabilecek tecrübelerden bahsetmek yerinde olur. Bir tavsiye sistemini canlı ortama sunmadan önce test ortamında denemek ve optimize ederek canlı ortama sunmak tavsiye sisteminin çoęu zaman prestijini arttırmaktadır. Bunun için öncelikle test verileri hazırlanmalıdır. Test verileri homojen olmamalı, yani farklı sınıflardan ve test deęerlendirmesi açısından anlamlı sonuçlar çıkartacak verilerden seçilmelidir. Bu verileri kullanacak kullanıcı kümesi de farklı özellikleri olan kişilerden oluşmalıdır. En azından gerçek ortamdaki kullanıcılara yakın özellikler oluşturulmalıdır. Kullanıcı davranışları modellenmelidir.

Test verileri üzerinden yapılan deęerlendirme yeterli olmadığına ya da canlı ortama çıkmadan önce farklı bir yöntem izlenecekse *kullanıcı çalışmaları* [65] yapılabilir. Bu aşamada kullanıcılara anket veya bazı sorular sorularak deęerlendirme yapması istenebilir. Bu anketler çalışma öncesinde, esnasında veya sonrasında uygulanabilir.

Ne var ki, bu çalışmalar da yeterli olmadığı takdirde veya modellenen ortamı bulamadığınız durumlarda testleri canlı ortamda da yapabilirsiniz. Bu durumda başarıyı ölçebilmeniz gerekmektedir. Bir sonraki kısımda başarıyı ölçecek deęerlendirme metrikleri anlatılmaktadır.

### 5.1.2. Değerlendirme Metrikleri

Tavsiye sisteminin başarısını test edebilecek birçok yöntem mevcuttur. Bu metotlardan bazılarını şu şekilde sıralayabiliriz:

- **Root Mean Squared Error (RMSE):** En popüler metriklerdendir. Tahmin edilen oy ile verilen oy arasındaki ilişkiye dayanarak hesaplama yapar. Gözlemlenen sonuçlar ile tahmin edilen sonuçlar arasındaki standart sapmaya dayanır. Normalize edilmiş olan RMSE'de hesaplamalarda kullanılabilir. Sistemimizde bu metrik kullanılmamıştır.

$$RMSE = \sqrt{\sum_{i=1}^k \frac{(x_{obs,i} - x_{model,i})^2}{2}}$$

$$N-RMSE = \frac{RMSE}{x_{obs,max} - x_{obs,min}}$$

$x_{obs,i}$  ; Gözlemlenen veri iken,  $x_{model,i}$  modellenen veridir.

- **Mean Absolute Error (MAE):** Popüler metriklerdendir.  $MAE = \frac{1}{n} \sum_{i=1}^n |e_i|$  RMSE, MAE ye göre büyük hatalara daha büyük ceza verir. MAE, formülden de anlaşılacağı gibi mutlak hataların ortalamasıdır. MAE ve RMSE'nin farklılaştırılmış ve normalize edilmiş versiyonları bulunmaktadır. Karşılaştırmaları ile ilgili çalışmalar mevcuttur [67].
- **Precision and Recall:**

	Tavsiye edilen	Tavsiye edilmeyen
Kullanılan	TP (True Positive)	TN (True Negative)
Kullanılmayan	FP (False Positive)	FN (False Negative)

Tablo 5.1 Precision-Recall

Precision ve Recall' da onaylanmayan veya tavsiye edilmeyen nesnelere için de ceza kesilir. Tablo 5.1 de tablo olarak gösterilmektedir. RMSE ve MAE' de oylanan nesnelere için sadece işlem yapılmaktadır. Şöyle ki elinizde çok fazla örneklem olduğunu ve bu örneklem arasında kullanıcılara ne kadar doğru

tavsiyelerde bulunduğunuzu ölçmeye çalıştığınızı farz edin. Eğer sadece tavsiye ettiğiniz nesnelere üzerinden değerlendirme yapar ve tavsiye etmediğiniz nesnelere duydukları ilgiyi ölçmezseniz başarılı bir sistem değerlendirmesi yapmış olamazsınız.

Bu tarz bir değerlendirmeyi Precision ve Recall ile yapabilirsiniz.

$$Precision = TP/(TP+FP)$$

$$Recall = TP/(TP+FN)$$

Precision ve recall sonuçlarını doğrudan değerlendirebileceğimiz gibi Precision ve Recall kıvrımını göz önüne alarak inceleme de yapabiliriz. ROC (Receiver Operating Characteristics) curve metodu [68] bu değerlendirme için idealdir. Precision ve Recall çoğu zaman istediğimiz sonucu vermeyecektir.

Precision ve Recall yöntemini ROC curve ile birlikte kullanmak daha uygun olabilir, fakat ikisini ayrı ayrı kullanmak ta uygun olabilir. Şöyle ki, bizim sistemimizde olduğu gibi kullanıcıya kaç haber tavsiye etmeniz gerektiği ile ilgili bir çalışma yapıyorsanız eşik değeri bulmak için bir grafik oluşturmanız gerekir. Bu durumda ROC curve kullanmak daha uygun olabilir. ROC curve ile Precision-Recall kullanacaksanız, F-Measure kullanmanız yerinde olacaktır. Accuracy ve F-Measure ayrı bir metrik olarak da uygulanabilir;

$$Accuracy : (TP+TN)/(TP+TN+FP+FN).$$

$$F-Measure : 2RP/R+P, P : Precision, R : Recall.$$

- **Ranking Metrikleri:** Eğer kullanıcıya belli sayıda tavsiyede bulunmayacaksanız yani kullanıcıya sayfalama yapıp sonraki nesnelere de sırayla veya kullanıcının isteğiyle gösterecekseniz bu durumda Normalized Distance-based Performance Measure (NDPM) kullanılabilir [69]. Bunlara ek olarak Spearman's  $\rho$  veya Kendall's  $\tau$  kullanılabilir.

### 5.1.3. Problem Sahaları

*Soğuk Başlangıç:* Bazı problem sahalarını metrikler içerisinde değerlendirmekle birlikte bir kısmını da burada ele alabiliriz. *Soğuk Başlangıç* bu problem daha önce çözülmediyse ve değerlendirme esnasında ele alınacaksa bu durumda eşik

değerler uygulayarak bu nesnelere değerlendirmenizi etkilemesini engellememesini sağlayabilirsiniz.

*Nesneye Güven:* Kullanıcılar aynı oy almış iki nesne arasında seçim yapacaksa daha çok güvenileni tercih etmek isteyebilir. *Nesneye güven*, sistemin nesneye verdiği güvendir.

*Kullanıcıya Güven:* *Kullanıcıya güven* de kullanıcıların nesneye verdiği güven oranıdır. Aynı oy almış iki nesneye oy veren kullanıcıların güvenilir olması aslında oylamanın güvenini de etkiler. Bu durumu da kullanıcılar dikkate alabilirler. Örneğin iki filme beş üzerinden dört oy veren kullanıcı gruplarından biri filmin sadece yarısını izleyerek oy vermişse ve ikinci grup ise tamamını izleyerek oy vermişse bu durumda iki film de dört oy almışken ikinci grubun güveni daha yüksek olacaktır.

*Çeşitlilik:* *Çeşitlilik* benzerliğin tersi olarak tanımlanır ve ele alınması gereken konulardandır. Örneğin bir kullanıcıya her zaman benzer haberler tavsiye ederseniz bir süre sonra sıkılma ve farklı haberler okuma ihtiyacı duyabilir. Bu durumda o kullanıcıya farklılık arz eden haber tavsiyesinde bulunmanız yerinde olacaktır.

*Faydacılık:* Faydacılık ta bazı durumlarda ele alınması gereken konulardan birisidir. Örneğin bir kullanıcıya film tavsiye ederken beş oy alanları ve sırasıyla dört ve üç oy alanları tavsiye ederken, diğer kullanıcıya sadece beş oy alanları tavsiye etmeniz gerekebilir. Bu durum ender olarak yaşansa da uyguladığınız sistem için göz önüne almanız gereken hususlardandır.

*İstikrar:* Sisteminize dışarıdan müdahale olmamalı ve sisteminiz istikrarlı olmalıdır. Kullanıcılara gerçekten sisteminizin tavsiye ettiği nesnelere tavsiye etmelisiniz. Bu aynı zamanda sistemin güvenliğini artırır.

*Gizlilik:* Kullanıcıya tavsiye edeceğiniz nesnelere gizli olan veriyi ifşa etmemeli veya kullanıcıyı mağdur etmemelidir.

*Ölçeklenebilirlik:* Veri setleri büyüdükçe tavsiye sisteminiz hesap yapmada güçlük yaşayabilir. Bu durumda kullanıcı kesinlikle mağdur olmamalıdır. Sisteminizin her durumda istikrarlı kalmasını sağlamalısınız.

## 5.2. Tavsiye Sistemi Deęerlendirmesi

Haber analizi sisteminin deęerlendirmesini yapmadan önce oluřturduęumuz veri setlerinden bahsetmek yerinde olur. Sistemimiz yaklaşık olarak üç ay boyunca test edilmiřtir. Kullanıcıların statik ve dinamik profillerine bakılarak haber sitelerinden her gün yaklaşık olarak çekilen bin beř yüz haberden elli tanesi tavsiye edilmiřtir. Kullanıcıların da her gün bu haberlere oy vermesi saęlanmıřtır.

Bu oylar; *ilgili deęil*, *az ilgili*, *yorum yok*, *ilgili* ve *çok ilgili* dir. Eęer kullanıcı *yorum yok*, *ilgili* veya *çok ilgili* durumunu seçerse tavsiyenin olumlu olduęunu deęerlendirdik. Kullanıcılardan aktif olarak yirmi gün boyunca yaklaşık yirmi bin oy topladık. Topladığımız oy neticesinde başarılı sonuçlar elde ettik. *Accuracy* 0.59 olarak ortaya çıktı.

	İlgili Deęil	Az İlgili	Yorum Yok	İlgili	Çok İlgili	Tavsiye edilen Haber Sayısı	Accuracy
<b>Eęitim</b>	210	267	380	193	65	1115	0.57
<b>Ekonomi</b>	343	442	613	435	145	1978	0.60
<b>Gündem</b>	638	815	1103	1006	500	4062	0.64
<b>Kültür-Sanat</b>	345	267	225	295	125	1257	0.51
<b>Politika</b>	575	703	585	897	215	2975	0.57
<b>Saęlık</b>	223	137	145	115	65	685	0.47
<b>Spor</b>	975	1370	1670	1572	820	6407	0.63
<b>Tekno-Bilim</b>	336	610	245	195	135	1521	0.38
<b>Toplam</b>	3645	4611	4966	4708	2070	20000	0.59

**Tablo 5.2 Tavsiye Deęerlendirme Sonuçları (20000 haber için)**

Tablo 5.2 incelendięinde ilginç sonuçlar gözükmetedir. Kullanıcıların çoęu spor, gündem ve politika haberlerini genellikle çok ilgili olarak iřaretlemiřlerdir. Bunun nedeni olarak spor ile ilgilenen kiřinin rakip takımları da takip ettięi, politika ile

ilgilenenlerin genel politikayı incelediği ve magazin kısmında ise taraf olmadığı için her haberin dikkat çekici olduğu söylenebilir.

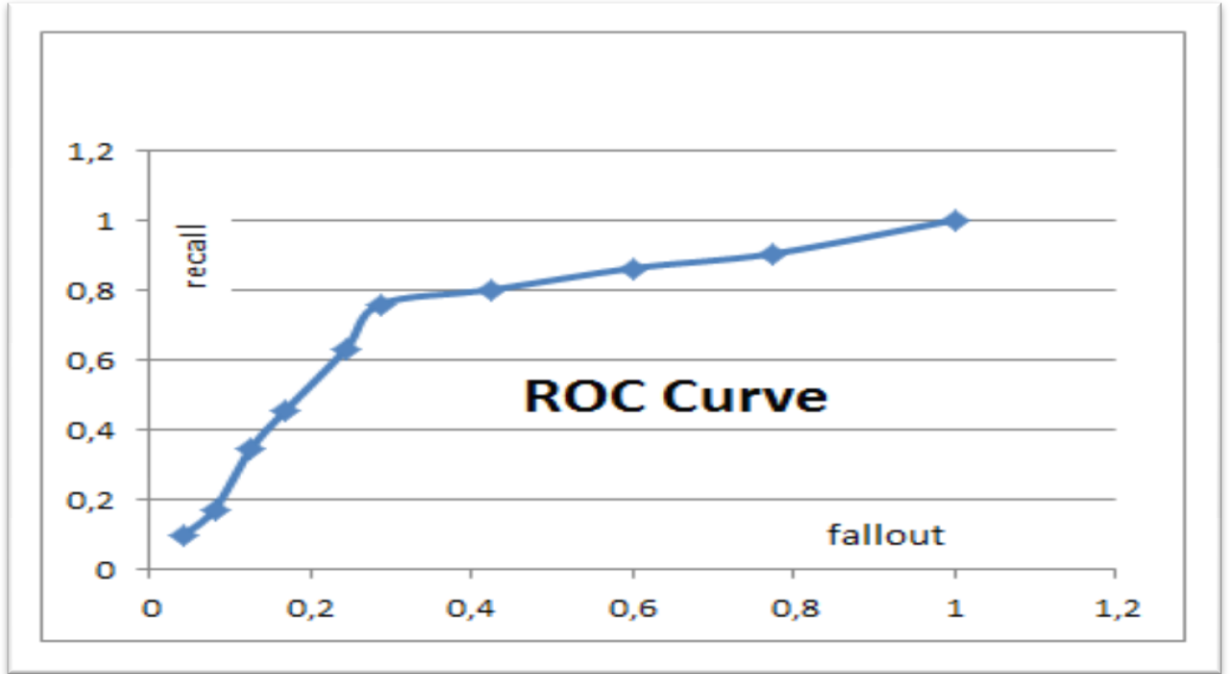
### **5.2.1. Zaman-bazlı tavsiye Sisteminde Optimum Haber Sayısının Tespiti**

Haber tavsiye sistemleri için tavsiye edilecek haber sayısını tespit etmek son derece zordur. Eğer tavsiye edilecek sayı kişinin okuyacağı ve ilgi duyduğu haber sayısından fazla ise önemli görülebilecek haberler gözden kaçabilir. Benzer şekilde tavsiye edilen sayı ilgi duyulandan az ise, kişi başka haberler okumak için başka kaynaklara yönelmek durumunda kalabilir. Optimum sayıyı tespit edebilmek için, kullanıcılara başlangıçta elli haber tavsiye ettik. Her haberin de kullanıcı için oluşan önemine göre ağırlık almasını sağladık. Sistemimizde mevcut olan geri besleme yöntemini kullanarak, tavsiye ettiğimiz haberlerin kullanıcılar tarafından oylanması neticesine göre değerlendirmesini yaptık. Değerlendirmeye başlamadan önce yirmi gün boyunca, yirmi kullanıcıya, elli haber tavsiyesinde bulduk. Tavsiyelerimizi kullanıcıların dinamik profillerine göre yaptık. Kullanıcılar haberleri oyladıktan sonra 50 haberi 5 sayfaya bölerek 10, 20, 30, 40, 50 haber için precision, recall ve F-Measure değerlerini oluşturduk. Bu değerleri oluştururken kullanıcıya 50 haber tavsiye ettiğimizde, ilk 10 haber arasında kaçını olumlu olarak değerlendirmiş, ilk 20 içinde kaçını olumlu olarak değerlendirmiş vb. şeklinde devam ederek sonuçları birleştirdik. Örnek bir değerlendirme verecek olursak, kullanıcı 20 haberin 10'unu olumlu, 30 haberin 15 sini olumlu, 40 haberin 17 sini olumlu ve toplamda 50 haberin 25'ini olumlu olarak değerlendirmiş olabilir. Amaç, F-Measure'ın en fazla olduğu değeri bulmaktır. Bunun için de tavsiye sayısı, kullanıcının okuyabileceği haberden fazla olmamalı, fakat kullanıcıyı da başka bir kaynağa sevk etmemelidir. Elde ettiğimiz sonuçlar Tablo 5.3'tedir. Bu değerlendirmeye göre 30 haberin ideal olduğu görülmektedir. Bu değer tespiti için de *Precision-Recall* ve *ROC* analizi kullanılmıştır.

Tavsiye Edilen Haber Sayısı	Precision	Recall	F-Measure	Accuracy
10	0.60	0.20	0.30	0.44
20	0.70	0.45	0.55	0.60
30	0.73	0.76	0.75	0.70
40	0.63	0.86	0.72	0.62
50	0.59	1.00	0.73	0.59

Tablo 5.3 Optimum Haber Sayısı Tespiti

Tablo 5.3 te elde ettiğimiz sonuçlara göre **F-Measure** değerlerini incelediğimizde 30 tavsiyenin yerinde olduğu sonucuna vardık. Sayılar arasında ara değerlerin ve



Şekil 5.1 Haber Sayısı İçin ROC Analizi

kırılma noktalarının da önemli olduğunu düşünerek ROC analizini de yaptık.

Şekil 5.1 deki ROC grafiği RECALL ve FALLOUT değerlerinden oluşmaktadır. Bu grafikte de görüleceği gibi recall değerinin 0.77-0.8 aralığında olduğu grafiğin eğiminde ani değişim görülmektedir. Bu ani eğim değişikliği de Tablo 5.3 üzerinde incelendiğinde yaklaşık olarak 30 değerine denk gelmektedir. Başka bir ifadeyle, 30 haberinin seçimi F-Measure değerine bakılarak yapılmasına rağmen ROC grafiğinde de bu değer anlamlı olduğu sonucu çıkarılmıştır.

### 5.3. İki Tavsiye Sisteminin Karşılaştırılması

Tavsiye Edilen Haber Sayısı	Precision	Recall	F-Measure	Accuracy
10	0.60	0.20	0.30	0.44
20	0.70	0.45	0.55	0.60
30	0.73	0.76	0.75	0.70
40	0.63	0.86	0.72	0.62
50	0.59	1.00	0.73	0.59

Tablo 5.4 Zaman Bazlı Tavsiye Sistemi Değerlendirmesi

Number of Recommended Articles	Precision	Recall	F-Measure	Accuracy
10	0.80	0.28	0.41	0.54
20	0.75	0.52	0.61	0.62
30	0.73	0.76	0.75	0.70
40	0.63	0.86	0.72	0.62
50	0.59	1.00	0.73	0.59

Tablo 5.5 Keyphrase Bazlı Tavsiye Sistemi Değerlendirmesi



Öncelikle Tablo 5.5'teki sonuçları nasıl oluşturduğumuzu belirtecek olursak; zaman-bazlı haber tavsiye sistemi neticesinde tavsiye ettiğimiz haberler, anahtar kelime bazlı tavsiye sistemi algoritması kullanarak tavsiye edildiğinde sıra değişecek ve yeni sıraya göre değerlendirme yapılacaktır. Bu durumda da Tablo 5.5 elde edilmiş olurdu. Örneğin zaman-bazlı tavsiye sisteminde 19'uncu sırada tavsiye edilecek haber anahtar kelime-bazlı sistemde 5'inci sırada tavsiye edilebilirdi. Bu durumda, eğer kullanıcı da bu haberi okumuş olsa, anahtar kelime-bazlı sistem için ilk 10 haber için olumlu bir sonuç oluşmuş olurdu.

Tablo 5.4 ve Tablo 5.5 deki sonuçlarını bu esaslar çerçevesinde karşılaştırarak olursak, iki sistem için de Accuracy ve F-Measure sonuçlarının 30 haber için en yüksek seviyeye ulaştığını söyleyebiliriz. Bu sonuç da bize iki sistem için de 30 haberin ideal olduğu sonucunu çıkarır. Bunun yanında precision değerlerinin ilk 30 haber için daha yüksek olduğunu söylemek yerinde olur. İki sistemde de F-Measure değeri 30 haberde en üst seviyeye çıkmaktadır.

## 6. SONUÇ VE GELECEK ÇALIŞMALAR

### 6.1. Sonuç

Sonuçlarımızı değerlendirmek ve anlamlı sonuçlar oluşturabilmek için sistemi uyguladığımız süreyi uzun tutmaya çalıştık. Kullanıcıların geri besleme de bulunmalarını sağladık. Farklı geri besleme metotları kullanılabilir. Adomavicious ve Kwon A+ ve F arasında değerlendirilmek üzere bir sınıflandırma kullanmışlardır [70]. Bizim için olumlu olan, yani kullanıcı için uygun tavsiye olarak değerlendirdiklerimiz kullanıcının yorum yok, ilgili ve çok ilgili olarak işaretledikleriydi. Bu çerçevede zaten sistemimizin başarısını ve tavsiye edilecek sayısı belirleyebildik. Elde ettiğimiz neticeleri şu şekilde özetleyebiliriz;

- Kullanıcılara haber tavsiyesinde bulunurken benzerlikler kadar farklılıkları da göz önüne almak gerekir. Kullanıcıya aynı haberi farklı kaynaklardan vermek yerine haberi sadece bir kaynaktan vererek farklı haberi de kullanıcıya göstermek gerekir.
- Haber tavsiyelerinde anahtar terimleri (keyphrase) de dikkate almak gerekir.
- Kaynakların güvenilirliğini belirleyerek nesne-güven ilişkisini kurmak yerinde olur. Bunun için kullanıcılardan geri dönüş olarak kaynağa güven değeri de

alınabilir. Ya da benzer haberler arasında farklılıklara bakarak doğru haber veren kaynaklar tespit edilebilir.

- Eğer işbirlikçi filtreleme kullanılacaksa kullanıcıların oyları dikkate alınacağından mutlaka kullanıcıların da *kullanıcı-güven* durumunu belirlemek gerekir.

## 6.2. Gelecek Çalışmalar

Haber tavsiye sistemleri genel hatlarıyla incelendiğinde emekleme döneminindedir. Bu aşamada karşılaşılabilecek sorunları tek tek ele almak yerinde olacaktır: Örneğin, soğuk başlangıç sorununu çözmek için KNN veya statik profil kullanılabilir. Bununla birlikte bu sorun önemli sorunlardan kabul edilmiş olup farklı çözüm yolları denenmiştir. Kullanıcıların profillerinin çeşitli kaynaklardan tespit edilerek (sosyal paylaşım sistemleri gibi) sistemleri bütünlük çalıştırmak kullanıcıdan profil doldurmasını istemekten daha başarılı olacaktır. Bu sayede, kullanıcılardan profillerini devamlı güncellemesi de beklenmeyecektir. Ayrıca bir sonraki aşamada haberler ile ilgili kullanıcılardan yorumlar yapılması beklenebilir ve bu yorumlara dayanarak farklı haber grupları tavsiye edilebilir.

Haber özetlemesi ve haberlerin ön izlemesinin kullanıcılara gösterilmesi de kullanıcıların haberi okumadan değerlendirme yapmasını sağlayacaktır. İyi oluşturulmuş ve anahtar kelimelerin vurgulandığı bir özet sistemi kullanıcıların haberleri daha iyi yorumlamasını ve algılamasını sağlayacaktır.

## KAYNAKLAR

- [1] Cleger-Tamayo, S., Fernández-Luna, J. M., Huete, J. F., Top-N news recommendations in digital newspapers, *Knowledge-Based Systems*, 27, 180-189, **2012**.
- [2] Schafer, J. B., Konstan, J. A., Riedl, J. T., Recommender Systems for the Web. In *Visualizing the Semantic Web* (pp. 102-123), Springer London, **2006**.
- [3] Li, Q., Kim, B. M., Clustering approach for hybrid recommender system, *Web Intelligence, 2003. WI 2003, Proceedings IEEE/WIC International Conference on*. IEEE, **2003**.
- [4] Asanov, D., Algorithms and methods in recommender systems, *Berlin Institute of Technology, Berlin, Germany*, **2011**.
- [5] Resnick, P., Varian, H. R., Recommender systems, *Communications of the ACM*, 40(3), 56-58, **1997**.
- [6] Lee, T. Q., Park, Y., Park, Y. T., A time-based approach to effective recommender systems using implicit feedback, *Expert systems with applications*, 34(4), 3055-3062, **2008**.
- [7] KG, S., Sadhasivam, G. S., A Personalized Online News Recommendation System, *International Journal of Computer Applications* 57 ,**2012**.
- [8] Pazzani, M. J., Billsus, D., Content-based recommendation systems, *The adaptive web* (pp. 325-341), Springer Berlin Heidelberg, **2007**.
- [9] Kazienko, P., Musiał, K., *Recommendation framework for online social networks*, Springer Berlin Heidelberg, **2006**.
- [10] Radev, D. R., Fan, W., Zhang, Z., Webinessence: A personalized web-based multi-document summarization and recommendation system, *Ann Arbor*, 1001, 48103, **2001**.
- [11] Su, X., Khoshgoftaar, T. M., A survey of collaborative filtering techniques, *Advances in artificial intelligence*, **2009**.
- [12] Ungar, L. H., Foster, D. P., Clustering methods for collaborative filtering, *AAAI workshop on recommendation systems* (Vol. 1), **1998**.
- [13] Chen, C. C., Wan, Y. H., Chung, M. C., Sun, Y. C., An effective recommendation method for cold start new users using trust and distrust networks, *Information Sciences*, 224, 19-36., **2013**.
- [14] Balabanović, M., Shoham, Y., Fab: content-based, collaborative recommendation, *Communications of the ACM*, 40(3), 66-72., **1997**.

- [15] Adomavicius, G., Tuzhilin, A., Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *Knowledge and Data Engineering, IEEE Transactions on*, 17(6), 734-749., **2005**.
- [16] Burke, R., Hybrid recommender systems: Survey and experiments, *User modeling and user-adapted interaction*, 12(4), 331-370, **2002**.
- [17] Belkin, N. J., Croft, W. B., Information filtering and information retrieval: two sides of the same coin?, *Communications of the ACM*, 35(12), 29-38, **1992**.
- [18] Konstan, J. A., Riedl, J., Borchers, A., Herlocker, J. L., Recommender systems: A groupLens perspective, *Recommender Systems: Papers from the 1998 Workshop (AAAI Technical Report WS-98-08)* (pp. 60-64), **1998**.
- [19] Billsus, D., Pazzani, M. J., User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10(2-3), 147-180, **2000**.
- [20] Montes-García, A., Álvarez-Rodríguez, J. M., Labra-Gayo, J. E., Martínez-Merino, M., Towards a journalist-based news recommendation system: The Wesomender approach, *Expert Systems with Applications*, 40(17), 6735-6741, **2013**.
- [21] Rich, E., User modeling via stereotypes\*, *Cognitive science*, 3(4), 329-354., **1979**.
- [22] Schafer, J. B., Konstan, J. A., Riedl, J., E-commerce recommendation applications, *Applications of Data Mining to Electronic Commerce* (pp. 115-153), Springer US, **2001**.
- [23] Schafer, J. B., Konstan, J., Riedl, J., Recommender systems in e-commerce, *Proceedings of the 1st ACM conference on Electronic commerce* (pp. 158-166), ACM., **1999**.
- [24] Liu, B., Dai, Y., Li, X., Lee, W. S., Yu, P. S., Building text classifiers using positive and unlabeled examples, In *Data Mining, 2003, ICDM 2003, Third IEEE International Conference on* (pp. 179-186), IEEE, **2003**.
- [25] Salton, G., Lesk, M. E., Computer evaluation of indexing and text processing, *Journal of the ACM (JACM)*, 15(1), 8-36, **1968**.
- [26] Turney, P. D., Pantel, P., From frequency to meaning: Vector space models of semantics, *Journal of artificial intelligence research*, 37(1), 141-188, **2010**.
- [27] Singh, A., Yadav, A., Rana, A., K-means with Three different Distance Metrics, *International Journal of Computer Applications*, 67(10), 13-17., **2013**.
- [28] Pan, X., Cheng, J., Xia, Y., Zhang, X., Wang, H., Which Feature is Better? TF\* IDF Feature or Topic Feature in Text Clustering, *Multimedia Information*

- Networking and Security (MINES), 2012 Fourth International Conference on* (pp. 425-428). IEEE, **2012**.
- [29] Huang, A., Similarity measures for text document clustering, *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand* (pp. 49-56), **2008**.
- [30] Liu, F., Xiong, L., Survey on text clustering algorithm, *Software Engineering and Service Science (ICSESS), 2011 IEEE 2nd International Conference on* (pp. 901-904), IEEE, **2011**.
- [31] Yonghong, Y., Wenyang, B., Text clustering based on term weights automatic partition, *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on* (Vol. 3, pp. 373-377). IEEE, **2010**.
- [32] Miyahara, K., Pazzani, M. J., Improvement of Collaborative Filtering with the Simple Bayesian Classifier 1, **2002**.
- [33] Miyahara, K., Pazzani, M. J., Collaborative filtering with the simple Bayesian classifier, *PRICAI 2000 Topics in Artificial Intelligence* (pp. 679-689), Springer Berlin Heidelberg, **2000**.
- [34] Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., Scholkopf, B., Support vector machines, *Intelligent Systems and their Applications, IEEE*, 13(4), 18-28, **1998**.
- [35] Sutton, O., Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction, *University lectures, University of Leicester*, **2012**.
- [36] Madsen, R. E., Hansen, L. K., Winther, O., Singular value decomposition and principal component analysis, *Neural Networks*, 1, 1-5, **2004**.
- [37] Zheng, L., Li, L., Hong, W., Li, T., PENETRATE: Personalized news recommendation using ensemble hierarchical clustering *Expert Systems with Applications*, 40(6), 2127-2136, **2013**.
- [38] Xue, G. R., Xing, D., Yang, Q., Yu, Y., Deep classification in large-scale text hierarchies, *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*(pp. 619-626), ACM, **2008**.
- [39] Liu, J., Dolan, P., Pedersen, E. R., Personalized news recommendation based on click behavior, *Proceedings of the 15th international conference on Intelligent user interfaces* (pp. 31-40), ACM, **2010**.
- [40] Tan, A. H., Teo, C., Learning user profiles for personalized information dissemination, *Neural Networks Proceedings, 1998, IEEE World Congress on Computational Intelligence, The 1998 IEEE International Joint Conference on* (Vol. 1, pp. 183-188), IEEE, **1998**.

- [41] Li, L., Wang, D., Li, T., Knox, D., Padmanabhan, B., Scene: a scalable two-stage personalized news recommendation system, *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 125-134), ACM, **2011**.
- [42] Billsus, D., Pazzani, M. J., *A hybrid user model for news story classification* (pp. 99-108), Springer Vienna, **1999**.
- [43] Chandramouli, B., Levandoski, J. J., Eldawy, A., Mokbel, M. F., StreamRec: a real-time recommender system, *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data* (pp. 1243-1246), ACM, **2011**.
- [44] Das, A. S., Datar, M., Garg, A., Rajaram, S., Google news personalization: scalable online collaborative filtering, *Proceedings of the 16th international conference on World Wide Web* (pp. 271-280), ACM, **2007**.
- [45] Liang, T. P., Lai, H. J., Discovering user interests from web browsing behavior: An application to internet news services, *System Sciences, 2002. HICSS, Proceedings of the 35th Annual Hawaii International Conference on* (pp. 2718-2727), IEEE, **2002**.
- [46] Gong, S., A collaborative filtering recommendation algorithm based on user clustering and item clustering. *Journal of Software*, 5(7), 745-752, **2010**.
- [47] Xia, Z., Xu, S., Liu, N., Zhao, Z., Hot news recommendation system from heterogeneous websites based on bayesian model, *The Scientific World Journal*, **2014**.
- [48] Jiang, S., Hong, W., A vertical news recommendation system: CCNS—An example from Chinese campus news reading system, *Computer Science & Education (ICCSE), 2014 9th International Conference on*(pp. 1105-1114), IEEE, **2014**.
- [49] Li, L., Chu, W., Langford, J., Schapire, R. E., A contextual-bandit approach to personalized news article recommendation, *Proceedings of the 19th international conference on World wide web* (pp. 661-670), ACM, **2010**.
- [50] Wan, X., Xiao, J., Single Document Keyphrase Extraction Using Neighborhood Knowledge, *AAAI* (Vol. 8, pp. 855-860), **2008**.
- [51] Cho, Y. H., Kim, J. K., Kim, S. H., A personalized recommender system based on web usage mining and decision tree induction, *Expert systems with Applications*, 23(3), 329-342, **2002**.
- [52] Iacob, I. E., Dekhtyar, A., Zhao, W., Xpath extension for querying concurrent xml markup, *University of Kentucky Department of Computer Science*, **2004**.
- [53] Akin, A. A., Akin, M. D., Zemberek, an open source nlp framework for turkic languages, *Structure*, 10, **2007**.

- [54] Kutlu, M., Cığır, C., Cicekli, I., Generic text summarization for Turkish, *The Computer Journal*, bxp124, **2010**.
- [55] Hovy, E., Lin, C. Y., Automated text summarization and the SUMMARIST system, *Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998* (pp. 197-214), Association for Computational Linguistics, **1998**.
- [56] Ozsoy, M. G., Cicekli, I., Alpaslan, F. N., Text summarization of turkish texts using latent semantic analysis, *Proceedings of the 23rd international conference on computational linguistics* (pp. 869-876), Association for Computational Linguistics, **2010**.
- [57] Hahn, U., Mani, I., The challenges of automatic summarization, *Computer*, 33(11), 29-36., **2000**.
- [58] Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A., User profiles for personalized information access, *The adaptive web* (pp. 54-89), Springer Berlin Heidelberg, **2007**.
- [59] Sugiyama, K., Hatano, K., Yoshikawa, M., Adaptive web search based on user profile constructed without any effort from users, *Proceedings of the 13th international conference on World Wide Web* (pp. 675-684), ACM, **2004**.
- [60] Kim, H. R., Chan, P. K., Learning implicit user interest hierarchy for context in personalization, *Proceedings of the 8th international conference on Intelligent user interfaces* (pp. 101-108), ACM, **2003**.
- [61] Lv, Y., Moon, T., Kolari, P., Zheng, Z., Wang, X., Chang, Y., Learning to model relatedness for news recommendation, *Proceedings of the 20th international conference on World wide web* (pp. 57-66), ACM, **2011**.
- [62] Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C., Nevill-Manning, C. G., KEA: Practical automatic keyphrase extraction, *Proceedings of the fourth ACM conference on Digital libraries* (pp. 254-255), ACM, **1999**.
- [63] Pala, N., Çiçekli, I., Turkish keyphrase extraction using KEA, *Computer and information sciences, 2007, iscis 2007, 22nd international symposium on* (pp. 1-5), IEEE, **2007**.
- [64] Ozdemir, B., Cicekli, I., Turkish keyphrase extraction using multi-criterion ranking, *Computer and Information Sciences, 2009, ISICIS 2009, 24th International Symposium on* (pp. 269-273), IEEE, **2009**.
- [65] Shani, G., Gunawardana, A., Evaluating recommendation systems, *Recommender systems handbook* (pp. 257-297) Springer US, **2011**.
- [66] Sarkar, K., A hybrid approach to extract keyphrases from medical documents, *arXiv preprint arXiv:1303.1441.*, **2013**.

- [67] Chai, T., Draxler, R. R., Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature, *Geoscientific Model Development*, 7(3), 1247-1250, **2014**.
- [68] Davis, J., Goadrich, M., The relationship between Precision-Recall and ROC curves, *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240), ACM, **2006**.
- [69] Yao, Y. Y., Measuring retrieval effectiveness based on user preference of documents, *JASIS*, 46(2), 133-145., **1995**.
- [70] Adomavicius, G., Kwon, Y., New recommendation techniques for multicriteria rating systems, *Intelligent Systems, IEEE*, 22(3), 48-55., **2007**.



## EKLER

### EK-A StopList

a	b	c	e	j	o	ö	ş	u
acaba	bağlı	ç	eden	k	on	öbürü	şayet	ü
altı	bana	çoğu	elbette	kaç	ona	ön	şekilde	üç
ama	bazen	çoğuna	en	karşı	ondan	önce	şey	üzere
ancak	bazı	çoğunu	f	kadar	onlar	öte	şeyden	üst
artık	bazıları	çok	fakat	kendi	onlara	ötürü	şeye	üzerine
asla	bazısı	çünkü	falan	kendine	onlardan	öyle	şeyi	v
aslında	belki	d	felan	kendini	onların	p	şeyler	var
az	ben	da	filan	kez	onların	r	şimdi	ve
ardından	beni	daha	g	ki	onu	rağmen	şöyle	veya
artık	benim	de	gene	kim	onun	s	şu	veyahut
aynı	beş	değil	gibi	kime	orada	sana	şuna	y
ı	bile	demek	ğ	kimi	oysa	sekiz	şunda	ya
i	bir	devam	h	kimin	oysaki	sen	şundan	ya da
için	birçoğu	diğer	hâlâ	kimisi	n	senden	şunlar	yanı
içinde	birçok	diğeri	hangi	l	nasıl	seni	şunu	yandan
ilgili	birçokları	diğerleri	hangisi	m	ne	senin	şunun	yani
ilk	biri	diye	hani	madem	ne kadar	sıra	t	yedi
iki	birisi	dokuz	hatta	mı	ne zaman	siz	tabi	yerine
ile	birkaç	dolayı	hem	mı	neden	sizden	tamam	yine
ise	birkaçı	dört	henüz	mi	nedir	size	tüm	yoksa
işte	birşey		hep	mu	nerde	sizi	tümü	z
	birşeyi		hepsi	mu	nerede	sizin		zaten
	biz		hepsine	mü	nereden	son		zira
	bize		hepsini	mü	nereye	sonra		
	bizi		her	mümkün	nesi			
	bizim		her biri		neyse			
	boş		herkes		niçin			
	böyle		herkese		niye			
	böylece		herkesi					
	bu		hiç					

	buna		hiç kimse					
	bunda		hiçbiri					
	bundan		hiçbirine					
	bunu		hiçbirini					
	bunun							
	burada							
	bütün							

## ÖZGEÇMİŞ

### Kimlik Bilgileri

Adı Soyadı : Servet TAŞCI  
Doğum Yeri : İstanbul  
Medeni Hali : Evli  
E-Posta : servettasci@cs.hacettepe.edu.tr  
Adresi :

### Eğitim

Lisans : Kara Harp Okulu  
Yüksek Lisans :  
Doktora :

### Yabancı Dil ve Düzeyi

İngilizce: KPDS, 83,75

### İş Deneyimi

2009 – 2015 : Kara Kuvvetleri Komutanlığı, Web Tabanlı Yazılım Geliştirme  
Uzmanı

### Deneyim alanları

### Tezden Üretilmiş Projeler ve Bütçesi

### Tezden Üretilmiş Yayınlar

Tasci, S., Cicekli, I., A Media Tracking and News Recommendation System, *KDIR 2014, 6th International Conference on Knowledge Discovery and Information Retrieval*, 2015.

### Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar