Türkçe Paylaşım Yapan Kullanıcılar İçin Sosyal Ağ Tabanlı Analiz ve Tavsiye Sistemi

[*1] Onur SEVLİ, [2] Ecir Uğur KÜÇÜKSİLLE
[1] Mehmet Akif Ersoy Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Ana Bilim Dalı,
[2] Süleyman Demirel Üniversitesi, Bilgisayar Mühendisliği Ana Bilim Dalı

Geliş Tarihi: 2016-09-05 Kabul Tarihi: 2016-10-07

Öz

Kişiler arası bilgi paylaşımında yaygın olarak kullanılan sosyal ağlar, hızla artan kullanıcı sayıları ile iletişimde yeni bir çığır açmıştır. Kullanıcı sayılarındaki artışa bağlı olarak sosyal ağlarda üretilen içeriğin devasa boyutlara ulaşması, gerekli bilgilerin ilgili hedef kitleye ulaşmasını zorlaştırmaktadır. Bu noktada içerik analiz ve filtreleme sistemlerine ihtiyaç doğmaktadır. Bu çalışmada Twitter üzerinde kullanıcıların ilgi alanlarını dinamik olarak analiz eden ve bu doğrultuda takip edilebilecek diğer kullanıcıları tavsiye olarak sunan bir sistem geliştirilmiştir. Sistem tasarımında doğal dil işleme, büyük veri analizi, şartlı olasılık teoremi ve tabu arama yaklaşımlarından yararlanılmıştır. Geliştirilen yazılım Twitter'ı aktif olarak kullanan bir grup ile test edilmiş ve alınan geri bildirimler doğrultusunda sistemin %86 başarılı olduğu ortaya konmuştur.

Anahtar kelimeler: Sosyal ağ analizi, içerik filtreleme, tavsiye sistemi

Social Network Based Analysis and Recommendation System for Users Who Share Turkish Content

[*1] Onur SEVLİ, [2] Ecir Uğur KÜÇÜKSİLLE
[1] Mehmet Akif Ersoy University, Department of Computer Education & Instructional Technologies
[2] Süleyman Demirel Üniversitesi, Department of Computer Engineering

Received Date: 2016-09-05 Accepted Date: 2016-10-07

Abstract

Social networks which are widely used for interpersonal information sharing has opened a new era in communication with rapidly increasing number of users. Depending on the increase in the number of users the generated content has reached huge dimensions, so it is difficult the required information to reach its target audience. At this point a need arises for content analyse and filtering systems. In this study a system has been developed which analyses users' interests dynamically on Twitter and in this direction advices appropriate users that can be followed. In system design natural language processing, big data analysis, conditional probability theorem and tabu search approach were used. The developed software was tested with a group that uses Twitter actively and according to received feedbacks it has been shown the system is %86 successful.

Keywords: Social network analyse, content filtering, recommendation system

1. Giriş

İnsanlar sürekli diğer insanların düşüncelerini merak etmiş ve bu düşüncelerden esinlenerek hayatlarına yön vermişlerdir. İnternetin ortaya çıkması bireylerin kendilerini ifade etmeleri ve diğer bireylerden haberdar olabilmeleri için yeni ortamlar oluşturmuştur. Sosyal ağlar, fikir paylaşımında kullanılan günümüzün en popüler, çevrimiçi ortamlarıdır.

İnsanlar, genel hayat görüşlerini, içinde bulundukları duruma ilişkin izlenimlerini, ilgi alanlarını, yönelimlerini kısaca diğer bireylerin bilmesini istedikleri her şeyi sosyal ağ paylaşımlarına yansıtmaktadır. Sosyal ağ kullanımı ve sosyal ağlar üzerinden fikir paylaşımı her geçen gün hızla artmaktadır. Sosyal ağlarda bir kullanıcı tarafından üretilen fikirler diğer kullanıcıların da ilgisini çekebilir. Sosyal ağ kullanıcıları ilgi duydukları

*Sorumlu Yazar: Adres: Mehmet Akif Ersoy Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Ana Bilim Dalı, İstiklal Yerleşkesi, 15030 BURDUR, E-mail: onursevli@mehmetakif.edu.tr, Tel:+902482134130

Doi:10.21541/apjes.69981

diğer kullanıcıları takip altına alarak, paylaşımlarından haberdar olabilmektedir.
Sosyal ağlar üzerinde, gerçek hayattakine benzer şekilde, ortak görüşleri paylaşan kullanıcılar birbirleri ile yakın ilişkiler kurmaktadır. Sosyal ağ kullanıcılarının birbirlerini takip etmesinin temelinde de ortak ilgi alanları yer almaktadır. Hızla artan kullanıcı sayıları ve sürekli büyüyen paylaşım miktarı ile sosyal ağlar üzerinde kullanıcıların, kendilerine hitap eden içeriklere ulaşmaları zorlaşmaktadır. Sosyal ağlar üzerinde her dakikada yaklaşık 2 terabayt boyutunda veri üretilmektedir [1] ve "bilgi çöplüğü" olarak tabir edilen bu veri yığını içerisinde, pek çok yararlı içerik, hedefine ulaşamadan yok olmaktadır.

Sürekli büyüyüp, çeşitlenen veri yığınları içerisinde anlamlı bilgilerin elde edilerek doğru hedeflere sunulması bir ihtiyaç haline gelmiştir. Ancak her kullanıcının ilgi alanı farklıdır ve bir kullanıcı için anlamlı olan bilgi diğer bir kullanıcı için gereksiz olabilir. Bu nedenle öncelikli olarak, kullanıcı ilgi alanlarının analiz edilmesi ve ilgi alanları doğrultusunda kişiye özel içeriklerin filtrelenmesi gerekmektedir. Bu noktada sosyal ağ analiz ve tavsiye sistemlerine ihtiyaç duyulmaktadır. Tavsiye sistemleri büyük bilgi yığınları içerisinden kullanıcıların ilgilerini çekebilecek bilgileri bulmalarına yardımcı olur [2].

Sosyal ağların kendi bünyelerinde belirli analiz ve tavsiye mekanizmaları yer almaktadır. Ancak bu mekanizmalar kullanıcıların ilgi alanındaki değişimleri ele almada başarısız olmaları ve çoğunlukla dış etkenleri temel alarak sonuç üretmeleri nedeniyle yetersiz kalmaktadır. Bu eksikliği ortadan kaldırmak için üçüncü parti, destek yazılımlara ihtiyaç doğmaktadır.

Bu çalışmada mevcut sosyal ağ analiz ve tavsiye sistemlerindeki eksikliklerden yola çıkılarak, kullanıcı ilgi alanlarını dinamik bir şekilde analiz edip bu doğrultuda kullanıcı ilgi alanına en uygun, takip edilebilecek diğer kullanıcıları tavsiye olarak sunan bir sistem geliştirilmiştir.

Hedef platform olarak, son dönemde en hızlı büyüme sergileyen Twitter tercih edilmiştir. Sistemin geliştirilmesinde Türkçe doğal dil işleme, büyük veri analiz teknikleri ve şartlı olasılık teoremi ve tabu arama yaklaşımı kullanılmıştır. Geliştirilen sistem, "Tweetlyzer" isimli bir web uygulaması olarak, tüm internet kullanıcılarının erişim ve kullanımına açılmıştır. Sistemin testi Twitter'ı aktif kullanan yüz kişilik bir grup ile gerçekleştirilmiştir. Bu kullanıcılardan alınan geri bildirimler doğrultusunda sistemin %86 oranında başarılı olduğu sonucuna varılmıştır.

2. Literatür Özeti

Son zamanlarda web ortamı daha sosyal ve gerçek zamanlı hale gelmiştir [3]. Özellikle sosyal ağlar bilgi paylaşımında popülerliği hızla artan ortamlardır [4]. İnsanlar sosyal ağlar vasıtasıyla çeşitli konularda bilgiler paylaşmakta, birbirlerine sorular sorup yardımlaşabilmektedirler. Sosyal ağ paylaşımları bireyler için birer yol gösterici olabilmektedir [5]. Lee ve Chien çalışmalarında, sosyal ağlarda sıkça bahsedilen konuların, toplumsal fikir dinamikleri hakkında ön bilgi verebileceğini ifade etmişlerdir [6]. Bu anlamda sosyal ağlar üzerinde paylaşılan fikirlerin etkisi, bireyden ilham alarak geniş bir kitleye yayılmaktadır.

Sosyal ağlar, hızla artan kullanıcı sayıları ve veri miktarlarıyla, her an, muazzam boyutta ve karmaşık bilgi yığınları üretmektedir. Kern vd., bir grup Facebook kullanıcısına ait veri kümesini inceleyerek, yaşlara göre kullanımını değişen kelime ve konular hakkında bir çıkarım yapmışlardır [7]. Zhao vd. çalışmalarında, restoran ve tüketici elektroniği alanındaki belirli bir ürün grubu üzerinde, tüketicilerin sosyal ağlarda paylaştıkları görüşleri içeren büyük veri kümesini inceleyerek, alışveriş esnasında tüketicilerin doğru karar vermesine yardımcı olacak sonuçların üretilmesini hedeflemişlerdir [8].

Compton vd., Twitter üzerinden halka açık paylaşımların belli bir kısmını ele alarak, çeşitli anahtar kelimeler ile yaptıkları filtrelemeler sonucunda Latin Amerika'da meydana gelebilecek olası toplumsal olaylar konusunda çıkarım yapmaya çalışmışlardır [9]. Bruns vd., 2011 yılı "Arap Baharı"nda Facebook, Twitter gibi sosyal ağların toplum üzerinde etkili bir role sahip olduğu görüşüyle, Ocak – Kasım 2011 döneminde Mısır'da yaşanan devrim ve Libya'daki sivil savaş sürecinde Twitter üzerindeki trafiği izlemiş, paylaşımlar üzerinde farklı grupların görüş ve rollerini analiz etmeye çalışmışlardır [10].

Sosyal platformlar üzerindeki hızla artan içerik miktarı kullanıcıları, çoğu zaman, büyük miktarda karmaşık veri yığınlarıyla karşı karşıya getirmektedir [11]. Son dönemde yapılan pek çok çalışma bu büyük veri yığını içerisinden kullanıcıya hitap eden içerikleri filtrelemeye odaklıdır [12]. İçerik filtreleme işlemi ise kullanıcıların sosyal ağ üzerindeki aktivitelerine bağlı olarak tespit edilen yönelimlere göre gerçekleştirilmelidir [13]. Yang ve Rim geliştirmiş oldukları topik analiz modeli ile Twitter paylaşımları içerisindeki ilgi çeken konuları belirlemeye çalışmışlardır [14]. Kang vd. yaptıkları

çalışmada, gündemdeki olaylar ve olaylar arası bağlantıları belirlemeye çalışmışlardır [15].

Twitter üzerinde paylaşılan bir mesajın başka bir kullanıcı tarafından yeniden yayılmasına "retweet" adı verilmektedir. Uysal ve Croft, retweet sayısının, bir mesajın popülerlik ve etkinliği yansıttığını belirterek, yeni paylaşımların ilerideki retweet durumlarını tahmin eden bir sistem önermişlerdir [16]. Yu bir dizi araç kullanarak tweet'leri bağımsız olarak puanlamış daha sonra bir karşılaştırma algoritması ile gruplamıştır [17]. Sun vd. ise geliştirdikleri puanlama sistemi ile tweet'leri önem düzeyine göre gruplamışlardır [18].

Sosyal ağlarda pek çok kullanıcı takip edilebilecek diğer kullanıcı ve konulardan habersiz kalmaktadır [19] bu nedenle sosyal ağ tavsiye mekanizmaları ön plana çıkmaktadır. Jamil vd. Twitter kullanıcısının profilinden alınan konum bilgisi ile kullanıcıya yakın coğrafyadaki benzer kullanıcıları tavsiye eden bir uygulama geliştirmişlerdir [20] Islam vd., Twitter kullanıcılarının geçmişteki arkadaşlıkları üzerinden elde ettikleri veriler doğrultusunda tavsiyelerde bulunan bir strateji ortaya koymuşlardır [21]. Etiketler (hashtag) sosyal ağlarda birbiriyle ilişkili mesajlara erişmenin önemli yollarından biridir. Lee vd. tweet, retweet ve etiketler üzerinden elde ettikleri anahtar sözcüklerle Twitter kullanıcılarına haber tavsiyeleri sunan bir servis geliştirmişlerdir [22]. Lu ve Lee etiket tavsiyeleri üreten bir model geliştirmişlerdir [23]. Jonnalagedda ve Gauch ise Twitter'da sık kullanılan kelimelere göre popüler haberleri belirleyip bunlar içerisinden kullanıcının kendi tanımladığı profil verileri ile uyumlu olanları tavsiye olarak sunan hibrit bir uygulama tasarlamışlardır [24].

3. Sosyal Ağlar, Analiz ve Tavsiye Sistemleri

Küresel ölçekte bilgi akışının en gözde bileşeni olan sosyal ağlar, zaman ve mekân kavramlarını bir kez daha yapı bozumuna uğratarak bireysel ve toplumsal hayatın yeniden şekillenmesini sağlamıştır [25]. Sosyal ağlar, yeni nesil web teknolojilerinin ortaya çıkması ile birlikte tek taraflı bilgi paylaşımından, karşılıklı ve eş zamanlı bilgi paylaşımına geçişi sağlayan, insanların birbiri ile içerik ve bilgi paylaşmalarına imkân sunan uygulamaların geneline verilen addır. Diğer bir ifadeyle, yüksek erişilebilir ve ölçeklenebilir iletişim araçları üzerinden oluşturulan sosyal etkileşim ve paylaşım ortamlarına sosyal ağ denir [26].

Sosyal ağlar bilgi üretim ve yayımı konusunda, geleneksel medyaya göre pek çok üstün özelliğe sahiptir. En önemli fark sosyal ağların geleneksel medya araçlarından çok daha hızlı olmasıdır. Geleneksel medyada yayımlanan içerikler durağanken, sosyal ağlar canlı ve dinamik bir ortamdır. Sosyal ağlar, geleneksel medyadan daha geniş bir erişim ve kullanıma sahiptir. Geleneksel medyada bilgi yaymak için bir altyapı oluşturmak gereklidir ve bu altyapı belirli kuruluşların elindedir. Oysa sosyal ağlarda herhangi bir altyapı yatırımı yapmaksızın isteyen her birey bilgi yayabilir. Sosyal ağlar kullanıcı odaklıdır ve özgürdür, hiçbir sınırlama olmaksızın herkesin söz hakkı vardır [27]. Bu anlamda "kullanıcının ürettiği içerik" ön plandadır.

Sosyal ağ platformları, kullanıcıların sosyal çevresini genişletme, birbiriyle iletişim kurma, bilgi ve deneyimlerini paylaşma, boş zamanlarını değerlendirme gibi farklı amaçlarla kullandıkları yazılımlardır [28]. Sosyal ağlar ortak ilgi alanlarına sahip kullanıcıları bir araya getirmekte, yeni bir tanışma ortamı oluşturmaktadır. Son dönemde mobil cihaz kullanımının yaygınlaşması ile birlikte sosyal ağ platformlarının kullanımı da artmıştır. Bu platformlar üzerinde kullanıcılar, zaman ve mekândan bağımsız olarak, o an içinde bulundukları durumla ilişkili güncellemeler paylaşabilmektedir. Günümüzde çok sayıda sosyal ağ platformu bulunmaktadır. Facebook, Flickr, Foursquare, Instagram, LinkedIn, Twitter yaygın olarak kullanılan platformlardır. Bu platformlar içerisinde, kendine has yapısı ve terminolojisi bulunan Twitter, son dönemdeki büyüme ivmesi ile dikkat çekmektedir.

3.1. Twitter

Twitter 2006 yılında Jack Dorsey tarafından, Amerika Birleşik Devletleri'nde geliştirilen bir sosyal ağ sitesidir. Kullanıcılarına 140 karakterden oluşan "tweet" adı verilen mesajlar yazma imkânı sağlayan Twitter, bu özelliğinden dolayı internet dünyasının "kısa mesaj servisi" olarak anılmaktadır. Twitter kullanıcıları kendi ilgi alanlarına yakın buldukları diğer kullanıcıları takip listelerine ekleyebilmekte, bu sayede takip ettikleri kullanıcıların tweet'lerini kendilerine sunulan zaman tünelinde izleyebilmektedirler. Kullanıcılar Twitter'ı, kendine ait bir web sayfası üzerinden kullanabilmektedir. Bunun yanında akıllı telefonlar için geliştirilmiş Twitter uygulamaları da mevcuttur. Mobil kullanımın da artması ile birlikte, Twitter kullanım oranı da gün geçtikçe artmıştır. Twitter son dönemde büyüme ivmesi en yüksek olan sosyal paylaşım platformlarındandır. 2010 yılının ilk çeyreğinde 30 milyon civarında olan küresel aktif kullanıcı sayısı 2014 yılının son çeyreğinde 288 milyon, 2015 yılının son çeyreğinde ise 316 milyona ulaşmıştır [29].

3.1.1. Twitter terminolojisi

Twitter platformu kendine ait bir terminolojiye sahiptir. En yaygın kullanılan terimleri şunlardır:

- **Tweet:** 140 karakterden oluşan metin tabanlı mesajlardır. Tweet'ler içerisinde fotoğraf, videolara ilişkin URL bilgileri, web sayfası adresleri, atıflar gibi yine metin formunda özel veriler de yer alabilmektedir.
- **Retweet:** Bir kullanıcının, başka bir kullanıcının yazmış olduğu tweet'i kendi hesabında yeniden paylaşmasıdır.
- **Zaman tüneli (Timeline):** Herhangi bir Twitter kullanıcısının, yapılan paylaşımları görüntüleyebildiği sayfadır.
- **Etiket (Hashtag):** Atılan tweet'leri belirli başlıklar altında toplamak için kullanılan, "#" sembolü ile işaretlenmiş kelimelerdir.
- **Ekran adı:** Twitter platformunda yer alan her bir kullanıcıya ait eşsiz addır. Başına @ işareti konularak da ifade edilir.
- **Atıf (Mention):** Bir tweet içerisinde bir başka kullanıcıyı referans göstermek için kullanıcı adının @ sembolü ile birlikte konulması yoluyla metin içerisinde yapılan atıftır.
- **API:** Twitter platformunun sunmuş olduğu verilere üçüncü parti uygulamaların yazılımsal olarak erişimine imkân sağlayan uygulama programlama ara yüzleridir.

3.2. Sosyal Ağ Analizi

Sosyal ağlar, barındırdıkları sosyal aktörlerin fikir ve ilişkilerine dair bilgiler içerir. Bu yapı ve ilişkilerin analizi sonucu; kullanıcılar arası etkileşimler, eğilimler gibi bilgilere ulaşılarak ağdaki ilişkiler hakkında çeşitli yorum ve tahminlere varılabilir. Sosyal ağ analizi bireyler, nesneler ya da kurumlar arasındaki ilişkilerin sayısallaştırarak sunulmasını sağlayan bir yöntemdir [30].

Sosyal ağ analizi, genel olarak, ağdaki bireylerin ve ilişkilerin yapısına odaklanmıştır. Analiz sonucu elde edilen tespitler ile bireyler ve ilişkilere yönelik yeni çıkarımlar elde edilip, öneriler üretilebilir.
Sosyal ağ analizi temel üç adımda gerçekleşir. Bu adımlar şöyledir:

- **Veri elde etme:** Sosyal ağların sunduğu uygulama programlama ara yüzleri üzerinden, gerekli verilerin alınıp, analize hazır hale getirilmesine kadar devam eden süreçtir.
- **Analiz:** Sosyal ağdan elde edilen ve işlenmeye hazır hale getirilen verilerin, belirlenen çıktıları üretecek şekilde yorumlandığı işlem adımıdır.
- **Uygulama:** Analiz sürecinde yorumlanan verinin kullanıcıya sayısal ve grafiksel olarak sunulduğu aşamadır.

3.3. Tavsiye Sistemleri

Tavsiye sistemleri, kullanıcının kişisel tercih ve özelliklerine göre uygun içerikleri filtreleyip öneri olarak sunan sistemlerdir. Arama motorlarından farklı olarak kullanıcı istediği içeriği aramamakta, sistem kullanıcıya uygun öğeleri kendiliğinden bulmaya çalışmaktadır [31]. Günümüzde kullanıcıların ulaşabildiği dijital veriler çok büyük boyutlara ulaşmıştır. Bu nedenle bireylerin, bu devasa veri yığınları içerisinden, kendileri için uygun bilgileri elde etmeleri zorlaşmaktadır. Tavsiye sistemleri büyük bilgi yığınları içerisinden kullanıcıların ilgilerini çekebilecek olanları bulmalarına yardımcı olur [2].

3.3.1. Tavsiye yöntemleri

Genel olarak tavsiye sistemleri iki farklı tavsiye yöntemi kullanır. Bunlardan biri içerik tabanlı filtreleme diğeri ise işbirlikçi filtrelemedir [32].

İçerik tabanlı filtreleme: İçerik tabanlı filtreleme yönteminde öğelerin özellikleri kullanılır. Kullanıcının geçmişteki tercihlerinden yola çıkılarak önceki tercihlere benzer olan öğeler tavsiye olarak sunulur. Öğelerin ilişkilerinin belirlendiği veri tablosu üzerinden ilişki oranı yüksek olanlar tavsiye edilir. Tablo 1'de 4 adet ürünün birbiriyle olan ilişkileri verilmiştir. A ve C ürünleri arasında güçlü bir ilişki olduğu görülür. Bu ilişkiden yola çıkarak geçmişte A ürününü beğenen bir kullanıcıya C ürünü tavsiye olarak sunulabilir.

Tablo 1 - Ürün benzerlik oranları

	A	**B**	**C**	**D**
A	%100	%20	%80	%40
B	%20	%100	%35	%70
C	%80	%35	%100	%55
D	%40	%70	%55	%100

İşbirlikçi filtreleme: İşbirlikçi filtreleme kullanıcı tercihlerinin benzerliğinden yararlanan yöntemdir. Benzer tercihlere sahip kullanıcılar arasında ilişki kurularak, bir kullanıcının beğendiği herhangi bir ürün diğerine tavsiye olarak sunulur. Tablo 2, iki adet kullanıcının belirli ürünler hakkındaki beğenilerini, 1 ile 10 arasında verdikleri puanlarla göstermektedir.

Tablo 2 - Kullanıcıların ürünlere ilişkin beğenileri

	Kullanıcı 1	**Kullanıcı 2**
A	7	7
B	6	
C	2	5

D	8	8
E	9	

Kullanıcı 1 ve Kullanıcı 2'nin A ve D ürünlerine ait beğenileri aynıdır. Bundan yola çıkarak Kullanıcı 1 ve 2 arasında bir benzerlik olabileceği sonucuna ulaşılabilir. Kullanıcı 1'in yüksek oranda beğenip, Kullanıcı 2'nin henüz bir fikir sahibi olmadığı E ürünü, bu ilişkiden yola çıkılarak Kullanıcı 2'ye tavsiye edilebilir.

4. Yöntem

Geliştirilen yazılım temel beş adımdan oluşmaktadır. Çalışmaya ait iş akış şeması şekil 1'de verilmiştir. İlk adım Twitter platformundan verilerin elde edildiği "Twitter işlem" sürecidir. İkinci adım elde edilen verilerin işlenmeye hazır hale getirildiği "Kelime işlem" sürecidir. Üçüncü adım verilerin Türkçe doğal dil işleme sürecine tabi tutulmasıdır ve bu süreç de "Dil işlem" olarak ifade edilmiştir. Dördüncü adım elde edilen çıktıların büyük veri analiz tekniklerinden en yaygın kullanılanı olan MapReduce tekniği ile gruplanmasıdır. Analiz sürecinin ardından da kullanıcı tavsiyelerinin gerçekleştirildiği adıma geçilmektedir.

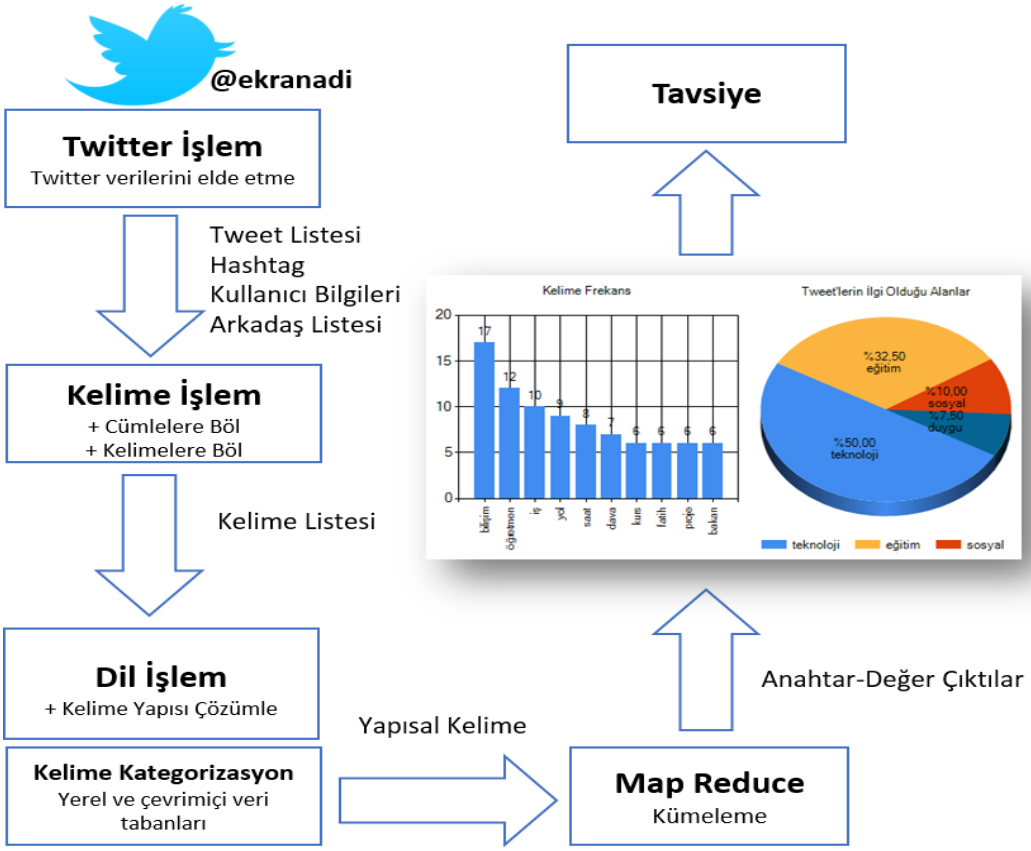
4.1. Twitter işlem

Twitter platformuna bağlanılarak, kullanıcılar ve paylaşımlara ilişkin verilerin, programa, uygun veri formlarında aktarıldığı iş süreçlerinin tümünü ifade eder. Bu iş sürecinin yerine getirilmesi için Twitter üzerinde bir API oluşturulmuştur. Bu API'ye program içerisinden uygun parametreler gönderilerek gerekli veriler elde edilir. Twitter işlem sürecinde yerine getirilen işler şunlardır:

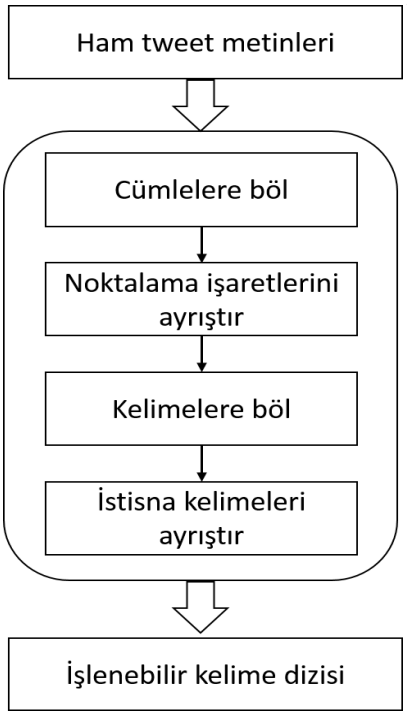
- **Kullanıcı profil bilgilerinin elde edilmesi:** Twitter ekran adı parametre olarak kullanılarak, ilgili kullanıcının ad, konum, hashtag, arkadaş listesi, profil resmi gibi bilgilerin gerçek zamanlı olarak programa aktarılmasıdır.
- **Kullanıcıya ait tweet'lerin elde edilmesi:** Twitter ekran adı ile o kullanıcıya ait tweet'lerin paylaşımlarının programa aktarılmasıdır.
- **Benzer kullanıcıların tespiti:** Bir Twitter kullanıcısını tanımlayan anahtar sözcükler ile bu kullanıcıyla benzeşen diğer kullanıcıların gerçek zamanlı olarak aranıp, bulunmasıdır.

4.2. Kelime işlem

Twitter platformundan alınan, belirli bir kullanıcıya ait, ham mesajlar dizisinin dil işleme sürecine hazırlandığı, analize konu olmayan verilerin ayıklandığı ve süreç sonunda işlenebilir kelime dizilerinin elde edildiği işlem adımıdır. Bu sürecin iş akışı Şekil 2'de gösterilmiştir.



Şekil 1. Çalışmanın iş akışı



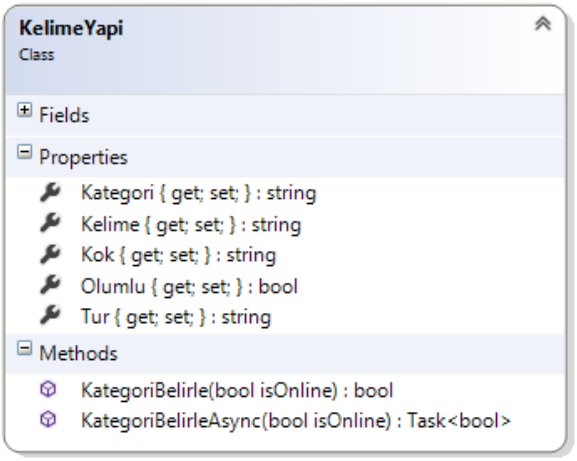
Şekil 2. Kelime işlem süreci iş akışı

Kelime işlem sürecinde yerine getirilen işlemler şunlardır:

- **Cümlelere bölme:** Bir tweet metninin, cümle sonu karakterlerine göre, yapı taşlarını oluşturan cümlelere parçalandığı süreçtir.
- **Noktalama işaretlerini ayrıştırma:** Cümleler içerisindeki noktalama işaretlerinin temizlendiği işlem adımıdır.
- **Kelimelere bölme:** Cümlelere ayrılan tweet verileri bu aşamada, kelimelerine ayrıştırılır.
- **İstisna kelimeleri temizleme:** Tweet'lerin parçalanması sonucu elde edilen kelimeler içerisinden, analizde kullanılmayacak olan bağlaç, edat, URL, atıf gibi verilerin ayrıştırıldığı süreçtir. Ayrıca çalışmaya ait veri tabanına daha önce istisna kelime olarak eklenmiş kelimeler de temizlenir.

4.3. Dil İşlem

Kelime işlem adımından elde edilen çıktıların doğal dil işleme sürecine tabi tutulup yapısal kelime formlarının elde edildiği işlem adımıdır. Yapısal bir kelimeyi ifade eden sınıf diyagramı Şekil 3'te verilmiştir. Doğal dil işleme sürecinde Zemberek isimli, açık kaynak kodlu kütüphaneden yararlanılmıştır. Kütüphane içerisinde kelime köklerini içeren ikili ağaç yapısı, kelimeye ait eklerin ve türlerinin yer aldığı yardımcı veri dosyaları yer almaktadır.

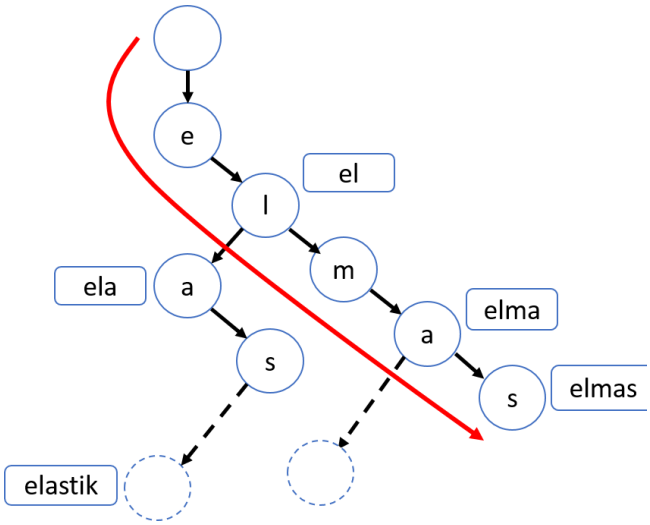


Şekil 3. Kelime yapı sınıf diyagramı

Dil işleme sürecinde gerçekleşen işlemler şunlardır:

4.3.1. Kelime kök ve eklerinin bulunması

Dil işleme kütüphanesi içerisinde yer alan ikili ağaç yapısı üzerinde yapılan aramalar ile kelime kök ve eklerinin tespit edildiği süreçtir. Süreç Şekil 4'te şematik olarak gösterilmiştir. Kelime kökünün aranması esnasında, kelimenin harfleri izlenerek ağaç üzerinde takip edilen yoldaki tüm kök adayları toplanır. Ardından çözümleme sürecine geçilerek tüm adaylar tek tek denenir. Genelde en uzun kök uygun seçim olacağından denemeye en uzun kökten başlanır. Örneğin "elmasları" kelimesi için "el", "elma" ve "elmas" kök adayları bulunur [33]. Eklerin tanımlı olduğu bir dosya üzerinden bu köklere getirilen kalıp halindeki eklerle doğru sonuç elde edilmeye çalışılır. En uzun kök olan "elmas" ele alınarak "-lar" çoğul eki kelimeye eklenir. Ardından aitlik eki olan "-ı" eklendiğinde doğru sonucun bulunduğu görülür. Diğer kök adaylarına bu ekler getirildiğinde aramaya esas olan "elmasları" kelimesi sağlanamaz.

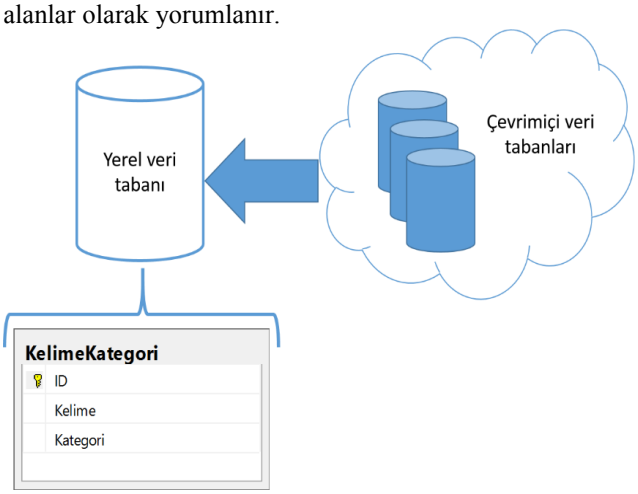


Şekil 4. Kelime kökü arama süreci [33]

4.3.2. Kelime yapısı çözümleme

Parametre olarak alınan kelimenin kökü, türü, olumlu ya da olumsuzluk anlamı ve kategori bilgisi ile paketlenerek "KelimeYapi" nesnesi olarak ifade edildiği süreçtir. Kelime köküne bağlanan bir olumsuzluk eki bulunmaması halinde kelimeye ait "Olumlu" özelliği mantıksal olarak doğru, olumsuzluk eki bulunursa yanlış değerine kurulur. Kelime türü kelime kökünün isim, fiil gibi Türkçe'deki hangi kelime türüne karşılık geldiğini ifade eder. Bu iş için de doğal dil işleme sınıfına ait özelliklerden yararlanılır. Kategori özelliği ise bir kelimenin hangi ilgi alanı ya da disipline ait olduğunu ifade eden bir özelliktir. Kelime kategorilerinin belirlenmesi için yazılımın yerel veri tabanı ya da çevrimiçi veri tabanlarından yararlanılır. Kategori tespit sürecinde çevrimiçi arama işlemi yapılmadan önce yerel veri tabanı üzerinde yer alan kelime-kategori ikilileri kontrol edilir. Eğer bu çevrimdışı arama sonucu istenilen kategori elde edilemezse, kategorisiz kelime için çevrimiçi arama yoluna gidilir (Şekil 5). 21

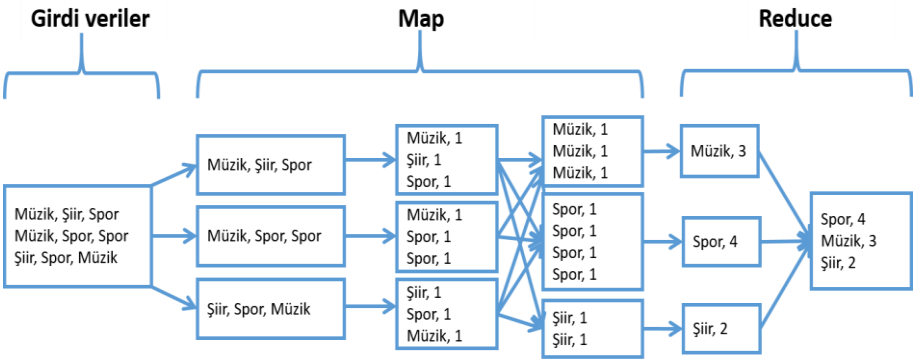
Bir kelimeye ait kategori bilgisi, analiz edilen kullanıcının ilgi alanının belirlenmesinde kullanılır. Sık tekrar eden kategoriler, kullanıcının en çok ilgi duyduğu alanlar olarak yorumlanır.



Şekil 5. Kelime kategorilerinin tespiti

4.4. MapReduce İşlemi

Analiz edilen Twitter kullanıcısının yaygın kullandığı kelimeler ve ilgi alanlarının anahtar-değer ikilileri halinde gruplandığı ve sıralandığı işlem adımıdır. Bu süreç, çok miktarda veriyi işlenebilir parçalara bölüp, belirli özelliklere göre özetleyerek bir küme oluşturmayı sağlar. Map ve Reduce olmak üzere iki farklı alt yordamdan oluşur (Şekil 6). Map verilerin ne olduğunun adreslendiği, Reduce ise adreslenen verilerin gruplandığı bir iş parçacığıdır.



Şekil 6. MapReduce süreci

Öncelikli olarak girdi veriler bloklara ayrılır. Map fonksiyonu her bir kelime ve kategoriyi ele alıp, anahtar-değer ikilileri formuna dönüştürür. Anahtar verinin kendisi iken, değer verinin frekansıdır. Anahtar ve değer ikilileri oluşturulduktan sonra sonuçlar gruplanır. Elde edilen veri seti Reduce işlevine aktarılır. Bu işlemin ardından elde edilen özet veri setinde verinin kendisi anahtar iken, frekanslara ait toplam ise değeri teşkil eder. Sonuçlar anahtar ya da değere bağlı olarak sıralı halde geri döndürülür.

Çalışmanın MapReduce işlem adımında, analiz edilen kullanıcının sık kullandığı kelimeler ve ilgi alanlarını haritalamak için iki ayrı MapReduce işlevi yer alır. İlkinde kelime kökleri adreslenerek toplam frekanslarına göre gruplanır; ikincisinde ise kelime kategorileri gruplanarak kullanıcı ilgi alanları elde edilir.

4.5. Tavsiye İşlemi

Sık kullandığı kelimeler ve ilgi alanları tespit edilen bir Twitter kullanıcısı için aynı kelimeleri sık tekrar eden ya da aynı ilgi alanında tweet'ler paylaşan kullanıcılardan en yüksek benzerliğe sahip olanlarının belirlendiği adımdır. Bu süreç, yöntem içerisinde sözü geçen diğer işlem adımlarının üzerine inşa edilir. Bu adımda, ilgili kullanıcının takip edebileceği diğer Twitter kullanıcıları gerçek zamanlı olarak aranıp, tavsiye olarak sunulur. Tavsiye işlemine hazırlık sürecinde sık kullanılan kelimeler ve kullanıcı ilgi alanlarının oransal tespiti gerçekleştirilir. Ardından şartlı olasılık teoremi ve tabu arama yöntemlerinden yararlanılarak en uygun tavsiyeler elde edilir.

4.5.1. Sık kullanılan kelimelerin oransal tespiti

Analiz edilen kullanıcının sık tekrar ettiği kelimelerin MapReduce süreci ile frekans dağılımları elde edilir. Bu dağılımlardan yola çıkılarak her bir kelimenin kullanım oranı, küme içindeki diğer tüm kelime kullanımlarına bağlı olarak belirlenir. Kelimenin tekrar oranı tavsiye edilecek kullanıcıların tespitinde kullanılan benzerlik katsayısını belirler. Bu katsayı tavsiye edilecek kullanıcılara ilişkin yapılan aramada ilgili kelimeyi tekrar eden kullanıcıya, katsayı oranınca artı değer kazandırır.

4.5.2. Kullanıcı ilgi alanlarının oransal tespiti

Kullanıcı ilgi alanları, tweet'ler içerisinde geçen kelime desenlerinin kategorizasyonu ile tespit edilir. Kelimeleri kategorize edebilmek için, kelime-kategori ikililerini barındıran veri setinden yararlanılır. Bu veri seti yazılımın yerel veri tabanı içerisinde bulunmakta olup, zaman içerisinde yapılan analizler ile elde edilen yeni kelimelere göre, web üzerinde yapılan aramalarla sürekli olarak genişlemektedir.

Türkçe' de bazı kelimeler birden çok anlam ifade edebileceği için tweet içerisinde kullanılan bir kelime birden çok kategoriye karşılık gelebilir. Bu durumda kullanıcının paylaşımlarının geneli üzerinden olasılıksal bir değerlendirme yapılmalıdır. Herhangi bir kelime için yanlış kategori eşleşmesi ile yanlış ilgi alanının ortaya çıkmaması için diğer paylaşımlarda yer alan kelime kategorilerinin ağırlıklarından yararlanılır. Kelime-kategori eşleşmeleri ile elde edilen kategori havuzu içerisinde, daha önce kelime frekanslarında da yapıldığı gibi, oransal bir hesap yapılır. İlgili kategorinin, havuz içerisindeki diğer tüm kategoriler toplamındaki kullanım oranına bakılır. Bu oran da tavsiye edilecek kullanıcıların, ilgili kullanıcıya benzerliğinin tespitinde bir katsayı olarak kullanılır.

4.5.3. Şartlı olasılık teoremi

Olasılık kuramı üzerine kurulu ve bir olasılık dizisi içerisindeki koşullu olasılıkları ifade etmeye yönelik bir teoremdir. Olasılık hesaplamaları ile sisteme sunulan verilerin sınıfını yani kategorisini tespit etmeyi amaçlar [34]. Hâlihazırda kategorize edilmiş veri seti temel alınarak, girdi verinin var olan veriye göre hangi sınıfta yer aldığı oransal olarak tespit edilir. A değerinin B'ye göre şartlı olasılığını ifade eden denklem şu şekildedir:

$$P(A\backslash B) = (P(B\backslash A)\, P(A)) / P(B) \qquad (1)$$

Bu teoremin işletilebilmesi için daha önceden bir veri kümesi bulunmalı ve bu verilerin de birer kategorisi olmalıdır. Sisteme sunulan yeni verilerle mevcut veriler kıyaslanarak, yeni verinin hangi kategori içerisinde yer aldığı tespit edilir. Tablo 3'te yer alan veri seti incelenirse şarkı ve gitar kelimeleri müzik kategorisiyle, şiir ve roman kelimeleri ise edebiyat kategorisiyle ilişkilidir. Bundan yola çıkarak 1, 2 ve 3 numaralı satırlarda oransal olarak müzik kategorisindeki veriler ağırlıklı olduğu için bu satırın kategorisi 2/3 oranla

müzik olarak belirlenmiştir. Benzer şekilde 2 nolu satır 2/3 ve 3 nolu satır da 2/2 oranla müzik kategorisinde yer alır. 4 nolu satır ise 2/3 oranla edebiyat kategorisindedir. Tüm tablo ele alındığında müzik kategorisi 3/4 oranla daha baskındır.

Tablo 3. Şartlı olasılık örnek veri seti

Sıra No	İçerik	Kategori
1	şarkı, şiir, şarkı	müzik
2	şarkı, şarkı, şiir	müzik
3	şarkı, gitar	müzik
4	şiir, roman, şarkı	edebiyat

Tweet'ler incelenirken, içerisinde geçen kelimelerin kategorik oranlarına göre puanlanır. Kullanıcıya ilişkin analize konu olan tüm tweet'lerin değerlendirilmesi ile de genel kategorik bilgi yani kullanıcının ağırlıklı ilgi alanları tespit edilmiş olur.

Tespit edilen kategoriler ve olasılık değerlerine bakılarak Twitter üzerindeki diğer kullanıcılara ilişkin son paylaşımlar analiz edilir. İlgili kategorilerle örtüşen tweet'ler toplanarak, bu tweet'leri gönderen kullanıcılar, ilgili kategori ve olasılık değeri ile tavsiye kullanıcı havuzuna alınır. Tüm kullanıcılar için bu işlem gerçekleştirilir. Bir kullanıcı birden çok kategoriye girecek şekilde tweet paylaşmış olabilir. Bu durumda tüm ilgili kategorilerin olasılıkları ilgili mesajın gönderenine artı puan olarak eklenir. Ardından analiz edilen kullanıcının tekrar ettiği kelimelerin, benzer şekilde hesaplanan şartlı olasılık değerlerine göre, bu kelimeleri içeren paylaşımları gönderen kullanıcılar puanlanır. Daha sonra en iyi adayların seçimi işlemine geçilir.

4.5.4. Tabu arama yaklaşımı ve en iyi adayların seçimi

Şartlı olasılıklar göz önüne alınarak Twitter üzerinde yapılan gerçek zamanlı aramalar sonucu elde edilen ve analize konu olan kullanıcıya takip etmesi için önerilecek adaylar içerisinden en uygun olanların seçiminde Tabu arama yaklaşımından yararlanılır. Bu yaklaşımın temelinde, en iyi çözüme ulaşırken, uygun olmayan çözümlerin tabu belleği isimli bir yasaklı listesine alınması ve ilerleyen aşamalarda bu liste dışındaki çözümlerin değerlendirilmesi mantığı vardır. Ancak olasılıklar değerlendirilirken uygun görülmeyen bir çözüm hemen yasaklanmamalı, en iyiye aday olabilecek ara çözümlere şans verilmelidir.

Twitter kullanıcısının analizi sonucu tavsiye edebilecek adaylar için ilgi alanı ve tekrar edilen kelimelerin benzerliği olmak üzere iki ana benzerlik durumu söz konusudur. Ayrıca bu iki ana durum için kategori ve kelimenin şartlı olasılığı ya da diğer ifadeyle benzerlik katsayısı da devreye girmektedir. Buna bağlı olarak tavsiye edilecek bir adayın uygun kategoriye girmemesi ya da zayıf olasılıklı bir kategori içermesi gerekçesi ile çözüm kümesinden direkt çıkarılması doğru bir yaklaşım olmayacaktır.

Tavsiye edilecek adaylar, birden çok kategoriye ilişkin paylaşımda bulunmuş olabilir ya da sık tekrar edilen kelimelerden birden çoğunu kullanmış olabilir. Bu nedenle aday bireyin her bir kategori ve kelimenin olasılık değeri ile puanlaması yapılır. Her bir adım sonucunda çözüm kümesi içerisinde belirli sayı ya da orandaki aday uygun çözümler olarak kabul edilirken, arada kalan kesim olası adaylar kümesinde yer alır. Olası adaylar bir sonraki adımda alacakları puanlarla ideal çözüm kümesi içerisinde yer alabilir. Olası tüm çözümler puan değerlerine göre sıralandığında, benzerlik oranı çok düşük ya da sıfır olan adaylar yasaklı listesine alınarak çözüm kümesinden çıkarılır.

Kategorik puanlama süreci ardından, anahtar kelime bazlı değerlendirme gerçekleşir. Kelime bazlı gerçekleşen bu olasılıksal puanlama ile nihai aday listesi ortaya çıkar. Kategoriler kelimelerin üzerinde bir kapsayıcı olup genellenmiş ifadelerdir. Anahtar kelimeler ise kategorik benzerlikleri daha da özele indirgeyerek detaylı puanlama yapılmasını sağlar.

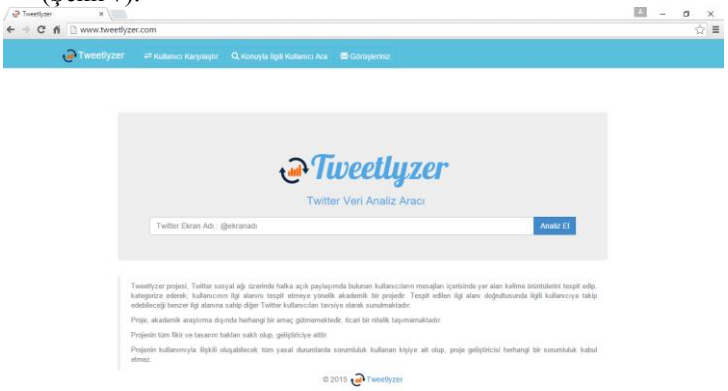
Değerlendirme süreçleri tamamlandığında elde edilen nihai aday listesi son puan durumuna göre tersten sıralanır. En yüksek puana sahip 100 adet aday kullanıcıya tavsiye olarak sunulur.

5. Çalışmanın Web Arayüzü

Geliştirilen yazılım "Tweetlyzer" adında bir web uygulaması olarak kullanıma sunulmuştur. 2015 yılının ikinci çeyreğinden itibaren altı aylık süreçte iki bin iki yüz dolayında profil analizi gerçekleştirilmiştir. Web uygulaması iki ana ekrandan meydana gelmektedir. Bunlar kullanıcı analiz sayfası ve kullanıcı profil sayfasdır.

5.1. Kullanıcı analiz sayfası

Bu sayfa uygulamanın karşılama ekranı olup Twitter kullanıcı analiz sürecinin başlangıç noktasıdır. Sayfa, ilgili kullanıcının Twitter ekran adının girilerek analiz sürecinin başlatılmasını sağlayan bir yapıya sahiptir (Şekil 7).



Şekil 7. Tweetlyzer kullanıcı analiz sayfası

Analiz edilecek kullanıcının ekran adı girilerek "Analiz Et" butonuna tıklandığında iş katmanında yer alan metotlar devreye girerek gerçek zamanlı profil analiz süreci başlatılır. Girilen ekran adı geliştirilen yazılım ile ilişkilendirilen Twitter API'ye parametre olarak gönderilir. Twitter API tarafından gerçekleştirilen sorgulama işlemi sonucu gelen veriler uygulamaya aktarılır. Bu aşamadan sonra verilerin yapılandırılması ve analizine yönelik işlemler gerçekleşmeye başlar. Ardından kullanıcı, Twitter'dan elde edilen veriler ve analiz sonuçlarını görmek üzerine kullanıcı profil sayfasına yönlendirilir.

5.2. Kullanıcı profil sayfası

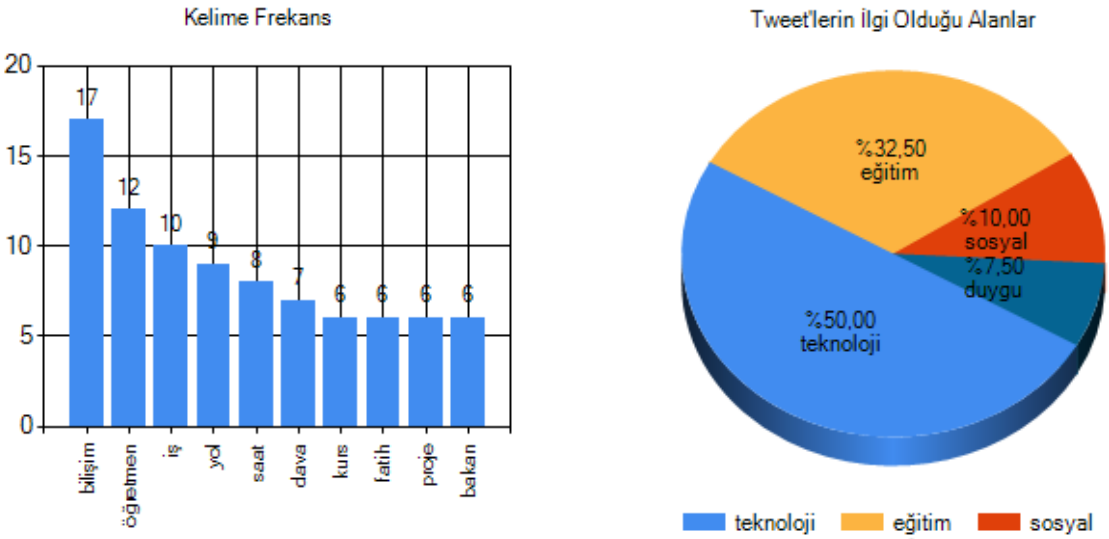
Analiz edilen Twitter kullanıcısına ilişkin temel bilgiler, analiz sonuçları ve kullanıcı tavsiyelerinin yer aldığı sayfadır (Şekil 8). Temel üç kısımdan meydana gelmektedir.

Sayfanın üst kısmını oluşturan ilk bölüm, kullanıcının profiline ilişkin temel bilgileri ve Twitter üzerindeki son aktivitelerini içerir. Sayfanın sol kısmında temel profil bilgileri, son kullanılan Twitter etiketleri ve yakın geçmişteki arkadaş olduğu diğer kullanıcılar görüntülenir. Ortada yer alan panel içerisinde kullanıcının paylaşmış olduğu en son tweet'ler yer alır. Sağ kısımdaki panelde ise paylaşılan tweet'ler içerisinde en sık tekrar eden kelimeler ve kelimelerin kullanım sayıları görüntülenir.



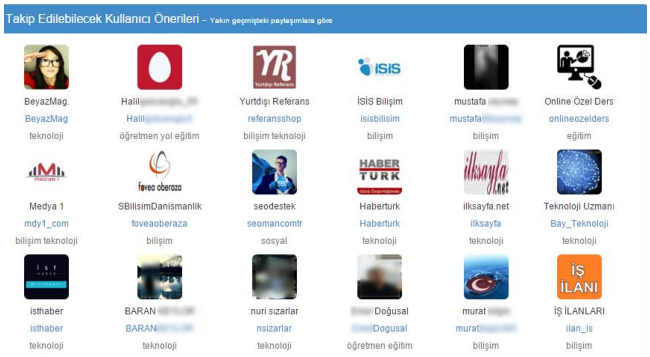
Şekil 8. Kullanıcı analiz sayfası

Profil sayfasının orta bölümünde tweet'lerin analizi ile elde edilen kelime frekans ve ilgi alanı grafikleri yer almaktadır (Şekil 9). Kelime frekans grafiği, kullanıcının yaygın olarak kullandığı on adet kelimeyi bir çubuk grafik üzerinde görüntüler. Yaygın kullanılan kelimeler ve frekanslarının tespiti esnasında yöntem başlığı altında anlatılan iş süreçleri kullanılır. Tweet'ler içerisinde geçen kelimelerin kategorizasyonu ile elde edilen verilerin işlenmesi sonucunda her bir kategorinin oranı belirlenir. Kategorizasyon süreci sonunda elde edilen veriler kullanıcının ilgi alanlarını yansıtmaktadır. En yüksek orana sahip ilk dört kategori ilgi alanı grafiğinde dairesel yüzdelik dilimlerle ifade edilir.



Şekil 9. Kelime frekans ve ilgi alanı grafikleri

Profil sayfasının alt bölümünde analiz edilen kullanıcıya, ilgi alanları doğrultusunda, takip etmesi önerilen kullanıcıların listelendiği panel yer alır (Şekil 10). Kullanıcı öneri listesinde ilgili kullanıcıların profil resimleri, Twitter kullanıcı adları ve profil erişim linkleri ve ayrıca ilgili profilin analiz edilen kullanıcı ile hangi açıdan daha çok benzeştiğine ilişkin anahtar sözcük ya da kategoriler yer alır.



Şekil 10. Takip edilebilecek kullanıcı önerileri listesi

6. Araştırma Bulguları ve Tartışma

Geliştirilen yazılım iki binin üzerinde Twitter kullanıcı ile test edilmiştir. Twitter'ı aktif kullanan yüz kişilik bir kullanıcı grubundan alınan düzenli geri bildirimler ile elde edilen sonuçların yorumlanması sonucu sistemin %86 oranında başarılı olduğu tespit edilmiştir.

6.1. Kullanıcı Örneklemi

Uygulama test sürecinin temelini Twitter'ı aktif olarak kullanan, yaşları 18 ile 25 arasında değişiklik gösteren üniversite düzeyinde yüz adet kullanıcı oluşturmaktadır. Kullanıcıların geneli günde 1 ile 7 arasında değişen sayıda tweet paylaşımında bulunmaktadır.

Bu kullanıcılar, belirli zamanlarda sistem üzerinde kendi Twitter profillerini analiz etmişlerdir. Analiz sonucu sistem tarafından çıkarsanan, yaygın kullanılan kelimeler, ilgi alanları, tavsiye kullanıcılara ilişkin bilgiler ve sistemin geneli hakkındaki görüşlerini geri bildirim olarak iletmişlerdir.

6.*2. Sistem başarı ölçütleri*

Geliştirilen uygulama iki ana unsur üzerine kuruludur. İlki kullanıcı ilgi alanlarının anlık ve doğru tespiti, diğeri ise bu doğrultuda doğru tavsiyelerin sunulmasıdır. İlgi alanlarının doğru tespiti ise kelimelerin doğru analiz edilmesine bağlıdır. Bu anlamda sistemin ana başarı ölçütlerini şu şekilde sıralamak mümkündür:

- Yaygın kullanılan kelimelerin belirlenmesi
- İlgi alanlarının doğru tespiti
- Tavsiyelerin ilgi alanına uygunluğu
- Anlık değişimlere cevap verebilme

Yaygın kullanılan kelimeler paylaşımlar içerisinde yer alan kelimelerin biçimsel olarak analiz edilmesinden sonra elde edilen kök gruplarıdır. Bu süreç temelde doğal dil işleme kütüphanesi olmak üzere, yazılımın kelime işleme sürecinin başarısı ile ilgilidir.

İlgi alanlarının doğru tespit edilmesi kelime-kategori ikililerinin doğru eşleşmeleri ve veri tabanındaki kategori veri setinin genişliği ile ilintili olup, çevrimiçi ve yerel veri tabanlarından elde edilen sonuçlarının yeterliliği doğru tespit ihtimalini arttırmaktadır. Ayrıca MapReduce ve şartlı olasılık analiz süreçlerinin başarısı da ilgi alanlarının doğru tespitini etkilemektedir.

Sistem tarafından üretilen kullanıcı tavsiyelerinin analize konu olan kullanıcının ilgi alanına uygunluğu Twitter API vasıtası ile platform üzerinde gerçek zamanlı olarak yapılan aramaların başarısına bağlıdır. Arama işleminin başarısını doğru arama kriterinin kullanılması belirler. Yazılım içerisindeki arama kriterlerini yaygın kullanılan kelimeler ve tespit edilen kategorilere ilişkin anahtar sözcükler belirler. Twitter'dan dönen sonuç kümesinden en uygun bireylerin seçilmesinde ise yine şartlı olasılık analizi ve tabu arama tekniklerinin uygulanmasındaki başarı etkilidir.

Twitter kullanıcılarının zaman içerisinde değişen eğilimleri paylaşımlarına da yansımaktadır. Değişen eğilimlere sistemin cevap verebilmesindeki en önemli etken çevrimiçi ve yerel veri tabanlarındaki kelime-kategori ikililerinin miktarı ve doğruluğudur. Veri miktarının çeşitlenmesi doğru sonuçları hassasiyetle bulabilmeyi kolaylaştırmaktadır. Ayrıca zaman içerisinde gerçekleştirilen iyileştirme çalışmaları da sistemin doğru sonuç üretme karakteristiğini arttırmaktadır.

6.3. Sistemin değerlendirilmesi

Geliştirilen yazılımın başarılı sayılabilmesi için, belirlenen başarı ölçütleri temelinde, %70 oranında olumlu dönüt alınması kriteri konmuştur. Yayımlandığı ilk günden bu yana belirli profiller farklı aralıklarla birden çok kez olmak üzere, toplamda iki binin üzerinde analiz işlemi gerçekleştirilmiştir. Her analiz işlemi sonucunda sistem tarafından üretilen verilere ilişkin kullanıcılardan geri bildirimler alınmıştır. Test grubundaki yüz adet kullanıcının, sistemin ilk kullanıma açıldığı dönemden itibaren üç ayrı periyotta sistemi yeniden test etmeleri istenmiştir. Periyot aralarında, gelen geri bildirimler doğrultusunda sistem üzerinde iyileştirmeler yapılmıştır. Her periyodun sonunda gelen dönütler değerlendirilerek Tablo 4'te yer alan sonuçlar elde edilmiştir

İlk değerlendirme döneminde elde edilen başarı oranları zaman içerisinde yazılım üzerinde yapılan iyileştirmeler ve veri tabanının genişlemesi ile birlikte yükselme göstermiştir. Test grubundan üç ayrı dönemde alınan geri bildirimler sonucu sistemin %86 oranında başarılı olduğu sonucuna ulaşılmıştır. Ulaşılan bu başarı oranı ile sistemin başarılı sayılabilmesi için konulan %70 kriterinin üzerine çıkılmıştır.

6.4. Twitter Veri Analizinde Ortaya Çıkan Olası Sorunlara İlişkin Değerlendirme

Twitter veri analizi büyük oranda bir dil analiz süreci olup kullanılan dilin yapısı veya dilin doğru kullanılması ile yakından ilgilidir. Türkçe geniş bir dil olmanın yanında sabit kuralları olan bir dil değildir. Ayrıca dillerin kural çerçeveleri ne kadar belirgin bir şekilde çizilse de pratikte bu kuralların pek çoğuna uyulmamaktadır. Bu noktada veri analizi konusunda karşılaşılabilecek pek çok sorun dil yapısı ve kullanımıyla ilişkilidir. Bu çalışmanın geliştirilmesi esnasında en sık karşılaşılan sorunlar ve çözüm önerileri şu şekilde değerlendirilebilir:
.

Tablo 4. Değerlendirme sonuçları

Ölçüt/ Periyot başarı yüzdesi	**1. periyot**	**2. periyot**	**3. periyot**	**Ortalama**
Yaygın kullanılan kelimelerin doğru tespiti	87	93	94	91
İlgi alanlarının tespiti	83	85	90	86
Tavsiye kullanıcıların ilgi alanlarına uygunluğu	80	86	92	86
Değişen yönelimlere cevap verebilme	78	80	89	82
Ortalama	**82**	**86**	**91**	**86**

6.4.1. Yazım hataları

Sosyal ağ paylaşımlarında kullanıcılar bilinçsiz ya da bilinçli olarak kelimeleri yanlış yazabilmektedir. Bu hatalı yazımlar dil işleme sürecini zorlaştırmakta, belirsizliklere ve yanlış çıkarımlara sebep olabilmektedir. Bu sorunun çözümü için, doğru yazım önerileri sunabilen bir doğal dil işleme kütüphanesinin kullanılması yararlı olacaktır. Geliştirilen yazılımda kullanılan doğal dil işleme kütüphanesi, içerisinde barındırdığı kelime kök ağacı ile yanlış yazılan kelimeler için doğru alternatifler sunabilmektedir. Ancak kelime yapısının büyük oranda deforme edildiği durumlarda ilgili kelime için doğru yazım önerisi bulunamamakta ve bu halde kelime değerlendirmeye alınamamaktadır.

6.4.2. İroni

İroni, söylenenin tam tersinin ima edildiği durumdur. Türkçe yapı itibari ile ironik cümleler kurmaya müsait bir dildir. İronik cümlelerdeki imayı cümlenin biçimsel ögelerinden anlayabilmek zor olup, ancak cümlenin zihinsel süreçle yorumlanmasından sonra kastedilen mana sezilebilmektedir.

6.4.3. Çok sayıda anlamı olan kelimeler

Türkçe'de birden çok anlamı olan kelimeler bulunmaktadır. Bu durum kelime-kategori eşleştirmelerinde, ilgili kelimenin hangi alana ait olduğunun tespitinde zorluk çıkarır. Böyle bir halde tek başına o kelimeye ilişkin kategori bilgisine odaklanmak yanlış bir çıkarıma sebep olabilir. Bu nedenle gerçekleştirilen yazılım içerisinde şartlı olasılık teoremi ile çok sayıda kategori içerisinden en uygun olanının tespiti yoluna gidilmiştir.

6.4.4. Yabancı sözcük kullanımı

Twitter kullanıcıları, Türkçe paylaşımları içerisinde zaman zaman yabancı sözcükler de kullanmaktadır. Bu durumda Türkçe odaklı gerçekleştirilen bu çalışmada yabancı kelimenin yorumlanamaması sonucu ortaya çıkmaktadır. Ancak bazı yabancı sözcükler yazım itibariyle Türkçe'deki bir başka sözcüğe benzemektedir. Bu durumda kelime Türkçe gibi algılanıp, yanlış kategorizasyona sebep olabilmektedir. Diller arası çok anlamlılıktan kaynaklı bu sorun, yine, şartlı olasılık teoremi ile tolere edilmektedir.

7. Sonuç

Teknolojinin gelişmesi ve internet kullanımının yaygınlaşması ile birlikte web üzerinde paylaşılan bilgi miktarı da hızlı bir şekilde artış göstermiştir. Özellikle sosyal ağlar günümüzün en çok bilgi üretilen ortamlarıdır. Sosyal ağlarda çok miktarda verinin üretilmesi, karmaşık bir bilgi yığını ortaya çıkarmakta ve pek çok yararlı bilgi hedefine ulaşamadan bu karmaşa içerisinde kaybolmaktadır. Sosyal ağ kullanıcıları gün içerisinde çok miktarda gerekli ya da gereksiz bilgiye maruz kalmakta, bu durum ciddi bir bilgi kirliliği yanında, zihinsel yorgunluk ve yanlış yönelimlere de sebep olabilmektedir. Bu noktada sosyal ağlar üzerinde doğru bilginin elde edilmesine yönelik çalışmaların yapılması gerekli hale gelmektedir.

Son dönemin en hızlı büyüyen sosyal ağ platformu Twitter üzerinde her dakika 100 binden fazla paylaşım gerçekleştirilmektedir. Tweet adı verilen bu paylaşımlar 140 karakterden oluşan metin formunda ve yapısal olmayan verilerdir. Üretilen bu verilerin içerisinde belirli disiplinlere ait, belirli kullanıcı kitlelerine hitap eden yararlı içerikler yer alabileceği gibi, anlamsız pek çok öge de bulunabilir. Twitter kullanıcılarının platform üzerinde üretilen ve kendi ilgi alanlarına hitap eden yararlı içeriklere erişmesinin sağlanması önem taşımaktadır.

Twitter platformu kendi içerisinde belirli tavsiye mekanizmaları barındırmaktadır. Bu mekanizmalar temelde iki farklı mantık üzerine kuruludur.

İlki içerik tabanlı filtreleme diğeri ise işbirlikçi filtrelemedir. İçerik tabanlı filtrelemede kullanıcının geçmişteki tercihlerinden yola çıkılarak tavsiyeler üretilir. İkincisi olan işbirlikçi filtreleme tekniğinde ise birbirini takip eden kullanıcıların herhangi birinin ilgilendiği bir içerik diğer ilişkili kullanıcının da dikkatini çekebileceği düşüncesi ile tavsiye olarak sunulur.

Bu iki tavsiye yaklaşımı kullanıcının zaman içerisinde eğilimlerinin değişebileceğini ya da bir kullanıcının ilişkili diğer kullanıcının ilgilendiği her şeyle alakadar olamayacağını göz ardı etmektedir. Bu nedenle kullanıcının kendine ait düşünce ve eğilimlerini dinamik olarak çıkarsayan ve bu doğrultuda kullanıcıya gerçek zamanlı tavsiyeler sunabilen bir sisteme ihtiyaç vardır.

Geliştirilen yazılım, bir Twitter kullanıcısının profilini gerçek zamanlı olarak analiz etmekte ve bu doğrultuda tespit edilen ilgi alanlarına göre takip edebileceği benzer kullanıcıları öneri olarak sunmaktadır. Kullanıcı ekran adının parametre olarak gönderilmesi ile Twitter uygulama programlama üzerinden anlık olarak elde edilen veriler doğal dil işleme, büyük veri analizi, şartlı olasılık teoremi ve tabu arama yaklaşımından yararlanılarak analiz edilir. Belirli zamanlarda alınan sabit verilerle değil, dinamik elde edilen verilerle yapılan bu çıkarımlar sayesinden kullanıcının gerçek zamanlı ilgi alanları ve beklentileri elde edilebilmektedir. En çok ilgi duyulan alanlar şartlı olasılık teoremi uyarında tespit edilmektedir. Tespit edilen kullanıcı ilgi alanları doğrultusunda, Twitter üzerinde yapılan ve yine gerçek zamanlı aramalarla, benzer ilgi alanına sahip diğer kullanıcılar belirlenmektedir. En çok benzerliğe sahip kullanıcıların belirlenmesinde sık kullanılan kelimeler ve ilgi alanları temelinde, tabu arama yaklaşımından yararlanılmaktadır.

Çalışma bir web sayfası olarak kullanıma açılmış, yüz adet aktif Twitter kullanıcısı ile düzenli aralıklarla test edilmiştir. Sistemin başarılı sayılabilmesi için %70 oranında bir başarı kriteri konulmuştur. Alınan geri bildirimlerde sistemin başarı oranının %86 olduğu tespit edilmiştir. Bu durum belirlenen ön başarı ölçütünün üstüne çıkıldığını göstermektedir.

Yapılan çalışma genel anlamda Twitter kullanıcılarına, Twitter üzerinde yer alan fakat henüz ulaşamadıkları, kendilerine hitap eden kullanıcıları filtreleyip tavsiye etme işlevini yerine getirmektedir. Bunu yaparken de ilgili kullanıcının paylaşımlarından çıkarsadığı anlık eğilimleri kullanmaktadır. Yapılan çalışma gerçekleştirdiği bu fonksiyonun yanında eğitim, ekonomi, güvenlik gibi pek çok alana hitap edecek şekilde de kullanılabilir niteliktedir. İlgi alanı çıkarsanan kullanıcıya kendi eğilimi doğrultusunda içeriklerin sunumu, ilgi duyduğu konudaki entellektüel seviyesinin artmasına katkı sunabilir. Bu altyapı ile sosyal medyada yayınlanacak olan reklamlar kullanıcı ilgi alanına uygun olarak sunulabilir. Ortaya çıkabilecek herhangi bir toplumsal olaya ilişkin fikirler yahut kullanıcının belirli bir fikir alanına olan eğilimleri tespit edilebilir. Bu anlamda yapılan çalışma pek çok alana katma değer sağlayabilir niteliktedir.

Teşekkür

Bu çalışmayı 4083-Dl-14 No`lu proje ile destekleyen Süleyman Demirel Üniversitesi Bilimsel Araştırma Projeleri Yönetim Birimi Başkanlığı'na teşekkür ederiz.

Kaynakça

[1] R. Kalakota, "Sizing 'Mobile + Social' big data stats," Business Analytics 3.0, URL: https://practicalanalytics.co/2012/10/22/sizing-mobile-social-big-data-stats (Erişim zamanı: Ocak, 10, 2016).
[2] T. Aytekin, Tavsiye Sistemleri (Öneri Sistemleri), URL:http://akademik.bahcesehir.edu.tr/~tevfik/tavsiye _sistemle ri_oneri_sistemleri.html(SonErişim:5.5.2015)
[3] J. Hannon, M. Bennett, and B. Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," Fourth ACM conference on Recommender systems, pp. 199–206, (2010).
[4] Z. Yu, C. Wang, J. Bu, X. Wang, Y. Wu, and C. Chen, "Friend recommendation with content spread enhancement in social networks," Information Sciences, vol. 309, pp. 102–118, (2015).
[5] CC. Chang and KH. Chu, "A Recommender System Combining Social Networks for Tourist Attractions", Fifth International Conference on Computational Intelligence, Communication Systems and Networks, pp. 42-47, (2013).
[6] CH. Lee and TF. Chien, "Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking," Journal of Information Science, vol. 39, no. 4, pp. 523–543, (2013).
[7] ML. Kern, JC. Eichstaedt, HA. Schwartz, G. Park, LH. Ungar, DJ. Stillwell, M. Kosinski, L. Dziurzynski, M. Seligman, "From "Sooo Excited!!!" to "So Proud":

Using Language to Study Development", Developmental Psychology, vol. 50 no. 1, pp. 178-188, (2014).
[8] LJ. Zhao, YJ. Ren, J. Wang, LS. Meng, CL. Zou, "Research on the Opinion Mining System for Massive Social Media Data", Natural Language Processing And Chinese Computing, vol. 400, pp. 424-431, (2013).
[9] R. Compton, C. Lee, TC. Lu, L. De Silva, M. Macy, "Detecting future social unrest in unprocessed Twitter data Emerging Phenomena and Big Data", IEEE International Conference on Intelligence and Security Informatics: Big Data, Emergent Threats, And Decision-Making In Security Informatics, pp. 56-60, (2013).
[10] A. Bruns, T. Highfield, and J. Burgess, "The Arab spring and social media audiences," American Behavioral Scientist, vol. 57, no. 7, pp. 871–898, (2013).
[11] XT. Vu, MH. Abel, P. Morizet-Mahoudeaux, "A user-centered and groupbased approach for social data filtering and sharing", Computers in Human Behavior, vol. 51, pp. 1012-1023, (2014).
[12] L. Wu, N. Luo, "Social streams recommendation in sina microblog with relation of user and interest", IEEE International Conference on Information Science and Technology, pp. 480-483, (2014).
[13] S. Puglisi, J. Parra-Arnau, F. Jordi, D. Rebollo-Monedero, "On contentbased recommendation and user privacy in social-tagging systems" Computer Standards & Interfaces vol. 41, pp. 17-27, (2015).
[14] MC. Yang and HC Rim, "Identifying interesting Twitter contents using topical analysis", Expert Systems with Applications, vol. 41, pp. 4330-4336, (2014).
[15] D. Kang, D. Han, N. Park, S. Kim, U. Kang, S. Lee, "Eventera: Real-Time Event Recommendation System from Massive Heterogeneous Online Media". IEEE International Conference on Data Mining Workshop, pp. 1211-1214, (2014).
[16] I. Uysal and WB. Croft, "User oriented tweet ranking: a filtering approach to microblogs", 20th ACM international conference on Information and knowledge management, 2261-2264, (2011).
[17] SJ. Yu, "The dynamic competitive recommendation algorithm" Social network services, vol. 187, pp. 1-14, (2012).
[18] AR, Sun, J. Cheng, DD. Zeng, "A novel recommendation framework for micro-blogging based on information diffusion", 19th Annual Workshop on Information Technologies and Systems, (2009).
[19] B. Liang, Y. Liu, M. Zhang, S. Ma, L. Ru, K. Zhang, "Searching for people to follow in social networks", Expert Systems with Applications vol. 41, no. 16, pp. 7455-7465, (2014).
[20] N. Jamil, AC. Alhadi, SA. Noah, "A collaborative names recommendation in the Twitter environment based on location", In Semantic Technology and Information Retrieval (STAIR) 2011 International Conference on, pp. 119-124, (2011).
[21] M. Islam, C. Ding, C. Chi, "Personalized Recommender System on Whom to Follow in Twitter" Big Data and Cloud Computing (BdCloud), IEEE Fourth International Conference on, pp. 326-333, (2014).
[22] WJ. Lee, KJ. Oh, CG. Lim, HJ. Choi, "User profile extraction from Twitter for personalized news recommendation", In Advanced Communication Technology (ICACT) 16th International Conference on, pp. 779-783, (2014).
[23] H. Lu and C. Lee, "The Topic-Over-Time Mixed Membership Model (TOTMMM): A Twitter Hashtag Recommendation Model that Accommodates for Temporal Clustering Effects", Web Intelligence (WI) and Intelligent Agent Technologies (IAT) 2013 IEEE/WIC/ACM International Joint Conferences on, pp. 21-25, (2015).
[24] N. Jonnalagedda and S. Gauch, "Personalized News Recommendation Using Twitter", In Web Intelligence (WI) and Intelligent Agent Technologies (IAT) IEEE/WIC/ACM International Joint Conferences on, pp. 21-25, (2013).
[25] ME. Babacan, İ. Haşlak, İ. Hira, "Sosyal Medya ve Arap Baharı", Akademik İncelemeler Dergisi, 6(2), 63-91, (2011).
[26] YE., Sarıgül, "Sosyal Medya Nedir?", Sosyal Pedia, URL: http://www.sosyalpedia.com/sosyal-medya-nedir-1 (Erişim zamanı: Mart, 18, 2015).

[27] S. Tezgüler, "Sosyal Medya Nedir?", URL: sosyalmedyakulubu.com.tr/sosyalmedya/sosyal-medya-nedir.html (Erişim zamanı: Şubat, 3, 2015).
[28] S. Barutçu ve M. Tomaş, "Sürdürülebilir Sosyal Medya Pazarlaması Ve Sosyal Medya Pazarlaması Etkinliğinin Ölçümü" Journal of Internet Applications & Management/İnternet Uygulamaları ve Yönetimi Dergisi, 4(1), (2013).
[29] S. Eyidilli, "Twitter'ın aylık aktif kullanıcı sayısı 316 milyona yükseldi", Webrazzi, URL: http://webrazzi.com/2015/07/29/twitterin-aylik-aktif-kullanici-sayisi-316-milyon (Erişim zamanı: Eylül, 4, 2015).
[30] B. Kaya, "Sosyal Ağ Modelleme Kullanılarak Tıbbi Veri Ağında Bağlantı Tahmini Yöntemleri", Doktora tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ, 2015.
[31] Ö. Özgöbek ve RC. Erdur, "Öneri Sistemleri ve Bir Uygulama Alanı Olarak Haber Öneri Sistemleri", Akademik Bilişim Konferansları, Eskişehir, 31 Ocak-6 Şubat 2015.
[32] H. Jafarkarimi, ATH. Sim, R. Saadatdoost, "A naive recommendation model for large databases" International Journal of Information and Education Technology, vol. 2, no. 3, pp. 216, (2012).
[33] ZemberekNlp, "Zemberek Nasıl Çalışır?", URL: http://zembereknlp.blogspot.com.tr/2007_04_01_archive.html (Erişim zamanı: Nisan, 27, 2014).
[34] R. Usta, "Naïve Bayes Sınıflandırma Algoritması", URL: http://kodcu.com/2014/05/naive-bayes-siniflandirma-algoritmasi (Erişim zamanı: Haziran, 24, 2015).