



**SOSYAL AĞLARDAKİ HABER PAYLAŞIMI İÇERİKLERİNİN
SINIFLANDIRILMASI**

M. Sedef GÜNDÜZ

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

AĞUSTOS 2014

Sedef GÜNDÜZ tarafından hazırlanan “Sosyal Ağlardaki Haber Paylaşımı İçeriklerinin Sınıflandırılması” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ / OY ÇOKLUĞU ile Gazi Üniversitesi Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. Şeref SAĞIROĞLU

Bilgisayar Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Başkan : Prof. Dr. Nazife BAYKAL

Bilişim Sistemleri, Orta Doğu Teknik Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Üye : Yrd. Doç. Dr. Mehmet DEMİRCİ

Bilgisayar Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Tez Savunma Tarihi:/...../.....

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

.....
Prof. Dr. Şeref SAĞIROĞLU
Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

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

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

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

M. Sedef GÜNDÜZ

29.08.2014

SOSYAL AĞLARDAKİ HABER PAYLAŞIMI İÇERİKLERİNİN SINIFLANDIRILMASI

(Yüksek Lisans Tezi)

M. Sedef GÜNDÜZ

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ağustos 2014

ÖZET

Günümüzde sosyal ağların özellikle haber kaynağı olarak kullanılmaya başlaması, internet kullanıcılarının bu platformlar üzerindeki büyük veri miktarı ve akışı içerisinde ihtiyaç duydukları haber bilgisine daha kolay ve daha hızlı erişim yapma ihtiyaçlarını gündeme getirmiştir. Bu tez çalışması kapsamında, haber kuruluşları ve gazeteciler tarafından sosyal ağlar aracılığıyla yapılan haber paylaşımlarını gerçek zamanlı bir şekilde toplayan ve bu paylaşımları içerdikleri haber bilgisinin türüne göre sınıflandıran bir yöntem ile otomatik bir sistem önerilmiştir. Ayrıca, Twitter üzerinden yapılan haber içerikli tweet paylaşımlarını gerçek zamanlı bir şekilde toplayarak sınıflandıran ve ait oldukları kategori başlıkları altında görsel bir web sitesi yazılımı üzerinden internet kullanıcılarına sunan bir yazılım platformu geliştirilmiştir. Böylelikle internet kullanıcıları, belirlenen sosyal ağ üzerinde profil hesabına sahip olmasalar dahi, o sosyal ağ üzerinden yayınlanan haber bilgisine istedikleri kategori başlıkları altından erişebileceklerdir. Bu çalışma kapsamında geliştirilen sistem, sosyal ağ üzerinden yapılan haber paylaşımlarının sınıflandırılması işlemini gerçekleştirebildiği gibi, bu platform üzerinden yayınlanan her türlü metin tabanlı paylaşım verisini arzu edilen kategori başlıkları altında sınıflandırabilecek şekilde tasarlanmıştır. Bu çalışmanın, sosyal ağlar üzerinde yayınlanan haber içeriklerinin daha kolay, hızlı ve etkin bir şekilde takip edilmesini sağlayacağı özellikle haber ajansları tarafından etkin bir şekilde kullanılacağı değerlendirilmektedir.

Bilim Kodu : 902.1.014

Anahtar Kelimeler : Sosyal Ağ, haber, sınıflandırma, Twitter, platform, tweet, içerik, kategori, profil hesabı

Sayfa Adedi : 135

Danışman : Prof. Dr. Şeref SAĞIROĞLU

CLASSIFICATION OF CONTENT OF NEWS POSTS ON SOCIAL NETWORKS

(M. Sc. Thesis)

M. Sedef GUNDUZ

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

August 2014

ABSTRACT

Nowadays, social networks are started to be used in many areas, especially as a news source. On the other hand, internet users cannot access the information they need in an easier and fast way due to the huge amount of meaningless data that could not be perceived as news information in these platforms. In this sense, the need for accessing news information with high accessibility and less effort has risen. In this thesis, a method and an automatic system which collecting the news posts shared over social networks by news agencies and journalists, in real time; and also classifying these posts according to their news content are proposed. In addition, a software collecting and classifying the tweets that have news content over Twitter in real time, and presenting them under the most probable news category to the internet users is developed. In this way, internet users can access the news shared over social networks under the news category they desire to get information from, even they do not have profile account on that social network. The developed system is designed in a flexible way that could easily transform into a classifier that can assign any type of text-based social networking data into any desired category. Also, it is anticipated that this study will contribute on following news shared over social media in an easier, fast and more effective way by internet users, especially news agencies and journalists.

Science Code : 902.1.014

Key Words : Social network, news, classification, Twitter, platform, tweet, content, category, profile account

Page Number : 135

Supervisor : Prof. Dr. Şeref SAĞIROĞLU

TEŞEKKÜR

Yüksek Lisans eğitimim boyunca ilminden faydalandığım, hayata bakış açısı ve insani değerleri ile her zaman kendime örnek aldığım, değerli yardım ve katkılarıyla beni her zaman yönlendiren danışman hocam Sayın Prof. Dr. Şeref SAĞIROĞLU'na tüm kalbimle teşekkür ederim. Ayrıca değerli görüşleri ile çalışmamın gelecek hedeflerine katkılar sağlayan jüri üyeleri Sayın Prof. Dr. Nazife BAYKAL ve Sayın Yrd. Doç. Dr. Mehmet DEMİRCİ'ye ve bugünlere gelmemde büyük pay sahibi olan ve maddi manevi destekleriyle beni hiçbir zaman yalnız bırakmayan babam Atilla GÜNDÜZ'e, annem Mine GÜNDÜZ'e ve kardeşim Ahmet GÜNDÜZ'e teşekkürü bir borç bilirim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ.....	xi
RESİMLERİN LİSTESİ	xii
SİMGELER VE KISALTMALAR.....	xiii
1. GİRİŞ.....	1
2. SOSYAL AĞ KAVRAMI	5
2.1. Sosyal Ağların Sınıflandırılması	6
2.1.1. Akademik sosyal ağlar (social networks for academics/researchers)	6
2.1.2. Etnik ve bölgesel sosyal ağlar (ethnic/regional social networks)	6
2.1.3. Sosyal çevre edinme amacıyla kullanılan sosyal ağlar (general purpose social networks).....	7
2.1.4. İş ve kariyer amacıyla kullanılan sosyal ağlar (proffesional social networks)	7
2.1.5. Multimedya paylaşımı amacıyla kullanılan sosyal ağlar (multimedia sharing social networks)	7
2.1.6 . Evlilik ve flört amacıyla kullanılan sosyal ağlar (online dating social networks)	8
2.1.7. Eğitim amaçlı sosyal ağlar (educational social networks)	8
2.1.8. Konum tabanlı sosyal ağlar (location based social networks).....	8
2.1.9. Askeri sosyal ağlar (military social networks)	8

Sayfa

2.1.10. Dini sosyal ağlar (religious social networks)	9
2.1.11. Tıbbi sosyal ağlar (medical social networks)	9
2.1.12. Engelli bireyler için sosyal ağlar (social networks for people with a disability)	9
2.1.13. Seyahat amaçlı sosyal ağlar (social networks for travel)	10
2.1.14. Dil eğitimi amaçlı sosyal ağlar (social networks for language learning)	10
2.1.15. Aile ve akrabalara özel sosyal ağlar (family oriented social networks)	10
2.1.16. Hayır amaçlı sosyal ağlar (social networks for charity)	10
2.1.17. Şirket içi sosyal ağlar (social networks for office colleagues)	11
2.1.18. Mesleki sosyal ağlar (vocational social networks)	11
2.1.19. Bloglar (blogs)	11
2.1.20. Forumlar (forums)	12
2.2. Sosyal Ağlardan Veri Elde Etme Yöntemleri	12
2.2.1. Uygulama programlama ara yüzleri (application programming interfaces)	12
2.2.2. Veri kaynağı şirketleri	14
2.2.3. Anketler	14
2.2.4. Arama Botları	15
2.3. Sosyal Ağ Verilerinin Kullanım Alanları ve Gerçekleştirilen Projeler	16
2.3.1. Marka satış ve pazarlama stratejilerinin belirlenmesi	16
2.3.2. Devlet stratejilerinin belirlenmesi ve yürütülmesi	18
2.3.3. Ülkelerin tartıştığı konuların tespiti	21
2.3.4. Kullanıcı profillerinin çıkarılması ve kişilik tahmini	23
2.3.5. Tavsiye sistemlerinin geliştirilmesi	25
2.3.6. Suçların çözümü ve aydınlatılması	26

2.3.7. Eğitim alanında değerlendirmelerin yapılması.....	29
2.3.8. Duygu ve düşünce analizi.....	30
2.3.9. Kişiyeye uygun reklamların üretilmesi.....	32
2.3.10. Çıkarım yapma	34
3. TWITTER	37
3.1. Twitter kullanıcı kitlesi.....	42
3.2. Twitter’a ait temel kullanım özellikleri.....	43
3.3. Haber Kaynağı olarak Twitter	45
4. LİTERATÜR TARAMASI.....	51
5. SOSYAL AĞLARDAKİ HABER PAYLAŞIMLARININ SINIFLANDIRILMASI.....	59
5.1. Önerilen Yaklaşım.....	59
5.1.1. Geliştirilen modüller	62
5.2. Geliştirilen Web Sitesi Yazılımı.....	88
5.2.1 Geliştirilen modüller	89
5.2.2 Çevrimiçi performans testi	99
6. SONUÇ VE ÖNERİLER	103
EKLER	125
KAYNAKLAR.....	109

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Blog ve mikroblogların özelliklerinin karşılaştırılması.....	40
Çizelge 3.2. Mikroblog sitelerine ait özelliklerin karşılaştırılması	40
Çizelge 5.1. Modül #1'in işlevi sonucunda toplanan Twitter verisi.....	66
Çizelge 5.2. Veri toplama işlemi sonucunda her bir kategoriden alınan toplam tweet sayısı	67
Çizelge 5.3. Veritabanında her bir kategoriye ait son tweet sayısı.....	68
Çizelge 5.4. Zemberek kütüphanesinde tanımlanmış durak kelime listesinden bir kesit	74
Çizelge 5.5. Her bir kategoriye ait toplam kelime sayısı	75
Çizelge 5.6. Kelime köklerinin kaydedildiği veri tabanı tablosundan bir kesit.....	78
Çizelge 5.7. Ağırlıklandırılmış örnek kelime grubu.....	81
Çizelge 5.8. Sistemin test edilmesi sonucu elde edilen sınıflandırma başarımları yüzdeleri.....	87
Çizelge 5.9. Gerçek-zamanlı Twitter paylaşımları sistem tarafından dinlenen resmi haber kuruluşlarına ait Twitter hesap bilgileri	91
Çizelge 5.10. Belirlenmiş hesaplara ait gerçek zamanlı bir şekilde toplanmış olan tweet bilgileri	92
Çizelge 5.11. Kategorisi belirlenmiş tweetler.....	93
Çizelge 5.12. Geliştirilen yazılıma ait çevrimiçi performans testi sonuçları	100

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 5.1. Önerilen yaklaşıma ait akış diyagramı	60
Şekil 5.2. Modül 1'e ait akış diyagramı	64
Şekil 5.3. Modül #2'ye ait akış diyagramı	69
Şekil 5.4. Kırmızı ve mavi renkli objelerden oluşan sınıfların gösterimi	84
Şekil 5.5. Güven aralığı analiz sonucu.....	88
Şekil 5.6. Geliştirilen yazılıma ait akış diyagramı	89
Şekil 5.7. Performans testi sonuçlarının grafiksel gösterimi	100

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 3.1. Twitter’a ait güncel kullanıcı ara yüzü	42
Resim 3.2. NASA’nın resmi Twitter hesabı üzerinden yayınladığı haber tweeti	47
Resim 5.1. Twitter üzerinden yapılan bir paylaşımın içerdiği bağlantı	71
Resim 5.2. Örnek Zemberek uygulaması	76
Resim 5.3. tweetajans.com’un ana sayfasına ait bir ekran görüntüsü	94
Resim 5.4. Spor kategorisine ait Twitter haberleri	96
Resim 5.5. Sağlık ve Kültür-Sanat kategorilerinin çoklu seçimi.....	98

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
DVM	Destek Vektör Makineleri
SA	Sosyal Ağ
TDF	Ters Doküman Frekansı
TF	Terim Frekansı
TF-TDF	Terim Frekansı-Ters Doküman Frekansı
UPA	Uygulama Programlama Arayüzü

1. GİRİŞ

İnternet, birçok amaçla kullanıldığı gibi şüphesiz sosyal ilişkilerimiz için de bir dönüm noktası olmuştur. İnsanlar internet ağını bilgi edinme ve araştırma amaçlı kullanabildikleri gibi arkadaşlarıyla iletişim kurma, hızlı bilgi paylaşımı yapma ve yeni arkadaşlar edinme gibi değişken amaçlarla da kullanabilmektedir. Bu istekler ve gereksinimler dahilinde “insanların kendilerini tanımlayarak oluşturdukları profil bilgilerinin internet ortamında paylaşılmasına imkan sağlayan, bireylerin çevrimiçi olarak birbirlerini tanımalarına ve sosyal ilişkiler kurmalarına izin veren ve aynı zamanda sisteme kayıtlı bulunan kullanıcıların birbirlerinden haberdar olmasını sağlayan sanal topluluklar” olarak tanımlanan, Sosyal Ağlar kavramı ortaya çıkmıştır [2]. Günümüz dünyasında sosyal ağlar, internet kullanıcıları tarafından iletişim amacıyla kullanılan en son ve en yeni trend olarak değerlendirilmektedir. Fakat sosyal ağlar, internet kullanıcıları tarafından yalnızca kişisel iletişim kurma ve sosyal çevre edinme amacıyla değil, bunların yanı sıra ülke ve dünya gündeminden haberdar olma ve güncel haberleri takip etme amacıyla da kullanılmaktadır. Sosyal ağların, günümüzde özellikle haber kaynağı olarak kullanılmasını tetikleyen birçok özelliği mevcuttur. Bu çerçevede,

- Sosyal ağlarda haberlerin geleneksel medyaya nispeten daha hızlı yayılıyor olması,
 - İnternet kullanıcılarının ülke ve dünya gündeminden daha hızlı haberdar olmasını sağlaması,
 - Birtakım haberlerin ilk olarak sosyal ağlardan duyurulması,
 - Sosyal ağların resmi haber kuruluşları tarafından kabul görmüş olması,
 - Gazeteciler tarafından kullanılıyor olması,
 - Geleneksel medyanın aksine sosyal ağlardan yayınlanan haberlerin doğrulanma zorunluluğu olmadığından haber kuruluşları ve gazetecilerin haberleri ilk olarak sosyal ağlardan yayınlama eğiliminde olmaları ve
 - Son olarak sosyal ağların çift yönlü bir iletişim kanalı olmasından dolayı kullanıcılara ülke ve dünya gündemini daha aktif bir şekilde takip etme imkânı sunuyor olması
- bu platformların birer haber kaynağı olarak değerlendirilmesinde büyük rol oynamaktadır [100, 108, 116-135].

Sosyal ağların sahip olduğu bu özelliklerden yola çıkılarak yapılan bir çalışmada elde edilen sonuçlar, internet kullanıcılarının gündemden haberdar olmak amacıyla haber kuruluşlarının web siteleri yerine sosyal ağ hesaplarını takip ettiklerini göstermektedir [124]. Ayrıca sosyal ağların kullanıcı kitlesinin çok geniş olmasından dolayı bu platformlar üzerinden yayınlanan haberleri kullanıcıların güvenilir bulduğu ve geleneksel medya tarafından doğrulanmadan da bu haberlerin kullanıcıları ikna ettiği tespit edilmiştir [120].

Bunun yanı sıra sosyal ağlar üzerinden çok farklı kullanıcı kitleleri pek çok alanda haber paylaşımı yapabilmektedir. Bu durum da sosyal ağlar üzerinde büyük veri boyutlu verilerin bulunmasına veya toplanmasına sebebiyet vermektedir. Bu platformları kullanan kişilerin ihtiyaç duydukları bilgiye kısa sürede erişim sağlayabilmeleri ve yalnızca bilgi sahibi olmak istedikleri alanda haber bilgisine erişmek üzere sosyal ağ verisini filtreleyebilmeleri ise bu alanda yaşanan en büyük zorluk olarak değerlendirilmektedir. Ayrıca, kişilerin ihtiyaç duydukları haber bilgisinin, sürekli ve dinamik veri akışı içerisinde eriyip gitmesi ve kullanıcının takip etmek istediği haberi büyük veri akışı içerisinde gözden kaçırmaması da bu çerçevede yaşanan problemler arasındadır [116, 120]. Bu nedenle haber bilgisine erişimin kolaylaşması ve hızlı erişim sağlanarak zamandan tasarruf edilmesi amacıyla; sosyal ağlarda mevcut bulunan büyük veri boyutundaki haber verisinin toplanması, işlenmesi, bilgiye dönüştürülmesi ve içeriğinin saptanması büyük önem arz etmektedir.

Bu tez çalışması kapsamında, haber kuruluşları ve gazeteciler tarafından yayınlanan haber içerikli metin tabanlı sosyal ağ paylaşımlarının izlenmesi ve sınıflandırılması işlemlerini kolaylaştırmak, iş ve işlemleri hızlandırmak ve en önemlisi etkin kullanımı artırmak için yeni bir yöntem önerilmiş ve buna bağlı olarak bu işlemleri otomatik yapan bir sistem geliştirilmiştir. Önerilen yöntem ile herhangi bir sosyal ağa ait metin tabanlı haber paylaşımı verisi, kabul görmüş haber kategorisi başlıkları altında sınıflandırılarak kullanıcıya görsel bir web sitesi yazılımı üzerinden sunulmaktadır. Böylelikle sosyal ağlar üzerinden yayınlanan haberleri takip etmek isteyen internet kullanıcıları, sosyal ağ hesabına sahip olmasalar dahi, arzu ettikleri haber sınıfına ait sosyal ağ haberlerine daha kısa sürede ve daha kolay bir şekilde erişim sağlayabileceklerdir. Geliştirilen sistemin çalışma ve başarımlarını göstermek amacıyla, günümüzde en sık kullanılan sosyal ağlardan biri olan Twitter üzerinden haber kuruluşları ve gazeteciler tarafından yapılan haber paylaşımlarının 7 farklı haber kategorisi altında sınıflandırılması işlemi gerçekleştirilmiştir. Geliştirilen sistem, sosyal ağ verisinin sınıflandırılması işlemini farklı

amaçlarla farklı alanlarda gerçekleştirebilecek esnek bir yapıda tasarlanmıştır. Dolayısıyla geliştirilen sistem ile yalnızca haber içerikli sosyal ağ paylaşımları değil, herhangi bir sosyal ağa ait metin tabanlı paylaşım verisi, arzu edilen kategori başlıkları altında sınıflandırılarak kullanıcıya görsel bir web sitesi yazılımı üzerinden sunulabilmektedir.

Bu tez çalışması 6 bölümden oluşmaktadır. 2. Bölümde sosyal ağ kavramı; farklı sosyal ağ türleri bakımından ele alınarak literatürde mevcut bulunan sosyal ağ türleri genişletilmiş ve ilk kez 20 farklı başlık altında gruplandırılarak tanımlanmış ve örneklendirilmiştir. Bunun yanı sıra sosyal ağlardan veri elde etme yöntemleri ve sosyal ağ verisinin kullanım alanları; her bir alanda gerçekleştirilen akademik çalışma ve sektörel projeler ışığında detaylı bir şekilde incelenerek ilk kez 10 farklı başlık altında ele alınmıştır. 3. Bölümde bu çalışmada kullanılan Twitter sosyal ağı; sahip olduğu kullanıcı kitlesi, temel kullanım özellikleri ve haber kaynağı değeri bakımından bu alanda yapılmış olan literatür çalışmaları ve ülkemizde ve dünyada yaşanmış güncel örnekler kapsamında değerlendirilmiştir. 4. Bölümde sosyal ağ üzerinden yayınlanan haber ve farklı içerikteki metin paylaşımlarının sınıflandırılması üzerine yapılmış olan çalışmalar ele alınmıştır. 5. Bölümde ise bu tez çalışması kapsamında sosyal ağlar üzerinden yapılan haber içerikli paylaşımların sınıflandırılması amacıyla bir yaklaşım önerilmiş, önerilen yaklaşımda kullanılan Zemberek kütüphanesine katkılar sağlanmış ve böylelikle literatürde bu alanda önerilmiş yaklaşımlar geliştirilmiştir. Ayrıca önerilen yaklaşımı kullanarak Twitter üzerinden yapılan haber içerikli metin tabanlı paylaşımları sınıflandıran bir yazılım platformu geliştirilmiştir. Yazılımın başarısı test verisiyle yapılmış olan çevrimdışı test ve gerçek zamanlı veriyle yapılmış olan çevrimiçi performans testi olmak üzere 2 farklı şekilde test edilmiştir. Son olarak 6. Bölümde, sonuç ve değerlendirmeler tartışılarak gelecek çalışmalara konu olabilecek öneriler sunulmuştur.

2. SOSYAL AĞ KAVRAMI

Uzak mesafelerdeki aile bireyleri ve arkadaşlarla iletişim kurmak, insanlar için yüzyıllardır üzerinde düşünülen, çözüm aranan ve pek çok çözüm geliştirilmiş bir konudur. Yüz yüze görüşmelerin mümkün olmadığı durumlarda birçok yaratıcı çözüm üretilerek uzak mesafelerdeki kişiler ile olan ilişkiler korunmaya ve güçlendirilmeye çalışmıştır. İnternet, birçok amaçla kullanıldığı gibi şüphesiz sosyal ilişkilerimiz için de bir dönüm noktası olmuştur. İnsanlar internet ağını bilgi edinme ve araştırma amaçlı kullanabildikleri gibi arkadaşlarıyla iletişim kurmak, hızlı bilgi paylaşımı yapmak ve yeni arkadaşlar edinmek gibi değişken amaçlarla da kullanabilmektedir. Bu istekler ve gereksinimler dâhilinde “insanların kendilerini tanımlayarak oluşturdukları profil bilgilerinin internet ortamında paylaşılmasına imkân sağlayan, bireylerin çevrimiçi olarak birbirlerini tanımalarına ve sosyal ilişkiler kurmalarına izin veren ve aynı zamanda sisteme kayıtlı bulunan tüm kullanıcıların birbirlerinden haberdar olmasını sağlayan sanal topluluklar” olarak tanımlanan Sosyal Ağlar kavramı ortaya çıkmıştır [2].

İnternet ve sosyal ağ kullanımına ait günümüz verilerine bakılacak olursa, 2013 yılının ilk üç ayı itibarı ile (Ocak-Mart) Türkiye’de internet kullanımı % 47,4’ten % 48,9’a yükselmiştir [3]. İnternet kullanan bireylerin interneti % 75,6’lık bir oranla haber, gazete ya da dergi okumak için kullandıkları ifade edilmiştir. 2. Sırada ise %73,2 ile sosyal ağlara katılma yer almaktadır [3].

2013 yılı dünya genelindeki istatistiklere bakıldığında ise dünya nüfusunun % 85’inin internet erişimine sahip olduğu ve sosyal medya kullanımının % 18’lik bir artışla %25’lere ulaştığı görülmektedir [4]. 2012 yılı itibarıyla dünya çapında 1,47 milyar kullanıcıya sahip olan sosyal ağlar için bu rakam 1,73 milyara yükselmiştir. 2017 yılında ise bu rakamın 2,55 milyara ulaşacağı öngörülmektedir. Bu durum, geçmişten günümüze sosyal ağ kullanımının ülkemizde ve dünyada internet kullanımına paralel olarak arttığını ve artmaya devam edeceğini gözler önüne sermektedir [5].

Sosyal ağ kullanımı internet kullanımına paralel olarak arttıkça bu ortamlarda paylaşılan bilgi miktarı da artmakta ve paylaşılacak istenen bilgi türünde farklılıklar görülebilmektedir. Dolayısıyla sosyal ağ yapıları da kendi içlerinde çeşitlenmeye ve farklılık göstermeye ihtiyaç duymuşlardır. Bu kapsamda farklı bilgi türüne, bilginin

paylaşılma yöntemine, hedef kitlesine ve odak noktasına göre birden fazla sosyal ağ türü ortaya çıkmıştır.

2.1. Sosyal Ağların Sınıflandırılması

1990'lı yıllardan itibaren sosyal ilişkilerimize çevrimiçi olarak yön veren sosyal ağlar genel olarak eski arkadaşları bulma, yeni arkadaşlıklar edinme ve hızlı bilgi paylaşımı yapma gibi temel amaçlara hizmet etmektedirler [2, 6, 7]. Sosyal ağlar, bu ortamlarda paylaşılan bilginin türüne, odak noktasına, paylaşılma yöntemine ve hedef kitlesine göre değişkenlik göstererek farklı başlıklar altında sınıflandırılmaktadırlar. Bu tez çalışması kapsamında sosyal ağlar ilk kez özelleştirilmiş ve kullanım amaçlarına göre tarafımızdan tanımlanıp örneklendirilerek 20 farklı başlık altında toplanmış ve aşağıdaki alt başlıklarda kısaca özetlenmiştir.

2.1.1. Akademik sosyal ağlar (social networks for academics/researchers)

Akademik alanda çalışan ve farklı derecelerde akademik kariyere sahip olan akademisyenlerin bilimsel anlamda çalışmalarını ve tecrübelerini paylaştıkları ve aynı ağda hesabı olan diğer akademisyenlerin çalışmalarını takip edebildikleri sosyal ağlara akademik sosyal ağlar adı verilmektedir. Academia.edu bu alanda hizmet vermek üzere 2008 yılında açılmış ve günümüz itibarıyla 3 milyonun üzerinde kullanıcıya sahip olan akademik bir sosyal ağıdır [8]. ResearchGate ve Mendeley de bu kapsamda verilebilecek örnekler arasında yer almaktadır.

2.1.2. Etnik ve bölgesel sosyal ağlar (ethnic/regional social networks)

Belirli etnik kökene sahip olan kişilerin üye olmasını teşvik ederek aynı bölgede yaşayan veya aynı ulusal kökene sahip olan kişilerin birbirleriyle iletişime geçmelerine olanak sağlayan yapılara Etnik ve Bölgesel Sosyal Ağlar adı verilmektedir. Örneğin AsianAvenue, Asyalı Amerikanların kullanımı için tasarlanmış bir Etnik Sosyal Ağ'dır. BlackPlanet, Biip.no, Cloob, CozyCot, Cyworld, Douban, Draugiem.lv, DXT.cn, IRC-Galleria, iWiW, Jiebang, Kaixin001, LaiBhaari, LiveJournal, Mixi, Nasza-klasa.pl, Nexopia, Odnoklassniki, Qzone, Sina Weibo, Skoob, Skyrock, Spaces, Trombi.com ve Tuenti, Zoo.gr bu kapsamda verilebilecek örnekler arasında yer almaktadır.

2.1.3. Genel amaçlı sosyal ağlar (general purpose social networks)

Sosyal çevre edinme amacıyla kullanılan sosyal ağlar belirli bir amaca hizmet etmekten ziyade, kişilerin genel olarak çevrelerini genişletmek amacıyla yapabilecekleri bütün etkinliklere olanak sağlayan yapılardır. Genel amaçlı sosyal ağlar olarak da adlandırılan bu platformlar, kullanıcılarına yazı, resim, fotoğraf, video veya müzik gibi içerik paylaşma, paylaşımlara yorum yapma, gruplara üye olarak bilgi paylaşımında bulunma gibi birçok fonksiyonu bir arada sunmaktadır. Facebook, Twitter, Google+ ve Myspace kullanıcı sayıları ve popülerlikleri bakımından bu tür sosyal ağların başında gelmektedir. Bebo, Bolt.com, Cloob, Cyworld, Draugiem.lv, Faceparty, Friendster, Frühstückstreff, hi5, Hyves, Jaiku, Kiwibox, Makeoutclub, MEETin, Meetup, MillatFacebook, Netlog, Orkut, Quechup, Tagged, Tribe.net gibi sosyal ağlar bu kategoride ele alınabilecek örnekler arasındadır.

2.1.4. İş ve kariyer amacıyla kullanılan sosyal ağlar (proffesional social networks)

Kişilerin çevrimiçi profesyonel kimliklerini oluşturmalarına imkân tanıyarak iş ve okul arkadaşları ile bağlantılar kurmalarına olanak sağlayan sosyal ağlar iş ve kariyer amacıyla kullanılan sosyal ağlar kategorisinde incelenmektedir. Kullanıcılarının profesyonel fırsatları, iş anlaşmalarını ve yeni girişim olanaklarını keşfetmelerine imkân sağlayan bu yapılar aynı meslek alanında çalışan kişilerin birbirleriyle fikir alışverişinde bulunmalarını da kolaylaştırmaktadır [9, 10]. Dünyanın en geniş profesyonel ağı olarak bilinen LinkedIn bu kapsamda ortaya çıkmış olan ilk sosyal ağıdır. Elixio, Focus.com, Meettheboss, Ryze, Talkbiznow, XING bu kategoride yer alan sosyal ağlardan birkaçıdır.

2.1.5. Multimedya paylaşımı amacıyla kullanılan sosyal ağlar (multimedia sharing social networks)

Kullanıcılarının fotoğraf, video, ses ve animasyon gibi multimedya dosyalarını paylaşımlarına olanak sağlayarak aynı ağdaki kişilerin bu içerikleri izleyip hakkında yorum yapmalarına imkân tanıyan platformlar multimedya paylaşımı amacıyla kullanılan sosyal ağlar olarak tanımlanmaktadır. Bu kategorinin başında ise sahip olduğu kullanıcı sayısı itibarıyla Youtube gelmektedir. DailyBooth, Fotki, Instagram, Vine, Vimeo, İzlesene.com ve Picasa bu kategoride verilebilecek örnekler arasında yer almaktadır.

2.1.6. Evlilik ve flört amacıyla kullanılan sosyal ağlar (online dating social networks)

Kişilerin kendilerini tanımlayarak oluşturdukları profil bilgileri dâhilinde ruh ikizlerini bulmalarını sağlayan, kendileriyle aynı görüş ve düşüncelere sahip insanlarla tanışarak duygusal ilişkiler kurmalarını kolaylaştıran ve bunu çevrimiçi ortama taşımalarına imkân tanıyan sosyal ağlar bu kategoride ele alınmaktadır. AnastasiaDate, Badoo, BeautifulPeople.com, Compatible Partners, DateMySchool ve GayRomeo gibi sosyal ağlar bu kategoriye örnek olabilecek yapıya sahiptir [11].

2.1.7. Eğitim amaçlı sosyal ağlar (educational social networks)

Öğrencilerin okul ödevleri ve akademik projeleri konusunda birbirleriyle iletişime geçmelerine olanak tanıyarak aynı zamanda öğretmenleriyle de fikir alışverişi yapabilecekleri çevrimiçi platformlar eğitim amaçlı sosyal ağlar olarak ifade edilmektedir. Günümüzde bu yapılar kullanıcı sayılarını artırarak git gide popülerleşmektedir [12]. The Student Room, The Math Forum, ePALS School Blog, StudiVZ ve TeachStreet bu kategorideki sosyal ağlara verilebilecek örnekler arasındadır [12].

2.1.8. Konum tabanlı sosyal ağlar (location based social networks)

Son 5 yılda gelişimine hız verilerek kullanıcıların bulundukları konum bilgilerini, gerçek zamanlı olarak internet üzerinden bildirmelerine olanak sağlayan sosyal ağlara “konum tabanlı sosyal ağlar” adı verilmiştir. Bu platformlar kişilerin fiziksel dünyalarında bulundukları yerleri, sanal ortama taşımalarına imkân tanıyarak bahsedilen ortamlar arasındaki boşluğu kapatma amacıyla ortaya çıkmışlardır [13, 14]. Foursquare, Yelp, Facebook Places, Google Latitude, FullCircle ve Jiebang bu kategoride sayılabilecek örnekler arasındadır.

2.1.9. Askeri sosyal ağlar (military social networks)

Askeri personelin diğer askeri personellerle iletişim kurmasını amaçlayarak bu iletişimin çevrimiçi ortama taşınmasını hedefleyen yapılara askeri sosyal ağlar adı verilmektedir. Askeri sosyal ağlar, diğer sosyal ağlarda olduğu gibi eski arkadaşları bulma, yeni arkadaşlıklar edinme, fotoğraf paylaşımı yapma gibi olanaklar da sunmaktadır. Service-

Connected, MilitarySpot, ve TroopSpace bu kategoride örnek olarak gösterilebilecek sosyal ağlar arasındadır [15].

2.1.10. Dini sosyal ağlar (religious social networks)

Dini sosyal ağlar; aynı dine mensup olan kişilerin inanışları hakkındaki bilgilerini, görüşlerini ve fikir ayrılıklarını ifade edebilecekleri, dini konular hakkında aynı ağda bulunan arkadaşlarıyla resim, yazı, fotoğraf ve video benzeri içerikleri paylaşabilecekleri ve birlikte daha birçok aktivite yapabilecekleri çevrimiçi platformlar olarak tanımlanmaktadır. Aynı zamanda, bu yapıların diğer sosyal ağlara göre kişilerin yardımlaşma duygularını daha geliştirdiği ve kullanıcılarını olumlu yönde etkilediği var olan görüşler arasındadır [16]. Cross.tv, iChristianLife, Northland Church, HisHolyPlace, HolyPal, Xt3, Ummaland ve Muxlim bu kapsamda örnek gösterilebilecek sosyal ağlar arasındadır [17].

2.1.11. Tıbbi sosyal ağlar (medical social networks)

Tıbbi sosyal ağlar, doktorların tıp alanındaki bilgi ve birikimlerini birbirleriyle ve hastalarıyla paylaşmalarına olanak sağlayan ve aynı zamanda hastaların da birbirleriyle iletişim kurmalarını sağlayarak hastalıkları ve doktorları hakkındaki görüşlerini ve yaşadıklarını birbirleriyle paylaşmalarını kolaylaştıran çevrimiçi platformlardır. Bu platformlar sayesinde kişilerin hastalıklar hakkında pek çok şeyi bildikleri, daha bilinçli oldukları ve para ve zamandan tasarruf ederek karşılaşılan olumsuzlukları kısa sürede giderdikleri gözlemlenen sonuçlar arasındadır. PatientsLikeMe, DailyStrength, DoctorsHangout.com, doc2doc ve doktorsitesi.com bu kategoride ele alınabilecek sosyal ağlar arasındadır.

2.1.12. Engelli bireyler için sosyal ağlar (social networks for people with a disability)

Engelli bireyler için özel olarak oluşturulmuş bu tür sosyal ağlar, engelli bireylerin yaşadıkları zorlukları ve tecrübelerini birbirlerine aktarabilecekleri ve diğer sosyal ağlarda olduğu gibi problemleri hakkında multimedya benzeri içerikleri paylaşabilecekleri platformlardır. Disaboom, Disabilinet.com ve Disability Social Network bu kategoride verilebilecek örnekler arasındadır.

2.1.13. Seyahat amaçlı sosyal ağlar (social networks for travel)

Seyahat amaçlı sosyal ağlar, kişilerin ziyaret etmek istedikleri mekânlar ve yerler hakkında bilgi edinerek tavsiye alabilecekleri, ziyaret edilen yerler hakkında farklı içeriklerde paylaşım yapabilecekleri çevrimiçi platformlardır. Bu ağlara üye olan kullanıcılar birbirleriyle ve aynı zamanda seyahat acenteleriyle iletişime geçebilmektedirler. CouchSurfing, Exploroo, Gapyear.com, TravBuddy.com, Travellerspoint ve WAYN bu kategorideki sosyal ağlara örnek gösterilebilmektedir [18].

2.1.14. Dil eğitimi amaçlı sosyal ağlar (social networks for language learning)

Dil eğitimi amaçlı sosyal ağlar, kişilerin yabancı dil yeteneklerini geliştirmek için dünyanın birçok yerinden farklı kişilerle çevrimiçi pratik yapma olanağına sahip oldukları çevrimiçi platformlardır. Birçok dil desteği sunan bu platformlar arasında kullanıcılarına akıcı konuşma garantisi veren sosyal ağlar da mevcuttur. Busuu, italki.com, Livemocha ve TermWiki bu kategoride örnek olarak gösterilebilecek olan sosyal ağlar arasındadır [19].

2.1.15. Aile ve akrabalara özel sosyal ağlar (family oriented social networks)

Aileye özel sosyal ağlar, kişilerin aile bireyleriyle iletişime geçmek amacıyla aile bireyleri arasında ağ kurabilecekleri, soy ağaçlarını belirli bir seviyeden oluşturmaya başlayarak tanımadıkları aile bireyleriyle iletişime geçebilecekleri ve ailelerinin geçmişe dair hikâyelerini ve anılarını paylaşabilecekleri çevrimiçi platformlardır. Ayrıca diğer sosyal ağlarda olduğu gibi, kullanıcılarına aile fotoğraflarının paylaşılması ve özel günlerin etiketlenmesi gibi özellikler de sunmaktadır. MyHeritage, bu tür ağlar arasından en popüler olanıdır. Famiva, Piqniq, Family Crossings, Geni.com, Commily, MyFamily.com ve Family.com bu kategoride gösterilebilecek diğer örnekler arasındadır [20].

2.1.16. Hayır amaçlı sosyal ağlar (social networks for charity)

Kullanıcıların kredi kartları ile diledikleri organizasyona bağışta bulunabilecekleri ve aynı ağda kayıtlı olan kişilerle iletişime geçerek yapılan bağışlar ve yardımlar hakkında bilgi sahibi olarak farkındalık yaratabilecekleri sosyal ağlar hayır amaçlı sosyal ağlar olarak

nitelendirilmektedir. Başta HuTerra olmak üzere Crowdrise, Jumo ve Causecast gibi sosyal ağlar bu kategoride verilebilecek örnekler arasındadır.

2.1.17. Şirket içi sosyal ağlar (social networks for office colleagues)

Ofis içi kullanıma uygun olarak geliştirilmiş, şirket çalışanlarının ekip olarak yapmaları gereken işlerde ekip üyelerini takip etmelerine imkân tanıyan, grup oluşturma özelliği ile kullanıcılarına proje sürecini her an her yerde takip etme olanağı sunarak birlikte çalışmayı kolaylaştıran ve ofis içi iletişimi çevrimiçi ortama taşıyan sosyal ağlara şirket içi sosyal ağlar denilmektedir [21]. Yammer, Socialcast, Socialtext, Tibbr, Jive, Convo, Kaltura, Chatter, Zyncro ve Connections şirket içi sosyal ağların en çok bilinen örnekleri arasında yer almaktadır [21].

2.1.18. Mesleki sosyal ağlar (vocational social networks)

Yalnızca aynı meslek grubuna ait kişilerin üye olduğu ve birbirleriyle mesleki bilgi ve tecrübelerini paylaşarak meslekleri hakkında güncel gelişmeleri internet ortamından takip edebildikleri sosyal ağlara mesleki sosyal ağlar adı verilmektedir. Örneğin, ressam.com ressamların resimlerini paylaşarak diğer kullanıcılarla iletişim kurdukları ve aynı zamanda yapılan yorum ve oylamalar ile gelecekteki eserlerine yön verebildikleri bir mesleki sosyal ağıdır [22]. Virb, Taltopia, Fandalism, Soundirect ve Giggem gibi platformlar mesleki sosyal ağlara verilebilecek diğer örneklerdir.

2.1.19. Bloglar (blogs)

Çevrimiçi kişisel günlükler olarak da adlandırılan bloglar, kullanıcıların bir olay, ürün ya da ideoloji hakkındaki duygu ve düşüncelerini periyodik olarak paylaşabildikleri bir sosyal ağ türüdür. Ayrıca, blogu takip eden kişilerin de yapılan paylaşımlar hakkında yorum yaparak fikirlerini beyan edebildikleri bir platformdur [23]. Blogster, Fuelmyblog, LiveJournal, My Opera, Open Diary, Vox, Wooxie, Xanga, Blogger ve Blogcubu kapsamda verilebilecek örnekler arasında yer almaktadır.

2.1.20 Forumlar (forums)

Forumlar kişilerin belirli bir konuda var olan soru ve sorunlarını paylaşarak diğer üyelerle çözüm aradıkları, birbirleriyle fikir alışverişinde bulundukları, bulunan çözümleri paylaştıkları, birbirlerine yardım ettikleri çevrimiçi platformlardır. Aynı zamanda bu tür sosyal ağlarda, var olan probleme çözüm bulunduktan sonra aynı problemi yaşayanlara rehber olması açısından yapılan yorumlar ve açılan konular kaldırılmamaktadır. Forum TR, ShiftDelete.net ve Donanım Haber Türkiye’de en çok kullanılan forum siteleri arasındadır. Bunun yanı sıra Gaia Online, Crunchyroll, StackOverflow ve SuperUser gibi forumlar da dünya çapında en çok kullanılan sosyal ağlar arasındadır.

2.2. Sosyal Ağlardan Veri Elde Etme Yöntemleri

Günümüzde kişilerin iletişim bilgileri, fotoğrafları, bulundukları konumlar, çalıştıkları yerleri, mezun oldukları okullar, arkadaşlık ilişkileri, sosyal yaşantıları, hobileri, siyasi ve dini görüşleri gibi birçok kişisel veri sosyal ağlar üzerinde toplanmaktadır. Dolayısıyla, insan hayatının her alanına dair yürütülmek istenilen ticari, sanatsal veya bilimsel her türlü çalışmada sosyal ağ verilerinin kullanılması oldukça akılcı bir yöntemdir. Bu kapsamda sosyal ağ verileri pek çok alanda farklı şekillerde yorumlanarak bilgiye dönüştürülmektedir. Fakat sosyal ağ verilerinin bilgiye dönüştürülebilmesi için öncelikle veri çıkarma veya toplama olarak da adlandırılabilen veri elde etme işleminin gerçekleştirilmesi gerekmektedir. Bu tez çalışması kapsamında günümüzde sosyal ağlardan veri elde etme amacıyla kullanılan yöntemler 4 başlık altında sınıflandırılarak, elde edilen bu verilerin kullanıldığı alanlar, takip eden alt başlıklarda detaylı bir şekilde incelenmiştir.

2.2.1. Uygulama programlama ara yüzleri (application programming interfaces)

Uygulama programlama ara yüzü (UPA), herhangi bir programın yürüttüğü belirli görevleri veya kullandığı verileri diğer uygulamalarında kullanabilmesi için geliştirilmiş ara yüz olarak adlandırılan bir modül olarak tanımlanmaktadır. Sosyal ağlardan veri toplamanın ve bu verileri kullanarak yeni uygulamalar geliştirmenin en kolay ve en güvenilir yollarından biri sosyal ağların sunduğu uygulama programlama ara yüzü hizmetleridir. Günümüzde neredeyse her sosyal ağın uygulama programlama ara yüzü bulunmaktadır. Bunlar arasından en çok bilinen ve uygulama programlama ara yüzü

bulunan sosyal ağlar Facebook, Twitter, Foursquare, Pinterest, Bebo, Brightkite, Change.org, Friendster, Google+, Gowalla, Hi5, LinkedIn, MySpace, NetLog, Orkut ve Yammer olarak sıralanabilir. Bu tez çalışması kapsamında geliştirilen sistem bu tür uygulama yazılımlarını kullanmaktadır. Bu sistem Twitter sosyal ağına ait verileri ve UPA'yı kullanarak geliştirildiğinden bu bölümde Twitter'ın sağladığı uygulama programlama ara yüzleri daha detaylı bir şekilde ele alınacaktır. Twitter uygulama geliştiricilere Rest ve Streaming olmak üzere 2 farklı UPA sunmaktadır [24]. Bu UPA'lar;

- kişinin attığı tweetlerin geliştirilen uygulama tarafından geçmişe yönelik ve gerçek zamanlı olarak toplanabilmesine,
- çekilen her bir tweetin coğrafi konum bilgisine,
- atıldığı zamana,
- diline,
- tweet sahibinin Twitter üzerindeki kullanıcı adına,
- fotoğraflarına,
- arka plan resimlerine,
- ne zaman Twitter'a katıldığı bilgisine,
- belirli zaman aralıklarında atılan tweetlerin ayıklanmasına,
- hashtaglere
- tweetin atıldığı platforma,
- tweetin içerdiği bağlantılara,
- cevap yazılan kullanıcı bilgilerine,
- re-tweetlere ve
- tweetlere atılan cevaplara vb.

geliştirilen uygulama tarafından erişim izni sunmaktadır. Dolayısıyla sosyal ağ verilerine ihtiyaç duyan bir çalışmada, verisi toplanmak istenen sosyal ağın UPA'sını kullanarak ihtiyaç duyulan verileri çekmek bu kapsamda kullanılabilecek en kolay ve en mantıklı yöntemdir.

2.2.2. Veri kaynağı şirketleri

Sosyal ağlar üzerinde var olan veriyi elde etmenin bir diğer yolu ise, sosyal veri platformları olarak da bilinen veri kaynağı şirketleridir. Bu şirketler sosyal ağlarla anlaşmalı bir şekilde veriyi alıp satmaktadır. Bu şekilde sosyal medyayı takip etmek isteyen kurum ve kuruluşların plan ve karar verme süreçlerine yardımcı olarak sosyal veri kavramının önemini artırmaktadırlar. Bu tür şirketler sadece ham sosyal ağ verisini toplayarak bunu satmakla kalmayıp, aynı zamanda kategorize edilmiş veya filtrelenmiş veriyi satın almak isteyen kuruluşlara da sosyal ağ verilerini kategorize edip veya filtreleyip daha sonra satmaktadırlar [25]. Bu şirketler arasından en çok sesini duyurmuş olanlar ise GNIP ve DATASIFT'tir [25, 26].

Kendisini dünyanın en büyük ve en güvenilir sosyal veri kaynağı olarak tanımlayan GNIP, birçok sosyal ağın anlaşma yaptığı ilk şirkettir. Twitter'ın ilk veri ortağı olmakla birlikte bütün arşivlerine ve geçmiş verilerine sahiptir. Tumblr, WordPress, Foursquare, Disqus, IntenseDebate, StockTwits ve GetGlue sosyal ağları da ilk veri ortaklıklarını GNIP ile başlatmışlardır [26]. Bunların yanı sıra, GNIP ile anlaşmalı olan ve bu şirkete verilerine tamamen erişim izni veren diğer sosyal ağlar etsimize ve newsgator platformlarıdır. Ayrıca GNIP; Bitly, Dailymotion, del.icio.us, Facebook, Flickr, Google+, Instagram, Metacafe, Panoramio, Photobucket, PLURK, Reddit, Stackoverflow, Vimeo ve Youtube sosyal ağ verilerine de kontrollü erişim sağlamaktadır [26].

DATASIFT ise günümüzde büyük verinin büyük bir kısmını oluşturan sosyal verinin önemine vurgu yaparak aynı amaçlarla hareket etmektedir. DATASIFT şirketinin Twitter, Facebook, Sina Weibo, WordPress, Intense Debate, Tumblr, Google+, Youtube, bitly, Instagram, NewsCred, Flickr, Reddit, Wikipedia, DailyMotion, Topix, IMDb, Videos, Blogs ve MessageBoards sosyal ağları ile anlaşması bulunmaktadır. Ayrıca, DELL, Dachis Group, shrethrough, Yum! ve local response gibi şirketler DATASIFT'in müşterileri arasında yer almaktadır [25].

2.2.3. Anketler

Sosyal ağ verilerini eğitim, araştırma veya şirket stratejilerini belirleme gibi alanlarda kullanmak üzere elde etmek için kullanılan bir diğer yöntem de bu platformlar üzerinden

çevrimiçi anketler düzenlemektir. Sosyal ağ kullanıcılarına bu platformlar üzerinden yöneltilen anket soruları ile elde edilmek istenen bilgiler toplanmaktadır. Bu kapsamda, sosyal ağ kullanıcılarının ankete katılmasına teşvik amacıyla, katılan kişilere farklı ödüller sunularak sosyal ağ verileriyle anket düzenleme işlemi daha kolay ve ilgi çekici hale getirilebilmektedir.

İlgili kişilerin veya kurumların çevrimiçi anket düzenleyebilmeleri için çeşitli yazılımlar mevcuttur. Bu alanda, SurveyMonkey, e-Research-Global, Anketix ve Surveygizmo gibi yazılımlar örnek olarak verilebilmektedir. Fakat bunlar arasından SurveyMonkey, sosyal ağlarla entegre bir şekilde çalışabilmekte ve düzenlenmek istenen anketlerin sosyal ağlar üzerinden yapılabilmesine olanak sağlamaktadır. Bu kapsamda Surveymonkey kullanıcılarına çeşitli sosyal ağ anket şablonları sunmaktadır. Bilgi almak istenen kişilere ulaşmak için hangi sosyal ağların en uygun ve etkili olduğuna karar verildikten sonra her türlü amaç için kullanmak üzere SurveyMonkey'in anket yelpazesinden seçim yapılabilmektedir. SurveyMonkey anketleri doğrudan Facebook sosyal ağına yerleştirilebilmekle birlikte Twitter, LinkedIn, Foursquare, ve Google+ sosyal ağlarında bir bağlantı olarak da paylaşılabilmektedir. SurveyMonkey anket hizmetinin Facebook ile entegre bir şekilde kullanılmak üzere sunduğu anket hizmetleri birçok alanı kapsamaktadır. Bunlar çevrimiçi oylamalar, Facebook kullanımı, eğlence amaçlı anketler, etkinlik planlama, pazar araştırması, müşteri memnuniyeti, mobil anketler, akademik anketler, teknoloji anketleri ve seyahat ve konaklama anketleri olarak sıralanabilmektedir [27].

2.2.4. Arama Botları

Sosyal ağ arama botları, verisine ulaşılmak istenilen sosyal ağları tarayarak ilgili içerikleri otomatik alan yazılımlardır. Dolayısıyla sosyal ağlar için geliştirilmiş bir arama botunun görevi ziyaret edilen bütün hesaplarda herkese açık olarak paylaşılan verileri daha sonra daha kolay erişmek ve farklı amaçlarla kullanmak üzere toplama veya sadece belirli amaca uygun verileri toplama olarak ifade edilmektedir [28].

Günümüzde sosyal ağ verilerini toplamaya yarayan beevolve, 80legs, NodeXL ve prompt cloud gibi birçok arama botu bulunmaktadır. Bu arama botları şirketlerle yaptıkları anlaşma dâhilinde hangi sosyal ağ verisine ihtiyaç duyuluyorsa o platform üzerinde bulunan hesapların herkese açık olarak belirtilen bilgilerini toplamaktadır. Daha sonra bu

verilere ihtiyaç duyan kurum ve kuruluşlara istenilen formatta satarak gelir elde etmektedir.

Sosyal ağlar için geliştirilen arama botları, taramaya başlamadan önce hangi veriye ihtiyaç duyulduğunu ve ihtiyaç duyulan veriyi elde etmek için kullanılması en uygun tarama metodunun ne olduğunu tespit etmektedirler. Tarama işlemi sırasında ise işlemin ne zaman sonlanması gerektiğini ve tarama sonrasında hangi verilerin göz ardı edilmesi gerektiğinin tespiti de bu yazılımların görevleri arasında yer almaktadır [29].

2.3. Sosyal Ağ Verilerinin Kullanım Alanları ve Gerçekleştirilen Projeler

Sosyal veri olarak da ifade edilen sosyal ağ verileri günümüzde çok çeşitli alanlarda birçok amaçla kullanılmaktadır. Bu çalışma kapsamında sosyal ağ paylaşımlarının içerdiği sosyal veri kullanım alanları ve amaçlarına göre sınıflandırılarak gerçekleştirilen akademik çalışmalar ve sektör projeleri 10 farklı grup altında incelenmiştir. Ayrıca, her bir alanda sayıca en çok çalışmaya sahip olan kategori belirlenerek (akademik veya sektörel) bu kategorilerde yapılan çalışmalar referansları ile birlikte; kullandıkları sosyal ağ türü, sosyal ağ adı, veri türü, veri büyüklüğü ile veri toplama ve analiz araçları bakımından incelenmiştir ve EK-1 Çizelge 1.1.'de sunulmuştur. Literatürdeki makaleler gözden geçirildiğinde yapılan çalışmalar farklı açılardan değerlendirilmiş ve temel olarak 10 alana ayrılmıştır. Bu değerlendirmeler, günümüzde sosyal ağ verisinin en çok kullanıldığı alanlar temel alınarak yapılmıştır. Böylelikle sosyal ağ verisi ile yapılabilecek olan çalışmalara daha çok ihtiyaç duyulduğu belirlenerek, yeni ufukların açılması ve çalışma alanlarının belirlenmesine yardımcı olunması amaçlanmıştır.

2.3.1. Marka satış ve pazarlama stratejilerinin belirlenmesi

Sosyal ağ paylaşımlarından veri çıkarımı, markaların satış ve pazarlama kriterlerinin belirlenmesinde büyük rol oynamaktadır. Günümüz dünyasında, kişilerin kendi ürünleri ve servisleri hakkında ne düşündüklerini önemseyen markalar sosyal ağları takip etmek zorundadırlar. Çünkü sosyal ağ hesapları kişilerin herhangi bir konu hakkındaki fikir ve düşüncelerini özgürce beyan edebilecekleri ortamlar olarak algılanmaktadır [2]. Markaların sosyal ağ hesaplarına ait sayfaların gösterim sayısı, bu sayfalara kayıtlı üye sayısı, e-posta ve e-bülten okunma oranları ve takipçi ve beğeni sayıları, markaların satış ve pazarlama

stratejilerini belirlerken kullanabilecekleri sosyal ağ verileri arasında yer almaktadır [30]. Bu kapsamda sektörde, özellikle büyük şirketlerin kullandığı sosyal medya analizi ve ölçümlemesi yapan birçok yazılım aracı bulunmaktadır [30, 31, 32, 33]. Google Analytics, BoomSocial, AddServer, BrandWatch ve Dekatlon Buzz bu araçlara verilebilecek örneklerden sadece birkaçıdır. Örneğin Google Analytics hizmetleri ile markalar, isimlerinin Twitter’da kaç kez ve nasıl geçtiğini bilerek kampanya faaliyetlerine yön verebilmektedirler. Bu tür raporlama ve ölçüm araçlarını kullanarak yapılan kampanyaların performansını ölçmek veya sosyal ağlarda markayı ve kampanyaları diğer kullanıcılara tanıtan “alakalı etkileyicilerin” metriklerini çıkarmak markaların sosyal ağ paylaşımlarından veri çıkarımını uygulayabilecekleri alanlar arasındadır. Sosyal ağ verilerinden gerekli çıkarımlar yapıldıktan sonra elde edilen bilgiler ise markanın rakiplerine stratejik mesajlar iletmesinde veya yeni kampanyaların oluşturulmasında kullanılabilir.

Pazarlamanın kollarından biri olan ağızdan ağza yayılma (word of mouth marketing) stratejisi için en uygun ortamlardan biri sosyal ağlardır. Çünkü ağızdan ağza yayılma müşterileri marka hakkında olumlu yönde konuşturmak için ürünler ve servisler hakkında önerilen çözümler olarak tanımlanmaktadır ve günümüzde müşterilerin satın aldıkları ürünler hakkında en çok konuştukları platform sosyal ağlardır. Bu nedenle sosyal ağlardan veri çıkarımı, pazarlama stratejilerine yön vermek adına uygulanabilecek en akılcı yöntemler arasındadır [34, 35] .

Bunun yanı sıra sosyal ağlardaki veri, kullanıcıların niyetini anlayarak kriz yönetiminde proaktif olma amacıyla kullanılmaktadır [36, 37]. Markalar sosyal ağlardaki veriyi kullanarak analiz ve raporlama yazılımları ile ürünleri veya hizmetleri hakkındaki olumsuz paylaşımlardan zamanında haberdar olarak kriz durumunu doğru kampanyalar ile fırsata dönüştürebilmektedirler [36, 37].

Aynı şekilde markalar sosyal ağ paylaşımlarını dikkate alarak ürünleri neden satıyor ya da neden satamıyor gibi sorulara cevap bulabilmektedirler. Örneğin, bir markanın Twitter sosyal ağında paylaştığı satış kanalı linkinin o markanın çevrimiçi satışlarını ve satış hacmini ne kadar artırdığı bu kapsamda değerlendirilmesi gereken veriler arasındadır. Bunlara ek olarak, sosyal ağ verileri sayesinde markalar kalıcı müşteri sayılarını ve hangi müşterilerin markaya yeni müşteriler kazandırdığını da tespit edebilmektedirler [30, 31, 32, 33, 36, 37].

Sosyal ağlardaki verilerin kurum ve şirketler için satış ve pazarlama amacıyla bir diğer kullanım alanı da iş planlarının oluşturulmasıdır. İş planı aşamaları; amaçları belirleme, hedef kitlesine ulaşma, etki alanını belirleme, marka hakkındaki duygu ve düşünceleri saptama, harekete geçme, yapılan yorumları ve etkileri dinleme ve sonuçlara ulaşma olmak üzere 7 ayrı başlık altında ifade edilmektedir [38, 39]. Bu aşamaların her birinde de sosyal ağ paylaşımlarındaki mevcut verilere ihtiyaç duyulmaktadır.

Sonuç olarak sosyal ağ paylaşımlarından çıkarılan veriler, satış ve pazarlama alanında hedef kitlenin marka hakkındaki düşünce, niyet ve yorumlarını öğrenme, kriz oluşumuna engel olma veya krizlere anında müdahale etme, kurum itibarının ne durumda ve seviyede olduğunu tespit etme, müşterinin dikkatini çekebilecek etkili kampanyalar düzenleme ve müşteri profili çıkarma amaçlarıyla kullanılabilmektedir [30-39].

2.3.2. Devlet stratejilerinin belirlenmesi ve yürütülmesi

Devletin vatandaşlarına sunduğu kamusal hizmetlerin elektronik ortama taşınmasına olanak sağlayan yapılara e-devlet platformları denilmektedir [40]. E-devlet sistemleri ile kamu hizmetlerinin çevrimiçi, yüksek bilgi işleme kapasitesine sahip, kolay, hızlı, etkin, kaliteli ve kesintisiz bir şekilde gerçekleştirilmesi hedeflenmektedir [41]. Sosyal ağların kişilerin günlük yaşantısında bu denli etkili olduğu bir dijital çağda ise bireylere sundukları içerik yaratma, etkileşim kurma ve ortak hareket etme özellikleri, ülke vatandaşlarının bu yapıları e-devlet uygulamalarında da aramasına ve talep etmesine yol açmıştır [40]. Bu kapsamda, sosyal ağların e-devlet uygulamalarına iki farklı alanda katkı sağlayabileceği öngörülmektedir. Bunlardan birincisi; sosyal ağların e-devletin idari boyutu yönü ile kamusal bilgi ve hizmet sunumuna olan katkısı, ikincisi ise; sosyal ağların e-devletin siyasi boyutu yönü ile demokratik katılım, denetim, şeffaflık ve hesap verilebilirliğin artırılmasına olan katkısı olarak değerlendirilmektedir [42]. Bu sayede e-devlet kullanıcıları olan vatandaşlar pasif katılımdan aktif katılıma geçerek yönetime katılım ve yönetim ile işbirliği aşamalarına gelebileceklerdir.

Bu kapsamda, Avrupa CELTIC Araştırma ve Geliştirme Programı tarafından desteklenen ve Türkiye'yi TÜBİTAK desteği ile SAMPAS'ın temsil ettiği CRUMBS isimli projede sosyal ağlar ve e-devlet uygulamaları bir araya getirilmiş ve bu proje ile Sosyal Ağ Temelli Arttırılmış Gerçeklik ile Yeni e-Devlet ve Turizm Uygulamaları geliştirilmiştir [43].

Sosyal ağlara erişim için mobil cihazlarını kullanan kişi sayısının artmasından yola çıkılarak geliştirilen bu proje ile bir sosyal ağ üzerinden konum tabanlı kullanıcı verilerini toplama ve bu verileri turizm alanında kullanılmak üzere organize eden ve hizmet veren bir artırılmış gerçeklik platformu oluşturulmuştur. CRUMBS uygulaması ile mobil kullanıcı, herhangi bir koordinat üzerinde herhangi bir zaman diliminde oluşturulmuş sosyal ağ paylaşımı kırıntılarına (crumbs) erişerek bulunduğu yerde artırılmış gerçeklik ile etkileşim kurabilmekte ve geriye dönük içerik taraması yapabilmektedir. Aynı zamanda, kullanıcının mobil cihazını doğrulttuğu obje ile ilgili olan internetteki bütün veriler ve bu objenin 360 derece çevresinde oluşturulmuş tüm sosyal içerikler kullanıcıya sunulmaktadır [43].

Devlet stratejilerinin yürütülmesi adına sosyal ağ kullanımına önem verilmesine örnek olabilecek bir olayda Avustralya’da yaşanmıştır [44]. Avustralya hükümeti, devlet kurumları ile halk arasındaki iletişim kopukluğunu sosyal ağ kullanımı ile telafi etmeyi planlamaktadır. Devlet memurlarını Facebook, Myspace veya Twitter gibi sosyal ağları kullanmaları yönünde teşvik ederek, sosyal ağları devlet ve halk arasındaki ilişkileri geliştirme ve sağlamlaştırma adına uygun bir yöntem olarak gördüklerini ispat etmişlerdir. Bu kapsamda, devlet birimleri, sosyal ağlar üzerinden vatandaşları gelişmeler hakkında haberdar etmekte ve planlanan tasarıları yoruma açmaktadırlar. Vatandaşlar da devletin verdiği hizmetler hakkında geri bildirimler sunarak sosyal ağ kullanımı ile devlet stratejilerinin yürütülmesinde yardımcı olmaktadır [44].

Türkiye Cumhuriyeti Devleti’nin sosyal ağlar üzerinden belirlediği, yürüttüğü ve denetlediği devlet stratejileri arasında ülke içerisinde yaşanan karışıklıkları ve ayaklanmaları tespit etme ve önleme politikaları yer almaktadır. Bu kapsamda verilebilecek en güzel örnek henüz yaşanmış olan Gezi Parkı olaylarıdır. Gezi Parkı olaylarında sesini devlet yetkililerine duyurmak üzere bir grup ülke vatandaşı sosyal medya üzerinden gruplaşarak protestolarını gerçekleştirmişlerdir. Bu durum, günümüzde devlet yönetimi ile vatandaş arasında ülke konuları hakkında iletişim kurulabilecek en güçlü yolun sosyal ağlar olduğunu ispatlar niteliktedir. Türkiye Cumhuriyeti Devleti’nin bu kapsamda gerçekleştirdiği bir “Mavi Oda” projesi mevcuttur [45]. Takip edilecek olan sosyal ağların ortak rengi maviden adını alan bu özel birim, daha önce yayınlanan eylem planı doğrultusunda devletin siber saldırıları önlemek ve yalan bilgilerle toplumu yanlış yönlendirmekle suçladığı sosyal medya üzerinde bazı izleme ve denetim ekipleri oluşturma

hedefine sahiptir. Bu proje kapsamında, Facebook ve Twitter gibi sosyal ağlarda suç unsuru taşıyan mesajlar ve bunları paylaşan kişilerin ortaya çıkarıldığı bilinmektedir. Ayrıca aynı proje kapsamında Ankara Emniyet Müdürlüğü bünyesinde bulunan Siber Suçlarla Mücadele Daire Başkanlığı'nda yapılan sosyal medya takibi sonucunda Gezi Parkı protestoları boyunca devlet yetkililerine küfür veya hakaret içerdiği düşünülen 170 bin mesaj tespit edilmiştir [46].

Devlet stratejilerinin yürütülmesinde sosyal ağ kullanımını inceleyen akademik çalışmalarda literatürde mevcuttur. Bu kapsamda ülkemizde gerçekleştirilen bir çalışmada sosyal ağ kullanıcılarının bu platformlar üzerinden hızlı bir şekilde organize olabilme durumları dikkate alınmıştır. Twitter kullanıcılarının hesapları üzerinden yayınladıkları tweetler analiz edilerek herhangi bir durum veya olay karşısında o kişi için 'destekçi' veya 'destekçi değil' çıkarımı yapılmaktadır. Geliştirilen sistemin Türkiye'de yaşanan Gezi Parkı protestoları ile ilişkilendirilerek %90 oranında doğru sonuçlar ürettiği görülmüştür [47].

Hoffman ve arkadaşları tarafından yapılan bir çalışmada, devletlerin vatandaşları ile sosyal medya üzerinden nasıl iletişime geçtiği araştırılmıştır [48]. Bu kapsamda, Almanya'nın en büyük 3 şehrinde yaşayan vatandaşların Facebook hesapları incelenmiş ve bunlar arasından 2.522 Facebook paylaşımı değerlendirilmiştir. Sosyal ağ analizi tekniklerinden biri olan içerik analizi yönteminin kullanıldığı çalışmada, sosyal ağlar üzerinden yapılan iletişimin çevrimdışı iletişime göre çok daha fazla fırsat sunması ve kolaylık sağlamasına rağmen henüz daha tam anlamıyla gerçekleştirilemediği gözlemlenmiştir. Ayrıca, Almanya hükümetinin vatandaşlarıyla iletişim kurma adına sosyal ağları etkin bir şekilde kullanmadığı görülmüştür [48].

Amerika ve Kore devletlerinin devlet stratejileri doğrultusunda sosyal ağ kullanımını inceleyen bir diğer çalışmada her iki devletin vatandaşlarıyla etkileşime geçmek için sosyal ağ kullanımları ve politikalar değerlendirilmiştir [49]. Kore devletinin sosyal ağları, vatandaşlarıyla iletişime geçmek için yeni bir ortam olarak değerlendirdiği ve bu ortamları etkin kullandığı görülmüştür. Amerika'nın ise pek çok sosyal ağı bünyesinde bulundurması ve aktif olarak kullanmasına rağmen bu kullanımın gizlilik ve güvenlik boyutunda kaldığı görülmüştür. Bu nedenle, devletlerin sosyal ağları nasıl etkili ve yararlı bir şekilde

kullanacaklarına dair çalışmaların yapılması gerekliliğine işaret edilmiş ve bazı öneriler sunulmuştur [49].

Mossberger ve arkadaşları tarafından yapılan diğer bir çalışmada, Amerika'nın en büyük 75 şehrinin 2009 ve 2011 yılları arasındaki sosyal ağ kullanımları değerlendirilmiştir [50]. Vatandaşların sosyal ağ kullanımlarının artmasıyla birlikte, devletle bu platformlar üzerinden yürüttükleri ilişkilerinin daha açık bir hale geldiği, devlet kurumları, yöneticileri ve çalışanları ile olan etkileşimlerinin sosyal ağ kullanımı ile birlikte %11'den %96'ya çıktığı görülmüştür [50].

Abdelsalam ve arkadaşları tarafından Mısır devletinin web sayfalarındaki sosyal ağ kullanımını incelemek üzere bir çalışma gerçekleştirilmiş ve devletin kullandığı sosyal ağların sırasıyla mevcudiyetleri, kullanımları ve etkinlik değerleri araştırılmıştır [51]. Devlete ait bütün kayıtlı domainlerin incelendiği araştırma sonucunda, Mısır devletinin web sayfaları üzerinden kurulan sosyal ağ iletişimlerinin çok düşük boyutlarda olduğu görülmüştür [51].

2.3.3. Ülkelerin tartıştığı konuların tespiti

Günümüz dünyasında ülke vatandaşlarının hangi konuları tartıştığı, neler üzerinde kafa yorduğu ve yorum yaptığı sorularının cevabı sosyal ağlarda saklıdır. Bu kapsamda Twitter ağının ana sayfasında yer alan “Trend Topic” yani “Gündem” bölümü ülke yöneticilerine ve vatandaşlarına fikir verebilecek niteliktedir. Twitter'ın özel algoritmaları tarafından belirlenen ve sürekli olarak güncellenen Gündem, tartışılan diğer konulara göre daha kısa sürede hakkında daha çok kişinin konuştuğu yani tweet attığı konular olarak tanımlanmaktadır. Tartışmaya açılmak istenen bir konu # (hashtag) işareti ile etiketlendiğinde Gündem listesine girmeye aday haline gelmektedir. Özel algoritmalar sayesinde yapılan hesaplamalar doğrultusunda, belirlenen konu hakkında atılan tweet hacmi gittikçe artırıyor ve en kısa sürede en fazla tweete ulaşıyor ise de, o konu en çok konuşulan konu yani “Trend Topic” olmuş demektir [52]. Dolayısıyla, Twitter'ın kullanıcılarına sunduğu bu hizmetin takip edilmesiyle, belirli bir zamanda belirli bir ülkede hangi konular tartışılıyor ve insanlar o konu hakkında ne düşünüyor gibi sorulara cevap bulmak mümkündür. Bu da ülkenin o anki ve gelecekteki durumu hakkında yorum yapma ve fikir sahibi olma yetisini kişilere kazandırmaktadır.

Lee ve arkadaşları tarafından yürütülen bir çalışmada [53] ise Twitter'ın sağladığı gündem konuları daha genel kategorilerde sınıflandırılarak sosyal ağlardan veri çıkarımı işleminin kolaylaştırılması amaçlanmıştır. Yapılan çalışma Twitter gündem konuları; spor, politik, eğitim gibi 18 genel kategoride sınıflandırılarak metin ve ağ tabanlı olmak üzere iki farklı yaklaşımla değerlendirilmiştir. Metin tabanlı sınıflandırma metodunda gündem listesindeki konular hakkında atılan tweetler kullanılarak kelime vektörleri oluşturulmuştur. Daha sonra, Terim Frekansı-Ters Doküman Frekansı ağırlıkları kullanılarak Naive Bayes sınıflandırıcısı ile sınıflandırma yapılmıştır. Ağ tabanlı sınıflandırma işleminde ise, en çok konuşulan 5 benzer konu belirlenerek C5.0 karar ağacı ile sınıflandırılmıştır. Yapılan deneyler sonucunda, 18'den fazla kategoriye ait olan 768 gündem konusunun; metin tabanlı sınıflandırma işleminde %65, ağ tabanlı sınıflandırma işleminde ise %70 başarı oranıyla sınıflandırıldığı görülmüştür. Bu tür sınıflandırma işlemleri, ülkeler genel olarak hangi konularda konuşuyor ya da belirli zamanlarda hangi kategoriye ait konular daha fazla tartışılıyor gibi soruların cevabını vererek kişilere kolaylık sağlamaktadır.

Literatürde, bu tür çalışmaları daha da özelleştiren ve ileriye dönük ülkede hangi konuların tartışılabilceğinin tespitini daha önceden yapabilen çalışmalar da mevcuttur [54, 55]. Bu kapsamda Nikolov ve Shah tarafından yapılan çalışmada [54], Twitter'ın sahip olduğu Gündem özelliği kullanılarak "Gündem konularının önceden tahmini için parametrik olmayan bir metot" isimli bir çalışma gerçekleştirilmiştir. Bu çalışma kapsamında, tartışılan konularda atılan tweet sayısı ve tweet sayısındaki artış sürekli olarak hesaplanmaktadır. Daha sonra hesaplanan rakamlar ile konular arasında ilişki kurularak ağ eğitilmektedir. 200 örnek gündem konusunun kullanıldığı bu çalışmanın gündem olabilecek konuları birkaç saat önceden % 95 doğrulukla tespit edebildiği görülmüştür. Geliştirilen bu sistem sayesinde bir ülkede gündem olabilecek konuların tahmin edilmesi; çıkabilecek ayaklanmalara karşı önceden önlem alınması veya ülkelerin birbirleriyle olan ilişkilerinin yürütülmesi açısından hayati bir öneme sahiptir [54, 55]. Ayrıca Twitter üzerinden yapılan bu takip ve tahminlerin diğer sosyal ağlar üzerinden de yapılarak ülke gündeminin nabzının tutulması yürütülmesi gereken projeler arasındadır.

2.3.4. Kullanıcı profillerinin çıkarılması ve kişilik tahmini

Literatürdeki çalışmalar incelendiğinde, sosyal ağlardaki verileri, kullanıcı profillerinin çıkarılması ve kişilik tahminlerinin yapılması gibi amaçlarla kullanan birçok çalışma mevcuttur [56-62].

Li ve Chen tarafından yapılan bir çalışmada konum bildirimi yapılmasına olanak sağlayan mekanizmalar tanımlanarak sosyal ağ paylaşımları tartışılmıştır [56]. Kullanıcıların gerçek ve sanal dünyaları arasında köprü görevi gören bu servisler tanımlanarak konum algılama ve paylaşma özellikleri arasında kapsamlı bir değerlendirme yapılmıştır. Ayrıca kullanıcılar tarafından yapılan sosyal konum bildirimleri ticari amaçlı bir sosyal ağ şirketinden alınarak analiz edilmiştir. Yazarlar çalışma sonucunda, kişilerin özel hayatın gizliliği konusundaki hassasiyetlerinin yaş, cinsiyet, hareketlilik ve coğrafi bölge gibi etkenlerle yakından ilişkili olduğu sonucuna varmışlardır.

He ve arkadaşları tarafından yapılan bir çalışmada Foursquare ve Gowalla gibi konum tabanlı sosyal ağlarda kişilerin bilinçli olarak yaptıkları sahte konum bildirimleri incelenmiştir [57]. Yazarlar, kişilerin sahte konum bildirimi yapmalarına sebep olan davranışların büyük oranda, daha fazla kullanıcıya sahip olmak isteyen ve yer bildirimlerine sanal veya gerçek ödüller sunan servis sağlayıcılarının etkili olduğunu değerlendirmiştir. Ayrıca, sahte konum bildirimlerine karşı sistemlerin konum doğrulama mekanizmalarının iyileştirilmesine yönelik öneriler sunularak profil tarama analizleriyle sınırlandırma yöntemlerinin kullanılabileceği vurgulanmıştır [57].

Noulas ve arkadaşları tarafından yapılan bir çalışmada [58] ise, araştırmacıların hareketlilik (mobility) konusunda daha kapsamlı bir anlayış kazanabilmeleri için konum tabanlı sosyal ağlardan elde edilen verilerden nasıl faydalanabileceği araştırılmıştır. Bu çalışma kapsamında sosyal ağlarda geniş bir kullanıcı kitlesi bulunan Foursquare ağının veri seti kullanılmıştır. Yazarlar, önerdikleri metot ile kullanıcı konum bildirimlerini analiz ederek kişilerin günlük ve haftalık rutinlerini göstermişlerdir.

Cheng ve arkadaşlarının yaptığı bir çalışmada [59], insanların konum tabanlı sosyal ağ kullanımının sosyal ve zamansal özellikleri incelenmiştir. Gelecekte mobil ve lokasyon tabanlı servislerin oluşturulmasında, trafik tahminlerinde ve kent planlamasında

kullanılmak üzere, kişilerin hareketlilik rutinlerinin (Mobility Pattern) nasıl olduğu, konum tabanlı sosyal ağlardan elde edilen veriler yardımıyla değerlendirilmiştir. Elde edilen sonuçlara göre, coğrafi ve ekonomik kısıtların kişilerin hareketlilik rutinlerini ve sosyal statülerini etkilediği görülmüştür. Ayrıca, kişilerin yer bildirimleriyle birlikte bulundukları mekân hakkında yaptığı yorumlar, bağlamsal ve duygusal açıdan değerlendirildiğinde insanların bu servisleri neden kullandıkları ve nasıl etkileşim halinde oldukları konusunda daha detaylı bilgi elde edilebileceği sonucuna varılmıştır.

Fusco ve arkadaşları tarafından yapılan bir çalışmada [60] ise, konum tabanlı sosyal ağların toplumdaki rolü ve arkadaşların birbirlerine duydukları güven duygusu üzerindeki etkisi değerlendirilmiştir. Çalışmada gerekli olacak veriyi toplamak için belirli bir insan grubu üzerinde anket çalışması yapılmıştır ve verileri analiz etmek için nitel bir yöntem kullanılmıştır. Elde edilen sonuçlar, konum tabanlı sosyal ağların kişilerin birbirlerine duydukları güven üzerinde tehdit teşkil ettiği ve kullanıcıların güvenliği açısından bu hizmetlerin daha güvenilir ve gürbüz geliştirilmesi gerektiği yönündedir.

Quercia ve arkadaşları tarafından yapılan çalışmada [61], kullanıcıların anlık iletiler yayımlayabildikleri bir sosyal paylaşım sitesi olan Twitter'daki veriler kullanılarak bu servisin kullanıcılarının karakteristik özellikleri tahmin edilmeye çalışılmıştır. 335 Twitter kullanıcısının yayınladığı tweetler değerlendirilerek bu tweetlerin taşıdığı kişisel bilgiler analiz edilmiştir. Elde edilen sonuçlar, Twitter'daki popüler ve etkili kullanıcıların çoğunun duygusal olarak dışa dönük ve hayal gücü kuvvetli insanlar oldukları yönündedir. Ayrıca, bir kullanıcının 'takip edilen', 'takipçi' ve 'tweet' listesinin bilinmesiyle o kişinin açıklık, dürüstlük, dışa dönüklük, geçimlilik ve sinirlilik gibi 5 önemli karakteristik özelliğinin tahmin edilebileceği elde edilen sonuçlar arasındadır.

Hamburger ve arkadaşları tarafından yapılan bir çalışmada [62] ise, kişilerin karakterleri ile sosyal ağ kullanımları arasında nasıl bir ilişki olduğu saptanmaya çalışılmıştır. Kişilerin Facebook üzerinden yayınladığı temel bilgiler, (cinsiyet, doğum tarihi, yaşadığı yer, vb.) kişisel bilgiler (ilgi alanları, kitaplar, müzikler, hakkımda, vb.) iletişim bilgisi (e-posta, telefon, web sitesi, vb.) ve eğitim ve iş bilgileri (lise, üniversite, kurum, unvan, vb.)'nin değerlendirilmeye alındığı çalışmada, bu bilgileri paylaşma ile içe-dönük ve dışa-dönük karakterlere sahip olma arasındaki ilişki saptanmıştır. Elde edilen sonuçlara göre, dışa-dönük kişilerin daha çok arkadaşına sahip oldukları ve daha fazla grubun üyesi oldukları

görülmüştür. Duygusal bozukluk yaşayan kişilerin ise, özel mesajları kullanmaktansa açık mesaj paylaşımına daha yatkın oldukları gözlemlenmiştir. Sonuç olarak, yapılan bu çalışma kişilerin sosyal ağ kullanım alışkanlıkları ile karakterleri arasında güçlü bir bağ olduğunu göstermiştir.

2.3.5. Tavsiye sistemlerinin geliştirilmesi

Tavsiye sistemleri, belirli bir kullanıcının alışkanlıkları, beğenileri ya da rutinleri ışığında o kişinin belirli bir alanda ilgisini çekebilecek durumların ya da varlıklarının belirlenerek sunulması mantığını taşıyan kişiselleştirilmiş bilgi filtreleme teknolojisi olarak tanımlanmaktadır [63]. Bu sistemler sayesinde çevrimiçi ortamda seçim yapması her geçen gün biraz daha zorlaşan çok fazla miktardaki içerikler arasından kullanıcı için en anlamlı olanlar otomatik ayrıştırılarak, kullanıcının işi kolaylaştırılmakta ve harcanan zamandan tasarruf edilmektedir. Tavsiye sistemlerinin geliştirilmesinde kişiler hakkında bilgi taşıyan birçok veri kaynağı bulunmaktadır. Bu kaynakların başında e-ticaret siteleri (amazon.com vb.) gelmektedir. Tavsiye motorları genellikle e-ticaret sitelerinde kişilerin yaptığı alışveriş alışkanlıkları göz önünde bulundurularak bu veriler ışığında tasarlanmaktadır. Sosyal ağ verileri ise, bu kapsamda oldukça fazla veri taşımaya rağmen pek kullanılmayan bir alandır. Literatürde geleneksel tavsiye sistemlerinden farklı olarak, farklı türdeki sosyal ağ verilerini kullanarak geliştirilmiş tavsiye sistemleri mevcuttur.

Feitosa ve arkadaşları [64], konum tabanlı sosyal ağların içerdiği verileri kullanarak metin madenciliği yöntemi ile sosyal tavsiyeler veren bir sistem tasarlamışlardır. Foursquare kullanıcılarının bulundukları mekânları, mekân ismini kullanarak etiketleme (tagging) özelliğinin kullanıldığı bu çalışmada, metin madenciliği yöntemi ile etiketin içerdiği kelimeler ayrıştırılarak o mekânın yakınında bulunan mekânlar, semtler ve bölgeler kullanıcıya tavsiye olarak sunulmaktadır.

Kwoon ve arkadaşları [65] tarafından yapılan bir çalışmada ise, mobil sosyal ağ verileri kullanılarak kişilere uygun gerçek zamanlı yer etiketli içerikler tavsiye olarak sunulmaktadır. Geliştirilen tavsiye sistemi, kullanıcının bulunduğu yer bilgilerini GPS ile alarak uzaklık ve mekân tercihleri verilerini filtreleme metotları ile eşleştirmekte ve kullanıcıya tercih edebileceği yer tavsiyesini uzaklık bilgisi ile birlikte sunmaktadır.

Özcan ve Öğüdücü tarafından yapılan bir çalışmada [66], sosyal ağ verileri kullanılarak mobil telefonlar için bir tavsiye sistemi örneği geliştirilmiştir. Özcan ve Öğüdücü, mobil servis sağlayıcıların kullanıcılarına arama yapma veya mesaj atma gibi özelliklerden çok daha fazlasını sunmaları gerektiğinden yola çıkarak, sunabilecekleri bir özelliğin de tavsiye sistemleri olduğunu tartışmışlardır. Bu kapsamda, kullanıcının tavsiyeye ihtiyaç duyduğu anda o kişinin tercihlerini dikkate alarak ve içerik filtreleme (content filtering), işbirlikçi filtreleme (collaborative filtering) ve kişinin sosyal ağ kullanımını temel alarak kullanıcıya mekân tavsiyesi yapan bir sistem geliştirmişlerdir. Geliştirilen sistemin performansı bir mobil servis sağlayıcısından alınan veriler ışığında değerlendirilmiştir.

Tavsiye sistemleri ile sosyal ağ verilerini birleştirerek bu alanda fark atmış ve ACM tarafından düzenlenen “Strands Call for Recommender Start-ups” yarışmasında birincilik ödülünü kazanan bir Türk projesi mevcuttur. “İletken” olarak isimlendirilen projenin diğerlerinden farkının tavsiye sistemlerini sosyal ağ verileriyle birleştirdiği noktalar olduğunun ifade edildiği proje, e-ticaret, medya içeriği, mobil içerik ve kurumsal ve kişisel tavsiyeler veren bir sistemdir [67].

2.3.6. Suçların çözümülenmesi ve aydınlatılması

Günümüzde sosyal ağ verilerinin kullanıldığı alanlardan biri de suçların çözümülenmesi ve aydınlatılmasıdır. Brezilya’da sosyal suçlar ve suçun önlenmesi ve azaltılması gibi konularda polis ofislerine destek veren Robert Muggah “sosyal ağların geleneksel suç tespiti yöntemlerine nazaran daha hızlı, daha şeffaf, ölçeklenebilir ve spesifik” veri sunduğundan bahsetmiştir [68]. Bu kapsamda, 2007 yılında Fortaleza Üniversitesi Bilgisayar Bilimleri öğretim üyesi Prof. Dr. Vasco Furtado tarafından “WikiCrimes” isimli bir sosyal ağ geliştirilmiştir [69]. Suçların tespit edilmesinin önündeki en büyük engelin kişilerin başına gelen kötü olayların suç kapsamına girmediğini düşünmeleri ve polise bu nedenlerden dolayı bildirmemeleri olduğunu ifade eden Brezilya Polisi, geliştirilen bu yazılım ile kişilerin çevrimiçi olarak hangi bölgede başlarına ne geldiklerini paylaşılabileceklerini ifade etmiştir. Bu kapsamda sosyal ağ özelliğinin yanı sıra haritalandırma ile görsel bir suç haritalama metodu olarak da düşünülebilecek olan WikiCrimes suçların çözümülenmesi için sosyal ağ kullanımına önem veren en büyük projeler arasında yer almaktadır. Kullanıcıların, yanlış bilgi vermemesi ya da suçluların hedef şaşırtmak üzere sistemi kötüye kullanmasının önüne geçilmek için de birden fazla

önlem alınmıştır. Örneğin, herhangi bir kullanıcı bir yeri bir suç ile ilişkilendirirken video, fotoğraf veya link paylaşımı ile tezini destekleyebilmektedir. Aynı zamanda bir suçun onaylanabilmesi için, en az iki kişi tarafından onaylanması gerekmektedir.

LexisNexis tarafından yapılan bir araştırmaya [70] göre ise, ABD’de 1,221 kolluk kuvveti üyesinin %80’inin suç aydınlatmada sırasıyla Facebook ve Youtube başta olmak üzere sosyal ağ yapılarını kullandıkları ifade edilmiştir.

ABD’de polis departmanları artık bir şüpheliyi ararken sosyal ağ hesapları üzerinden paylaştıkları yazı, resim, video vb. içerikleri kontrol etmektedirler. Herkese açık olan bilgilere erişim yapıp kişi hakkında bilgi toplanırken, erişime kapatılmış özel bilgilere de farklı tekniklerle ulaşılmaktadır. Örneğin, aranan kişinin bilgileri saklıysa arkadaş listesinden kişilere erişilerek bilgi vermeleri istenmektedir. Ya da sahte hesaplar açılıp, aranan kişiyle ortak ilgi alanları varmış gibi gösterilerek arkadaşlık istekleri gönderilmektedir. Kabul edilen arkadaşlıklar sayesinde kişinin paylaşımlarına erişilerek suç tespiti konusunda gerekli olan kanıtlar sosyal ağlar üzerinden toplanmaktadır. Bu duruma örnek teşkil edebilecek bir olay “Melvin Colon” örneğinde yaşanmıştır. Cinayet ve narkotik suçlarından aranan şüpheli çete üyesi Melvin Colon, Facebook hesabı üzerinden çete üyesi olduğunun sinyallerini veren bir fotoğraf yayınlamıştır. Polis bu fotoğrafla birlikte sadece arkadaşlarının görmesine izin verdiği daha anlamlı fotoğrafları da ele geçirmiştir. Bu kapsamda şüphelinin arkadaşlarıyla iletişime geçerek onların hesaplarından da gerekli olan kanıtları toplayan New York polisi, sosyal ağ verileri ile Melvin Colon’u yakalamayı başarmıştır [71]. ABD’nin Cincinnati şehrinde 71 çete üyesinin tutuklanması da yine sosyal ağlardaki veri üzerinden gerçekleştirilmiştir. Bu kapsamda ABD’de yaşanan örnekler Darrin Anderson ve Ronnie Tienda gibi durumlarla artırılabilir. Dolayısıyla, sosyal ağların acil durumlarda ve devlet yetkilileri tarafından suç kapsamında araştırılması gerektiğinde paylaşılan veriler açık olsun ya da olmasın ele geçirilerek suçların çözülmesi ve aydınlatılması alanında kullanılmaktadır.

Özellikle Amerikan polisinin sosyal ağlardaki verileri kullanmasından sonra, bu başlık altında incelenebilecek bir durum da ülkemizde yaşanmıştır. Gezi parkı olaylarından sonra suç takibinde sosyal medyanın büyük önem taşıdığını fark eden hükümet daha önce de belirtildiği gibi Mavi Oda isimli bir birim kurarak sosyal ağları takip etme kararı almıştır.

Bu kapsamda başlatılan projeye Twitter ve Facebook ağları dâhil edilerek bu yapılar üzerinden tertip edilen illegal faaliyetlerin deşifre edileceği bildirilmiştir [46].

Sosyal ağ verilerini kullanarak suçların açığa çıkarılması, günümüzde artık yaygın bir şekilde uygulanırken bu alanda yapılmış akademik çalışmaların sayısı oldukça azdır. Bu kapsamda Zainudin ve arkadaşları [72] tarafından yapılan bir çalışma da bu eksiklikten bahsedilerek yeni bir yaklaşım ortaya konulmuştur. Araştırmacılar çalışma ortamlarını fiziksel ve dijital ortam olmak üzere 2'ye ayırmışlardır. Fiziksel ortamda yetkili birimler tarafından suç duyurusunda bulunulması, gerekiyorsa özel bilgilere erişim için sosyal ağ şirketinden izin alınması ve yeterli bilginin fiziksel dünyadan toplanması gibi adımlar yer almaktadır. Dijital ortamda ise elde edilen bilgiler taranmak istenen sosyal ağda otomatik olarak aratılarak sosyal ağ verileri toplanmaktadır. Daha sonra bu veriler arasından alakasız veriler ayrıştırılarak odak veriler belirlenmektedir. Belirlenen veriler üretilen hipotezler ışığında analiz edilerek suç araştırmasında sosyal ağ verilerinin nasıl kullanılabileceğine dair bir yaklaşım sunulmuştur.

Bu tez çalışmasının başında tarafımızdan yürütülen bir çalışmada ise hakkında suç duyurusu veya izleme kararı verilen bir şüphelinin teknik takibinin yapılabilmesi için yeni bir yöntem önerilmiştir [2]. Önerilen yöntem ile kolluk kuvvetleri, savcılık tarafından hakkında teknik takip talebi bulunan kişiyi yazılımsal olarak izleyebileceklerdir. Ayrıca geliştirilen yöntem sayesinde, kolluk kuvvetlerinin suçlu profili belirlemede yaşadığı zorluklar azaltılarak bu süreçte harcanan zamanın minimize edilmesi de hedeflenmektedir. Gerçekleştirilen uygulamanın teknik takip sürecinde kullanılabilmesi için kişinin kullandığı bir Twitter hesabının olması gerekmektedir. Çalışmamızda teknik takibe alınacak kişinin Twitter üzerinden yayınladığı Foursquare konum bildirimleri değerlendirmeye alınmıştır. Hakkında teknik takip talebi bulunan kişinin Twitter sosyal ağında kullandığı kullanıcı adına erişildikten sonra, bu bilgi geliştirilen yazılıma girdi olarak verilerek o kişinin belirli zaman aralıklarında bulunduğu yerler gerçek zamanlı haritalandırma ile görsel bir biçimde yetkililere sunulmaktadır [2].

Yapılan diğer bir çalışmamızda ise [73], hakkında teknik takip talebi bulunan kişilerin bir sonraki adımında bulunması muhtemel olan yerlerin tahmini yapılarak, o kişinin hangi bölgede aranmasının daha makul olacağı öngörülerek yetkili kişilere görsel bir sonuç sunulmaktadır [73].

2.3.7. Eğitim alanında değerlendirmelerin yapılması

Sosyal ağların ve sosyal ağlar üzerinden yapılan paylaşımların içerdiği veriler günümüzde eğitim amaçlı ve eğitim alanında, eğitim değerlendirilmesi ve kalitenin artırılması için sıklıkla kullanılmaktadır.

Bu kapsamda 2013 yılı için yükseköğretimde sosyal medya kullanımı isimli bir çalışma gerçekleştiren Pearson, çalışma bulgularını bir rapor halinde sosyal medya ve eğitim alanındaki ilgililere sunmuşlardır [74]. 8 bin üniversiteden akademisyenlerin katılımıyla yürütülen çalışmada, öğretim üyelerinin %41'i sosyal ağları eğitim amaçlı kullandıklarını belirtmiştir. 2012 yılı itibarıyla ise bu oranın %34'lerde olduğu görülmektedir [74]. Dolayısıyla, sosyal ağ kullanımının eğitim alanında da büyük ölçüde etkili olduğu ve rakamların her geçen yıl artış gösterdiği görülmektedir.

Koç ve Karabatak tarafından yapılan bir çalışmada [75], Fırat Üniversitesi Teknik Eğitim Fakültesi Bilgisayar Öğretmenliği Bölümü öğrencilerinin sosyal ağları kullanma seviyeleri ve öğrencilerin eğitimi üzerinde ne gibi etkileri olduğu incelenmiştir. Anket yoluyla toplanan veriler veri madenciliği tekniklerinden biri olan birliktelik kuralı ile sonuçlar değerlendirilmiştir. Elde edilen sonuçlara göre, öğrencilerin %61,07'si sosyal ağları dersleri için kullandıklarını ifade etmiştir. Ayrıca öğrencilerin %66,41'i sosyal ağlar sayesinde diğer öğrenciler ve öğretmenleriyle ders ile ilgili konularda daha kolay iletişime geçtiklerini söylemiştir. Bu da sosyal ağların eğitim alanında öğrenci-öğrenci, öğrenci-ders ve öğrenci-öğretmen arasındaki bağları daha da sağlamlaştırdığını ve öğrencilerin sosyal ağlarda öğretmenleri ile arkadaş olma fırsatını yakalayarak kendilerini daha rahat ifade etme olanağına sahip olduklarını kanıtlar niteliktedir [76].

Ajan ve Harsthone tarafından yapılan bir çalışmada [77], ABD'de bulunan büyük bir üniversitenin sosyal ağ kullanımını derslerine entegre etmelerinin altında yatan gerçek sebep araştırılmıştır. Planlanmış Davranış Teorisinin (Theory of planned behavior) kullanıldığı çalışmada, akademik personelin sosyal ağ kullanımının öğrencilere pozitif etkilerinin olduğunu bildikleri gözlemlenmiştir. Akademik personel, özellikle blog türü sosyal ağların öğrencilerin öğrenme yeteneklerini geliştirdiğinin, profesör-öğrenci ilişkilerini yakınlaştırdığını ve öğrencilerin derse olan ilgilerini artırdığını düşünmektedir.

Ülkemizde bu kapsamda Milli Eğitim Bakanlığı'nın yürürlüğe sunduğu bir proje mevcuttur. “MEB’e özel sosyal medya” olarak da bilinen bu proje eğitim bilişim ağı (EBA) ismini taşımaktadır. FATİH (Fırsatları Artırma ve Teknolojiyi İyileştirme Hareketi) projesi kapsamında öğretmen ve öğrenciler için geliştirilmiş olan bu sosyal eğitim platformu, öğrenciler için öğretici videolar izleyip görselleri tarayabilecekleri, dergi indirip okuyabilecekleri ses dosyalarını dinleyip haberleri görebilecekleri ve öğretmenleriyle iletişim kurarak bilgilerini artırıp öğrendiklerini pekiştirebilecekleri bir sosyal ağıdır [50]. Aynı amaçla İzmir Bayraklı Nedret İlhan Keten Ortaokulu öğrencileri için düzenlenmiş bir sosyal ağ da mevcuttur [78]. Öğrencilerin Facebook kullanımına olan alışkanlıklarından dolayı Facebook’un açık kaynak kodları kullanılarak geliştirilen proje, bahsedilen ortaokulun eğitim süreçlerine sosyal ağları da dâhil ederek daha etkileşimli bir öğrenme modeline sahip olmaları için geliştirilmiştir. Ayrıca bu proje Milli Eğitim Bakanlığı ve Microsoft işbirliği ile gerçekleştirilen Yenilikçi öğretmenler Türkiye forumunda ve Fas’ta düzenlenen Orta Doğu ve Afrika forumunda içeriğin kalitesi dalında birincilik ödülü kazanmıştır. Aynı zamanda Prag’da yapılan Dünya forumunda da ülkemizi temsil etmiştir [78, 79].

Ayrıca, dünya çapında da sosyal ağların eğitimdeki yerini araştırmak ve gerekli destekleri sunmak üzere SMILE (Social Media in Learning and Education) isminde bir araştırma projesi başlatılmıştır. Bu proje kapsamında, okulların sosyal ağ politikaları, pedagojik prensipler, profesyonel sosyal ağ geliştiricileri, internet ve sosyal ağ güvenliği gibi birçok konu üzerinde çalışılmaktadır [80].

2.3.8. Duygu ve düşünce analizi

Günümüzde sentiment analizi, düşünce analizi veya düşünce çıkarımı gibi birçok farklı isimle anılan duygu analizi çok yeni bir kavram olmasına rağmen ihtiyaç duyduğu veriyi, yaygın olarak kullanılan sosyal ağlardan temin etmektedir. Dolayısıyla günümüzde sosyal ağ verilerinin kullanıldığı en önemli uygulama amaçlarından biri de duygu ve düşünce analizidir.

Sentiment analizi, “kişilerin birbirlerine veya objelere karşı kalıcı olarak eğilimleri ve duyguları” olarak tanımlanmakta ve genel olarak, metin halinde olan içeriklerin taşıdığı

duygu ve düşüncelerin pozitif, negatif veya nötr yönde olup olmadığı belirlenmeye çalışılmaktadır [81].

Sentiment analizi genellikle şirketler tarafından müşterilerinin ürünler, servisler ve şirket hakkında ne düşündüklerini analiz etmek amacıyla kullanılmaktadır. Bunun yanı sıra, oyuncular veya yazarlar gibi ünlü kişiler tarafından da kullanılarak belirlenen toplulukların duygu ve düşünceleri hakkında fikir edinilebilmektedir.

Sentiment analizi yapılırken, sosyal ağ üzerinden yapılan metin paylaşımları alınarak kelime bazında ayrıştırılmaktadır (isimler, hashtagler, büyük harfle başlayan kelimeler, surat ifadeleri, vb.). Daha sonra farklı türdeki her bir kelime (sıfat, isim, zarf, vb.) taşıdığı anlam bakımından işaretlenmekte ve belirlenen paylaşım, içerdiği kelimelerin ağırlığına göre pozitif, negatif veya nötr olarak sınıflandırılmaktadır. Bu işlem oldukça zordur. Kelimelerin anlamı, kullanılan içeriğe, dile ve kültüre göre değişkenlik gösterebilmektedir. Örneğin, bir kelime bir cümlede olumlu anlamda kullanılırken bir diğerinde olumsuz anlam taşıyabilmekte, bir kelime bir kültürde pozitif anlama sahipken farklı bir kültürde kötü bir anlam taşıyabilmektedir. Bu nedenle, sentiment analizi tam anlamıyla güvenilir bir sistem değildir ve hala üzerinde çalışmalar devam etmektedir [81].

Ceron ve arkadaşları tarafından yapılan bir çalışmada [82], sentiment analizi ile sosyal ağ verileri, İtalya ve Fransa'da yaşayan vatandaşların politik görüşlerini öğrenmek amacıyla kullanılmıştır. Yazarlar sentiment analizi ile farklı senaryoları analiz etmeye çalışmışlardır. Bunlar İtalyan liderlerin 2011 yılı içerisindeki çevrimiçi popülerliğinin analizi ve Fransız internet kullanıcılarının 2012 yılındaki başkanlık seçimlerinde ve sonraki yasal seçimlerdeki oylarının analizi olarak sıralanmaktadır. Yöntem olarak denetimli HK (Hopkins-King) metodunun kullanıldığı çalışmada Twitter sosyal ağı üzerinden yapılan paylaşımlar ele alınmıştır. Elde edilen sonuçlara göre sosyal ağ üzerinden yapılan sentiment analizinin geleneksel anketlerle aynı sonuçları ürettiği ve bu yöntem ile seçim sonuçlarının doğru bir şekilde tahmin edildiği görülmüştür.

Bir başka çalışmada ise, farklı araba markalarına ait araba modelleri hakkında sosyal ağ kullanıcılarının fikirlerini analiz edebilen bir sentiment analizi yazılım prototipi önerilmiştir [83]. Bu kapsamda öncelikle arabalar için sıklıkla kullanılan kelimeler belirlenerek bir sözlük oluşturulmuştur. Daha sonra, sosyal ağ kullanıcılarına yöneltilen

“hangi araba daha güzel?” veya “hangisinin motoru daha güçlü?” gibi sorulara verilen cevapları analiz ederek, araba marka ve modelleri hakkındaki görüşler pozitif ve negatif olarak sınıflandıran bir sistem tasarlanmıştır. Elde edilen sonuçların markalarla paylaşarak üretim ve pazarlama stratejilerine yön vermelerini sağlamak ve bu sistemi daha geniş ürün yelpazesine uygulamak çalışmanın planlanan sonuçları arasında yer almaktadır [83].

Sosyal ağ verilerini kullanarak sentiment analizi yapmak üzere geliştirilmiş birçok araç bulunmaktadır. Bunlar arasında Standford Üniversitesi Bilgisayar Bilimleri Bölümü lisansüstü öğrencileri tarafından geliştirilen Sentiment140, herhangi bir marka, ürün veya konu hakkında Twitter verilerini tarayarak sentiment analizi yapmaktadır. Ayrıca programcıların da kullanarak yeni projeler geliştirebilmeleri için ücretsiz Uygulama Programlama Arayüz (UPA)’ü mevcuttur [84].

Repustate isimli bir başka sistem ise birçok dil desteği (İngilizce, Fransızca, İtalyanca, Almanca, İspanyolca, Arapça ve Çince) ile herhangi bir konuda sentiment analizi yapabilmektedir. Ayrıca, belirlenen duygu ve düşüncelere göre ileriye dönük tahminler de geliştirilen bu sistemin alanları arasındadır [85].

Bu kapsamda başka bir proje de ülkemizde geliştirilmiştir. 2007 yılında kurularak doğal dil işleme teknolojileri üreten bir AR-GE şirketi olan Botego ülkemizde kullanılmak üzere Türkçe sentiment analizi yapan bir sistem geliştirmiştir [86]. Bu sistem ile herhangi bir konuda analiz edilecek metin belirlenen sosyal ağdan çekildikten sonra ilk öğretim adı verilen sınıflandırma işlemi Botego editörleri tarafından el ile (manuel) yapılmaktadır. Sistem öğrendikten sonra ise, kelimeler pozitif, negatif ve nötr olarak belirlenir. Metinde hem olumlu ve olumsuz hem de nötr ifadelerin bulunduğu durumda ağırlıklarına göre ağır basan taraf metnin genel sentimentini olarak seçilmektedir. Daha sonra sınıflandırılan sonuçlar standart raporlarla ilgili birimlere sunulmaktadır [86].

2.3.9. Kişiyeye uygun reklamların üretilmesi

Sosyal ağ verilerinin sıklıkla kullanıldığı alanlardan birinin pazarlama ve reklam stratejilerinin belirlenmesi olduğu daha önceki başlıklarda detaylı bir şekilde ele alınmıştır.

Bu bölümde ise sosyal ağ verilerinin kişiye uygun reklamlar üretilmesindeki rolü açıklanacaktır.

Sosyal ağların en büyük gelir kaynaklarından biri olan standart reklamlar yerini, günümüzde kişiye özel reklamlara bırakmıştır. Kişiye özel reklam stratejisi, sosyal ağlar üzerinden yapılan paylaşımların içerdiği veriler çekilerek bu verilerin yorumlanması ve böylelikle kullanıcıların kişisel tercihlerine, ülkelerine, dillerine, internet kullanım alışkanlıklarına siyasî ve dinî görüşlerine ve hatta belirli bir zamanda sosyal ağ üzerinden paylaştığı ürün içerikli bir cümleye göre yine sosyal ağlar üzerinden reklam yapma anlamına gelmektedir.

Bu kapsamda yapılan bir çalışmada, internet üzerinden ikili ilişkiler kurmaya çalışan orta yaşlı, yaşlı ve çok yaşlı kesime sunulması gereken çevrimiçi reklamların neler olabileceği analiz edilmiştir. “Yahoo Personals” üzerindeki 450 kişisel reklamın incelendiği çalışmada belirlenen yaş gruplarındaki kişilerin ikili ilişkiler hakkında yayınlanan reklamlara yaptığı yorumlar analiz edilmiştir ve sonuç olarak orta yaşlı ve yaşlı kesime sunulması gereken reklamlar arasında fazla fark olması gerekmezken, yaşlı ve çok yaşlı kesime sunulması gereken reklamlar arasında oldukça fark olması gerektiği saptanmıştır. Çok yaşlı kesim ile karşılaştırıldığında orta yaşlı ve yaşlı kesime sunulması gereken çevrimiçi reklamların macera, romantizm ve ruh eşi arama konularında daha net olması gerektiği belirtilmiştir. Çok yaşlı kesime sunulan reklamlarda ise sağlık içerikli ibarelerin daha fazla yer alması gerektiği ifade edilmiştir [87].

Twitter ve Facebook sosyal ağlarının, bu yapılar üzerinden paylaşılan verileri kullanarak kişiye uygun reklam projeleri de mevcuttur [88-90]. Örneğin Facebook, kişiye özel reklam yayınlama stratejisini 3 farklı yolla gerçekleştirmektedir. Bunlardan ilki “Facebook Power Editor Chrome” eklentisidir. Bu eklenti sayesinde şirketler sahip oldukları müşteri verilerini kullanarak Facebook reklamlarını direk kişi bazlı olarak yayınlayabilmektedirler. Herhangi bir reklam için hedeflenen kişilerin Facebook ID, telefon numarası veya e-posta adresleri Facebook Power Editor’e yüklenerek bu kişilere gösterilmek istenen reklamlar Facebook sayfalarında görüntülenebilmektedir [88]. İkinci olarak, kişilerin Facebook üzerinden beğenip paylaştığı fotoğraf içeriklerinden bu kişilerin ilgisini çekebilecek reklamların gösterilmesi hedeflenmektedir. New York Üniversitesi’nde yapılması planlanan bir çalışmada, dijital görüntü işleme konusunda hazırlanacak özel bir algoritma

ile kişiler hakkında bilgi toplanarak bu bilgilerin ilgili reklamcılarla paylaşılacağı yönündedir [89]. Üçüncü olarak ise, Facebook kişiye özel gerçek zamanlı reklam uygulaması ile sosyal ağ verilerini bu yönde kullanmaktadır. Bu kapsamda kullanıcıların durum güncellemelerinde veya arkadaşlarının duvarlarına yaptıkları yorumlarda kullanılan kelimeler baz alınmaktadır. Geliştirilen algoritma sayesinde, herhangi bir paylaşımda ayakkabı kelimesini kullanan Facebook kullanıcılarında gerçek zamanlı olarak ayakkabı reklamı sunulabilmektedir. Hazırladığı reklam havuzundaki reklamları anında kullanıcıya sunabilen bu algoritma, kişilerin sadece beğenilerini ya da aktivitelerini değerlendirmekle kalmayıp yapılan yer bildirimlerine kadar değerlendirmeye almaktadır [90]. Twitter sosyal ağı da kişilerin yaptığı paylaşımların içerdiği verileri kişiye özel reklam amacıyla kullanmaktadır. Kişilerin internette gezindiği sitelerin kullanıldığı bu projede, reklam vermek isteyen şirketler kendi web sitelerini ziyaret eden kullanıcı bilgilerini Twitter ile paylaşarak, bu kişilerin Twitter sayfalarında kendilerine ait reklamların görünmesini sağlayabilmektedir.

2.3.10. Çıkarım yapma

Sosyal ağ verilerinin sıklıkla kullanıldığı alanlardan biri de eksik bilgi olduğu durumlarda mevcut bilgilerin kullanımıyla çıkarım yapılarak istenen bilgiye ulaşılması durumudur. Bu kapsamda, kişilerin sosyal ağ üzerinden herkese açık olarak paylaştığı bilgiler toplanarak genel anlamda bir ağ topolojisi oluşturulmaktadır. Toplanamayan bilgilerin neler olduğu da tespit edilerek eldeki verilerle bilinmeyen verilere ulaşabilmek için mantıklı çıkarımlar yapılmaktadır ve böylelikle eksik bilgiye ulaşılabilir.

Tang ve arkadaşları tarafından yapılan bir çalışmada [91], Facebook kullanıcılarının sadece isimleri kullanılarak bu kişilerin cinsiyetleri ve Facebook davranışları konusunda çıkarım yapılmıştır. New York'taki Facebook kullanıcılarının bu sosyal ağ üzerindeki herkese açık olan bilgileri toplanarak bir isim listesi oluşturulmuştur. Listedeki her bir isim, toplamda kaç kişinin bu isme sahip olduğu, bunlardan kaçının bayan olduğu ve kaçının erkek olduğu bilgileriyle ilişkilendirilmiştir. Daha sonra ABD devletinin yıllara göre popüler isimler listesi kullanılarak isimler erkek veya bayan diye etiketlenmiştir. Makine öğrenmesi metotları ile isimden cinsiyet tahmini yapabilen metotlar kullanılarak çıkarım yapılmıştır. Elde edilen sonuçlara göre, sistemin %95,2'lik bir başarı oranı ile çıkarım yapabildiği görülmüştür. Cinsiyet tespiti yapıldıktan sonra bayanların ve erkeklerin Facebook

davranışları analiz edilmiştir. Sonuç olarak, erkeklerin ve bayanların çelişkili bilgileri Facebook üzerinden rahatlıkla paylaştıkları fakat cinsiyet, yaş ve cinsel tercih gibi bilgileri gizledikleri görülmüştür. Ayrıca, bayanların çevrimiçi gizlilik konusunda erkeklere göre daha bilinçli oldukları tespit edilmiştir.

Bu kapsamda yapılan bir başka çalışmada [92] ise, sosyal ağlardaki ilişkilerin isimlendirilmesi (meslektaş, samimi arkadaş, vb.) üzerine çalışılmıştır. Kişiler arasındaki bağları çözümlmek üzere yarı denetimli PLP-FGM (Partially-labeled Pairwise Factor Graph Model) öğrenme modeli kullanılmıştır. Model, yayın veri seti, e-posta veri seti ve mobil ağ veri seti olmak üzere 3 farklı veri seti üzerinde denenmiştir. Elde edilen sonuçlara göre, yayın veri seti için danışman-öğrenci ilişkisinin %92,7, e-posta veri seti için yönetici-çalışan ilişkisinin %88 ve mobil ağ veri seti için arkadaşlık ilişkisinin %83,1 doğruluk oranıyla tespit edildiği görülmüştür [92].

3. TWITTER

Web 2.0'ın hayatımıza girmesi ile birlikte web sayfaları statik yapılardan dinamik yapılara dönüşerek kullanıcı ile iletişimde daha etkileşimli çözümler sunar hale gelmiştir. Örneğin Web 2.0 teknolojilerinden JavaScript ve AJAX, web sitelerinin masaüstü uygulamalara benzer şekilde kullanıcı ile etkileşimli bir şekilde çalışarak içeriğin dinamik bir şekilde güncellenmesi ve kullanıcının bilgisayarından istek gönderilen sayfa yüklendikten sonra da farklı türdeki içeriklerin (flash, reklam vb.) kullanıcıya sunulmasını sağlamaktadır [93]. Web 2.0 çağının internet kullanıcılarına sunduğu en önemli kavramlardan biri de sosyal ağlardır. Sosyal ağlar kişilerin kendilerini tanımlayarak oluşturdukları profil bilgilerinin internet ortamında paylaşılmasına imkân sağlayan, bireylerin çevrimiçi olarak birbirlerini tanımlarına ve sosyal ilişkiler kurmalarına izin veren ve aynı zamanda sisteme kayıtlı bulunan tüm kullanıcıların birbirlerinden haberdar olmasını sağlayan sanal topluluklar olarak tanımlanmaktadır [73]. Web 2.0, kullanıcıların sosyal ağ sitelerinde içerik yaratmasına izin vererek bu yapıların birçok alanda etkili bir şekilde kullanılmasını sağlamıştır [94]. Marshall ve Shipman'a göre ise sosyal ağ kullanıcılarının bu platformlar üzerinden yapabilecekleri aktiviteler 4 grupta sınıflandırılabilir [95]. Bunlar;

1. Dijital içeriğin kaydedilmesi ve saklanması,
2. Oluşturulmuş dijital içeriğin belirli kişi ve gruplarla paylaşılması,
3. Oluşturulmuş içeriğin geniş kitleler tarafından erişilebilir olması ve
4. İçeriğin istenilen zaman diliminde sosyal ağ hesabından kaldırılabilmesi

olarak sıralanmaktadır [95].

Sonuç olarak, geleneksel e-posta ve kısa mesaj uygulamaları ile kıyaslandığında sosyal ağlar gibi Web 2.0 uygulamalarını kullanan kişilerin daha aktif oldukları ve insan doğasının gereklerini daha iyi yansıttıkları görülmüştür [96].

Yukarıda 20 farklı başlık altında sınıflandırılan sosyal ağlar içerisinde, günümüzde en sık kullanılan sosyal ağ türlerinden biri bloglardır [94]. “Web Log” kelimelerinin birleştirilmesinden meydana gelen blog kelimesi çevrimiçi günlükler olarak tanımlanmaktadır [97]. Divine ve arkadaşlarına göre ise bloglar; kişilerin veya grupların, herhangi bir olay veya durum karşısındaki fikir ve görüşlerini ters kronolojik sıraya göre

yayınlayabilmeleri için tasarlanmış çevrimiçi defterlerdir [93]. Bunların yanı sıra, blogu takip eden kişilerin de yapılan paylaşımlar hakkında yorum yaparak fikirlerini beyan edebildikleri platformlar olarak da tanımlanmaktadır. Mark ve arkadaşlarına göre bloglar aracılığı ile paylaşılan bilgi küresel anlamda bütün insanlığa açık olmakla birlikte geleneksel medyanın ulaşamadığı birçok alana ulaşabilmekte ve yine geleneksel medyanın yaymakta yetersiz kaldığı birçok fikir ve görüşü yayabilmektedir [98]. Gaonkar ve arkadaşları mikroblogları; dünya çapında görüş sağlayan yüksek çözünürlüklü bir sosyo-fiziksel mikroskop olarak tanımlamıştır [99].

2006 yılında ortaya çıkarak bir blog türü olarak sayılabilecek olan mikrobloglar ise; kullanıcılarına, sınırlandırılmış karakter özelliği ile yaşamları ve düşünceleri hakkında kısa mesajlar yayınlama şansı veren ve bu mesajları arkadaşları ve diğer ilgili kişilerle paylaşımlarına olanak sağlayan anlık mesajlaşma platformları olarak tanımlanmaktadır [100-102].

Mikrobloglarda da bloglarda olduğu gibi kullanıcıların düşüncelerini ve görüşlerini paylaşmak üzere yayınladıkları içerikler yer almaktadır. Yine kullanıcıların her iki platformu da güncel olayları tartışma, fikir alma ve verme ve ağızdan ağza yayılma (word-of-mouth) ile gündem hakkında bilgi sahibi olma amacı ile kullandıkları bilinmektedir [103].

Ayrıca, her iki sosyal ağ yapısında da en sık kullanılan birincil paylaşım türü metin içerikleridir ve ters kronolojik sırada yayınlanmaktadır. Bütün bunların yanı sıra hem bloglarda hem de mikrobloglarda ilgi duyulan kişilerin paylaştıkları içeriklerin takip edilmesi mümkün olmaktadır [102].

Fakat mikrobloglar her ne kadar blog ismini taşısalar ve bloglarla ortak özelliklere sahip olsalar da bloglardan ayrıştıkları birçok özellikleri de mevcuttur. Bunlardan ilki mikrobloglardaki karakter sınırıdır. Bloglarda kişiler istedikleri uzunluk ve karakter sayısına sahip metinler yayınlayabilirken mikrobloglarda yayınlanmasına izin verilen metin uzunluğu belirli karakter sayıları ile sınırlandırılmıştır. Günümüzde mevcut olan mikrobloglarda karakter sınırı en fazla 200'dür. Fakat Twitter ve Sina Weibo ağında olduğu gibi genellikle 140 karakterlik paylaşımlara izin verilmektedir [102, 104].

İkincisi ise mikrobloglar genel anlamda mobil terminaller için geliştirilmiştir [102]. Her ne kadar günümüzde her türlü web içeriğine mobil cihazlar aracılığı ile erişilebiliyor olsa da blogların çıkış fikri mobil terminaller için bu tür platformlara ihtiyaç duyulduğu değildir. Fakat mikrobloglar içerdikleri karakter sınırlaması ile cep telefonlarının fonksiyonlarından biri olan kısa mesaj özelliğini taşımaktadır [100]. Dolayısıyla mikroblogların çıkış fikri mobil cihazlar üzerinden sadece çift yönlü iletişim değil çok yönlü iletişim (one-to many) kurulmasını sağlamaktır. Ayrıca bu hizmet karakter sınırlaması özelliğini kısa mesaj hizmetinden, altyapı hizmetini de internet üzerinden sunan blog yapısından alarak kısa mesaj ve blog hizmetlerinin birleştirilmesi ile oluşturulmuş hibrid bir yapı olarak nitelendirilebilmektedir.

Blog ve mikroblogların farklılık gösterdiği bir diğer nokta ise, kullanıcılar tarafından bu platformlara zaman bazında içerik sağlanma sıklığıdır. Yapılan bir araştırmaya göre bloglarda içerik paylaşma ve güncellenme sıklığı 1 güne tekabül eden 86.400 saniye iken mikrobloglarda bu süre 1800 saniye yani yarım saate kadar düşmektedir. Dolayısıyla mikroblogların mobil platformlar için kısa mesaj niteliğinde tasarlanmış olması da bu sonucu kanıtlar niteliktedir [96].

Yukarıda bahsedilenlere ek olarak blog ve mikroblogların bir diğer farkı da içerik yayınlamada yaşanan zorluklardır. Blog yazarlarının blogların yapısı gereği herhangi bir içerik yayınlarken daha çok zaman harcadıkları ve yayınlayacakları paylaşımların içeriklerine ve yapısal görünüşlerine daha fazla dikkat gösterdikleri görülmüştür. Fakat bu durum mikrobloglarda bir hayli farklıdır. Mikrobloglarda kullanıcıların daha çok gelişigüzel paylaşımlar yaptıkları, paylaşımları için fazla zaman harcamadıkları ve herhangi bir olay veya durum karşısındaki görüşlerini, düşünmek için fazla zaman harcamadan anlık olarak paylaştıkları görülmüştür [96].

Çizelge 3.1. Blog ve mikrobloggerların özelliklerinin karşılaştırılması

Özellikler/Blog ve Mikrobloggerlar	Blog	Mikroblogger
Geçmiş Paylaşımları Görme	✓	✓
Karakter Sınırı		✓
Metin Paylaşma	✓	✓
Resim Paylaşma	✓	✓
Video Paylaşma	✓	✓
Kullanıcı Takip etme	✓	✓
Yorum Yazma	✓	✓
Mobil ve Web Platformlarından Erişim	✓	✓
Anlık İleti Yayınlama		✓
İçeriği Gizleyebilme		✓
Profil Editörü	✓	✓
Arkadaşlıklar		✓
3. Parti Uygulamalar	✓	✓
Gizlilik Ayarı		✓
Kullanıcı Engelleme		✓
Kişileri Etiketleme		✓
Grup Oluşturma		✓

Çizelge 3.1’de blog ve mikrobloggerların temel farklılıkları özetlenerek karşılaştırma yapılmıştır. Çizelge 3.1’den de görülebileceği gibi mikrobloggerların bir blog türü olduğu fakat kullanım alanları ve amaçları bakımından pek çok noktada farklılık gösterdikleri ve pek çok özelliği destekledikleri görülmektedir.

Çizelge 3.2. Mikroblogger sitelerine ait özelliklerin karşılaştırılması

Özellikler/Mikrobloggerlar	Twitter	Tumblr	Plurk	Hictu	IRateMyday
2014 yılı yaklaşık aktif kullanıcı sayısı	255 Milyon	183.3 Milyon	210.000	430.920	15.199
Dünya Çapında Kullanım Yüzdesi	> 23 ülke	> 22 ülke	> 10 ülke	1 ülke	1 ülke

Çizelge 3.2’de ise en sık kullanılan mikroblogger sitelerinin karşılaştırması yapılmıştır. Karşılaştırma yapılırken her bir mikroblogger ağının resmi sayfalarından ve web siteleri hakkında analitik bilgiler sunan Alexa’dan alınan veriler kullanılmıştır [105]. Tüm

sitelerde UPA desteği olduğu ve iOS, Android, Windows Phone ve BlackBerry platformlarının mikroblogları destekledikleri bilinmektedir. Çizelgeye bakıldığında en çok kullanıcıya sahip olan mikroblog ağının Twitter olduğu görülmektedir. Bu tez çalışmasında da Twitter verileri kullanılacağından bu bölümde Twitter, tüm özellikleri ile birlikte detaylı bir şekilde ele alınacaktır.

Twitter ortaya çıktığı 2006 yılından bu yana; kişilerin anlık olarak yaptığı faaliyetleri; ekonomik, sosyal, çevresel veya politik durumları, düşüncelerini, çeşitli konulardaki bilgilerini ve haberleri 140 karakter ile; bir web tarayıcısı, SMS, e-posta veya 3. parti uygulamalar yardımıyla paylaşımlarına olanak sağlayan, bu paylaşımların kişilerin kullanıcı profilinde görüntülenmesine izin veren ve birden fazla kullanıcı grubundan çok sayıda farklı kullanıcıyı bünyesinde barındıran, dünya çapındaki en popüler mikroblog ağı olarak tanımlanmaktadır [100, 106-108].

Humphreys ve arkadaşları Twitter’da içerik analizi yaparak, bu yeni iletişim aracının geçmişteki insan alışkanlıkları ile bağlantısını araştırmışlardır [109]. Bu kapsamda geçmişi 18. Yüzyıla kadar uzanan kişisel günlükler ile günümüzde “çevrimiçi kişisel günlük” olarak tanımlanan Twitter sosyal ağını karşılaştırmışlardır. Sonuç olarak, Twitter ve kişisel günlükler arasındaki benzerlikleri

1. Doğaları gereği halka yarı-açıktır.
 2. Günlük veya belirli bir zamana bağlı periyodik aktivitelerin kayıtlarını tutmaktadır.
 3. Hikâye ve anlatı formundadır.
- gibi sıralamışlardır [100].

İkisi arasındaki fark ise Twitter’da sosyal ilişkiler kurarak sistematik bir şekilde istenilen kişiler takip edilebilirken, kişisel günlükler de böyle bir özelliğin mümkün olmayışı olarak ifade edilmiştir [100, 109].

Twitter’da 140 karakter ile yapılan metin tabanlı paylaşımlara tweet adı verilmekte ve kullanıcılar tarafından yayınlanan bu tweetler otomatik olarak Twitter hesap sahibinin profil sayfasında görüntülenmektedir. Fakat bu paylaşımlar, diğer kullanıcılar tarafından ancak Twitter hesap sahibinin verdiği izin ölçüsünde okunabilmektedir [110]. Resim 3.1’de Twitter mikroblog ağına ait güncel kullanıcı ara yüzü gösterilmiştir [52].



Resim 3.1. Twitter’a ait güncel kullanıcı ara yüzü

3.1. Twitter kullanıcı kitlesi

Ülkemizde ve Dünyada Twitter ağına üye olan kullanıcı sayısı her geçen gün biraz daha artmaktadır. 35’den fazla dil desteği sağlayan Twitter’da ait aktif kullanıcı sayısı, 2014 yılı itibariyle 255 milyona ulaşmıştır [52]. Ayrıca 2014 yılı Mayıs ayı itibari ile günlük gönderilen tweet sayısının 500 milyona ulaştığı Twitter şirketinin yayınladığı bilgiler arasındadır. Bunun yanı sıra istatistikler, Twitter kullanıcılarının %78’lik bölümünün Twitter’a mobil platformlar üzerinden ulaştığı ve hesapların %77’sinin Amerika Birleşik Devletleri dışından kişilere ait olduğunu göstermektedir.

Bu kapsamda Twitter kullanıcı kitlesinin oldukça zengin ve çeşitli gruplardan oluştuğu bilinmektedir. Temel anlamda Twitter kullanıcı kitlesi, bireyler ve kurum-kuruluşlar olarak sınıflandırılmaktadır [100]. Twitter kullanan ünlüler, sanatçılar, şarkıcılar, sporcular, politikacılar ve ulusal liderler bireysel kullanıcılar arasında değerlendirilirken haber kuruluşları, özel amaca hizmet eden şirketler ve devlet daireleri ise Twitter’ın kurumsal kullanıcıları olarak değerlendirilmektedir.

Krishnamurthy ve arkadaşlarına göre Twitter kullanıcıları “yayın yapan kullanıcılar”, “arkadaşlık ilişkilerini esas alan kullanıcılar” ve “kötü niyetli kullanıcılar” olarak 3 başlık altında sınıflandırılmaktadır [111]. Yayın yapan kullanıcılar genellikle kendisini takip edenlerin sayısının kendisinin takip ettiklerinin sayısından fazla olan hesaplar olarak ifade edilmiştir ve yayın istasyonları, haber kuruluşları ve ünlü kişiler bu kapsamda değerlendirilmiştir. Arkadaşlık ilişkilerini esas alan kullanıcılar grubunda ise takip edilen ve takipçi sayısının neredeyse eşit olduğu ve aynı kişilerden oluştuğu görülmektedir.

Bireysel kullanıcılar da genellikle bu grup altında ele alınmaktadır. Kötü niyetli kullanıcılarda ise takip edilen sayısının takipçi sayısına göre oldukça fazla olduğu ve bot ve istenmeyen e-posta (spam) hesaplarının bu gruba dâhil olduğu belirtilmiştir.

3.2. Twitter’a ait temel kullanım özellikleri

Twitter mikroblog ağı, diğer mikrobloglarda olduğu gibi metin tabanlı 140 karakterli paylaşım yapma, diğer hesapları takip etme, arkadaşlık ilişkileri kurma gibi genel mikroblog özelliklerine sahip olsa da kendine özgü birçok özelliği de bünyesinde barındırmaktadır. Bu başlık altında Twitter’a ait temel kullanım özellikleri detaylı bir şekilde ele alınarak bu özellikler aşağıda kısaca tanıtılacaktır.

Tweet: Kelime anlamı “kuş cıvıltısı” olan tweet, Twitter’ın hayatımıza girmesi ile birlikte “Twitter üzerinden atılan mesaj” olarak sözlüklerde yerini almıştır [112]. Twitter’da 140 karakter uzunluğundaki metin tabanlı paylaşımlara “tweet” adı verilmekte ve kullanıcılar tarafından yayınlanan bu tweetler otomatik olarak Twitter hesap sahibinin profil sayfasında görüntülenmektedir [52, 107, 113, 114].

Retweet: Herhangi bir kullanıcı tarafından yayınlanmış bir tweetin başka bir kullanıcı tarafından kendi hesabı üzerinden tekrar yayınlanması ve yönlendirilmesi işlemine “retweet” adı verilmektedir. Retweet işlemi yayınlanan her bir tweetin altındaki “Retweetle” bağlantısı ile gerçekleştirilmektedir. Bir tweet retweetlendiğinde o paylaşımın gerçek sahibinin kullanıcı adı, tweetin başındaki “RT” ifadesi ve orijinal tweet metni ile birlikte retweet yapan kişinin profil sayfasında yayınlanmaktadır. Yapılan çalışmalar bir tweetin retweet edilme sebebinin genellikle retweet eden kişinin düşünceleri ile örtüşmesi, önemli veya ilginç bulunması ile ilişkili olduğunu göstermektedir [52, 115].

Takip: Twitter, kullanıcılarına bu ağ üzerindeki diğer kullanıcıları takip etme ve dolayısıyla diğer kullanıcılar tarafından da takip edilme özelliği sunmaktadır. Diğer sosyal ağlar üzerindeki arkadaşlık ilişkisi olarak da değerlendirilebilecek olan bu özellik; kullanıcılara diledikleri kişilerin Twitter paylaşımlarını görme, takip etme, bu kişilere tweet atma gibi fonksiyonlar sunmaktadır [52].

Hashtag: Kullanıcıların bir tweetin konusunu veya kategorisini göstermek üzere tweetlerine “#” işaretiyle birlikte ekledikleri etiketlere hashtag adı verilmektedir. Diğer

sosyal ağlardaki etiketleme özelliği ile benzerlik göstermektedir. Genellikle özel durumlarda veya tartışmalarda bir görüşü desteklemek ve fikir beyan etmek amacıyla kullanılmaktadır. Ayrıca herhangi bir konu hakkında atılmış tweetleri araştırırken bu özellik oldukça etkili bir şekilde kullanılabilir [52].

Mention: Kullanıcıların başka bir kullanıcıya cevap vermek, kişiyi de konuşmaya dahil etmek veya dikkatini çekmek amacıyla “@” işaretinden sonra o kişinin Twitter kullanıcı adını kullanarak paylaştıkları tweetlere mention adı verilmektedir. Ayrıca mention sayıları kullanıcıların birbirleri ile ve markalar hakkındaki konuşma oranlarını ve sıklıklarını belirlemede kullanılmaktadır [52, 110].

Gündem: Twitter’da en çok kullanılan hashtaglerin oluşturduğu listeye gündem adı verilmektedir. Twitter’ın özel algoritmaları tarafından belirlenen ve sürekli olarak güncellenen Gündem, tartışılan diğer konulara göre daha kısa sürede hakkında daha çok kişinin konuştuğu yani tweet attığı konular olarak da tanımlanmaktadır. Tartışmaya açılmak istenen bir konu # (hashtag) işareti ile etiketlendiğinde Gündem listesine girmeye aday haline gelmektedir. Özel algoritmalar sayesinde yapılan hesaplamalar doğrultusunda, belirlenen konu hakkında atılan tweet hacmi gittikçe artırıyor ve en kısa sürede en fazla tweete ulaşıyor ise de, o konu en çok konuşulan konu yani Gündem (Trend Topic) olmuş demektir [52].

Liste: Twitter liste özelliği, kullanıcıların grup oluşturarak istediği tweetleri o listeden yayınlamasına imkan tanıyan bir özelliktir. Diğer sosyal ağlardaki grup oluşturma fonksiyonu ile oldukça benzerdir [52].

Gizlilik ayarı: Twitter gizlilik ayarı, kullanıcıların diğer kullanıcılar tarafından fotoğraflarda etiketlenmesi, profil sayfasının görüntülenme seviyesi, atılan tweetlerde arka planda konum bilgisinin iletilmesi durumu, kişinin diğer kullanıcılar tarafından Twitter’da aranma kısıtı, Twitter hesabının kişiselleştirilmesi ve reklam içeriklerinin yayınlanması gibi birçok özelliği koruma altına alma şansı sunmaktadır [52].

3.3. Haber Kaynağı olarak Twitter

Twitter verisi ile yapılan çoğu çalışmada bu platformun alternatif bir haber kaynağı olarak kullanıldığı görülmüştür [100, 108]. Her gün milyonlarca kişi çeşitli konularda paylaşım yaptığından dolayı kullanıcılar Twitter'ı en güncel olayları ve haberleri takip edebilecekleri bir platform olarak görmektedirler [116, 117]. Naaman ve arkadaşları; Twitter'ın bir mikroblog olmasının ve mikrobloglarda var olan karakter sınırlamasının gazete manşet ve sürmanşetlerindeki karakter sayısından ileri gelmesi durumunun, bu platformun bir haber paylaşım kaynağı olarak kullanılmasının temelini oluşturduğunu ileri sürmüştür [118]. Kwak ve arkadaşları tarafından yapılan çalışmada da Twitter topolojisi ve gündem konuları analiz edilmiştir. Elde edilen sonuçlar gündem konularının %85'inin haber manşetlerinden oluştuğunu göstermektedir. Dolayısıyla bu sonuçlar dâhilinde Twitter'ın sosyal ağdan çok bir bilgi ağı olduğu sonucuna varılmıştır [119]. Bu bölümde Twitter'ın bir haber kaynağı olarak değerlendirilmesinin altında yatan sebepler literatürdeki çalışmalar ışığında farklı başlıklar altında ele alınmıştır. Bunlar alt başlıklarda sunulmuştur.

Sık kullanılan öncelikli bir platform olma: Yaşanan gerçek olaylar dâhilinde Twitter'da haberlerin geleneksel medyaya nispeten daha hızlı yayıldığı ve genellikle herhangi bir kötü durum hakkındaki haberlerin, geleneksel medyadan daha önce Twitter'dan kişilere ulaştığı bilinmektedir. Bu durumun en güzel örneği 2009 yılında Michael Jackson'ın ölümü olayında yaşanmıştır. 2009 yılının Haziran ayında ilk olarak Twitter'da Michael Jackson'ın öldüğüne dair bir haber yer almıştır fakat daha sonra Google, bunun kötü niyetli kişiler tarafından yayınlanan bir yalan haber olduğunu duyurmuştur. Ancak 25 dakika sonra Google tarafından Michael Jackson'ın ölümü doğrulandığında durum netlik kazanmıştır. Dolayısıyla bu durum tüm dünyaya Twitter'ın haber kaynağı olarak rolünün değerlendirilmesinde örnek bir olay teşkil etmiştir [100].

Geleneksel medyadan önce Twitter'da yayınlanan ve dünyayı sarsan olayların bir diğer örneği de Usame bin Ladin'in ölüm haberidir. 1 Mayıs 2011 tarihinde Barack Obama tarafından Usame bin Ladin'in öldüğüne dair yapılan resmi açıklamadan 1 saat 11 dakika öncesinde bu haberin Twitter'da yayınlandığı bilinmektedir [120].

İlk olarak Twitter'dan yayınlanan haberlere birçok örnek ülkemizde de yaşanmıştır. Bu kapsamda 30 Mart 2014 yerel seçimlerinde henüz televizyon kanalları için yayın yasağı kalkmadan önce açılmaya başlanan sandıklar için oy oranlarının belediye başkan adayları tarafından Twitter üzerinden yayınlandığı bilinmektedir [121, 122]. Yine spor dünyasında da Beşiktaş takımının futbolcusu olan Julien Escude'nin takımdan ayrılış haberi spor kanalları ve haber bültenlerinden önce futbolcunun kişisel Twitter hesabı üzerinden duyurulmuştur [123].

Bir sosyal medya optimizasyon şirketi olan SocialFlow tarafından yapılan bir çalışmada, haberleri internetten takip eden kullanıcıların haber taşıyan herhangi bir tweete tıklama olasılığının haber kuruluşlarının internet sayfalarındaki haberlere tıklama olasılıklarından daha yüksek olduğu yönündedir [124].

Ayrıca ülke vatandaşlarının bölgesel olayları öncelikle Twitter'dan paylaştığını ve bu haberlerin kitlesel medyadan önce Twitter'dan yayıldığını gören Zubiaga ve arkadaşları, gazetecilerin Twitter'dan haber toplamalarına yardımcı olmak amacıyla TweetGathering isimli bir sistem geliştirmişlerdir [117]. Geliştirilen arama motoruna gazetecilerin anahtar kelime girmesiyle Tweetleri tarayan ve bu tweetler arasından haber niteliğinde olanları kullanıcıya sunan bu sistem 5 profesyonel gazeteciye denetlenmiştir. Belirlenen kullanıcılardan gelen olumlu yorumlar sistemin gelecek vaat ettiğini göstermektedir.

Güvenilir bir platform olma: Usame bin Ladin'in ölüm haberi üzerine Hu ve arkadaşları tarafından yapılan bir durum analizi çalışması sonucunda Twitter'dan yayınlanan bu haberin henüz geleneksel medya tarafından doğrulanmadan Twitter kullanıcılarını ikna ettiği görülmüştür [120]. Bu durum da günümüzde Twitter'ın haber kaynağı olarak nitelendirildiğini ispatlar niteliktedir. Kullanıcıların resmi bir açıklama yapılmadan dahi bu habere inanmasının ve ikna olmasının sebebi araştırıldığında ise, haberi yayan tweetin Amerika Birleşik Devletlerinin önceki savunma bakanı Donald Rumsfeld tarafından atılmış olması olduğu görülmüştür. Dolayısıyla Twitter'ın gazetecilerden ünlü kişilere kadar birçok kullanıcıya sahip olması, bu platformun haber kaynağı niteliğini artıran bir özelliği olarak değerlendirilmektedir.

Hızlı paylaşım yapma olanağı sunma: Twitter'da haberlerin geleneksel medyadan daha önce yayınlanmasının yanı sıra tartışılan bir diğer konu da bu platformda haberlerin diğer

haber kaynaklarına göre daha hızlı yayılıyor olmasıdır. Bu durumun sebebi ise paylaşım yapan kullanıcıların haberi doğrulama veya kaynak gösterme gibi bir zorunluluklarının olmayışıdır. Haber kuruluşlarında ise bu durum tam tersine işlemektedir. Bu kuruluşlar haberi kendileri üretmediklerinden dolayı haberi aldıkları kaynaklarını araştırmak, göstermek ve indekslemek durumundadırlar. Bu işlemler zaman alıcı olduğundan dolayı Twitter’da haberlerin daha önce yayınlanması kaçınılmaz bir durumdur [116, 125].

Çift yönlü bir iletişim kanalı olma: Twitter’ın bir haber kaynağı olarak algılanmasının başlıca sebeplerinden biri de çift yönlü iletişime imkân tanıyan bir platform olmasıdır. Bu özelliğini sosyal ağ kimliğinden alan Twitter, kullanıcılarına haber kuruluşları veya gazetecilere ait sayfalar üzerinden yapılan paylaşımlara birebir cevap yazma, yorum yapma ve görüş belirtme gibi özellikler sunmaktadır. Televizyon, radyo, gazete ve dergi gibi geleneksel medya organları ise yapılarından dolayı böyle bir özelliğe sahip değildir ve tek yönlü iletişim araçlarıdır. Dolayısıyla sosyal ağ kavramı ile kullanıcıların hayatına giren Twitter, kişilere okudukları haberlere etkileşimli bir şekilde katkıda bulunma imkânı tanıdığı için bir iletişim aracı olarak kullanılmasının yanı sıra bir haber kaynağı olarak da sıkça kullanılmaktadır [126, 127].

Twitter’ın yukarıda bahsedilen haber kaynağı özelliklerine sahip olmasının sonucu olarak 2008 yılında Mars Phoenix tarafından Mars’ta su belirtisi bulunduğunda NASA bu bilgiyi resmi Twitter hesabı üzerinden duyurmayı tercih etmiştir. Bu durum, haberi saatler sonra yayınlamak durumunda kalan birçok haber kuruluşunu oldukça şaşırtmıştır [100, 128]. Bahsedilen Resim 3.3’de gösterilmiştir.



Resim 3.2. NASA’nın resmi Twitter hesabı üzerinden yayınladığı haber tweeti

Dolayısıyla gerçek yaşamdan geçmiş ve güncel örnekler dikkate alınarak incelendiğinde, Twitter'ın bir sosyal medya ve iletişim aracı olarak kullanılmasının yanı sıra ülkemizde ve dünyada bir haber kaynağı olarak da sıklıkla kullanıldığı ve haber kaynağı değerinin her geçen gün biraz daha arttığı gözlemlenmektedir.

Hızlı yayılım özelliğine sahip olma: Twitter'ın bir haber kaynağı olarak kullanılmasının bir diğer sebebi de retweet özelliği ile bilgiyi yönlendirme imkânı sunmasıdır [129]. Haber değeri taşıyan bir tweet kullanıcılar tarafından fazla bir çaba sarf etmeden yalnızca 'retweet' bağlantısına tıklanarak kendi sayfalarında görüntülenmektedir. Dolayısıyla haber hızlı bir şekilde paylaşılmasının yanı sıra, aynı zamanda kısa süre içerisinde daha çok kullanıcıya ulaşmaktadır. Aynı zamanda herhangi bir tweeti retweet eden kişilerin takipçi sayıları da haberin yayılımında büyük rol oynamaktadır. Retweet eden kişinin takipçi sayısı ne kadar fazla ise haberin yayılma oranı da o kadar yüksektir. Fakat takipçi sayısı az olan bir kişi retweet etse bile takipçileri arasında takipçi sayısı fazla olan kullanıcıların olması, haberin yayılma olasılığını yine artırmaktadır. Dolayısıyla retweet özelliği her türlü haberin yayılım oranını artırdığı için Twitter'ı vazgeçilmez bir haber kaynağı haline getirmiştir ve Twitter'ın bir haber kaynağı olarak değerlendirilmesinde büyük rol oynamaktadır.

Resmi haber kuruluşları tarafından kabul görmüş olma: Bütün bunların yanı sıra, günümüzde Twitter resmi bir haber kaynağı olarak da kullanılmaktadır. Haber kuruluşlarının televizyon ve radyo gibi geleneksel medya olarak adlandırılan kanallarının yanı sıra artık internet ortamından erişilen web siteleri ve sosyal ağ hesapları bulunmaktadır. Bu sosyal ağ hesapları arasından Twitter ise manşet uzunluğundaki karakter sınırlaması ve bir mikroblog ağı olmasından dolayı haber kuruluşlarının amaçlarına en uygun yapıya sahiptir [130]. Ülkemizde ve dünyada Twitter üzerinden haber paylaşımı yapan birçok haber kuruluşu mevcuttur. CNN, NTV ve BBC küresel çapta ve aynı zamanda ülkemiz için ulusal çapta yayın yapan ve resmi Twitter hesabına sahip olan haber kuruluşları arasında gösterilebilmektedir [131-133]. Ülkemiz sınırları içerisinde ulusal çapta resmi Twitter hesaplarına sahip olan haber kuruluşları ise, TGRT, Habertürk, Anadolu Ajansı, Zaman Gazetesi, Türkiye Gazetesi, Samanyolu Haber, Radikal Gazetesi, Sabah Gazetesi, Hürriyet Gazetesi vb. olarak örneklendirilebilmektedir. Dolayısıyla haber kuruluşlarının da bizzat kendilerine ait ve onaylanmış Twitter hesaplarının olması, Twitter'ın sosyal ağ olmasının yanı sıra önemli bir haber kaynağı olarak da değerlendirilebileceğini gösterir niteliktedir.

Gazeteciler tarafından kullanılıyor olma: Bahsedilen bu durumu kanıtlayan bir diğer örnek ise resmi haber kuruluşlarında çalışan gazetecilerin de Twitter'ı haber paylaşmak için uygun bir ortam olarak görüyor olmalarıdır. Gazeteciler sadece haber kuruluşlarının Twitter hesaplarından değil, kendi kişisel hesaplarından da gazeteci kimlikleri ile köşe yazıları veya ülke gündemiyle ilgili haber paylaşımı yapmaktadırlar [100]. Gazetecilerin kişisel hesaplarından paylaşılan bu haberler ise geleneksel haber standardının dışında daha çok kişisel görüşleri de içinde barındırdığı için Twitter kullanıcılarının ilgisini daha çok çekmektedir [134]. Ayrıca bu şekilde bir haberin sahip olması gereken en önemli özelliklerden biri olan şeffaflık prensibi de sağlandığı için Twitter kullanıcıları tarafından gazetecilerin kişisel hesapları oldukça yakından takip edilmektedir [100, 130].

Bunlara ek olarak, günümüzde haber kuruluşları ve gazeteciler arasında herhangi bir haberi en hızlı ve en kısa sürede yayınlama rekabetinin olduğu bilinmektedir. Bu bağlamda Twitter bir mikroblog ağı olması, daha çok mobil platformlara hitap etmesi ve internet üzerindeki kısa mesaj hizmeti gibi algılanması özelliklerinden dolayı hızlı ve kısa sürede paylaşım yapma imkânı sunmaktadır. Bu nedenle haber kuruluşları ve gazetecilerin en çok tercih ettikleri haber paylaşım platformu haline gelmiştir [135].

Yukarıda Twitter mikroblog ağına haber kaynağı olma özelliği kazandıran durum ve şartlar 7 başlık altında ele alınmıştır. Bunlar arasından, Twitter'ın resmi haber kuruluşları tarafından kabul görerek bu kuruluşlara ait resmi hesapları bünyesinde barındırması bu tez çalışmasında Twitter'daki haberlerin kullanılmasında temel oluşturan özelliğidir. Daha öncede belirtildiği gibi, bu tez çalışmasında ülkemizde yayın yapan resmi haber kuruluşları ve gazetecilere ait Twitter hesapları üzerinden yayınlanan tweetler kullanılmıştır.

4. LİTERATÜR TARAMASI

Günümüzde sosyal veri olarak adlandırılan sosyal ağ verisinin önem kazanmasıyla birlikte metin formatında olan sosyal ağ verilerini konularına göre sınıflandırarak bu paylaşımların ilgili kategorilerde ele alınması üzerine birçok çalışma yapılmıştır. Bu çalışmalar, sosyal ağlar üzerinden yayınlanan haberlerin, yorumların ve diğer tüm metin formatındaki paylaşımların metin madenciliği yöntemleri ile sınıflandırılması işlemini gerçekleştirmiştir. Böylelikle ülkelerin gündeminin zamana bağlı olarak tespit edilmesi, gündem olabilecek konuların tahmini, en çok okunan haber konularının belirlenmesi ve herhangi bir marka veya ürün hakkında kullanıcı duygu ve düşüncelerinin tespit edilmesi gibi probleme çözüm sunulmaktadır.

Sosyal ağ kullanıcılarının ilgi alanlarının ve örüntülerinin çıkarılması, sosyal ağ gruplarının analiz edilmesi, sosyal ağlar üzerinde mevcut olan bot gibi zararlı hesapların tespit edilmesi ve kullanıcının yaptığı paylaşımlardan paylaşım sahibinin tanınarak tespit edilmesi de bu kapsamda çözüme kavuşan problemler arasında yer almaktadır. Bu bölümde literatürde yer alan ve metin madenciliği yöntemlerini kullanarak yukarıda bahsedilen amaçlarla yürütülmüş olan çalışmalar değerlendirilmiştir. Bu kapsamda sosyal ağlar üzerinden yayınlanmış haber içeriklerinin sınıflandırması üzerine yapılmış çalışmalar ise yok denecek kadar azdır.

Sosyal paylaşım ağları üzerinden yayınlanmış haber içeriklerinin sınıflandırılması üzerine yapılan bir çalışma Dilrukshi ve arkadaşları tarafından gerçekleştirilmiştir [101]. Bu çalışmada Sri Lanka'da bulunan haber kuruluşlarının, bir mikroblog sitesi olan Twitter üzerinden yayınladıkları haberler ele alınmıştır. Bu haberler konularına göre 12 farklı kategori (savaş-terörizm, ekonomi-iş, sağlık, spor, devlet, politika, trafik, eğlence, felaket, eğitim, toplum ve uluslararası) altında sınıflandırılarak, herhangi bir zaman aralığında belirlenen ülkedeki en popüler haber içeriği ve gruplarının neler olduğu tespit edilebilmektedir. Ayrıca bir ülkenin sosyal ağlar üzerinden yapılan paylaşımlarla mevcut durumunun ne olduğu hakkında bilgi edinilmesi de bu çalışmanın amaçları arasındadır. Resmi haber kuruluşlarının Twitter üzerinden yayınladıkları haber içerikleri toplanarak, 140 karakterlik kısa mesaj formatında olan bu metinler için kelime torbası modeli (bag of words model) ile metni oluşturan kelimeler öznitelik olarak seçilmiştir. Daha sonra elde edilen yüksek boyuttaki veri için kelime bazında frekans analizi ile boyut indirgeme işlemi

gerçekleştirilmiştir. Veri seti bu şekilde oluşturulduktan sonra yüksek boyuttaki veriler için etkili bir şekilde çalışabilen denetimli öğrenme tekniklerinden biri olan Destek Vektör Makineleri sınıflandırma işlemi için kullanılmıştır. Sonuç olarak DVM ile eğitilen bu sistem test edildikten sonra haber sınıflandırma işlemindeki başarısının her bir haber kategorisi için %90'ın üzerinde olduğu görülmüştür.

Twitter üzerinden yayınlanan haber içeriklerinin sınıflandırılması üzerine yapılan bir diğer çalışma da Wattanavaekin ve arkadaşları tarafından gerçekleştirilmiştir [136]. Çalışmada, Twitter Uygulama Programlama Arayüzü yardımıyla Twitter üzerinden yayınlan belirli sayıda haber tweeti toplanmıştır. Daha sonra, toplanan tweetler bağlantıların elenmesi, hashtaglerin silinmesi, durak kelimelerin ayrıştırılması ve bu işlemler sonucunda elde edilen kelimelerin eklerinden arındırılarak köklerinin elde edilmesi gibi ön işleme metotlarına tabi tutulmuştur. Son olarak elde edilen kelimeler ile Destek Vektör Makineleri, Naive Bayes ve Çok Terimli Naive Bayes modelleri eğitilerek Twitter haberleri, önceden belirlenmiş olan 'Politika', 'Eğlence', 'Spor', 'Teknoloji' ve 'İş' kategorileri altında sınıflandırılmıştır. Kullanılan 3 sınıflandırıcıdan elde edilen sonuçlar kıyaslandığında ise %80,1'lik başarı oranıyla en etkili sınıflandırıcının Çok Terimli Naive Bayes modeli olduğu görülmüştür.

Sosyal ağ verilerinin metin madenciliği yöntemleri ile sınıflandırılarak kullanıcıların herhangi bir marka, ürün, olay veya durum hakkındaki duygularının tespit edilmesi amacıyla yapılan çalışmalardan biri Tseng ve arkadaşları tarafından yürütülmüştür [137]. Tseng ve arkadaşları, Twitter üzerinden yapılan paylaşımların genel olarak pozitif, negatif, öznel ve nesnel olarak sınıflandırılması üzerine çalışmışlardır. Çalışma kapsamında mobil platformlar için geliştirilen yazılım ile son kullanıcı arama yapmak istediği kelimeyi geliştirilen arama motoruna girdi olarak sunmaktadır. Girilen kelimeyi içeren tweetler geliştirilen sistem tarafından ayrıştırıldıktan sonra bu paylaşımlar, içerdikleri bilgi kapsamında sınıflandırıcı tarafından pozitif, negatif, öznel ve nesnel olarak sınıflandırılmaktadır. Sınıflandırıcı olarak Naive Bayes algoritmasının kullanıldığı bu çalışma ile kullanıcılar, arama yapmak istedikleri konu hakkında Twitter kullanıcılarının düşüncelerinin ne yönde olduğu konusunda fikir sahibi olabilmektedirler. Kullanıcıların Twitter sosyal paylaşım ağı üzerinde tartışılan en güncel konuların ne olduğu konusunda herhangi bir çaba sarf etmeden otomatik olarak bilgi sahibi olabilmeleri de bu çalışmanın katkıları arasında yer almaktadır.

Medevet ve Bartoli tarafından yapılan bir çalışmada ise 3 önemli teknoloji markası hakkında yapılan Twitter paylaşımlarının tespit edilmesi, markaların ve markalar hakkındaki olayların popülerliklerinin değerlendirilmesi, markalar ve olaylar hakkında Twitter kullanıcılarının fikir kutuplarının pozitif veya negatif olarak belirlenmesi ve bu paylaşımların özetlenmesi üzerinde çalışılmıştır [138]. Bu kapsamda Apple, Google ve Microsoft markalarını tanımlamada yardımcı olacak anahtar kelimeler belirlenmiştir. Bu kelimelerden en az iki tanesini içerdiği tespit edilen olaylar popülerliklerine göre düşük, orta ve yüksek olacak şekilde derecelendirilmiştir. Daha sonra derecelendirilen tweetler içerdikleri duygulara göre Destek Vektör Makineleri sınıflandırıcısı ile pozitif, negatif ve nötr olarak sınıflandırılmıştır. En son ise belirlenen bir olayı en iyi ifade edecek olan tweetler Manhattan uzaklığı (Manhattan distance) yöntemi ile seçilerek özetleme işlemi gerçekleştirilmiştir. Elde edilen sonuçlara göre Google hakkındaki ürün ve olayların Microsoft ve Apple'a göre daha olumlu duygular içerdiği yönündedir.

Mostafa tarafından yapılan bir çalışmada da Twitter verileri markalar hakkındaki duyguların sınıflandırılması amacıyla kullanılmıştır [139]. 3516 tweet kullanılarak 16 marka hakkındaki duygu ve düşüncelerin sınıflandırıldığı çalışmada sözcük tabanlı yöntem (lexicon-based method) kullanılarak veri sözlüğünde bulunan kelimelerin frekans ağırlıkları hesaplanmıştır. Elde edilen sonuçlar kıyaslandığında, Twitter kullanıcılarının Nokia hakkındaki görüşlerinin Pfizer markasına göre daha olumlu olduğu yönündedir. Ayrıca Lufthansa ve DHL markaları hakkındaki görüşlerin çoğu olumlu iken T-Mobile hakkındaki görüşlerin çoğunun olumsuz olduğu görülmüştür.

Literatürde sosyal ağlar üzerindeki zararlı ve kötü niyetli hesapların yayınladıkları içeriklerin sınıflandırılması amacıyla yürütülmüş çalışmalar da mevcuttur. Buna örnek olarak verilebilecek olan Cingiz ve Diri tarafından yapılan bir çalışmada Twitter ağı üzerinden paylaşım yapan normal ve bot kullanıcıların tweetleri ekonomi, spor, eğlence ve teknoloji olmak üzere 4 farklı kategori altında sınıflandırılmıştır [104]. Sınıflandırıcı olarak, Çok Terimli Naive Bayes ve Destek Vektör Makineleri 'nin kullanıldığı çalışmada eğitim için BBC, CNN ve SKYNews gibi haber sağlayıcılarının bahsedilen 4 kategoriye ait toplamda 2105 RSS besleme verisi kullanılmıştır. Test amacıyla ise, wefollow.com aracılığıyla kategorisi belirli olan 26 bot kullanıcı ve 27 normal kullanıcıya ait tweetler kullanılmıştır [140]. Eğitim ve test verileri ön işleme ve boyut indirgeme gibi işlemlerden geçirildikten sonra yapılan sınıflandırma işleminde Çok Terimli Naive Bayes

sınıflandırıcısının DVM'ne göre daha iyi sonuçlar ürettiği gözlemlenmiştir. Yapılan kategorizasyon işlemi sonucunda bot kullanıcıların normal kullanıcılara göre daha kategorik tweetler yayınladığı görülmüştür.

Mevcut literatürde yaygın olmamakla birlikte ülke gündemini tespit etmek amacıyla sosyal ağ verisini zamana bağlı olarak sınıflandırma işlemini gerçekleştiren bir çalışma Kamath ve Caverlee tarafından gerçekleştirilmiştir [141]. Bu çalışmada metin madenciliği yöntemlerinden farklı olarak sosyal ağ paylaşımlarını sınıflandırmak için zamana bağlı adaptif bir konu sınıflandırma sistemi geliştirilmiştir [141]. Geliştirilen sistemin temel özellikleri, eğitim seti olarak her bir kategori için (politika, teknoloji, spor ve eğlence) uzmanlara ait sosyal ağ verilerinin kullanması ve zamana bağlı sınıflandırma yaparak zaman bazında sosyal medyadaki konu değişimlerini gözlemlemesi olarak ifade edilmiştir. Maximum Entropi sınıflandırıcısının kullanıldığı çalışmada, konuların zamana bağlı kategorize edilebilmesi ve gün ve saat bazında değişimlerin incelenebilmesi için sınıflandırıcı kayan pencere (sliding window) mantığı ile zamansal olarak eğitilmiştir. Sonuç olarak kayan pencere boyutunun büyüklüğünün 1 ile 8 günlük eğitimler için sistem performansını etkilediği ve ne kadar büyük olursa o kadar iyi sonuçlar üretilebildiği görülmüştür. Fakat bunun yanı sıra 8 ile 14 günlük eğitimler için bu boyutun büyüklüğünün sınıflandırma başarısını pek fazla etkilemediği gözlemlenmiştir.

Literatürde sosyal ağ paylaşımlarının yanı sıra bu paylaşımlara yapılan yorumları kategorize eden çalışmalar da bulunmaktadır. Bu kapsamda Santos ve arkadaşları tarafından yapılan bir çalışmada ünlü bir İspanyol haber sosyal ağı olan Meneame verileri kullanılmıştır [142]. Bu sosyal ağ üzerinden haber paylaşımlarına yapılan yorumlar; yorumun baz aldığı kaynak (haber veya başka bir yorum), içerdiği bilgilerin türü (katkı yorumu, ilgisiz yorum ve düşünce yorumu) ve tartışma derecesi (normal, tartışmalı, tartışma derecesi yüksek) olmak üzere 3 farklı öge bazında sınıflandırılmıştır. Çalışmanın amacı yapılan yorumların sınıflandırılması ile bu yorumları yapan kullanıcıların da kategorize edilerek ilgi alanlarının belirlenmesidir. Bu kapsamda sınıflandırma işleminde kullanılmak üzere istatistiksel, sözdizimsel ve yorumun yansıttığı fikir bazında özellik çıkarımı işlemi gerçekleştirilmiştir. Sınıflandırma işlemi için ise Naive Bayes, Destek Vektör Makineleri, K-En Yakın Komşu (K-Nearest Neighbor) ve Karar Ağaçları sınıflandırıcıları kullanılmıştır. Elde edilen sonuçlar yorumun gövdesine göre yapılan sınıflandırmada kelime tabanlı yaklaşımın n-gram tabanlı yaklaşıma göre daha iyi sonuçlar

ürettiği yönündedir. Fakat yorumun tartışma seviyesine göre sınıflandırma yapıldığında Rastgele Orman (Random Forest) algoritmasının n-gram tabanlı yaklaşımla daha iyi sonuçlar ürettiği gözlemlenmiştir.

Twitter paylaşımları üzerinden kullanıcıların kimliklerinin belirlenmesi amacıyla yapılan bir çalışma Keretna ve arkadaşları tarafından gerçekleştirilmiştir [143]. Sosyal paylaşım ağlarının kullanımı arttıkça ortaya çıkan sahte hesap ve kimlik hırsızlığı problemlerine çözüm sunmayı amaçlayan bu çalışma, Twitter paylaşımlarının sahibinin kim olduğunu tahmin etmektedir. Bu tespit işleminde metin madenciliği yöntemleri ile kullanılan sıfat sayısı, fiil sayısı, en sık kullanılan kelimeler gibi kişilerin yazım stilleri konusunda belirleyici olduğu düşünülen özellikler belirlenmiştir. Daha sonra eğitilmiş olan sistem Jaccard katsayı yöntemi ile test verisi için benzerlik analizi yaparak yapılan paylaşımın sahibinin kim olduğunu tahmin etmiştir. 1000 tweet içeren veri setiyle yapılan çalışma sonucunda elde edilen sonuçların umut verici nitelikte olduğu ve daha büyük veri setleri için daha iyi sonuçlar elde edilebileceği vurgulanmıştır.

Huang ve arkadaşları tarafından yapılan bir çalışmada ise diğerlerinden farklı olarak sınıflandırma amacıyla video paylaşım ağı verileri kullanılmıştır [144]. Böylelikle video içerikleri, videolara ait olan başlıklar, açıklamalar ve yorumlar gibi metin içerikleri kullanılarak sınıflandırılmıştır. Metinlere ait sözcüksel, sözdizimsel ve içerik odaklı özelliklerin çıkarıldığı çalışmada sınıflandırma işlemi için C4.5, Naive Bayes ve Destek Vektör Makineleri kullanılmıştır. Elde edilen sonuçlar önerilen yöntem ile video sınıflandırma işleminin başarısının %87,2'lere ulaşabildiği yönündedir. Ayrıca DVM sınıflandırıcısının C4.5 ve Naive Bayes yöntemlerine göre sınıflandırma işlemini daha yüksek bir performansla gerçekleştirdiği görülmüştür. Son olarak bu çalışma ile video paylaşım ağlarındaki siber toplulukların ilgi alanları ve paylaşımlarına göre rahatlıkla tespit edilebileceği ve kullanıcı örüntüsünün çıkarılabileceği gözlemlenmiştir.

Sriram ve arkadaşları tarafından yapılan bir çalışmada, Twitter verileri genel olarak Haber, Olay, Düşünce, Alışveriş-Ticaret ve Özel Mesaj başlıkları altında sınıflandırılmıştır [145]. Bu çalışmada tweetlerde var olan karakter kısıtlamasından dolayı sınıflandırma işlemi sırasında kelime frekansının yanı sıra önerilen 8 farklı özellik de değerlendirmeye tabi tutulmuştur. Bu özellikler metnin yazarı, kısaltılmış veya argo kelimelerin bulunup bulunmaması, zaman-olay deyimlerinin varlığı, düşünce ifade eden kelimelerin varlığı,

kelimelerdeki vurgular, para birimi ve miktarı ifadeleri, tweetin başka bir tweete cevap olup olmaması ve başka kişilerden bahsedip bahsetmemesi durumları olarak sıralanmaktadır. Bu özellikler dikkate alınarak yapılan sınıflandırma işleminde Naive Bayes sınıflandırıcısı kullanılmıştır. Elde edilen sonuçlar, kelime frekansı ve önerilen 8 özelliğin değerlendirilmesiyle yapılan sınıflandırma işleminin sadece kelime frekansı yöntemi kullanılarak yapılan sınıflandırma işlemine göre sırasıyla her bir kategori için (Haber, Olay, Düşünce, Alışveriş-Ticaret ve Özel Mesaj) %35,2, %103,4, %12,2, %9,9 ve %87 daha başarılı sonuçlar ürettiği yönündedir.

Mevcut literatürde, genel olarak kişi veya grupların sosyal medya üzerinden konuştuğu konuların ve yaptığı paylaşımların konu bazında sınıflandırmasını yapan çalışmalara yukarıda değinilmiştir.

Horn tarafından yapılan tez çalışmasında da tweetler genel olarak haber, kullanıcı ve şirket, tweetleri olmak üzere 3 başlık altında sınıflandırılmıştır [106]. Destek Vektör Makinelerinin kullanıldığı çalışmada duygu tespiti yapan bir modül de geliştirilmiştir. Bu modül de olay ve fikir-düşünce tweetlerini objektif ve sübjektif olarak sınıflandırmaktadır. Tez çalışması kapsamında oluşturulan veri seti ile elde edilen sınıflandırma işleminin başarısının %80'den fazla olduğu tespit edilmiştir. Ayrıca; N-gram modeli, eğitim setinin boyutu, kelimelerin dönüştürülmesi (word conversion) ve köklerine ayrıştırılması işlemlerinin sistem başarısı üzerindeki etkisi analiz edilmiştir. Elde edilen sonuçlar; N-gram modeli ve kelime köklerinin belirlenmesi işlemlerinin sınıflandırıcı başarısı üzerinde etkisi olmadığını fakat kelime dönüştürme işleminin sistem başarısını etkilediğini göstermiştir. Tezin çıktısı, Twitter kullanıcılarının kullanımına hazır bir web uygulaması olarak sunulurken oluşturulan veri setinin de paylaşılacağı bildirilmiştir.

Bhattacharya ve arkadaşları tarafından yapılan bir çalışmada ise yukarıda bahsedilen çalışmalardan farklı olarak daha spesifik konuları tartışan ve jeoloji, nöroloji, astrofizik ve karate gibi kendi alanlarında özelleşmiş konularda paylaşım yapan grupların ağ yapıları ve tweet paylaşma özellikleri analiz edilmiştir [146]. Çalışmanın katkısı; Twitter'da belirli bir konu hakkında görüş belirten grupların uzmanları, yetkili kullanıcıları ve takipçilerin tespit edilmesi ve bu grupların takip etme, takip edilme ve tweet paylaşma davranışlarının analiz edilmesi olarak sıralanmıştır. Uzman tespiti işleminde, grupların isimleri ve tanımları kısmında en sık kullanılan kelimeler belirlenerek grubun takipçileri arasından bu

kelimeleri, paylaştığı tweetlerde en çok kullanan kişiler uzman olarak değerlendirilmiştir. Herhangi bir konu ile ilgili olan kişilerin tespitinde ise o konuya ait en az 3 uzmanı takip eden kişiler belirlenen konu ile ilgili kişiler olarak nitelendirilmiştir. Sonuç olarak, çalışma kapsamında tweetleri toplanan 38,4 milyon kişiden 584.759 kişinin herhangi bir konuda uzman olduğu görülmüştür. Ayrıca uzmanların çoğunun takipçi olarak da nitelendirilebileceği elde edilen sonuçlar arasındadır.

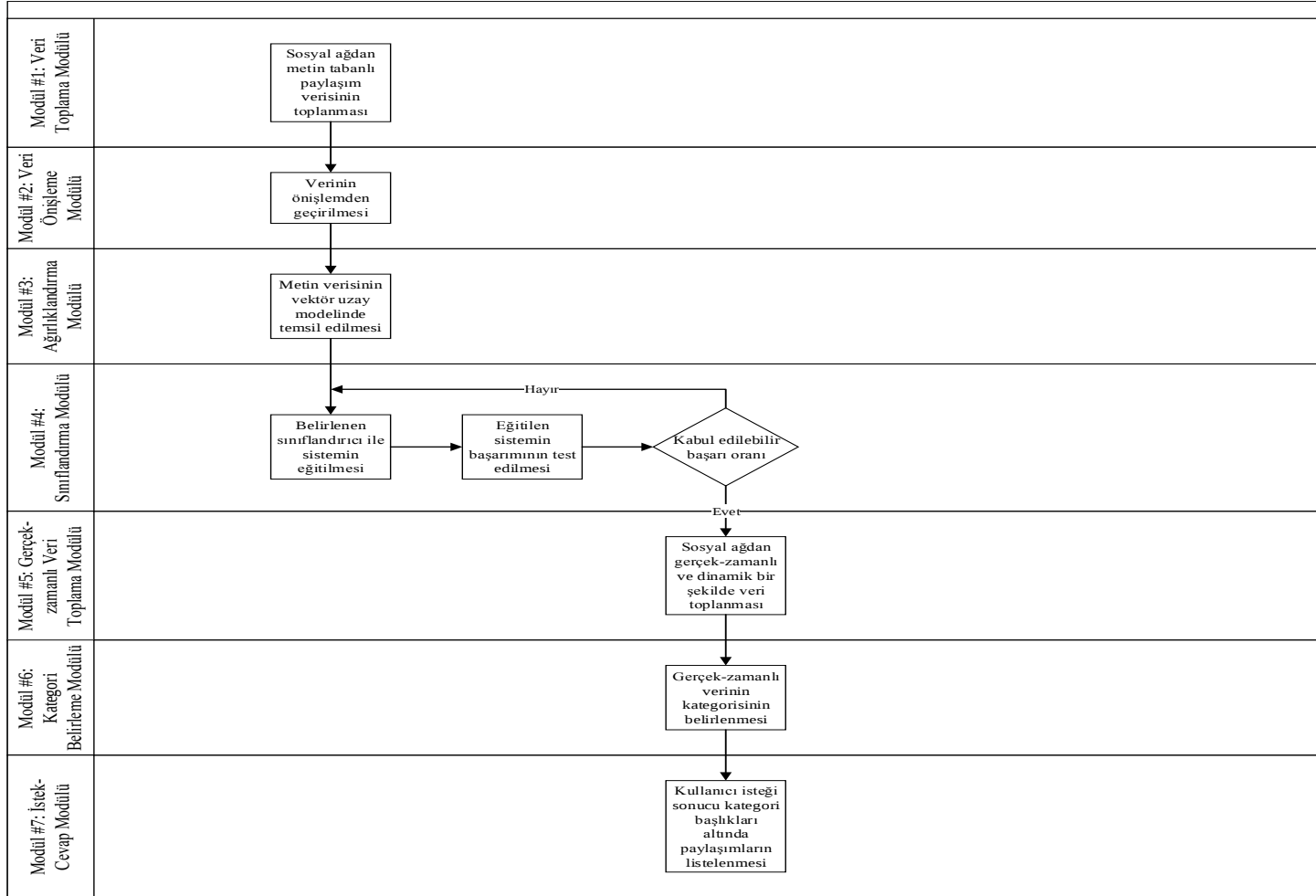
Bu tez çalışmasında ise, literatüre bir katkı niteliğinde metin tabanlı ve haber içerikli sosyal ağ paylaşımlarının nasıl sınıflandırılabileceğini gösteren bir yaklaşım önerilmiş, literatürde bu alanda mevcut bulunan diğer yöntemler geliştirilmiş ve işlemleri otomatik gerçekleştiren bir sistem sunulmuştur. Ayrıca literatüre bir katkı niteliğinde, ilk defa olarak Türkçe metin tabanlı ve haber içerikli Twitter paylaşımlarını dinamik ve gerçek zamanlı bir şekilde toplayan ve kabul görmüş 7 farklı haber kategorisi başlığı altında sınıflandırarak görsel bir şekilde internet kullanıcılarına sunan ve tweetajans.com alan adı üzerinde çalışan bir yazılım platformu geliştirilmiştir.

5. SOSYAL AĞLARDAKİ HABER PAYLAŞIMLARININ SINIFLANDIRILMASI

Bu tez çalışmasının amacı, günümüzde özellikle haber kaynağı olarak kullanılan sosyal ağlar üzerinden yayınlanan ve özellikle haber niteliği taşıyan sosyal ağ paylaşımlarını sahip oldukları içeriklere göre sınıflandırarak internet kullanıcılarının kaliteli bilgiye daha kolay erişim yapabilmelerini sağlamaktır. Bu çerçevede sosyal ağlar üzerinden yayınlanan ve haber içeriğine sahip olan paylaşımların nasıl sınıflandırılabileceğini gösteren bir yaklaşım sunulmuştur. Ayrıca önerilen bu yaklaşımı kullanarak, haber kuruluşlarının sosyal ağlar üzerinden sürekli olarak yaptıkları haber paylaşımlarını dinamik ve gerçek zamanlı bir şekilde sınıflandıran ve sınıflandırılan veriyi görsel bir şekilde internet kullanıcılarının hizmetine sunan bir web sitesi yazılımı geliştirilmiştir. Bu tez kapsamında önerilen yaklaşım ile bu yaklaşıma ait modüller ve bu modüllerin çalışma mekanizması ile tasarlanmış web tabanlı yazılım sonraki bölümlerde detaylı bir şekilde açıklanmıştır.

5.1. Önerilen Yaklaşım

Bu tez çalışmasında önerilen yaklaşım ile internet kullanıcılarının sosyal ağlar üzerinden yaptıkları metin paylaşımları toplanabilmektedir. Bu paylaşımlar içerdikleri bilginin türüne göre farklı kategori başlıkları altında sınıflandırılarak sistemin başarımı test edilmektedir. Bu çalışma kapsamında yapılan literatür çalışması sonucunda Türkçe sosyal ağ paylaşımlarının sınıflandırılmasına dair herhangi bir çalışma bulunmamasına karşın, İngilizce sosyal ağ paylaşımlarının sınıflandırıldığı çalışmalarla kıyaslandığında sistemin sınıflandırma performansının yüksek olduğu görülmüştür. Bunun sonucu olarak sürekli artan, akan ve dinamik yapıda olan ve haber kuruluşları tarafından sosyal ağlar üzerinden paylaşılan Türkçe metin tabanlı haber paylaşımları, içerdikleri haber bilgisinin kategorisine göre sınıflandırılarak bir web sitesi yazılımı üzerinden görsel bir şekilde, sosyal medya haberlerini takip etmek isteyen internet kullanıcılarının hizmetine sunulmuştur. Geliştirilen yaklaşıma ait akış diyagramı Şekil 5.1’de sunulmuş olup detayları aşağıda verilmiştir.



Şekil 5.1. Önerilen yaklaşıma ait akış diyagramı

Şekil 5.1’de verilen akış diyagramındaki modüller ve bu modüllerin görev ve içerikleri aşağıda özetlenmiştir.

Modül #1: Sistem başarımının testi için sosyal ağ üzerinden belirli sayıdaki metin paylaşımı verisinin toplanması

Modül #2: Metin içerisindeki anlamlı kelimelerin belirlenebilmesi için nitelikli olmayan içeriklerin elenmesi

Modül #3: Kelimelerin önceden belirli olan her bir kategori için önem derecesinin tespit edilebilmesi için sayısallaştırılıp vektör uzay modelinde ifade edilmesi

Modül #4: Test verisinin sınıflandırılması ve sistemin başarımlı testi

Modül #5: Sürekli güncellenen ve yenilenen sosyal ağ verisinin dinamik ve gerçek-zamanlı bir şekilde sürekli olarak depolanması

Modül #6: Gerçek-zamanlı bir şekilde depolanan dinamik ve akan verinin ait olduğu kategorinin bulunması

Modül #7: Sınıfı belirlenmiş olan sosyal ağ paylaşımının web sayfası üzerinden kullanıcının istem yaptığı kategori başlıkları altında sunulması

Bir metnin önceden belirli olan kategorilerden hangisine ait olduğunu belirlenebilmesi için o metin içerisinde geçen kelimelerin her bir kategori için önem derecesinin belirlenmesi gerekmektedir. Bu nedenle eğitim öncesinde, test edilecek verinin diline ait olmayan karakterler, harf grupları ve kelimeler elenmelidir. Ayrıca herhangi bir kategori için özel bir anlamı olmayan veya her kategori için eşit derecede önem taşıyan kelimelerin elenmesi de sınıflandırma performansını ve başarımlı artıracaktır. Daha sonra, belirlenen kelimeler her bir sınıf için ayırt edici olma özelliklerine göre ağırlıklandırılarak sayısal hale getirilir. Böylece sistem, kelimeler ve kategoriler arasındaki sayısal ilişkiyi öğrenerek test verilerini ait oldukları en uygun sınıfa atayacaktır.

Bu tez çalışmasında sunulan yöntem ile geliştirilen sistem, kelimeler ve kategoriler arasındaki ilişkiyi öğrendikten sonra öğrendiği bu örüntü ve ilişki bilgilerini kullanarak sosyal ağlar üzerinden sürekli ve dinamik olarak paylaşılan veriyi sınıflandırmaktadır. Bu çerçevede verinin içerdiği kelimelerin her bir sınıf için görülme sıklığını modellemekte ve metin tabanlı sosyal ağ paylaşımının ait olduğu kategoriyi belirlemektedir. Bu sayede özellikle büyük veri boyutundaki sosyal verinin sınıflandırılması yapılarak internet kullanıcılarının sosyal medya üzerindeki bilgilere daha kısa sürede ve daha kolay bir şekilde erişmesi sağlanmaktadır. Önerilen bu yaklaşımı kullanarak geliştirilen web sitesi

yazılımında ise Türkiye’de varlık gösteren haber kuruluşları ve gazetecilerin resmi Twitter hesapları üzerinden yayınladıkları metin tabanlı haber paylaşımları sınıflandırılarak, bu paylaşımlar ait oldukları haber kategorilerine atanmaktadır. Daha sonra kullanıcının web sitesi üzerinden istem yaptığı kategoriye ait Twitter haberleri görsel bir şekilde kullanıcıya sunulmaktadır. Böylelikle sosyal medya üzerinden yayınlanan haberler, ait oldukları kategorilere göre filtrelenerek sosyal medya hesabı olmayan internet kullanıcılarının dahi haber paylaşımlarına çok daha kısa sürede ve daha kolay bir şekilde erişim yapmalarına imkân tanınmaktadır.

5.1.1. Geliştirilen modüller

Aşağıdaki alt başlıklarda Şekil 5.1’de verilen akış diyagramındaki modüller detaylı bir şekilde açıklanmıştır.

Modül #1: Veri toplama modülü

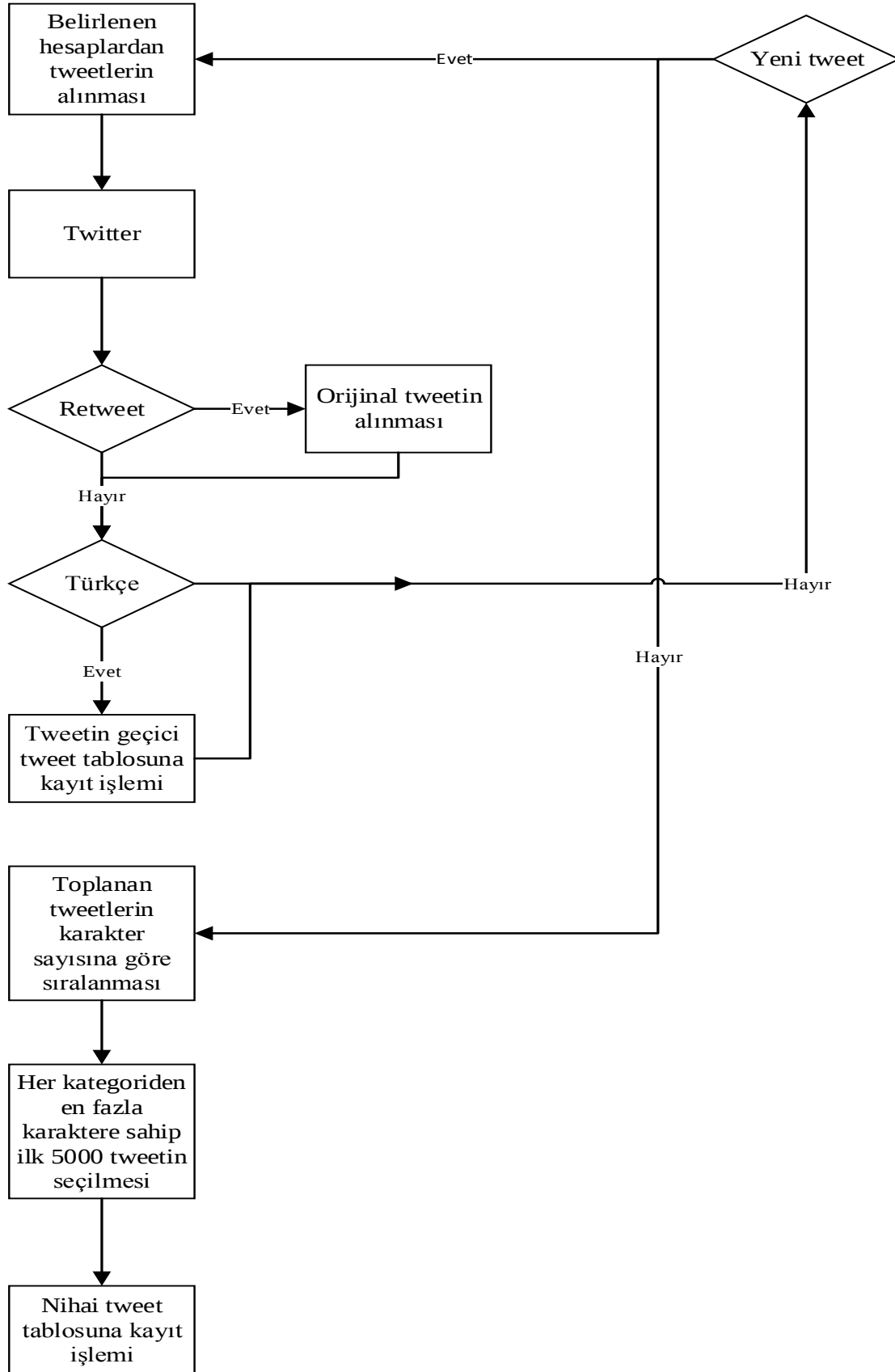
Bu modül ile sistemin sınıflandırma başarısının ölçülmesi amacıyla gerekli olan belirli sayıdaki eğitim ve test verisi belirlenen sosyal ağ üzerinden toplamaktadır.

Bu tez çalışması kapsamında önerilen ve bütün sosyal ağlar için geçerli olan yaklaşımın uygulanabilirliğini göstermek amacıyla geliştirilen yazılımda, önceki bölümlerde özellikleriyle birlikte detaylı bir şekilde açıklanan Twitter sosyal ağına ait metin paylaşımlarının sınıflandırması yapılmıştır. Bu nedenle bu modülde veri toplama amacıyla Twitter’ın geliştiricilere sunduğu Twitter REST UPA hizmeti kullanılmıştır.

Twitter REST UPA

Twitter REST UPA, geliştiricilere belirledikleri hesaplara ait geçmişe yönelik tweetleri, durum güncellemelerini, profil resimlerini, takip edilen ve takipçi ilişkileri gibi o hesaplara ait birçok bilgiyi elde etme imkânı sunan bir uygulama programlama ara yüzüdür [147, 148]. Twitter REST UPA ile geliştiriciler tek bir sorguda, bir hesaba ait son 200 tweeti çekebilmektedirler. Bunun yanı sıra Twitter REST UPA bir hesaba ait maksimum son 3200 tweet ’in çekilmesine imkân tanımaktadır. Dolayısıyla bir hesaba ait maksimum 3200 tweet ’in elde edilebilmesi en fazla 16 sorgu ile gerçekleştirilebilmektedir [148].

Bu çalışma kapsamında günümüzde özellikle haber kaynağı olarak kullanılmaya başlanan sosyal ağlar üzerinden yayınlanan haber paylaşımlarının sınıflandırılması amaçlanmıştır. Bu nedenle, geliştirilen yazılımda sistemin başarımını ölçmek için kullanılan veri, Twitter üzerinden haber paylaşımı yapan haber kuruluşlarının resmi Twitter hesapları üzerinden toplanmıştır. Sınıflandırma amacıyla ‘Teknoloji’, ‘Spor’, ‘Ekonomi’, ‘Politika’, ‘Magazin’, ‘Sağlık’ ve ‘Kültür-Sanat’ olmak üzere 7 kategori belirlenmiştir. Bu kategoriler belirlenirken haber kuruluşlarının resmi web sitelerinde kullandıkları kategoriler ve literatürde var olan diğer çalışmalarda kullanılan haber kategorileri göz önünde bulundurulmuştur [101, 136]. Bu bağlamda her bir kategori için verisi toplanan Twitter hesapları EK-2’de sunulmuştur. EK-2’de görüldüğü gibi belirlenen kategorileri için NTV, TGRT, Habertürk, Cumhuriyet, CNN TÜRK, Anadolu Ajans, Sabah, Radikal, Türkiye, Hürriyet ve Mynet gibi ülkemizde yayın yapan haber kuruluşlarının resmi Twitter hesapları veri kaynağı olarak kullanılmıştır. Şekil 5.2’de bu modülde gerçekleştirilen işlemlere ait akış diyagramı sunulmuştur.



Şekil 5.2. Modül 1'e ait akış diyagramı

Şekil 5.2'deki akış diyagramında gerçekleştirilen adımlar aşağıda detaylı bir şekilde açıklanmıştır.

Belirlenen hesaplardan tweetlerin alınması:

Verisi kullanılacak olan resmi Twitter hesapları EK-2'de görüldüğü şekilde belirlendikten sonra Twitter REST UPA'nın Java uygulamaları için var olan Twitter4J kütüphanesi yardımı ile bu hesaplardan tweet çekme işlemi başlatılmıştır [136]. Twitter REST UPA her bir hesap için o hesaptan atılan yalnızca son 3200 tweetin toplanmasına imkân tanımaktadır. Bu nedenle, tweet sayısı 3200'den az olan hesaplar için o hesaba ait bütün tweetler çekilebilirken, 3200'den fazla olan hesaplardan yalnızca son 3200 tweet çekilebilmiştir.

Retweet kontrolü:

Tweetlerin toplanması işlemi sırasında alınacak olan tweetin orijinal mi yoksa retweet mi olduğu bilgisi göz önünde bulundurulmuştur. Tweet orijinal ise direk olarak alınmış, retweet ise de tweetin orijinal hali alınmıştır. Bu şekilde retweet kontrolü yapılmasının sebebi ise Twitter'ın retweet olan paylaşımların başına RT ibaresi eklemesidir. Bu ibare Türkçe dili için anlamsız bir harf grubu olduğundan ön işleme sırasında elenmesi gerekecektir. Ayrıca karakter sınırında olan tweetlerin başına bu ibare eklendiğinde karakter sınırı aşılabacağından tweetin içerdiği kelimelerde harf kaybına yol açmaktadır. Dolayısıyla sistemin ön işleme sırasında performans kaybına uğramaması ve ayrıca kelime kaybına yol açılmaması amacıyla veri toplama aşamasında retweet olan tweetlerin orijinal halleri seçilmiştir.

Türkçe tweet kontrolü:

Bu çalışma kapsamında Türkçe metin tabanlı sosyal ağ paylaşımlarının sınıflandırılması işlemi gerçekleştirilmektedir. Bu nedenle veri toplama işlemi sırasında bahsedilen hesaplardan belirli sayıda tweetler çekilirken göz önünde bulundurulması gereken önemli bir husus da çekilen tweetlerin Türkçe olması gerektiğidir. Belirlenen hesaplardan atılan tweetler incelendiğinde özellikle 'Teknoloji' alanında yapılan haber paylaşımlarının İngilizce kelimeleri de içerdiği görülmüştür. Bunun sebebi ise henüz bu alanda bazı

kelimelerin Türkçe literatürde tam olarak karşılığının bulunmaması olarak değerlendirilmiştir. Diğer kategorilerde de böyle bir durumla karşılaşılabilceği göz önünde bulundurularak veri toplama işlemi sırasında tweetlerin içerdiği kelimelerin dilini tespit edebilen ve Java platformu için geliştirilmiş olan özel bir kütüphane kullanılmıştır. 53 farklı dili destekleyen ve %99 doğruluk oranı ile dil tespiti yapabilen 'langdetect.jar' kütüphanesi yardımı ile tweetlerin Türkçe olup olmadığı belirlenmiştir [149]. Bir tweetin Türkçe olduğu tespit edilmişse bu tweet veri olarak kabul edilmiş ve veritabanına kaydedilmiştir. Türkçe olmayan tweetler ise veri olarak kabul edilmeyip göz ardı edilmiştir.

Çekilen tweetler veritabanına kaydedilirken Tweet ID, Tweet İçeriği ve Tweet 'in ait olduğu kategori şeklinde kaydedilmiştir. Bu aşamada veritabanından bir kesit Çizelge 5.1'de sunulmuştur.

Çizelge 5.1. Modül #1'in işlevi sonucunda toplanan Twitter verisi

<i>Tweet_ID</i>	<i>Tweet</i>	<i>Kategori</i>
483533759606558720	Altın Koza'lı Gözümün Nuru Moskova'dan ödülle döndü http://t.co/LyyVoAagm5 @zamancomtr	7
483554012902346755	Danimarka'da doktor iki hastasının ölümüne sebep oldu http://t.co/bS49ITpbuz	5
483562867250065409	Çini ve seramiklerle kahvenin serüveni http://t.co/Jts796WYuO @zamancomtr aracılığıyla	7
483568552331313152	Tolstoy'un itirafları http://t.co/J5t9finLL5 @zamancomtr	7
483573283422347264	Dekorasyon sorununa etkili çözümler http://t.co/JTYfyM1quR	5
483582435473387520	Gupse Özay 'Deliha' ile beyazperdeye adım atıyor http://t.co/YeWUKCqM20	5
483592251860197376	Cemil İpekçi hastaneye kaldırıldı http://t.co/MYL246no8R	5
483601701954269184	Elif Batı'dan yeni kitap: Gökten Üç Aşk Düştü http://t.co/zjzcH8U2Ez	5

Çizelge 5.1'de görüldüğü gibi, Tweet ID her bir tweete özgü olan kimlik numarasını, Tweet içeriği tweetin kendisini, kategori ise bu tweetin içerdiği haberin hangi kategoriye ait olduğunu göstermektedir.

Bahsedilen şekilde Türkçe tweetler Twitter REST UPA'nın Java Platformu için geliştirilmiş olan kütüphaneleri yardımıyla çekilerek mysql veritabanında depolanmıştır. Veri toplama işlemi sona erdiğinde veritabanında her bir kategoriye ait olan toplam tweet sayısı Çizelge 5.2'de verildiği gibidir.

Çizelge 5.2. Veri toplama işlemi sonucunda her bir kategoriden alınan toplam tweet sayısı

Kategoriler	Toplam Tweet Sayısı
Teknoloji	6,590
Spor	70,976
Ekonomi	12,018
Politika	12,753
Magazin	10,256
Sağlık	12,811
Kültür-Sanat	10,200

Nihai tweetlerin belirlenmesi:

Çizelge 5.2'den görülebileceği gibi her bir kategoriden alınan tweet sayısı birbirinden bağımsızdır. Bu durum resmi Twitter hesaplarına sahip olan haber kuruluşlarının bazı kategorilerden daha çok paylaşım yaparken bazılarında daha az paylaşım yapmasından kaynaklanmaktadır. Ayrıca bir kategoride (örn. spor) Twitter üzerinden haber paylaşımı yapan çok fazla sayıda haber kuruluşu mevcutken başka bir kategoride (örn. Kültür-Sanat) Twitter üzerinden haber paylaşımı yapan daha az sayıda haber kuruluşunun bulunması da böyle bir tablonun ortaya çıkmasında etkilidir. Fakat bu durum sınıflandırma performansı ve başarımını olumsuz etkileyecek sebeplerden biridir. Çünkü metin sınıflandırma işleminde kelimelerin ağırlıklandırılması temel alındığından, her kategori için farklı sayıda tweet alınması kelime ağırlıklarını değiştirecek ve sınıflandırma işleminin dengesiz ve yanlış yapılmasına sebebiyet verecektir. Bu nedenle, her bir kategoride değerlendirilmesi gereken toplam tweet sayısı, en az veri içeren Teknoloji kategorisi göz önünde bulundurularak 5000'e düşürülmüştür. Bu eleme işlemi sırasında mysql veritabanında tutulan tweetler, tweet içeriklerinin sahip olduğu karakter sayısına göre sıralanarak en fazla karaktere sahip olan tweetler veritabanında kalacak şekilde her bir kategoriden en fazla karakter içeren 5000 tweet alınmıştır. Karakter sayısına göre eleme yapılmasının amacı daha fazla öznitelik içeren tweetleri kullanarak sistemin daha çeşitli bir veri kümesi ile

eğitim ve test edilmesini sağlamaktır. Sonuç olarak, veri toplama işlemi sona erdiğinde her bir kategoride 5000 tweet olacak şekilde toplamda 35000 tweet içeren bir veritabanı oluşturulmuştur.

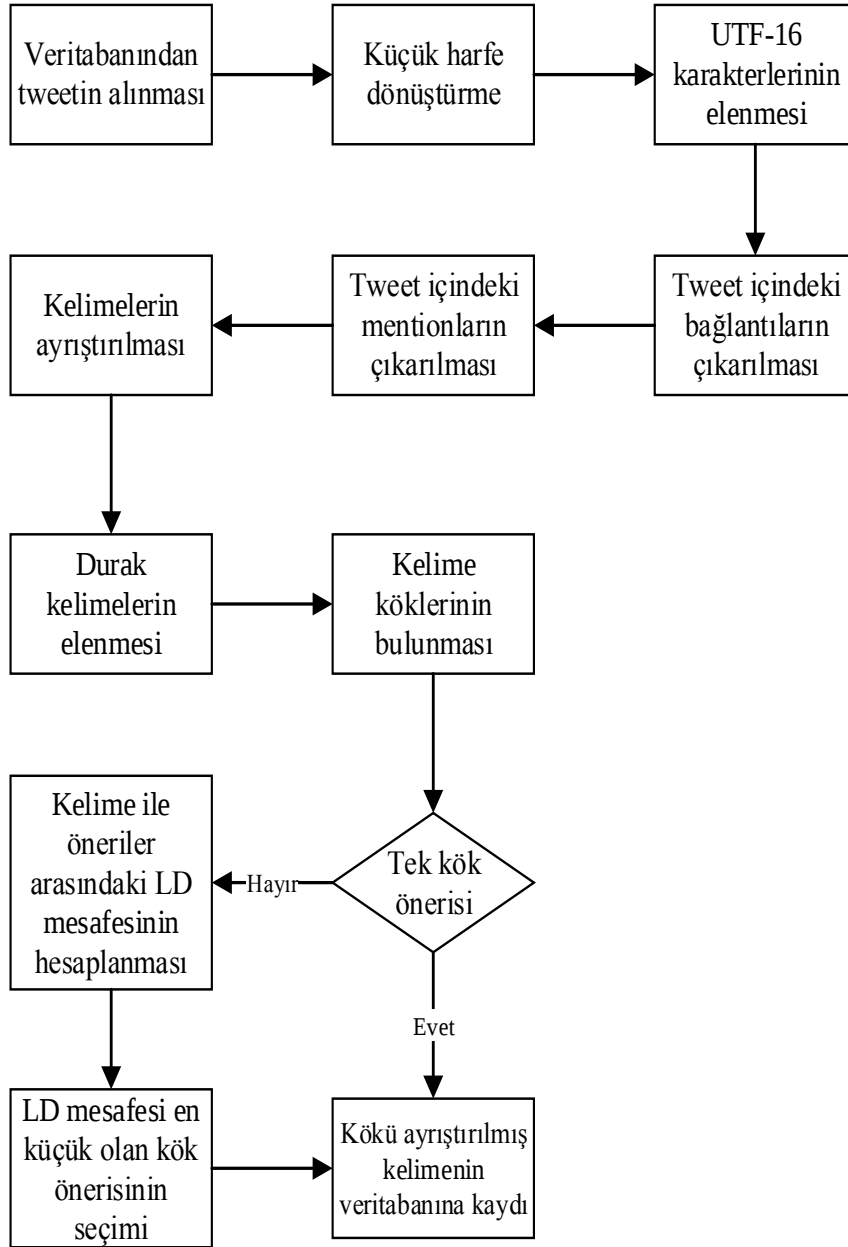
Çizelge 5.3. Veritabanında her bir kategoriye ait son tweet sayısı

<i>Kategoriler</i>	<i>Toplam Tweet Sayısı</i>
Teknoloji	5000
Spor	5000
Ekonomi	5000
Politika	5000
Magazin	5000
Sağlık	5000
Kültür-Sanat	5000

Modül #2: Veri önileme modülü

Bu modülün görevi, veri toplama modülü sonucunda depolanan metin tabanlı sosyal ağ paylaşımlarının ön işleminden geçirilerek makine öğrenmesi yöntemleri ile gerçekleştirilecek olan sınıflandırma işlemine hazır hale getirilmesidir [148, 150]. Literatürde metin sınıflandırma öncesinde yapılan önileme adımı gerçekleştirilen işlemler genellikle küçük harfe çevirme, noktalama işaretlerinin silinmesi, kelimelerin ayrıştırılması, ayrıştırılmış kelimeler arasından durak kelimelerin elenmesi ve kelimelerin köklerinin bulunması olarak sıralanmaktadır [104, 151, 152, 153].

Bu yaklaşımı kullanarak geliştirilen yazılımda yürütülen önileme adımları Şekil 5.3’de verilen akış diyagramında sunulmuştur.



Şekil 5.3. Modül #2'ye ait akış diyagramı

Şekil 5.3'deki akış diyagramında gösterilen adımların her biri aşağıda detaylı bir şekilde açıklanmıştır.

Küçük harfe dönüştürme

Bu aşamada, veri toplama modülü sonucunda veritabanına kaydedilmiş 7 farklı kategorideki her bir tweetin içeriğindeki tüm kelimelerin içerdiği harfler Java'daki String sınıfına ait özel metotların kullanımı ile küçük harfe dönüştürülmüştür.

UTF-16 karakterlerinin elenmesi

Bu aşamada ise sosyal paylaşım ağlarında sıklıkla görülen, kullanıcıların duygularını, görüşlerini ve hatta bulundukları mekânları ifade etmelerine yarayan emoji adı verilmiş UTF-16 karakter kümesinde tanımlı olan küçük resimler elenmiştir. Çünkü bu ifadeler kelime değeri taşımamakla birlikte haber kategorileri için ayırt edici özelliklere sahip değildirler. UTF-16 karakterleri sisteme Düzenli İfadeler Yardımı ile tanıtılmıştır.

Düzenli ifadeler

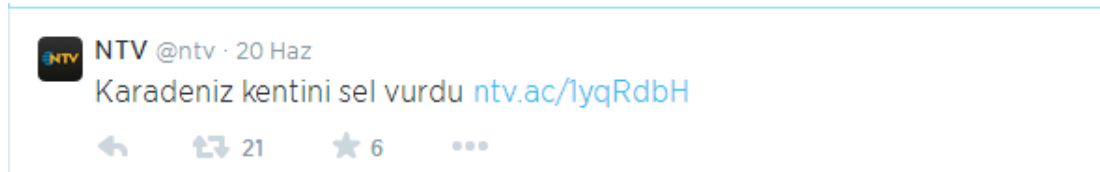
Düzenli İfadeler Jeffrey Friedl tarafından “metinlerin tanımlanmasını ve çözümlenmesini sağlayan küçük programlama dili” olarak tanımlanmıştır [106, 154]. Düzenli İfadeler ile sistem tarafından bir metinden istenilen özellikler çıkarılabilmekte, metin değiştirilebilmekte veya istenilen örüntünün metin içerisinde olup olmadığı kontrol edilebilmektedir. Küçük bir örnekle açıklanacak olursa posta kodu ve bölge isminden oluşan “1234 Maltepe” gibi bir adreste sadece posta kodunun seçilerek çıkarılması işlemi “/([0-9]f4g) .*/.” düzenli ifadesi ile gerçekleştirilebilmektedir. Bu örüntü aranan metin içerisinde, kendisini bir boşluk ve sonrasında herhangi bir karakterin takip edebileceği 4 rakamlık bir sayının arandığını ifade etmektedir. Parantezler ise metin içerisinde aranan kısmın çıkarılmasının istendiğini göstermektedir [106]. Düzenli İfadeler neredeyse her programlama dili için mevcuttur [106]. Bu tez kapsamında geliştirilen yazılımda Java dili kullanıldığından Java için geliştirilmiş olan regex kütüphanesi kullanılmıştır. UTF-16 karakterlerinin elenmesi için kullanılan Düzenli İfade ise aşağıda belirtildiği gibidir:

```
“([^\u0000-\uFFFF]”, “”).replaceAll(“\r”, “ ”).replaceAll(“\n”, “ ”).replaceAll(“””, “”)”
```

Bu ifade yardımıyla tweetlerin içerdiği UTF-16 karakterleri tweet içeriklerinden silinmiştir.

Bağlantıların silinmesi

Sosyal ağlardan yapılan haber paylaşımlarında haberin alındığı kaynağın gösterilmesi amacıyla metin içerisinde bağlantı paylaşma oldukça popülerdir. Özellikle mikroblog türündeki sosyal ağlarda karakter sınırlaması bulunduğundan kullanıcılar sıklıkla bağlantı paylaşma eğilimi göstermektedir. Bunların yanı sıra haber paylaşımlarında diğer paylaşımlara göre daha çok bağlantı paylaşıldığı görülmüştür. Bu durumun sebebi ise haber kuruluşlarının sosyal ağlar üzerinden paylaştıkları haberleri kendi resmi web sayfalarındaki haberlere yönlendirme eğilimleri olarak yorumlanmıştır. Paylaşılan bağlantılar birer internet adresi olduğu için metin içerisinde kelime bazında hiçbir anlam ifade etmemektedir [138]. Bu çalışma kapsamında geliştirilen yazılımda Twitter mikroblog ağı verileri kullanıldığından ön işleme sürecinin bu adımında tweet içerisindeki bağlantılar elenmiştir [155]. Resim 5.1’de Twitter üzerinden paylaşılan bir tweetin içerdiği bağlantının nasıl gösterildiği ifade edilmiştir.



Resim 5.1. Twitter üzerinden yapılan bir paylaşımın içerdiği bağlantı

Şekil 5.1’den görülebileceği gibi, yapılan Twitter paylaşımında haber özetlenmiştir fakat daha detaylı bilgi almak isteyen kullanıcılar için haberin kaynağına yönlendirme amacı ile bir bağlantı paylaşılmıştır. Fakat bu bağlantı metin olarak değerlendirildiğinde, harf gruplarının Türkçe dili için anlamlı birer kelimeye karşılık gelmediği görülmektedir. Dolayısıyla ön işleme sırasında bu tür bağlantıların elenmesi hem sistem performansını artıracak hem de sınıflandırma başarımını olumlu yönde etkileyecektir. Geliştirilen sistem de bağlantıların silinmesi için UTF-16 karakterlerin elenmesinde olduğu gibi Düzenli İfadeler yöntemine başvurulmuştur. Sistemin tweetlerin içerisindeki bağlantıları tanıyabilmesi için yazılan Düzenli İfade ise aşağıda görüldüğü gibidir.

"((https?|ftp|gopher|telnet|file|Unsure|http):(((/)|(\|))) +[\\w\\d:#@%/\$()~_?\\+ -=\\\\\\\\. &]*)"

Düzenli İfadeler yardımı ile tweetlerin içerdiği bağlantılar tweet içeriklerinden silinmiştir. Her bir kategoride eşit sayıda tweet olmak üzere 7 farklı kategoride toplam 35000 tweetin bulunduğu veri tabanından bağlantıların elenmesi işlemi sonlandırıldığında toplamda 37,979 bağlantının elendiği görülmüştür. Bu durum özellikle haber kuruluşlarının sosyal ağ hesapları üzerinden yaptıkları haber paylaşımlarında bol sayıda bağlantı paylaşarak, kullanıcıları kendi resmi haber sitelerine yönlendirdiklerini kanıtlar niteliktedir.

Mentionların çıkarılması

Twitter'daki metin paylaşımlarının sınıflandırılması üzerine yapılan çalışmaların oldukça az bir kısmında mentionların metin içeriğinden elenmesi gerekliliğine vurgu yapılmıştır [143]. Fakat mentionlar sadece kullanıcı adından oluştuğundan ve kullanıcı isimleri haber kategorileri bakımından herhangi bir anlam ifade etmediğinden dolayı bu çalışma kapsamında tweet içeriklerinden mentionlar da elenmiştir. Mentionların sisteme tanıtılması için "@\S" Düzenli İfadesi kullanılmıştır. Bu ifade sisteme @ ile başlayan ve bir sonraki boşluğa kadar olan bütün karakterleri mention olarak algılaması gerektiğini belirtmektedir. Toplamda 35000 tweet içeriğinden oluşan veri tabanından bu aşamanın sonunda 5476 mention'ın elendiği görülmüştür. Bu rakam haber kuruluşlarının resmi sosyal ağ hesapları üzerinden yaptıkları haber paylaşımlarının oldukça az bir kısmında diğer kullanıcılardan bahsettiklerini göstermektedir.

Kelimelerin ayrıştırılması

Metin tabanlı sosyal ağ paylaşımları içerisinden sınıflandırma başarısına etki etmeyeceği tespit edilen alanlar elendikten sonra yapılması gereken işlem paylaşım içerisinde geçen her bir kelimenin ayrıştırılmasıdır [104]. Bu adım sınıflandırma işlemine yardımcı olacak kelime ağırlıklandırma işleminin yapılabilmesi için gereklidir. Dolayısıyla sistem kelimeleri ağırlıklandırmadan önce kelimelerin neler olduğunu tespit etmelidir.

Bu işlem genellikle sisteme 2 boşluk karakteri arasındaki dizgelerin tanıtılması ve bu dizgelerin ayrıştırılmış birer kelime olarak kabul edilmesiyle gerçekleştirilmektedir [104]. Fakat bu adımda 2 boşluk karakteri arasındaki her dizgenin bir kelime ifade etmeyeceği göz önünde bulundurulmalıdır. Sistemin kelime olarak algılayabileceği dizgeler; noktalama işaretleri, rakamlar, belirlenen dil için kelime olarak algılanmayacak olan harf

grupları ve hatta rakam, harf ve noktalama işareti kombinasyonlarından oluşabilmektedir. Bu nedenle bu tez çalışması kapsamında sistemin kelimeleri doğru bir şekilde ayrıştırıp yalnızca Türkçe kelime anlamı taşıyan harf gruplarını algılayabilmesi için Türkçe 'ye özgü, açık kaynak kodlu ve başarısı test edilmiş bir Doğal Dil İşleme kütüphanesi olan Zemberek yazılımı kullanılmıştır [156].

Zemberek

Java ile gerçekleştirilmiş olan Zemberek yazılımı, Türkçe metin tespiti, Türkçe kelimelerin ayrıştırılması ve çözümlenmesi, kelimelerin köklerinin bulunması, heceleme, kelime yazımlarının denetlenmesi ve hatalı kelimeler için doğru olabilecek önerilerin yapılması gibi birçok işlemi gerçekleştirebilmektedir [157]. Zemberek Türkçe Doğal Dil İşleme Kütüphanesi, Türkçe metinlerin sınıflandırıldığı birçok akademik çalışmada kullanıldığı gibi Open Office'te de yazım denetleyicisi olarak kullanılmaktadır [156, 158, 159, 160].

Bu tez çalışması kapsamında geliştirilen sistemde mention, bağlantı ve UTF-16 karakterleri gibi metin sınıflandırmada başarıya etki etmeyecek olan alanlar çıkarılarak Twitter metin paylaşımları temizlendikten sonra Zemberek kütüphanesi ile her bir tweetin içerdiği kelimeler ayrıştırılmıştır. Bu ayrıştırma işlemi sırasında ağırlıklandırma ve sınıflandırma için seçilecek olan kelimelerin, Zemberek 'in yaptığı sınıflandırmaya göre yalnızca Türkçe kelimelerden oluşmasına dikkat edilmiştir. Türkçe kelime olduğu tespit edilen dizgelerin kesme işareti ile ayrılmış bir şekilde yazılabileceği de göz önünde bulundurularak Zemberek tarafından yalnızca kesme işareti içeren Türkçe kelimelerin belirlenmesi ve ayrıştırılması işlemi gerçekleştirilmiştir. Böylelikle yalnızca kesme işareti içeren Türkçe kelimelerin seçilmesiyle, metin içerisindeki noktalama işaretleri ve rakamlar gibi kelime değeri olmayan içerikler de elenmiş olmuştur. Bu işlemin sonucunda 35,000 tweet içeriğinden toplamda 253,030 kelimenin ayrıştırıldığı görülmüştür.

Durak kelimelerin çıkarılması

Metin sınıflandırma amacıyla verinin temizlenmesi ve ön işleme adımıyla gerçekleştirilen bir diğer işlem de metin içerisindeki durak kelimelerin çıkarılmasıdır [151]. Belirlenen dil için tek başına herhangi bir anlam ifade etmeyen ve sınıflandırma işleminde belirlenmiş olan kategorilerden her birinde aynı sıklıkta kullanılacak şekilde yaygın olan kelimeler durak kelime olarak tanımlanmaktadır [148]. Dolayısıyla durak kelimeler metnin

içeriğinden bağımsızdır ve herhangi bir anlamsal bilgi taşımadıkları gibi veride ekstra gürültü oluşmasına da sebebiyet vermektedirler [142, 161].

Geliştirilen sistemde durak kelimelerin çıkarılması için Zemberek kütüphanesinde Türkçe için tanımlanmış olan durak kelimelerin listesi kullanılmıştır. Çizelge 5.4’de Zemberek tarafından tanımlanmış olan ve sistemin ayrıştırdığı kelimeler arasından seçerek elediği durak kelime listesinin bir bölümü sunulmuştur [156].

Çizelge 5.4. Zemberek kütüphanesinde tanımlanmış durak kelime listesinden bir kesit

<i>Durak Kelime</i>
acaba
acayip
ait
ama
artık
ancak
bazı
bazen
belki
ben
çoğuna
çoğunu
çok
çünkü
çav
diğer
diğeri
diğerleri
dikkat
diye

Bir önceki adımda elde edilen 253,030 kelime arasından toplamda 37,923 adet durak kelime olduğu tespit edilmiştir ve bu kelimeler veri tabanından elenmiştir. Bu eleme işlemi sonucunda ise toplam 215,107 kelime kalmıştır. Bu rakam veri setindeki her bir haber kategorisi için değerlendirildiğinde her bir kategoride içerilen toplam kelime sayısı Çizelge 5.5’de görüldüğü gibidir.

Çizelge 5.5. Her bir kategoriye ait toplam kelime sayısı

<i>Haber Kategorisi</i>	<i>Toplam Kelime Sayısı</i>
Teknoloji	30,497
Spor	32,916
Ekonomi	28,271
Politika	32,305
Magazin	28,033
Sağlık	34,078
Kültür-Sanat	29,007

Fakat 35,000 tweet içeriğinden toplamda elde edilen 215,107 kelime arasında aynı kelimeler de bulunmaktadır. Bir kelimenin veri tabanında birçok kez yer alması sistemin performansını olumsuz yönde etkileyeceğinden birden fazla görülen kelimelerin sayısı veri tabanında bire düşürülmüştür. Böylelikle veri tabanının birbirinden farklı kelimelerden oluşması sağlanmıştır. Bu işlem sonlandırıldığında ise veri tabanında 12,620 farklı kelime kaldığı tespit edilmiştir.

Kelime köklerinin bulunması

Sınıflandırılacak olan metin içerisindeki anlamlı ve belirleyici kelimeler ayrıştırıldıktan sonra yürütülmesi gereken ön işleme adımlarında biri de kelimelerin köklerinin bulunması işlemidir [158, 159]. Türkçe sondan eklemeli bir dil olduğu için bir kelimeye farklı ekler getirilerek yeni sözcükler türetilabilmektedir [157]. Türkçe dilbilgisine ait ek türleri arasından çekim ekleri, eklendikleri kelimenin anlamını değiştirmemektedirler. Bu durumda sistemin de çekim eki olsa da anlamı değişmeyen kelimeleri tek bir kelime olarak tanınması, ağırlıklandırma ve sınıflandırma işleminin başarı ile sonuçlandırmasını sağlayacaktır. Bir örnekle açıklanacak olursa ‘kitap’ kelimesi ile ‘kitaplarım’ kelimesi söz dizimsel olarak bakıldığında farklı kelimeler olsa da anlam olarak her ikisi de kitap anlamını taşımaktadır ve özellikle otomatik metin sınıflandırma işlemlerinde sistem başarısını olumlu yönde etkilemesi açısından sistemin de aynı kelimeler olduğunu saptaması gerekmektedir. Bu tez çalışması kapsamında geliştirilen yazılımda kelime köklerinin bulunması için yukarıda detaylı bir şekilde özellikleri açıklanan Zemberek Doğal Dil İşleme Kütüphanesi kullanılmıştır. Zemberek bir kelimenin kökü için kelime türüne göre birden fazla öneride bulunarak genellikle en uzun kelime kökünü ilk sırada

çıktı olarak sunmaktadır ve yapılan çalışmalarda genellikle başarılı sonuçlar ürettiği gözlemlenmiştir [158, 160, 162]. Resim 5.2’de Zemberek ‘in web uygulamasına girdi olarak sunulmuş bir cümlenin çözümlenmiş hali gösterilmiştir [163].



Resim 5.2. Örnek Zemberek uygulaması

Şekil 5.2’de görüldüğü gibi Zemberek yazılımının web ortamı için geliştirilmiş platformu kendisine sunulan cümlenin içerdiği kelimeleri başarılı bir şekilde köklerine ayırmıştır. Bu aşamada, Zemberek ‘in bir kelime için birden fazla kök bulduğu durumlarda, sunduğu ilk kök o kelimenin kökü olarak kabul edilmiştir. Bunun sebebi kelimenin yapım eki olarak anlamının değişmiş olmasına karşın Zemberek ‘in sunduğu ilk kökün kelimeye en yakın ve en uzun kök olmasıdır [158].

Zemberek ‘in bazı kelimeler için hiçbir kök bulamadığı durumlar da mevcuttur. Kökü bulunamayan kelimeler incelendiğinde ise bu kelimelerde yazım hatası olduğu görülmüştür. Bu durumda yapılması gereken Zemberek ‘ten bir kök önerisinde bulunmasını istemektir. Fakat bu sefer de Zemberek ‘in birden fazla kök önerisinde bulunduğu gözlemlenmiştir. Literatürdeki diğer çalışmalar incelendiğinde, kolaylık sağlaması açısından bu öneriler arasından ilk önerinin o kelimenin kökü olarak kabul edildiği görülmüştür [158, 159]. Fakat Cengiz ve Diri, çalışmalarında ilk kök önerisinin kabul edilmesinden dolayı bir takım hataların oluştuğunu belirtmişlerdir [158]. Bu nedenle bu çalışma kapsamında, kök önerileri arasından asıl kelimeye en yakın mesafede olan kök önerisi o kelimenin kökü olarak kabul edilmiştir. Kelime ve kök önerileri arasındaki mesafe hesaplaması ise Levenshtein Mesafe (Levenshtein Distance) algoritması ile gerçekleştirilmiştir.

Levenshtein mesafe algoritması

Levenshtein Mesafe Algoritması iki dizilim arasındaki benzerliği ölçmek için ikisi arasında mesafe ölçümü yapılması gerektiğini öngörmektedir [164]. Bu algoritma genellikle kelimeler arasındaki benzerliğin derecelendirilmesinde kullanılmaktadır [165]. Algoritma iki kelime arasındaki benzerliğe bakarken, kelimeleri birbirine eşitlemek için harfler arasında yapılması gereken değişim ve ekleme işlemlerinin sayısına göre bir mesafe hesaplaması yapmaktadır [164]. Küçük bir örnekle açıklanacak olursa, “Kitap” ve “Hitap” kelimeleri arasındaki Levenshtein Mesafesi 1’dir.

K-i-t-a-p

H-i-t-a-p

Yukarıda görüldüğü gibi algoritma bu 2 kelime arasındaki mesafeyi hesaplarken yalnızca k harfinin h harfi ile değiştirilmesine ihtiyaç duyacaktır çünkü kelimelerin geri kalan harfleri birbiri ile aynıdır. Dolayısıyla bu 2 kelimeyi birbirine eşitlemek için yalnızca 1 değişime ihtiyaç duyulduğundan bu kelimeler arasındaki Levenshtein Mesafesi 1’dir. “Arap” ve “araba” kelimeleri ele alındığında ise Levenshtein Mesafesinin 2 olduğu görülmektedir.

A-r-a-p

A-r-a-b-a

Bu örnekte kelimeleri birbirine eşitlemek için p ve harfinin b harfi ile değişimi ve a harfinin eklenmesi gerekmektedir. Dolayısıyla 1 değişim ve 1 ekleme gerçekleştiğinden bu kelimeler arasındaki Levenshtein Mesafesi algoritma tarafından 2 olarak bulunmaktadır.

Bu çalışma kapsamında geliştirilen yazılımda Zemberek ‘in sunduğu kök önerileri ile gerçek kelime arasındaki Levenshtein Mesafesi hesaplanarak en doğru kök önerisinin seçilmesi ve böylelikle sınıflandırma başarımının artırılması amaçlanmıştır.

Kelimelerin veri tabanına kaydedilmesi

Bu basamakta, yukarıda detaylı bir şekilde açıklanan veri ön işleme adımları veri setindeki metin tabanlı sosyal ağ paylaşımlarına uygulandıktan sonra elde edilen kelime kökleri ağırlıklandırmak üzere veri tabanına kaydedilmiştir. Bu işlem sonucunda 35,000 tweet için

toplamda 12,620 kelime kökünden oluşan veri tabanının son halinden bir kesit Çizelge 5.6’da görüldüğü gibidir.

Çizelge 5.6. Kelime köklerinin kaydedildiği veri tabanı tablosundan bir kesit

<i>ID</i>	<i>Kelime</i>
1	Erzade
2	süper
3	süpür
4	sakin
5	Levanten
6	kristal
7	hiç
8	vokal
9	kobra
10	kobay
11	aksalı
12	tatmin
13	şişe
14	şifa
15	şile
16	şiir
17	şike
18	şive
19	şist
20	şirk
21	ofis
22	RTÜK
23	Molotof
24	pür
25	poz
26	poy
27	pos

Modül #3: Ağırlıklandırma modülü

Bu modülün işlevi, metin tabanlı sosyal ağ paylaşımlarında görülen kelimeleri, her bir kategori için sahip oldukları önem derecesine göre ağırlıklandırarak bir başka deyişle

doküman analizi yapmaktır [166]. Metnin belirlenen kategorilerden hangisine ait olduğunun sistem tarafından belirlenebilmesi için metin içerikli her bir dokümanın vektör uzay modelinde bir vektör olarak temsil edilmesi gerekmektedir [104, 106]. Vektör uzay modeline göre her bir metin tabanlı sosyal ağ paylaşımı bir dokümana, paylaşımın içerdiği her bir kelime ise bir terime tekabül etmektedir. Bu durum eşitliklerle gösterilecek olursa D doküman kümesini, d her bir dokümanı ve t her bir terimi ifade etmek üzere:

$$D=(d_1, d_2, d_3, \dots, d_n) \quad (5.1)$$

$$d_1=(t_1, t_2, t_3, t_4, t_5, \dots, t_m) \quad (5.2)$$

olarak gösterilmektedir [106].

Vektör uzay modelinde her terim bir ağırlık değerine sahiptir. Bu nedenle öncelikle sınıflandırma amacıyla kullanılacak her metin paylaşımının içerdiği kelimelerin ağırlıklandırılması gerekmektedir. Literatürde metin sınıflandırma alanında yapılmış olan çalışmalar incelendiğinde terim ağırlıklandırma işlemi için genellikle 3 farklı yöntemin kullanıldığı görülmüştür. Bunlar bit ağırlıklandırma, terim frekansı ve terim frekansı-ters doküman frekansı ağırlıklandırma yöntemleri olarak sıralanmaktadır [106, 167, 168]. Bit ağırlıklandırma yöntemi terimi ağırlıklandırırken yalnızca terimin dokümanda geçip geçmediği bilgisini dikkate alır. Belirlenen dokümanda geçen terimlerin ağırlığı 1, geçmeyen terimlerin ağırlığı ise 0 olarak belirlenir [106]. Terim frekansı yönteminde ise dokümanda geçen terimlerin kaç kez geçtiği bilgisi de önem kazanır. Bu yöntemle göre, bir doküman içerisinde görülen iki terimin ağırlığı birbirlerinden farklıdır ve terimin ağırlığı dokümanda görülme frekansına eşittir [168].

Bu tez çalışması kapsamına geliştirilen sistemde ise terim frekansı-ters doküman frekansı ağırlıklandırma yöntemi kullanılmıştır.

Terim frekansı – ters doküman frekansı

Terim Frekansı-Ters Doküman Frekansı (TF-TDF), doküman içerisinde geçen terimleri ağırlıklandırırken, bir terimin her bir kategori için ne derece önemli ve belirleyici olduğu bilgisini kullanan bir ağırlıklandırma yöntemidir [166]. Bu yöntemle göre, bir terim kendisiyle en çok bağlantılı olan kategoriye ait dokümanlarda yüksek bir sıklıkta

görülürken, ait olduğu kategoriye ait az sayıda dokümanda görülmelidir. Böylelikle kelimenin o kategori için belirleyici olmaması durumu ortadan kaldırılmaktadır. Bir başka deyişle bir terimin ait olduğu kategori için az sayıda dokümanda daha sık görülmesi durumu, çok sayıda dokümanda aynı sıklıkta görülmesi durumuna göre o terimin o kategori için daha belirleyici olduğunu göstermektedir [104].

TF-TDF ağırlıklandırılması yönteminin formülü Eş. 5.3'de görüldüğü gibidir [142, 166]. Formülde i her bir terimi, j ise her bir dokümanı temsil etmektedir.

$$TF-TDF_{i,j} = TF_{i,j} \times TDF_i \quad (5.3)$$

Eş. 4.3'den görülebileceği gibi her bir terimin Terim Frekansı (TF) ve Ters Doküman frekansı (TDF) değerlerinin çarpımı o terimin TF-TDF ağırlığını vermektedir. TF-TDF değerlerinin bulunması için kullanılan formüller ise aşağıda sırasıyla verilmiştir.

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (5.4)$$

$$TDF_i = \log \frac{|D|}{|\{d: t_i \in d\}|} \quad (5.5)$$

Eş. 5.4 ve Eş. 5.5'de ifade edildiği gibi, bir terimin bir kategori için TF'si hesaplanırken, o terimin belirlenen kategoride kaç kere görüldüğünü ifade eden n sayısı, aynı terimin diğer kategorilere ait dokümanlarında bulunduğu bütün dokümanlarda toplamda kaç kere görüldüğünü ifade eden sayıya bölünmektedir. Böylelikle bir terimin bir kategori için görülme sıklığı elde edilir. Bir terimin bir kategori için terim frekansı ne kadar fazla ise o terim o kategori için o kadar değerlidir. Ağırlığı ise o derece yüksektir.

Bir terimin TDF'si hesaplanırken ise $|D|$ ile ifade edilen toplam doküman sayısı o terimin belirlenen kategori için kaç dokümanda görüldüğü bilgisini temsil eden d sayısına bölünmektedir. Elde edilen sonucun logaritma değeri hesaplanmaktadır. Bir terim bir kategoride ne kadar az sayıda dokümanda görülüyorsa o kategori için o kadar belirleyicidir. Sonuç olarak TF-TDF yöntemi ile bir kategoride az sayıda dokümanda daha sık görülen kelimelerin ağırlığı daha yüksek olmaktadır. Böylelikle, kategoriler için daha

anlamalı ve belirleyici kelimeler belirlenerek sınıflandırma başarımının artırılması hedeflenmektedir.

Bu çalışma kapsamında metin tabanlı tweetler ön işlemden geçirildikten sonra elde edilen kelimeler TF-TDF yöntemine göre ağırlıklandırılmıştır. Bu ağırlıklandırma için kelimelerin hesaplanan TF ve DF değerleri veri tabanına eklenmiştir. Çizelge 5.7.'de bu aşamadaki veri tabanından bir kesit sunulmuştur.

Çizelge 5.7. Ağırlıklandırılmış örnek kelime grubu

<i>Ağırlık Değeri/ Kelime</i>	<i>Süper</i>	<i>Şike</i>	<i>Ofis</i>	<i>Poz</i>	<i>Mobil</i>
<i>Teknoloji_TF</i>	3	1	11	1	134
<i>Spor_TF</i>	162	52	1	10	1
<i>Ekonomi_TF</i>	10	2	10	1	2
<i>Politika_TF</i>	4	12	6	2	2
<i>Magazin_TF</i>	7	1	3	32	1
<i>Sağlık_TF</i>	5	1	2	2	4
<i>Kültür-Sanat_TF</i>	4	1	1	5	3
<i>Teknoloji_DF</i>	3	1	11	1	119
<i>Spor_DF</i>	147	43	1	10	1
<i>Ekonomi_DF</i>	9	2	10	1	2
<i>Politika_DF</i>	4	11	5	2	2
<i>Magazin_DF</i>	7	1	2	31	1
<i>Sağlık_DF</i>	5	1	2	2	4
<i>Kültür-Sanat_DF</i>	4	1	1	5	3

Modül #4: Sınıflandırma ve test modülü

Geliştirilen sistemin son basamağı olan bu modülün görevi, metin tabanlı sosyal ağ paylaşımlarını sınıflandırarak ait oldukları kategoriye atamaktır.

Bilgi teknolojilerinde kullanılan veri seti boyutlarının hızlı bir şekilde artmasıyla birlikte çevrimiçi verilerin büyüklüğü de petabayt, exabayt ve hatta zetabaytlar mertebesine ulaşmıştır. Veriden bilgi çıkarımı yapma amacıyla bu tür yapılandırılmamış verilerin manuel bir şekilde sınıflandırılması günler, aylar hatta yıllar süren bir çaba gerektirmektedir. Böylesi bir ihtiyaç sonucunda ortaya çıkan otomatik sınıflandırma işlemi, büyük miktardaki veriyi etkili ve kullanışlı şekilde kategorize ederek veriden bilgi çıkarımı

olarak tanımlanmaktadır [137, 145]. Otomatik sınıflandırma işlemi ile doğru bilgiye erişimin hızlandırılması ve böylelikle kullanıcıların bilgi ihtiyaçlarının daha kısa sürede karşılanması amaçlanmaktadır [169]. Sınıflandırma teknolojileri büyük veriyi farklı perspektiflerden ele alma ve organize etme alanlarında umut verici sonuçlar üretmektedir [144]. Sınıflandırma problemi matematiksel olarak tanımlandığında D kayıtlar kümesini ve X_n her bir kaydı göstermek üzere, $D=\{X_1, X_2, \dots, X_n\}$ kümesindeki her bir kaydın $\{1, \dots, k\}$ kümesindeki 1'den k 'ya kadar olan sınıf etiketleri ile etiketlenmesi işlemi olarak ifade edilmektedir [136]. Otomatik veri sınıflandırmanın alt alanlarından biri olan metin sınıflandırma işlemi, doğal dil metinlerini tematik kategoriler ile etiketleme olarak tanımlanmaktadır [106]. Böylelikle metin verisinin belirlenen sınıflardan hangisine ait olduğu otomatik olarak bulunabilmektedir.

Makine öğrenmesi yöntemleri ile gerçekleştirilen metin sınıflandırma işleminde eldeki metin verisinin eğitim ve test kümesi olacak şekilde ayrıştırılması gerekmektedir. Eğitim kümesi sistemin metin ve ait olduğu kategori arasındaki ilişkiyi öğrenmesinde kullanılırken, test verisi sistemin öğrendiği bilgileri test etmek amacıyla daha önce karşılaşmadığı metin verisinin sınıfını bulması amacıyla kullanılmaktadır. Daha sonra test verisinin sistem tarafından bulunan sınıfı ile gerçekte ait olduğu sınıf kıyaslanarak sistemin başarımlı ölçümü gerçekleştirilmektedir.

Bu tez çalışması kapsamında oluşturulan veri setinde, veri toplama modülü yardımıyla toplanan Twitter sosyal ağına ait 35,000 metin tabanlı tweet paylaşımı bulunmaktadır. Sınıflandırma modelinin oluşturulması amacıyla veri setinin eğitim ve test verisi olarak ayrıştırılmasında çapraz geçerlilik yöntemi kullanılmıştır [136, 142].

Çapraz geçerlilik yöntemi

Çapraz geçerlilik yöntemi; çift çapraz geçerlilik ve çoklu çapraz geçerlilik olmak üzere 2 başlık altında incelenmektedir [170]. Çift çapraz geçerlilik yönteminde veri seti rastgele iki eşit parçaya ayrılmaktadır. Birinci aşamada ilk parça eğitim ikinci parça test amacıyla kullanılırken, ikinci aşamada ikinci parça eğitim ilk parça test amacıyla kullanılmaktadır. Sistemin genel başarısı ölçülürken ise bu iki test işleminin başarısının ortalaması alınmaktadır [106]. Çoklu çapraz geçerlilik yönteminde ise veri seti, her biri eşit sayıda veri içeren k tane gruba ayrılır. İlk aşamada birinci grup test, geri kalan diğer gruplar

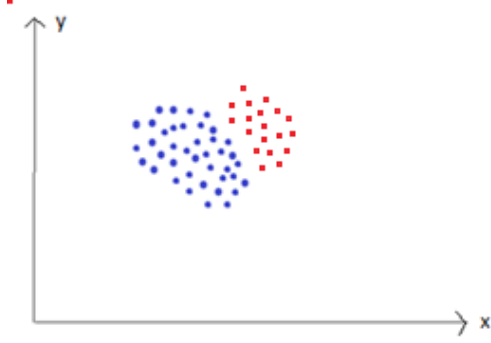
eğitim amacıyla kullanılır. Bu işlem grup sayısı kadar devam ettirilir ve her defasında bir grup test verisi diğer gruplar ise eğitim verisi olarak kullanılır. Sonuçta elde edilen hata oranı ise her bir test işlemi sonucunda elde edilen hatanın ortalaması olarak kabul edilmektedir [142, 170]. Böylelikle test verisinde oluşabilecek sapmalar engellenerek, yapılan sınıflandırma işleminin geçerliliği sağlanmış olmaktadır [101]. Kohavi tarafından yapılan bir çalışmada sınıflandırma işleminde genel olarak en makul parça sayısının 10 olduğundan bahsedilmiştir [171]. Literatürde var olan diğer metin sınıflandırma çalışmaları incelendiğinde, çoğunda 10-parçalı çapraz geçerlilik yönteminin kullanıldığı görülmüştür [101, 142, 144, 172, 173, 174]. Bu nedenle bu çalışma kapsamında 10-parçalı çapraz geçerlilik yöntemi kullanılarak 35,000 tweetten oluşan veri seti 10 parçaya bölünmüştür. Böylelikle sistem, her defasında 3,500'ü test 31,500'ü eğitim tweetinden oluşan veri setleri ile toplamda 10 kez eğitime ve test işlemleri için hazır hale getirilmiştir.

Bu aşamadan sonra yapılması gereken işlem sınıflandırıcının seçilmesi olayıdır. Yapılan literatür çalışması sonucunda metin tabanlı içeriklerin sınıflandırılmasında en çok kullanılan sınıflandırıcılardan birinin Naive Bayes sınıflandırıcısı olduğu görülmüştür [104, 137, 142, 144, 145, 153, 155, 161, 174, 175, 176, 177]. Bunun sebebi ise yapılan çalışmalarda, algoritmanın uygulama kolaylığı, ürettiği yüksek başarımlar, kolay tasarımı ve karmaşık problemleri çözmedeki başarısı olarak ifade edilmiştir [136, 137, 161]. Özellikle metin sınıflandırma alanında, metni sınıflandırma işlemine hazır hale getirmek için yapılan ön işleme adımları ile birlikte zaman ve performans bakımından kullanılabilecek en etkili sınıflandırıcının Naive Bayes olduğunu ispatlayan çalışmalar literatürde mevcuttur [153, 175, 177]. Bundan dolayı bu çalışma kapsamında Türkçe metin tabanlı sosyal ağ paylaşımlarının sınıflandırılmasında ve ilgili kategorilere atanmasında Naive Bayes sınıflandırıcısı kullanılmıştır.

Naive Bayes

Naive Bayes algoritması, metin sınıflandırma işleminde kullanılmak üzere ilk olarak 1998 yılında Lewis tarafından önerilmiştir [174]. Naive Bayes olasılık tabanlı bir sınıflandırıcıdır ve bir verinin öznitelik değerlerini kullanarak her bir kategoriye ait olma olasılığını hesaplamaktadır [144]. Naive Bayes sınıflandırıcısı, Bayes teoremini uygularken özniteliklerin birbirinden tamamen bağımsız olduğunu varsaymaktadır [106, 136].

Naive Bayes sınıflandırıcısının çalışma mantığı basit bir örnek üzerinden açıklanacak olursa; Şekil 5.4’de görüldüğü gibi 2 boyutlu koordinat düzleminde kırmızı-kare ve mavi-yuvarlak objelerden oluşan bir küme düşünüldüğünde, sınıflandırıcının görevi yeni gelen bir objeyi en uygun sınıfa atamaktır.



Şekil 5.4. Kırmızı-kare ve mavi-yuvarlak objelerden oluşan sınıfların gösterimi

Bu atama işlemi ise kümede var olan objelerin bilgilerinden yola çıkılarak gerçekleştirilecektir. Şekil 5.4’e bakıldığında mavi-yuvarlak objelerin kırmızı-kare objelerden sayıca daha üstün olduğu görülmektedir. Bu kapsamda Naive Bayes sınıflandırıcısına göre kümeye yeni gelen bir objenin mavi-yuvarlak obje olma olasılığı kırmızı-kare obje olma olasılığından daha yüksektir. Aynı şekilde kırmızı-kare ve mavi-yuvarlak objelerin sayıca eşit olduğu durumda ise yeni gelen objenin kırmızı-kare veya mavi-yuvarlak obje olma olasılığının birbirine eşit ve %50 olması beklenmektedir. Bu durumda Bayes sınıflandırıcısı yeni objenin her 2 obje sınıfı için de olasılıklarını hesaplamaktadır. Bu hesaplama işlemi Eş. 5.6 ve Eş. 5.7’de ifade edildiği gibi gerçekleşmektedir [137].

$$\text{Objenin mavi-yuvarlak olma olasılığı} = \text{Mavi objelerin sayısı} / \text{Toplam obje sayısı} \quad (5.6)$$

$$\text{Objenin kırmızı-kare olma olasılığı} = \text{Kırmızı objelerin sayısı} / \text{Toplam obje sayısı} \quad (5.7)$$

Naive Bayes algoritması metin sınıflandırma işleminde kullanıldığında ise kelimeler öznitelik olarak değerlendirilmekte ve olasılık hesaplaması her bir kelimenin her bir sınıf için frekans değerini temsil eden ağırlık değerlerinin kullanılması ile bulunmaktadır. Sisteme, sınıflandırılması istenen yeni bir metin geldiğinde, metnin içerdiği kelimelerin her bir sınıf için ağırlık değerleri bulunmaktadır ve bu ağırlık değerlerinin çarpımından elde edilen değer, metnin belirlenen sınıfa ait olma olasılığını temsil etmektedir [174, 178].

Daha önceki bölümlerde açıklandığı gibi bu tez çalışması kapsamında geliştirilen sistemde, Türkçe metin tabanlı Twitter paylaşımlarının sınıflandırma işlemine hazır hale getirilmesi amacıyla her bir tweet vektör uzay modelinde ifade edilebilmesi açısından sayısallaştırılmıştır. Bu işlem ağırlıklandırma modülü ile gerçekleştirilmiş ve her bir kelimenin her bir sınıf için TF-TDF değeri hesaplanmıştır. Sisteme uygun kategoriye yerleştirilmesi için bir tweet geldiğinde ise Naive Bayes metodu kullanılarak tweetin her bir sınıfa ait olma olasılığı hesaplanmış ve tweet, en yüksek olasılık değerine sahip olan kategoriye atanmıştır. Örnek olarak S tweetinin Teknoloji Haberi sınıfına ait olan bir paylaşım olma olasılığı hesaplanırken öncelikle tweetin içerdiği her bir kelimenin Teknoloji Haberi sınıfına ait TF-TDF ağırlıkları bulunur ve bu ağırlıklar birbirleriyle çarpılır. S tweeti için bu işlem her bir kategori için tekrarlanır. S tweetinin her bir kategori için genel ağırlık değeri hesaplandıktan sonra bu tweetin Teknoloji Haberi sınıfına ait olma olasılığı, tweetin Teknoloji Haberi sınıfı için bulunmuş genel ağırlığının yine aynı tweetin bütün kategorileri ait genel ağırlıklarının toplamına bölünmesiyle bulunur [155, 161]. Bu durum matematiksel olarak ifade edilecek olursa;

S: Sisteme yeni gelen tweeti,

t: tweetin içerdiği her bir kelimeyi,

X_1 : S tweetinin Teknoloji Haberi sınıfına ait genel ağırlığını

x_1 : t_1 terimin Teknoloji Haberi sınıfına ait TF-TDF ağırlığını

Y: S tweetinin tüm kategorilere ait ağırlık değerlerinin toplamını ifade etmek üzere $S=(t_1, t_2, \dots, t_n)$ olarak ifade gösterilmektedir. Bu durumda $X_1=(x_1 * x_2 * \dots * x_n)$ $Y=(X_1+X_2+X_3+X_4+X_5+X_6+X_7)$ olmak üzere S tweetinin Teknoloji Haberi sınıfına ait olma olasılığı

$$\text{Teknoloji Haberi olma olasılığı} = X/Y \quad (5.8)$$

olarak ifade edilmektedir.

Kategorisi belirlenmek istenen tweetin her bir kategoriye ait olma olasılığı bu şekilde hesaplandıktan sonra, tweet en yüksek olasılık değerine ait sınıfa atanmıştır.

10-parçalı çapraz geçişleme yöntemi ile organize edilmiş veri kümesinin bu şekilde eğitim ve test işlemleri tamamlandıktan sonra her bir parça için elde edilen başarı oranı Çizelge

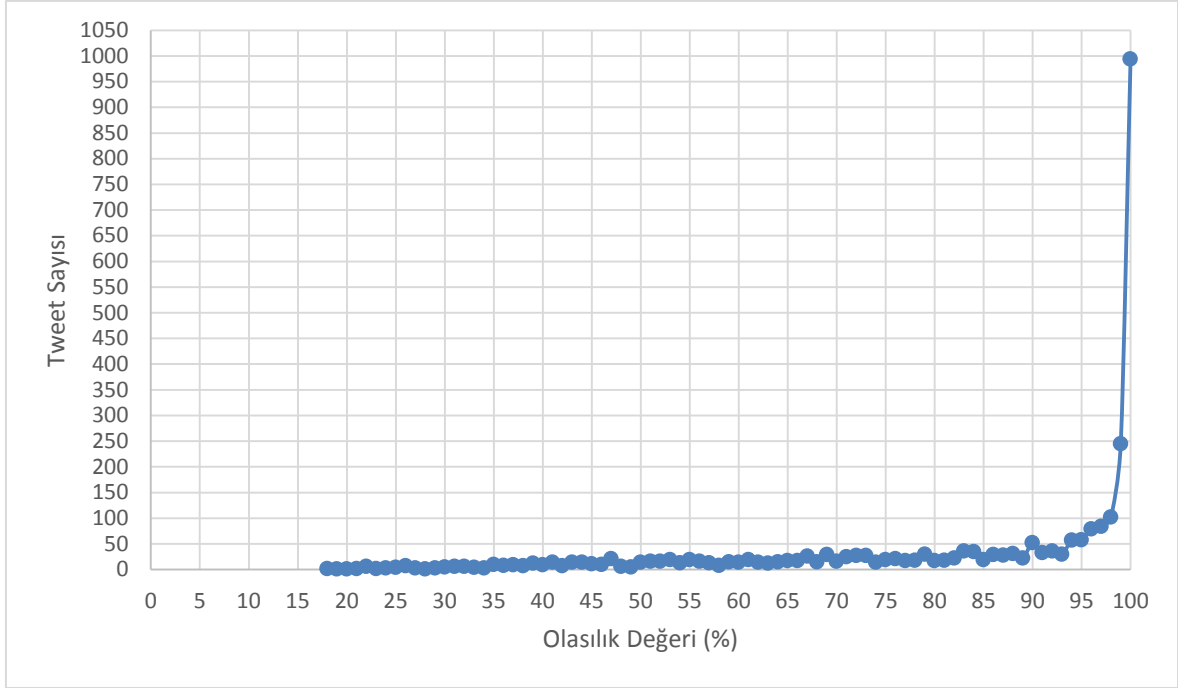
5.8’de sunulmuştur. Başarı oranı; sistemin test aşamasında kendisine sunulan tweetlerden doğru sınıflandırdığı tweet sayısının tüm test tweeti sayısına bölünmesi ile elde edilmektedir.

Çizelge 5.8. Sistemin test edilmesi sonucu elde edilen sınıflandırma başarımlarının yüzdeleri

<i>Parça/Kategori</i>	<i>Teknoloji</i>	<i>Spor</i>	<i>Ekonomi</i>	<i>Politika</i>	<i>Magazin</i>	<i>Sağlık</i>	<i>Kültür-Sanat</i>	<i>Genel</i>
<i>1. parça</i>	84,8	87,0	76,6	78,4	67,8	92,2	70,6	79,6
<i>2. parça</i>	81,8	85,8	74,2	75,2	68,0	90,4	75,8	78,7
<i>3. parça</i>	85,4	87,6	68,8	75,8	65,6	84,2	71,6	77,0
<i>4. parça</i>	81,2	88,0	71,8	81,2	68,8	89,4	71,6	78,9
<i>5. parça</i>	81,0	86,0	73,2	78,6	69,8	86,2	69,2	77,7
<i>6. parça</i>	81,0	84,6	70,0	79,0	68,8	88,0	73,4	77,8
<i>7. parça</i>	83,4	84,6	71,8	76,8	66,6	87,2	73,6	77,7
<i>8. parça</i>	83,8	82,8	76,6	71,6	69,0	87,0	72,2	77,6
<i>9. parça</i>	84,6	83,6	74,8	77,6	70,6	85,6	71,2	78,3
<i>10. parça</i>	80,8	85,8	76,8	77,4	68,4	86,2	70,0	77,9
<i>Ortalama</i>	<i>82,78</i>	<i>85,58</i>	<i>73,46</i>	<i>77,16</i>	<i>68,34</i>	<i>87,64</i>	<i>71,92</i>	<i>78,12</i>

Çizelge 5.8’de görüldüğü gibi her bir kategori için elde edilen sınıflandırma başarımlarının oranı %70’in üzerindedir. Ayrıca genel olarak tüm kategoriler için sistemin başarımlarını %78,12 olarak bulunmuştur. Daha öncede açıklandığı gibi literatürde Türkçe sosyal ağ paylaşımlarını sınıflandıran herhangi bir çalışmaya rastlanmamıştır. İngilizce diline ait sosyal ağ paylaşımlarını sınıflandıran çalışmalarda ise sınıflandırma başarımlarının oranları %20 ile %99 arasında değişmektedir [101, 104, 106, 136, 141, 142, 144]. Dolayısıyla elde edilen sonuçlar bu çalışmalar ile kıyaslandığında İngilizce oldukça az sayıda ek yapısına sahip kelimeler içermesine rağmen, bu çalışma kapsamında geliştirilen sistemin başarımlarının oranı oldukça yüksek ve tatmin edicidir.

Sistemin başarımlarının oranı tespit edildikten sonra güven aralığının belirlenmesi için veri toplama modülü aracılığıyla toplanmış olan 35,000 veri analiz edilmiştir. Şekil 5.5’de sistemin sınıflandırdığı tweetlerin sayısı ile sınıflandırma aşamasındaki olasılık değeri arasındaki ilişki sunulmuştur.



Şekil 5.5. Güven aralığı analiz sonucu

Şekil 5.5’de görüldüğü gibi sistem %90 ve %100 arasındaki olasılık değerleri ile sınıflandırdığı tweetleri genel olarak doğru kategorilerine atamıştır. %90’dan küçük olasılık değerleri ile sınıflandırdığı tweet sayıları ise birbirine çok yakın olduğundan herhangi bir yorum yapılamamaktadır. Bu nedenle sistemin güven aralığı %90 ve üzeri olarak belirlenmiştir. Dolayısıyla bir tweet bir kategoriye %90 ve üzeri olasılık değeri ile atanmışsa sistemin güven aralığı içerisinde yer almaktadır.

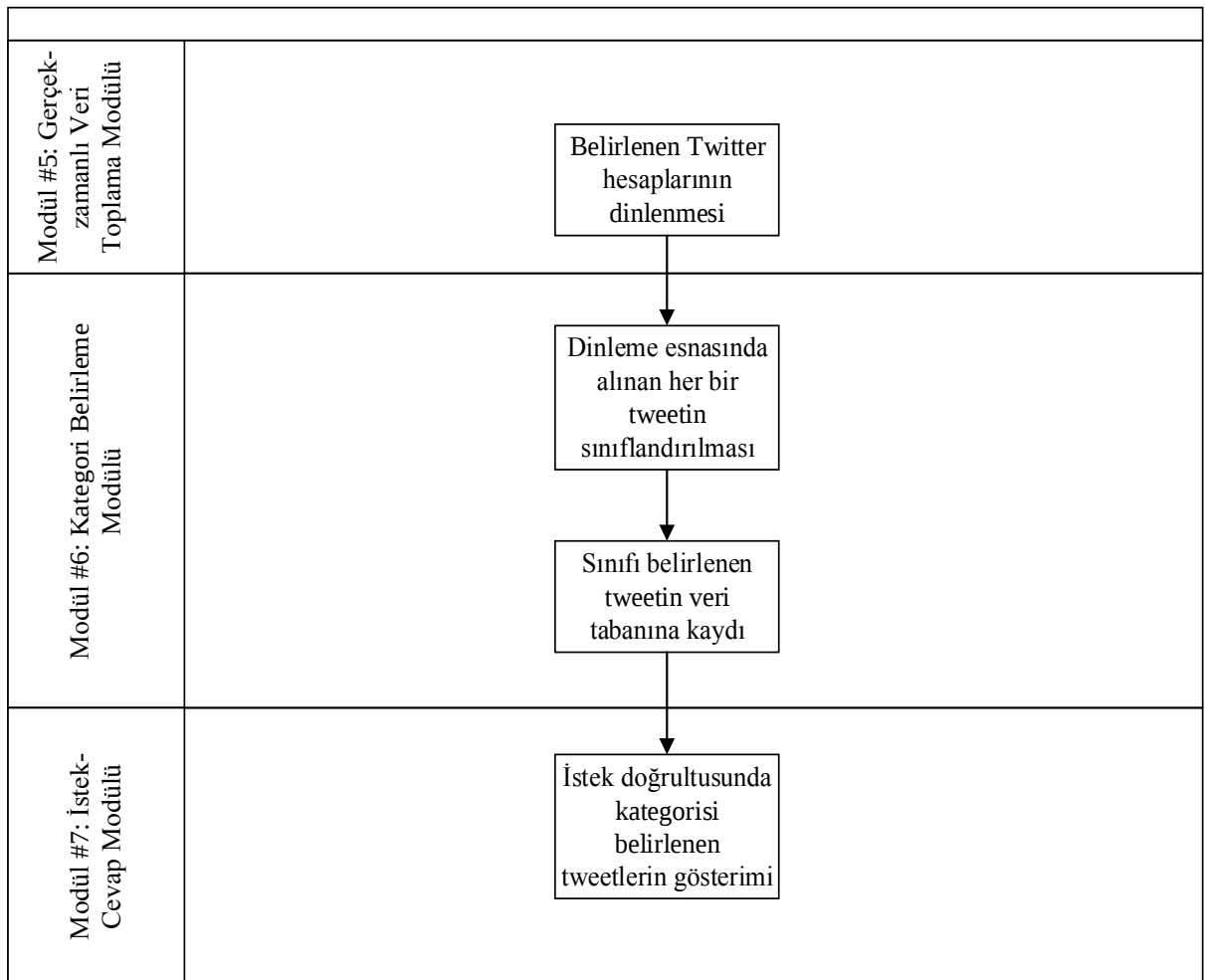
5.2. Geliştirilen Web Sitesi Yazılımı

Bu tez çalışması kapsamında, sınıflandırılarak ait oldukları haber kategorisine yerleştirilen Twitter haber paylaşımlarının internet kullanıcılarının hizmetine sunulması amacıyla bir web sitesi yazılımı geliştirilmiştir. Bu yazılım sayesinde paylaşımları takip edilmek istenen resmi haber kuruluşları ve gazetecilere ait Twitter hesapları, gerçek zamanlı ve dinamik bir şekilde sistem tarafından dinlenmektedir. Dinleme esnasında bu hesaplardan atılan her bir tweet yine geliştirilen sistem tarafından sınıflandırılarak ait olduğu kategori başlığı altına yerleştirilmektedir. Kullanıcı ise geliştirilen bu web sitesi yazılımı sayesinde hangi kategori başlığı altındaki Twitter haberlerine erişim yapmak istiyorsa o kategoriye seçerek sistemden istekte bulunmaktadır ve sistem kullanıcının istem yaptığı kategori altındaki

Twitter haberlerini kullanıcıya sunmaktadır. Bu sayede Twitter hesabı olmayan internet kullanıcılarının da Twitter haberlerini takip edebilmeleri, böylelikle kaliteli ve filtrelenmiş bilgiye daha kısa sürede erişim yapmaları sağlanmaktadır. Yazılım geliştirilirken Java web teknolojileri kullanılmıştır. Bunlar JSF teknolojisinin içerdiği Prime Faces Ana çatıları ve HTML, CSS, AJAX gibi temel web teknolojilerini içermektedir. Bu tez kapsamında geliştirilen yazılıma ait modüller aşağıda kısaca açıklanmıştır.

5.2.1 Geliştirilen modüller

Şekil 5.6’da geliştirilen yazılımın çalışma mantığını özetleyen akış diyagramı sunulmuştur. Akış diyagramındaki her bir modülün görevi aşağıda detaylı bir şekilde açıklanmıştır.



Şekil 5.6. Geliştirilen yazılıma ait akış diyagramı

Şekil 5.6'daki akış diyagramında verilen modüller ve yazılımın çalışma mantığı aşağıda detaylı bir şekilde açıklanmıştır.

Modül #5: Gerçek-zamanlı veri toplama modülü

Geliştirilen yazılımın ilk basamağı olan bu modül ile Twitter üzerinden yaptıkları metin tabanlı haber paylaşımlarının takip edilmek istendiği resmi haber kuruluşları, gazeteciler ve Twitter kullanıcılarının Twitter hesapları dinlenmeye alınır. Bu dinleme işlemi gerçek zamanlı olarak gerçekleştirilmektedir. Yani geliştirilen bu yazılım ile tweet paylaşımları dinlenen hesaplardan herhangi bir zaman aralığında herhangi bir paylaşım yapıldığında bu paylaşımı yakalayarak veri tabanına kaydetmektedir. Twitter üzerinden yapılan paylaşımların gerçek zamanlı olarak alınarak kaydedilmesine olanak sağlayan teknoloji ise yine Twitter tarafından geliştiricilerin hizmetine sunulmuş olan Twitter Streaming UPA'dır [179].

Twitter Streaming UPA

Twitter Streaming UPA, ismindeki 'stream' kelimesinin Türkçe anlamından da anlaşılabilirliği gibi, kullanıcıların Twitter üzerinden 'sürekli ve aralıksız bir şekilde' yaptıkları paylaşımlara, geliştiricilerin gerçek zamanlı olarak erişim yapmasına ve bu paylaşımları depolamasına imkân tanımaktadır [179]. Streaming UPA sayesinde herhangi bir kullanıcı tarafından yalnızca 'herkesin erişimine açık' olacak şekilde paylaşılmış tweetlere, paylaşım yapıldığı anda geliştiriciler tarafından erişim yapılabilmektedir. Dolayısıyla Twitter kullanıcısı, yaptığı paylaşımları gizlilik ayarlarından korumaya almışsa ve gizlemişse bu durumda Twitter'ın sunduğu UPA hizmetleri ile o kullanıcının tweetlerine hiçbir şekilde erişim yapılamamaktadır [148].

Twitter Streaming UPA, geliştiricilere paylaşımları depolanmak istenen kullanıcıların tweetlerine gerçek zamanlı erişim imkânı sunduğundan, bu hizmetin kullanılabilmesi için Streaming UPA ile http bağlantısının sürekli açık olacağı bir veya birden fazla sunucu bilgisayara ihtiyaç vardır. Dolayısıyla Streaming UPA'dan sürekli istekte bulunan bağlantı kapandığında gerçek zamanlı tweet erişimi de sona ermiş olacaktır [179].

Bu tez çalışması kapsamında geliştirilen yazılımda resmi haber kuruluşları ve gazetecilerin Twitter üzerinden yapmış oldukları haber paylaşımları dikkate alınmıştır. Dolayısıyla bu hesaplar kişisel hesaplar olmadığından genellikle yaptıkları tweet paylaşımlarının ‘herkese açık’ olduğu tespit edilmiştir. Çizelge 5.9’da Twitter paylaşımları sistem tarafından dinlenen resmi haber kuruluşları ve gazetecilere ait Twitter hesap bilgilerinin bulunduğu veri tabanı tablosundan bir kesit sunulmuştur.

Çizelge 5.9. Gerçek-zamanlı Twitter paylaşımları sistem tarafından dinlenen resmi haber kuruluşlarına ait Twitter hesap bilgileri

<i>Twitter_ID</i>	<i>Hesap Adı</i>
15016209	NTV
30631061	Cüneyt Özdemir
61168062	CNN
63796949	Ahmet Hakan
65642163	Hürriyet
82943239	Sabah
462921513	Şamil Tayyar
219248343	TRT
556713042	Habertürk
26763070	Fatih Altaylı
1283193666	Fox

Çizelge 5.9’dan görülebildiği gibi, veritabanı tablosunda Twitter paylaşımları sistem tarafından dinlenen hesapların Twitter ID numaraları ve hesap adları bulunmaktadır. Bu hesaplara ait paylaşımlar sistem tarafından dinlenirken Twitter ID numaraları kullanılacaktır. Bunun sebebi ise, Twitter ID numarasının her hesaba özgü olması ve veritabanında birincil anahtar olarak kullanılabilecek olması ve aynı zamanda sayı kullanımının programlama açısından isim kullanımına göre daha kolaylık sağlayacak olmasıdır. Örneğin Çizelge 5.9’dan görülebileceği gibi tweetleri gerçek zamanlı bir şekilde dinlenecek olan NTV haber kuruluşunun resmi Twitter hesabına ait Twitter ID numarası veri tabanına kaydedilmiştir. Dolayısıyla bu aşamadan sonra Streaming API ile veri tabanında bilgileri bulunan Twitter hesapları dinlenmeye başlanarak bu hesaplar üzerinden yapılan her türlü tweet paylaşımı gerçek zamanlı bir şekilde sistem tarafından dinlenerek veri tabanına kaydedilmektedir. Çizelge 5.10’da ise bu hesaplardan gerçek zamanlı bir

şekilde dinlenmiş olan tweetleri ve bu tweetlere ait bilgileri tutan veri tabanı tablosundan bir kesit sunulmuştur.

Çizelge 5.10. Belirlenmiş hesaplara ait gerçek zamanlı bir şekilde toplanmış olan tweet bilgileri

<i>Tweet_ID</i>	<i>Twitter_ID</i>	<i>Tweet</i>	<i>Profil_Foto</i>	<i>Kul_Adı</i>	<i>Hesap_Adı</i>	<i>Bağlantı</i>	<i>Zaman</i>
4840852..	15016209	Baba o..	http://pbs.t..	Ntv	NTV	http://nt..	2014-07..
4858803..	63796949	bunu ya..	http://pbs.t..	ahmethc	ahmet hakan		2014-07..
4843012..	219248343	Kahven..	http://pbs.t..	trthaber	TRT HABER	http://g..	2014-07..
4852018..	12413037..	BİMER..	http://pbs.t..	TCBasb..	TC Başbakan..	http://ba..	2014-07..
4857021..	95248383	Güncell..	http://pbs.t..	TC_Dis..	Dışişleri Bak..	http://m..	2014-08..

Çizelge 5.10'dan görülebildiği gibi belirlenen hesaplardan gerçek zamanlı olarak çekilen her bir tweete ait

- ID numarası,
- Paylaşıldığı hesabın ID numarası,
- Tweet içeriği,
- Paylaşıldığı hesaba ait profil fotoğrafının bulunduğu bağlantı,
- Paylaşıldığı hesaba ait Twitter kullanıcı adı,
- Paylaşıldığı hesaba ait kullanıcı hesap adı,
- Tweet içeriğindeki her bir bağlantı (varsa) ve
- Paylaşım zamanı (gün-ay-yıl, saat-dakika-saniye olarak)

bilgileri veritabanına kaydedilmiştir. Bu bilgilerin depolanma amacı, sistemin işlevi olan; haber paylaşımı olan tweetleri içeriklerine göre sınıflandırarak ilgili olduğu haber kategorisine atamak ve geliştirilen web sitesi yazılımında tweeti tüm bilgileri ile birlikte kullanıcılara sunmaktır.

Modül #6: Kategori belirleme modülü

Geliştirilen web sitesi yazılımının bir parçası olan bu modülün amacı, belirlenen resmi haber kuruluşlarına ait Twitter hesaplarından veri dinleme modülü yardımıyla gerçek zamanlı bir şekilde dinleyerek toplanan tweetleri sınıflandırarak ilgili haber kategorisi başlığı altında

kullanıcıya sunmaktır. Bu kapsamda, veri dinleme modülünün belirlenen hesaplardan gerçek zamanlı olarak yakaladığı her bir tweet, Geliştirilen Yaklaşım başlığı altındaki Modül #4: Sınıflandırma Modülünün geliştirilen web yazılımına entegre edilmesiyle sınıflandırılmıştır. Yapılan sınıflandırma işlemi tüm detaylarıyla birlikte Modül #4: Sınıflandırma Modülü başlığı altında açıklanmıştır. Web yazılımına entegre edilen bu modül Veri Dinleme Modülünün topladığı her bir tweetin kategorisini belirleyerek veri tabanı tablosuna tweeti kategorisi ile birlikte kaydederek işlevini tamamlamaktadır. Çizelge 5.11’de bu aşamadaki veri tabanı tablosundan bir kesit sunulmuştur.

Çizelge 5.11. Kategorisi belirlenmiş tweetler

<i>Tweet_ID</i>	<i>Twitter_ID</i>	<i>Tweet</i>	<i>Profil_Foto</i>	<i>Kul_Adi</i>	<i>Hesap_Adi</i>	<i>Bağlantı</i>	<i>Zaman</i>	<i>Kategori</i>
48408..	150162..	Baba o..	http://pbs.t..	ntv	NTV	http://nt..	2014-07..	7
48588..	637967..	bunu ya..	http://pbs.t..	ahmethc	ahmet hakan		2014-07..	2
48430..	219248..	Kahven..	http://pbs.t..	trthaber	TRT HABER	http://g..	2014-07..	3
48520..	124130..	BİMER..	http://pbs.t..	TCBasb..	TC Başbakan..	http://ba..	2014-07..	7
48570..	952483..	Güncell..	http://pbs.t..	TC_Dis..	Dışişler i Bak..	http://m..	2014-08..	4

Çizelge 5.11’den görülebildiği gibi veri dinleme modülü yardımıyla dinlenen her bir tweet, sınıflandırma modülü yardımıyla sistem tarafından uygun görülen haber kategorisine atanmıştır. Programlama sırasında kolayca erişim sağlanabilmesi açısından sırasıyla ‘Teknoloji’, ‘Spor’, ‘Ekonomi’, ‘Politika’, ‘Magazin’, ‘Sağlık’ ve ‘Kültür-Sanat’ kategorileri 1’den 7’ye kadar numaralandırılmıştır.

Modül #7: İstek-cevap modülü:

Geliştirilen yazılımın son basamağı olan bu modül, web sitesi yazılımı üzerinden kullanıcıdan kategori bilgilerini istek olarak almaktadır. Ve gerçek zamanlı olarak dinlenen ve kullanıcının istem yaptığı kategoriye ait olduğu belirlenen tweetleri o kategori başlığı altında kullanıcıya cevap olarak sunmaktadır. Bu tez çalışmasında geliştirilen web sitesi <http://tweetajans.com/> adresi üzerinde çalışmaktadır [180]. Aşağıdaki şekilde tweetajans.com’un Ana sayfasına ait bir ekran görüntüsü sunulmuştur.



[Anasayfa](#)
[Hakkında](#)
[Biz Kimiz?](#)
[İletişim](#)



Twitter Haberleri

Twitter'da paylaşılan haberleri kategorilerine göre okuyun.

Teknoloji

Spor

Ekonomi

Politika

Magazin

Sağlık

Kültür-Sanat

✓ Gönder

	Habertürk TV @HaberturkTV 07/07/2014 00:25 YAYINDA! Cem Dizdar: '80 milyonuz ve çok yetenekliyiz sanıyoruz' #ÖtekiGündem @ciftpelin @yilmazvuraltr @ebruyuc @OtekiGundemTV	Kategori : Magazin
	Hurriyet.com.tr @Hurriyet 07/07/2014 00:21 Unutmayın, bütün yağlar zararlı değildir. İşte en faydalı 3 yağ http://t.co/fZZs1ta6yG http://t.co/8avaH192o8	Kategori : Sağlık
	Hurriyet.com.tr @Hurriyet 07/07/2014 00:19 Tarihe geçen penaltılar... http://t.co/Z8ApQxkHYg http://t.co/4UQECzQfPH	Kategori : Spor
	Habertürk TV @HaberturkTV 07/07/2014 00:19 YAYINDA! 'Türkiye'de futbolu kim yönetiyor?' @ÖtekiGündem @ciftpelin @yilmazvuraltr @ebruyuc @OtekiGundemTV http://t.co/j1WBAiAUW6	Kategori : Spor

Resim 5.3. tweetajans.com'un ana sayfasına ait bir ekran görüntüsü

Resim 5.3'den de görüldüğü gibi geliştirilen yazılımın açılış sayfasında kullanıcıya resmi haber kuruluşları tarafından kabul görmüş haber kategorileri görsel butonlar halinde sunulmuştur. Ayrıca ana sayfada, kullanıcı herhangi bir kategori seçmemişken, sistemin Twitter'dan çekerek sınıflandırdığı en son atılmış 20 adet gerçek-zamanlı tweet rastgele bir şekilde ana sayfada sıralanmaktadır. Örneğin Şekil 5.3'de kullanıcı tarafından herhangi bir kategori seçilmediğinde sistem rastgele bir şekilde Magazin, Sağlık ve Spor kategorilerine ait paylaşımları ana sayfada göstermektedir.

Kullanıcı haber almak istediği kategoriye ait butona tıkladığında ise paylaşımları gerçek-zamanlı bir şekilde dinlenen haber kuruluşlarının resmi Twitter hesapları üzerinden yayınladığı o kategoriye ait ve zamana göre sıralandığında en son atılmış 20 tweet kullanıcının erişimine sunulmaktadır. Örneğin Resim 5.4'de Spor kategorisi tıkladığında, paylaşımları dinlenen hesaplar üzerinden en son atılmış 20 tweet kullanıcıya sunulmuştur.

Twitter Haberleri

Twitter'da paylaşılan haberleri kategorilerine göre okuyun.

Teknoloji	Spor	Ekonomi	Politika	Magazin	Sağlık	Kültür-Sanat	✓ Gönder
	Habertürk TV @HaberturkTV 07/07/2014 00:27						
	YAYINDA! Yılmaz Vural: 'Türkiye'de hangi takımı seyrederken, futbol seyrediyorum diyorsunuz?' #ÖtekiGündem @ciftpelin @yilmazvuraltr						Kategori : Spor
	Hurriyet.com.tr @Hurriyet 07/07/2014 00:19						
	Tarihe geçen penaltılar... http://t.co/Z8ApQxkHYg http://t.co/4UQECzQfPH						Kategori : Spor
	Habertürk TV @HaberturkTV 07/07/2014 00:19						
	YAYINDA! 'Türkiye'de futbolu kim yönetiyor?' @ÖtekiGündem @ciftpelin @yilmazvuraltr @ebruyuc @OtekiGundemTV http://t.co/j1WBAiAUW6						Kategori : Spor
	Habertürk TV @HaberturkTV 07/07/2014 00:17						
	YAYINDA! 'Futbol nasıl bir güç?' @ÖtekiGündem @ciftpelin @yilmazvuraltr @ebruyuc @OtekiGundemTV http://t.co/j1WBAiAUW6						Kategori : Spor
	Habertürk TV @HaberturkTV 07/07/2014 00:17						
	YAYINDA! Yılmaz Vural: 'Türkiye'de futbolcular cahil bıraktırılıyor' @ÖtekiGündem @ciftpelin @yilmazvuraltr @ebruyuc @OtekiGundemTV						Kategori : Spor
	Habertürk TV @HaberturkTV 07/07/2014 00:15						
	YAYINDA! Yılmaz Vural: 'Biz de eğitilmiş insan yok, futbol zeka isteyen bir oyundur' @ÖtekiGündem @ciftpelin @yilmazvuraltr @ebruyuc						Kategori : Spor

Resim 5.4. Spor kategorisine ait Twitter haberleri

Kullanıcıların tek seferde birçok kategoriye ait haber bilgisine erişebilmesi için kategori butonları çoklu seçilebilir şekilde tasarlanmıştır. Örneğin Şekil 5.5’de Sağlık ve Kültür-Sanat kategorilerine çoklu tıklandığında bu 2 alanda, dinlenen hesaplar üzerinden atılmış son 20 tweet kullanıcıya gösterilmektedir.

Teknoloji	Spor	Ekonomi	Politika	Magazin	Sağlık	Kültür-Sanat	✓ Gönder
	Hurriyet.com.tr @Hurriyet 07/07/2014 00:21						Kategori : Sağlık
Unutmayın, bütün yağlar zararlı değildir. İşte en faydalı 3 yağ http://t.co/fZZs1ta6yG http://t.co/8avaH192o8							
	Sabah Gazetesi @Sabah 07/07/2014 00:13						Kategori : Sağlık
İstanbul'da sahura kadar su yok http://t.co/5tor3ZrkOn							
	Habertürk TV @HaberturkTV 06/07/2014 23:50						Kategori : Sağlık
YAYINDA! Cem Dizdar: 'Avrupa Birliği'nde neden yoksak, Dünya Kupası'nda o yüzden yokuz' @ÖtekiGündem @ciftpelin @yilmazvuraltr @ebruyuc							
	Hurriyet.com.tr @Hurriyet 06/07/2014 23:39						Kategori : Kültür-Sanat
İşte 'çatı aday' #Ekmeleddinİhsanoğlu'nun 43 yıllık eşi Füsun Hanım... http://t.co/RbVHXCWMBj http://t.co/qAgEx3ipXF							
	ahmet hakan @ahmethc 06/07/2014 23:02						Kategori : Kültür-Sanat
#OTdergi / O mu Bu mu Niçin'de bu ay @ahmethc var. @otdergi Temmuz sayısı çıktı. http://t.co/Z37w7FQI9A							
	NTV @ntv 06/07/2014 23:00						Kategori : Kültür-Sanat
Dinlerini değiştirmemek için ormanlara sığındılar; Mişer Tatarları.Türk ve İslam tarihinin az bilinen öyküleri.Zaman Yolcusu 23.10'da NTV'de							
	NTV @ntv 06/07/2014 23:00						Kategori : Kültür-Sanat
Zaferlerine bir yenisini daha ekledi http://t.co/NdNwY4Hkx http://t.co/6CvLFibK2S							
	Sabah Gazetesi @Sabah 06/07/2014 22:00						Kategori : Kültür-
Kartal'dan güzel başlangıç http://t.co/vin3wPnxvI							

Resim 5.5. Sağlık ve Kültür-Sanat kategorilerinin çoklu seçimi

tweetajans.com'da ana sayfa dışında “Hakkında”, “Biz Kimiz?” ve “İletişim” sayfaları bulunmaktadır. EK-3'deki Resim 3.1 (a) ve Resim 3.1 (b)'de sırasıyla “Hakkında” ve “Biz Kimiz?” sayfalarına ait ekran görüntüleri bulunmaktadır.

Ayrıca internet kullanıcılarının sistem hakkındaki fikir ve görüşlerini iletebilmeleri için EK-3'deki Resim 3.2'de ifade edildiği üzere bir iletişim sayfası hazırlanmıştır. Bu iletişim sayfası üzerinden kullanıcıların fikir ve görüşleri alınarak değerlendirilmek üzere veri tabanına kaydedilmektedir.

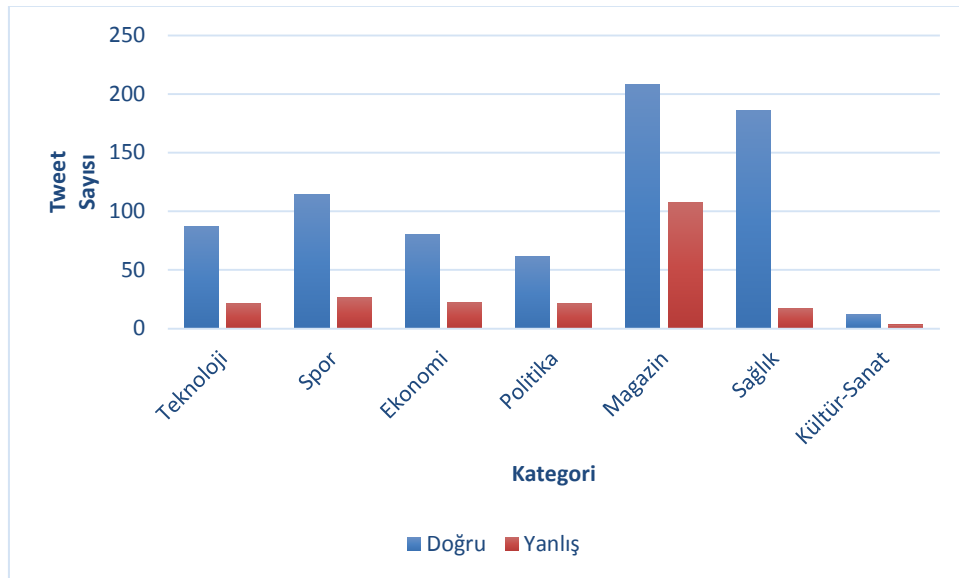
5.2.2. Çevrimiçi performans testi

Geliştirilen web sitesi yazılımının başarısının test edilebilmesi için, sistemin gerçek zamanlı olarak topladığı tweetleri doğru haber kategorilerine atama yüzdesinin tespit edilmesi gerekmektedir. Bu amaçla, tweetajans.com sisteminin, özellikle kategori bazında haber paylaşımı yapan Twitter hesaplarını dinlemesi sağlanmıştır. Böylelikle belirli kategori sayfalarından atılan tweetleri, sistemin aynı kategori başlığı altına yerleştirme yüzdesi hesaplanarak sistemin başarıyı test edilmiştir. EK-4 Çizelge 4.1'de 23.06.2014 - 29.06.2014 tarihleri arasındaki 1 haftalık süre içerisinde kategori bazında paylaşım yaptığı tespit edilen ve bu nedenle sistemin testi için takibe alınan Twitter hesaplarına ait bilgiler verilmiştir. Bu hesaplar dinlemeye alındıktan sonra, 1 haftalık süre içerisinde her bir kategori için belirlenen sayfalardan atılan tweetlerin toplam sayısı, dinlenen hesaplar ile sistem tarafından aynı kategoriye atanmış tweet sayısı ve yanlış kategoriye atanan tweet sayısı hesaplanmış, elde edilen rakamlar Çizelge 5.13'de verilmiştir.

Çizelge 5.13. Geliştirilen yazılıma ait çevrimiçi performans testi sonuçları

İstatistik/Kategori	Teknoloji	Spor	Ekonomi	Politika	Magazin	Sağlık	Kültür-Sanat	Genel
Toplam tweet sayısı	108	140	102	82	315	203	15	965
Doğru tweet sayısı	87	114	80	61	208	186	12	748
Yanlış tweet sayısı	21	26	22	21	107	17	3	217
Çevrimiçi performans testi başarı oranı (%)	80,56	81,43	78,43	74,39	66,03	91,63	80	77,51
Çevrimdışı test başarı oranı (%)	82,78	85,58	73,46	77,16	68,34	87,64	71,92	78,12

Çizelge 5.13’de verilen bilgiler, Şekil 5.7’de grafik halinde sunulmuştur.



Şekil 5.7. Performans Testi sonuçlarının grafiksel gösterimi

Elde edilen sonuçlar sistemin çevrimdışı test aşamasındaki rakamlar ile kıyaslandığında, sistemin sınıflandırma başarısı,

- Teknoloji kategorisi için %2,22,
- Spor kategorisi için %4,15,
- Politika kategorisi için %2,77 ve
- Magazin kategorisi için %2,31

beklendiği üzere daha düşük bulunmuştur. Bunun yanı sıra başarımlar oranları,

- Ekonomi kategorisi için %4,97,
- Sağlık kategorisi için %3,99 ve
- Kültür-Sanat kategorisi için %8,08

daha yüksektir. Sistemin genel sınıflandırma başarısının ise %1 daha düşük olduğu görülmüştür.

6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında sosyal ağların günümüzde özellikle haber kaynağı olarak kullanılmasından yola çıkılarak, herhangi bir sosyal ağ üzerindeki metin tabanlı haber paylaşımlarını haber kuruluşları tarafından kabul görmüş kategori başlıkları altında sınıflandıran ve bu sınıflandırılmış haberleri kullanıcılara görsel bir web sitesi yazılımı üzerinden internet kullanıcılarına sunan bir yaklaşım önerilmiştir. Önerilen yaklaşımın başarımları ve performansının kabul edilebilir olduğunu göstermek için de ayrıca bir sistem geliştirilmiştir. Geliştirilen sistemde, Twitter sosyal ağı üzerinden resmi haber kuruluşları ve gazeteciler tarafından yayınlanan haber paylaşımlarının gerçek zamanlı ve dinamik bir şekilde toplanması, sınıflandırılması ve ait olduğu haber kategorisinin tespit edilmesi işlemi gerçekleştirilmiş ve görsel olarak kullanıcıya sunulmuştur. Ayrıca geliştirilen yazılım sosyal ağlar üzerinden yayınlanan haber paylaşımlarının sınıflandırılması işlemini gerçekleştirebildiği gibi, bu platformlar üzerinden yayınlanan her türlü metin tabanlı paylaşım verisini arzu edilen kategori başlıkları altında sınıflandırabilecek şekilde esnek bir yazılım olarak tasarlanmıştır.

Geliştirilen sistemin bir parçası olan sınıflandırma modülü, öncelikle belirli sayıdaki Twitter haber paylaşımı ile eğitilerek, sistem tweetlerin içerdiği kelimeler ile ait oldukları kategoriler arasındaki örüntüyü öğrenmiştir. Daha sonra eğitim verisi dışında kalan Twitter haber paylaşımları sistemin başarımlarını test etmek amacıyla kullanılmıştır. Elde edilen sonuçlar bu çalışmada geliştirilen sınıflandırma modülünün haber içerikli metin tabanlı Twitter paylaşımlarını sırasıyla,

- Teknoloji kategorisi için %82,78
- Spor kategorisi için %85,58
- Ekonomi kategorisi için %73,46
- Politika kategorisi için %77,16
- Magazin kategorisi için %68,34
- Sağlık kategorisi için %87,64
- Kültür-Sanat kategorisi için %71,92

başarım oranıyla sınıflandırmıştır. Sistemin ortalama sınıflandırma başarımı ise %78,12 olarak belirlenmiştir. Literatürde Türkçe metin tabanlı sosyal ağ paylaşımlarının sınıflandırılması üzerine herhangi bir çalışma bulunmadığından elde edilen sonuçlar Türkçe literatür için kıyaslanamamıştır. Fakat İngilizce dil yapısı Türkçe'den oldukça farklı olmasına karşın İngilizce metin tabanlı sosyal ağ paylaşımlarının sınıflandırıldığı çalışmalar ile kıyaslandığında, geliştirilen sistemin %78'lik bir oranla oldukça makul, yüksek ve kabul edilebilir başarımla sahip olduğu görülmüştür. Kategoriler arasında en yüksek başarımla sahip olan sağlık, spor ve teknoloji kategorilerinin kendi içerisinde oldukça belirleyici kelimeler içermelerinin bu sonucun elde edilmesinde etkili olduğu yorumlanmaktadır. En düşük başarımla sahip olan magazin ve kültür-sanat kategorilerinde elde edilen sonuçların ise bu kategori başlıkları altında yayınlanan haberlerin birbirleriyle ilişkili olmasından kaynaklandığı olarak değerlendirilmektedir.

Geliştirilen sistemin sınıflandırma modülünün ürettiği başarımlar kabul edilebilir olduğundan Twitter üzerinden yayınlanan metin tabanlı haber paylaşımları gerçek zamanlı bir şekilde sistem tarafından çekilmektedir. Çekilen bu paylaşımların içerdikleri haber bilgisinin türü sistem tarafından tespit edilerek ait olduğu haber kategorisi başlığı altında görsel bir web sitesi yazılımı üzerinden internet kullanıcılarına sunulmaktadır. Geliştirilen yazılımın çevrimiçi ve gerçek zamanlı sınıflandırma başarısının test edilmesi için ise sisteme çevrimiçi performans testi uygulanmıştır. Bu test neticesinde elde edilen sonuçlar; sistemin teknoloji, spor, politika ve magazin kategorileri için sınıflandırma oranının daha düşük olduğunu gösterirken, ekonomi, sağlık ve kültür-sanat kategorileri için daha yüksek sonuçlar ürettiğini göstermiştir.

Bu durumun sebebi değerlendirildiğinde ise:

- Test başarımının seçilen verilere göre artış veya düşüş gösterebileceği,
- tweetajans.com'un test edildiği 23.06.2014-29.06.2014 tarihleri arasında gündemin değiştiği,
- buna bağlı olarak sistemin test edildiği 1 haftalık süre içerisinde atılan haber içerikli tweetlerin içerdiği kelimelerin, gündeme bağlı olarak eğitim kümesinin içerdiği kelimelerden farklılaştığı ve

- gün geçtikçe sistemin öğrenmediği kelimelerle karşılaşma ihtimalinin artmasına bağlı olarak sınıflandırma başarısında düşüş yaşanabileceği

görülmektedir.

Sistemin genel başarıml oranı literatürde kabul edilebilir bir yere sahip olsa da karakter sınırlaması olmayan bir sosyal ağ üzerinden yapılan haber paylaşımlarının daha yüksek başarıml oranları ile sınıflandırılabilceği öngörülmektedir. Twitter üzerinden yapılan haber paylaşımları Twitter'a ait karakter sınırlamasından dolayı çok az kelime içermektedir. Bu durum da ağırlıklandırma işlemin, doğrudan etkilediğinden sistemin başarıml oranını aşağı çekmektedir. Dolayısıyla bu çalışma kapsamında önerilen yöntemin Twitter dışındaki diğer sosyal ağlar üzerinde uygulanması ve başarıml oranlarının kıyaslanması araştırmlara kaynak olabilecek bir çalışma olabileceği değerlendirilmektedir.

Bütün bunların yanı sıra, Türkçe bir kelimenin birçok anlamı bulunabilmektedir. Dolayısıyla bir kelimenin her bir kategori için taşıdığı anlam değişken olabilmektedir. Bu durum da kategoriler arasındaki ortak kelimelerin sistemin başarıml performansını etkilediği görülmüştür.

Bu tez çalışması kapsamında geliştirilen sistemin en büyük avantajı her türlü sosyal ağa ait Türkçe metin tabanlı paylaşımları arzu edilen kategori başlıkları altında sınıflandırabilecek esnek bir şekilde tasarlanmış olması ve görsel bir şekilde kullanıcıya sunulabilecek yapıda geliştirilmesidir. Dolayısıyla geliştirilen sistem, yalnızca Twitter sosyal ağına ait haber paylaşımlarını değil, herhangi bir sosyal ağa ait paylaşımları arzu edilen kategori başlıkları altında sınıflandırıp kullanıcıya görsel bir web sitesi yazılım üzerinden sunabilecek şekilde tasarlanmıştır. Ayrıca bu sistem sayesinde internet kullanıcılarının belirlenmiş sosyal ağ üzerinde hesap açmasına gerek kalmadan, o sosyal ağ üzerinden yayınlanan haber paylaşımları internet kullanıcıları tarafından takip edilebilecektir.

Bu tez çalışması kapsamında Türkçe metin tabanlı sosyal ağ haber paylaşımlarını gerçek zamanlı bir şekilde yayınlandığı sosyal ağ üzerinden çekip sınıflandıran ve haberleri ait oldukları kategori başlıkları altında bir web sitesi yazılımı üzerinden görsel bir şekilde

internet kullanıcılarına sunan zeki bir sınıflandırıcı geliştirilmiştir. Geliştirilen sisteme dair elde edilen sonuçlar aşağıda maddeler halinde sunulmuştur.

- Naive Bayes sınıflandırıcısının Türkçe metin sınıflandırmada başarılı olduğu,
- Metnin içerdiği kelimelerin ağırlıklandırılması işleminin sınıflandırma başarısına doğrudan etki ettiği ve TF-TDF ağırlıklandırma yöntemi ile yüksek başarımlarına ulaşıldığı,
- Sınıflandırma başarımı açısından eğitim kümesinin büyük önem arz ettiği ve eğitim kümesinin içerdiği metin tabanlı sosyal ağ paylaşımlarının ait olduğu kategorilerin el ile belirlenmesinin sınıflandırma başarımını artıracacağı,
- Tasarlanan sistemin gerçek-zamanlı sınıflandırma başarımı test edildiğinde, test kümesinden elde edilen başarımlarından ortalama %1’lik bir oranda düşüş yaşandığı,
- Bu düşüşün, her geçen gün gündemin değişmesiyle birlikte sistemin karşılaştığı kelimelerin farklılaşmasına bağlı olabileceği,
- Geliştirilen sistemin farklı sosyal ağlara ve farklı alanlarda farklı amaçlarla yapılacak olan sınıflandırma işlemlerine kolaylıkla uyarlanabileceği,
- Ülkemizde ve dünyada sosyal ağlar üzerinden yayınlanan haber paylaşımlarının sınıflandırılarak ait oldukları haber kategorisi başlıkları altında görsel bir şekilde internet kullanıcılarına sunulduğu herhangi bir çalışmaya rastlanmadığı ve dolayısıyla bu çalışmanın ülkemiz ve dünya için bir ilk niteliğinde olduğu,
- Önerilen yöntemin dünya üzerinde Türkçe dışındaki diğer diller için uyarlanmasının yeni çalışmalara kaynak oluşturacağı ve
- Bu çalışmanın bu kapsamda yapılacak olan diğer bütün çalışmaların önünü açacağı

görülmüştür.

Geliştirilen yazılımın testi amacıyla Twitter üzerinden gerçek zamanlı bir şekilde çekilerek sınıflandırılan paylaşımlardan elde edilen başarımlarının, sistemin bir defaya mahsus olmak üzere eğitildikten sonra gerçekleştirilen test işlemi sonucunda ürettiği başarımlarından ortalama %1 oranında daha düşük olduğu görülmüştür. Bu durumun sistemin test edildiği veri kümesinin içerdiği paylaşımlardan kaynaklandığı düşünülmektedir. Bunun yanı sıra, gerçek zamanlı olarak sınıflandırılan haber paylaşımları manuel bir şekilde incelendiğinde, gündemin değişmesine bağlı olarak bu paylaşımlarda sistemin daha

önce öğrenmediği kelimelerin bulunduğu bu durumun da başarı oranlarını değiştirdiği gözlemlenmiştir.

Bu çalışmanın,

- İnternet kullanıcıları tarafından sosyal ağlar üzerinden yapılan haber paylaşımlarının o sosyal ağa ait profil hesabı olmadan da takip edilebilmesine,
- Büyük veri boyutundaki sosyal ağ verisinin sınıflandırılarak içeriğinin belirlenmesine,
- Böylelikle internet kullanıcılarının ihtiyaç duydukları bilgiye daha kısa sürede daha kolay bir şekilde erişim yapabilmesinin sağlanmasına,
- İnternet kullanıcılarının ihtiyaç duydukları bilgiyi büyük veri akışı içerisinde gözden kaçırma durumlarının ortadan kaldırılmasına,
- İnternetin ve sosyal ağların daha verimli kullanılmasına

katkıları sağlayacağı değerlendirilmektedir.

Ayrıca sistem, farklı alanlarda farklı amaçlarla gerçekleştirilebilecek olan sınıflandırma işlemleri için esnek bir yapıda tasarlandığından kullanıcıların yalnızca arzu ettikleri kategori başlıklarını belirleyerek sahip oldukları veriyi sisteme sunmaları ile birlikte sistemin istenilen sınıflandırma işlemini gerçekleştirebileceği değerlendirilmektedir.

Bu çalışma kapsamında,

- Sistem başarımını yükseltmek için eğitim ve test kümesinin içerdiği paylaşımların ait oldukları kategorilerin manuel olarak etiketlenmesi,
- Hatalı sonuçların kullanıcıdan alınacak geri bildirimler doğrultusunda eğitime dâhil edilerek sistemin dinamik eğitiminde kullanılması ve böylelikle başarımın iyileştirilmesi,
- Paylaşımların içerdiği kelimelerin kök ve eklerinin daha etkili bir şekilde ayrıştırılması

işlemlerinin sistemin başarısını artıracakları öngörülmektedir.

Son olarak, sistemin henüz kullanıcılardan geri bildirim alma ve kendi kendine öğrenme süreci bulunmamaktadır. Bu çalışma kapsamında geliştirilen sisteme geri bildirim mekanizması eklenmesi ve sınıflandırma işlemi için kullanıcılardan alınan geri bildirim doğrultusunda sistemin öğrenme sürecinin dinamikleştirilmesi işlemi geliştirilen sistem ve yazılımın ileri seviyelere taşınması amacıyla gelecek çalışmalarda gerçekleştirilecektir. Böylelikle, sistemin güncel konulardan da haberdar olması ve başarı oranının artırılması hedeflenmektedir. Ayrıca, geliştirilen web sitesi yazılımı üzerinde haber tweetlerinin gösteriminde zamana göre ve önem derecesine göre kategori seçeneklerinin oluşturulması ve kullanıcının haber içerikli tweetleri istediği şekilde takip edebilmesinin sağlanması da çalışmanın gelecek hedefleri arasında yer almaktadır. Bu kapsamda kullanıcıların tweetajans.com üzerinde kelime tabanlı haber tweeti araması yapabilmesi, belirli zaman aralıklarında yayınlanan haber tweetlerini görüntüleyebilmesi, gündem konularını baz alarak tweetleri önem derecesine göre sıralayabilmesi ve sistemin tweetleri haber kaegorilerine atarken belirlediği olasılık değerleri bazında da kullanıcıların eldeki bilgiyi filtreleyebilmeleri gerçekleştirilecek çalışmalar arasındadır.

KAYNAKLAR

1. Odlyzko, A. (2000). The history of communications and its implications for the Internet. *AT&T Labs – Research Preliminary Version*, 1-160.
2. Yavanoğlu, U., Gündüz, S. and Sağiroğlu, Ş. (2013, April). *Sosyal ağ paylaşımlarından adli profil çıkarılması için yeni bir yaklaşım*. First International Symposium on Digital Forensics and Security, Elazığ, 440-445.
3. İnternet: Türkiye İstatistik Kurumu Hanehalkı Bilişim Teknolojileri Kullanım Araştırması. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Ftuik.gov.tr%2FPreHaberBultenleri.do%3Fid%3D13569&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
4. İnternet: Ahmad, I. Social Media Today Global Internet, Mobile and Social Media Engagement and Usage Stats and Facts. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fsocialmediatoday.com%2Firan-ahmad%2F1993606%2Fglobal-overview-internet-mobile-and-social-media-engagement-and-usage-infographi&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
5. İnternet: Social Media Predictions for 2014: What Do You Need to Watch?. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fsocialmediatoday.com%2F2009796%2Fsocial-media-predictions-2014-what-do-you-need-watch-webinar-replay&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
6. Wang, T. and Krim, H. (2012, March). *Statistical classification of social networks*. International Conference on Acoustics, Speech and Signal Processing, Kyoto, 3977-3980.
7. Heatherly, R., Kantarcioglu, M. and Thuraishingham, B. (2009, June). *Social network classification incorporating link*. International Conference on Intelligence and Security Informatics, Arizona, Dallas, TX, 19-24.
8. İnternet: Academia.edu. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fwww.academia.edu%2F&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
9. İnternet: Hook, S. R. V. Professional Social Networks. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.saylor.org%2Fsite%2Fwp-content%2Fuploads%2F2013%2F03%2FCUST105-4.3-Networks-FINAL.pdf&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
10. İnternet: Türkiye|LinkedIn. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Ftr.linkedin.com&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.

11. Couch, D. and Liamputtong, P. (2009). Online dating and mating. *Qualitative Health Research*, 18(2), 268-279.
12. İnternet: Mary, G. W. What Types of Social Networks Exist?. URL: [http://www.webcitation.org/query?url=http%3A%2F%2Fsocialnetworking.lovetoknow.com%2FWhat Types of Social Networks Exist&date=2014-08-18](http://www.webcitation.org/query?url=http%3A%2F%2Fsocialnetworking.lovetoknow.com%2FWhat+Types+of+Social+Networks+Exist&date=2014-08-18), Son Erişim Tarihi: 18.08.2014.
13. İnternet: Yu, Z. and Xie, X. Microsoft Research Location Based Social Networks. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fprojects%2Flbsn%2Fdefault.aspx&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
14. Sui, D. and Goodchild, M. (2011). The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*, 25(11), 737-1748.
15. İnternet: Military Social Networking Service-Connected. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.service-connected.com%2F&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
16. Lewis, V. A., MacGregor, C. A. and Putnam, R. D. (2013). Religion, networks, and neighborliness: the impact of religious ocial networks on civic engagement. *Social Science Research*, 42(2), 331-346.
17. İnternet: Religious Social Networking Sites. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fipost.christianpost.com%2Fnews%2Freligious-social-networking-sites-10043%2F&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
18. Axhausen, K. W. (2005). Social networks and travel: some hypotheses. *Social Dimensions of Sustainable Transport: Transatlantic Perspectives*, 90-108.
19. Harrison, R. And Thomas, M. (2009). Identity in online communities: social networking sites and language learning. *International Journal of Emerging Technologies & Society*, 7(2), 09-124.
20. Chen, Y., Xu, X. And Wang, Z. (2012, May). *Family-oriented social network and services*. International Joint Conference on Service Sciences, Shanghai, 217-21.
21. İnternet: Fee, J. 10 Enterprise Networks to Improve Company Communication. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fmashable.com%2F2013%2F06%2F14%2Fenterprise-social-networks%2F&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.

22. İnternet: Ressamlara ve resim severlere yönelik sosyal platform. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.ressam.com%2F+&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
23. Chau, M. and Xu, J. (2007). Mining communities and their relationships in blogs: a study of online hate groups. *International Journal of Human-Computer Studies*, 65 (1), 57-70.
24. İnternet: Twitter Developers. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fdev.twitter.com%2F&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
25. İnternet: DataSift, Power any decision with social data. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fdatasift.com%2F+&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
26. İnternet: Gnip, The source for social data. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fgnip.com%2F+&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
27. İnternet: SurveyMonkey Facebook anketleri. URL: <https://tr.surveymonkey.com/mp/facebook/>, Son Erişim Tarihi: 18.08.2014.
28. Catanese, S. A., Meo, P. D., Ferrara, E., Fiumara, G. and Provetti, A. (2011, May). *Crawling Facebook for social network analysis purposes*. International Conference on Web Intelligence, Mining and Semantics, Sogndal.
29. İnternet: Markapoulou, A., Gjoka, M., Kurant, M. and Butts, C. T. Crawling online social networks. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fodysseas.calit2.uci.edu%2Flib%2Fexe%2Ffetch.php%2Fpublic%3Aathina-nmmc-may2011.pdf&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
30. İnternet: Google Analytics ile Sosyal Medya Ölçümü. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.google.com.tr%2Fintl%2Ftr%2Fanalytics%2Ffeatures%2Fsocial.html&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
31. İnternet: BoomSocial, Sosyal Medya Ölçümleme Raporlama Analiz. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.boomsocial.com&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
32. İnternet: Dekatlon Buzz, Sosyal Medya İletişim Ajansı. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.dekatlonbuzz.com%2Ftr&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
33. İnternet: Brandwatch, Social Media Monitoring Tool, URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.brandwatch.com+&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.

34. İnternet: Temizel, T. T. Electronic Word of mouth communication. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fonline.metu.edu.tr%2FnetclassR%2Flecturenotes%2F155311%2Findex.html&date=2014-08-18>, Son Erişim Tarihi: 15.02.2014.
35. Brown, J., Broderick, A. J. and Lee, N. (2007). Word of mouth communication within online communities: conceptualizing the online social network. *Journal of interactive marketing*, 21(3), 2-20.
36. İnternet: Monitera. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.monitera.com+&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
37. İnternet: Sense Monitoring. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.sensekit.com&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
38. İnternet: Temizel, T. T. Social media as business. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fonline.metu.edu.tr%2FnetclassR%2Flecturenotes%2F155311%2Findex.html+&date=2014-08-18>, Son Erişim Tarihi: 15.02.2014.
39. Sterne, J. (2010). *Social media metrics: how to measure and optimize your marketing investment*, John Wiley & Sons, Inc, 1-256.
40. Zappen, J. P., Harison, T. M. and Watson, D. (2008, May). *A new paradigm for designing e-government: web 2.0 and experience design*, International conference on Digital Government Research, Montreali.
41. Bhuiyan, M. S. H. (2010, October). *E-Government applications in Bangladesh -status and challenges*. 4th International Conference on Theory and Practice of Electronic Governance, 255-260.
42. İnternet: Demirbaş, M. E-devlet ve sosyal medyanın kullanımı. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.slideshare.net%2Fmmdemirbas%2Fe-devlet-ve-sosyal-medya&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
43. İnternet: Crumbs, Places and Augmented reality in Social Networks. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fcrumbs.tid.es&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
44. İnternet: ntvmsnbc, Devlet memurlarına Facebook tavsiyesi. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.ntvmsnbc.com%2Fid%2F25029312&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
45. İnternet: Radikal, Devlet sosyal medyayı Mavi Oda'dan izleyecek. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.radikal.com.tr%2Fturk>

- [iye%2Fdevlet sosyal medyayi mavi odadan izlenecek-1140191&date=2014-08-18](http://www.webcitation.org/query?url=http%3A%2F%2Fsosyalmedya.co%2Fdevlet-sosyal-mediyayi-mavi-odadan-denetleyecek+&date=2014-08-18), Son Erişim Tarihi: 18.08.2014.
46. İnternet: Devlet Sosyal Medyayı ‘Mavi Oda’dan Denetleyecek. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fsosyalmedya.co%2Fdevlet-sosyal-mediyayi-mavi-odadan-denetleyecek+&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
 47. Yavanoglu, U., Caglar, B., Milletsever, O., Colak, M., Cakir, S. and Sagiroglu, S. (2013, December). *Intelligent approach for identifying political views over social networks*. International Conference on Machine Learning and Applications, Miami, Florida, 281-287.
 48. Hoffman, S., Rackers, M., Beverungen, D. and Becker, J. (2013, January). *Old blunders in new media? How local governments communicate with citizens in online social networks*. International Conference on System Sciences, Hawaii, 2023-2032.
 49. Yi, M., Oh, S. G. and Kim, S. (2013). Comparison of social media use for the U.S. and the Korean governments. *Government Information Quarterly*, 30(3), 310-317.
 50. Mossberger, K., Wu, Y. and Crawford, J. (2013). Connecting citizens and local governments? Social media and interactivity in major U.S. cities. *Government Information Quarterly*, 30(4), 351-358.
 51. Abdelsalam, H. M., Reddick, C. G., Gamal, S. and Alshaar, A. (2013). Social media in Egyptian government websites: Presence, usage, and effectiveness. *Government Information Quarterly*, 30(4), 406-416.
 52. İnternet: Twitter. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Ftwitter.com&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
 53. Lee, K., Palsetia, D., Narayanan, R., Patwary, M. A., Agrawal, A. and Choudhary, A. (2011, December). *Twitter trending topic classification*. International Conference on Data Mining Workshops, Vancouver, 251-258.
 54. Nikolov, S. and Shah, D. (2012). A nonparametric method for early detection of trending topics. *Department of EECS, Massachusetts Institute of Technology*, 1-2.
 55. İnternet: MIT, Predicting what topics will trend on Twitter. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fweb.mit.edu%2Fpress%2F2012%2Fpredicting-twitter-trending-topics.html+&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
 56. Li, N. and Chen, G. (2010). Sharing location in online social networks. *Journal IEEE Network*, 24 (5), 20-25.

57. He, W., Liu, X., Ren, M. (2011, June). *Location cheating: a security challenge to location-based social network services*. International Conference on Distributed Computing Systems, Minneapolis, 740-749.
58. Noulas, A., Scellato, S., Mascolo, C. and Pontil, M. (2011, July). *An empirical study of geographic user activity patterns in Foursquare*. International AAAI Conference on Weblogs and Social Media, Barcelona, 570-573.
59. Cheng, Z., Caverlee, J., Lee, K. and Sui, D. (2011, July). *Exploring millions of footprints in location sharing services*. International AAAI Conference on Weblogs and Social Media, Barcelona, 81-88.
60. Fusco, S., Michael, K., Michael, M. and Abbas, R. (2010, June). *Exploring the social implications of location based social networking*. International Conference on Mobile Business, Athens, 230-237.
61. Quercia, D., Kosinski, M., Stillwell, D. and Crowcroft, J. (2011, October). *Our twitter profiles, ourselves: predicting personality with Twitter*. International Conference on Privacy, Security, Risk, and Trust, Boston, MA, 80-185.
62. Hamburger, Y. A. and Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, 26(6), 1289-1295.
63. Han, E. H. S. and Karypis, G. (2005, October). *Feature Based recommendation system*. International Conference on Information and knowledge management, Bremen, 446-452.
64. Feitosa, R. M., Santos, A. L. S., Labidi, S. and Santos, N. (2013, January). *Social recommendation in location-based social network using text mining*. International Conference on Intelligent Systems, Modelling and Simulation, Bangkok, 67-72.
65. Kwon, H. J. and Hong, K. S. (2012, January). *Personalized real-time location-tagged contents recommender system based on mobile social networks*. International Conference on Consumer Electronics, Las Vegas, 558-559.
66. Özcan, A. and Ögüdücü, Ş. G. (2010, June). *A recommendation framework for mobile phones based on social network data*. International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, Berlin, 139-150.
67. İnternet: İletken, Kurumsal ve Size Özel Çözümler. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.iletken.com.tr%2Findex.php%2Ftr%2Fsolutions%2Fenterprisecustom.html+&date=2014-08-18>, Son Erişim Tarihi: 15.02.2014.
68. İnternet: Cisco, Social Media: The New Crime Fighting Tool. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fnewsroom.cisco.com%2Ffeed>

[turecontent%3Ftype%3Dwebcontent%26articleId%3D1289540+&date=2014-08-18](http://www.webcitation.org/query?url=http%3A%2F%2Fwww.wikicrimes.org%2Fmain.html%3Bjsessionid%3D645AD909E3150C69919C03265695F552&date=2014-08-18),
Son Erişim Tarihi: 15.02.2014.

69. İnternet: WikiCrime, Mapping crimes collaboratively. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.wikicrimes.org%2Fmain.html%3Bjsessionid%3D645AD909E3150C69919C03265695F552&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
70. İnternet: Guenther, A. J. LexisNexis, Role of Social Media in Law Enforcement Significant and Growing. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.lexisnexis.com%2Fen-us%2Fabout-us%2Fmedia%2Fpress-release.page%3Fid%3D1342623085481181&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
71. İnternet: Kelly, H. CNN, Police embrace social media as crime-fighting tool. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.cnn.com%2F2012%2F08%2F30%2Ftech%2Fsocial-media%2Ffighting-crime-social-media&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
72. Zainudin, N. M., Merabti, M. and Jones, D. L. (2011, November). *Online social networks as supporting evidence: A digital forensic investigation model and its application design*. International Conference on Research and Innovation in Information Systems, Kuala Lumpur, 1-6.
73. Gunduz, S., Yavanoglu, U. And Sagiroglu. S. (2013, December). *Predicting next location of Twitter users for surveillance*. International Conference on Machine Learning and Applications, Miami, FL, 267-273.
74. Seaman, J. and Tinti-Kane, H. (2013). Social media for teaching and learning. *Pearson Survey Research Group*, 1-32.
75. Koç, M. and Karabatak, M. (2011, September). *Sosyal ağların öğrenciler üzerindeki etkisinin veri madenciliği kullanılarak incelenmesi*. International Computer & Instructional Technologies Symposium, Elazığ, 1-6.
76. Gülbahar, Y., Kalelioğlu, F. ve Madran, R. O. (2010, Aralık). *Sosyal ağların eğitim amaçlı kullanımı*. Türkiye’de İnternet Konferansı’nda sunuldu, İstanbul.
77. Ajjan, H. And Hartshorne, R. (2008). Investigating faculty decisions to adopt Web 2.0 technologies: Theory and empirical tests. *Internet and Higher Education*, 11(2), 71-80.
78. İnternet: EBA, Eğitim Bilişim Ağı, URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.eba.gov.tr+&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.

79. İnternet: Digital Yerliler. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fdigitalyerliler.com&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
80. İnternet: SMILE-Social Media in Learning and Education. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.eun.org%2Fteaching%2Fsmile&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
81. İnternet: Temizel, T. T. Sentiment Analysis. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fonline.metu.edu.tr%2FnetclassR%2Flecturenotes%2F155311%2Findex.html+%&date=2014-08-18>, Son Erişim Tarihi: 15.02.2014.
82. Ceron, A., Curini, L., Mlacus, S. and Porro, G. (2013). Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, 1(1), 1-19.
83. Baracho, R. M. A., Silva, G. C., Ferreira, L. G. F. (2012, September). *Sentiment analysis in social networks: a study on vehicles*. CEUR Workshop Proceedings, Germany, 132-143.
84. İnternet: Sentiment140, Discover the Twitter sentiment for a product or brand. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.sentiment140.com&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
85. İnternet: Repustate, Sentiment analysis and social media analytics. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fwww.repustate.com%2F+%&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
86. İnternet: Türkçe sentiment analizi – Botego. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.sentimentanalizi.com+%&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
87. Alterovitz, S. S. R. And Mendelsohn, G. A. (2013). Relationship goals of middle-aged, young-old, and old-old internet daters: An analysis of online personal ads. *Journal of Aging Studies*, 27(2), 159-165.
88. İnternet: Facebook Power Editor. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fwww.facebook.com%2Fads%2Fmanage%2Fpowereditor+%&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
89. İnternet: Türkiye Gazetesi, Facebook kişiye özel donatılacak. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.turkiyegazetesi.com.tr%2Fbilimteknoloji%2F115251.aspx+%&date=2014-08-18>, Son Erişim Tarihi: 18.08.2014.
90. İnternet: AdAge blogs, Let Facebook Test New Ads Before Writing Them off. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fadage.com%2Farticle%2Fed>

[itorials%2Ffacebook-test-ads-writing%2F149688%2F+&date=2014-08-18,](#) Son
Erişim Tarihi: 18.08.2014.

91. Tang, C., Ross, K., Saxena, N. and Chen, R. (2011, April). *What's in a name: a study of Names, gender inference, and gender behavior in Facebook*. International conference on Database systems for advanced applications, Hong Kong, 344-356.
92. Tang, W., Zhuang, H., Tang, J. (2011, September). *Learning to infer social ties in large networks*. European conference on Machine learning and knowledge discovery in databases, Athens, Greece, 6913, 381-397.
93. Hailpern, J., Reid, L. G., Boardman, R., and Annam, S. (2009, April). *Web 2.0: blind to an accessible new world*. Proceedings of the 18th international conference on World wide web, Madrid, Spain, 821-830.
94. Akcay, A. (2012). The views of Turkish language teachers about blogs. *Procedia - Social and Behavioral Sciences*, 46, 1654-1657.
95. Marshall, C. C. and Shipman, F. M. (2011, May). *Social media ownership: using Twitter as a window onto current attitudes and beliefs*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Canada, 1081-1090.
96. Song, Y., Zhang, C. and Wu, M. (2010, October). *The study of human behavior dynamics based on blogosphere*. International Conference on Web Information Systems and Mining, Sanya, China, 87-91.
97. Fu, F., Liu, L. and Wong, L. (2008). Empirical analysis of online social networks in the age of Web 2.0. *Physica A: Statistical Mechanics and its Applications*, 387(2), 675-684.
98. Mark, G., Bagdouri, M., Palen, L., Martin, J., Al-Ani, B., and Anderson, K. (2012, February.), *Blogs as a collective war diary*. Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, Seattle, Washington, 37-46.
99. Gaonkar, S. and Choudhury, R. R. (2007, November). *Micro-blog: map-casting from mobile phones to virtual sensor maps*. Proceedings of the 5th international conference on Embedded networked sensor systems, Sydney, Australia, 401-402.
100. Lau, C. H. (2014). *Detecting news topics from microblogs using sequential pattern mining*, Doctoral Thesis, Queensland University of Technology, Science and Engineering Faculty, Queensland.
101. Dilrukshi, I., Zoysa, K. D. and Caldera, A. (2013, April). Twitter news classification using SVM. International Conference on Computer Science & Education, Colombo, Sri Lanka, 287-291.
102. Oulasvirta, A., Lehtonen, E., Kurvinen, E., and Raento, M. (2010). Making the ordinary visible in microblogs. *Personal and Ubiquitous Computing*, 14(3), 237-249.

103. Tse, T. S. M. and Zhang, E. Y. (2013). Analysis of blogs and microblogs: a case study of chinese bloggers sharing their Hong Kong travel experiences. *Asia Pacific Journal of Tourism Research*, 18(4), 314-329.
104. Cingiz, M. Ö. and Diri, B. (2012, August). *Content mining of microblogs*. International Conference on Advances in Social Networks Analysis and Mining, İstanbul, 835-838.
105. İnternet: Alexa-Actionable Analytics for the Web. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.alexa.com%2F&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
106. Horn, C. (2010). Analysis and classification of Twitter messages, Master's Thesis, Graz University, Graz.
107. Ha, I., Park, H., and Kim, C. (2014, February). *Analysis of Twitter research trends based on SLR*. 16th International Conference on Advanced Communication Technology, Pyeongchang, South Korea, 774-778.
108. Castillo, C., Mendoza, M. and Poblete, B. (2011, March). *Information credibility on Twitter*. Proceedings of the 20th international conference on World wide web, Hyderabad, India, 675-684.
109. Humphreys, L., Gill, P., Krishnamurthy, B., and Newbury, E. (2013). Historicizing new media: A Content Analysis of Twitter. *Journal of Communication*, 63(3), 413-431.
110. Torrente, E., Marti, T., and Escarrabill, J. (2012). A breath of Twitter. *Revista Portuguesa de Pneumologia (English Edition)*, 18(3), 137-141.
111. Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. (2008, June). *A few chirps about Twitter*. International Workshop on Software and Performance, NY, USA 19-24.
112. İnternet: Tureng. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Ftureng.com%2F&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
113. Morris, M. R., Counts, S., Roseway, A., Hoff, A. and Schwarz, J. (2012, February). *Tweeting is believing? understanding microblog credibility perceptions*. 2012 Conference on Computer Supported Cooperative Work, Seattle, Washington, 441-450.
114. Watanabe, M. and Suzumura, T. (2013, May). *How social network is evolving? - a preliminary study on billion-scale Twitter network*. 22nd international conference on World Wide Web companion, Rio de Janeiro, Brazil, 531-534.

115. Kwak, H., Lee, C., Park, H., and Moon, S. (2010, April). *What is Twitter, a social network or a news media?*. 19th international conference on World Wide Web, NC, USA, 591–600.
116. Jackoway, A., Samet, H., and Sankaranarayanan, J. (2011, November). *Identification of live news events using Twitter*. International Workshop on Location-Based Social Networks, Chicago, USA, 25-32.
117. Zubiaga, A., Ji, H. and Knight, K. (2013, March). *Curating and contextualizing Twitter stories to assist with social newsgathering*. International conference on Intelligent user interfaces, LA, USA, 213-224.
118. Naaman, M., Becker, H., and Gravano, L. (2011). Hip and trendy: characterizing emerging trends on Twitter. *Journal of the American Society for Information Science and Technology*, 62(5), 902-918.
119. Kwak, H., Lee, C., Park, H., and Moon, S. (2010, April). *What is Twitter, a social network or a news media?*. 19th international conference on World wide web, NC, USA, 591-600.
120. Hu, M., Liu, S., Wei, F., Wu, Y., Stasko, J., and Ma, K. (2012, May). *Breaking news on Twitter*. SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 2751-2754.
121. İnternet: Mansur Yavaş Resmi Twitter Hesabı. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Ftwitter.com%2Fmansuryavas06&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
122. İnternet: İbrahim Melih Gökçek Resmi Twitter Hesabı. URL: <http://www.webcitation.org/query?url=http%3A%2F%2F+https%3A%2F%2Ftwitter.com%2F06melihgokcek&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
123. İnternet: Spornx. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.sporx.com%2Ffutbol%2Ftransfer%2Fhaber%2F385845.php%2Chttps%3A%2F%2Ftwitter.com%2FEscudeOfficiel&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
124. Castillo, C., El-Haddad, M. and Pfeffer, J. (2014, February). *Characterizing the life cycle of online news stories using social media reactions*. 17th ACM conference on Computer supported cooperative work & social computing, BC, Canada, 221-223.
125. İnternet: Poulsen, K. Firsthand reports from california wildfires pour through twitter. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.wired.com%2F2007%2F10%2Ffirsthand-repor%2F&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.

126. Bruns, A. and Highfield, T. (2012). Blogs, Twitter, and breaking news: the produsage of citizen journalism. *Producing Theory in a Digital World: The Intersection of Audiences and Production in Contemporary Theory*, 80, 15–32.
127. Stassen, W. (2010). Your news in 140 characters: exploring the role of social media in journalism. *Global Media Journal African Edition*, 4(1).
128. İnternet: 10 Moments on Twitter that will go down in history. URL: http://www.webcitation.org/query?url=http%3A%2F%2Fmashable.com%2F2013%2F09%2F23%2Ftwitter-history-moments%2F%3Futm_cid%3Dmash-com-fb-mainlink%23gallery%2Ftwitter-tweets%2F525c003112d2cd222d00222d&date=2014-08-19, Son Erişim Tarihi: 19.08.2014.
129. Bhattacharya, D., Ram, S. (2012, August). *Sharing news articles using 140 characters: A diffusion analysis on Twitter*. International Conference on Advances in Social Networks Analysis and Mining, İstanbul, 966-971.
130. Ienco, D., Bonchi, F., and Castillo, C. (2010, December). *The meme ranking problem: maximizing microblogging virality*. International Conference on Data Mining Workshops, Sydney, NSW, 328–335.
131. İnternet: CNN Türk URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.cnnturk.com%2F&date=2014-08-19nk%23gallery%2Ftwitter-tweets%2F525c003112d2cd222d00222d&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
132. İnternet: ntvmsnbc. URL: <http://www.webcitation.org/query?url=http%3A%2F%2F+http%3A%2F%2Fwww.ntvmsnbc.com%2F&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
133. İnternet: BBC Türkçe. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.bbc.co.uk%2Fturkce%2F&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
134. Hayes, A. S., Singer, J. B., and Ceppos, J. (2007). Shifting roles, enduring values: the credible journalist in a digital age. *Journal of Mass Media Ethics*, 22(4), 262–279.
135. Deuze, M. and Marjoribanks, T. (2009). Newswork. *Journalism*, 10(5), 555–561.
136. Wattanavaekin, U. and Amornwat, W. (2013). Classification of Twitter news. Thammasat University Sirindhorn International Institute of Technology, Thailand, 1-24.
137. Tseng, C., Patel, N., Paranjape, H., Lin, T. Y. and Teoh, S. (2012, August). *Classifying Twitter data with Naive Bayes classifier*. International Conference on Granular Computing, Hangzhou, China, 294-299.

138. Medvet, E. and Bartoli, A. (2012, December). *Brand-related events detection, classification and summarization on Twitter*. International Conferences on Web Intelligence and Intelligent Agent Technology, Macau, 297-302.
139. Mostafa, M. M. (2013). More than words: social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10), 4241-4251.
140. İnternet: Wefollow: Discover prominent people. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwefollow.com%2F+&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
141. Kamath, K. Y. and Caverlee, J. (2011, October). *Expert-driven topical classification of short message streams*. International Conference on Privacy, Security, Risk, and Trust, and International Conference on Social Computing, Boston, MA, USA, 388-393.
142. Santos, I., Sordo, J., Lopez, I. P., Garcia, P. G., and Pringas, P. G. (2012). Automatic categorisation of comments in social news websites. *Expert Systems with Applications*, 39, 13417-13425.
143. Keretna, S., Hossny, A. and Creighton, D. (2013, October). *Recognising user identity in Twitter social networks via text mining*. International Conference on Systems, Man, and Cybernetics, Manchester, United Kingdom, 3079-3082.
144. Huang, C., Fu, T. and Chen, H. (2009). Text-based video content classification for online video-sharing sites. *Journal of the American Society for Information Science and Technology*, 61(5), 891-906.
145. Sriram, B., Fuhry, D. and Demirbas, M. (2010, July). *Short text classification in Twitter to improve information filtering*. International ACM SIGIR conference on Research and development in information retrieval, Geneva, Switzerland, 841-842.
146. Bhattacharya, P., Mondal, M., Ghosh, S., Zafar, M. B., Gummadi, K. P., Kulshrestha, J. and Ganguly, N. (2014, February). *Deep Twitter diving: exploring topical groups in microblogs at scale*. ACM conference on Computer supported cooperative work & social computing, BC, Canada, 197-210.
147. İnternet: Twitter Developers. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fdev.twitter.com%2Fstart&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
148. Kumar, S., Morstatter, F. and Liu, H. (2013). *Twitter Data Analytics* (2014th Edition), USA: Springer, 23-33.
149. İnternet: Language Detection Library for Java. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fcode.google.com%2Fp%2Flanguage-detection%2F&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.

150. Uysal, A. K. and Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing and Management*, 50, 104-112.
151. Zubiaga, A., Fresno, V., Spina, D. and Martinez, R. (2011, October). *Classifying trending topics: a typology of conversation triggers on Twitter*. ACM international conference on Information and knowledge management, Glasgow, United Kingdom, 2461-2464.
152. Wang, X. and Wang, J. (2013). A method of hot topic detection in blogs using N-gram model. *Journal of Software*, 8(1), 184-191.
153. Zakzouk, T. S. and Mathkour, H. I. (2012). Comparing text classifiers for sports news. *Procedia Technology*, 1, 474-480.
154. Friedl, J. E. F. (2006). *Mastering Regular Expressions*. (3rd Edition). USA: O'Reilly Media, 35-77.
155. Fiaidhi, J., Mohammed, S., Islam, A., Fong, S. and Kim, T. (2013). Developing a hierarchical multi-label classifier for Twitter trending topics. *International Journal of u- and e- Service, Science and Technology*, 6(3), 1-13.
156. İnternet: Zemberek Library. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fcode.google.com%2Fp%2Fzemberek%2F&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
157. Varol, M. (2011). *Metin Madenciliği Yöntemlerini Kullanarak Türkçe Dökümanlarda Tür Ve Yazar Tanıma*, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.
158. Cengiz, C. and Diri, B. (2013, April). *Türkçe tweetlerden soru ifadelerini bulmak*. Signal Processing and Communications Applications Conference, Cyprus, 1-3.
159. Tüfekçi, P., Uzun, E. and Sevinç, B. (2012, April). *Text classification of web based news articles by using Turkish grammatical features*. Signal Processing and Communications Applications Conference, Mugla, 1-4.
160. İlhan, S., Duru, N., Karagöz, Ş. ve Sağır, M. (2014, Kasım). *Metin madenciliği ile soru cevaplama sistemi*, ELECO Elektronik - Bilgisayar Mühendisliği Sempozyumu, Bursa.
161. Asyarie, A. D. and Pribadi, A. W. (2009, December). *Automatic news articles classification in Indonesian language by using Naive Bayes classifier method*. International Conference on Information Integration and Web-based Applications & Services, Kuala Lumpur Malaysia, 658-662.

162. Durmaz, O. and Bilge, H. Ş. (2011, April). *Metin sınıflandırmada boyut azaltmanın etkileri ve özellik seçimi*. Signal Processing and Communications Applications Conference, Kemer, Antalya, 21-24.
163. İnternet: Zemberek Demo. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fzemberek-web.appspot.com%2F+&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
164. Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics-Doklady*, 10(8), 707-710.
165. Rane, S. and Sun, W. (2009, December). *Privacy preserving string comparisons based on Levenshtein distance*. International Workshop on Information Forensics and Security, London, United Kingdom 1-6.
166. Klieber, P. (2009). Document Clasification Through Data Mining Social Media Networks. Lisans Bitirme Tezi, Stetson University, Florida.
167. Kazkılınç, S. and Adalı, E. (2013, October). *Labeling Turkish news stories with CRF*. International Conference on Application of Information and Communication Technologies, Baku, Azerbaijan, 1-5.
168. Karadağ, A. ve Takçı, H. (2010, Şubat). *Metin madenciliği ile benzer haber tespiti*. Akademik Bilişim 2010, Muğla.
169. Güven, E. N., Onur, H. ve Sağiroğlu, Ş. (2008). Yapay sinir ağları ile web içeriklerini sınıflandırma. *Bilgi Dünyası*, 9(1), 158-178.
170. Terzi, Ö., Küçüksille, E. U., Ergin, G. ve İlker, A. (2011). Veri madenciliği süreci kullanılarak güneş ışıınımı tahmini. *SDU International Technologic Science*, 3(2), 29-37.
171. Kohavi, R. (1995, August). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. International Joint Conference on Artificial Intelligence, Canada, 1137-1143.
172. Haider, S. A. and Mehrotra, R. (2011, June). *Corporate news classification and valence prediction: A Supervised Approach*. Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Portland, USA, 175-181.
173. Ma, Z., Pant, G. and Sheng, O. R. L. (2011). Mining competitor relationships from online news: a network-based approach. *Electronic Commerce Research and Applications*, 10(4), 418-427.
174. Haruechaiyasak, C., Jitkrittum, W., Sangkeettrakarn, C., and Damrongrat, C. (2008, December). *Implementing news article category browsing based on text categorization technique*. International Conference on Web Intelligence and Intelligent Agent Technology, Sydney, NSW, 143-146.

175. Wu, L., Li, Z., Li, M., Ma, W., and Yu, N. (2007). *Mutually beneficial learning with application to on-line news classification*. First Ph.D. workshop in CIKM, Shanghai, 85-92.
176. Zhang, Y., Dang, Y., Chen, H., Thurmond, M. and Larson, C. (2009). Automatic online news monitoring and classification for syndromic surveillance. *Decision Support Systems*, 47(4), 508-517.
177. Toraman, C., Can, F. and Koçberber, S. (2011, June). *Developing a text categorization template for Turkish news portals*. International Symposium on Innovations in Intelligent Systems and Applications, İstanbul, 379-383.
178. Amasyalı, M. F. and Yıldırım, T. (2004, April). *Automatic text categorization of news articles*. Signal Processing and Communications Applications Conference, 224-226.
179. İnternet: Twitter Developers The Streaming APIs. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fdev.twitter.com%2Fdocs%2Fapi%2Fstreaming&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.
180. İnternet: Twitter Haberleri. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Ftweetajans.com%2F&date=2014-08-19>, Son Erişim Tarihi: 19.08.2014.

EKLER

EK-1. Sosyal ağ verisi kullanılarak gerçekleştirilen akademik ve sektörel çalışmalar

Çizelge 1.1. Sosyal ağ verisi kullanılarak gerçekleştirilen akademik ve sektörel çalışmaların sınıflandırılması

Sosyal Ağ					Veriler			
Alan	Çalışma Türü	Kaynak	Türü	Adı	Türü	Büyükülüğü	Toplama Aracı	Analiz Aracı
1	Sektörel proje/ Sosyal Medya Analiz Aracı	[2-5], [8-9]	Genel, İş ve Kariyer, Multimedya Paylaşımı	Twitter, Facebook, Google + LinkedIn, Youtube	Metin ve Multimedya	-	Twitter API, Facebook API, Google + API, LinkedIn API, Youtube API	Geliştirilen Sistem
2	Akademik / Konferans Bildirisi	[12], [13], [19], [20]	Genel	Twitter, Facebook	Metin	10,000 tweet, 2,522 Facebook Paylaşımı	Twitter API, Facebook API	Geliştirilen Sistem
3	Sektörel Proje	[24], [27]	Genel	Twitter	Metin	400 Twitter Gündem Konusu (Trending Topic)	Twitter API	Geliştirilen Sistem
4	Akademik / Konferans Bildirisi	[29-33]	Konum Tabanlı, Genel	Foursquare, Gowalla, Twitter, Echofon, Gravity	Konum	220,000 – 1,89 Milyon Kullanıcı Hesabı, 12 - 22 Milyon Konum Paylaşımı	Arama Botu (Crawler), Foursquare API, Twitter API, Anket	Geliştirilen Sistem
5	Akademik / Konferans Bildirisi	[35-38]	Konum Tabanlı	Foursquare	Konum	10 Milyon Kullanıcı Hesabı	Foursquare API	Geliştirilen Sistem
6	Sektörel Proje	[40-43]	Suç Bilgisi Paylaşımı, Genel, Multimedya Paylaşımı	WikiCrimes, Facebook Youtube	Metin ve Multimedya	-	Geliştirilen SA, Manuel Veri Toplama	Geliştirilen SA
7	Sektörel Proje	[50-52]	Eğitim, Genel	Eğitim Bilişim Ağı, Dijital Yerliler, Smile, Facebook	Metin ve Multimedya	-	Geliştirilen SA, Facebook Polls, SurveyMonkey	Geliştirilen SA

SA: Sosyal Ağ

EK-1.(devam) Sosyal ağ verisi kullanılarak gerçekleştirilen akademik ve sektörel çalışmalar

Çizelge 1.1. (devam) Sosyal ağ verisi kullanılarak gerçekleştirilen akademik ve sektörel çalışmaların sınıflandırılması (devam)

Sosyal Ağ					Veriler			
Alan	Çalışma Türü	Kaynak	Türü	Adı	Türü	Büyüküğü	Toplama Aracı	Analiz Aracı
8	Sektörel Proje	[56-58]	Genel	Facebook Twitter	Metin	-	Geliştirilen SA, Facebook API, Twitter API	Geliştirilen SA, Geliştirilen Sistem
9	Sektörel Proje	[60]-62]	Genel	Facebook Twitter	Metin ve Multimedya	-	Twitter API, Facebook API	Geliştirilen Sistem
10	Akademik / Konferans Bildirisi	[63], [64]	Genel	Facebook	Metin	1.67 Milyon Kullanıcı Hesabı	Arama Botu (Crawler)	Geliştirilen Sistem

SA: Sosyal Ağ

EK-2. Paylaşımları veri olarak kullanılan haber kuruluşlarının resmi Twitter hesaplarına ait bilgiler

Çizelge 2.1. Teknoloji kategorisinde haber paylaşımı yapan haber kuruluşlarına ait resmi Twitter hesapları

<i>Hesap Adı</i>	<i>Kullanıcı Adı</i>	<i>Twitter ID</i>	<i>Tweet Sayısı</i>
NTV Tekno Bülten	@ntvteknobulden	1279906620	425
Hürriyet Teknoloji	@HurriyetTekno	78828166	7,876
Türkiye Teknoloji	@Turkiye_Teknolo	1613134129	2,684
Zaman Teknoloji	@ZamanTeknoloji	935268350	281

Çizelge 2.2. Ekonomi kategorisinde haber paylaşımı yapan haber kuruluşlarına ait resmi Twitter hesapları

<i>Hesap Adı</i>	<i>Kullanıcı Adı</i>	<i>Twitter ID</i>	<i>Tweet Sayısı</i>
Habertürk Ekonomi	@HTGazeteEkonomi	450080875	402
Zaman Ekonomi	@ZamanEkonomi	322554469	4.728
CNN TÜRK Ekonomi	@CNNTURKEko	255504289	8.554
TürkiyeEkonomi	@Turkiye_Ekonomi	1613022960	2.016
Hürriyet Ekonomi	@HurriyetEkonomi	91967764	20.568

Çizelge 2.3. Politika kategorisinde haber paylaşımı yapan haber kuruluşlarına ait resmi Twitter hesapları

<i>Hesap Adı</i>	<i>Kullanıcı Adı</i>	<i>Twitter ID</i>	<i>Tweet Sayısı</i>
Zaman Politika	@ZamanPolitika	322575899	4.986
NTV Politika	@ntvpolitika	802284644	3.850
HBR Politika	@HBR_Politika	1450591112	1.919
Politikahaberleri	@politikahaber	304089903	1.244
Politik Gündem	@PolitikGundem	596632313	3.190

Çizelge 2.4. Spor kategorisinde haber paylaşımı yapan haber kuruluşlarına ait resmi Twitter hesapları

<i>Hesap Adı</i>	<i>Kullanıcı Adı</i>	<i>Twitter ID</i>	<i>Tweet Sayısı</i>
NTV Spor	@ntvspor	15016249	66.702
TGRT Haber Spor	@tgrthaberspor	270749957	6.973
Habertürk Spor	@HaberturkSpor	377644615	10.736
AA SPOR	@aa_spor	547622332	9.273
Zaman Spor	@ZamanSpor	322592478	4.316
CNN TÜRK Spor	@CNNTURKSpor	255505370	11.061
TürkiyeSpor	@Türkiye_Spor	1613088704	12.015
Samanyolu Spor	@SamanyoluSpor	304308128	9.028
Radikal Spor	@Radikal_Spor	551709379	5.795
Sabah Spor	@sabahspor1	568214350	13.775
Hürriyet Spor	@HurriyetSpor	91969277	37.702
CumhuriyetSpor	@SporCumhuriyet	2438142068	576
Fanatik	@fanatikcomtr	224176099	58.190
Amkspor	@Amkspor	583863506	40.795
SporOku	@Spor_Oku	397082335	10.880
SPORX.COM	@sporex	18426124	60.571
Ajansspor	@Ajanssporcom	31094717	21.537
Mynet Spor	@mynetspor	91126526	44.622
360Spor	@360spor	360578984	16.000
Maraton.com.tr	@maraton_com_tr	108278967	45.226
TRT Spor	@TRT3SPOR	368948189	100.381
LİGTV	@LigTV	155280601	28.011
Medyaspor	@Medyaspor	117730774	24.194

Çizelge 2.5. Magazin kategorisinde haber paylaşımı yapan haber kuruluşlarına ait resmi Twitter hesapları

<i>Hesap Adı</i>	<i>Kullanıcı Adı</i>	<i>Twitter ID</i>	<i>Tweet Sayısı</i>
Magazin Haberleri	@magazinler	304554706	1.030
Hürriyet Magazin	@HurriyetMagazin	91968524	14.032
TürkiyeMagazin	@Türkiye_Magazin	1613182226	2.826
Birbak Magazin	@MynetMagazin	108964268	26.768

Çizelge 2.6. Sağlık kategorisinde haber paylaşımı yapan haber kuruluşlarına ait resmi Twitter hesapları

<i>Hesap Adı</i>	<i>Kullanıcı Adı</i>	<i>Twitter ID</i>	<i>Tweet Sayısı</i>
Hürriyet Sağlık	@HurriyetSaglik	91970390	6.830
NTV Sağlık	@NTV_Saglik	353826788	4.693
Sağlık Bakanlığı	@saglikbakanligi	556130277	3.237
Sağlık Haberleri	@Saglik_Insan	356827903	11
Medimagazin	@Medimagazin	359847298	19.356

Çizelge 2.7. Kültür-Sanat kategorisinde haber paylaşımı yapan haber kuruluşlarına ait resmi Twitter hesapları

<i>Hesap Adı</i>	<i>Kullanıcı Adı</i>	<i>Twitter ID</i>	<i>Tweet Sayısı</i>
Hürriyet KültürSanat	@KulturSanat	78829163	6.269
NTV Kültür Sanat	@NTVKSanat	377964190	1.909
Sabah Kültür-Sanat	@sabahsanat	94073108	1,049
Zaman Kültür & Sanat	@Zaman_Kultur	929710746	541
Kültür Sanat Sitesi	@kultursitesi	2364641574	130
İSTApin Kültür Sanat	@istapinks	1111042956	537
Koç Kültür Sanat	@kockultursanat	359932719	1.278
Akbank Sanat	@AkbankSanat	45819962	1.556

EK-3. Geliştirilen web sitesi yazılımına ait ekran görüntüleri



(a)

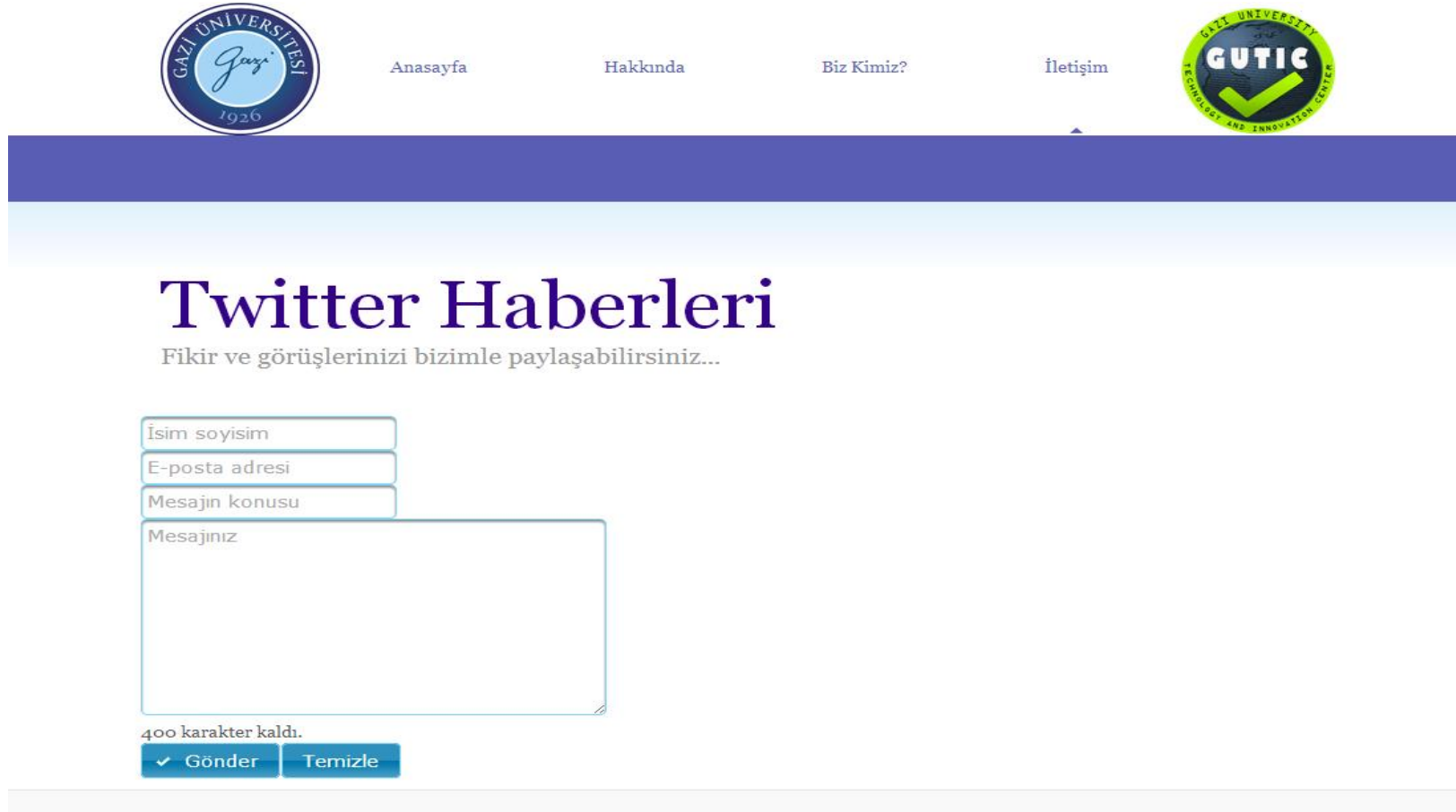
Resim 3.1. (a) tweetajans.com'a ait "Hakkında" sayfası
(b) tweetajans.com'a ait "Biz Kimiz?" sayfası

EK-3.(devam) Geliştirilen web sitesi yazılımına ait ekran görüntüleri



Resim 3.1.(devam) (a) tweetajans.com'a ait "Hakkında" sayfası
(b) tweetajans.com'a ait "Biz Kimiz?" sayfası

EK-3.(devam) Geliştirilen web sitesi yazılımına ait ekran görüntüleri



The screenshot displays the contact page of the tweetajans.com website. At the top, there is a navigation bar with the Gazi University logo on the left and the GUTIC logo on the right. The navigation menu includes links for 'Anasayfa', 'Hakkında', 'Biz Kimiz?', and 'İletişim'. The main heading is 'Twitter Haberleri' in a large, bold, purple font. Below the heading is a subtitle: 'Fikir ve görüşlerinizi bizimle paylaşabilirsiniz...'. The contact form consists of four input fields: 'İsim soyisim', 'E-posta adresi', 'Mesajın konusu', and a larger 'Mesajınız' text area. Below the text area, it indicates '400 karakter kaldı.' (400 characters left). At the bottom of the form are two buttons: '✓ Gönder' (Send) and 'Temizle' (Clear).

Resim 3.2. tweetajans.com'a ait iletişim sayfası

EK-4. 23.06.2014 - 29.06.2014 tarihleri kategori bazında paylaşım yapan Twitter hesaplarına ait bilgiler

Çizelge 4.1. Test sürecinde dinlenen ve kategori bazında paylaşım yapan Twitter sayfaları

<i>Hesap Adı</i>	<i>Kullanıcı Adı</i>	<i>Twitter ID</i>	<i>Dinlenen Tweet Sayısı</i>
Hürriyet Teknoloji	@HurriyetTekno	78828166	76
Türkiye Teknoloji	@Turkiye_Teknolo	1613134129	32
TGRT Haber Spor	@tgrthaberspor	270749957	29
AA SPOR	@aa-spor	547622332	111
Zaman Ekonomi	@ZamanEkonomi	322554469	71
CNN TÜRK Ekonomi	@CNNTURKEko	255504289	31
Zaman Politika	@ZamanPolitika	322575899	82
Hürriyet Magazin	@HurriyetMagazin	91968524	92
TürkiyeMagazin	@Turkiye_Magazin	1613182226	82
Birbak Magazin	@MynetMagazin	108964268	141
Hürriyet Sağlık	@HurriyetSaglik	91970390	8
NTV Sağlık	@NTV_Saglik	353826788	33
Sağlık Bakanlığı	@saglikbakanligi	556130277	2
Medimagazin	@Medinagazin	359847298	160
Hürriyet KültürSanat	@KulturSanat	78829163	1
Sabah Kültür-Sanat	@sabahsanat	94073108	4
Akbank Sanat	@AkbankSanat	45819962	5
Zaman Kültür & Sanat	@Zaman_Kultur	929710746	5

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : Merve Sedef, GÜNDÜZ
 Uyuğu : T.C.
 Doğum tarihi ve yeri : 23.06.1991, Erzincan
 Medeni hali : Bekâr
 Telefon : 0 (507) 4428424
 Faks : -
 E-Posta : sedefgunduz@gazi.edu.tr



Eğitim

Derece	Okul/Program	Mezuniyet tarihi
Yüksek Lisans	Gazi Üniversitesi/Bilgisayar Mühendisliği	2014
Lisans	Fatih Üniversitesi/Bilgisayar Mühendisliği	2012
Lise	Özel Otlukbeli Hacı Ali Akın Anadolu Lisesi	2007

İş Deneyimi

Yıl	Çalıştığı Yer	Görev
2012-Halen	Gazi Üniversitesi	Araştırma Görevlisi

Yabancı Dil

İngilizce

Yayınlar

1. Yavanoğlu, U., Gündüz, S. and Sağiroğlu, Ş. (2013, April). *Sosyal ağ paylaşımlarından adli profil çıkarılması için yeni bir yaklaşım*. First International Symposium on Digital Forensics and Security, Elazığ, 440-445.
2. Gunduz, S., Yavanoglu, U. and Sagiroglu. S. (2013, December). *Predicting next location of Twitter users for surveillance*. International Conference on Machine Learning and Applications, Miami, FL, 267-273.
3. Gunduz, S., Demirhan, F. (2014, September). *Analyzing the effects of personality traits on Facebook use for education: a study on university students*. International Conference on Computer Science Education: Innovation and Technology, Singapore, in-press.