



**T.C.**

**ULUDAĞ ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ  
EKONOMETRİ ANABİLİM DALI  
İSTATİSTİK BİLİM DALI**

**DUYGU ANALİZİ ve SOSYAL MEDYA ALANINDA  
UYGULAMA**

**(DOKTORA TEZİ)**

**Yusuf Murat KIZILKAYA**

**BURSA - 2018**





**T.C.  
ULUDAĞ ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ  
EKONOMETRİ ANABİLİM DALI  
İSTATİSTİK BİLİM DALI**

**DUYGU ANALİZİ ve SOSYAL MEDYA ALANINDA  
UYGULAMA**

**(DOKTORA TEZİ)**

**Yusuf Murat KIZILKAYA**

**Danışman Prof. Dr. Ayşe OĞUZLAR**

**BURSA - 2018**

**T.C.**  
**ULUDAĞ ÜNİVERSİTESİ**  
**SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE**

Ekonometri Anabilim Dalı, İstatistik Bilim Dalı'nda 711417005 numaralı Yusuf Murat KIZILKAYA'nın hazırladığı “**DUYGU ANALİZİ VE SOSYAL MEDYA ALANINDA UYGULAMA**” konulu Doktora Tezi ile ilgili tez savunma sınavı, .../.../2018 günü ..... saatleri arasında yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin/çalışmasının ..... (başarılı/başarısız) olduğuna ..... (oybirliği/oy çokluğu) ile karar verilmiştir.

	Öğretim Üyesinin Adı Soyadı	Üniversitesi	İmzası
1	Tez Danışmanı	Prof. Dr. Ayşe OĞUZLAR – Uludağ Ün.	
2	Jüri Üyesi	Prof. Dr. Şahamet BÜLBÜL – Marmara Ün.	
3	Jüri Üyesi	Prof. Dr. Dilek ALTAŞ – Marmara Ün.	
4	Jüri Üyesi	Prof. Dr. Aysan ŞENTÜRK – Uludağ Ün.	
5	Jüri Üyesi	Doç. Dr. Sevda GÜRSAKAL – Uludağ Ün.	

## YEMİN METNİ

Doktora tezi olarak sunduđum “**DUYGU ANALİZİ VE SOSYAL MEDYA ALANINDA UYGULAMA**” başlıklı alıřmamın bilimsel arařtırma, yazma ve etik kurallarına uygun olarak tarafımdan yazıldıđına ve tezde yapılan bütn alıntuların kaynaklarının usulne uygun olarak gsterildiđine, tezimde intihal rn cmle veya paragraflar bulunmadıđına řerefim zerine yemin ederim.

01/07/2018

Adı Soyadı : Yusuf Murat KIZILKAYA

đrenci No : 711417005

Anabilim Dalı: Ekonometri

Programı : İstatistik

Stats : Doktora

## ÖZET

Yazar : Yusuf Murat KIZILKAYA  
Üniversite : Uludağ Üniversitesi  
Enstitü : Sosyal Bilimler Enstitüsü  
Anabilim Dalı : Ekonometri  
Bilim Dalı : İstatistik  
Tezin Niteliği : Doktora Tezi  
Sayfa Sayısı : XVI + 137  
Mezuniyet Tarihi : 18/07/2018  
Tez Danışmanı : Prof. Dr. Ayşe OĞUZLAR

### DUYGU ANALİZİ VE SOSYAL MEDYA ALANINDA UYGULAMA

Bilgisayar teknolojilerinde meydana gelen değişiklik hayatın her aşamasında yenilikleri de beraberinde getirmiştir. Bunlardan en önemlileri farklı tiplerde ki verilerin derlenip saklanabilmesine, işlenmesine imkan tanınması ve geleneksel medya araçlarına alternatif olarak sosyal medyanın kullanılmaya başlanmasıdır. Geçmişte bir verinin işlenebilmesi için o verinin yapılandırılmış olma ön koşulu varken, günümüzde yapılandırılmamış verilerde son derece kolay bir şekilde işlenebilmektedir. Metin madenciliği tam da burada karşımıza çıkmaktadır. Yapılandırılmamış halde bulunan metinler artık rahatlıkla işlenebilmektedir. Duygu analizi son yıllarda hızla popüleritesi artan bir araştırma alanı olmuştur. Duygu analizinde temel amaç, metinde yer alan duygunun ortaya çıkarılabilmesidir. Sosyal medya özellikle Web 2.0 teknolojisini getirmiş olduğu ve insanların yaşamlarını tamamen değiştiren bir kavramdır. Sosyal medya sayesinde, bireyler artık kendilerine sunulan içerikleri tüketen pasif kullanıcılar değil, kendileri de içerik üretip bunu paylaşabilen aktif bireyler haline gelmiştir. Günümüzde akıllı telefona sahip olan herkes aynı zamanda birer içerik üreticisidir. R yazılımı dünya da son yıllarda en sık kullanılan ve pek çok işlemi kolaylıkla yerine getirebilen önemli bir açık kaynak kodlu programdır. Ancak ülkemizde henüz değeri çok iyi algılanamamış ve kullanım düzeyi istenilen seviyeye ulaşamamıştır. Bu tezde R yazılımı ile duygu analizi yapılmıştır. Duygu analizi için Twitter'dan, 24 Haziran 2018

de Türkiye’de yapılan Cumhurbaşkanlığı ve 27. Dönem Millet Vekilli Genel Seçimi’ne ilişkin en yüksek oy alması beklenen aday ve partilere ilişkin bir ay boyunca atılan tweetlerden ele alınmıştır. Tweetlerden hareketle duygu analizi yapılmış, bu amaçla duygu skoru hesaplanmış ve makine öğrenmesi ile hesaplamalar yapılmıştır. Elde edilen veriler ile seçim sonuçları karşılaştırılmıştır.

**Anahtar Sözcükler:** Duygu Analizi, Sosyal Medya, Twitter, Makine Öğrenmesi



## **ABSTRACT**

Name and Surname : Yusuf Murat KIZILKAYA  
University : Uludağ University  
Institution : Social Science Institution  
Field : Econometrics  
Branch : Statistics  
Degree awarded : PhD  
Page Number : XVI+137  
Degree Date : 18/07/2018  
Supervisor : Prof. Dr. Ayşe OĞUZLAR

### **SENTIMENT ANALYSIS AND SOCIAL MEDIA APPLICATION**

The change that took place in computer technology brought innovations with every step of life. The most important of these is the compilation of data in different types, the possibility of processing them, and the introduction of social media as an alternative to traditional media tools. In the past, when a given data is processed, it can be processed very easily in the unstructured data of today, while the data is pre-configured. Text mining is exactly here. Unstructured texts can now be easily processed. Emotional analysis has been a rapidly growing field of research in recent years. The main purpose of emotion analysis is to be able to reveal the sensation in the text. Social media is a concept that especially brings Web 2.0 technology and changes people's lives completely. With social media, individuals are no longer passive users who consume content that is presented to them, but become active individuals who can produce and share content themselves. Today, everyone who owns a smartphone is also a content producer. R software is one of the most frequently used open source programs in the world in recent years and can easily handle many operations. However, in our country, the value has not been perceived very well and the level of use has not reached the desired level. In this thesis, emotion analysis was done with R software. Essential texts



for emotion analysis were obtained from Twitter, an important social sharing site. For sentiment analysis on Twitter, June 24, 2018 also receive the highest number of votes related to the presidential and 27. Period Vekille Nations General elections in Turkey were taken from discarded tweets related to the expected candidates and parties. Sentiment analysis was performed by tweets, the emotion score was calculated for this purpose and calculations were made by machine learning. We compared the selection results with the obtained data.

**Keywords:** Sentiment Analysis, Social Media, Twitter, Machine Learning



## ÖNSÖZ

Bu tezin ortaya çıkması sürecinde manevi desteğini hiçbir zaman esirgemeyen, her zaman bir abla şefkatiyle yaklaşan, bana kattıklarını ömür boyu unutamayacağım değerli danışmanım Prof. Dr. Ayşe OĞUZLAR'a teşekkürümü sunarım. Gerek Lisans gerekse doktora eğitimim boyunca bilgilerini paylaşmakta cömert davranan ve yetişmemde çok büyük emekleri olan, Uludağ Üniversitesi İ.İ.B.F. Ekonometri Bölümü Öğretim Üyelerine teşekkürü bir borç bilirim. Doktora eğitimim boyunca her türlü kahrımı çeken Uludağ Üniversitesi İ.İ.B.F. Ekonometri Bölümü araştırma görevlilerini burada anmazsam onlara haksızlık etmiş olurum, onların da emeklerinin karşılığı ödeyemem. Tez konumun belirlenmesinde yardımcı olan ve her aşamada fikirleriyle yolumu aydınlatan Uludağ Üniversitesi Emekli Öğretim Üyesi Prof. Dr. Necmi GÜRSAKAL'a, her sohbetimizde ufkumu açan ve tezimin gelişmesinde katkılarını benden esirgemeyen değerli arkadaşlarım, Dr. Arif Cem TOPUZ'a ve Dr. Emrah AKDAMAR'a teşekkür ederim.

Eğitim hayatımın her anında yanımda olan, maddi ve manevi desteğini hiçbir zaman esirgemeyen çok kıymetli, anneme, babama, ablama, ağabeyime çok teşekkür ederim, onların haklarını hiçbir zaman ödeyemem. Sevgili eşim, hayat arkadaşım, en umutsuz günlerimde bile bana enerji veren değerli yoldaşım Pınar'a çok teşekkür ederim. Sevgili Ailem olmasaydı bu günlere gelemezdim.

Bu tezin uygulama kısmı Uludağ Üniversitesi BAP Birimi tarafından desteklenen DDP(İ) - 2017/8 numaralı proje kapsamında yapılmıştır. Desteklerinden dolayı Uludağ Üniversitesi BAP Birimine teşekkür ederim.

## İÇİNDEKİLER

TEZ ONAY SAYFASI .....	iv
YEMİN METNİ.....	v
ÖZET.....	vii
ABSTRACT.....	viii
ÖNSÖZ .....	x
İÇİNDEKİLER .....	vii
TABLolar .....	xiv
KISALTMALAR .....	xv
GİRİŞ .....	1
BİRİNCİ BÖLÜM .....	4
METİN MADENCİLİĞİ .....	4
1. VERİ OLARAK METİN .....	4
1.1. Metin Madenciliği Nedir? .....	4
1.2. Metin Madenciliğinin Aşamaları.....	7
1.2.1. Veri Setinin Oluşturulması.....	8
1.2.2. Ön İşleme .....	8
1.2.3. Özellik Seçme .....	11
1.2.4. Sınıflandırma.....	11
1.2.5. Sonuçların Değerlendirilmesi.....	22
1.3. Metin Madenciliğinin Kullanım Alanları .....	22
1.4. Metin Madenciliği Alanında Geliştirilmiş Olan Paket Programlar .....	25
1.5. Doğal Dil İşleme (DDİ).....	27
1.6. Makine Öğrenmesi (MÖ) .....	30
1.6.1. Denetimli Öğrenme.....	31
1.6.2. Denetimsiz Öğrenme .....	32

1.6.3. Yarı Denetimli Öğrenme .....	33
1.7. R ile Metin Madenciliği .....	33
İKİNCİ BÖLÜM.....	35
DUYGU ANALİZİ.....	35
2. DUYGU ANALİZİ.....	35
2.1. Duygunun Önemi .....	36
2.1.1. Kansei Mühendisliği (KM) .....	36
2.1.2. Duygusal Zekâ .....	37
2.2. Literatür Özeti .....	38
2.3. Duygu Analizi Yöntemleri .....	42
2.3.1. Sözlüğe Dayalı Yöntemler .....	44
2.3.2. Makine Öğrenmesine Dayalı Yöntemler .....	46
ÜÇÜNCÜ BÖLÜM .....	56
SOSYAL MEDYA .....	56
3. SOSYAL MEDYA .....	56
3.1. Sosyal Medyanın Tanımı.....	57
3.2. Sosyal Medya Araçları .....	58
3.2.1. Bloglar.....	59
3.2.2. Wikiler.....	60
3.2.3. Sosyal Ağlar .....	60
3.2.4. İçerik Paylaşım Toplulukları.....	61
3.2.5. Vloglar.....	62
3.2.6. Podcasting .....	62
3.2.7. Sosyal İmlleme Siteleri (Socila Bookmarking Systems) .....	63
3.2.8. Mikrobloglar .....	63
3.3. Önemli Bir Sosyal Medya Aracı Olan Twitter.....	64

3.3.1. Takip (Follow) / Takipçi (Follower).....	65
3.3.2. Etiket (Hashtag) .....	66
3.3.3. Trend Konu (Trend Topic).....	66
3.3.4. Mention (Bahsetmek).....	67
3.3.5. Yeniden Teyweet (Retweet).....	67
3.3.6. Direkt Mesaj (Direct Message) .....	67
3.3.7. Troll Hesap.....	67
3.4. Sosyal Medyanın Önemi .....	68
DÖRDÜNCÜ BÖLÜM.....	74
UYGULAMA: R İLE DUYGU ANALİZİ.....	74
4.1. Seçim Sonuçları İçin Geleneksel Tahmin Yöntemi .....	75
4.2. Verilerin Twitter'dan Çekilmesi .....	78
4.3. Verilerin Ön İşlemesi .....	80
4.4. Verilerin Analizi.....	90
4.4.1. Duygu Skoru .....	91
4.4.2. Makine Öğrenmesi İle Duygu Analizi.....	102
SONUÇ .....	105
KAYNAKÇA .....	110
EKLER.....	122
Ek 1 Tezde Kullanılan Kodlar .....	122
Ek 2: Gruplarda İlk 6 Teyweetin Duygu Puanı Ortalaması .....	130
ÖZGEÇMİŞ .....	134

## TABLULAR

<b>Tablo 2.1.</b> Kontenjans Tablosu.....	50
<b>Tablo 4.1.</b> Kamuoyu Araştırma Firmalarının Cumhurbaşkanlığı Seçimine İlişkin Son Tahminleri.....	76
<b>Tablo 4.2.</b> Kamuoyu Araştırma Firmalarının 27. Dönem Millet Vekili Seçiminde Siyasi Partilerin Oy Oranlarına İlişkin Son Tahminleri.....	77
<b>Tablo 4.3.</b> Adayların Duygu Skorlarının Aritmetik Ortalaması.....	91
<b>Tablo 4.4.</b> R. Tayyip ERDOĞAN'a İlişkin Tweet Puanları ve Frekansları.....	93
<b>Tablo 4.5.</b> Muharrem İNCE'ye İlişkin Tweet Puanları ve Frekansları.....	94
<b>Tablo 4.6.</b> Meral AKŞENER'e İlişkin Tweet Puanları ve Frekansları.....	95
<b>Tablo 4.7.</b> Selahattin DEMİRTAŞ'a İlişkin Tweet Puanları ve Frekansları.....	96
<b>Tablo 4.8.</b> Siyasal Partilerin Duygu Skorlarının Aritmetik Ortalaması.....	97
<b>Tablo 4.9.</b> Ak Parti'ye İlişkin Tweet Puanları ve Frekansları.....	98
<b>Tablo 4.10.</b> CHP'ye İlişkin Tweet Puanları ve Frekansları.....	99
<b>Tablo 4.11.</b> HDP'ye İlişkin Tweet Puanları ve Frekansları.....	100
<b>Tablo 4.12.</b> MHP'ye İlişkin Tweet Puanları ve Frekansları.....	100
<b>Tablo 4.13.</b> İyi Parti'ye İlişkin Tweet Puanları ve Frekansları.....	101
<b>Tablo 4.14.</b> Cumhurbaşkanı Adaylarının Makine Öğrenmesi Sonuçları.....	102
<b>Tablo 4.15.</b> Partilerin Makine Öğrenmesi Sonuçları .....	103

## KISALTMALAR

**AID** Automatic Interaction Detector (Otomatik Etkileşim Dedektörü)

**API** Application Programming Interface (Uygulama Programlama Arayüzü)

**CART** Classification And Regression Trees (Sınıflandırma ve Regresyon Ağaçları)

**CHAID** Chi-squared Automatic Interaction Detector (Ki-Kare Otomatik Etkileşim Dedektörü)

**DDİ** Doğal Dil İşleme

**DM** Direct Message (Direk Mesaj)

**DVM** Destek Vektör Makinesi

**HTML** Hyper Text Markup Language (Hiper Metin İşaretleme Dili)

**İBB** Türkiye'deki İstatistiksel Bölge Birimleri

**KDT / KDD** Knowledge Discovery in Textual Databases (Metin Veritabanlarında Bilgi Keşfi)

**KEYK** K-En Yakın Komşuluk

**KNN** K-Nearest Neighbor (K-En Yakın Komşuluk)

**MÖ** Makine öğrenmesi

**NB** Navie Bayes

**NLP** Natural Language Process (Doğal Dil İşleme)

**NUTS** Nomenclature d'Unités Territoriales (Bölgesel Birimlerin İsimlendirilmesi)

**QUEST** Quick-Unbiased-Efficient-Statistical-Tree (Hızlı-Yansız-Etkili-İstatistiksel-Ağaç)

**RF / RO** Random Forest (Rastgele Orman)

**RT** Retweet (Yeniden Tweet)

**SVM** Support Vector Machine (Destek Vektör Makinesi)

**TDK** Türk Dil Kurumu

**TDM** Terim Doküman Matrisi

**TT** Trend Topic (Trend Konu)

**YSK** Yüksek Seçim Kurulu

**XML** EXtensible Markup Language (Geniřletilebilir İşaretleme Dili)





## GİRİŞ

Bilgisayar teknolojilerinde meydana gelen deęişikliklerin, pek çok alanda yeni çıęırlar açtıęı gözlenmektedir. Bu süreçte verinin depolanmasının ve veriye erişilmesinin kolaylaşması bilim dünyasında da yeniliklere yol açmış, dünya artık “büyük veri” (big data) kavramı etrafında dönmeye başlamıştır. Beş önemli bileşenden oluşan büyük veri kavramı, bu bileşenlerin İngilizcelerinin kısaltmalarına atıfla 5V üzerine inşa edilmiştir. Çeşitlilik (Variety); verilerin çok farklı formata sahip olduğunu, sadece yapılandırılmış verilerin ele alınmadığını ifade eder. Hız (Velocity); verilerin her geçen gün hatta her geçen saat büyük bir hızla artmaya devam ettiğine işaret eder. Bu kadar büyük bir hızla artan verileri depolayabilecek yazılımsal ve donanımsal alt yapının oluşturulması gerekmektedir. Veri Büyüklüğü (Volume); veri miktarı sürekli arttığı için ileride bu kadar büyük veri yığınıyla nasıl baş edilebileceğinin planlanması gerekmektedir. Doğrulama (Verification); hızla büyüyen bu verilerin güvenli olup olmadığının, doğruluğunun kontrol edilmesi gerekir. Değer (Value); bu kadar büyük veriden işe yarar bilginin üretilmesi gerekir ki veri bizim için değerli olabilsin.

Veri dünyasında meydana gelen bu deęişiklikler; veri madenciliğinde de evrilmelere neden olmuştur. Yapılandırılmış veriyi kullanan veri madencilięi, yapılandırılmamış veri sayısında ki muazzam artışla birlikte, yapılandırılmamış verilerin de analiz edilmesi çalışmalarına maruz kalmış ve bu süreçte metin madencilięi ortaya çıkmıştır. Metin madencilięi son yılların en önemli araştırma alanlarından biri olmuştur.

Metin madencilięinin ilgi alanlarından biri bu tezin de konusu olan duygu analizidir. Duygu analizi herhangi bir metin içerisindeki; bu metin bir kitap gibi çok büyük de olabilir, bir tweet gibi 280 karakterle sınırlandırılmış kısa metin de olabilir, duyguyu ortaya çıkarmayı amaçlamaktadır. Pazarlamacılardan siyasetçilere kadar hedef kitlesi insan olan herkes duygu analizinin önemini farkındadır. Literatürde ise Türkçe metinler üzerine yapılan duygu analizlerine ilişkin çok büyük bir eksiklik vardır. Bu tezin çıkış noktalarından biri literatürdeki eksikliğin giderilmesi için katkıda bulunabilmektir.

Sosyal medya, internetin insanların cebine kadar girmesinin getirmiş olduđu alternatif bir medya aracı olarak karşımıza çıkmaktadır. Sosyal medya ile akıllı telefona sahip olan tüm bireyler birer medya üreticisi olmuş, önceden sadece kendisine sunulanı tüketmekle yetinen bireyler, artık birer aktif içerik üreticisi konumuna geçmiştir. Sosyal medyanın toplum yaşamında meydana getirdiđi muazzam deđişiklik, onu pek çok araştırmanın da merkezi konumuna koymuştur. Bu yüzden literatürde sosyal medya üzerine pek çok çalışma yer almaktadır. Buna karşın sosyal medya üzerine duygu analizi alanında literatürde yine büyük bir açık yer almaktadır. Sosyal medya da duygu analizi üzerine yapılan çalışmaların tamamına yakını, uygulama kolaylığından dolayı İngilizce metinler üzerine yapılmıştır. Türkçe metinlerden hareketle duygu analiziyle karşılaşmamız pek de mümkün olmamaktadır.

Özellikle gençler arasında sosyal medya kullanımı o kadar yaygınlaşmıştır ki günümüzde sosyal medyayı kullanmak deđil kullanmamak bir ayrıcalık gibi algılanmaya başlanmıştır. Bu tezin ortaya çıkış noktalarından birisi de toplumlarda bu kadar büyük öneme sahip olan sosyal medyadan hareketle Türkçe verilere duygu analizi yapmaktır.

Bu çalışma dört ana bölümden oluşmaktadır. Birinci bölümde metin madenciliđi hakkında detaylı bir bilgilendirme yapılmış, metin madenciliđinin aşamaları anlatılmıştır. Metin madenciliđinin önemli süreçlerinden olan sınıflandırma algoritmaları tek tek açıklanmış, metin madenciliđinin kullanım alanları deđerlendirilmiştir. Metin madenciliđin olmazsa olmazları olan doğal dil işleme ve makine öğrenmesi yöntemlerine yer verilmiştir.

İkinci bölümde duygunun önemi açıklanarak duygu analizine yer verilmiştir. Duygu analizine dair yurt dışında ve yurt içinde yapılmış çalışmalar incelenmiş, duygu analizi yöntemleri sırasıyla açıklanmıştır. Duygu analizinde kullanılan makine öğrenmesi tekniklerinin karşılaştırılmasında yoğun olarak kullanılan sınıflandırma algoritmalarının başarılarını ölçmek amacıyla geliştirilmiş olan teknikler açıklanmış, aralarındaki ilişkilere değinilmiştir. Ayrıca model performans deđerlendirme teknikleri ayrıntılı bir şekilde açıklanmıştır.

Üçüncü bölümde sosyal medya kavramı tüm yönleriyle ele alınmıştır. Sosyal medyanın çeşitli tanımları verilmiş, sosyal medya araçları ayrıntılı bir şekilde açıklanmıştır. Bu tezin uygulama alanı, Twitter'dan elde edilen veriler olduğu için, Twitter'a ayrı bir sayfa açılmış, Twitter hakkında detaylı bir inceleme yapılmıştır. Sosyal medyanın önemi tarihteki olaylardan hareketle açıklanmış, Türkiye ve yurtdışından örnekler verilmiştir.

Tezin son bölümü olan dördüncü bölümde uygulama yapılmıştır. Uygulama amacıyla 24 Haziran 2018'de Türkiye'de yapılan Cumhurbaşkanı ve Milletvekili seçimine ilişkin Twitter'dan yaklaşık 1 ay boyunca çekilen toplam 720.000 adet tweet kullanılmıştır. Uygulama kapsamında kizilkayamurat.com isimli bir web sitesi oluşturulmuş bu siteye Türkçe'de eksikliği hissedilen; pozitif duygu ifade eden kelimeler, negatif duygu ifade eden kelimeler ve durak kelimeler oluşturularak yüklenmiştir. Bu sayede ileride bu alanda çalışma yapmak isteyen araştırmacılar için hazır bir veri tabanı kullanıma sunulmuştur. Çekilen tweetler açık kaynak kodlu olan ve son yıllarda kullanım alanı hızla artan R yazılımı ile analiz edilmiştir.

# BİRİNCİ BÖLÜM

## METİN MADENCİLİĞİ

### 1. VERİ OLARAK METİN

Veri, bizi bilgiye götürecektir olan ham bilgi parçacıklarıdır. 2011 yılında, tüm verilerin yaklaşık yüzde doksanı, web ortamında ki verilerinse yaklaşık yüzde doksan dokuzu metin biçimindedir.<sup>1</sup> Bilgisayar teknolojisinin gelişmesi, özellikle sosyal medyanın insanların hayatlarına girip yaygınlaşmasıyla beraber, web üzerindeki veri miktarı her geçen saniye hızla artmaktadır. Veriler yapılandırılmış (structured) veya yapılandırılmamış (unstructured) olmak üzere iki farklı formda olabilirler. Günlük yaşamda karşılaştığımız verilerin çok büyük bir kısmı yapılandırılmamış dolayısıyla herhangi bir kurala uymayan verilerdir. Yapılandırılmamış olan bu verilerin analiz edilmesi pek çok farklı bilim dalı için çok büyük önem göstermektedir. Bu verilerin, kullanılabilir bir hale gelmesi için öncelikle bir dizi işleminden geçmesi gerekir.<sup>2</sup> Bu verilerin analiz edilebilmesi için kullanılan metotlardan birisi de metin madenciliğidir.

Yaşadığımız çağ artık büyük verinin çağı olarak adlandırılabilir. Büyük veri 100 terabayt ve üzerindeki veriyi ifade etmektedir. Büyük veri, bilim insanları için çok farklı bir boyut getirmiştir. Büyük verinin analizi ile aslında birbiri arasında hiç ilişki olmadığı düşünülen şeyler arasında, çok güçlü ilişkiler keşfedilmeye başlanmıştır. Büyük veri ile yapılandırılmamış çok büyük hacimde ve sürekli artmakta olan veriler, makine öğrenmesi sayesinde analiz edilmeye başlanmış, bu şekilde klasik, hipoteze dayalı, az sayıda yapılandırılmış, statik haldeki veriden hareketle yapılan analizlerin bir adım ilerisine geçilmiştir.

#### 1.1. Metin Madenciliği Nedir?

Metin madenciliği ile yapısal olmayan veriler yapısal bir hale getirilerek, analitik analize uygun bir kaynak elde edilir.<sup>3</sup> Yazılı metinlerin tamamı veya yazılı olmayan fotoğraf, harita gibi görseller metin madenciliğinin konusunu oluşturabilir. Bunun yanı sıra; makale, gazete, kitap, akademik yayınlar, internet siteleri gibi tüm topluma açık

<sup>1</sup> Ayşe Oğuzlar, *Temel Metin Madenciliği* Bursa, Dora Basım Yayın, 2011, s. 2.

<sup>2</sup> Filiz Erten, *Metin Madenciliği Tabanlı Bir Web Sitesi Sınıflandırma Aracı Tasarımı*, Maltepe Üniversitesi Fen Bilimler Enstitüsü, İstanbul, 2015, s. 3.

<sup>3</sup>Ramesh Sharda, Dursun Delen ve Efraim Turban, *Business Intelligence and Analytics Systems for Decision Support, 10th Edition*, 10 th (U.S.A: Pearson, 2015), s 289

olan metinler, e posta, hastalar raporları, adli sicil kayıtları, mektuplar, gibi bireye özel metinler de, metin madenciliğinin kaynağı olabilirler. Günümüzde sosyal medyanın yaygınlığı nedeniyle metin madenciliği denilince ilk akla gelen şey sosyal medyadaki paylaşımlar olmaktadır. Gün içinde pek çok kişi sosyal medyadan o ana dair, güncel konular hakkında, izlediği bir film veya diziye ilişkin, arkadaşlarının yapmış oldukları paylaşımlara ilişkin, pek çok yorum ve paylaşımlarda bulunmaktadır. Bu paylaşımların tamamı metin madenciliği için muazzam bir kaynak oluşturmaktadır.

Metin madenciliği; çok fazla miktardaki metinlerden, yüksek kaliteli bilgileri ayıklamak için kullanılan hesaplama yöntem ve teknikleridir.<sup>4</sup> Metin madenciliği ile metinlerde yer alan gereksiz kısımlar atılarak istenilen bilgiye erişmek amaçlanmaktadır. Yapılandırılmamış veri ham haliyle bize çok fazla bilgi vermezken, metin madenciliği sayesinde, bu veriler bizi çok önemli bilgilere ulaştıracak bir kaynak haline gelir. Metin madenciliği aslında büyük bir dağ kitlesinin altında yatan küçük bir pırlantayı elde etmek için sarf edilen çaba diye tanımlanmıştır. Verilerin çok büyük bir kısmı yapılandırılmamış veri yani dağ kitlesi halindedir, bu verilerin kullanılabilmesi için pek çok yöntem geliştirilmiştir. Metin madenciliği geliştirilmiş olan bu yöntemlerin bir bütünüdür.

Veri madenciliği büyük miktardaki veriyi analiz ederek bu veriden anlamlı, kullanılabilir bilgi elde etme sürecidir. Metin madenciliği ise veri madenciliği alanında çok farklı ve yeni çıktılar açan, veri madenciliğine ilham ve yön veren bir alan olmuştur. Bu yüzden ki metin madenciliği ile veri madenciliği arasında birçok üst düzey mimari benzerlik vardır.<sup>5</sup> Basit bir tanımla; veri madenciliği; veri ambarlarından hareketle, verilerden anlamlı bilgiler ve ilişkiler çıkaran, ancak bunu yaparken yapılandırılmış verileri kullanan, çok büyük miktardaki veri arasındaki ilişkiyi analiz etmeye çalışan tekniktir.<sup>6</sup> Tanımdan da anlaşılacağı üzere veri madenciliği çalışma alanı yapılandırılmış veridir. Metin madenciliği ise yapısal olmayan verilerden hareketle veri

---

<sup>4</sup> Jelena Jovanovic, "Introduction To Text Mining", <http://ai.fon.bg.ac.rs/wp-content/uploads/2015/04/Intro-to-TM.pdf>, 11.10.2016, s.3

<sup>5</sup> Ronen Feldman, James Sanger, *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*, Cambridge, 2006. s.1

<sup>6</sup> Haldun Akpınar, 'VERİ TABANLARINDA BİLGİ KEŞFİ VERİ MADENCİLİĞİ', İÜ İşletme Fakültesi Dergisi, C. 29 S. 1, 2000, s.2.

madenciliğinin yaptığı işleri yapar.<sup>7</sup> Aslında veri madenciliği ile metin madenciliği arasında ki temel fark kullanılan verilerin yapılandırılmış olup olmamasıdır. Metinler insanlar tarafından yazılıp okunabilirken, bilgisayarların yapılandırılmış verileri işlemesi için veri tabanları programlanmıştır. Metinleri insanlar gibi okuyup anlayabilecek bilgisayar programları henüz yoktur.<sup>8</sup> Ancak bu alanda ciddi çalışmalar sürdürülmektedir. Gelecekte insanlar gibi metinleri okuyup anlayan bilgisayar programlarının geliştirilmesi sürpriz olmayacaktır.

Veri madenciliğiyle metin madenciliği arasındaki güçlü ilişki nedeniyle gelişim süreçleri de bir birlerine paraleldir. Bilgisayarların günlük hayatta hızla yaygınlaşması, veri depolamanın ve veriye ulaşmanın kolaylaşması, ucuzlaması veri madenciliğini olduğu gibi metin madenciliğini de geliştiren temel neden olmuştur. 1960'lı yıllarda başlayan bu süreç 2000'den sonra hızlanarak gelişimini sürdürmektedir.<sup>9</sup> 1990'lardan sonra metin veri tabanlarında bilgi keşfi (Knowledge Discovery in Textual Databases – KDT) alanında yoğun bir şekilde çalışılmaya başlanmış,<sup>10</sup> 1995 yılında Knowledge Discovery in Textual Databases (KDD) konferanslarının birincisi düzenlemiş, metnin yapısal özelliği nedeniyle işlenmesi zor olduğundan, metin analizleri için algoritmaların geliştirilmesi 1990'ların sonlarını bulmuştur.<sup>11</sup> 90'lardan sonra metin analizi üzerine çalışmalar hızlanmış ve günümüze kadar gelmiş, bu alanda çok çeşitli yöntemler geliştirilmiştir.

Sosyal medyanın, yaşamın bir vazgeçilmezi hale gelmesiyle, gün içinde özellikle internet ortamında, milyonlarca metin halindeki veri ortaya çıkmaktadır. Bu verilerin işlenmesi için geliştirilen metotlar metin madenciliğinin çok fazla gelişmesine yol açmıştır. Yapılan araştırmaya göre Twitter'da, 1 dakikada 347.222 adet tweet atılmaktadır.<sup>12</sup> Facebook gibi Twitter'dan daha fazla kullanılan diğer sosyal mecralar da

---

<sup>7</sup> M. Özgür Dolgun ve Tülin Güzel Özdemir, "Veri Madenciliği'nde Yapsal Olmayan Verinin Analizi: Metin ve Web Madenciliği", İstatistikçiler Dergisi 2, 2009, s. 49.

<sup>8</sup> Dursun Delen ve Martin D. Crossland, "Seeding the Survey and Analysis of Research Literature with Text Mining", *Expert Systems with Applications*, C.34, 2008, s. 1710.

<sup>9</sup> Oğuzlar, a.g.e., s. 4.

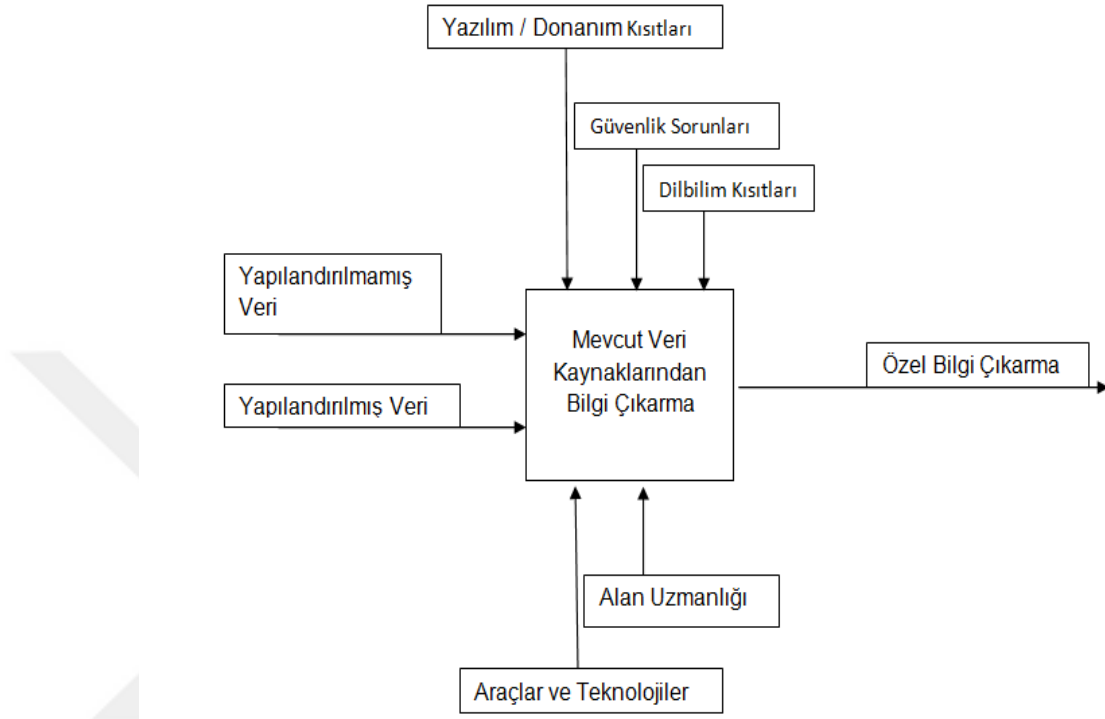
<sup>10</sup> Ronen Feldman ve Ido Dagan, "Knowledge Discovery in Textual Databases (KDD).", *International Conference on Knowledge Discovery and Data Mining (Kdd)*, 1995, s. 112.

<sup>11</sup> Marti A Hearst, 'Untangling Text Data Mining', Association for Computational Linguistics, Stroudsburg, US, 1999, s.3.

<sup>12</sup> <http://www.digitalinformationworld.com/2014/10/what-happens-in-a-twitter-minute-social-media-infographic.html>, 24.10.2016

düşünüldüğünde anlık üretilen verinin büyüklüğü daha da çok ortaya çıkmaktadır. Dolayısıyla bu metinler, araştırmacılar için büyük bir nimet haline gelmiştir.

Şekil 1.1. Metin Madenciliği için Bağlam Diyagramı



**Kaynak:** Delen ve Crossland (2008, 1711)

Yapılandırılmış veya yapılandırılmamış halde bulunan veriler, metin madenciliği için birer girdidirler. Metin madenciliği sürecinin çıktısı ise; karar vermek için kullanılacak olan özel bilgilerdir. Süreç; yazılım donanım kısıtları, güvenlik sorunları ve dilbilim kısıtları (doğal dil işleme) gibi zorlukları içermektedir. Metin madenciliğinin temel amacı verilerden hareketle, metinden anlamlı sonuçlar çıkarmak için metni işlemektir. Bu amaçla metinler çeşitli araç ve alan uzmanlığı (istatistik ve makine öğrenme) ile işlenirler.<sup>13</sup>

## 1.2. Metin Madenciliğinin Aşamaları

Literatürde metin madenciliğinin aşamalarına ilişkin pek çok farklı yaklaşım söz konusudur. Genel hatlarıyla metin madenciliği aşamaları 5 adımda özetlenebilir. Bu adımların bir kısmı farklı aşamaları da içinde barındırmaktadır.

<sup>13</sup> Delen ve Crossland. s. 1711

### 1.2.1. Veri Setinin Oluřturulması

Metin madenciliğinin temel verisini metinler oluřturmaktadır. Veri setini oluřturmak için ilgilenilen konuya dair metinlere ulařmak ilk ařamadır. Günümüzde en önemli metin kaynağı olarak internet karřımıza çıkmaktadır. Google, Yandex gibi arama motorları ile Twitter, Facebook gibi sosyal ağılar en temel kaynaklar olarak görülebilir. Aynı řekilde kurumların oluřturmuř oldukları kendi veri ambarları, yazılmıř raporlar çevirim içi araçlar da veri seti oluřturabilirler. Teknolojinin bize sağılamıř olduđu büyük nimetlerden biri de veri setlerine rahatlıkla ulařabilmektir. Bu konuda hali hazırda var olan veri setleri kullanılabilceğı gibi internet ortamından anlık veri çekebilmek amacıyla geliřtirilmiř olan paket programlar da kullanılabilir. Bazı firmalar ise kendi derledikleri verileri açık kullanıma sunmakta, bireyler veri setine eriřerek bu verileri alıp kullanma řansına sahip olmaktadırlar.

### 1.2.2. Ön İřleme

Veri setinde yapılacak ilk iřlem ön iřlemedir. Çođu zaman elde edilen metinler hemen kullanılabilcek bir yapıda olmamaktadır. Metinler çeřitli nedenlerle uzun bir ön iřleme adımına tabi tutulmak zorundadır. Veriyi kelimelere ayırma, kelimelerin anlamsal deđerlerini bulma, kelime köklerini bulup gereksiz kelimeleri temizleme, imla hatalarını ortadan kaldırma ve yazım yanlıřlarını düzeltme gibi iřlemler bütününe ön iřleme denir.<sup>14</sup> Metin madenciliğinde kullanılabilcek doküman kaynağı için olası biçim ve ham görünüm sayısı oldukça fazladır. Bu durum metinlerin dönüřtürülmesi için güçlü teknikler üretilmesini zorunlu kılmıřtır.<sup>15</sup> Bu zorunluluk ön iřleme adımlarının çok fazla zaman alması sonucunu doğurmuřtur. Ön iřleme ařamasında çok titiz davranılması analiz sonucunun sağılıklı olabilmesi için řarttır. Ön iřleme süreci ařağıda kısaca açıklanan adımları içermektedir.

#### 1.2.2.1. İřaretleme (Tokenization)

Metin içinde yapılması gereken ilk iřlem iřaretlemedir. İřaretleme karakter dizileri olan metinlerin makine öğrenme algoritması için uygun bir hale getirilmesi çabasıdır. Bunun için özellikle web tabanlı bir veri ile çalışılıyorsa webin kendine has karakterleri olan; XML (EXtensible Markup Language) ve HTML (Hyper Text Markup

---

<sup>14</sup> Kemal Kuzucu, Müřteri Memnuniyeti Belirlemek İçin Metin Madenciliğı Tabanlı Bir Yazılım Aracı, Maltepe Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, 2015, s. 13.

<sup>15</sup> Oğuzlar. a.g.e. s.31



Language) gibi etiketlerin atılması olmalıdır.<sup>16</sup> XML, HTML gibi uzantılar her ne kadar web dili için anlamlı olsa da metin analizinde anlamlı birer kelime veya kök değildir. Metinde yer alan noktalama işaretleri ile satır sonu karakteri vb. okunabilir olmayan tüm karakterler boşluk karakteri ile değiştirilir.<sup>17</sup> Bu şekilde metin anlamın çıkarılmasında katkısı olmayacak kısaltma ve noktalama işaretleri gibi unsurlardan ayıklanmış olur.

### 1.2.2.2. Kök Bulma (Stemming)

Kök başlı başında anlama sahip olan kelimedir. “Gel”, “git” gibi herhangi bir ek almadan da anlama sahip olan kelimeleri bulma çalışmasına kök bulma denir. Türkçe sondan eklemeli bir dildir. Dolayısıyla kelime kökleri, çeşitli yapım ekleri ve/veya çekim ekleri almaktadır. Bu sayede aynı kelime köküne sahip pek çok farklı kelime ortaya çıkmaktadır. Kök bulma sayesinde çoğul ekler (evler → ev) ve fiil çekim ekleri (okuyor → oku) ortadan kalkar. Kök bulmada iki farklı sorunla karşılaşılabilir; bunlar kök bulmak istenirken çok ileri gidilerek anlam olarak çok farklı bir kelimeye ulaşılması veya az sayıda ek çıkarılarak asıl köke ulaşamamasıdır.<sup>18</sup> Bu tür hatalara düşmemek için iyi bir dilbilgisine ihtiyaç vardır. Kelimenin özel yapısı gereği kök bir ek alırken; bazen harf düşmesi, bazen harf yumuşaması, bazen kaynaştırma harfi alması gibi veya büyük ünlü uyumu küçük ünlü uyumu gibi süreçlerde söz konusu olabilmektedir. Dile hakim olmadan yapılacak kök bulma işlemi beklenenin aksine yaradan çok zarar getirebilir bu yüzden kök bulurken son derece dikkatli olmak gerekir.

Türkçe için kelime kökenini bulmak amacıyla geliştirilmiş olan, açık kaynak kodlu Zemberek programı mevcuttur.<sup>19</sup> Bu program sayesinde kelime köklerine ayırma işlemi yapılabilmektedir.

### 1.2.2.3. Durak Kelimeleri Çıkartma

Durak kelimeler (stop words); edat, bağlaç, zamir gibi başlı başına anlamı olmayan ancak yazım dilinde çok sık kullanılan kelimelerdir. Durak kelimelerin

---

<sup>16</sup> Emine Kübra Çelikyay, ‘Metin Madenciliği Yöntemiyle Türkçe’de En Sık Kullanılan Ve Birbirini Takip Eden Harflerin Analizi Ve Birliktelik Kuralları’ (Beykent Üniversitesi Fen Bilimleri Enstitüsü, 2010)., Yüksek Lisans Tezi, İstanbul, 2010, s. 60.

<sup>17</sup> Aykut Demirel, Metin Madenciliği Yöntemleri İle Sosyal Medyadan Toplanan Fotoğraflı Paylaşımların, Metin – Fotoğraf Eşleşmesinin İncelenmesi, Beykent Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 2015, s. 12.

<sup>18</sup> Çelikyay, a.g.e., s. 61

<sup>19</sup> Mehmet Dündar Akın ve Ahmet Afşin Akın, "Türk Dilleri İçin Açık Kaynaklı Doğal Dil İşleme Kütüphanesi: ZEMBEREK", Elektrik Mühendisliği, 431. Sayı, 2007, s. 38

metinden çıkarılması gerekir.<sup>20</sup> Durak kelimeler çıkarılarak, metin içinde yer alan ancak herhangi bir bilgi taşımayan kelimeler metinden atılmış olur. Bu sayede işlenecek kelime sayısında ciddi anlamda azalma görülebilir. “İle”, “çünkü”, gibi kimi kelimeler Türkçe’de çok sık tekrarlanmaktadır. Analize başlanmadan önce; “sadece”, “ile”, “çünkü”, “gibi”, kelimelerin atılması bile analiz edilen kelime sayısında ciddi bir düşüşe neden olacak ve yapılan analizin doğruluk oranını yükseltecektir.

#### 1.2.2.4. Terim Ağırlıklandırma

Bir metinde yer alan her bir kelimenin metnin içeriğine olan katkısı değişiktir. Bir terim metin içinde sadece bir kez de geçebilir, defalarca da geçebilir. Bir terimin ağırlığı o terimin frekansının metinde geçen diğer tüm terimlerin frekanslarına göre göreceli değeriyle ölçülür. Bir sözcük, doküman içinde ne çok geçiyorsa dokümanın ilgili kategoriye atanmasında o kadar etkili olur, bir sözcük birden çok dokümanda geçiyorsa, o sözcüğün ayırt edici özelliği o ölçüde ortadan kalkar.<sup>21</sup> Aynı şekilde belli terimler bir arada kullanılıyor olabilirler. Bir arada geçen terimler genellikle aynı kategoriye atanırlar.

#### 1.2.2.5. Terimleri Ayıklama

Bir terim bir metinde sadece bir kez geçmişse o terimin göz ardı edilmesi mümkündür. Frekansı düşük olan kelimeler metinlerden ayıklanabilirler, bu sayede çalışılacak olan metinde boyut azaltılmış olur. Bir kelime metinde sadece bir kez kendine yer bulmuş ve bir daha aynı kelime metinde yer almıyorsa bu tür kelimeler metinde ayıklanabilirler. Terim ayıklamanın altında yatan temel neden, düşük frekanslı kelimelerin metin bağlamında daha az öneme sahip olduğunun kabul edilmesidir.<sup>22</sup> Terim ayıklaması sayesinde; zamandan büyük bir tasarruf sağlanacaktır çünkü işlem yapılacak olan matris fazlasıyla küçülecektir; unutmamak gerekir ki yapılan çalışmayla ilgisi olmayan kelimeler daha fazla emek harcanmasına, daha fazla zaman kaybına sebep olmaktadır.

---

<sup>20</sup> Ahmet Haltaş, Ahmet Alkan ve Mustafa Karabulut, ‘Metin Sınıflandırmada Sezgisel Arama Algoritmalarının Performans Analizi’, *Journal of the Faculty of Engineering and Architecture of Gazi University Cilt*, 30.3, 2015, s. 419.

<sup>21</sup> Timo Lahtinen, *Automatic Indexing: An Approach Using an Index Term Corpus and Combining Linguistic and Statistical Methods*, University of Helsinki Department of General Linguistics, Helsinki, 2000.s. 119

<sup>22</sup> Haltaş, Alkan ve Karabulut., a.g.e. s. 419.

### 1.2.3. Özellik Seçme

Ön işleme aşaması tamamlandıktan sonra özellik seçme aşamasına geçilir. Özellik seçme de aslında bir ayıklama aşamasıdır. Veri kümesi özellik seçiminden önce çok yüksek boyutlu ancak içinde nitelikli ve niteliksiz terimleri barındıran bir yapıdadır.<sup>23</sup> Özellik seçme sürecinde; ön işleme aşamasından geçen verilerdeki önemli kelimeleri belirleme ve ilişkili olmayan özelliklerin çıkarılması, az sayıda dokümanda yer alan özelliklerin ayıklanması, çok sayıda dokümanda yer alan özelliklerin azaltılması işlemleri yapılır.<sup>24</sup>

Özellik seçimi sayesinde büyük boyutlu veriler daha küçük bir boyuta indirgenmiş, içinden niteliksiz terimler ayıklanmış, metin üzerinde çalışmaya uygun bir hal almış olur. Bu süreç zaman tasarrufu yanı sıra sınıflandırmanın başarılı yapılması ihtimalini de arttırmaktadır.

### 1.2.4. Sınıflandırma

Sınıflandırma süreci benzer dokümanların aynı sınıflara ayrılması süreci olarak ifade edilebilir. Temel amaç; metinleri anlamsal olarak, önceden belirlenmiş sınıflara otomatik olarak ayırmaktır. Dokümanların otomatik olarak sınıflandırılabilmesi için vektörel olarak ifade edilmesi gerekir. Bunun için de “vektör uzay modeli” oluşturulmalıdır.<sup>25</sup>

#### 1.2.4.1. Vektör Oluşturma

Tüm belgeler bir kelime vektörü olarak gösterilirler. Bunun temel amacı bilgisayar üzerinde bir “kelime \* belge” gibi iki boyutlu dizi elde etmektir. Bu şekilde işlemler bilgisayar tarafından çok daha kolay yapılabilmektedir.<sup>26</sup>

Bu iki boyutlu bir matristir, bu matrise A matrisi dersek, matrisin elemanlarının her biri; D belgesinde geçen T kelimesinin belgenin analizi için taşıdığı önemi gösterir. Bu rakam en yalın anlatımıyla kelimenin belgedeki frekansının kaç olduğunu gösterir.<sup>27</sup>

---

<sup>23</sup> Haltaş, Alkan ve Karabulut., s. 420

<sup>24</sup> Kadriye Ergün, Metin Madenciliği Yöntemleri İle Ürün Yorumlarının Otomatik Değerlendirilmesi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya, 2012. s. 51

<sup>25</sup> Pilavcılar İsmail, "Metin Madenciliği İle Metin Sınıflandırma 1", <http://www.csharpnedir.com/articles/read/?id=731>, 08.11.2016

<sup>26</sup> Michael W Berry, Zlatko Drm ve Elizabeth R Jessup, "Matrices, Vector Spaces, And Information Retrieval", <https://www.cs.colorado.edu/~jessup/SUBPAGES/PS/matrices.pdf>, 09.11.2017, s.4.

Vektör oluşturmada üç farklı alternatif kullanılabilir. Bunlar; kelime vektörü, N-Gram kelime modeli ve boyut ağırlıklandırmasıdır.

Kelime vektörü; kelimelerin tekil olarak ele alınıp sınıflarıyla beraber gösterilmesi ile elde edilen vektördür. Kelime vektörü sayesinde verileri en iyi temsil edecek öznitelikler kümesinin oluşturulması amaçlanır.<sup>28</sup>

N - Gram Kelime Modeli; belgede yer alan kelimelerin “n” âdeti bir arada kullanılarak kelime vektörünün oluşturulmasıdır. Tek başına bir kelime ele alındığında genellikle anlam bütünlüğü sağlanamaz ancak kelime yanındaki diğer kelimelerle birlikte ele alınırsa daha sağlıklı sonuçlar ortaya çıkabilir. Burada N sayısı kelimenin yanında yer alan “kaç kelime ile birlikte” ele alındığını gösterir. N iki olarak ele alınırsa metinde yer alan kelimeler ikişerli grup halinde ele alınır, n üç olarak ele alınırsa kelimeler üçerli gruplar halinde ele alınır. Uygulamada genellikle kelimeler bir, iki veya üçerli gruplar halinde ele alınır.

Boyut ağırlıklandırma; kelime vektörü oluşturulurken kelimelerin sıklıklarının baz alınmasıdır. Geliştirilmiş yöntemler özetle şu şekildedir; kelime varsa 1 yoksa 0 değerinin verilmesi veya kelimenin frekansının vektörde yer alması şeklinde uygulanabilmektedir.

#### **1.2.4.2. Sınıflandırma Yöntemleri**

Kelime vektörleri oluşturulduktan sonra uygun sınıflandırma yöntemi kullanılarak sınıflandırma işlemine geçilebilir. Metin madenciliğinde pek çok farklı sınıflandırma algoritması yer almaktadır. Bu alanda en fazla kullanılan algoritmalar aşağıda açıklanmıştır.

##### **1.2.4.2.1. Navie Bayes Sınıflandırıcı**

Navie Bayes sınıflandırıcı adını İngiliz matematikçi Thomas Bayes'ten almaktadır.<sup>29</sup> Navie Bayes sınıflandırıcı varsayımları Bayes Teoremine dayanan, olasılıksal bir sınıflayıcıdır. Navie Bayes sınıflandırıcı ile özellikle boyutları çok büyük veri setleri için parametre tahmini yapmak kolaydır. Navie Bayes sınıflandırma, sade bir

<sup>27</sup> Güven Ahmet, Türkçe Belgelerin Anlam Tabanlı Yöntemlerle Madenciliği, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul 2007, s. 46.

<sup>28</sup> Topaçan Ümit, ‘Sosyal Medya Paylaşımlarında Duygu Analizi: Makine Öğrenimi Yaklaşımı Üzerine Bir Araştırma’, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, İstanbul, 2016, s 83.

<sup>29</sup> Tom M. Mitchell, “Machine Learning”, McGraw-Hill Science, USA, 1997 s.155

tasarıma sahip olmasına rağmen, gelişmiş sınıflandırma yöntemleriyle kıyaslandığında çok daha iyi performans göstermektedir. Navie Bayes sınıflandırıcı bu özelliğinden dolayı çok yaygın bir kullanıma sahiptir.<sup>30</sup> Yapılan çalışmalarla Navie Bayes sınıflandırıcının sınıflandırmada çok etkili olduğu defalarca kanıtlanmıştır.<sup>31</sup>

Bayes teoremi rassal değişkenler için koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi gösterir.<sup>32</sup>

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1.1)$$

$P(A|B)$ : B olayı gerçekleşirse A olayının meydana gelme olasılığı

$P(B|A)$ : A gerçekleşirse B olayının meydana gelme olasılığı

$P(A)$ : A olayının ön olasılığı

$P(B)$ : B olayının ön olasılığı

Navie Bayes sınıflandırma modeli; birçok özellikten ve bir sonuç değişkeninden oluşur. Hesaplamalar, sınıflar düzeyinde gerçekleştirilir, her bir sınıf için olasılık değeri hesaplanarak en yüksek olasılık değerine sahip sınıf, sınıflandırılması yapılacak olan dokümanın sınıfıdır.<sup>33</sup>

Bayes Teoreminden hareketle Navie Bayes sınıflandırmaya geçmek için; elimizde n adet sınıf olduğunu varsayalım. Sınıfları S harfi ile, herhangi bir sınıfa ait olmayan bir veriyi X harfi ile ifade edecek olursak, X'in elimizde bulunan n adet sınıftan hangisine ait olduğu Navie Bayes sınıflandırma ile hesaplanabilir. Yapılan hesaplama neticesinde sınıfı belli olmayan "X" isimli veri, en yüksek olasılığa sahip olan S sınıfına atanacaktır. Navie Bayes sınıflandırma sayesinde X'in atanacağı S sınıfı

<sup>30</sup> Navie Bayesian, [http://www.saedsayad.com/naive\\_bayesian.htm](http://www.saedsayad.com/naive_bayesian.htm), 10.11.2016

<sup>31</sup> Wenyuan Dai vd diğerleri, "Transferring Naive Bayes Classifiers for Text Classification", AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence - Volume 1, Canada, 2007, s 541.

<sup>32</sup> Paola Sebastiani, 'A Tutorial on Probability Theory', <http://www.math.umass.edu>, s. 6.

<sup>33</sup> Karaca Mehmet Fatih, Metin Madenciliği Yöntemi İle Haber Sitelerindeki Köşe Yazılarının Sınıflandırılması, Karabük Üniversitesi Fen Bilimleri Enstitüsü, 2012, s.27.

tespit edilmiş olacaktır. Burada X'in sınıfını belirleyen değer aslında olasılıksal bir yaklaşımdır. Hesaplanan şey, X'in hangi sınıfa ait olabileceğinin olasılığıdır. X'ler m boyutlu özellik vektörü ile gösterilir, özelliklerin hepsi aynı derecede önemli ve birbirinden bağımsızdır, bir özelliğin değeri başka bir özellik değeri hakkında bilgi içermez.<sup>34</sup>

Bayes teoremini bu sürece uyarlırsak;

$$P(S_i|X) = \frac{P(X|S_i)P(S_i)}{P(X)} \quad (1.2.)$$

Eşitlikte P(X) ler bütün sınıflar için sabitse X'in S<sub>i</sub> sınıfında olma olasılığı P(X|S<sub>i</sub>).P(S<sub>i</sub>) olur.

P(S<sub>i</sub>), her bir sınıfın olasılığını ifade eder. S<sub>i</sub> sınıfına ait eğitilen örnek sayısına O<sub>i</sub>, toplam eğitilen örnek sayısına O dersek P(S<sub>i</sub>) nin değeri aşağıdaki gibi olur;<sup>35</sup>

$$P(S_i) = \frac{O_i}{O} \quad (1.3.)$$

Sınıfların ön olasılıkları bilinmiyorsa, o zaman genel olarak sınıflar eşit kabul edilir. P(S<sub>1</sub>) = P(S<sub>2</sub>) = ..... = P(S<sub>n</sub>), bu yüzden P(X|S<sub>i</sub>) ifadesi, X'in S<sub>i</sub> sınıfında olma olasılığını bulmak için kullanılır. Aksi taktirde P(X|S<sub>i</sub>).P(S<sub>i</sub>) ifadesi bizim için en anlamlı ifadedir. Bilinmeyen örnek X'i sınıflandırmak için, her S<sub>i</sub> sınıfı için P(X|S<sub>i</sub>).P(S<sub>i</sub>) hesaplanır, örnek X'i en yüksek değere sahip S<sub>i</sub> sınıfına atanır.<sup>36</sup> Bu hesaplama için aşağıdaki bağıntıdan yararlanır;

---

<sup>34</sup> Doğan Sibel, Türkçe Dokümanlar İçin N-Gram Tabanlı Sınıflandırma: Yazar, Tür Ve Cinsiyet, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, 2006, s. 22.

<sup>35</sup> Doğan, a.g.e. s 22.

<sup>36</sup> Doğan, a.g.e. s 23.

$$P(X \setminus S_i) = \prod_{k=1}^m P(X_k \setminus S_i) \quad (1.4.)$$

#### 1.2.4.2.2. K-En Yakın Komşuluk (KEYK)

K-en yakın komşuluk (K-Nearest Neighbor – KNN olarak da ifade edilmektedir), Navie Bayes sınıflayıcıdan daha basit bir tahmincidir. Yapılan, X değerlerinin en yakınlardaki K adet komşuların sınıf etiketlerine bakılarak sınıflandırmadır. Bu sınıflandırmayı yapabilmek için ihtiyaç duyulan tek şey gözlem çiftleri arasındaki mesafedir. Bu mesafenin simetrik olma zorunluluğu yoktur.<sup>37</sup>

KNN'nin temelinde birbirine yakın olan nesnelere muhtemelen aynı kategoriye ait olabileceği mantığı vardır.<sup>38</sup> Genellikle aynı sınıfa ait olan terimler bir arada kullanılır. Birbirine yakın olan terimlerin aynı sınıfta değerlendirilebilmesi için önceden tanımlanmış olan eğitim verilerinin özellikleri ile sorgu örneği arasındaki mesafeye bakılır. Hangi sınıfa ait olduğu belirlenmek istenen sorgu örneği, eğitim verilerinin en yakınındaki k adet örnek çoğunlukla hangi sınıfa aitse, sorgu örneği de o sınıfa atılır denir. Amaç sınıflandırılmak istenen sorgu örneğinin daha önceden belirlenmiş olan eğitim verilerinde k tanesine yakınlığını hesaplamaktır. Sorgu örneği eğitim verilerinden herhangi k tanesine yakın değilse veya hepsine aynı uzaklıkta ise bir sınıfa atanma yapılmaz.

KNN'nin basit olmasına karşın, çeşitli avantaj ve dezavantajları vardır. Avantajları; uygulamasının basit olması, eğitim verilerinde yer alan gürültülere karşı etkili olması ve eğitim dokümanlarının sayısı arttıkça etkinliğinin artmasıdır. Dezavantajları ise; eğitim seti büyük olduğu zaman performansı düşük olabilir, tüm özellikler sınıflandırmaya katkı sunduğu için ilgisiz veya gereksiz özelliklere karşı çok duyarlıdır, algoritmanın başlangıçta K parametresine ihtiyaç duyması, en iyi sonucun alınabilmesi için hangi uzaklık ölçümünün uygulanacağını açık olmaması (çünkü soru

---

<sup>37</sup> Alex Smola ve S. V. N. Vishwanathan, *Introduction To Machine Learning*, Cambridge, Cambridge University, 2008, s 24.

<sup>38</sup> Erten, a.g.e. 13.

örneğin tüm eğitim örneklerine mesafesinin ölçülmesi gerekir) ve hesaplama maliyetinin yüksek olması sayılabilir.<sup>39</sup>

KNN'nin adımları:<sup>40,41</sup>

1. Öncelikle K değeri belirlenir. Bu değer verilen bir noktaya en yakın komşuların sayısıdır.
2. Diğer nesnelere hedef nesneye olan uzaklıkları hesaplanır.
3. Uzaklıklara göre satırlar sıralanır, bunlardan en küçük olan k tanesi seçilir.
4. Seçilen satırların, hangi kategoriye ait oldukları belirlenir ve en çok tekrar eden kategori değeri seçilir.
5. Seçilen kategori, en uygun komşu kategorisi olarak kabul edilir.

KNN'de en yakın komşuluk, Öklit, Manhattan, Minkowski ve Chebyshev Uzaklıkları hesaplanarak bulunur.

Öklit Uzaklığı; kümeleme ve sınıflandırma algoritmalarında en sık kullanılan uzaklık ölçütü, öklit uzaklığıdır. Öklit uzaklığı doğrusaldır.

$X=(x_1, x_2, \dots, x_n)$  ve  $Y=(y_1, y_2, \dots, y_n)$  arasındaki öklit uzaklığı aşağıdaki bağıntıdan hareketle hesaplanır.<sup>42</sup>

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1.5.)$$

Manhattan Uzaklığı; iki nokta arasındaki uzaklığın mutlak değerlerinin toplanması ile hesaplanan yöntemdir.

$X=(x_1, x_2, \dots, x_n)$  ve  $Y=(y_1, y_2, \dots, y_n)$  arasındaki city-blok uzaklığı da denilen manhattan uzaklığı aşağıdaki bağıntıdan hareketle hesaplanır.<sup>43</sup>

---

<sup>39</sup> Sadegh Bafandeh Imandoust ve Mohammad Bolandraftar, 'Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background', *Journal of Engineering Research and Applications*, Vol 3, 2013 s 608.

<sup>40</sup> Sayali D Jadhav ve H P Channe, 'Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques', *International Journal of Science and Research*, Vol. 5, 2016, s. 1842.

<sup>41</sup> Imandoust ve Bolandraftar age s. 607-608

<sup>42</sup> Doğan Sibel a.g.e. s 28.



$$D(X, Y) = |x_i - y_i| \quad (1.6.)$$

Minkowski Uzaklığı; öklit ve Manhattan uzaklıklarının geliştirilmiş halidir.

$X=(x_1, x_2, \dots, x_n)$  ve  $Y=(y_1, y_2, \dots, y_n)$  arasındaki city-blok uzaklığı da denilen minkowski uzaklığı aşağıdaki bağıntıdan hareketle hesaplanır<sup>44</sup>.

$$D(X, Y) = [\sum_{k=1}^n |x_{ik} - x_{jk}|^m]^{1/m} \quad (1.7.)$$

Chebyshev Uzaklığı; Minkowski uzaklığında özel bir durum olarak;  $n \rightarrow \infty$  iki nokta arasındaki farkın mutlak değerinin maksimumu olarak tanımlanır.<sup>45</sup> Chebyshev uzaklığı aşağıdaki bağıntı ile hesaplanır.

$$\lim_{n \rightarrow \infty} [\sum_{k=1}^n |x_{ik} - x_{jk}|^m]^{1/m} = \max_{i=1}^n |x_i - y_i| \quad (1.8.)$$

#### 1.2.4.2.3. Destek Vektör Makinesi (DVM)

Viladimir N. Vapnik tarafından geliştirilmiş olan destek vektör makinesi (Support Vector Machine – SVM) basit ve son derece etkili yöntemlerden biridir. Güçlü teorik temellere ve mükemmel ampirik başarılarla sahip olan DVM’de sınıflandırma için doğrusal bir alanda bulunan iki grup arasında bir sınır çizilerek iki grubu ayırmak mümkündür. Bu sınırın çizileceği yer ise iki grubunda üyelerine en uzak olan yerdir. DVM bu sınırın nasıl çizileceğini belirler.<sup>46</sup>

DVM’nin belirlenebilmesi için her iki gruba eşit uzaklıkta ve paralel iki alan çizgisi çekilir, bu alan çizgileri birbirine yakınlaştırılır ve ortak bir alan çizgisi meydana getirilir.<sup>47</sup>

DVM’nin avantajlı yönleri; özellikle optimizasyon problemleri için benzersiz çözümler sunar, öğrenme süreci son derece hızlıdır, aynı anda hem destek vektör

<sup>43</sup> Haşiloğlu Selçuk Burak, ‘Algı Haritalarının Değerlendirilmesi İçin Kümeleme Algoritmalarına Dayalı Yeni Bir Model Geliştirilmesi’ (Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, 2017). 2017, Pamukkale Üniversitesi, s.26

<sup>44</sup> Haşiloğlu Selçuk Burak, ‘Algı Haritalarının Değerlendirilmesi İçin Kümeleme Algoritmalarına Dayalı Yeni Bir Model Geliştirilmesi’ (Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, 2017). s.29

<sup>45</sup> <http://yazilimagiris.com/2017/11/k-en-yakin-komsu-methodu-k-nearest-neighborhood/?print=print>

<sup>46</sup> Saygılı Ahmet, ‘Veri Madenciliği İle Mühendislik Fakültesi Öğrencilerinin Okul Analizlerinin Başarıları’ (Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, 2013).

<sup>47</sup> Kuzucu a.g.e. s 19.

kümesi hem de karar kuralı inşa eder ayrıca karar fonksiyonları için yeni bir dizi uygulaması olabilir.<sup>48</sup>

#### 1.2.4.2.4. Karar Ağacı

En basit sınıflandırma tekniklerinden birisi de karar ağaçlarıdır. Bir ağaç görünümüne sahip ve uygulanması kolay bir tekniktir. Karar ağaçları son yıllarda yaygın olarak kullanılmaktadır, bunun temel nedeni, karar ağaçlarında kullanılan kuralların sade ve anlaşılabilir olmasıdır.<sup>49</sup>

Bir karar ağacının yapısı; düğüm, dal ve yaprakтан oluşmaktadır. Karar ağacı yapısı, kök düğümünden başlar ve yaprağa ulaşıncaya kadar devam eder. Karar düğümünden iki tane dal çıkar bu dallardan bir tanesi istenilen terimin belgede bulunduğu durumu, diğeri ise bulunmadığı durumu gösterir.<sup>50</sup> Her bir düğümde dallara ayrılma işlemi yapılır, eğer bir dalın ucunda artık sınıflama yapılamıyorsa o noktada karar düğümü oluşur, belirli bir sınıfa ulaşıyorsa o dalın sonunda yaprak vardır. Bu yaprak, verilerden hareketle belirlenmek istenilen sınıflardan birisidir.<sup>51</sup>

Karar ağaçlarının yoğun olarak kullanıldıkları alanlar:<sup>52</sup>

- Belli bir sınıfın üyesi olacak elemanın belirlenmesi,
- Çeşitli vakaların risk kategorilerine ayrılması,
- Gelecekteki olayların tahmin edilmesi için kurallar oluşturulması,
- Parametrik modellerin kurulmasında kullanılmak üzere çok miktardaki değişken ve veri kümesinden faydalı olacakların seçilmesi
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması,
- Bazı kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikliye dönüştürülmesi.

Karar ağaçları her ne kadar uygulanması kolay olsa da dal sayısı arttıkça kullanım kolaylığını yitirebilir. Ayrıca karar ağaçlarını kullanırken alternatif

<sup>48</sup> Vladimir N Vapnik, "An Overview of Statistical Learning Theory", *IEEE TRANSACTIONS ON NEURAL NETWORKS*, Vol 10. No 5, 1999, s 997.

<sup>49</sup> Kavzoğlu Taşkın ve Çölkesen İsmail, "Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği", *Harita Teknolojileri Elektronik Dergisi*, C.2., No:1, 2010, s 39.

<sup>50</sup> Tantuğ A. Cüneyd, 'Metin Sınıflandırma', *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5.6 (2012).

<sup>51</sup> Tantuğ A. Cüneyd, 'Metin Sınıflandırma', *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5.6 (2012).

<sup>52</sup> Akpınar, a.g.e.,s.14

algoritmalarından hangisinin kullanılacağına da karar verilmesi gerekir. Karar ağaçları algoritmalarının atası olarak AID (Automatic Interaction Detector) karar ağacı kabul edilmektedir. 1970’li yıllarda Morgan ve Sonquist tarafından geliştirilmiştir.<sup>53</sup>

AID karar ağaçları ilk geliştirilen algoritma olmasına karşın temel bilgiyi edinmek için çok fazla kullanılmamıştır. 1980 yılında G.V. Kass tarafından geliştirilen CHAID (Chi-squared Automatic Interaction Detector – Ki-kare Otomatik Etkileşim Dedektörü) algoritması bağımlı değişken sürekli ise F testi, kategorik ise Ki Kare testini kullanır. Veriler homojen alt gruplara ayrılır, optimal bölünmenin teşhisi için ki-kare istatistiğinden faydalanılır. İstatistiksel testin anlamlılığı sayesinde ön kestiriciye ait değişkenler değerlendirilir. Bağımlı değişkene göre homojen olan tüm değerler birleştirilir, karar ağacındaki ilk dalın formuna göre en iyi ön kestirici değişkenin seçilmesi ile birlikte her bir düğümün seçilen değişkenin homojen değerinin bir grubu oluşturulur. Ağaç tamamen büyüyene kadar işleme devam edilir.<sup>54</sup> CHAID algoritmasının diğer karar ağacı algoritmalarından en önemli farkı, diğer algoritmalar ikili ağaçlar üretirken, CHAID algoritması çoklu ağaçlar üretebilmektedir.<sup>55</sup>

1984 yılına basılmış olan “Classification And Regression Trees” adlı kitap ile CART (C&RT) isimli yeni bir karar ağacı algoritması ortaya konulmuştur. CART sınıflandırma ve regresyon modellerinin üretilmesi için kullanılan bu algoritma, Leo Breiman, Charles J. Stone, Jerry Friedman ve R. Olshen tarafından geliştirmiştir. CART ile ağaç oluşturulurken, veriler iki alt kümeye ayrılarak ilerlediği için alt küme bir üst kümeye göre daha homojen bir yapıya sahip olur, homojenlik tam olarak sağlanana dek işlem devam eder.<sup>56</sup> Homojenliğin tam olarak sağlanmasına, kategorik değişkenlerde Gini Katsayısı, sürekli değişkenlerde ise en küçük kareler sapması gibi hesaplamalara dayanarak karar verilir. Homojen yapıya ulaşıldıktan sonra bu sefer tersi yönde bir budama işlemi yapılarak işlem sonuçlandırılır.<sup>57</sup>

<sup>53</sup> Gül Gökay Emel Ve Çağatan Taşkın, ‘Veri Madenciliğinde Karar Ağaçları Ve Bir Satış Analizi Uygulaması’, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6.2 (2005) s.228.

<sup>54</sup> Çalış Aslı, Kayapınar Sema ve Çetinyokuş Tahsin, ‘Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama’, *Endüstri Mühendisliği Dergisi*, 25.21 (2014), 2–19.

<sup>55</sup> Çalış ve diğerleri a.g.e.

<sup>56</sup> Oğuzlar Ayşe, ‘Cart Analizi İle Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi’. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, 2010.

<sup>57</sup> Oğuzlar Ayşe, ‘Cart Analizi İle Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi’. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, 2010

QUEST Karar ağacı 1997 yılında Loh ve Shih tarafından geliştirilmiştir. Quick (hızlı), Unbiased (yansız), Efficient Stastical Tree (etkili istatistiksel ağaç) ikili karar ağacı yapısı kullanılan bir algoritmadır. QUEST algoritması, ağaç oluşturulurken ön yargıya dayalı seçimin genel hale getirilmesi ve hesaplama maliyetinin düşürülmesini amaçlar. QUEST algoritması tek değişkenli ve doğrusal kombinasyon bölünmeleri için kullanılır. Her bölme için, her bir girdi özniteliği ile hedef özniteliği arasındaki ilişki F testi veya Levene testi (sıralı ve sürekli öznitelikler için) veya Pearson'ın ki-kare (nominal öznitelikler için) kullanılarak hesaplanır. Hedef özniteliği çok terimli ise iki süper küme oluşturmak için iki yönlü kümeleme kullanılır.<sup>58</sup> QUEST karar ağacı algoritması ihmal edilebilir bir önyargıya sahiptir ve ikili karar ağaçları üretir. Quest karar ağaçları algoritması tek değişkenli veya lineer kombinasyon bölünmeleri ile kullanılabilir.

ID3 (Iterative Dichotomizer 3) Karar ağaçları, 1983 yılında, CART ile aynı dönemde fakat birbirlerinden bağımsız olarak çıkmıştır.<sup>59</sup> J.R. Quinlan tarafından geliştirilen entropi kavramından faydalanarak karar ağacını oluşturur. Entropi, bir veri kümesindeki rassalığı, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir. Beklentisizliğin maksimumlaştırılmasıdır. Eğer veri kümesindeki tüm değişkenler homojen ise, entropi değeri 0 çıkacaktır. Bütün olasılıklar eşitse entropi maksimum değerine ulaşır ve entropi bir olasılığı gösterdiği için alabileceği değer 0'la 1 arasındadır. S sınıf kümesi için ortalama entropi değeri aşağıdaki bağıntı ile hesaplanır.

$$E(S) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (1.9)$$

Sadece kategorik veriler için kullanılan ID3 algoritmasının arkasındaki temel fikir, her bir düğümden her bir özellik için yukarıdan aşağıya doğru belli bir kümenin karar ağacını oluşturabilmektir.<sup>60</sup> Veri seti bölünmeden önce elde edilen bilgi ile veri seti bölündükten sonra gelen bilgi arasındaki fark kazanım (gani) olarak adlandırılır. Kazanım miktarını hesaplamak için entropiden faydalanılır. Kazanımlar aşağıdaki formülden hareketle hesaplanır.

<sup>58</sup> Loh Wei-Yin and Shih Yu-Shan, 'Split Selection Methods For Classification Trees', *Statistica Sinica*, 7 (1997), 815-40.s.819

<sup>59</sup> Loh Wei-Yin and Shih Yu-Shan, 'Split Selection Methods For Classification Trees', *Statistica Sinica*, 7 (1997), 815-40.

<sup>60</sup> Oshoibhor Emmanuel O., John-OtumuAdetokunbo M. and Ojieabu Clement E., 'An Implementation of ID3 Decision Tree Learning Algorithm for Tax Fraud Control and Prevention System', *Journal of Computer Science and Information Technology*, 4.42 (2016), 73-93 s.75.

$$Gain(T, X) = Entropy(T) - Entropi(T, X)$$

(1.10)

Bölünmeden önceki entropi ile her bir alt bölümün entropisi arasındaki fark alınır. Maksimum entropi bölünmede en uygun özelliktir. Entropinin maksimum olduğu dallanma yapılıdır.<sup>61</sup>

C4.5 (Commercial Version 4.5) karar ağaçları, 1993 yılında tıpkı ID3'te olduğu gibi Ross Quinlan tarafından geliştirilmiştir. ID3 kullanışlı bir algoritma olmasına rağmen çok sayıda değeri olan özelliklere karşı aşırı duyarlıdır, C4.5'in amacı ise bu duyarlılığı ortadan kaldırmaktır.<sup>62</sup> Pek çok karar ağacı sınıflandırma algoritması olmasına karşın, C4.5 uygulama kolaylığı, diğer sınıflandırma algoritmalarına nazaran daha kolay anlaşılması, ağaç üzerinde kolaya karar verebilme gibi artıları nedeniyle diğer karar ağacı algoritmalarından daha fazla kullanılır.<sup>63</sup>

C4.5 karar ağaçları, ID3'teki gibi bir dizi eğitim verisi veya veri seti kullanılarak inşa edilmiştir. Ağacın her bir düğümünde, C4.5, bir dizi veya zenginleştirilmiş alt kümeler, örnek kümesini en etkili şekilde ayıran verilerin bir özneliğini seçer. Seçimi yaparken kullandığı ölçütler, normalleştirilmiş bilgi kazanımı yani entropi farkıdır. En yüksek normalleştirilmiş bilgi kazancı olan özellik (entropi farkı), karar vermek için seçilir.<sup>64</sup> C4.5 karar ağacının geliştirilmiş hali olarak C5.0 algoritması ortaya çıkarılmıştır. C5.0 ile C4.5 arasında temelde tahmin hatasında büyük bir fark yoktur fakat işlem süresi ve karar ağacının sadeliği konularında C5.0'in C4.5'e bir üstünlük sağlamaktadır.<sup>65</sup>

Literatürde karar ağaçlarıyla ilgili pek çok çalışmayla karşılaşmak mümkündür. Tüm karar ağaçlarının uygulanabilmesine karşın en fazla yapılan çalışma, kullanım kolaylığı nedeniyle ID3 ve C4.5 karar ağaçlarıdır.

---

<sup>61</sup> Nishant Mathur ve diğerleri, 'The Base Strategy for ID3 Algorithm of Data Mining Using Havrda and Charvat Entropy Based on Decision Tree', *International Journal of Information and Electronics Engineering*, 2.2 (2012) s.253.

<sup>62</sup> Hssina Badr ve diğerleri, 'A Comparative Study of Decision Tree ID3 and C4.5', *International Journal of Advanced Computer Science and Applications*, 2014, s.15.

<sup>63</sup> Harvinder Chauhan ve Anu Chauhan, 'Implementation of Decision Tree Algorithm C4.5', *International Journal of Scientific and Research Publications*, 3.10 (2013), s.1.

<sup>64</sup> Hssina ve diğerleri s.15.

<sup>65</sup> Akkçük Ulaş, *Veri Madenciliği Kümeleme Ve Sınıflama Algoritmaları* İstanbul: Yalın Yayıncılık, 2011 s.82.

#### 1.2.4.2.5. Rastgele Orman (RO)

Rastgele Orman (RO), ağaç tipi bir sınıflandırma metotlarından biridir. RO'da, her düğüm, rassal olarak seçilen değişkenlerden en iyisi kullanılarak dallara ayrılır. Her veri seti orijinal veri setinden, yer değiştirmeli olarak türetilir daha sonra rastgele özellik seçimi kullanılarak ağaçlar geliştirilir. Geliştirilen ağaçların budanmaması gerekir ki bu strateji RO'nun doğruluğunu eşsiz yapar.<sup>66</sup>

#### 1.2.5. Sonuçların Değerlendirilmesi

Metin madenciliğinin önemli bir aşaması da sonuçların değerlendirilmesi kısmıdır. Sonuçların alanında uzman bireyler tarafından değerlendirilmesi metin madenciliğinin etkinliğini arttıracaktır. Sonuçların değerlendirme aşamasında istenmeyen bir durum karşımıza çıkmışsa sınıflandırma algoritmaları tekrar gözden geçirilir gerekli hallerde analiz tekrar yapılır.<sup>67</sup> Sonuçların doğru değerlendirilmesi uzmanlık ve deneyim gerektirmektedir. Sonuçları değerlendirecek araştırmacı çalışmayı ilk kez yapıyorsa alanında uzman kişilerin görüşlerine başvurmalıdır. Aksi halde sınıflandırma tam anlamıyla doğru yapılsa bile sonuçlar yanlış değerlendirilip, çalışmanın tekrar yapılmasına sebep olunabilir.

### 1.3. Metin Madenciliğinin Kullanım Alanları

Veri madenciliğinin özel bir alanı olan metin madenciliği, içinde metin geçen tüm verilerin işlenmesinde kullanılır. Metin madenciliğini veri madenciliğinden ayıran en temel özellik yapılandırılmamış veriyle de çalışmasıdır. Yapılandırılmamış veriye sahip olan tüm alanlar metin madenciliği için bir kaynak olabilmektedir. Başlıca kullanım alanları; bilgi çıkarma, konu izleme, özetleme, kategorizasyon, kümeleme, kavram bağlantısı, bilgi görselleştirme, soru cevaplama, tahminde bulunma olarak sayılabilir.<sup>68,69</sup> Aynı şekilde son zamanlarda metin madenciliğinin en sık kullanım alanları arasında; müşteri ilişkileri yöntemleri, sahtekârlıkların tespiti, bilimsel

<sup>66</sup> Akar Özlem ve Güngör Oğuz, 'Rastgele Orman Algoritması Kullanılarak Çok Bantlı Görüntülerin Sınıflandırılması', *Jeodezi ve Jeoinformasyon Dergisi*, Ankara, 2013, s141 .

<sup>67</sup> Kuzucu.a.g.e., s. 15

<sup>68</sup> Weiguo Fan ve Diğerleri, "Tapping into the Power of Text Mining", *Communications of the ACM*, New York, V. 49, 2006, s.77-80

<sup>69</sup> Jelena Jovanovic.s.6

arařtırmalar, gvenlik-istihbarat, pazar arařtırmaları, biyomedikal alıřmalarını da sıralamak mmkndr.<sup>70</sup>

Kısaca aıklamak gerekirse;<sup>71,72</sup>

Bilgi ıkarma; bilgisayarların yapılandırılmamıř veriden hareketle analiz yapma srecindeki ilk ařaması bilgi ıkarmadır. Bunu yaparken metin iinde nceden tanımılı dizileri arar, desen eřleme adı verilen sre sayesinde, metin iinde yer alan bilgi ıkarımında bulunur. Sistem bunu otomatik yaparken grnt eřleme sayesinde bilgiye eriřmektedir.

Konu izleme; bir kullanıcının profillerinden hareketle o kullanıcının ilgi duyabileceėi diėer konular saptanmaya alıřılır. Gnmzde pek ok web sitesi bunu uygulamaya ynelik aralar kullanıcılara sunmaktadır. rneėin youtube'da izlediėiniz videodan hareketle otomatik olarak size ilgi duyabileceėiniz benzer sınıftaki videolar nerilir. Konu izleme, makine ėrenmesinde denetimsiz ėrenmenin bir unsuru olarak karřımıza ıkmaktadır.

zetleme; uzun bir belgenin, kullanıcının ihtiyaını karřılayıp karřılayamayacaėını anlamak iin zetlenmesi srecidir. Buradaki ama kullanıcının tm metni okumadan sadece bir paragraf sayesinde, bu metnin ihtiyaını karřılayıp karřılayamayacaėını grebilmesidir. zetleme yapılırken dikkat edilmesi gereken husus, metnin ana temasının, btnlėnn korunabilmesidir. Burada temel ama okuyucuya zaman kazandırmaktır. Bilgisayarlar zetleme yaparken kiři isimleri, yer isimleri, tarihler gibi bilgileri belirlemede bařarılı iken, tm metni analiz edip anlamlandırmakta yetersizdirler. Bilgisayarlar, insanlar gibi bir dil yeteneėine sahip olmadıklarından, henz insanlar kadar bařarılı zetleme yapamamaktadır. İnsanlar kadar bařarılı olmasa da kazandırdıėı zaman nedeniyle zetleme zelliėi ok kullanıřlı olabilmektedir.

Kategorizasyon; bilgisayarlar kategorizasyon yaparken, genellikle belgeyi szck paketi (bag of words) olarak deėerlendirir. Kategorizasyon, gerek bilgileri ortaya koymaktan ziyade, sadece kelimelerin frekanslarından hareketle belgenin ana

---

<sup>70</sup> Jelena Jovanovic. s.51

<sup>71</sup> Jelena Jovanovic.

<sup>72</sup> Dolgun ve zdemir.

konularını tanımlamaya çalışır. Yapılan bu işlem sayesinde belge, önceden tanımlanmış belli bir kümeye yerleştirilerek belgenin ana teması ortaya çıkarılmış olur. Terimlerin frekansları hangi kümeye belgenin yerleştirileceğini belirleyen temel unsurdur. Metinde sadece bir kez tekrarlanan terimler metnin kategorizasyonu için önemsizdir.

Kümeleme; amaç benzer belgelerin ortak gruplarda toplanmasıdır. Kategorizasyondan temel farkı kümeler önceden belirlenmiş değildir, işlem anında ortaya konulmaktadır. Önemli bir yararı, belge içinde birden çok alt konu varsa bunların hepsinin görülebiliyor olmasıdır. Bu durum bizim için, işimize yarayan bir belgenin arama sonuçlarında atlanmamasını sağlamaktadır. Kümelemede vektör uzayı mantığıyla vektörler oluşturulur ve belgenin her bir kümeye ne kadar uyduğu bu şekilde belirlenir. Belge uyum sağladığı tüm kümelerde kendine yer bulur. Bu sayede belge kendine sadece tek bir kümede değil de alakalı olan iki veya daha fazla kümede aynı anda yer bulabilir.

Kavram bağlantısı; belgelerdeki ortak kavramların tanımlanmasını sağlar. Bu sayede belki de araştırmacıya, geleneksel arama metotlarıyla bulamayacağı bilgileri bulmasında yardımcı olunabilir. Bilgiyi aramadan ziyade tarama yapılmasını sağlayan bir yöntemdir.

Bilgi görselleştirme; çok geniş metinsel kaynakların belli bir hiyerarşiyle basit bir harita vb. görselliğe kavuşturulmasını sağlar. Kullanıcı bu sayede; çok büyük miktardaki metnin eşleşmelerini ölçeklendirme, bölge haritalarını kurabilme gibi işlemler yapabilir. Bunun sonucunda, kullanıcı geniş bir dokümanı daraltarak ilgilendiği konuya yönelebilir. Bu alanda bilgi görselleştirme son derece yararlıdır. Özellikle sosyal ağ analizinde kullanımı geniş bir yere sahiptir.

Soru cevaplama; metin madenciliğinin kullanım alanlarından biri de verilen bir soruya en iyi cevabın nasıl bulunacağıdır. Soru cevaplama yönteminde birden fazla teknik bir arada kullanılabilir. Örneğin bilgi çıkarma tekniği ile insanları, yerleri, olayları ayıklayabilir, sınıflandırma tekniği ile kim, nerede, ne zaman vb. soru özellikleri sınıflandırabilir. Tüm bu süreçlerden hareketle sistem sorulara en doğru yanıtı vermeye çalışır.



Tahminde bulunma; kullanıcı profili ve geçmiş işlemlerinden hareketle kullanıcıya özel, o kişinin ilgisini çekebileceği düşünülen diğer dokümanlar tahmin edilmeye çalışılır. Sosyal paylaşım siteleri bu uygulamaya çok sık başvurarak kişiye tanıyor olabileceği kişileri önermekte veya ilgi alanına uygun olduğu tahmin edilen reklamlar sık sık karşısına çıkarılmaktadır.

Müşteri ilişkileri yönetimi; tüm müşterilere ait erişilebilecek tüm bilgiler; iletişim bilgileri, kişisel bilgiler, anket verileri, alışveriş hareketleri gibi verilerden hareketle, nitelikli bilgi elde edilip, müşterilere, çapraz satışları tahmin etmek gibi süreçlerde kullanılır. Müşterilerin ihtiyaçları tespit edilip satışın cazip hale getirilmesi için müşteri bazlı kampanyalar düzenlenebilmektedir.

Sahtekârlık tespiti; sağlık, sigorta gibi sektörlerde toplanmış olan büyük hacimdeki verilerden hareketle çeşitli kalıpların ve anormalliklerin saptanarak sahtekârlıkların tespit edilmesinde kullanılır. Özellikle sigorta şirketlerinin çok sık karşılaştıkları sahtekârlıkların önüne geçilebilmesi için modeller oluşturulmaya çalışılmaktadır. Modellerden hareketle hasar kayıtları incelenip sahte işlem yapıp yapılmadığı saptanabilmektedir.

Güvenlik – istihbarat; kişiler, gruplar arasındaki bağlantılar incelenerek, terör olaylarının tahmin edilmesi, bunların engellenmesi için çalışmalar yapılır. Bunun için çok büyük miktardaki suç ve suçlu profili, kriminal verilerden faydalanılır. Bu verilerden hareketle aralarında bağlantı olan veya bağlantılı olma olasılığı olan kişiler tespit edilmeye çalışılır.

Pazar araştırması; yayınlanmış belgeler, makaleler, internet siteleri, basında çıkan haberler, bültenler, dergiler, anket sonuçları pazar etkisini ölçmek için sürekli olarak takip edilir. Satış rakamları, pazar payları, pazarlama için çok büyük veriler oluşturmaktadır. Bu verilerden hareketle pazar ve pazarlama sürecine ilişkin çeşitli tahminlerde bulunulur.

#### **1.4. Metin Madenciliği Alanında Geliştirilmiş Olan Paket Programlar**

Metin madenciliği alanında geliştirilmiş olan pek çok paket program mevcuttur. Fakat henüz bu programlardan herhangi birisi istenilen tüm sonuçları elde etmek için tek başına yeterli değildir. Kullanıcılar yapmak istedikleri işleme göre en uygun

program veya programları tercih etmelidir. Bu programların bir kısmı açık yazılımken bir kısmı ise ticari yazılım olarak piyasada bulunmaktadır. Bu programların nerdeyse tamamı bilgi çıkarma ve kategorizasyon işlemini yaparken, her biri farklı uygulama alanlarına da sahiptir.

En yaygın olarak kullanılan paket programlardan bazıları ve bu programların kimi özellikleri aşağıda sıralanmıştır.<sup>73,74</sup>

Ticari Yazılımlar;

ClearForest (Bilgi çıkarma, kategorizasyon, kavram bağlantısı)

IBM Intelligent Miner Data Mining Suite (Bilgi çıkarma, özetleme, kategorizasyon, kümeleme)

Megaputer (Bilgi çıkarma, özetleme, kategorizasyon, kümeleme, soru cevaplama)

NodeXL Pro (Sosyal ağ analizi, görselleştirme)

ODM (Risk analizi, soru cevaplama, sınıflandırma)

SAS Text Miner (Bilgi çıkarma, kategorizasyon, kavram bağlantısı)

SPSS Text Mining For Clementine (Bilgi çıkarma, kategorizasyon, bilgi görselleştirme)

STATISTICA Text Mining

Vantage Point

WorldStat

Açık Kaynak Yazılımlar;

GATE

---

<sup>73</sup> Seçkin Keziban, 'Metin Madenciliğinde Kullanılan Yöntemlerin Karşılaştırılması: Siyasi Parti Liderlerinin Grup Genel Toplantı Konuşmaları İle Bir Uygulama'. Sakarya Üniversitesi Sosyal Bilimler Enstitüsü, Sakarya, 2011, s.12

<sup>74</sup> Melek Cemile, 'Metin Madenciliği Teknikleri İle Şirketlerin Vizyon İfadelerinin Analizi', Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, İzmir, 2012, s.30

RapidMiner (görselleştirme)

R

Perl (metin işleme, görüntü tanıma)

LingPipe

WEKA (Sınıflandırma, kümeleme, bilgi görselleştirme)

### 1.5. Doğal Dil İşleme (DDİ)

İstatistiksel yöntemler, sadece pratik bir mühendislik yaklaşımı olsaydı eğer, bilimin henüz çözemediği, dilin zorlu problemleri için istatistiğe ilgi daha az olurdu. Oysa dil bilimi araştırmalarında istatistiğin son derece önemli bir yeri vardır.<sup>75</sup> İstatistik bilimi hemen her alanda olduğu gibi dil bilim çalışmalarında da çok geniş kullanım ağına sahiptir. İstatistiğin, dil bilim alanında çalışmalar yaparken, en büyük yardımcısı her zaman olduğu gibi bilgisayarlardır. Bilgisayar programları ve istatistik yardımıyla dile dair çözümler yapılmaya çalışılmaktadır.

Doğal dil işlemenin (DDİ) temel mantığı, insan eliyle üretilmiş olan metinden hareketle bilgisayarların rahatlıkla işlem yapabilmesi, o metinden çeşitli sonuçlar çıkarabilmesidir. Buradaki en büyük zorluk dilin karmaşık yapısıdır. Dilin uçsuz bucaksız yapısına rağmen bilgisayarın son derece sınırlı yetisi vardır ve bu iki durum arasında bir çelişki söz konusudur. Yapılan çalışmaların temel amacı bu çelişkiyi ortadan kaldırarak, bilgisayarın dili en iyi şekilde işleyebilmesine yöneliktir.

Metin incelemesinde ortaya çıkan sorun açıktır; doğal dil insanların birbirleriyle iletişim kurabilmeleri amacıyla geliştirilmiştir, bilgisayarlar ise bu dili anlamaktan son derece uzaktır. İnsanlar dilsel kalıpları kolaylıkla ayırt edebilirken, bilgisayarlar bu kalıpları ayırt edememekte, argoları, kalıplaşmış deyimleri, imla hatalarını aşamamaktadırlar. Buna karşın bilgisayarların avantajı ise insanların işleyemeyeceği çok büyük hacimli metinleri yine çok büyük bir hızla işleyebilmeleridir. Metin

---

<sup>75</sup> Manning Christopher D, Hinrich Schiitze, *Foundations of Statistical Natural Language Processing*, MIT Press, 1999, s.8.

madenciliği insanların üstün yönleriyle bilgisayarların üstün yönlerini bir araya getiren bir teknoloji yaratmaya çalışır.<sup>76</sup>

DDİ'nin insanlara sağlayacağı faydalar; yazılı olan bir dokümanın otomatik olarak tercüme edilmesi, soru cevaplama, otomatik konuşma, komutların bilgisayarca anlaşılması, konuşma, otomatik metin özetleme, gibi başlıklarla özetlenebilir.<sup>77</sup> Tüm bunların yapılabilmesi bilgisayar teknolojisinin hızla gelişmesi ve yapay zekânın hayatımıza girmesiyle mümkün olmuştur. Zira DDİ, yapay zekânın bir alt disiplini olarak düşünülmektedir.<sup>78</sup>

DDİ'nin ilgi alanlarını kısaca özetlememiz gerekirse;<sup>79</sup>

- Yazım yanlışlarının düzeltilmesi; günümüzde hemen tüm metinler bilgisayar ortamında yazılmaktadır. DDİ'nin yapmaya çalıştığı işlemlerden biri de yazılan metinlerdeki hataların tespit edilip düzeltilmesidir. Burada dikkat edilmesi gereken husus dilbilgisi kurallarının bilgisayar sistemine önceden tamamen eksiksiz bir şekilde tanımlanmış olmasıdır. Türkçede çok sık karşılaşılan bir hata “uluslararası” kelimesidir. Bu kelime dilbilimi açısından ulular ve arası kelimesinin bitişik bir şekilde yazılmasını gerektirirken bilgisayar ortamında DDİ bu iki kelimeyi ayrı yazılması gerektiği şeklinde algılayıp bitişik yazılsa bile ayırmaktadır. Bu tür düzeltme esnasında yapılan hataların göz önüne alınması gerekir.
- Bul ve değiştir; metin içinde geçen herhangi bir kelimenin taranıp bulunması veya kelimenin alternatif bir kelimeyle değiştirilmesi amacıyla uygulanan yöntemdir. Bu yöntemle tek bir kelime değiştirilebileceği gibi metinde geçen tüm kelimelerin de değiştirilmesi mümkündür. Örneğin metinde geçen “cevap” kelimesi “yanıt” kelimesiyle değiştirileceği zaman metinde ki tek bir cevap kelimesi de değiştirilebilir veya isteğe bağlı olarak metinde ki tüm cevap kelimeleri yanıt kelimesi ile değiştirilebilir.
- Basılı bir metni okuma; bilgisayar ortamında bulunmayan bir metnin bilgisayara aktarılma işlemidir. Bu metin bir roman, makale olabileceği gibi

<sup>76</sup> Fan ve diğerleri. a.g.e. s.3

<sup>77</sup> Şeker Sadi Evren, Doğal Dil İşleme, YBS Ansiklopedi, Cilt 2, Sayı 4, Aralık 2015, s. 17

<sup>78</sup> Moreno Antonio ve Redondo Teófilo, ‘Text Analytics: The Convergence of Big Data and Artificial Intelligence’, *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol 3, 2016, s. 57 .

<sup>79</sup> Adalı Eşref, ‘Doğal Dil İşleme’, *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5.2 (2012).

doldurulmuş olan optik form da olabilir. Günümüzde özellikle optik form okuma sürecinde çok büyük gelişmeler kaydedilmiştir.

- Metnin özetini çıkarma; kişilerin çok uzun metinleri okuyarak o metnin gerçekten aradığı şeyle ilgili olup olmadığına karar vermeleri bazen çok fazla zaman kaybına neden olabilmektedir. Kaynağın tamamını okumak yerine metnin özetinin çıkarılması DDİ'nin ilgi alanlarında birisidir. Bu işlem yapılırken öncelikle varsa metnin içindekiler kısmı, bu yoksa konu başlıkları, metinde en sık geçen kelimeler taranır. Bu verilerden hareketle metnin neyle ilgili olduğu anlaşılmasına çalışılır.
- Metnin içerdiği bilgiyi çıkarma; metindeki bilgiyi çıkarmak için dil bilgisi kurallarından hareketle tek bir cümlede mi bilgi verilmek istenmiş, yoksa verilmek istenen mesaj metnin tamamına mı yayılmış buna bakılır. Temel amaç; metnin ana konusu dışında kalan konuları ayıklayarak içerdiği ana bilgiye erişmektir.
- Bilgiye erişim; internetin günlük yaşamımızda çok fazla yer edinmesiyle birlikte, internette aradığımız bilgiye kolayca erişebilmek amacıyla geliştirilmiştir. İnternet arama motorunda arama yaparken anahtar kelimeyi yazıp enter tuşuna bastığımızda bilgisayar bize yüzlerce alternatif sayfa seçeneği sunar. Bu sayfalardan hangisi veya hangilerinin işe yarayacağını bulmak ise araştırmacıya düşen görevdir.
- Metni anlama; temel amaç kısa metin veya cümlelerin tam, kesin olarak, anlaşılabilmesidir. Bilgisayara verilen veya söylenen bir cümlenin anlaşılabilir işlem yapılması temel hedeftir. Bilgisayardan arama motorunda bir şey araması istendiğinde aranacak konunun bir veya birkaç kelime ile söylenmesi ve bilgisayarın otomatik olarak bu kelimelere dair net üzerinde arama yapabilmesidir.
- Metin seslendirme; bilgisayara verilmiş olan bir metnin bilgisayar tarafından okunması hedeftir. Sesli gazete, sesli kitap gibi uygulamalarla görme engellilere yardım edilebileceği önemli bir alandır.
- Konuşmayı yazıya dökme; insanların konuştukları şeylerin yazıya dökülmesi, yazıya dökülen metnin anlaşılabilir işleme konulması temel amaçtır.

- Soru yanıtlama; özellikle çağrı merkezlerinde kullanılan, müşterinin isteklerini bilgisayarın anlayıp cevaplayabilmesi amacıyla geliştirilmiş süreçlerdir. Müşteri isteğini bir veya birkaç kelime ile söyler, bilgisayar müşteriye ulaşmak istediği bilgiye dair alternatifleri sunar, müşteri alternatiflerden kendisi için uygun olanı seçer. Bu şekilde müşterinin bir çalışanla görüşmeden bilgisayar yardımıyla istediği bilgiye kolaylıkla erişmesini sağlar.
- Çeviri; herhangi bir dilden başka bir dile sağlıklı ve anlaşılabilir bir şekilde çeviri yapılmasıdır. Basit bir kelimenin çevirisinden, karmaşık bir bilimsel metnin çevirisine kadar geniş bir kapsama alanı vardır. Nihai hedef ise bir kişinin kendi ana dilinde yaptığı konuşmayı karşısında ki kişinin kendi ana dilinde anlayabilmesini sağlayacak başarıda çevirilere imza atmaktır. Günümüzde bilgisayar tarafından yapılan çeviriler büyük başarı gösterebilmektedir.

DDİ'nin yoğun olarak kullanıldığı bir başka alan bu tezin de konusu olan duygu analizidir. Duygu analizi ilerleyen bölümde ayrıntılı bir şekilde ele alınacaktır.

DDİ'nin insanlara sağladığı ve sağlayacağı kolaylıkların yanı sıra üstesinden gelmesi gereken birçok zorlukta vardır. Bu zorlukların önemli bir kısmı yine insan kökenlidir. Konuşma dilinde ki lehçe farklılıkları, kullanılan argo kelimeler, devrik cümleler, eş anlamlı kelimeler, eş sesli kelimeler, imalar, kinayeler, DDİ'nin temel zorluk alanlarını oluşturmaktadır.

Tüm zorluklara rağmen, DDİ'nin pek çok sürecin üstesinden başarıyla geldiği yadsınamaz bir gerçek olarak karşımızda durmaktadır. Nihai amaçsa DDİ'nin tüm bu zorlukların da kendi başına üstesinde gelebilmesini sağlayabilmektir.

## 1.6. Makine Öğrenmesi (MÖ)

Makine öğrenmesi (MÖ) de tıpkı DDİ gibi yapay zekânın bir alt dalıdır. MÖ verilen bir problemi ortamdan edindiği bilgiye göre çözümler.<sup>80</sup> MÖ genellikle bir dizi görevi öğrenen, mantıksal veya ikili işlemlere dayanan, otomatik hesaplama prosedürlerini kapsamaktadır. MÖ insanlar tarafından kolaylıkla anlaşılacak basit

<sup>80</sup> Adalı Eşref, 'Doğal Dil İşleme', *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5.2 (2012).

sınıflandırıcı ifadeler üretmeyi amaçlar. MÖ'nün karar sürecinin arka planında istatistiksel yaklaşımlar vardır.<sup>81</sup> MÖ'nün en büyük avantajı insan müdahalesine ihtiyaç duymaksızın da sınıflandırma yapılabilmesidir.<sup>82</sup> MÖ sayesinde kısa sürede ve etkin bir biçimde istenilen sınıflandırmaların bilgisayar tarafından yapılabilmesidir.

MÖ, geçmiş tecrübelerden ve/veya örnek verilerden hareketle bir sürecin optimize edilebilmesi amacıyla bilgisayarları programlamadır. Bir model oluşturulur ve oluşturulan bu model ileride tahminlerde bulunmak için öngörülerde bulunabilir, bilgi edinmek için açıklamalarda bulunabilir veya eldeki verilerden hareketle süreçlere dair denetimlerde bulunabilir.

MÖ, matematiksel modeller oluşturmak için istatistik teorisini kullanır, çünkü temel görevi bir veri kümesinden hareketle çıkarım yapmaktır. MÖ'de bilgisayarın da iki farklı fonksiyonu vardır; bunlardan birincisi, ihtiyaç duyulan algoritmalar için muazzam sayıdaki verinin depolanıp etkili algoritmalar yaratılması, ikincisi ise model bir kez öğrenildiğinde etkili çözümler üretilmesidir.<sup>83</sup>

MÖ sağladığı pek çok avantaja rağmen bir takım dezavantajları da vardır. Şöyle ki MÖ; metin madenciliği için genellikle çok pahalı bir yöntemdir çünkü eğitim verisi oluşturmak için çok sayıda açıklanmış belgenin hazırlanıp sisteme tanıtılması gerekir. Metinlerde öznitelik alanı çok fazla olabilir, bu da analiz edilecek belgelerdeki çok fazla miktardaki terimlerden tümünün veya bir kısmının seçilmesini zorunlu kılar.<sup>84</sup> MÖ kimi zaman çok uzun bir zaman alabilmektedir.

MÖ üç farklı şekilde işlev görmektedir. Bunlardan ilki denetimli öğrenme, ikincisi denetimsiz öğrenme, üçüncüsü ise yarı-denetimli öğrenmedir.

### 1.6.1. Denetimli Öğrenme

Gözetimli öğrenme ya da kontrollü öğrenme de denen bu yöntemde, sisteme öğrenilmek istenen kavramla ilgili veriler ve bir eğitim kümesi girilir. Eğitim kümesinde her bir örnek için istenen çıktı değerleri de verilir (etiketleme yapılır). Bu

---

<sup>81</sup> D Michie, D J Spiegelhalter ve C C Taylor, *Machine Learning, Neural and Statistical Classification*, Overseas Press, 1994.s.3

<sup>82</sup> Fabrizio Sebastiani, 'Machine Learning in Automated Text Categorization', ACMDL V34, II, New York, 2002, s.2

<sup>83</sup> Alpaydın Ethem, *Introduction to Machine Learning Second Edition.*, The MIT Press, Cambridge, 2010, s.4

<sup>84</sup> Alpaydın, a.g.e..

bilgilerden hareketle girdi ve çıktı arasında bir ilişki oluşturulur. Girdi verilerinin değerleri kullanılarak çıktı değerleri tahmin edilmeye veya öğrenilmeye çalışılır. Sonuçları bilinen veriler üzerinde yapılan sınıflama ile sonuçları bilinmeyen veri kümesine dair tahminler yapılır.<sup>85</sup>

Etiketlenmiş gözlemlerde, etiketler algoritmaya gözlemleri nasıl etiketlemesi gerektiğini öğretir. Örneğin bir mailin spam olduğunu sistemin anlayabilmesi için, içinde “para kazan” ifadesi geçiyorsa sisteme bu maili spam olarak algıla şeklinde yol gösterilmesi gerekir.<sup>86</sup>

Denetimli öğrenmede girdi ve çıktı arasındaki bağıntı bilgisayara dışarıdan veri olarak kullanıcı tarafından sunulur. Bilgisayarın görevi ise bu bilgiler ışığında daha sonra kendine sunulacak verilerden hareketle tahminlerde bulunup, çıkarımlar yapabilmektir.

### 1.6.2. Denetimsiz Öğrenme

Gözetimsiz öğrenme olarak da adlandırılan bu sistemin denetimli öğrenmeden temel farkı, verilerin sisteme girilmesi ancak veriler arasında sebep – sonuç veya girdi – çıktı ilişkisinin işaretlenmemesidir (etiketleme yapılmamasıdır). Veri setinde yer alan ilişkiler bu işaretlemeler yapılmaksızın öğrenilmeye çalışılır. Sosyal ağlarda kullanıcılara sunulan “tanıyor olabileceğiniz kişiler” seçeneği denetimsiz öğrenmenin bir sonucudur.

Sonuç olarak kullanıcı sisteme sadece verileri yükler, algoritmanın kendi kendine keşifler yapıp, görünmeyen örüntüleri öğrenmesi bekler. Buradaki amaç girdideki düzensizlikleri bulmaktır. Giriş alanında belirli kalıpların diğerlerinde daha sık oluşacağı bir yapı vardır ve bu yapı dışına çıkan yapılar görülmeye çalışılacaktır.<sup>87</sup>

Denetimsiz öğrenme ile araştırmacının daha önceden öngöremeyeceği bilgilerin, ilişkilerin de keşfedilmesi mümkündür.

---

<sup>85</sup> Alpaydın Ethem, *Introduction to Machine Learning Second Edition*. The MIT Press Cambridge, Massachusetts London, England, 2010

<sup>86</sup> <http://www.emrealadag.com/makine-ogrenmesi-nedir.html>, 19.01.2017

<sup>87</sup> Alpaydın Ethem , s.11



### 1.6.3. Yarı Denetimli Öğrenme

Bazı durumlarda denetimli öğrenme veya denetimsiz öğrenmenin tek başına kullanılamaması söz konusu olmaktadır. Bu durumlarda yarı denetimli öğrenmenin uygulanması en ideal olandır. Yarı denetimli öğrenme, denetimli öğrenme ile denetimsiz öğrenmenin bir karmasıdır. Yarı denetimli öğrenme az sayıda etiketlenmiş veri ile çok fazla miktarda etiketlenmemiş verinin bir arada yer aldığı öğrenme problemlerinde kullanılır. Yarı denetimli öğrenmede etiketlenmemiş veri miktarı çok büyüktür, buna web siteleri ideal birer örnek oluşturmaktadır. Milyarlarca web sitesi vardır ancak insan müdahalesiyle bu sitelerin tamamını işaretlemek çok pahalı ve imkânsızdır. Bu gibi durumlarda yarı denetimli öğrenme ideal çözüm olarak karşımıza çıkmaktadır.<sup>88</sup> Yarı denetimli makine öğrenmesi çalışmada maliyetler açısından kısıtlı bütçe varsa uygulanabilecek ideal bir yöntem olarak karşımıza çıkmaktadır.

### 1.7. R ile Metin Madenciliği

R yazılımı açık kaynak kodlu; istatistikten makine öğrenmesine, zaman serisi analizinden görselleştirmeye varıncaya kadar pek çok analizi yapmaya olanak tanıyan son derece kullanışlı bir araçtır. R yazılımı, geçmişte kullanılan S dilinin geliştirilmesiyle ortaya çıkarılmıştır.<sup>89</sup> R klasik paket programlardan farklı bir kullanım sürecine sahiptir. R'yi kullanmak için önce R programını daha sonra da ihtiyaç duyulan paketleri indirmek gerekir. Paketler kullanıcılar tarafından geliştirilmiş olan diğer kişilere açık olarak yayınlanan hazır ortamlardır. Kullanıcı kendisi de paket geliştirip ihtiyaç duyduğu analizleri yapabilmekte ve geliştirdiği paket ile diğer kullanıcılara yardımcı olabilmektedir. Bu özelliğiyle R kullanıcıları arasında paylaşımı amaçlayan etkin bir araçtır. Nisan 2018 tarihi itibarıyla R'de bulunan paket sayısı; 12501'dir.<sup>90</sup> R yazılımı <https://cran.r-project.org/mirrors.html> adresinden istenilen bağlantılar sayesinde indirilebilmektedir. R yüklenirken kimi paketler öngörülen olarak yüklenmektedir fakat öngörülen paket sayısı çok sınırlıdır. Bu yüzden kullanıcı ile ihtiyaç duyacağı paketleri analiz yapmadan önce indirmek zorundadır. Paket indirmek için alternatif yöntemler mevcuttur, kullanıcı kullanım alışkanlıklarına göre alternatiflerden birini seçebilmektedir. R de çok sayıda paket olduğu için bu paketlerin

<sup>88</sup> Chapelle Olivier, Weston Jason ve Scholkopf Bernhard, 'Cluster Kernels for Semi-Supervised Learning'. NIPS'02 Proceedings of the 15th International Conference on Neural Information Processing Systems, 2002, 8

<sup>89</sup> <https://www.r-project.org/about.html>, erişim tarihi; 04.04.2018

<sup>90</sup> <https://cran.r-project.org/web/packages/>, erişim tarihi; 04.04.2018

tamamının aynı sürücüde olması indirme yükünü arttıracaktır. Bu yüzden farklı konumlarda bulunan sürücülerden istifade edilerek paketler farklı konumlarda bulunmaktadır.<sup>91</sup>



---

<sup>91</sup> <https://www.quora.com/What-is-the-CRAN-mirror>, erişim tarihi; 05.04.2018

## İKİNCİ BÖLÜM

### DUYGU ANALİZİ

#### 2. DUYGU ANALİZİ

Türk Dil Kurumu (TDK) duyguyu “belirli nesne, olay veya bireylerin insanın iç dünyasında uyandırdığı izlenim” şeklinde tanımlamaktadır.<sup>92</sup> Tanımdan da anlaşılacağı gibi duygu kavramı içinde pozitif öğelerle beraber negatif öğeleri de barındırmaktadır. Yani duygular; aşk, sevgi, hoşgörü, sadakat gibi pozitif olabileceği gibi nefret, öfke, tiksinti gibi olumsuzlukta barındırabilir.

Pozitif duygular insanların yaşam kalitesini yükseltip, olumlu davranışta bulunmaları için motive edici bir rol oynarken, negatif duygular bunun tam tersine, kişilerin yaşam kalitesini düşürür, entelektüel zekâlarını bloke eder ve sürekliliği durumunda pek çok soruna yol açar. Bu yüzden negatif duygu durumunun sürekli bir hal alması istenmeyen ve bir an önce mücadele edilmesi gereken bir haldir.

Duygu analizi üç farklı düzeyde yapılabilir. Bunlar doküman (belge) düzeyinde, cümle düzeyinde ve öznitelik düzeyinde diye adlandırılabilir.<sup>93</sup> Doküman düzeyinde yapılan duygu analizinde; doküman, bir bütün olarak ele alınır ve bu dokümanın olumlu mu yoksa olumsuz mu duygu taşıdığı araştırılır. Cümle düzeyinde ise her bir cümle tek tek ele alınarak cümlenin taşıdığı duygu analiz edilmeye çalışılır. Cümle düzeyinde yapılan çalışma esasında doküman düzeyinde yapılan çalışmanın küçük bir boyuta indirgenmesinden başka bir şey değildir. Gerek doküman düzeyinde gerekse cümle düzeyinde yapılan duygu analizlerinde hedeflenen bilgiye erişilemeyebilir. Bu yüzden, öznitelik temelli duygu analizine ihtiyaç duyulur. Öznitelik düzeyli duygu analizinde, duygular varlıkların özel yönlerine göre sınıflandırılabilir. Bir kişi bir olay karşısında duygularını ifade ederken olayın kimi yönleri hakkında pozitif, kimi yönleri hakkında negatif duygulara sahip olabilir. Örneğin bir film hakkında yorum yapan kişi; “filmi beğendim ancak başrol oyuncusunun performansı yeterli değildi” diyebilir. Bu cümle

<sup>92</sup>[http://www.tdk.gov.tr/index.php?option=com\\_gts&arama=gts&guid=TDK.GTS.5891b60fb4f438.66021402,01.02.2017](http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.5891b60fb4f438.66021402,01.02.2017)

<sup>93</sup> Chapelle Olivier, Weston Jason ve Scholkopf Bernhard, ‘Cluster Kernels for Semi-Supervised Learning’. NIPS’02 Proceedings of the 15th International Conference on Neural Information Processing Systems, 2002, 8

içinde hem olumlu hem de olumsuz duyguları barındırmaktadır. Cümle düzeyli analizde bu cümleyi analiz etmek zorken öznitelik düzeyli analizde olumlu duygu ayrı, olumsuz duygu ayrı ele alınacağından dolayı çok daha sağlıklı sonuç verebilmektedir.

İnsanlar yaşamları boyunca tüm kararlarını mantıklarıyla almazlar. Duyguların da karar verme sürecinde çok büyük bir etken olduğu bilinmektedir. Günümüzde “duygu”ya olan ilgi gittikçe artmakta ve bu konu üzerine pek çok araştırma yapılmaktadır. Duygu analizi de bu süreçte karşımıza çıkan, gün geçtikçe kullanım alanı yaygınlaşan bir alandır. Bu alana dair yapılmış olan çalışmaların çok büyük bir kısmı bilgisayar tabanlı olsa da temelinde metin verisi olan bu alana istatistiğin uzak kalması düşünülemez. Bu tezin temel amacı da duygu analizi konusunda, literatürde istatistik açısından var olan bir eksikliğin giderilmesine katkıda bulunmaktır.

## **2.1. Duygunun Önemi**

Duygunun önemi ve duygu analizi yapmaya neden ihtiyaç duyulduğunun daha iyi kavranabilmesi için duyguya dair bazı kavramlar aşağıda açıklanacaktır.

### **2.1.1. Kansei Mühendisliği (KM)**

Kansei mühendisliği (KM) literatürde kimi zaman duygu mühendisliği olarak ele alınmakta, kimi zamanda ayrı bir kavram olarak karşımıza çıkmaktadır.

Kansei kelimesi ilk olarak M. Nadamachi'nin “Emotional Engineering (Duygusal Mühendislik” adlı eserinde ürün geliştirme yöntemi olarak sunulmuştur.<sup>94</sup> Kansei kelimesinin Japonca'da; duyu organlarının hissetme becerisi, bilinç ve irade ile kontrol edilen duygusal istek gibi anlamları vardır.<sup>95</sup> Kansei mühendisliği kavramı ise ilk olarak, 1986 yılında Japon otomobil şirketi olan Mazda Motor A.Ş. nin başkanı, Kenichi Yamamoto tarafında Michigan Üniversitesi'nde yapmış olduğu konuşmada kullanılmıştır. Bu konuşmada Yamamoto, Kansei kelimesini, duyarlılık, duyu, his kelimelerini kullanarak İngilizceye tercüme etmek istemiş ancak bu kelimelerin tamamının Kansei terimini açıklamakta yetersiz kaldığını vurgulamıştır.<sup>96</sup> KM içinde,

<sup>94</sup> Olivier Chapelle, Jason Weston and Bernhard Scholkopf, ‘Cluster Kernels for Semi-Supervised Learning’.

<sup>95</sup> Fiğlali Nilgün ve diğerleri, ‘Kansei Mühendisliği Ve Uygulamaları’, Doğuş Üniversitesi Dergisi, İstanbul, Sayı 5, 2002, 85-96.

<sup>96</sup> Shinya Nagasawa, ‘Present State of Kansei Engineering in Japan’, in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2004, s. 333-338

his, duygu, duygunun dışı vurumu (duygulanım), izlenim, algı gibi pek çok kavramı bir arada barındırmaktadır.

KM, son derece geniş bir kullanım alanı bulmuş, Japon Eğitim Bakanlığı tarafından desteklenmiş, bu alanda çeşitli sempozyumlar düzenlenmiştir. Kansei Mühendisliğinde temel amaç tasarım yapılırken duyguların da işin içine katılmasıdır. Bu sayede insanlar için renkleri ve şekli ile daha çekici gelen, kullanım özellikleri ile daha ergonomik ürünler üretilmesi amaçlanmıştır. Buradaki temel amaç ürünün sadece beyne değil duygulara da hitap etmesini sağlamaktır.

Sevim'e göre,<sup>97</sup> duyguların ekonomik faaliyetlerdeki yeri gittikçe artmakta ve artık "duygu mühendisliği" çağına girilmektedir. Bu süreçte tasarımdan pazarlamaya varıncaya kadar firmalar artık duyguları göz önünde bulundurmalı bu şekilde insanlara ulaşmanın yollarını aramalıdır.

Bu tezi doğrulayan temel unsurlardan birisi de Mazda firması tarafından üretilmiş olan, Mazda Miata (MX 5) isimli çift kişilik spor otomobildir. Bu otomobil dünya da en çok satan çift kişilik spor otomobil olma özelliğine sahiptir ve KM teknolojisi ile üretilmiş olan ilk otomobillerden biridir. Miata (MX5), 2000 yılı mayıs ayında 531.890 adet üretimiyle Guinness Dünya Rekorları Kitabına girmiş ve kendi rekorunu her seferinde yenileyerek alanında 900.000 adet üretime ulaşmış ilk otomobil unvanını almıştır.<sup>98</sup>

Özellikle Japon otomobil üreticilerinin başarılı KM kullanımları dikkat çekmektedir. 1990'lardan bu yana başta Mazda olmak üzere diğer Japon üreticiler olan Mitsubishi, Honda ve Toyota KM'yi kullanmış, Mazda ve Mitsubishi'nin ürettiği ürünler dünya piyasasında en çok satanlar listesine girmiştir.<sup>99</sup>

### 2.1.2. Duygusal Zekâ

Tarih boyunca insanlar zekâyâ ilgi duymuş ve bir şekilde zekâ seviyelerini ölçmeye çalışmıştır. Bu amaçla yapılan çalışmalar çok uzun süre sadece mantıksal zekânın ölçülmesine dayanmıştır. IQ diye kısaltılan bu zekânın ölçülebilmesi için 1912

<sup>97</sup> Bilgen Aydın Sevim, 'Postkapitalist Dönemde Küresel Markaların Yeni Pazarlama Stratejisi: Kitlelerin Yeni Yaşam Koçu Olarak Coca Cola', Uluslararası İktisadi ve İdari İncelemeler Dergisi, sayı 15, 2015 s.141.

<sup>98</sup> Ataalp Burak, Kansei Mühendisliği, <http://www.endustrimuhendisligim.com/kansei-muhendisligi>, 06.02.2017

<sup>99</sup> Mitsuo Nagamachi, 'Kansei Engineering: A New Ergonomic Consumer-Oriented Technology for Product Development', *International Journal of Industrial Ergonomics*, 1995, s.8.

yılında Alfred Binet ve Theodore Simon tarafından geliştirilmiş olan bir zekâ testi kullanılmıştır. IQ'nun aslında kişilerin sayısal düşünme yeteneğini ölçtüğü, entelektüel zekâyla ilgili olduğu bilinirken, bu test insanların ne kadar mutlu veya ne kadar başarılı olduğunu göstermediği bir gerçektir.

Kişilerin ne kadar mutlu veya başarılı olduğunu, yani duygusal zekâsını ölçebilmek için “duygusal zekâ” (emotional quotient – EQ) adı verilen test geliştirilmiştir. Duygusal zekâ bir kişinin, kendisi ve diğer kişilerin, hislerini ve duygularını izleyebilme, aralarında ayırım yapma ve bu bilgileri kişilerin düşüncelerini ve davranışlarını yönlendirmek için kullanabilme olarak tanımlanmıştır.<sup>100</sup> 1990 yılında Mayer ve arkadaşları tarafından ortaya atılan “duygusal zekâ” kavramı, 90'ların sonunda modern psikolojinin en çok kullandığı terim haline gelmiştir. En başarılı insanlar her zaman en zeki insanlar değildir. Bu süreç bize günümüzde duyguya verilen önem hakkında da ipuçları vermektedir. Günümüzde sadece mantık sürecinin değil duyguların da yaşamda analiz edilmeye muhtaç olduğunu göstermektedir.

## 2.2. Literatür Özeti

Literatür incelendiğinde duygu analizine dair yapılan çalışmaların çok büyük bir kısmı İngilizce üzerinedir. İngilizce dil yapısı nedeniyle analizin çok kolay olduğu ve web ortamında kullanıcılara hazır sunulan kelime setleri nedeniyle tercih edilmektedir. Diğer dillerde analiz yaparken dilin yapısı başlıca sorun olmakla birlikte, webte kullanıcılar için hazır bir halde kelimelerin duygu yüklerini gösteren veri tabanı sunulmamaktadır. Bu tezin amaçlarından bir tanesi de Türkçe pozitif, negatif ve nötr duygu yüklerini gösteren bir veri tabanını literatüre kazandırarak, bu alanda bundan yapılacak çalışmalara imkan tanımaktır. Büyük veri kavramı ile birlikte artık verinin bol olduğu alanlar araştırmacıları cezbeder hale gelmiştir. Sosyal medya bu alanda en ideal veri kaynaklarından biridir çünkü sosyal medya da hem veri hacmi çok büyüktür hem de veri miktarı her geçen saniye artmaktadır. Sosyal medya araçları arasında ise Twitter veri kullanımını açısından son derece uygundur. Twitter API sayesinde sınırlı da olsa tüm kullanıcılarla verilerini paylaşmaktadır ve Twitter'daki 280 karakter (2017 yılının son çeyreğine kadar 140 karakterdi) kısıtı araştırmacılar için ideal bir veri ortamı sağlamaktadır.

---

<sup>100</sup> Mayer John D., DiPaolo Maria ve Salovey Peter, ‘Perceiving Affective Content in Ambiguous Visual Stimuli: A Component of Emotional Intelligence’, *Journal of Personality Assessment*, 54, 1990, s. 772.

Duygu analizi temelde sözcük bazlı ve makine öğrenmesi ile yapılabilmektedir. MÖ ile yapılan çalışmalarda veri seti eğitim ve test olmak üzere ikiye bölünerek eğitim seti ile veriler öğretilir, model kurulur, test seti ile de modelin başarısı doğrulanılır. Bu bölümde, MÖ ile yapılan çalışmaların bir kısmı ele alınarak bulunan sonuçlar özetlenecektir.

Neethu ve Rajasree tarafından 2013 yılında yapılan çalışmada<sup>101</sup>, Twitter'dan elde edilen veriler MÖ teknikleri ile sınıflandırılmış, bunun yanı sıra MÖ teknikleri ile sözlük tabanlı teknik birleştirilerek toplamda dört farklı algoritmanın performansı test edilmiştir. Çalışma sonucunda kullanılan algoritmalarından SVM, NB ve ME ve bu üçünün bileşimi olan Ensemble yöntemi neredeyse aynı ölçüde performansı göstermiştir.

Pak ve Paroubek yaptıkları çalışmada<sup>102</sup> Twitter API kullanarak tweetleri otomatik olarak toplayıp, yine otomatik olarak bu tweetlerde kullanılan duygu ifadelerini açıklayan bir korpus geliştirdiler. Bu korpus sayesinde özellikle N-gram ve POS etiketlerini kullanan multinominal Navie Bayes sınıflandırıcı ile bir duygu analizi yaptılar. Bu yöntemde, belgelerin olumlu (pozitif), olumsuz (negatif) veya tarafsız (nötr) duyguya sahip oldukları belirlenmiştir.

Pang ve diğerleri yaptıkları çalışmada unigram, bigram, POS etiketi ve bunların kombinasyonlarını kullanarak denetimli makine öğrenmesi algoritmalarıyla sınıflandırma yapmışlardır. Veri seti olarak, duygu analizi yapılırken çok sık olduğu gibi, IMDb'den aldıkları film yorumlarını kullanmışlardır. Pang ve diğerleri, yorumları pozitif ve negatif olarak işaretlemişlerdir. Çalışma sonunda duygu sınıflandırmanın geleneksel konu sınıflandırmadan çok daha zor olduğunu sonucuna varmışlardır. En iyi sınıflandırma performansını ise destek vektör makinesi algoritması sağlamıştır.<sup>103</sup>

Ghag ve Shah 2013 yılında yaptıkları çalışmada, duygu analizi için kullanılan tekniklerin karşılaştırmasını yapmışlardır. Çalışmada çok dilli bir yaklaşım üzerinde

---

<sup>101</sup> Neethu M S And Rajesree R, 'Sentiment Analysis in Twitter Using Machine Learning Techniques', in *4th ICCNT* (Tiruchengode, India, 2013).

<sup>102</sup> Neethu M S And Rajesree R, 'Sentiment Analysis in Twitter Using Machine Learning Techniques', in *4th ICCNT* (Tiruchengode, India, 2013).

<sup>103</sup> Pang Bo, Lee Lillian and Vaithyanathan Shivakumar, 'Thumbs up? Sentiment Classification Using Machine Learning Techniques' (Philadelphia: Association for Computational Linguistics, 2002), pp. 79–86.

durulmuş ve mevcut hiçbir tekniğin dilden bağımsız olmadığını genelleştirilmiş bir duygu analizörü olmadığını ortaya koymuşlardır.<sup>104</sup>

Çelikler ve diğerleri yaptıkları çalışmada müşterilerin atmış oldukları tweetleri polar ve polar olmayan diye ikiye ayırmıştır. Polar gruba pozitif ve negatif duygu içeren tweetleri, polar olmayan gruba diğer tweetleri atamışlardır. Çalışma sonucunda günlük insan konuşmaları sırasında duygusal ifadeyi çıkarabilmek için çok geniş bir kutupsal sözlüğe ihtiyaç olmadığını ortaya koymuşlardır.<sup>105</sup>

Wang ve diğerleri yaptıkları çalışmada tweetleri 7 farklı duygu sınıfına göre ayırmıştır. Bu duygular, sevinç, sevgi, şükür, sürpriz gibi olumlu duygular ve üzüntü, öfke, korku gibi olumsuz duygulardır. Duygu kelimeleri çeşitli çekimleriyle genişletilmiştir, örneğin, sürpriz kelimesi “surprise” surprising ve surprised olarak çoğaltılmıştır ve üçü de aynı kategoride ele alınmıştır. Bu 7 duygu dışında kalan alanlar ve duygularsa belirsizliği ortadan kaldırmak için göz ardı edilmiş ve listeden çıkartılmıştır. Ortaya çıkan tweet veri setinin %93 ü söz konusu 7 duygudan birine karşılık gelmiştir. Veri seti n-gram, part of speech ve WordNet gibi önceden tanımlanmış birkaç sözcük kaynağının farklı kombinasyonları ile eğitilmiştir. Çalışmanın temel amacı farklı boyuttaki veri setleri için sınıflandırma performansını ölçmektir. Eğitim verilerinin büyüklüğü 1.000 den 2 milyona çıkınca %22’lik bir doğruluk kazancı elde edilmiştir.<sup>106</sup>

Literatürde Türkçe yapılan duygu analizi çalışmaları incelendiğinde son yıllarda bu alanda çok fazla çalışma yapılmaya başlandığı görülüyor. Çetin ve Amasyalı yaptıkları çalışmada çeşitli duygu analizi yöntemleri için WEKA kütüphanesinde bulunan algoritmaları kullanmışlardır. Telekom sektöründe faaliyet gösteren iki şirkete ait 6000 adet tweet çekilmiş, elle işaretleme yapılarak tweetler pozitif, negatif ve nötr

---

<sup>104</sup> Pang Bo, Lee Lillian and Vaithyanathan Shivakumar, ‘Thumbs up? Sentiment Classification Using Machine Learning Techniques’ (Philadelphia: Association for Computational Linguistics, 2002), pp. 79–86.

<sup>105</sup> Celikyilmaz Asli, Dilek Hakkani-Tür ve Junlan Feng, ‘Probabilistic Model-Based Sentiment Analysis of Twitter Messages’, in *2010 IEEE Workshop on Spoken Language Technology, SLT 2010 - Proceedings*, 2010.

<sup>106</sup> Wang Wenbo ve diğerleri, ‘Harnessing Twitter “Big Data” for Automatic Emotion Identification’, in *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*



olarak işaretlenmiştir. Yapılan çalışma sonucunda en iyi sınıflama algoritmasını SMO'nun verdiğini tespit etmişlerdir.<sup>107</sup>

Sevindi yaptığı çalışmada Türkçe film yorumları üzerine duygu analizi yapmış sözcük tabanlı ve makine öğrenmesi tekniklerini kıyaslamıştır. Makine öğrenmesi tekniklerinin daha başarılı sonuçlar ürettiğini tespit etmiştir.<sup>108</sup>

Akgül ve diğerleri ülkemizde sosyal medya alanında veri madenciliği yapabilen yazılım sayısının yetersiz olduğu öngörüsüyle Twitter verileri üzerinde otomatik duygu analizi yapabilmek amacıyla “Duygusal Twitter” adını verdikleri yazılımı geliştirmeye çalışmıştır. Geliştirilen yazılım belli bir anahtar kelime girilerek tweet çekilmesi ve bunların otomatik olarak etiketlenmesi amacını gütmektedir. Yazılım n gram ve sözlük tabanlı yaklaşımlarda %69 ve %70 oranında başarı elde etmiştir.<sup>109</sup>

Seçim sonuçlarını tahmin etme tüm toplumlarda son derece popüler ve ilgi uyandıran bir süreçtir. Son yıllarda geleneksel tahmin yöntemlerinin yanı sıra sosyal medya verileri de seçim sonuçlarının tahmini amacıyla kullanılmaktadır. Bu amaçla yapılan çalışmaların bir kısmı politik duygu analizi kavramının ortaya çıkmasına da neden olmuştur. 2009 yılında Almanya’da yapılan seçim sonuçlarını tahmin edebilmek için Tumasjan ve arkadaşları Twitter verilerini kullanmışlardır. Bu amaçla seçime giren altı partinin isimlerini geçtiği yüzbenden fazla tweeti ele almış ve tweet/etiket te partinin isminin geçme sayısı ile seçimleri kazanma olasılıklarının doğru orantılı olduğunu ortaya koymuşlardır.<sup>110</sup>

Singapurda yapılan cumhurbaşkanlığı seçimini tahmin etmek amacıyla Choy ve arkadaşları tweeter verileri üzerinden duygu analizi yapmışlar. Çalışma sonucunda iki aday arasında küçük oy farkı olacağını tahmin edebilmiş fakat galip gelecek adayı tahmin edememiştir. Nüfusun önemli bir kısmının Twitter kullanmamasına karşın

---

<sup>107</sup> Çetin Mahmut ve Amasyalı Fatih, ‘Eğitici ve Geleneksel Terim Ağırlıklandırma Yöntemleriyle Duygu Analizi’. 21. Sinyal İşleme ve İletişim Uygulamaları, IEEE 2013

<sup>108</sup> Çetin Mahmut ve Amasyalı Fatih, ‘Eğitici ve Geleneksel Terim Ağırlıklandırma Yöntemleriyle Duygu Analizi’.

<sup>109</sup> Akgül Eyüp Sercan, Ertano Caner ve Diri Banu, ‘Twitter Verileri İle Duygu Analizi’, *Pamukkale Univ Muh Bilim Derg*, 22.2 (2016), 106–10.

<sup>110</sup> Tumasjan Andranik ve diğerleri, ‘Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment’, in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010.

yapılan bu tahmin Twitter'dan elde edilen bilgilerin siyasi manzara hakkında oldukça doğru sonuçları yansıttığını ortaya koymuştur.<sup>111</sup>

### 2.3. Duygu Analizi Yöntemleri

İnsanların yaşamlarında duygu, her zaman mantık kadar önemli bir yer tutsa da günümüzde özellikle sosyal medyanın hayatımızda çok büyük bir yer edinmesiyle birlikte duyguların analiz edilmesi ihtiyacı daha çok karşımıza çıkmaktadır. Kişiler her türlü olay, durum, kurum vb. hakkında sürekli duygularını çeşitli mecralarda ifade etmekte ve bu ifadelerden hareketle, karar vericiler etkin kararlar alabilmek için çalışmalar yürütmektedir. Duygu kişilerin bir konu hakkında ne hissettikleri, bir şeye karşı tutumları veya görüşlerini ifade ederken, duygu analizi bu duyguları işlemeye çalışır.

Özellikle mikroblog alanlarının hızla yayılmış ve önemli bir iletişim aracı haline gelmiş olması duygu analizinin popülaritesini de arttırmıştır. Duygu analizi temelde, herhangi bir metinde çeşitli yöntemlerle kişilerin veya kitlelerin duygularını analiz etmeye çalışan yöntemler bütünüdür. İnsanlar günlük yaşamda; hangi telefonu alacakları, tatilde nereye gidecekleri, hangi partiye oy verecekleri, hangi filme gidecekleri, hangi otomobili tercih edecekleri, hangi semtte oturacakları gibi pek çok konuda karar vermek zorundadır. Karar verirken de en önemli yardımcılarını daha önce aynı konuda karar vermiş olan kişilerin duygu ve düşünceleridir. Diğer insanların duygu ve düşüncelerini öğrenmenin en pratik yolu sosyal medya yardımı ile yapılan paylaşımların analiz edilmesidir. Duygu analizi bu süreçte bize yardımcı olabilecek önemli bir araçtır. Örneğin bir ürün veya hizmeti hem satın alacak birey hem de satışı yapacak kişi veya kuruluş için karar sürecinde tüm yorumları okuyup incelemek bazen çok uzun bir zaman gerektiren bazense imkansız olan bir durumdur. Duygu analizi sayesinde çok büyük miktardaki veriler otomatik olarak analiz edilip kişilerin hedeflerine ulaşabilmeleri için yardımcı olunabilmektedir.

Duygu analizinde temel amaç bir metindeki duyguyu ortaya çıkarabilmektir. Duygular karşımıza olumlu, olumsuz veya nötr olarak çıkabilmektedir. Bir konu

---

<sup>111</sup> Choy Murphy ve diğerleri, 'A Sentiment Analysis of Singapore Presidential Election 2011 Using Twitter Data with Census Correction', 2011., s. 1093-1113.

hakkında ki ifadeler pozitif özellikler taşıyorsa metin olumlu, negatif özellikler taşıyorsa olumsuz, hiçbir duygu ifadesine yer vermiyorsa nötrdür.

Görünüşte basit gibi duran duygu analizi aslında son derece meşakkatli bir süreçtir. Metinde yer alan kelimelerin tamamı duygu ifade etmemektedir. Öncelikle duygu ifade eden kelimelerin ayırımına gidilmelidir. Eldeki metin Twitter verisi ise araştırmacının işi kısmen daha kolaydır çünkü 140 karakter arasında duygu ifade eden kelimeyi arayacaktır. Ancak bir makale veya bir paragrafsa araştırmacının ilgilendiği metin, duygu ifade eden kelimeyi veya kelimeleri bulup ayıklamak çok daha fazla zaman alacaktır. Ayrıca unutmamak gerekir ki araştırmacı asla tek bir paragraf veya tek bir tweetle ilgilenmemektedir. Bazen bu sayı binleri bulabilmektedir.

Duygu analizinin temel amacı metinlerde yer alan bilgilere erişip bu bilgileri işleyerek duyguları ortaya çıkarmaktır. Bunu yaparken de alternatif yöntemler kullanılabilir.

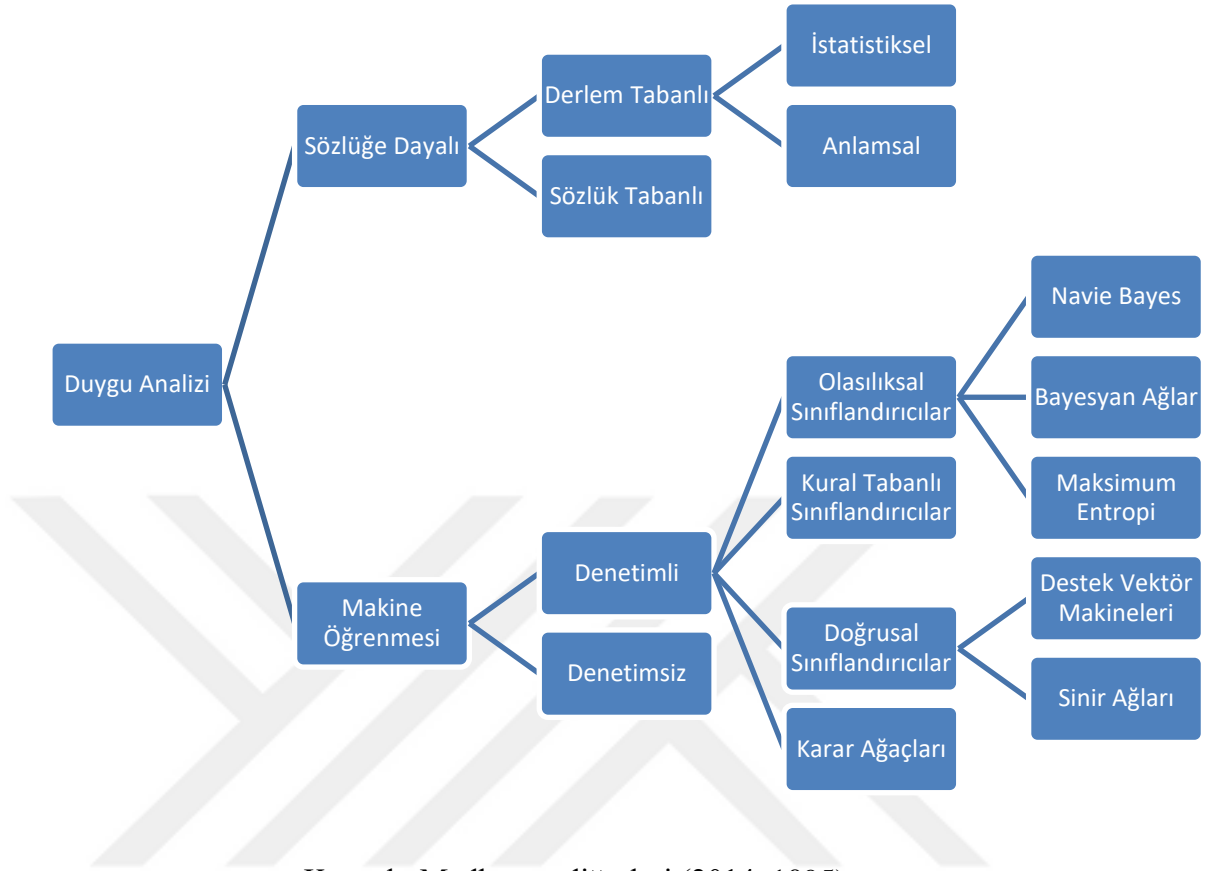
Duygu analizinin ham maddesi kullanılan dildir. Bu yüzden araştırmacının analizi yaptığı dile hakim olması analizin başarısı için son derece önemlidir. Literatürde özellikle İngilizce çalışmaların fazlalığı göze çarpmaktadır, bundaki en temel neden dilin yapısının basit olmasıdır. Türkçe gibi sondan eklemeli ve karmaşık yapıya sahip dillerde ise duygu analizi yapmak çok daha zordur. Türkçede bir kök; aldığı yapım ve çekim ekleri ile çok farklı anlamlara bürünebilmektedir. Duygu analizi bilgisayarlar sayesinde yapılmaktadır ve Türkçe esnek yapısı nedeniyle insanlar tarafından bile kimi zaman anlaşılammaktadır. Şöyle ki iğneleme veya ironi bilgisayar tarafından anlaşılması imkânsız olan süreçlerdir.

Duygu analizi yapılırken alternatif pek çok sınıflandırma teknikleri kullanılabilir. Bu teknikler Şekil 2.1 de gösterilmiştir.<sup>112</sup>

---

<sup>112</sup> Choy Murphy ve diğerleri, 'A Sentiment Analysis of Singapore Presidential Election 2011 Using Twitter Data with Census Correction', 2011., s. 1093-1113

**Şekil 3.1.** Duygu Analizi Sınıflandırma Teknikleri



Kaynak: Medhat ve diğerleri (2014, 1095)

Duygu analizi sözcüğe dayalı ve makine öğrenmesi teknikleriyle olmak üzere iki ana kategoriye ayrılmaktadır.

### 2.3.1. Sözlüğe Dayalı Yöntemler

Sözlüğe dayalı yöntemler, yarı denetimli yöntemlerdir. Bu yöntemler kendi içinde ikiye ayrılır. Bunlar; sözlük tabanlı ve derlem tabanlı olarak adlandırılırlar. Sözlüğe dayalı duygu analizinde temel mantık, duygu ifade eden anlamsal bir sözlük veri tabanı kullanılarak duygu analizi yapılmasıdır.<sup>113</sup> Bunun için hazır veritabanları kullanılabileceği gibi, kişinin kendisi tarafından oluşturulan bir derlem (corpus - külliyat) de kullanılabilir.

Sözlüğe dayalı yöntemde metinlerde yer alan ve duygu ifade eden kelimeler, önceden belirlenmiş duygu skoru ile değerlendirilirler. Sadece kelimelere değil görsel

<sup>113</sup> Kasaba Evrim ve Yıldıztepe Engin, 'Destek Vektör Makinesi Yöntemi İle Bir Duygu Çözümlemesi'. Akademik Bilişim 16, 2016, Aydın, s.2

karakter anlamına gelen emoji'lere de skorlar verilir. Bu duygu skorları -5 ile +5 arasında değişmektedir ve cümlenin veya metnin pozitif duygu mu yoksa negatif duygu mu içerdiği toplam skor değerine göre belirlenir. Cümlede veya metinde yer alan her kelimenin skor değeri yazılıp toplanır, ortaya çıkan değer 0 dan büyükse cümle pozitif duyguya sahip, 0 dan küçükse cümle negatif bir duygu içeriyor demektir. Toplam skoru 0 olan cümle ise nötr kabul edilir. Bu cümlelerin duygu ifadesi içermediği söylenir.

Sözlüğe dayalı duygu analizi yöntemlerinden birincisi sözlük tabanlı yöntemidir. Bu yöntemde duygu belirten kelimeler manuel olarak toplanır, belli bir veri tabanında arama yapılarak değerlendirilir.<sup>114</sup>

Sözlük tabanlı duygu analizi, özellikle Türkçe gibi sondan eklemeli bir dil için çok büyük bir önışleme aşaması gerektirmektedir. İngilizce gibi kelimenin çok fazla ek almadığı dillerde ise sözlük tabanlı duygu analiziyle çalışmak daha kolaydır.

Sözlüğe dayalı yöntem emek yoğun olduğu için terk edilmiş yerine derlem tabanlı yöntemler geliştirilmiştir. Derlem kelime anlamı itibariyle; büyük miktardaki metin koleksiyonlarının elektronik biçimde bilgisayar ortamına tutulmasıdır<sup>115</sup>. Derlemin temelinde içeriğe özgü sözcüklerin bir arada bulunması mantığı vardır. Derlem tabanlı yöntemde; içeriğe yönelik sözcükler bir araya getirilerek çok büyük bir derlem (korpus) oluşturulur, metinde yer alan, duygu ifade eden kelimeler bu derlem içindeki listede yer alan kalıplarla karşılaştırılır. Derlem tabanlı yöntemler istatistiksel veya anlamsal olabilirler.

İstatistiksel yöntemde; bir derlemde yer alan sıfatları bulmak için istatistikten yararlanılır. Dizindeki tüm belgelerin, web üzerinde sözlük yapımı için, derlem olarak kullanmak mümkündür. Oluşturulan derlem yeteri kadar büyük değilse aranan kelime derlem içinde bulunmayabilir. Bu aşamada istatistik devreye girer. Aranan kelimenin frekans kutupsallığına bakılır. Kelimenin pozitif kutupta mı yoksa negatif kutupta mı olduğu o kelime ile birlikte kullanılan kelimelerin frekansları ile tespit edilebilir. Şöyle ki sözcük, pozitif metinler arasında daha sık ortaya çıkarsa kutupsallığı pozitif olur, aynı şekilde negatif metinler arasında daha sık yer alıyorsa kutupsallığı negatiftir. Frekanslar

---

<sup>114</sup> Medhat, Hassan ve Korashy.s. 1102

<sup>115</sup> Tahiroğlu B. Tahir, 'Derlem Dil Bilimi ve Güncel Gelişmeler', Uluslararası Türk Dili Kurultayları Bildirileri, TDK Yayınları, 2008, s.4043

eşit olursa o zaman kelime için nötrdür denir. Benzer görüş bildiren kelimeler sık sık birlikte derlemde yer alırlar. Bu yüzden eğer iki kelime sıklıkla bir arada yer alıyorsa muhtemelen aynı duyguya sahip olacaklardır. Bu yüzden ki bilinmeyen bir kelimenin negatif mi yoksa pozitif mi olduğu birlikte yer aldığı kelimenin yer aldığı kutup hesaplanarak ta bulunabilir.<sup>116</sup>

Anlamsal yöntemde; anlamsal olarak yakın olan kelimelerin benzer duyguları ifade edeceği ilkesine dayanır.

### **2.3.2. Makine Öğrenmesine Dayalı Yöntemler**

Makine öğrenmesine dayalı yöntemler denetimli ve denetimsiz olmak üzere ikiye ayrılırlar.

Denetimsiz makine öğrenmesi bir önceki bölümde de ele alındığı gibi, sisteme sadece verilerin yüklendiği, işaretleme yapılmayıp sistemin kendi kendine keşif yapmasına dayanan süreçlerdir.

Denetimli öğrenme ise kendi içinde dörde ayrılmaktadır. Bunlar; karar ağaçları, kural tabanlı sınıflandırıcılar, doğrusal sınıflandırıcılar ve olasılıksal sınıflandırıcılarıdır.

Karar ağaçları; bir önceki bölümde ayrıntılı bir şekilde anlatıldığı gibi; düğüm dal ve yapraklardan oluşan bir yapıya sahiptir.

Kural tabanlı sınıflandırıcılar; belli bir kurala bağlı olarak yapılan sınıflandırmalardır. Örneğin, öğrencinin not ortalaması, 65 ve üstündeyse başarılı, 65'in altındaysa başarısız sayılması gibi bir kuralın ortaya konması ve sistemin bu kuralla çalışması gibi. Kural tabanlı sınıflandırıcılar karar ağaçlarına benzer bir mekanizmaya sahiptir, hatta tüm karar ağaçları kural tabanlı sınıflandırmaya, kural tabanlı sınıflandırmalar ise karar ağaçlarına dönüştürülebilir. Fakat dal sayısı arttıkça karar ağaçlarından ziyade kural tabanlı sınıflandırıcılar daha kullanışlı bir hal almaktadır. Karar ağaçları hiyerarşik bir yaklaşımla amacına ulaşma eğilimindedir.

Doğrusal sınıflandırıcılar; doğrusal bir fonksiyon tarafından yapılan sınıflandırma tekniklerdir. Doğrusal sınıflandırıcılar destek vektör makineleri ve sinir ağları olmak üzere ikiye ayrılırlar.

---

<sup>116</sup> Tahiroğlu B. Tahir, a.g.e. s. 4103

Destek vektör makineleri; sınıflandırma için bir alanda bulunan iki grup arasına doğrusal bir sınır çizilerek iki grubu ayırmak mümkündür. Bu sınır iki grubun üyelerine de en uzak olan yerdir.<sup>117</sup> Destek vektör makineleri bu sınırın nasıl çizileceğini belirleyen yönteme verilen addır.

Sinir ağları (yapay sinir ağları YSA); biyolojik sinir ağlarının yapısını ve işlevselliğini taklit etmeye çalışan bir matematiksel modeldir. YSA'nın üç basit kural kümesi vardır, bunlar; çarpma, toplama ve etkinleştirmedir. Sinir ağları tıpkı bir beynin çalışma sistemini baz alır. Beyinde nöron ismi verilen sinir hücreleri ve bu hücreleri birbirlerin bağlayan noktalar sinapslar vardır. İnsan beyninde öğrenme sinapsların ayarlanmasıyla olur. İnsan öğrendikçe sinapslar oluşur, yeni bağlantılar kurulur. Sinir ağları da çalışma prensibi olarak bunu baz alır. YSA'lar birbirlerine bağlı birçok birimden oluşur. Bu birimler birbirlerine çeşitli noktalarla bağlanırlar. YSA da girdi belli bir örüntüden hareketle çıktı haline gelir. YSA'nın girişinde girdilerin ağırlığı her girdi değerinin bireysel ağırlık ile çarpılması anlamına gelir, suni nöronun orta bölümünde tüm ağırlıklı girdileri önbilgilerle toplayan bir toplam fonksiyon vardır, YSA'nın çıkışında aktarım (transfer) fonksiyonu da denen aktivasyon fonksiyonu vardır. Her yeni girdi örüntüsü ve çıktı sinyallere göre tekrar ayarlanır. Bu ayarlanma süreci öğrenme olarak adlandırılır. Öğrenme tamamlandığında sistem stabil bir hal alır.<sup>118,119</sup>

Olasılıksal sınıflandırıcılar ise; olasılığa dayalı olarak geliştirilmiş sınıflandırma tekniklerinden en sık kullanılanlar; Navie Bayes, Bayesyan Ağlar ve Maksimum Entropidir.

Navie Bayes; olasılığa dayalı olan sınıflandırma türlerinden ilkidir. Sınıflandırma modeli; birçok özellikten ve bir sonuç değişkeninden oluşur. Hesaplamalar, sınıflar düzeyinde gerçekleştirilir, her bir sınıf için olasılık değeri hesaplanarak en yüksek olasılık değerine sahip sınıf, sınıflandırılması yapılacak olan dokümanın sınıfı olarak tespit edilmiştir. NB sayesinde etiketlenmiş verilerden hareketle, etiketi bilinmeyen verilerin hangi sınıfa dahil olabileceğinin olasılığı

---

<sup>117</sup> Şeker, a.g.e.

<sup>118</sup> Tahiroğlu, a.g.e.

<sup>119</sup> [https://tr.wikipedia.org/wiki/Yapay\\_sinir\\_a%C4%9Flar%C4%B1](https://tr.wikipedia.org/wiki/Yapay_sinir_a%C4%9Flar%C4%B1), 13.03.2017

hesaplanır. Bu konuya dair ayrıntılı bilgi MÖ bölümünde ayrıntılı bir şekilde ele alınmıştır.

Bayesian Ağlar; sonlu sayıdaki duruma sahip değişkenleri belirten düğümler ve bu değişkenler arasında ki koşullu olasılıkları gösteren oklardan oluşan bir yapıya sahiptirler.<sup>120</sup> Olasılıksal bir model olan Bayes Ağları, sistemde ki tüm değişkenlerin ve bu değişkenlerin arasındaki ilişkilerin görsel olarak ifade edilmesi, ağın yapısı oluşturulduktan sonra yapılan gözlemlerin çıkarım sürecine dahil edilebilmesi ve hesaplanan son olasılıkların güncellenebilmesi özelliklerine sahiptir.<sup>121</sup> Bayes Ağlarında sistemdeki tüm değişkenler birbirleriyle olasılıksal olarak bağımlılık ilişkisine sahiptirler.

Bayes Ağlarının yapısının belirlenmesinde iki alternatif yaklaşım söz konusudur. Birinci yaklaşıma göre değişkenler arasındaki bağımlılık ilişkisi ve ilişkinin yönü uzman görüşüne dayanılarak belirlenir, ikinci yaklaşıma göre ise ağ yapısı mevcut algoritmalar aracılığıyla veri setinden öğrenilir.<sup>122</sup>

Bayes Ağları ebeveyn ve çocuk düğümlerle bu düğümler arasındaki oklardan oluşurlar. Ağda bir değişkenden diğerine doğru ok çıkıyorsa ok çıkan değişken ebeveyn değişken, kendisine ok gelen değişkense çocuk değişken olarak adlandırılır.

---

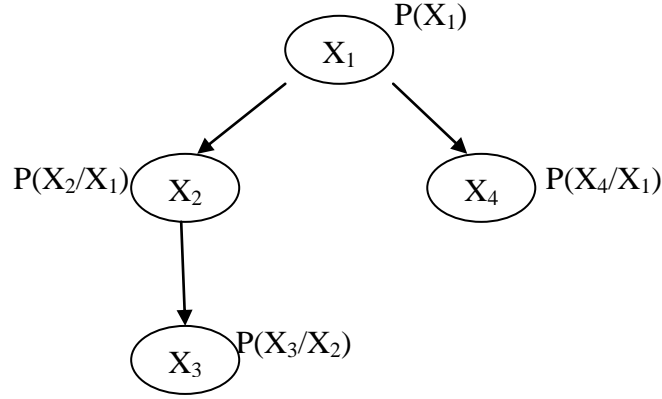
<sup>120</sup> Dereli Burak, 'Bayes Ağları İle Gelecek Odaklı Konumlandırma Analizi: Oyun Konsolları Pazarında Bir Uygulama', İstanbul Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, 2012, s. 15

<sup>121</sup> Çinicioğlu Esmâ Nur, Atalay Muhammet ve Yorulmaz Harun, 'Trafik Kazaları Analizi İçin Bayes Ağları Modeli', Bilişim Teknolojileri Dergisi, Cilt 6, Sayı 2, Mayıs 2013, s 42.

<sup>122</sup> Çinicioğlu Esmâ Nur, Ekici Şule Önsel ve Ülengin Füsün, 'Bayes Ağ Yapısının Oluşturulmasında Farklı Yaklaşımlar: Nedensel Bayes Ağları Ve Veriden Ağ Öğrenme' Siyasal Kitapevi, 2015, s. 267.



**Şekil 3.2.** Dört Değişkenden oluşan Basit Bir Bayes Ağı



**Kaynak:** Çinicioğlu ve Diğerleri (2015, 267)

Şekil 2.2 de dört değişkenden oluşan basit bir Bayes Ağı görülmektedir. Oklar X'ler arasındaki ilişkinin yönünü göstermektedir. Ağda  $X_1$  ebeveyn değişken,  $X_2$  ve  $X_4$   $X_1$ 'in çocuk değişkenleri,  $X_3$  ise  $X_2$ 'nin çocuk değişkeni olarak görünmektedir.  $P(X_1)$   $X_1$  olayının olasılığını,  $P(X_4/X_1)$   $X_4$  olayının  $X_1$  olayına bağlı koşullu olasılığını,  $P(X_2/X_1)$   $X_2$  olayının  $X_1$  olayına bağlı koşullu olasılığını,  $P(X_3/X_2)$  ise  $X_3$  olayının  $X_2$  olayına bağlı koşullu olasılığını ifade etmektedir. Bu niteliksel gösterimin niceliksel gösterim hali ise aşağıda (2.1)'de formülize edilmiştir;

$$P(X_1, \dots, X_N) = \prod_{i=1}^N P[X_i | Ebeveyn(X_i)] \quad (2.1)$$

Bayes Ağları, hesaplanması karmaşık olduğundan yoğun olarak kullanılmaz ancak belirsizlik altında olasılığa dayalı olarak akıl yürütmek için ihtiyaç duyulduğunda kullanışlıdır.

Maksimum Entropi; belirli bir kısıtlamayı sağlayan tekdüze modelleri tercih eden, özellik tabanlı sınıflandırıcılardır. Eğitim aşamasındaki etiketli veriler, sınıfı karakterize eden modelin kısıtlarını türetmek için kullanılır. Naive Bayes'in aksine özellikler arasında bağımsızlık varsayımı yapmaz. Bu sayede özelliklerin çakışması konusunda endişelenmeden rahatlıkla her türlü n-gram sınıflandırma yapılabilir.

Maksimum entropi sınıflandırıcılar ile çok zor sınıflandırmalarda bile iyi performans gösterebilirler.<sup>123</sup>

### 2.3.2.1. Sınıflandırma Algoritmalarının Karşılaştırılmalarında Kullanılan Kriterler

Sınıflandırma yöntemleri karşılaştırılırken aşağıda yer alan kontenjans tablosundan hareketle hesaplanan kriterlerden faydalanılır.

**Tablo 2.1.** Kontenjans Tablosu

Gerçek Değerler \ Tahmini Değerler	Pozitif	Negatif
Pozitif	Doğru Pozitif (TP)	Yanlış Pozitif (FP)
Negatif	Yanlış Negatif (FN)	Doğru Negatif (TN)

#### Kesinlik (Precision – P)

Tutarlılık olarak da ifade edilebilen kesinlik kavramı; sınıflandırıcının bir sınıf için yaptığı sınıflandırmada ne kadar hassas olduğunu gösterir. Sınıflandırıcı sonucunun kesinlik derecesini verir.

P pozitif olarak etiketlenen örneklerin sayısını, N negatif olarak etiketlenen örnekleri sayısını, TP (True Pozitif) doğru bir şekilde etiketlenmiş olan P'lerin sayısını, FP (False Pozitif) yanlış bir şekilde etiketlenmiş olan P'lerin sayısını göstermek üzere kesinlik formülü (2.2) gibidir;

$$Kesinlik = \frac{TP}{TP+FP} \quad (2.2)$$

<sup>123</sup> Isidoros Perikos ve Ioannis Hatzilygeroudis, 'Recognizing Emotions in Text Using Ensemble of Classifiers', *Engineering Applications of Artificial Intelligence*, 2016, s.195.

### **Duyarlılık (Recall – R)**

Bir sınıftaki örneklerin ne kadarının düzgün sınıflandırıldığı duyarlılıkla ölçülür. Duyarlılık pozitif olarak etiketlenmiş örneklerin gerçekten pozitif olan örneklerin toplamına oranıdır. Gerçekten pozitif olan örnekler; TP ile FN'nin (False Negatif - yanlış etiketlenmiş negatiflerin) toplamına eşittir. Duyarlılık (2.3) teki gibi formülüne edilebilir;

$$Duyarlılık = \frac{TP}{TP+FN} \quad (2.3)$$

### **Doğruluk (Accuracy – A)**

Doğruluk, sınıflandırmada en fazla kullanılan ölçüm değeridir. Sınıflandırıcının bütün örnekleri sınıflandırmada ne kadar başarılı olduğunun ölçüsüdür. Yukarıda yazılı olan ifadeler ek olarak, TN (True Negatif) doğru bir şekilde etiketlenmiş negatif örnekler olmak üzere (2.4) teki gibi formülüne edilebilir;

$$Doğruluk = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.4)$$

### **F- Ölçütü (F- Measure)**

Tek başına ne kesinlik ne de duyarlılık ölçütleri bir karşılaştırma yapmamız için yeterli değildir. Bu yüzden hem kesinlik hem de duyarlılığı bir arada kullanan, sistemi kesinlik veya duyarlılık yönüne optimize eden F-Ölçütü geliştirilmiştir. (2.5) F-Ölçütünün formülüne edilmiş şeklini göstermektedir;

$$F - Ölçütü = 2 \frac{Kesinlik \cdot Duyarlılık}{Kesinlik + Duyarlılık} \quad (2.5)$$

## Kappa İstatistiği

Cohen'in kappa katsayısı olarak ta ifade edilen kappa istatistiği, iki gözlemci arasındaki uyumu ölçen istatistik değeridir. Kappa istatistiği, İki gözlemci arasında ki ilişkinin tesadüfi olarak da ortaya çıkabileceğini göz önüne alarak hesaplanmaktadır.<sup>124</sup> Bu yönü ile basit yüzde orantılı olarak hesaplanan uyumdan çok daha güçlüdür. Kappa istatistiği (2.6) daki gibi fomülüze edilebilir.

$$k = \frac{Po - Pe}{1 - Pe} \quad (2.6)$$

Formülde Po gözlenen uyum oranını, Pe tesadüfi olarak ortaya çıkan uyum oranını göstermektedir. k değeri -1 ile +1 arasında yer alabilir. k değerinin 0 dan küçük olması iki gözlemci arasında bir uyum olmadığını gösterir. 0 dan büyük ancak +1'den çok küçük olan değerler için uyumun tesadüfi olduğu düşünülebilir. k değeri +1'e yaklaştıkça iki gözlemci arasındaki uyumun tesadüfi bir şekilde ortaya çıkamayacağı ve uyumun mükemmel olduğu, başka bir deyişle her iki gözlemcinin de homojen olduğu söylenebilir.

### 2.3.2.2. Model Performans Değerlendirme Yöntemleri

Model değerlendirme, model geliştirme sürecinin ayrılmaz bir parçasıdır. Verileri temsil eden en iyi modeli bulmaya ve seçilen modelin gelecekte ne kadar iyi çalışacağına model değerlendirme teknikleri yardımcı olur. Model performansının, tek başına eğitim için kullanılan verilerle değerlendirilmesi, veri madenciliğinde kabul edilemez. Bu yüzden veri madenciliğinde çeşitli model performans değerlendirme yöntemleri geliştirilmiştir.<sup>125</sup>

### Holdout (Dışarıda Tutma)

Holdout yöntemi ile veri seti; eğitim ve test olmak üzere ikiye ayrılır. Genellikle verilerin 2/3'ü eğitim ve 1/3'ü test verisi olarak ele alınır ancak verileri %50 eğitim,

<sup>124</sup> Kılıç Selim, 'Kappa Testi', *Journal of Mood Disorders*, Vol. 5, 2015, s. 142–144.

<sup>125</sup> Model Evaluation, [http://chem-eng.utoronto.ca/~datamining/dmc/model\\_evaluation.htm](http://chem-eng.utoronto.ca/~datamining/dmc/model_evaluation.htm), erişim tarihi; 02.02.2018

%50 test olarak ayırmakta mümkündür. Eğitim veri seti ile öğrenme gerçekleşirken (model kurulurken) test veri seti ile ne kadar öğrenmenin gerçekleştiği ortaya konularak modelin performansı ortaya çıkarılır.<sup>126</sup>

Holdout pratik bir yöntem olmakla birlikte bir takım dezavantajlara da sahiptir. Şöyle ki; zaman serisi ile çalışılıyorsa zaman serisinde verileri eğitim ve test seti olarak ikiye ayıramazsınız.<sup>127</sup> Çalışılan veri miktarı az sayıda ise bu veriyi ikiye ayırırken test veri seti için elinizde yeteri kadar veri kalmayacaktır, veri seti sadece bir kere eğitim ve test olmak üzere ikiye ayrılacaktır, veri seti ayrılırken yapılacak olası bir hata kalıcı olacaktır, modele pozitif katkıda bulunacak veriler test veri setinde kalmış olabilir.<sup>128</sup>

### **Tekrarlı Holdout**

Holdout yönteminin üst üste birkaç kez tekrarlanması esasına dayanan yöntemdir. Veri seti birkaç kez alt kümelere bölünür ve her bir kümeye holdout yöntemi uygulanır. Modelin hata oranı, her bir alt kümeye ait modellerin hata oranlarının ortalaması hesaplanarak bulunur.<sup>129</sup> Tekrarlı Holdout yönteminde her ne kadar alt kümelerin seçimleri tesadüfi olarak yapılsa da farklı test setlerinin üst üste binmesi mümkündür. Bu yüzden çok fazla tercih edilen bir yöntem değildir.

### **Tabakalı Örnekleme**

Hedef niteliği nominal veri tipinde olan veri setlerinde kullanılabilen bir yöntemdir. Bu tip verilerde bazı sınıfların verilerinin çok az olması veya hiç olmaması gibi bir durumla karşılaşılabilir. Böyle bir durumla karşılaşıncı, sınıf oranlarının korunabilmesi için, tabakalı örnekleme tercih edilmektedir. Ancak hedef nitelik nümerik ise bu yöntem kullanılamaz.

### **Üçlü Ayırma**

Veri seti aynı anda üç farklı kategoriye ayrılır. Bu kategoriler; eğitim, doğrulama ve test setleridir. Diğer yöntemlerde eğitim ve test olarak ikiye ayrılan verilere bu yöntemde üçüncü bir grup olan doğrulama veri seti eklenmiştir. Doğrulama veri setinin amacı, eğitim veri setinden hareketle oluşturulan modele ait parametrelerin ince

<sup>126</sup> Kartal Elif, 'Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri Ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulama' (İstanbul Üniversitesi Fen Bilimleri Enstitüsü, 2015). S:30

<sup>127</sup> <https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/training-and-testing-data-sets>

<sup>128</sup> Kartal, age. S.30

<sup>129</sup> Kartal Elif, 'Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri Ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulama' (İstanbul Üniversitesi Fen Bilimleri Enstitüsü, 2015). S.50

ayarlarının yapılmasıdır. Test veri seti ince ayarı yapılmış modelin performansını ölçmek amacıyla kullanılır. Üçlü ayırma yöntemi ile aynı anda model seçimi ve performans tayini yapılır.

### **K- Kat Çapraz Doğrulama**

Veri seti çok küçük olduğunda en sık kullanılan tekniklerden biri k-kat (k-fold) çapraz doğrulama tekniğidir. Önce veri seti k adet eşit paçaya bölünür, k parçadan her biri önce 1 tanesi test olarak ayrılıp geri kalan k-1 tanesi eğitim seti olarak kullanılarak işleme tabi tutulur. Toplam da k kez bu süreç tekrarlanmış olur ve k adet hata değeri elde edilir. Sonuç hata değeri ise tüm doğruluk değerlerinin ortalaması hesaplanarak bulunur. Daha doğru sonuçlar elde etmek için k sayısı büyük seçilir, holdout yönteminden daha sağlıklı sonuçlar doğurur. Yaygın bir şekilde k; 5 veya 10 olarak ele alınmaktadır.<sup>130</sup>

Birini dışarıda bırakarak çapraz doğrulama; k kat çapraz doğrulamanın özel bir halidir. k adet parçadan 1 tanesi her zaman doğrulama için kullanılır geri kalan k-1 tanesi ise eğitim veri seti olarak kullanılır. Veri seti çok az olduğunda birini dışarıda bırakarak çapraz doğrulamanın uygulanması önerilmektedir.<sup>131</sup>

### **Multiple Random Sampling (Çoklu Rassal Örnekleme)**

Veri kümesinin boyutu çok küçük ise eğitim ve test olarak bölünürken test kümesinin boyutu çok daha küçük olacağından elde edilen sonuçlar güvenilir olmayacaktır. Bu durumda n adet rassal örnekleme ile eğitim ve test kümeleri oluşturulur, bu sayede n adet doğruluk değeri elde edilir, elde edilen doğruluk değerlerinin ortalaması modelin doğruluk değerini ortaya koyar.<sup>132</sup>

### **Bootstrap Örnekleme**

Bootstrap örneklemede; “n” adet örnekten oluşan veri setinden, iadeli olarak “n” defa rassal örneklemin eğitim veri seti olarak seçilmesi ile oluşur. Eğitim veri seti oluşturulurken iadeli seçim uygulandığından herhangi bir örnek eğitim setinden birden fazla tekrar edebilmektedir. Eğitim seti seçme işlemi bittikten sonra test veri seti olarak,

---

<sup>130</sup> Tadayoshi Fushiki, ‘Estimation of Prediction Error by Using K-Fold Cross-Validation’, *Statistics and Computing*, Springer, 2011, s 138.

<sup>131</sup> <https://showmethenotes.blogspot.com.tr/2017/10/mkina-ogrenmesi-5.html>

<sup>132</sup> [http://w3.gazi.edu.tr/~akcayol/files/WM\\_L5ClassifierEvaluation.pdf](http://w3.gazi.edu.tr/~akcayol/files/WM_L5ClassifierEvaluation.pdf)

eđitim veri setinde yer almayan 6rnekler kullanılır, dolayısıyla test veri setinde kalan 6rnekler sadece 1 kez veri seti iinde yer alacaktır.



## ÜÇÜNCÜ BÖLÜM

### SOSYAL MEDYA

#### 3. SOSYAL MEDYA

Teknolojik gelişmeler yaşamımızın her alanında önemli değişikliklere yol açtığı bilinen bir gerçektir. Özellikle bilgisayarların yaygınlaşmasıyla birlikte; iletişim süreci de bu değişikliklerden payını almıştır. Geçmişte yazılı, sözlü ve sözsüz iletişim olarak adlandırılan iletişim kurma yollarına artık elektronik iletişim veya bilgisayarlar aracılığıyla iletişim adıyla yeni bir süreç daha eklenmiştir.

İnternet teknolojilerinin hızla gelişmesi, internetin kullanım alışkanlıklarını da değiştirmiştir. Geçmişte internet dünyasında sadece birer tüketici olan kullanıcılar Web 2.0 teknolojisiyle birlikte artık birer üretici konumuna geçmiştir. İnternet teknolojisi; Web 1.0, Web 2.0 ve Web 3.0 olarak ifade edilmektedir. Web 1.0, ilk geliştirilen internet teknolojisidir, Web 1.0'da kişilerin sadece içerik tüketicisi olarak yer aldığı, genel amacın kullanıcıya metin, video, fotoğraflar gibi içerikler sunulması ve kullanıcının pasif bir role sahip olduğu ortamlardır. Kullanıcı sunulan içeriği sadece okuyabilir, görüntüleyebilir, seyredebilir ve sayfalar arasında gezinebilir, tüm bunlara rağmen web ortamına kendisi müdahale edemez.<sup>133</sup> Web 2.0 kavramı ilk olarak O'Reilly Medya'nın sahibi Tim O'Reilly tarafından 2005 yılında bir konferansta dile getirilmiştir.<sup>134</sup> Web 2.0'dan önce bireyler kendilerine sunulan içeriği tüketen durumunda iken, Web 2.0'dan sonra bireyler artık sadece birer tüketici değil, kendileri de içerik üreticisi durumuna gelmiştir. Web 2.0 sayesinde internet kullanım anlayışı değişmiştir. Eskiden sadece kendine sunulanları okuyan bireyler artık internette kendilerini ifade edebilecek yazılar yazmaya başlamıştır, önceden sadece kendine sunulan videoyu izleyebilen kullanıcılar artık kendi ürettikleri videoları paylaşabilmektedir. Bu gibi yenilikler internet dünyasında yeni bir çığır açmıştır. Açılan bu çığırlardan biri de çevrimiçi toplumsal ağlar olarak da tanımlanabilecek olan sosyal medyadır. Bu yüzden Web 2.0 sosyal medya araçları olarak da adlandırılmaktadır. Web 3.0'ın temel amacı ise kişisel özellikleri göz önünde bulundurarak kullanıcılara özel

---

<sup>133</sup> Topaçan Ümit.s.9-10

<sup>134</sup> O'reilly Tim, (2005), What is web 2.0 Design Patterns And Business Models For The Next Generation Of Software, <http://oreilly.com/lpt/a/6228>



içeriklerin üretilmesini sağlamaktır. Bunu yaparken de temel olarak kişilerin geçmiş kullanım özelliklerinin veri olarak kullanılıp, kişisel tercihlerin ortaya konulması mantığına dayanmaktadır.

### 3.1. Sosyal Medyanın Tanımı

Sosyal medya kavramının iki ayağı vardır. Bunlardan ilki Web 2.0, ikincisi ise Kullanıcı Tarafından Oluşturulan İçeriktir. Web 2.0 kavramı, içerik ve uygulamaların sadece belli bireyler tarafından değil de sürekli bir katılımcı ve işbirliği çerçevesinde kullanıcıların tamamı tarafından oluşturulabilmesi ve/veya değiştirilebilmesini ifade eder. Kullanıcı Tarafından Oluşturulan İçerik kavramı ise bireylerin sosyal medyayı kullanma yolları olarak ifade edilebilir. Bu iki kavram birlikte ele alındığında; sosyal medya, Web 2.0'ın ideolojik ve teknolojik temelleri üzerine kurulu, kullanıcı tarafından üretilen içeriğin oluşturulmasına ve değiştirilmesine imkan tanıyan bir dizi internet tabanlı uygulamadır<sup>135</sup> şeklinde tanımlanmıştır. Kullanıcılar sosyal medya uygulamalarını kullanarak, web ortamında bilgi oluşturabilir ve bunları paylaşabilirler. Sosyal medya ortamlarının türü ne olursa olsun ortak amaçları iletişim kurup, paylaşımlarda bulunmaktır.

Çok farklı sosyal medya araçları olmasına karşın, bu sosyal medya araçlarının hepsinin ortak temel özellikleri vardır. Bunlar, kullanıcılar arasında, iletişim kurulması, paylaşımlarda bulunulması ve işbirliği yapılmasıdır. Mayfield'a göre sosyal medya aşağıda açıklanan özelliklerin çoğunu veya tamamını paylaşan online medya ortamlarının yeni türüdür.<sup>136</sup>

**Katılım:** Sosyal medya araçları ilgi duyan herkesin katkı sunmasını ve geribildirim sağlamasını teşvik eder. Bu sayede medya ile izleyici arasındaki çizgiyi ortadan kaldırır.

**Açıklık:** Çoğu sosyal medya araçları geribildirim ve katılıma açıktır. Kullanıcılarını oylamaya, yorum yapmaya ve bilgi paylaşmasına teşvik eder. Sosyal medya araçlarında içeriğe erişme ve bunlardan faydalanma nadiren engellenir.

---

<sup>135</sup> Topaçan a.g.e. .61

<sup>136</sup> Topaçan Ümit., e-book, Icrossing, 2008, s.5.

Konuşma: Geleneksel medya yayına yöneliktir yani tek taraflı bir süreç vardır. Oysa sosyal medya da iki yönlü bir süreç vardır, karşılıklı konuşmaya yer verir.

Topluluk: sosyal medya toplulukların hızla oluşmasını ve etkin iletişim kurmalarını sağlar. Ortak ilgi alanına sahip kişileri bir araya getirerek paylaşım yapmalarını sağlar.

Bağlantılı Olmak: Birçok sosyal medya araçları birbirlerine bağlı olmaları sayesinde gelişir. Diğer sitelerle, kaynaklarla ve insanlarla bağlantı kurarlar.

### **3.2. Sosyal Medya Araçları**

Geliştirilmiş pek çok sosyal medya aracı mevcuttur. Bunların tamamının ortak noktası kullanım kolaylığıdır. Ancak kullanıcı sosyal medya araçlarından ihtiyacına en uygun olanı seçmelidir. Çünkü tüm sosyal medya araçları aynı amaca hizmet etmemektedir. Örneğin kullanıcı video paylaşımları için youtube gibi sosyal medya araçlarını, kısa mesaj paylaşımı için twitter gibi mecraları, sadece alanında ki profesyonellerle iletişim kurmak için linkedln kullanmalıdır.

Dijital bir sosyal pazarlama ajansı olan “We Are Social” her yıl düzenli olarak internet ve sosyal medya kullanıcı istatistiklerini yayınlamaktadır. 2017 yılında “Hootsuite” ile işbirliği yaparak yayınladıkları “Digital In 2017 Global Overview A Collection Of Internet, Social Media, And Mobile Data From Around The World” raporunda dünya nüfusunun yarısından fazlasının internet kullandığını, internet kullanımının her yıl %8 oranında artış gösterdiğini ortaya koymuştur. Yine aynı rapor da mobil cihazların artışıyla birlikte dünya genelindeki web trafiğinin çoğunun mobil telefonlarla sağlandığı, dünya nüfusunun üçte birinden fazlasının düzenli olarak sosyal medya kullandığı ve son bir yılda sosyal medya kullanıcı sayısının %20 arttığını ortaya koymuştur. Raporda sosyal medya kullanıcı sayısındaki bu artışın temel nedeni mobil sosyal medya kullanımındaki artışa bağlanıyor.<sup>137</sup> Aynı raporun Türkiye’ye dair kısmı incelendiğinde; Türkiye’de nüfusun 80 milyon olduğu ve nüfusun %60’ının internet kullandığı ve internet kullanım oranı ile sosyal medya kullanım oranının eşit olduğu ifade edilmektedir. Türkiye’de son bir yıl içinde internet kullanıcı sayısı %4 artarken aynı dönem içinde sosyal medya kullanıcı sayısı %14 artmıştır. Türkiye’de kullanıcılar

<sup>137</sup> Kemp Simon, Digital In 2017 Global Overview, <https://www.slideshare.net/wearesocialsg/digital-in-2017-global-overview>, 27.11.2017

ortalama olarak günde 3 saat 1 dakikalarını sosyal medyada geçirdikleri saptanmıştır. Rapora göre dünyada en fazla kullanıcıya 1,8 milyar kişi ile facebook sahiptir. Facebooku 1 milyar kullanıcı ile youtube, 632 milyon kullanıcı ile qzone, takip etmektedir. En fazla kullanıcıya sahip olan mikroblog sitesi ise 317 milyon ile twitterdir.

Sosyal medya araçların pek çok başlık altında sınıflandırmak mümkündür, çalışma kapsamında en yaygın kullanılan sosyal medya araçları ele alınarak aşağıda teker teker açıklanacaktır.

### 3.2.1. Bloglar

Bloglar geçmişte tutulan günlüklerin web ortamında ki yansımaları olarak günümüzdeki karşılığı olarak düşünülebilir Türkçeye web günlüğü olarak çevrilebilirler. 1997 yılında web ve log kelimelerinin birleşimiyle weblog kelimesi kullanılmaya başlanmış ve bir kullanıcı bu kelimeyi şaka amacıyla “we blog” olarak ayırmıştır. Şakayla başlayan bu süreç blog kelimesini tek başına günlük konuşma dilinde kendine yer bulmasıyla literatürde yerini almıştır.<sup>138</sup>

Bloglar kullanıcıların herhangi bir konu üzerine kendi bilgilerini, fikirlerini veya deneyimlerini paylaştıkları alanlardır.<sup>139</sup> Blogların en büyük avantajları teknik bilgi ve donanım gerektirmemesidir. Bu yüzden kullanımı son derece kolaydır.<sup>140</sup> Bloglar sadece yazıların paylaşıldığı ortamlar değil aynı zamanda fotoğraf ve video gibi içeriklere de yer verilebilen ortamlardır. Blog kullanıcıları takipçileriyle etkileşim içindedir, çoğu zaman kullanıcılar takipçilerine paylaşımları altında yorum yapma imkânı da tanımaktadır.

Blog sahiplerine blogger denir. Bloglarda ters kronolojik bir sıralama vardır yani son yazılan en üstte görüntülenir. Bu sayede takipçi bloggerın son yazdıklarını ilk görüntüleme şansına sahip olur, özellikle bir bloggerın sürekli takipçileri için bu özellik son derece kullanışlıdır.

---

<sup>138</sup> Mayfield Antony, *What Is Social Media?E-book* (Icrossing, 2008).

<sup>139</sup> Mayfield a.g.e..

<sup>140</sup> Alikılıç Özlem ve Onat Ferah, ‘Bir Halkla İlişkiler Aracı Olarak Kurumsal Bloglar’, *Journal of Yasar University*, 2007, Cilt 2, Sayı: 8, s. 904.

### 3.2.2. Wikiler

Wikiler işbirliği projeleri veya kullanıcı üretimi içerik olarak ta adlandırılabilir. Temel amaç birden fazla kullanıcının bir araya gelerek işbirliği ile bir içerik üretmesidir. Wiki kelimesi İngilizcede “What I Know is” sözcüklerinin ilk harflerinin bir araya getirilmesiyle oluşturulmuştur.<sup>141</sup>

Wikilerin en belirgin özelliği daha önce başka bir kullanıcı tarafından oluşturulan içerikler zaman içinde diğer kullanıcılar tarafından değiştirilebilmektedir. Sosyal medya araçlarının genel özelliklerinden birisi sosyalleşme iken wikilerin genel özelliği sosyalleşme değil bilgi paylaşımıdır.

Dünyada en fazla kullanıcıya sahip ve en popüler wiki aracı 2001 yılında Jimmy Wales ve Larry Sanger tarafından kurulmuş olan wikipediadır. Wikipedia kar amacı gütmeyen, sadece bağışlarla yaşamını sürdüren, bu yüzden de hiçbir reklam almayan bir organizasyondur.

### 3.2.3. Sosyal Ağlar

Sosyal ağ siteleri (social network sites) olarak da adlandırılan sosyal ağlar, ilk başta günlük yaşamda insanların bir parçası olduğu ağların web ortamına yansımaları olarak algılanabilmektedir. Fakat web tabanlı sosyal ağları sanal olmayan ortamlardan ayıran pek çok özellik mevcuttur. Sosyal ağ siteleri; kullanıcılara, üye olarak belli bir profil oluşturma, bu profili sistemde yer alan herkesle veya sadece belli bir kesimle paylaşma olanağı tanıyan, kullanıcıların herhangi bir bağlantıyı paylaşabilecekleri listeleri olduğu, diğer kullanıcıların profillerini ve/veya paylaştıkları bağlantılarını görebildikleri web tabanlı ortamlar olarak tanımlanmaktadır.<sup>142</sup>

Günümüz sosyal ağ sitelerinin atası olarak, 1997 yılında kurulan SixDegrees.com sitesi kabul edilir. Bunun temel nedeni daha önce kurulan çeşitli web sitelerinin genellikle çevrimiçi sohbet odaklı olmasına karşın SixDegrees.com sitesinin kullanıcılara kendi profillerini oluşturma imkânı tanıması, arkadaş listesi oluşturulmasına ve bu kişilere mesaj göndermeye imkan tanınmış olmasıdır.

---

<sup>141</sup> Scoott, age, s. 37.

<sup>142</sup> Boyd Danah M. ve Ellison Nicole B., ‘Social Network Sites: Definition, History, and Scholarship’, *International Review of Research in Open and Distance Learning*, 2011 s. 211.

Sosyal ağ siteleri zamanla önemini baş döndürücü bir hızla arttırmış, hemen her alana ilişkin özel sosyal ağ siteleri oluşturulmuştur. Günümüzde dünya üzerinde en fazla ziyaret edilen siteler sosyal ağlar olmuştur. Sosyal ağ sitesi denilince ilk akla 2004 yılında kurulan Facebook gelmekle birlikte, yine Çinlilere ait olan facebook benzeri özelliklere sahip olan Qzone en önemli sosyal ağ sitelerindedir. İş hayatı ve profesyonelle yönelik olarak hizmet veren linkedIn yine önemli sosyal ağ sitelerinden birisi olarak kaşımıza çıkmaktadır.

### 3.2.4. İçerik Paylaşım Toplulukları

Sosyal medya kullanıcılarının kendi ürettikleri bir takım içerikleri veya başkası tarafından üretilmiş olan, beğendikleri içerikleri, insanlarla paylaşmaları amacıyla çeşitli web ortamları geliştirilmiştir. İçerik paylaşım toplulukları içerik paylaşım siteleri olarak da adlandırılırlar. İçerik paylaşım topluluklarında amaç; paylaşılan içeriklerin web üzerinden görünür hale gelmesi ve kullanıcılarca erişilebilmesi, kullanıcılar tarafında da başkalarıyla paylaşılabilmesine izin verilmesidir.<sup>143</sup> Paylaşılacak şey fotoğraf, video veya bir sunum dosyası gibi kullanıcıları ilgilendirebilecek veya ilgi duyabilecekleri her şey olabilir.<sup>144</sup> Özellikle akıllı telefonların yaygınlaşmasıyla içerik paylaşım toplulukları daha popüler hale gelmiştir çünkü her akıllı telefon sahibi artık rahatlıkla web ortamına aktarabileceği bir videoyu veya anı ölümsüzleştirmek için fotoğrafı çekebilmektedir.

En yaygın kullanılan içerik paylaşım topluluklarından biri video paylaşım platformu olan youtubedir. Youtube 1 milyar kullanıcısı ile dünya genelinde en fazla kullanıcıya sahip ikinci sosyal medya aracı olarak kaşımıza çıkmaktadır. Instagram ise 500 milyon kullanıcısı ile en fazla kullanılan fotoğraf paylaşma topluluğudur.<sup>145</sup> Bu sitelerin yanı sıra ses paylaşım siteleri, yer paylaşma siteleri gibi uygulamalar da kullanıcının karşısına çok fazla çıkmaktadır.

İçerik paylaşım toplulukların genel özellikleri birkaç başlık altında toplanabilir. Kolaylık bunlardan ilkidir, kullanıcıların paylaşımında bulunabilmesi için teknik bilgiye sahip olmasına gerek yoktur, kolaylıkla paylaşımında bulunabilirler. Geçmişte herhangi

<sup>143</sup> Gökçül Ahmet Neca, 'Sosyal Medyanın Siyasal İletişim Aracı Olarak Kullanımı: 7 Haziran 2015 Genel Seçimleri Twitter Örneği'. İstanbul Yeni Yüzyıl Üniversitesi Sosyal Bilimler Enstitüsü, 2016, s. 60

<sup>144</sup> Akkaya Talih Duygu, 'Sosyal Medya Reklamlarında Tüketici Algılarının Tutum, Davranış Ve Satın Alma Niyeti Üzerine Etkisi', Trakya Üniversitesi Sosyal Bilimler Enstitüsü, 2013, s.25.

<sup>145</sup> <https://wearesocial-net.s3.amazonaws.com/uk/wp-content/uploads/sites/2/2017/01/Slide048.png>, 28.11.2017

bir web sitesinde paylaşımda bulunabilmek için teknik bilgiye sahip olmak gerekirdi, günümüz de ise teknik bilgi olmaksızın rahatlıkla paylaşımlar yapılabilmektedir. Bu topluluklar sosyal bir özelliğe sahiptir çünkü paylaşımlar kullanıcının arkadaşları tarafından görülüp yorumlanabilirler, başkalarının görebilmesi için tekrar paylaşılabilirler. İçerik paylaşım toplulukları birbirlerine rahatlıkla link verir bu sayede yaygın etkiye sahiptirler ve genellikle içerik paylaşım toplulukları ücretsizdir.<sup>146</sup>

### 3.2.5. Vloglar

Vlog video ve log kelimelerinin birleşimiyle oluşmuş bir kelimedir. Teknolojinin hızla ilerlemesiyle birlikte bloglara alternatif olarak geliştirilmiştir. Vlogda temel amaç blogdaki gibi paylaşımlarda bulunmaktır ancak fark bloglarda paylaşımlar yazılı bir şekilde yapılırken vloglarda paylaşımlar videolar vasıtasıyla yapılır. Vlog çekip paylaşan kişilere vlogger adı verilir. Vlog kavramı yeni yeni yaygınlaşmakta olan bir kavram olmakla birlikte ileride hatırı sayılır bir sosyal medya aracı olacaktır.

### 3.2.6. Podcasting

İlk kez 2004 yılında, The Guardian yazarı tarafından kullanılan podcast kelimesi; Apple'ın taşınabilir dijital ses çalar cihazı olan İpod'daki küçük kapsül anlamına gelen "pod" kelimesi ile yayın anlamına gelen "broadcast" kelimesinde ki "cast" hecesinin bileşiminden oluşmuştur. Bu kelime o kadar beğenilmiştir ki 2005 yılında New Oxford American Dictionary editörleri tarafından yılın kelimesi seçilmiştir. Podcasting bir radyo programı veya ses dosyasının internette dağıtılmasına veya internette dağıtılan bu yayınların indirilmesine olanak tanıyan ortamlardır.<sup>147</sup>

Podcasting gibi teknolojiler gerek öğrenme gerekse eğlenmek için gelişmiş bir sanal ortam sunmaktadır. Podcasting son yıllarda hızlı bir şekilde artmaktadır. Podcasting sayesinde bir bilgisayar veya taşınabilir ses çalar sayesinde abone olunan ses dosyası veya video indirilerek kullanılabilir.<sup>148</sup> Podcastların en önemli avantajı aynı şeyi tekrar tekrar dinleyip, izlemeye imkan vermesidir. Bu özelliği sayesinde eğitim alanında da sıklıkla kullanılmaya başlanmıştır. Podcastlere abone olmak ücretsizdir, kullanımı son derece basittir.

<sup>146</sup> Ben Rigby, *Mobilizing Generation 2.0: A Practical Guide to Using Web 2.0*, San Francisco: Jossey-Bass A Wiley Imprint, 2008, s.99.

<sup>147</sup> Podcast, Wikipedia, <https://tr.wikipedia.org/wiki/Podcast>

<sup>148</sup> Parson Vanessa ve diğerleri, 'Educating an iPod Generation: Undergraduate Attitudes, Experiences and Understanding of Vodcast and Podcast Use', *Learning, Media and Technology*, 34.3 (2009), 215–228.

### 3.2.7. Sosyal İmlleme Siteleri (Socila Bookmarking Systems)

Sosyal imleme siteleri; kullanıcıların ilginç buldukları web sitelerine, bloglara, resimlere, wikilere, videolara ve podcastlere ait bağlantıları depolamalarını, sınıflandırmalarını ve paylaşmalarını sağlayan Web 2.0 araçlarıdır. Sosyal imleme siteleri de farklı türlere göre çeşitlenmiştir. Örneğin haberler için farklı siteler, resimler için farklı, bibliyografik referanslar için farklı sosyal imleme siteleri geliştirilmiştir.<sup>149</sup>

Sosyal imleme siteleri herhangi bir konu araştırılırken kaynak bulmaya yardımcı olabilecek mükemmel bir araçtır. Çünkü diğer kullanıcılar benzer etiketler altında çeşitli siteleri etiketlemiştir. Araştırmacılar kullanıcıların etiketledikleri bu sitelerden hareketle aradıkları bilgiye erişebilirler. Bu aynı alanda araştırma yapan kişiler tarafından oluşturulmuş ortak bir listeyi görmeyi sağlar.<sup>150</sup>

Sosyal imleme siteleri aynı zamanda birer sosyalleşme araçlarıdır. Kişiler ortak ilgi alanlarına sahip oldukları bireyleri sosyal imleme sitelerinden görebilirler.

### 3.2.8. Mikrobloglar

Bloglar uzun yazılardan oluşurken mikrobloglar anlık gönderilen, belli bir karakter kısıtı olan, kısa ve öz metinlerden oluşan iletilere izin veren web tabanlı uygulamalardır.

Mikrobloglar kullanıcıların herhangi bir konu hakkında anlık olarak düşüncelerini ortaya koymalarında yaradığı gibi bir linkin veya bir görselin paylaşılmasına da imkan tanırlar. Mikrobloglarda kullanıcılar başka kullanıcıların iletilerine kendi sayfalarında cevap verebilirler ancak ilk paylaşım yapan kişinin iletilerinin altına yorumlarını yazamazlar, bu özellik mikroblogları sosyal ağ sitelerinden ayıran en temel özelliklerden biridir.

Mikrobloglar bloglarla kısa mesajların bileşimi şeklinde organize edilmiştir. bloglar tüm takipçilerle paylaşılan uzun metinler, kısa mesajlar sadece muhatabına gönderilen harf kısıtı olan metinler iken, mikroblogta gönderiler tüm takipçilerle

---

<sup>149</sup> Parson Vanessa ve diğerleri, 'Educating an iPod Generation: Undergraduate Attitudes, Experiences and Understanding of Vodcast and Podcast Use', *Learning, Media and Technology*, 34.3 (2009), 215–228. s.176

<sup>150</sup> Parson Vanessa ve diğerleri, 'Educating an iPod Generation: Undergraduate Attitudes, Experiences and Understanding of Vodcast and Podcast Use', *Learning, Media and Technology*, 34.3 (2009), 215–228.s.66

paylaşılır, bu yüzden de tüm takipçilere aynı anda gönderilen bir kısa mesaj olarak düşünülebilir.

İlk mikroblog uygulaması 2004 yılında Amerika Birleşik Devletleri'ndeki seçimlerde, bir grup aktivistin kendi aralarında mesajlaşmaları amacıyla oluşturdukları TXTmob sitesidir. Günümüzde tüm dünyada mikroblog denince ilk akla gelen uygulama ise Twitter'dır.

### 3.3. Önemli Bir Sosyal Medya Aracı Olan Twitter

Mart 2006 yılında Jack Dorsey tarafından geliştirilen Twitter, kullanıcılarına en fazla 280 karakterden oluşan (2017 yılının son çeyreğine kadar karakter kısıtı 140 adetti ancak yapılan düzenleme ile artık kullanıcılar 280 adet karakter kullanabilmektedir) Türkçe cıvıdamak anlamına gelen mesaj (tweet) gönderme, fotoğraf, video paylaşma ve karşılıklı direk mesaj yollama olanağı sunmuştur.

2006 yılında ortaya konulan Twitter için başlıca iki sıçrama basamağı olduğu görülmektedir. Bunlardan ilki 2007 yılında Southwest'te düzenlenen interaktif konferansta katılımcıların konferans boyunca birbirleriyle iletişimde bulunabilmeleri ve panellere gerçek zamanlı yorum yapabilmeleri için Twitter kullanılmıştır. O zamana kadar günlük kullanım miktarı 20.000 adet tweetken söz konusu konferans sayesinde günlük atılma tweet sayısı 60.000 e yükselmiştir. Bir günde 3 katına çıkan kullanım trafiği Twitterın tarihinde önemli yapıtaşlarından biri olmuştur. O dönemde 111 farklı mikroblog uygulaması olmasına rağmen Twitter o zamana kadarki en popüler uygulama olmayı başarmıştır.<sup>151</sup> Ertesi yıl ise; ABD'de düzenlenen başkanlık seçimlerinde adaylardan Barac Obama seçim kampanyası boyunca sosyal medya kullanımına özen göstermiş, Twitter üzerinden aynı anda 150.000 kişi ile eş güdümlü olarak haberleşmeye başlamıştır. Bu sürecin Obama'nın başkanlık seçimini kazanmasında önemli bir etkiye sahip olduğu düşünülmektedir.<sup>152</sup> Bu durum aynı zamanda Twitter için de kuruluşundan itibaren ikinci yılında önemli bir sıçrama aşaması olarak görülebilir.

---

<sup>151</sup> Estellés Enrique, Moral Esther Del ve González Fernando, 'Social Bookmarking Tools as Facilitators of Learning and Research Collaborative Processes: The Diigo Case', *Interdisciplinary Journal of E-Learning and Learning Objects*, 6 (2010).s.33

<sup>152</sup> Borges Eugene Barsky ve Purdon Michelle, 'Introducing Web 2.0: Social Networking and Social Bookmarking for Health Librarians', *JCHLA / JABSC*, 27 (2006).s.18



Twitter’da kullanıcılar ücretsiz olarak üye olup kendi profillerini oluşturur ve diğer üyelerle iletişime geçerler. Son derece basit bir ara yüze sahip olan site kullanım kolaylığı ile cezp edicidir. Twitter 50’ye yakın farklı dil seçeneğiyle faaliyetlerini sürdürmektedir. Üyeler profillerini oluştururken ilk önce kullanıcı adlarını belirlerler burada dikkat edilmesi gereken daha önceden alınmamış olan bir kullanıcı adının belirlenmesidir. Kullanıcının almak istediği ad önceden başka bir kullanıcı tarafından edinilmişse sistem o ismi başkasının almasına müsaade etmeyecektir. Kullanıcı adı kişilerin diğer kullanıcılar tarafından aranırken bulunmasını sağlayacağı için önemlidir. Ayrıca tüm kullanıcı adlarının önüne otomatik olarak @ sembolü gelecektir. Ancak kullanıcı adı zaman için de değiştirilebilmektedir. Üye olabilmek için doğrulayıcı olarak en az bir aktif mail adresinin, cep telefonu gibi erişilebilir olunan bir bilginin sunulması esastır.

Twitter’da hesap oluştururken ikinci aşama profilin oluşturulmasıdır. Kullanıcı profilini oluştururken fotoğraf, kişisel bilgiler, kendine ait web sitesi varsa web sitesinin adresi, konum gibi bilgileri isteğe bağlı olarak girebilirler. Özellikle konum bilgilerinin paylaşılması bir tweetin atıldığı dil ne olursa olsun hangi ülkeden atıldığını, aynı ülkeden atılan tweetlerin hangi bölgeden veya hangi ilden atıldığını göstermesi bakımından önemlidir. Özellikle araştırmacılar ülke, bölge veya şehir bazında analiz yaptıklarında bu bilgilere ihtiyaç duyabilmektedirler.

Açılan hesaplar tüm kullanıcıların görebilecekleri şekilde kamuya açık olabilecekleri gibi sadece onaylanmış takipçilerin görebilecekleri şekilde kısıtlı da olabilirler. Yaklaşık 320 milyon aktif kullanıcıya sahip olan Twitter hesapları içinde sadece %5’lik bir kısmı korumalı, geriye kalan %95’i ise kamuya açıktır.<sup>153</sup>

Twitter’ın günlük yaşama kazandırmış olduğu pek çok farklı terim vardır. Bu terimlerden önemli olanlar aşağıda açıklanmıştır.

### **3.3.1. Takip (Follow) / Takipçi (Follower)**

Twitter’da kişilerin bir birlerinin hesaplarına üye olmalarına “takip” ismi verilir. Kamuya açık hesaplar dileyen herkes tarafından takip edilebilir, korumalı hesapları takip edebilmek için ise önce hesap sahibinin takip isteğine onay vermesi gerekir. Diğer

---

<sup>153</sup> Zarella Dan, *Social Media Marketing Book*, O’Reilly Books, 2010. s.34

pek çok sosyal medya araçlarının aksine Twitter'da karşılıklı takip koşulu yoktur. Twitter'da en fazla takipçiye sahip olan hesap yaklaşık 108 milyon takipçi ile ABD'li şarkıcı Katy Perry'ye ait hesaptır. En fazla takipçiye sahip siyasi lider, 98 milyonu aşkın takipçisi ile ABD Eski Başkanı Barac Obama, en fazla takipçiye sahip olan sporcu 66.7 milyon takipçi ile Cristiano Ronaldo, en fazla takipçiye sahip kurumsal adres ise; 62.3 milyon takipçi ile Twitter'ın kendi hesabıdır.<sup>154</sup> Twitter'da kişileri takip etmek gibi istenmeyen kişilerin takiplerinin engellenmesi de mümkündür. Engellenen kullanıcılar yapılan paylaşımları göremez, engelleyen kişilerle Twitter üzerinden iletişim kuramazlar.

### **3.3.2. Etiket (Hashtag)**

Kullanıcıların bir konu hakkında duygu ve düşüncelerini ifade ederken konuyu vurgulamak amacıyla kullandıkları isme hashtag (etiket) denir. # sembolü ile gösterilen etiketler kişilerin o konu etrafında toplanmasını sağlar. Etiket oluşturmak isteyen kullanıcılar; seçtikleri kelimenin veya kelime grubunun başına # sembolünü koyarak süreci başlatırlar sonrasında Twitter başında # sembolü olan kelimeyi bir linke dönüştür. Bu etiket altında neler konuşulduğunu görmek isteyen kullanıcılar bu linki tıklayarak yapılan paylaşımlara erişebilirler.

### **3.3.3. Trend Konu (Trend Topic)**

Twitter'da popüler olan konular trend topic (TT) başlığı altında toplanırlar. TT o esnada Twitter'da hakkında en fazla konuşulan, paylaşımda bulunan konuları (etiketleri) gösterir.

Twitter ana sayfada en fazla hakkında paylaşım yapılan etiketler (TT'ler) sıralanmış halde yer alırlar. Bu sayede gündemi takip etmek isteyen kullanıcılar o anda sosyal medya da popüler olan konular hakkında bilgi sahibi olabilirler. Herhangi bir etiket seçildiğinde en üstte kişinin kendi takip ettiği kullanıcılar o etiketle ilgili paylaşım yapmışsa onlar daha sonra da ters kronolojik bir şekilde diğer kullanıcıların yaptıkları paylaşımlar gösterilir. Yine bir etiket seçildiğinde o etikete dair popüler tweetler, en son atılan tweetler, o etiketle ilişkili kişiler, fotoğraflar, videolar, haberler ve yayınlar başlıkları altında çeşitli listelere ulaşmakta mümkündür.

---

<sup>154</sup> Twitter Top 100 Most Followers, <https://twittercounter.com/pages/100>, Erişim tarihi, 22.12.2017

Sosyal medyanın nabzını tutmak isteyen kullanıcılar için trend topicleri takip etmek son derece önemlidir.

#### **3.3.4. Bahsetmek (Mention)**

Twitter’da bir kişi veya kuruluştan bahsedilirken bahsedilen ismin önüne @ işareti konulur. Bu sayede tweette kimden bahsedildiği gösterildiği gibi kendinden bahsedilen kişi veya kurum da atılan bu tweetten haberdar edilmiş olunur. @ işaretinin ortaya çıkışı hakkında pek çok iddia olmakla birlikte Latince “tarafına, hususunda, halinde vb.” anlamlarına gelen ‘ad’ kelimesinden türemiş olan ‘at’ kelimesinde t harfinin a harfi üzerine yuvarlanması ile elde edildiği düşünülmektedir.<sup>155</sup> Daha önceden de vurgulandığı gibi Twitter’da orijinal metin üzerinde oynama yapılamaz ancak bir kişinin atmış olduğu tweete cevap veriliyorsa cevabın kime yönelik olduğunu göstermek için de yine mention ifadesi (@) kullanıcı adının önüne konularak cevabın kime verildiği belirtilmiş olunur.

#### **3.3.5. Yeniden Tweet (Retweet)**

Başkası tarafından atılmış olan herhangi bir tweetin üzerinde hiçbir değişiklik yapılmaksızın, kullanıcının kendi sayfasında yayınlamasına retweet (RT) denir. RT edilen tweetlerde tweetin ilk olarak kim tarafından atıldığı da tweetle birlikte görülür.

#### **3.3.6. Direkt Mesaj (Direct Message)**

Twitter kullanıcılarının kendi aralarında özel olarak mesajlaşması uygulamasına direct message (DM) denir. Bu mesajlar kamuya açık değildir sadece mesajlaşan kişiler tarafından görülmektedir. Geçmişte iki kullanıcının Twitter üzerinde DM yollayabilmeleri için her ikisinin de birbirlerini takip etmeleri gerekirken günümüze taraflardan sadece birini diğerini takip ediyor olması DM yollamak için yeterlidir.

#### **3.3.7. Troll Hesap**

Sosyal medya ortamlarında genellikle, bilgi kirliliği yaratmak, polemik çıkarmak, insanların huzurunu kaçırmak amacıyla tartışma yaratmak veya algı oluşturmak amacı güden kişilere troll denir. Troll hesaplar belli bir grup tarafından oluşturulabileceği gibi bir kişi tarafından da birden çok hesap oluşturularak

---

<sup>155</sup> <http://blog.milliyet.com.tr/----ne-demek-hic-dusundunuz-mu-/Blog/?BlogNo=20467>

kullanılabilirler.<sup>156</sup> Twitter’da troll hesaplar genellikle algı oluşturmak amacıyla, aynı mesajı dakikalar içinde defalarca farklı kullanıcılar tarafından atılmış gibi tekrar tekrar atarak yayarlar.

### 3.4. Sosyal Medyanın Önemi

Bilgisayar teknolojilerini hızla gelişmesinden medya da payına düşeni almıştır. Geleneksel medya araçları olan televizyon, gazete, dergi gibi ortamlar, karşılarında yeni medya denilen sosyal medyayı bulmuştur.

Geleneksel medyanın aksine sosyal medya sayesinde artık tüm kullanıcılar birer içerik üreticisidir. Dolayısıyla tüm sosyal medya kullanıcıları, amatörce olsa da; aynı zamanda bir haberci, aynı zamanda bir kameraman, aynı zamanda kurgucu, aynı zamanda senarist olarak karşımıza çıkmaktadır.

Geleneksel medya araçları büyük maddi yatırımlar gerektirirken sosyal medyanın böyle bir maddi yatırımı gerçekleştirilmesine ihtiyacı yoktur, bu da sosyal medyayı bir adım öne çıkarmaktadır.

Sosyal medya, sokaktaki vatandaşın gücü konumuna yükselmiştir. Bu sayede insanlar kolaylıkla organize edilebilmekte, olaylardan haberdar olmaları sağlanabilmekte ve hızlı bir şekilde yönlendirilebilmektedir. Bunun pek çok örneği gerek ülkemizde gerekse dünyada yaşanan sosyal olaylarda görülmüştür. Bu olayların ortak özellikleri sosyal medyanın gücünü gösterebilmesi açısından önemlidir, çünkü bu olayların büyük bir kısmı “Facebook Devrimi”, “Twitter Devrimi” veya “İnternet Devrimi” olarak adlandırılmıştır.<sup>157</sup>

Dünyada yaşanan örneklerden en önemlileri; 2004 yılında Ukrayna’da gerçekleşen turuncu devrim denilen süreç, 2010 yılından itibaren Arap Ülkelerinde gerçekleşen ve Arap Baharı denilen süreç, Türkiye’de gerçekleşen gezi olayları ve 15 Temmuz darbe girişimi süreci sosyal medyanın gücünü ortaya koyması açısından önemlidir. Aynı şekilde ABD’de yapılan başkanlık seçimleri de sosyal medyanın gücünü göstermesi açısından önemli bir yere sahiptir. Kısaca bu süreçler açıklanarak sosyal medyanın gücü ortaya konmaya çalışılacaktır.

<sup>156</sup> <https://www.tech-worm.com/troll-nedir-troll-hesap-ne-demektir/> erişim tarihi; 25.02.2018

<sup>157</sup> TelliAzime, ‘Mısır Devrimi’nde Sosyal Medyanın Rolü’, *Bilge Strateji*, cilt 4,sayı 7 Güz (2012). s. 65-66

31 Ekim 2004 yılında Ukrayna’da yapılan başkanlık seçimleri Batı’nın desteklediği aday Viktor Yuşçenko ile Rusya’nın desteklediği aday Viktor Yanukoviç arasında kıran kırana geçmiştir, ilk turda birbirlerine çok yakın oy alan iki aday ikinci turda en yüksek oyu alabilmek için çekişmiştir. İkinci turda Yanukoviç oyların yüzde 49,42’sini alırken Yuşçenko 46,69’unu almıştır. Ancak Yuşçenko taraftarları kimi bölgelerde seçimlere hile karıştırıldığını iddia ederek eylemler yapmıştır.<sup>158</sup> Eylemlere turuncu kıyafet giyerek katılan bu topluluk sosyal medya üzerinden organize olmuş, sosyal medya ile seslerini duyurarak daha fazla taraftar toplamıştır.

2009 yılı Haziran ayında İran’da yapılan başkanlık seçimini Mahmud Ahmedinejad kazanmıştı ancak muhalif gruplar seçimlere hile katıldığını düşünmekte idi. Twitter üzerinden örgütlenen gruplar protesto gösterileri düzenlemiştir. Devlet tarafından protesto gösterilerinin bastırılması için güç kullanılmış bunun yanı sıra geleneksel medya araçlarına sansür uygulanmış, dünyanın önde gelen basın kuruluşlarının çalışanları sınır dışı edilip basın kartları iptal edilmiştir, internet yasaklanmıştır. Fakat sosyal medyayı etkin olarak kullanan gruplar DNS ayarlarını değiştirerek internet üzerinden seslerini duyurmayı başarmış, organize olmak için #iranelection etiketi ile Twitter’ı kullanmıştır. Twitter’in yoğun olarak kullanılması nedeniyle bu sürece “Twitter Devrimi” denilmiştir. Her ne kadar bu protestolar başarısızlıkla sonuçlansa da burada devrim olarak görülen şey sosyal medyanın geleneksel medya karşısında ki başarısıdır.<sup>159</sup>

2010’un sonlarından itibaren zincirleme bir reaksiyon gibi Arap ülkelerinde art arda toplumsal olaylar patlak vermiştir. 2011 yılı Ocak ayında Tunus’ta başlayan olaylar hızla diğer Arap ülkelerine de yayılmış ve bu süreç Arap Baharı olarak anılmıştır. Arap Baharı sürecinde de sosyal medyanın etkin bir şekilde kullanıldığı görülmüştür. Arap Baharının başladığı yıllarda henüz Twitter’da sağdan sola doğru yazmaya destek yoktu bu yüzden de aktivistler İngilizce tweetler atmıştır.<sup>160</sup> Yinede bu

<sup>158</sup> <http://akademikperspektif.com/2013/04/22/ukraynadaki-turuncu-devrim/>, erişim 22.01.2018

<sup>159</sup> Bruns Axel, Highfield Tim ve Burgess Jean, ‘The Arab Spring and Social Media Audiences: English and Arabic Twitter Users and Their Networks’, *American Behavioral Scientist*, 2013, 57/7 s. 3

<sup>160</sup> Göçoğlu Volkan ve Aydın Mehmet Devrim, ‘Kamu Politikası Ve Sosyal Medya İlişkisinin Toplumsal Hareketler Bağlamında İncelenmesi’, *Uluslararası Sosyal Araştırmalar Dergisi*, 8.37 2015 s.894.

süreçte Mısır'daki Twitter kullanıcı sayısı hızla artmış ve 2011 Mart ayına gelindiğinde Mısır'da Twitter kullanıcı sayısı 1.100.000'i geçmiştir.<sup>161</sup>

Ukrayna'da 2013 yılı Kasım ayında başlayan ve 2014 yılında şiddetlene olaylar “Turuncu Devrim” olarak adlandırılmıştır. Hükümetin politikalarını protesto amacıyla düzenlenen protestolar ve sonrasında gelişen süreçte olaylar uluslararası arenada çok kısa süreçte yankılanmıştır. Bunun temel nedeni yine sosyal medya, özellikle de Twitter'ın yaygın olarak kullanılması vardır. Olayların en şiddetli şekilde yaşandığı gün olan 18 Şubat 2014'te atılan tweet sayısı rekor düzeye ulaşmıştır. Söz konusu günde Ukrayna'da saat başına ortalama 30.000 tweet atılmıştır. Ukrayna'da olan biteni dünya kamuoyuna duyurma amacı güdüldüğü için atılan tweetlerin %53'ünün İngilizce olduğu görülmüştür.

ABD başkanlık seçimlerini Barac Obama'nın seçildiği 2009 yılı ABD Başkanlık seçiminden bu yana kazanan aday her zaman sosyal medyayı en etkin kullanan aday olmuştur. Barac Obama gibi Donald Trump'ta seçim sürecinde sosyal medyayı etkin bir şekilde kullanmış, seçim günü henüz oylama yapılırken, Twitter verileri üzerinden yapılan analizle seçimi Donald Trump'ın kazanacağı ortaya konulmuştur.<sup>162</sup>

Türkiye'de sosyal medyanın dolayısıyla Twitter'ın ön plana çıktığı ilk süreç, Gezi Parkı olaylarıdır. 27 Mayıs 2013 Pazartesi sabahı Gezi Parkı'na giren dozerlerin parktaki ağaçları sökmeye başlamaları karşısında Taksim Dayanışma Platformu üyeleri Gezi Parkı'nda toplanmış ve ağaçların sökülmemesi için oturma eylemi başlatmıştır. 28 Mayıs'ta oturma eylemi yapanlara karşı polislerin sert müdahalesi barışçıl bir eylem için alanda bulunanların tartaklanması olayların büyümesine neden olmuş, sosyal medyada örgütlenen halk protestolara destek vermeye başlamıştır. Haziran ayının ilk günlerinde ise olaylar daha da büyüyerek hem ülkenin hem de dünyanın gündemi haline gelmiştir. Gezi Parkı Olaylarının ülke ve dünya gündemine gelmesinde ve uzun süre gündemde kalmasında sosyal medyanın rolü çok büyüktür. Keza aynı dönemde ana akım medya kuruluşları başta olmak üzere pek çok geleneksel medya organı olayları görmemezlikten

---

<sup>161</sup> Telli s.69.

<sup>162</sup> Oğuzlar Ayşe ve Kızılkaya Yusuf Murat, 'Is Donald Trump's Presidential Election Victory A Surprise According to Twitter Data', in *12th International Conference on Social Sciences* (Euser European Center For Science Education And Research, 2017), p. 568.

gelmiş, sosyal medya sayesinde dünya Gezi Parkı'nda olan olaylardan haberdar olmuştur.

YNKLABs (büyük veri yönetimi ve Lingustik Yapay Zeka üzerin çalışan bir yapıdır) hazırladığı Gezi Parkı Infografik isimli raporda Gezi Parkı Olayları sürecinde sosyal medyanın rolünü ortaya koymuştur. Raporda Gezi Parkı Olayları'nın en şiddetli dönemleri olarak 31 Mayıs – 24 Haziran 2013 yılları arasındaki verilere yer verilmiştir. Verilere göre söz konusu dönemde Twitter'ın gündeminde çok büyük bir değişiklik olmuştur. Normal dönemlerde (1 Mayıs – 31 Mayıs arasında) özel bir televizyon kanalında yayınlana Survivor isimli program hakkında dönem başına 44.471 adet tweet atılırken, 31 Mayıs – 24 Haziran tarihleri arasında söz konusu program hakkında bölüm başına 7.191 tweet atılmıştır. Yine aynı dönemde Twitter'da Gezi Parkı Olaylarıyla ilgili 1.385.733 tekil kişi tarafından 5.614.018 adet tweet atılmıştır ve bu sayıda retweetler (retweet sayısı sadece 1.182.613 tür) dahil değildir. Atılan tweetlerin görüntülenme sayısı 1.117.110.000'dir. Rapora göre aynı dönemde en fazla kullanılan hashtaglar ise; #direngezi, #occupygezi, #direnaksim olmuştur.<sup>163</sup> Raporu hazırlayan YNKLABs'ın verdiği bilgiye göre analizler online olarak yapılmıştır, bu yüzden de tüm gün verileri değil günün belli bir kesiti raporda yer almıştır. Gezi Parkı Olayları'nın en şiddetli olduğu gün olan 31 Mayıs 2013'te #direngezi dünyada en çok tweet atılan hashtag olmuştur.

Türkiye'de sosyal medyanın gücünü ortaya koyan önemli bir olayda 15 Temmuz 2016'da gerçekleşen darbe girişimidir. Darbe girişiminin engellenmesinde belki de geleneksel medyadan daha çok sosyal meydanında büyük bir payı vardır. 15 Temmuz 2016 gecesi tüm Türkiye tek bir yürek olarak darbe girişimine karşı direnmiş, demokrasisine sahip çıkmıştır. Darbe girişimini önleme de sosyal medyadan paylaşılan haberlerin, toplanma ve direniş çağrılarının büyük bir rolü olmuştur. Normal günlerde ortalama olarak 6 milyon civarında Türkçe tweet atılırken darbe girişimin ortaya çıktığı 15 Temmuz 2016 saat 22:00 dan itibaren Twitter kullanımı hızla artmış ve 16 Temmuz 2016 günü atılan Türkçe tweet sayısı %223 artmıştır. 15 – 17 Temmuz 2016 tarihleri

---

<sup>163</sup> Oğuzlar Ayşe ve Kızılkaya Yusuf Murat, 'Is Donald Trump's Presidential Election Victory A Surprise According to Twitter Data', in *12th International Conference on Social Sciences* (EUSER EUROPEAN CENTER FOR SCIENCE EDUCATION AND RESEARCH, 2017), p. 568.

arasında atılan tweet sayısı 34 milyon 818 bin 329, bu tweetlerin görülme sayısı ise 15 milyar olmuştur.<sup>164</sup>

Sosyal medyanın gücü sadece toplumsal olaylarda değil, turizm sektörü gibi farklı alanlarda da karşımıza çıkmaktadır. Ankara – Kars arası yıllardır sefer yapan “Doğu Ekspresi” treninin son yıllarda popüler bir hale gelmesi de sosyal medyanın gücünü ispat açısından önemlidir. Yöre halkının sıklıkla kullandığı yaklaşık 25 saatte Ankara’dan Kars’a ulaşmayı sağlayan Doğu Ekspresi treninde özellikle kış aylarında yer bulmak neredeyse imkânsız bir hale gelmiştir. Bunun en önemli nedenleri gezginleri sosyal medyada paylaşmış oldukları video ve fotoğraflardır. Yapılan paylaşımlar sayesinde Kars’a Doğu Ekspresi ile özellikle kış aylarında seyahat etmek bir trend haline gelmiş, seyahat etmek isteyenler boş yer bulabilmek için haftalarca beklemek zorunda kalmaktadır.

Sosyal medyanın gücü her zaman olumlu yönde kullanılmamaktadır. Yapılan araştırmalar sosyal medyada yalan haberlerin kendilerine doğru haberlerden daha fazla yer bulduklarını göstermektedir. Kullanıcıların yalan olan bir haberi paylaşma ihtimali doğru haberden %70 daha fazla, yalan haberin doğru habere kıyasla yayılma hızı 6 kat fazladır.<sup>165</sup>

Günümüzde sosyal medya geleneksel medyadan daha yaygın bir kullanım alanına sahiptir. Akıllı telefona sahip olan bir kişi aynı anda hem bir üretici hem de tüketici olarak sosyal medya da yer almaktadır. Sosyal medya hem olumlu hem de olumsuz yönde kullanılabilen etkili bir araçtır. Siyasetçisinden pazarlamacısına kadar toplumda karar verici olan hemen herkes sosyal medyanın gücünün farkındadır ve geleneksel medyadan çok sosyal medyaya yatırım yapılmaktadır. Geleneksel medya kanalları çeşitli yöntemlerle belli bir grubun elinde toplanabilir ancak sosyal medyanın bu şekilde belli grupların ellerinde toplanması mümkün değildir. Bu da sosyal medyayı ön plana çıkaran önemli unsurlardan birisidir. Geleneksel medya için çok büyük maddi yatırımlar gerekirken sosyal medyada üretim ve kullanım çok daha ekonomiktir, sosyal medyanın artıları nedeniyle geleneksel medyanın ilerleyen yıllarda sosyal medya ile

---

<sup>164</sup> Akçay Ekrem Yaşar, ‘15 Temmuz Darbe Girişimi Çerçevesinde Dijital Diplomasinin Kullanımı’. Darbe Sempozyumu, Aydın Adnan Menderes Üniversitesi, 2018

<sup>165</sup> <http://www.dw.com/tr/sosyal-medyada-yalan-haber-daha-%C3%A7ok-tutuluyor/a-42902833>, erişim tarihi 05.03.2018



rekabet edebilmek için farklı araçlara ihtiyaç duyacaktır. Tüm geleneksel medya araçlarının aynı anda sosyal medya hesaplarına sahip olması ve bu sosyal medya hesaplarını etkin bir şekilde kullanması bunun önemli bir kanıtıdır.



## DÖRDÜNCÜ BÖLÜM

### UYGULAMA: R İLE DUYGU ANALİZİ

Bu çalışmanın amacı; 24 Haziran 2018 de Türkiye’de yapılan Cumhurbaşkanlığı ve 27. Dönem Millet Vekilli Genel Seçimi’ne ilişkin, seçimlerden önce atılan tweetlerde duygu analizi yapmaktır. Literatürde yapılan çalışmaların çok büyük bir kısmının İngilizce tweetler üzerine olduğu görülmektedir. Bunun en temel nedenleri İngilizce’nin yapısı nedeniyle analizin kolay olması ve İngilizce olarak hazırlanmış olan pozitif, negatif duygu ifadelerini içeren korpusların var olmasıdır. Türkçe analiz yapabilmek için hazır bir korpus bulunmamaktadır. Bu yüzden analiz yapılmadan önce Türkçe’de pozitif duygu ifade eden kelimeler, negatif duygu ifade eden kelimeler ve bunlara ek olarak İngilizcede stopwords olarak ifade edilen cümleye duygu anlamında bir katkısı olmayan kelimelere karşılık gelen durak kelimeleri için birer korpus oluşturulmuştur. Stopwords Türkçeye “durak kelimeler” olarak çevrilebilir, bunlar cümlede bağlaç gibi çeşitli görevleri yerine getirmekle birlikte, duygu anlamında bir değeri olmayan kelimeledir. Oluşturulan bu korpuslara herkesin ulaşımını kolaylaştırabilmek için kizilkayamurat.com isimli web sitesi açılmış ve korpuslar bu siteye yüklenmiştir. Bu sayede ilerleyen dönemlerde Türkçe olarak duygu analizi yapmak isteyen kullanıcılar hazır bir korpusa erişebilecek hatta bu korpusu geliştirebilecektir.

Çalışmada R yazılımı kullanılmıştır. Çalışmada kullanılan paketler daha güncel olduğu düşünülen “France (Lyon 1)” konumundan elde edilmiştir. R de çalışma yapılırken, önce paketlerin indirilmesi ve sonra da indirilen paketlerin kütüphaneden çağrılması gerekir. Gerek paketleri indirirken gerekse kütüphaneden çağırırken tüm komutların doğru yazılması R’nin çalışması için önemlidir, örneğin paketler yüklenirken, parantez içinde paket ismi çift tırnak arasında gösterilmek zorundadır; paketler kütüphaneden çağırılırken çift tırnak kullanılmamalıdır. Çalışmada temel olarak tm (textmining) paketi kullanılmakla birlikte, gerek analizler gerekse görselleştirme için yardımcı olarak pek çok paket kullanılmıştır. Kullanılan paketlerin tamamı ekler kısmında ayrıntılı olarak gösterilmiştir.

#### 4.1. Seçim Sonuçları İçin Geleneksel Tahmin Yöntemi

Tüm dünyada seçimlerden önce hangi parti veya adayın hangi oranda oy alacağı merak konusudur. Adayların veya partilerin alacağı oy oranlarını tahmin edebilmek için geleneksel olarak kullanılan yöntem kamuoyu yoklamalarıdır. Alanında uzman kuruluşlar tarafından, oy kullanma yeterliliğine sahip kişilerden oluşturulan bir örneklemden hareketle adayların veya partilerin hangi oranda oy alacakları tahmin edilmeye çalışılır. Bu işlem yapılırken çıkarımsal istatistik kullanılır ve örneklemden hareketle anakütleye dair tahminler yapılır.

Türkiye’de yapılan geleneksel yöntemlerde; anket firmaları genellikle AB istatistik ofisi tarafından bölgesel bazda istatistik toplayabilmek amacıyla oluşturulan Türkiye’deki İstatistiksel Bölge Birimleri’ni (İBBS) (Fransızcası; Nomenclature d’Unités Territoriales-NUTS)<sup>166</sup> baz alarak tabakalı örnekleme yöntemi ile tahminler yürütmeye çalışmaktadır. İBBS sisteminde Türkiye ekonomik, sosyal, coğrafi olarak yakınlık vb kriterlere göre; düzey 1 olarak 12 alt gruba ayrılmış, bu 12 alt grup düzey 2 adında 26 alt düzeye ayrılmıştır. Günümüzde yapılan çalışmaların hemen hemen hepsin de bu 26 alt düzeye hareketle örneklemler seçilmektedir.

Geleneksel yöntemlerle seçim sonuçlarının tahmini amacıyla yapılan çalışmalarda, genellikle 21 – 36 arası şehirden, 2.000 - 4.250 kişilik bir örneklem seçilmekte ve seçilen örneklemden hareketle tahminler yapılmaktadır. Ancak geleneksel yöntemlerle seçim sonuçlarını tahmin etmek gittikçe zorlaşmaktadır. Bunun temel nedenlerinde biri; ankete katılanların kişisel bilgilerini paylaşmaktan kaçınmaları veya fişlenme kaygısı taşımalarıdır. Önemli araştırmacı firmalarından biri olan ve son yıllarda yapılan seçimlerin sonuçlarını doğru tahmin ettiklerini iddia eden Gezici Araştırma firmasının sahibinin vermiş olduğu mülakatlardan birinde daha önce 40 – 45 hane ziyaret edilerek 18 anket formu doldurulabilirken son dönemde 120 hane ziyareti ile 18 anket formunun ancak doldurulabildiğini iddia etmektedir.<sup>167</sup> Geleneksel yöntemlerle elde edilen sonuçlara ilişkin bir başka sorun da katılımcıların çeşitli sebeplerle cevap vermekten kaçınması veya gerçeğe aykırı beyanda bulunmasıdır.

<sup>166</sup> <https://ipfs.io/ipfs/tr.wikipedia-on-ipfs.org/wiki/NUTS.html> erişim tarihi; 06.06.2018

<sup>167</sup> <http://t24.com.tr/haber/gezici-secmen-konusmuyor-18-anket-icin-40-kapiya-giderdik-simdi-120-bu-dip-dalgasi.638931> erişim tarihi; 06.06.2018

2018 yılında yapılacak olan Cumhurbaşkanlığı ve 27. Dönem Millet Vekilli Genel Seçimi'ne ilişkin bazı kamuoyu araştırma firmalarının, seçimlerde en yüksek oy alması beklenen adaylar ve partilere ilişkin, seçimlerden önceki son tahminleri aşağıdaki tablolarda gösterilmiştir.

**Tablo 4.1.** Kamuoyu Araştırma Firmalarının Cumhurbaşkanlığı Seçimine İlişkin Son Tahminleri

Adaylar Firmalar	Recep Tayyip ERDOĞAN	Muharrem İNCE	Meral AKŞENER	Selahattin DEMİRTAŞ
SONAR	%48,3	%31,4	%9,5	%8,2
GENAR	%51,5	%28,4	%7,6	%10,5
METROPOL	%46,2	%24,5	%9,2	%11,3
KONDA	%51,9	%28,0	%10,2	%9,0
GEZİCİ	%48,0	%29,1	%11,4	%10,0
OPTİMAR	%51,6	%28,0	%8,5	%10,3
MEDİAR	%47,9	%29,8	%9,15	%10,1
A&G	%53,5	%29,1	%8,1	%8,8
MAK	%51,5	%24,4	%12,5	%8,8
ORC	%52,7	%29,2	%8,6	%8,4
VERA	%48,4	%28,8	%10,3	%11,7
ARGETUS	%51,3	%28,4	%8,1	%9,9
TÜSİAR	%50,7	%24,9	%11,7	%9,7

Tablo 4.1. incelendiğinde; Recep Tayyip ERDOĞAN'ın alacağı oy oranını en düşük tahmin eden firma %46,2 ile METROPOL iken, en yüksek oy oranını %53,5 A&G yapmıştır. Muharrem İNCE'nin oy oranları ise %24,4 ile %31,4 arasında çeşitli firmalarca tahmin edilmiştir. Meral AKŞENER'in alacağı oy oranları %7,6 ile %12,5, Selahattin DEMİRTAŞ'ın alacağı oy oranları ise %8,2 ile %11,7 arasında tahminler yapılmıştır.

Yüksek Seçim Kurulu (YSK) tarafından açıklanan seçim sonuçlarına göre; Recep Tayyip ERDOĞAN %52,6, Muharrem İNCE %30,6, Meral AKŞENER %7,2,

Selahattin DEMİRTAŞ ise %8,4 oy almıştır.<sup>168</sup> Kamuoyu arařtırma firmalarının genel performansları kıyaslanınca Cumhurbaşkanlığı seçimi için en etkili tahminleri; KONDA, A&G ve ORC firmalarının yaptığı görölmektedir.

**Tablo 4.2.** Kamuoyu Arařtırma Firmalarının 27. Dönem Millet Vekili Seçiminde Siyasi Partilerin Oy Oranlarına İliřkin Son Tahminleri

Partiler Firmalar	AK PARTİ	CHP	HDP	MHP	İYİ PARTİ
SONAR	%42,2	%28,2	%10,0	%7,1	%9,2
GENAR	%44,7	%26,1	%11,6	%6,1	%8,9
METROPOL	%48,0	%21,0	%10,0	%6,0	%12,0
KONDA	%45,5	%26,1	%11,6	%7,3	%8,5
GEZİCİ	%43,6	%26,7	%12,1	%5,3	%10,1
OPTİMAR	%46,7	%26,3	%11,2	%6,9	%6,8
MEDİAR	%41,5	%26,2	%12,1	%5,8	%10,7
A&G	%45,6	%25,5	%11,3	%8,8	%8,3
MAK	%43,6	%24,2	%9,2	%9,2	%10,3
ORC	%44,2	%24,5	%11,9	%10,5	%7,4
VERA	%43,6	%27,4	%12,3	%7,3	%8,2
ARGETUS	%44,8	%26,4	%10,3	%7,2	%8,3
TÜSİAR	%43,8	%25,5	%9,2	%7,5	%10,3

Tablo 4.2. incelendiğinden; Ak Partinin oy oranı çeřitli firmalar tarafından %41,5 - %48 arasında, CHP'nin oy oranı %21,0 - %28,2, HDP'nin oy oranı %9,2 - %12,3, MHP'nin oy oranı %5,3 - %10,5 ve İyi Parti'nin oy oranı %7,4 - %12,0 arasında tahmin edilmiştir.

YSK tarafından açıklanan seçim sonuçlarına partilerin oy oranları; Ak Parti %41,8, CHP %22,4, HDP %11,7, MHP %10,9 ve İyi Parti %9,9 oy almıştır.<sup>169</sup> Arařtırma firmalarının genel anlamda tüm partilerin oy oranlarını doğru tahmin edebilmek gibi bir başarı göstermedikleri ortaya çıkmıştır. SONAR ve MEDİAR Ak

<sup>168</sup> [https://www.ntv.com.tr/turkiye/ysk-kesin-olmayan-secim-sonuclarini-acikladi.9y2BK2\\_Y8kyM0-4-3rZeZA](https://www.ntv.com.tr/turkiye/ysk-kesin-olmayan-secim-sonuclarini-acikladi.9y2BK2_Y8kyM0-4-3rZeZA), erişim tarihi; 27.06.2018

<sup>169</sup> [https://www.ntv.com.tr/turkiye/ysk-kesin-olmayan-secim-sonuclarini-acikladi.9y2BK2\\_Y8kyM0-4-3rZeZA](https://www.ntv.com.tr/turkiye/ysk-kesin-olmayan-secim-sonuclarini-acikladi.9y2BK2_Y8kyM0-4-3rZeZA), erişim tarihi 27.06.2018

Parti'nin oy oranını diğer firmalara nazaran daha etkili tahmin etmiştir, METROPOL CHP'nin oy oranını daha etkili, HDP'nin oy oranı genelde doğru tahmin edilmiş, MHP'nin oy oranını sadece ORC etkin olarak tahmin edebilmiş, İyi Parti'nin oy oranını ise en etkin SONAR ve GEZİCİ firmaları tahmin edebilmiştir.

Kamuoyu araştırma firmalarının her iki seçim için performanları kıyaslanınca Cumhurbaşkanlığı seçim sonuçlarına dair daha isabetli tahminler yürüten firmala olduğu ancak firmaların bu başarılarını partilerin alacakları oy oranlarını tespit etmede sürdüremedikleri ortaya çıkmıştır.

Çalışmanın temel çıkış noktası geleneksel yöntemler dışında sosyal medya verilerinden hareketle duygu analizi yapıp, seçim sonuçlarına ilişkin çıkarımlar yapmaya çalışmaktır. Çalışmada sosyal medya verileri kullanılmıştır çünkü katılımcıların sosyal medya paylaşımlarında manipülasyonda bulunmayacakları ve kendilerini bir anket cevaplar gibi değilde daha rahat ifade edeceği cümleler kurması beklenmektedir.

#### **4.2. Verilerin Twitter'dan Çekilmesi**

Günümüzde sosyal medya alanında büyük veri kaynaklarından en önemlisi Twitter'dır. Analiz için kullanılan veriler Twitter'dan çekilmiştir. Twitter'dan veri elde etmek için açık bir Twitter hesabın olması ve veri çekebilmek için Twitter'dan bir nevi izin alınması gerekir. Bu amaçla API (Application Programming Interface – Uygulama Programlama Arayüzü) kullanılır. API bir uygulamaya ait yeteneklerin başka bir uygulamada kullanılabilmesi için oluşturulmuş olan arayüzdür. Twitter'daki yeteneklerin API ile bir başka uygulama da kullanılabilmesi için; <https://apps.twitter.com/> adresine girerek aşağıdaki adımları izleyip gerekli izinler alınır.

1. Create New App butonu tıklanarak açılan sayfada yer alan bilgiler doldurularak “Create your Twitter Application” sekmesi tıklanır.
2. İkinci aşamada Consumer Key'in karşısında bulunan “Key and Access Tokens” sekmesi tıklanarak Twitter tarafından kişiye özel olarak oluşturulan “Consumer Key (API Key)” ve “Consumer Secret (API Secret)” bilgilerine erişilir. Bu bilgilerin not edilmesi gerekir.

3. Your Access Token başlığı karşısında yer alan “Create my access” sekmesi tıklanır. Bu sayfada yer alan “Access Token” ve “Access Token Secret” değerleri not edilir.

Bu işlemler sonucunda dört adet bilgi elde edilmiş olmaktadır. Bu bilgilerin tamamı Twitter’a bağlanıp veri çekmek için kullanılır.

R’den Twitter’a bağlanıp tweet çekmek için iki pakete ihtiyacımız var. Bunlar; Twitter’ın web API arayüzüne bağlanabilmek için twitteR paketi ve kendi Twitter hesabımıza Twitter üzerinden bağlanabilmek için ROAuth paketidir.

Twitter’dan veri çekebilmek için öncelikle twitteR paketinin yüklenip, kütüphaneden çağrılması gerekir. Aynı şekilde kendi hesabımıza R üzerinden erişebilmek için de ROAuth paketi yüklenerek çağrılmalıdır. Daha sonra ROAuth paketinin hazır bir fonksiyonu olan `setup_twitter_oauth()` kodu kullanılarak R Twitter’a bağlanılır. Bu kod içine sırasıyla; `costumer_key`, `costumer_secret`, `access_token` ve `access_secret` değerleri girilir. Tüm bu işlemler tamamlandıktan sonra R Twitter’dan veri çekme adına yetkilendirilmiştir, “Using direct authentication” uyarısı artık Twitter’dan veri çekilebilceğini göstermektedir.

Twitter’dan istenilen konu hakkında veya tek bir kullanıcının attığı tweetleri çekebilmek gibi alternatifler mümkündür. Bunun yanı sıra istenilirse sadece belli bir dilde atılan tweetler çekilebilir, bunun için komutun sonunda dil seçeneğini eklemek yeterlidir.

Çalışmada hem Cumhurbaşkanlığı adaylarına ilişkin hem de milletvekilliği seçimine giren partilere ilişkin atılan tweetler, aday listelerinin kesinleştiği 30 Mayıs’tan seçimin yapıldığı 24 Haziran’a kadar düzenli olarak çekilip analiz edilmiştir. Cumhurbaşkanlığına toplam 6 aday girmekte, en yüksek oy potansiyeline sahip olan 4 aday için tweetler çekilmiştir. Milletvekilliği seçimi için 8 farklı siyasi parti seçime girmekte, bunlardan en yüksek oy potansiyelin sahip olan 5 parti için tweetler çekilmiştir. Çekilen toplam tweet sayısı; 720.000’dir. Her bir aday ve parti için 80.000 adet tweet çekilmiştir. Çekilen tweet sayısındaki en önemli kısıt kullanılan bilgisayarın kapasitesidir. R yazılımı bilgisayarın RAM ismi verilen belleğini kullanmaktadır,

işlemlerini RAM üzerinden yapmaktadır, bu yüzden de kullanılan bilgisayarın RAM kapasitesi işlenecek verinin hacmini belirlemede önemlidir.

### 4.3. Verilerin Ön İşlemesi

Twitter'dan çekilen veriler ham halleri ile çok fazla anlam ifade edemeyebilir çünkü içinde işe yaramayacak bir sürü kısaltma, sembol, fotoğraf, video gibi öğeleri bol miktarda barındırır. Veriler analiz edilmeden önce titiz bir ön işleme aşamasına tabi tutulmalıdır. Ön işleme aşaması çok uzun süren, sabır isteyen bir aşamadır ve analizin etkinliği açısından önemlidir.

Verilerin ön işleme yapılmadan önceki ham haline ilişkin örnek aşağıdaki gibidir. R yazılımında head komutu ilk altı satırı (veriyi) göstermektedir bu yüzden her bir grup için sadece ilk altı satır verilmiştir.

Recep Tayyip ERDOĞAN'a ilişkin atılan ham tweetler;

[1] "RT @HuseyinTuna\_tr: Muharrem İnce, Tayyip'in diplomasını göstermesini istemişti. Yandaş yalaka televizyon kanalında, aile boyu iktidar yala..."

[2] "RT @tugceseckin34: Tayyip 2.Tura İNCE'yi istiyor.\r\r\nİNCE'ye tüm medyalar sonuna kadar açık.16 yıldır SOLA laf ede ede seçim kaybetmedi.Tek um..."

[3] "RT @GizliOperasyon7: Recep Tayyip ERDOĞAN : Afrin'den sonra diğer terör bataklıklarını da kurutmak için gerekli adımları atmaya başladık..."

[4] "RT @AllBlg: Tayyip emeklilere 'gençlere eskiyi anlatın' demişti.\r\r\nAkşener'i dinlerken akla 90'lardaki siyasi üslup ne nezaket geldi.\r\r\n#MeralA..."

[5] "RT @kmIzn: Tayyip Erdoğan'ın bu ülkeye en büyük hizmeti abdestsiz gezmeyen bir maarrem kazandırmış olmasıdır.\r\r\nAdam daha naapsın he naapsın,..."

[6] "RT @SERHATLGEN4: Sayın @ErbakanFatih böyle bir zamanda hakkın yanında yer alarak Recep Tayyip Erdoğan ı desteklediğinizi açıklamanız Türk M..."

Muharrem İNCE'ye ilişkin atılan ham tweetler;

[1] "Ben anket falan anlamam ..\r\r\nMersin-Yenişehir ilçesini en az 10 kez dolaşıp el sıkılmış bir siyasetçi olarak söylüyorum... <https://t.co/crkjrhmxvB>"





[1] "RT @kadiratalaypiar: 33-Selahattin Demirtaş seçmene çok güzel hitap eden bir aday. Seçmen tarafından çok sevilen bir isim. Dediğim gibi sol..."

[2] "Selahattin Demirtaş bile MİT'i kapatacağım demedi. Doğu Perinçek bile TİKA'yı kapatacağım demedi. Çok ilginç bir vaka ile karşı karşıyayız."

[3] "#kılıçdaroğlunasoruyorum Enis Berberoğlu gibi Sn.Selahattin Demirtaş ta siyasi ESİR neden hiçbir zaman Selahattin D... <https://t.co/pfLmxZU2F1>"

[4] "RT @cem\_oezdemir: Der Horror in der #Türkei hat noch gar nicht begonnen – @dlfkultur berichtet über die Solidaritätslesung für @hdpdemirtas..."

[5] "RT @Fatma99848359: Selahattin Demirtaş'ın 6-7-8 Ekim Olaylarında Halkı Sokağa Çağırdığının İspatı <https://t.co/wJMHaKWAf3>\r\r\r\r\r#VakitTürkiyeVak..."

[6] "RT @Deli\_dalga\_ : Mutluluk paylaşınca güzel\r\r\r\r\rSeçim öncesi güzel bir haber <U+0001F447>\r\r\r\r\rBarzaniye yakın kaynaklar: #TSK'nın Kandile yaptığı hava saldır..."

Ak Parti'ye ilişkin atılan ham tweetler;

[1] "RT @mburakkizilelma: Bakanımız Sn @abdulhamitgul ile Nurdağı Belediyemizi ve Ak parti Nurdağı ilçe merkezimizi ziyaret ettik. \r\r\r\r\rNurdağı teş..."

[2] "RT @ESAREA: Ak Parti iktidarında aile ve kadın için yapılan düzenlemeleri sizlerle paylaşacağım. 1) Anayasa'da kadın-erkek eşitliği ilkesi..."

[3] "Ak Parti'nin Diyarbakır'da olduğu gibi Batman'da da işi çok zor. HDP dışında karşısında muhafazakâr kesimden Hüda P... <https://t.co/pIHxcd83of>"

[4] "RT @fatihtezcan: RT <ed><U+00A0><U+00BD><ed><U+00B4><U+00B4> UYARI: HDP Barajı geçemiyor. Ak Parti'ye 60 Mv. daha geliyor.\r\r\r\r\rDurumu gören CIA-FETÖ harekete geçiyor.\r\r\r\r\rKendine Milli..."

[5] "RT @cihatkomut: Bu Ülkede 45 Yaş Üstü Seçmenin % 70 i Neden \r\r\r@Akparti ye Oy Veriyor \r\r\rBiliyormusunuz ? \r\r\r\r\r\*Onlar Evlatlarının Geleceklerini D..."

[6] "RT @AkGncGazipasa: Ak Parti Gençlik Kolları olarak Cumhurbaşkanımız Recep Tayyip Erdoğan'ın talimatıyla kapı kapı dolaşmaya Teşkilatımız il..."

CHP'ye ilişkin atılan ham tweetler;

[1] "RT @enveryan: Şiddet politikasını savunanlar:\r\r\n\r\r\n- PKK özelinde HDP\r\r\n- FETÖ özelinde İP\r\r\n- Her iki örgütle de içli dışlı olan CHP"

[2] "RT @emrebagce: Saadet'in, CHP'nin ve diğer partilerimizin yer alacağı Yeni Meclis'te ve erdemli, vicdanlı yöneticilerin idare edeceği Yeni..."

[3] "RT @BirGun\_Gazetesi: CHP, Twitter hesabından yaptığı anketle Cumhur İttifakı dışında bir siyasi partiye oy veren vatandaşları terörist ilan..."

[4] "RT @Merve\_Ilgz: CHP'liyim Atatürkçüyüm ama PKK ile iş tutan Kılıçdaroğlu ile de HDP'ye oy isteyen İnce ile de artık işim yok.Bunlara inat m..."

[5] "RT @enveryan: 1] Diyorlar ki sen bir akademisyensin. Neden partizanlık yapıyorsun?\r\r\n\r\r\n"Partizanlık" dedikleri de özünde Erdoğan ve AK Parti'y..."

[6] "RT @ugurdundarsozcu: Mersin'den bir gazeteci kardeşim aradı. CHP'nin Cumhurbaşkanı adayı @vekilince 'nin konuştuğu miting alanının tüm siya..."

HDP'ye ilişkin atılan ham tweetler;

[1] "@edipyuksel: DİKKAT:\r\r\nHDP'nin baraja garantilediğini iddia eden her kişi, her anket, her haber AKP propagandasıdır. Böylece HDP'nin baraj..."

[2] "@Nesrinna: Erdoğan'ın öyle büyük oyun planları falan yok elinde. En büyük hamlesi HDP'nin barajın altında kalmasını sağlamak ve HDP'nin..."

[3] "@FidelOKAN: Muhalefetin Anayasa değişikliği yapabilmesi için toplam kaç vekil çıkarması gerekir? \r\r\n\r\r\nMillet İttifakı ve HDP'nin vekil sayısı..."

[4] "@yenergunes: PKK'ya bağlı yayın yapan Yeni Özgür Politika gazetesinin dağda ziyaret ettiği, HDP'nin Cumhurbaşkanı adayı Selahattin Demir..."

[5] "@HDPistanbul: Demi aş: Halk Erdoğan ve AKP'den bıktı ve bu seçimlerde "a ık yeter" diyecektir.\r\r\n\r\nhttps://t.co/MQKnYLIepJ https://t.co/vh..."

[6] "@ahakanbayrakci: 5 HAZİRAN itibarıyla Sonar'a göre pa ilerin ve adayların son oy oranları.\r\r\r\r\rR. T. Erdoğan (AKP).....48.3\r\r\r\r\rM. İnce (CHP)..."

MHP'ye ilişkin atılan ham tweetler;

[1] "@fatihceyhan70: MHP Kahramanmaraş Milletvekili Adayımız Sayın Zuhâl Karakoç Dora'nın düzenlediği çifköfte etkinliğine katıldık. Saha eki..."

[2] "@MKalayci42 Numan kurtulmuş af yok dedi devlet bahçeli nin sözleri kendi düşüncesidir diyor neyin desteğini verdi o zaman MHP"

[3] "RT @BozkurtCaps: Kayseri Şehidimiz'in yavrusu\r\r\r\r\rMHP Kayseri İl Başkanımıza emanet...\r\r\r\r\rKİN YUTTUK, KAN KUSTURACAĞIZ!  
<https://t.co/u3YI0wirZ9>"

[4] "RT @kacsaatoldunet: Meral Akşener: "Kerim Çoraklık ülkücü bir kardeşimiz. 17/25'ten sonra, MHP çatısı altındayken Bahçeli'nin sözlerini twe..."

[5] "RT @af\_var: MAHKUMLARIN YERİNDE TAŞ OLSA ÇATLARDI...\r\r\r\r\rAf Verin Artık...\r\r\r\r\rSABREDEMİYORUZ...\r\r\r\r\r@RT\_Erdogan\r\r\r\r\r@Akp parti @MHP\_Bilgi @dbdevletbahceli..."

[6] "@arzuerdemDB @mhpbcckmece2023 @1erdemaydin @MHP\_Sosyal @ISTANBULOCAK @kazimerdiaktunc @MhpTbmmGrubu @Emin\_Haber...  
<https://t.co/m3x4WvvnQU>"

İyi Parti'ye ilişkin atılan ham tweetler;

[1] "RT @AytunCiray: Farkında mısın? Kadınlar ve gençler onunla yürüdüler.  
<https://t.co/d1PiwCCjpP>"

[2] "<https://t.co/Mqb3OsCGup> adresinde form doldurarak, parti üyesi olmadan da gönüllü olabilir, sandık başındaki İYİler... <https://t.co/qxGN24jtr3>"

[3] "RT @korayaydintr: İYİ Parti'li sandık görevlileriyle birlikte and içtik... Milli irade hırsızlarına fırsat vermeyeceğiz. Sandıkları koruyac..."

[4] "Adaletİçin SonViraj\r\r\nAraba takla atmadan adalet gelsin yeter\r\r\n@Ynskprrt @magduriyetlertr @Akparti @iyiparti @CHPEU <https://t.co/7nCN4KpNWe>"

[5] "RT @meral\_aksener: Erzurum'da milletimizleyiz! #YüzünüGüneşeDönTürkiye <https://t.co/0nDNWfvfJX>"

[6] "<https://t.co/K4szEprc5b> adresinde form doldurarak, parti üyesi olmadan da gönüllü olabilir, sandık başındaki İYİler... <https://t.co/JBofjedwww>"

Twitter'da 2018 yılı itibariyle 280 karaktere kadar tweet atılabilmektedir, öncesinde ise sadece 140 karakterler tweet atılmaktaydı. R ile tweet analizi yapabilmek için 2016 yılının ağustos ayında geliştirilmiş olan twitteR paketi kullanılmaktadır. Bu paket sadece 140 karaktere kadar analiz yapabilmektedir.<sup>170</sup> Analizde en büyük dezavantajlardan biri budur. Çekilen tweetler yukarıdaki örneklerde de görülebileceği gibi analiz için anlamsız olan pek çok karakter barındırmaktadır, analize geçmeden önce bu karakterlerin ve diğer fazlalıkların ayıklanması gerekir. Veri setini temizlemek için R içinde yer alan "tm" kütüphanesi kullanılmaktadır. Tm paketi dışında kalan alanların temizlenebilmesi için çeşitli fonksiyonlar oluşturulmuş bu fonksiyonlar da EK1'de sunulmuştur.

Bu aşamalardan sonra tweetlerdeki semboller, büyük harfler, sayılar gibi analizde işimize yaramayacak olan tüm fazlalıklar ortadan kaldırılmıştır. Tweetlerde fotoğraf ve video paylaşılmış ise bunu R yazılımı çeşitli karakterler halinde algılamaktadır (Muharrem İNCE'yi ilişkin atılan 2 nolu tweette olduğu gibi) bu anlamsız duran karakterler temizlenmiştir. Herhangi bir link paylaşılmış ise bu paylaşıma dair link tweetten çıkarılmıştır. Tüm bu temizleme işlemine karşın edatlar, bağlaçlar gibi cümle içinde yer alan ancak analiz için bir anlamı olmayan kelimelerin de tweetlerden çıkarılması gerekmektedir. İngilizcede bu kelimelere stopwords denir ve R yazılımı içinde gerek İngilizce gerek başka dillerde hazır stopwords veritabanları vardır ancak Türkçe stopwords korpusu ne yazık ki henüz geliştirilip R kütüphanesine kazandırılmamıştır. Türkçede böyle bir hazır veritabanı olmadığı için 621 kelimedenden oluşan durak kelimeler veri tabanı oluşturulmuş, tezde kullanılmış ve diğer kullanıcıların da istifade edip geliştirebilmeleri için hazırlanan kizilkayamurat.com

<sup>170</sup> <https://cran.r-project.org/web/packages/twitteR/twitteR.pdf>, erişim tarihi; 26.06.2018

isimli web sitesine text dosyası olarak konulmuştur. Bu çalışmaların yanı sıra RT veya trol hesaplar nedeniyle birebir aynı olan tweetlerin ayıklanması yapılmıştır. Örneğin İyi Parti'ye ilişkin 2 nolu ve 6 nolu tweetler birebir aynıdır, bu tür tweetlerden sadece bir tanesi saklanmış geri kalanları ayıklanmıştır. Yukarıda ham halleri verilmiş olan tweetler ön işleme aşamasından sonra büyük bir değişikliğe uğramış, artık analize hazır haldedir.

Recep Tayyip ERDOĞAN'a ilişkin ön işleme tamamlanmış tweetler;

1 muharrem ince tayyipin diplomasını göstermesini istemişti yandaş yalaka televizyon kanalında aile boyu iktidar yala

2 tayyip tura inceyi istiyor inceye tüm medyalar sonuna kadar açık yıldır sola laf ede ede seçim kaybetmedi tek um

3 recep tayyip erdoğan afrinden sonra diğer terör bataklıklarını kurutmak gerekli adımları atmaya başladık

4 tayyip emeklilere gençlere eskiyi anlatın demişti akşeneri dinlerken akla lardaki siyasi üslup ne nezaket geldi merala

5 tayyip erdoğanın bu ülkeye en büyük hizmeti abdestsiz gezmeyen maarrem kazandırmış olmasıdır adam naapsın naapsın

6 sayın erbakanfatih böyle bir zamanda hakkın yanında yer alarak recep tayyip erdoğan ı desteklediğinizi açıklamanız türk m

Muharrem İNCE'ye ilişkin ön işleme tamamlanmış tweetler;

1 ben anket falan anlamam mersin yenişehir ilçesini az kez dolaşip el sıkışmış bir siyasetçi olarak söylüyorum

2 bir adet sevgili bir adet muharremince ince barışacağız

3 yakışıklılığa andama bakar mısınız yav muharremince

4 itfaiyeciler ekgoestergehakkıistiyor muharremince kemalkilictaroglu meralaksener temelkaramollaoğlu

5 muharremincenin seçim vaadleri ince eskitürkiye refahgöstergeleri çarşamba hatırla türkiyedeğiştiriyor vakittü

6 hayatım boyunca chpye oy vermedim sempati duymadım bırak muharremince yi tanımayı uzaktan gören biriyle

Ön işleme aşamasında Meral AKŞENER ve İyi Parti'ye ilişkin atılan tweetlerde diğer tweetler dışında ekstra olarak bir işlem daha yapılmıştır. “İyi” kelimesi pozitif anlama sahip olduğu ve gerek Akşener gerekse İyi Parti'ye ilişkin atılan tweetlerde parti isimden dolayı defalarca tekrarlandığı için analizin sağlıklı olabilmesi amacıyla ön işleme aşamasında bu iki grup tweetten “iyi” kelimesi ayıklanmıştır.

Meral AKŞENER'e ilişkin ön işleme tamamlanmış tweetler;

- 1 lütfen utanmadan maskaralık yapmaya devam ediniz ayrıca sizin birinin torununa ahlaklı bir öğretmesi beklenmezdi
- 2 böyle bunlar meral aksener
- 3 insan sinde neyse inde odur kelime haznesi kelimeden oluşan meral aksener vatan evlatlarına ahmak dedi si
- 4 yürürken birkaç çocuğun yanından geçiyordum öp öldür evlen oynuyorlardı isimler muharrem ince recep tayyip erdo
- 5 parti meral aksener üzerine ölü toprağı serpilmiş uyuyan siyaseti ay bir sürede uyandırdırması nedeniy
- 6 parti genel başkanı müstakbel cumhurbaşkanımız meral akşener seçilir seçilmez suriye ilişkileri düzeltip t

Selahattin DEMİRTAŞ'a ilişkin ön işleme tamamlanmış tweetler;

- 1 selahattin demirtaş seçmene güzel hitap eden bir aday seçmen tarafından çok sevilen bir isim dediğim gibi sol
- 2 selahattin demirtaş miti kapatacağım demedi doğu perinçek tikayı kapatacağım demedi ilginç bir vaka karşı karşıyayız
- 3 kılıçdaroğlunuyoruyorum enis berberoğlu selahattin demirtaş siyasi esir neden hiçbir zaman selahattin d
- 4 der horror in der türkei hat noch gar nicht begonnen dlfcultur berichtet über die solidaritätslesung für hdpdemirtas
- 5 selahattin demirtaş ekim olaylarında halkı sokağa çağırdığının ispatı vakittürkiyevak
- 6 mutluluk paylaşınca güzel seçim öncesi güzel haber barzaniye yakın kaynaklar tsk

kandile yaptığı hava saldırı

Ak Parti'ye ilişkin ön işleme tamamlanmış tweetler;

1 bakanımız abdulhamitgul nurdağı belediyemizi ak parti nurdağı ilçe merkezimizi ziyaret ettik nurdağı teş

2 ak parti iktidarında aile kadın yapılan düzenlemeleri sizlerle paylaşacağım anayasada kadın erkek eşitliği ilkesi

3 ak partinin diyarbakırda olduğu gibi batmanda işi çok zor hdp dışında karşısında muhafazakâr kesimden hüda p

4 uyarı hdp barajı geçemiyor ak partiye mv daha geliyor durumu gören cıa fetö harekete geçiyor kendine milli

5 bu ülkede yaş üstü seçmenin neden akparti oy veriyor biliyormusunuz onlar evlatlarının geleceklerini d

6 ak parti gençlik kolları olarak cumhurbaşkanımız recep tayyip erdoğanın talimatıyla kapı kapı dolaşmaya teşkilatımız il

CHP'ye ilişkin ön işleme tamamlanmış tweetler;

1 şiddet politikasını savunanlar pkk özelinde hdp fetö özelinde ip her iki örgütle içli dışlı olan chp

2 saadetin chpnin diğer partilerimizin yer alacağı yeni mecliste erdemli vicdanlı yöneticilerin idare edeceği yeni

3 chp twitter hesabından yaptığı anketle cumhur ittifakı dışında bir siyasi partiye oy veren vatandaşları terörist ilan

4 chpliyim atatürkçüyüm pkk iş tutan kılıçdaroğlu hdpye oy isteyen ince artık işim yok bunlara inat m

5 diyorlar sen bir akademisyensin neden partizanlık yapıyorsun partizanlık dedikleri özünde erdoğan ak partiy

6 mersinden bir gazeteci kardeşim aradı chpnin cumhurbaşkanı aday vekilince nin konuştuğu miting alanının tüm siya



HDP'ye ilişkin ön işleme tamamlanmış tweetler;

- 1 dikkat hdpnin baraja garantilediğini iddia eden kişi anket haber akp propagandasıdır böylece hdpnin baraj
- 2 erdoğanın öyle büyük oyun planları falan yok elinde en büyük hamlesi hdpnin barajın altında kalmasını sağlamak hdpnin
- 3 muhalefetin anayasa değişikliği yapabilmesi toplam kaç vekil çıkarması gerekir millet ittifakı hdpnin vekil sayı
- 4 pkya bağlı yayın yapan yeni özgür politika gazetesinin dağda ziyaret ettiği hdpnin cumhurbaşkanı adayı selahattin demir
- 5 demirtaş halk erdoğan akpden bıktı seçimlerde artık yeter diyecektir
- 6 haziran itibarıyla sonara göre partilerin adayların son oy oranları r t erdoğan akp m ince chp

MHP'ye ilişkin ön işleme tamamlanmış tweetler;

- 1 mhp kahramanmaraş milletvekili adayımız zuhal karakoç doranın düzenlediği çifköfte etkinliğine katıldık saha eki
- 2 numan kurtulmuş af yok dedi devlet bahçelinin sözleri kendi düşüncesidir diyor neyin desteğini verdi o zaman mhp
- 3 kayseri şehidimizin yavrusu mhp kayseri il başkanımıza emanet kin yuttuk kan kusturacağız
- 4 meral akşener kerim çoraklık ülkücü bir kardeşimiz ten sonra mhp çatısı altındayken bahçelinin sözlerini tve
- 5 mahkumların yerinde taş olsa çatlardı af verin artık sabredemiyoruz rt erdoğan akparti mhp bilgi dbdevletbahceli
- 6 mhpbcckmece erdemaydin mhp sosyal istanbulocak kazimerdiaktunc mhptbmmgrubu emin haber

İyi Parti'ye ilişkin ön işleme tamamlanmış tweetler;

- 1 farkında mısın kadınlar gençler onunla yürüdüler
- 2 adresinde form doldurarak parti üyesi olmadan gönüllü olabilir sandık başındaki

3 partili sandık görevlileriyle birlikte and içtik milli irade hırsızlarına fırsat vermeyeceğiz sandıkları koruyac

4 adalet sonviraj araba takla atmadan adalet gelsin yeter ynskpmagduriyetlertr akparti parti chpeu

5 erzurumda milletimizleyiz yüzünügüneşedöntürkiye

#### 4.4. Verilerin Analizi

Veriler ön işleme aşamasından sonra analize hazır hale gelmektedir. Ön işlemeden önce hem adayların her biri hem de partilerin her biri için 80.000 adet tweet varken ön işleme aşamasından sonra tweet sayılarında ciddi oranda azalmalar olmuştur. Erdoğan için analizde kullanılan tweet sayısı; 20.319, İnce için 28.652, Akşener için 29.818, Demirtaş için 16.689'dur. Ak Parti için 23921, CHP için 27701, HDP için 29.843, MHP için 30194 ve İyi Parti için 23.219 adet tweet analize katılmıştır.

Analizin yapılabilmesi için önce duygu skorunu hesaplayabilecek kodun yazılması gelmektedir. Bunun için alternatifler mümkün olmakla birlikte ortak nokta, duygu skorunu hesaplayabilmek için pozitif anlama sahip olan kelime sayısından negatif anlama sahip olan kelime sayısının çıkarılmasıdır. Bu aşamada pozitif veya negatif duygu ifade eden kelime veri setine ihtiyaç duyulmaktadır.

İngilizce duygu analizi çalışmalarında hazır korpuslardan yararlanılmaktadır. Araştırmacı pozitif ve negatif kelimeleri hazır veri setleri sayesinde web ortamından kolaylıkla edinebilmektedir. Bu alanda Finn Arup Nielsen tarafından geliştirilen AFINN, Bing Liu'nin öncülüğünde geliştirilmiş olan Bing ve Saif Mohammad ile Peter Turney tarafından geliştirilmiş olan NRC isimli 3 tane alternatif veri seti vardır.<sup>171</sup> Bu veri setlerinden en popüler olan ve en çok kullanılan Bing Lui tarafından geliştirilmiş olan "A List of English Positive and Negative Opinion Word for Sentiment Words" setidir.<sup>172</sup> Türkçede pozitif veya negatif anlama sahip olan kelimeler için geliştirilmiş kullanışlı bir korpus yoktur. Bu eksikliği gidermek adına Türkçe pozitif duygu ifade eden kelimeler ve negatif duygu ifade eden kelimelerden oluşan birer veri seti

<sup>171</sup> <https://www.r-bloggers.com/sentiment-analysis-of-the-lord-of-the-rings-with-tidytext/> erişim tarihi, 10.03.2018

<sup>172</sup> <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon> erişim tarihi, 14.04.2018

oluşturulmuş daha sonra bu alanda çalışma yapacak olan araştırmacıların da erişebilmeleri için kizilkayamurat.com isimli web sitesine yüklenmiştir.

#### 4.4.1. Duygu Skoru

R yazılımı ile önce Cumhurbaşkanı adayları için daha sonra seçime katılacak partiler için tek tek duygu skorları hesaplanıp sonuçlar tablo halinde verilmiştir. Hesaplama için kullanılan kodlar EK 1’de yer almaktadır.

Duygu skoru hesaplanırken; ön işleme sonucunda elde edilen tüm tweetler için ayrı ayrı birer duygu puanı hesaplanıp bulunan sonuçların ortalaması alınmıştır. Duygu puanları hesaplanırken; tweet içinde geçen ve oluşturmuş olduğumuz korpusta pozitif kelimeler arasında yer alan kelimeler için +1 puan, korpusta negatif kelimeler arasında yer alan kelimeler için -1 puan verilmiştir. Duygu puanı; tweetteki + puanlarla – puanların toplanması sonucunda elde edilmiştir. Bir tweetin 0 puana sahip olması ya o tweette duygu ifade eden hiç bir kelimenin yer almadığını ya da tweette yer alan pozitif kelime sayısı ile negatif kelime sayısının eşit olduğunu göstermektedir, dolayısıyla bu tweetlerin duyguları nötrdür. Her bir aday ve parti için hesaplanan duygu puanlarının dağılımları aşağıda gösterilmektedir.

Tüm adaylar için tweet skorları hesaplandıktan sonra tek tek ortalamalar alınarak aşağıdaki tabloda gösterilmiştir.

**Tablo 4.3.** Adayların Duygu Skorlarının Aritmetik Ortalaması

Adaylar	Ortalama Skorlar
Recep Tayyip ERDOĞAN	0,12
Muharrem İNCE	0,50
Meral AKŞENER	0,08
Selahattin DEMİRTAŞ	0,11

Cumhurbaşkanı adayları arasında en yüksek duygu skoruna sahip aday Muharrem İNCE’dir ve duygu skoru 0,50’dir. Muharrem İNCE’yi 0,12 ile Recep

Tayyip ERDOĞAN takip etmektedir. Selahattin DEMİRTAŞ'ın duygu skoru 0,11 ortalamaya sahipken Meral AKŞENER 0,08 ile en düşük ortalamaya sahiptir.

24 Haziran 2018 Cumhurbaşkanlığı seçiminde Recep Tayyip ERDOĞAN kullanılan oyların %52,6 sını alarak ilk turda Cumhurbaşkanı seçilmiştir. Yapılan analizde Recep Tayyip ERDOĞAN'ın Muharrem İNCE'nin arkasında kalmasının nedeni Tarhan ERDEM'in sahip olduğu ve 1986 yılında beri faaliyet gösteren Konda araştırma şirketinin Mayıs 2018'de yapmış olduğu "Seçmen Kümeleri" çalışması ile açıklamak mümkündür. Konda'nın araştırmasına göre; Türkiye'de seçmenlerin %29'u hiç internete girmezken bu oran Ak Parti seçmeninde %36 ile Türkiye'deki ortalamanın üstündedir. Yine aynı rapora göre internet kullanan Ak Parti seçmeninin %8'i hiç sosyal medya kullanmamaktadır, bu oran Türkiye ortalamasına eşittir. Ak Parti toplam seçmeninin sadece %13'ü Twitter kullanmaktadır.<sup>173</sup> CHP seçmeninde ise; internete hiç girmeyenlerin oranı %24, internet kullananlardan hiç sosyal medya kullanmayanların oranı ise %7'dir. CHP seçmeninin hem internet kullanma ortalaması hem de sosyal medya kullanma ortalaması Türkiye ortalamasının üstündedir. CHP seçmeninin %20'si Twitter kullanmaktadır.<sup>174</sup> Muhalefetin en güçlü adayı olan İNCE, yaptığı seçim çalışmalarında coşkulu bir ortam yaratmış, yıllardan beri ilk kez muhalif seçimde bir umut ve heyecan yaratmıştır. Bu durum kendisini sosyal medya mesajlarında da göstermiş, İNCE'nin duygu puanı ortalaması bu yüzden rakibinden çok daha yüksek çıkmıştır. İNCE'nin puanının çok yüksek çıkmasında seçmen kitlesinin sosyal medya kullanma konusunda rakiplerine nazaran daha etkin olması da rol oynamıştır.

Meral AKŞENER analizde en düşük ortalamaya sahip olan adaydır. Ortalamasının düşük çıkmasında ki en temel neden AKŞENER hakkında seçim sürecinde sık sık FETÖ irtibatı iddalarının ortaya atılması, AKŞENER'in yaptığı tüm toplantılarda bu bağlantıları reddedip bu iddayı ortaya atan kişilere iddalarının ispat edilmesi için meydan okuyan ifadeler kullanması, buna istinaden sosyal medyada bu ifadelerin çok sık geçmesidir. Keza FETÖ kelimesi negatif duygu korpusunda yer almakta, atılan her tweette bu ifadenin geçmesi nedeniye puan hesaplanırken -1 değeri eklenmesine neden olmaktadır.

<sup>173</sup> Uncu Baran Alp, Konda Seçmen Kümeleri, Ak Parti Seçmenleri, Mayıs 2018, s.24

<sup>174</sup> Doğan Sevinç, Konda Seçmen Kümeleri, CHP Seçmenleri, Mayıs 2018, s.18-19

Selahattin DEMİRTAŞ 0,11 ortalamaya sahiptir. DEMİRTAŞ hakkında en çok yabancı tweet atılan liderdir, ön işleme süreci sonunda analize dahil edilen tweet sayısında en düşük tweete sahip olma nedenlerinden biri de budur, analiz sonucunda Meral AKŞENER'in hemen üzerinde yer almıştır.

Adayların bireysel olarak puan dağılımları incelendiğinde tüm adaylar için nötr tweet sayısı çok fazladır. Nötr tweet sayısında ki aşırı yükseklik aritmetik ortalamaların da düşük çıkmasına neden olmuştur.

Cumhurbaşkanlığı adaylarından R. Tayyip ERDOĞAN'a ilişkin atılan tweetlerin puanları ve frekans dağılımları aşağıdaki gibidir.

**Tablo 4.4.** R. Tayyip ERDOĞAN'a İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-5	6
-4	30
-3	123
-2	569
-1	2570
0	11896
1	3796
2	1024
3	240
4	45
5	11
6	4
7	2
8	1

R. Tayyip ERDOĞAN'a ilişkin atılan tweetlerin puanları -5 ila 8 puan arasındadır. Atılan tweetlerin 6'sı -5 puan, 30'u -4 puan, 123'ü -3 puan, 569'u -2 puan, 2570'i -1 puandır. Yani işleme tabi tutulan toplam 20316 tweetin 3298'i yani yaklaşık %16,3'ü negatif duygu içermektedir. Atılan tweetlerden 11896'sı herhangi bir duygu içermemektedir. Bu tweetlerin toplam tweetlere oranı %58,5'tir. Tweetlerin 3796'sı 1 puan, 1'i 8 puan, 1024'ü 2 puan, 240'ı 3 puan, 45'i 4 puan, 11'i 5 puan 4'ü 6 puan ve 2 si 7 puandır. Pozitif tweetlerin toplam tweetlere oranı %25,2'si pozitif duygu içermektedir. Atılan pozitif tweet sayısı, negatif tweet sayının yaklaşık 1,5 katıdır.

Cumhurbaşkanlığı adaylarından Muharrem İNCE'ye ilişkin atılan tweetlerin puanları ve frekans dağılımları Tablo 4.5'te gösterilmiştir.

**Tablo 4.5.** Muharrem İNCE'ye İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-5	3
-4	24
-3	70
-2	443
-1	2013
0	13843
1	8282
2	2892
3	772
4	225
5	65
6	10
7	5
8	1
9	2

Muharrem İNCE'ye ilişkin atılan tweetlerin puanları -5 ila 9 arasında değişmektedir. Atılan tweetlerden 3 tanesi -5 puan, 24 tanesi -4 puan, 70 tanesi -3 puan, 443 tanesi -2 puan, 2013 tanesi -1 puana sahiptir. Atılan tweetlerin yaklaşık %8,9'u negatif duyguya sahiptir. Nötr olan tweet sayısı 13843'tür, bu oran analiz edilen tweet sayısının %48,3'üdür. Atılan tweetlerin 8282'si 1 puan, 2892'si 2 puan, 772'si 3 puan, 225'i 4 puan, 65'i 5 puan, 10'u 6 puan, 5'i 7 puan, 1'i 8 puan ve 2'si 9 puandır. Pozitif tweetlerin toplam tweete oranı %42,8'ine denk gelmektedir. Atılan pozitif tweet sayısı negatif tweet sayısının 4,8 katıdır.

Cumhurbaşkanlığı adaylarından Meral AKŞENER'e ilişkin atılan tweetlerin puanları ve frekans dağılımları Tablo 4.6'te gösterilmiştir.

**Tablo 4.6.** Meral AKŞENER'e İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-8	1
-5	5
-4	33
-3	192
-2	828
-1	3635
0	18563
1	4988
2	1200
3	275
4	73
5	16
6	4

Meral AKŞENER'e ilişkin atılan tweetlerin puanları -8 ila 6 arasındadır. Tweetlerden 1 tanesi -8 puan, 5 tanesi -5 puan, 33 tanesi -4 puan, 192 tanesi -3 puan, 828 tanesi -2 puan, 3635 tanesi -1 puan değerindedir. Atılan tweetlerin yaklaşık %15,8'i negatif puana sahiptir. 18563 tweet nötrdür, bu oran analiz edilen tweetlerin %62,2'sine denk gelmektedir. Pozitif duyguya sahip olan tweetlerin dağılımları incelendiğinde, 4988 tweet 1 puan, 1200 tweet 2 puan, 275 tweet 3 puan, 73 tweet 4 puan, 16 tweet 5 puan ve 4 tweet 6 puan değerindedir. Pozitif tweetlerin toplam tweetlere oranı yaklaşık %22 dir. Atılan pozitif tweet sayısı negatif tweet sayısının yaklaşık 1,4 katıdır.

Cumhurbaşkanlığı adaylarından Selahattin DEMİRTAŞ'a ilişkin atılan tweetlerin puanları ve frekans dağılımları Tablo 4.7'de gösterilmiştir.

**Tablo 4.7.** Selahattin DEMİRTAŞ'a İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-6	1
-5	4
-4	29
-3	101
-2	426
-1	2045
0	10048
1	3075
2	746
3	163
4	38
5	9
6	3

Selahattin DEMİRTAŞ'a ilişkin atılan tweetlerin, 1 tanesi -6 puan, 4 tanesi -5 puan, 29 tanesi -4 puan, 101 tanesi -3 puan, 426 tanesi -2 puan, 2045 tanesi -1 puan değerindedir. Negatif puana sahip olan tweet sayısı 2606, bu tweetlerin toplam tweetlere oranı ise %15,6'dır. Tweetlerden 10048'i nötrdür. Nötr tweetlerin analiz edilen toplam tweetlere oranı ise %60,2'dir. Atılan pozitif duygu puanına sahip tweetlerin dağılımı ise, 3075 tanesi 1 puan, 746 tanesi 2 puan, 163 tanesi 3 puan, 38 tanesi 4 puan, 9 tanesi 5 puan ve 3 tanesi 6 puandır. Pozitif tweetlerin analiz edilen tweetlere oranı ise %24,2'dir.

Tüm adaylarda ortak olan nokta nötr yani duygu ifade etmeyen tweet sayının çok yüksek olmasıdır. Recep Tayyip ERDOĞAN'a ilişkin atılan tweetlerin %58,5'i Muharrem İNCE'ye ilişkin atılan tweetlerin %48,3'ü, Meral AKŞENER'e ilişkin atılan tweetlerin %62,2'si ve Selahattin DEMİRTAŞ'a ilişkin atılan tweetlerin %60,2'si nötrdür. En yüksek pozitif duygu puanına sahip tweet oranı %42,8 ile Muharrem İNCE iken, en yüksek negatif puana sahip tweet oranı %16,3 ile Recep Tayyip ERDOĞAN'a aittir.

Milletvekilliği seçimine katılan ve en yüksek oy alması beklenen beş siyasal partiye ilişkin analiz sonuçları ayrıntılı bir şekilde incelenmiştir. Partilerin duygu skorlarının ortalaması Tablo 4.8'de gösterilmiştir.



**Tablo 4.8.** Siyasal Partilerin Duygu Skorlarının Aritmetik Ortalaması

Partiler	Duygu Skoru Ortalaması
Ak Parti	0,32
CHP	0,00
HDP	0,01
MHP	0,11
İyi Parti	0,09

Ak Parti'nin duygu skoru ortalaması 0,32 puan, CHP'nin 0 puan, HDP'nin 0,01 puan, MHP'nin 0,11 puan ve İyi Parti'nin 0,09 puandır. Siyasal partilerin duygu skorlarının ortalamaları Cumhurbaşkanı adaylarının duygu puanlarının ortalamalarından çok farklı dağılmaktadır. En yüksek ortalama iktidar partisi ve seçimden birinci çıkan parti olan Ak Parti'ye aitken en düşük ortalama ise ana muhalefet partisi olan CHP'ye aittir. Ortalamalar incelendiğinde sağ seçmen profiline hitap eden tüm partilerin duygu puanlarının ortalamalarının sol seçmen profiline hitap eden partilerden daha yüksek olduğu görülmektedir. Ak Parti seçimlere iktidar olmanın avantajı ve büyük bir medya desteği ile girmiştir. Seçim sürecinde geleneksel medya organlarında iktidar partisine ayrılan süre rakiplerine ayrılan sürenin en az 2-3 katı olmuştur.<sup>175</sup> Ak Parti meclise en yakın takipçisinin yaklaşık 2 katı kadar milletvekili sokabilmeyi başarmıştır. Partinin duygu skoru ortalaması da diğer partilerin ortalamasından bariz bir şekilde yüksektir. Partilerin bir önceki seçimlerde aldıkları oy oranları 24 Haziran 2018'de yapılan seçimdeki oy oranlarıyla kıyaslanınca, CHP'nin ciddi anlamda oy kaybettiği, seçmenin beklentilerini karşılayamadığı görülmüştür. Bu durum CHP'nin duygu skoru ortalamasının da 0 gibi çok düşük olmasına neden olmuştur. CHP'nin Cumhurbaşkanı adayı ülkede umut aşılarken partinin kendisi toplumda böyle bir coşku uyandıramamıştır. Bu yüzden CHP'nin cumhurbaşkanı adayı partisinden çok daha yüksek duygu skoru ortalamasına sahiptir. HDP toplumda çok geniş bir kesimde terör

<sup>175</sup> <http://t24.com.tr/haber/cnn-turk-haberturk-ve-ntv-hangi-partiye-ne-kadar-sure-ayirdi.648829>, erişim tarihi; 29.06.2018

örgütü ile arasına mesafe koymamakla eleştirilmektedir.<sup>176</sup> Bu durum duygu skorunda da kendisini ortaya koymaktadır. MHP seçim sürecinde sık sık af çağrısı yapmakla gündeme gelmiştir. Af kelimesi pozitif anlam ifade etmekte ve pozitif kelimeler korpusunda yer almaktadır. Bu durum MHP'nin duygu skorunun diğer rakiplerine nazaran yüksek çıkmasını açıklamaktadır. İyi Parti katıldığı ilk seçimde 44 milletvekili çıkarabilmiştir, bu durum toplumda bir karşılık bulabildiğini göstermektedir ve partinin duygu skoru ortalaması da en yüksek üçüncü ortalamadır.

Sosyal medyada özellikle de Twitter'da trol hesapların yaratmaya çalıştıkları algı operasyonlarının payı çok büyüktür. Siyasal partilerin duygu puanlarının ortalamaları ile gerçekte aldıkları oyların oranları arasında özellikle iki parti için (CHP ve HDP için) uçurum olması trol hesaplarının etkinliklerine de bağlıdır. Twitter, trol hesapların genele oranlayınca küçük bir rakamda olmasına rağmen bu hesapların Twitter'daki deneyimlerinin orantısız büyüklükte olduğu ve bununla mücadele etmeye başladığı açıklamıştır.<sup>177</sup>

Partilerin duygu puanları tek tek incelendiğinde; tüm partiler için mod 0 çıkmıştır. Yani nötr duyguya sahip tweet sayısı tüm partiler için en fazladır. Partilerin duygu skorları -9 ila 7 puan arasında değişmektedir. Partilerin her birine ilişkin atılan tweetlerin duygu puanları ve frekansları aşağıda ayrıntılı bir şekilde incelenmiştir.

**Tablo 4.9.** Ak Parti'ye İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-7	1
-5	6
-4	6
-3	93
-2	446
-1	2264
0	13396
1	5120
2	1743
3	737
4	80

<sup>176</sup> <https://odatv.com/hdpnin-pkky-mesafe-koymasi-lazim-0209151200.html>, erişim tarihi; 29.06.2018

<sup>177</sup> <http://www.mynet.com/haber/teknoloji/twitter-dan-onemli-aciklama-trol-hesaplara-savas-acti-4123334-1>, erişim tarihi 30.06.2018

5	25
6	3

Ak Parti'ye ilişkin atılan tweetlerin 1'i -7 puan, 6'sı -5 puan, 6'sı -4 puan, 93 tanesi -3 puan, 446'sı -2 puan, 2264'ü -1 puandır. Negatif tweetlerin toplam tweet sayısına oranı %11,7'dir. Atılan tweetlerin 13396'sı 0 puan yani nötrdür. Nötr tweetlerin oranı ise %56'dır. Geriye kalan tweetlerin, 5120'si 1 puan, 1743'ü 2 puan, 737'si 3 puan, 80'i 4 puan, 25'i 5 puan ve 3'ü 6 puan değerindedir. Atılan tweetlerin %32,3'ü pozitif duygu puanına sahiptir. Ak Parti'ye ilişkin atılan pozitif tweet sayısı negatif tweet sayısının yaklaşık 3 katıdır.

**Tablo 4.10.** CHP'ye İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-6	2
-5	7
-4	49
-3	265
-2	1176
-1	4503
0	15747
1	4573
2	1090
3	229
4	48
5	9
6	1
7	1

CHP'ye ilişkin atılan tweetlerden, 2'si -6 puan, 7'si 5 puan, 49'u -4 puan, 265 tanesi -3 puan, 1176'sı -2 puan, 4503'ü -1 puandır. Negatif tweetlerin oranı %21,7'dir. Tweetlerin 15747'si 0 puandır, nötr tweetlerin toplam tweetlere oranı ise %56,8'dir. Atılan tweetlerin 4573'ü 1 puan, 1090'ı 2 puan, 229'u 3 puan, 48'i 4 puan, 9'u 5 puan, 1'i 6 puan ve 1'1 7 puandır. Pozitif tweetlerin toplam tweetlere oranı %21,5'tir. CHP'ye ilişkin atılan pozitif tweet sayısı ile negatif tweet sayısı hemen hemen eşittir.

**Tablo 4.11.** HDP'ye İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-9	1
-5	6
-4	43
-3	208
-2	969
-1	3674
0	19509
1	4205
2	984
3	203
4	33
5	6
6	1

HDP'ye ilişkin atılan tweetlerin 1'i -9 puan, 6'sı -5 puan, 43'ü -4 puan, 208'i -3 puan, 969'u -2 puan, 3674'ü -1 puandır. Negatif tweetlerin toplam tweete oranı %16,4'tür. 19509 tweetin puanı 0'dır ve nötr tweetlerin toplam tweetlere oranı %65,4'tür. 4205'i 1 puan, 984'ü 2 puan, 203'ü 3 puan, 33'ü 4 puan, 6'sı 5 puan ve 1 i 6 puan değerindedir. Pozitif tweetlerin tüm tweetlere oranı ise %18,2'dir. Pozitif tweet sayısının negatif tweet sayısına oranı 1,1'dir.

**Tablo 4.12.** MHP'ye İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-5	5
-4	17
-3	120
-2	669
-1	2897
0	17439
1	4352
2	1152
3	271
4	45
5	20
6	1
7	1

MHP'ye ilişkin atılan tweetlerin 5'i -5 puan, 17'si -4 puan, 1202'si -3 puan, 669'u -2 puan, 2897'si -1 puan değerindedir. Negatif tweetlerin toplam tweetlere oranı %13,7'dir. Atılan tweetlerin 17439'u 0 puandır, bu tweetlerin toplam tweetlere oranı %64,6'dır. Pozitif tweetlerin dağılımı ise, 4352'si 1 puan, 1152'si 2 puan, 271'i 3 puan, 45'i 4 puan, 20'si 5 puan, 1'i 6 puan ve 1'i de 7 puandır. Pozitif tweetlerin oranı %21,7'dir. MHP'ye ilişkin atılan pozitif tweet sayısı negatif tweet sayısının yaklaşık 1,6 katıdır.

**Tablo 4.13.** İyi Parti'ye İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı
-5	2
-4	12
-3	105
-2	499
-1	2192
0	15951
1	3396
2	813
3	185
4	44
5	17
6	1
7	1

İyi Parti'ye ilişkin atılan tweetlerin, 2'si -5 puan, 12'si -4 puan, 105'i -3 puan, 499'u -2 puan ve 2192'si -1 puandır. Negatif tweetlerin toplam tweet sayısına oranı %12,1'dir. Nötr yani 0 puana sahip tweet sayısı 15951'dir ve bu tweetlerin toplam tweetlere oranı %68,7'dir. Pozitif tweetlerin dağılımı ise, 3396'sı 1 puan, 813'ü 2 puan, 185'i 3 puan, 44'ü 4 puan, 17'si 5 puan, 1'i 6 puan ve 1'i 7 puandır. Pozitif tweetlerin toplam tweet sayısına oranı 19,2'dir. Pozitif tweet sayısının negatif tweet sayısına oranı yaklaşık 1,6'dır.

Tıpkı Cumhurbaşkanı adaylarında olduğu gibi partilerde de ortak nokta nötr tweet sayının çok yüksek olmasıdır. Ak Parti'ye ilişkin atılan tweetlerin %56'sı CHP'ye ilişkin atılan tweetlerin %56,8'i, HDP'ye ilişkin atılan tweetlerin %65,4'ü, MHP'ye ilişkin atılan tweetlerin %64,6'sı ve İyi Parti'ye ilişkin atılan tweetlerin %68,7'si

nötrdür. En yüksek pozitif duygu puanına sahip olan parti %32,3 ile Ak Parti iken, en yüksek negatif puana sahip olan parti %21,7 ile CHP'dir.

#### 4.4.2. Makine Öğrenmesi İle Duygu Analizi

Duygu analizi yapılırken makine öğrenmesi ile hem Cumhurbaşkanı adaylarına ilişkin hem de seçime katılan partilere ilişkin analizler yapılmıştır. Bu amaçla sık kullanılan makine öğrenmesi algoritmalarından CART, Random Forest ve Navie Bayes ele alınmıştır. Analiz sürecinde önce terim doküman matrisleri oluşturulmuştur. Matris boyutunun indirgenmesi için frekansı 30'dan az olan terimler çıkarılarak modifiye edilmiştir. Algoritmaların performanslarının kıyaslanabilmesi amacıyla, kesinlik, duyarlılık, doğruluk ve F-ölçütü hesaplanıp tablo halinde verilmiştir. Analiz yapılırken tüm algoritmalar için veri seti %70 eğitim, %30 test olarak ikiye ayrılmış, analizde kullanılan kodlar EK 1'de sunulmuştur.

Cumhurbaşkanı adayları için makine öğrenmesi analizi sonuçları tablolaştırılarak karşılaştırmalı olarak analiz edilmiştir.

**Tablo 4.14.** Cumhurbaşkanı Adaylarının Makine Öğrenmesi Sonuçları

Adaylar	Algoritma	P	R	A	F
Recep Tayyip ERDOĞAN	CART	0,23	0,74	0,78	0,35
	NB	<b>0,82</b>	0,28	0,43	0,43
	RF	0,33	<b>0,77</b>	<b>0,80</b>	<b>0,46</b>
Muharrem İNCE	CART	0,67	0,85	0,81	0,75
	NB	0,40	0,65	0,65	0,50
	RF	<b>0,72</b>	<b>0,86</b>	<b>0,83</b>	<b>0,78</b>
Selahattin DEMİRTAŞ	CART	0,30	<b>0,82</b>	0,81	0,44
	NB	<b>0,73</b>	0,63	0,66	<b>0,68</b>
	RF	0,44	0,80	<b>0,84</b>	0,57
Meral AKŞENER	CART	0,25	<b>0,85</b>	0,82	0,39
	NB	<b>0,80</b>	0,29	0,66	0,43
	RF	0,32	0,79	<b>0,83</b>	<b>0,46</b>

Tablo 4.14.'te cumhurbaşkanı adayları için tek tek algoritma sonuçları, kesinlik (P), duyarlılık (R), doğruluk (A) ve F-ölçütü (F) hesaplanarak gösterilmiştir. Recep Tayyip ERDOĞAN için yapılan analizde, kesinlik ölçütü için en iyi sonucu NB,

duyarlılık ölçütü için en iyi sonucu RF, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu RF vermiştir.

Muharrrem İNCE için yapılan analizde, tüm ölçütlerde yani kesinlik, duyarlılık, doğruluk ve F-ölçütünde RF en iyi performansı göstermiştir.

Selahattin DEMİRTAŞ için yapılan analizde, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu CART, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu NB vermiştir.

Meral AKŞENER için yapılan analizde, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu CART, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu RF vermiştir.

Ölçütlerin performansları kıyaslandığında, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu RF, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu RF vermiştir.

Siyasal partilerin makine öğrenmesi çıktıları aşağıda yer alan tabloda bir araya getirilerek analiz edilmiştir.

**Tablo 4.15.** Partilerin Makine Öğrenmesi Sonuçları

Partiler	Algoritma	P	R	A	F
Ak Parti	CART	0,39	<b>0,86</b>	0,78	0,54
	NB	<b>0,88</b>	0,66	0,44	<b>0,75</b>
	RF	0,54	<b>0,86</b>	<b>0,82</b>	0,66
CHP	CART	0,22	<b>0,75</b>	<b>0,82</b>	0,34
	NB	<b>0,80</b>	0,70	0,59	<b>0,75</b>
	RF	0,40	0,65	0,65	0,50
HDP	CART	0,23	0,80	0,85	0,36
	NB	<b>0,73</b>	0,63	0,66	<b>0,68</b>
	RF	0,35	<b>0,81</b>	<b>0,86</b>	0,48
MHP	CART	0,21	<b>0,87</b>	0,80	0,34
	NB	<b>0,84</b>	0,31	0,45	<b>0,45</b>
	RF	0,26	0,76	<b>0,81</b>	0,39
İyi Parti	CART	0,17	<b>0,77</b>	0,82	0,29
	NB	<b>0,85</b>	0,28	0,65	<b>0,42</b>
	RF	0,25	0,73	<b>0,83</b>	0,38

Ak Parti için yapılan analizde, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu CART ve RF birlikte, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu NB vermiştir.

CHP için yapılan analizde, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu CART, doğruluk ölçütü için en iyi sonucu CART ve F-ölçütü için en iyi sonucu NB vermiştir.

HDP için yapılan analizde, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu RF, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu NB vermiştir.

MHP için yapılan analizde, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu CART, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu NB vermiştir.

İyi Parti için yapılan analizde, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu CART, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu NB vermiştir.

Ölçütlerin performansları kıyaslandığında, kesinlik ölçütü için en iyi sonucu NB, duyarlılık ölçütü için en iyi sonucu CART, doğruluk ölçütü için en iyi sonucu RF ve F-ölçütü için en iyi sonucu NB vermiştir.

Genel olarak MÖ sınıflandırmadaki başarının tüm algoritmaların tüm ölçütlerinde çok yüksek olmadığı görülmektedir. Bunun en temel nedeni veri yapısıdır. Tweetlerde nötr tweet sayısının çok yüksek çıkması, pozitif tweet sayısının negatif tweet sayısına oranla yüksek olması sınıflandırma başarısını önemli ölçüde belirlemektedir. Cumhurbaşkanı adayları için RF, siyasal partiler için CART'ın en iyi sınıflandırmayı yapan algoritmalar olduğu görülmektedir.



## SONUÇ

Günlük hayatın vazgeçilmezlerinden biri haline gelen bilgisayarlar yaşamın her alanına yenilikler getirmeye devam etmektedir. Bilim dünyasına getirdiği pek çok yeniliğin yanında belki de en önemli katkısı veri depolama kapasitesindeki muazzam artıştır. Veri depolama kapasitesindeki artış “Big Data” olarak karşımıza çıkmaktadır. Bu devasa büyüklükteki verilerin çok büyük bir kısmı ise yapılandırılmamış verilerden oluşmaktadır. Yapılandırılmamış verilerin analiz edilebilmesi önemli bir süreçtir ve bu alanda “Metin Madenciliği”nin katkıları yadsınamaz.

Metin madenciliği, istatistik bilimi için de önemli bir veri kaynağı olan metinlerin işlenip içindeki çok önemli bilgilere erişebilme çabalarının tamamına verilen genel isimdir. Metin madenciliği yapmak aslında iğne ile kuyu kazmaya benzetilmektedir. Arka planda çok büyük bir emek vardır, sonucunda ise çok önemli bilgiler, kazanımlar elde edilebilmektedir.

Günümüzde ebeveynlerin en çok dert yandığı noktalardan biri çocuklarının bilgisayar başından kalkmamasıdır. Bilgisayar dünyası özellikle internetin yaygınlaşması ve gelişmesiyle birlikte çok çekici bir yer haline gelmiş, yeni bir cazibe merkezi oluşturmuştur. Öyle ki insanlar tatil gittiklerinde bile, normal zamanlarda olduğu gibi vaktinin çoğunu bilgisayar veya onun küçük bir yansıması olan, küçük bilgisayarlık diyebileceğimiz cep telefonlarıyla geçirmektedir. Peki insanları için ailesindeki bireylerle sohbet etmekten bile alıkoyan, vaktinin büyük bir kısmını harcadığı bu cazibe merkezinin temelinde ne var? Bu sorunun en net yanıtı sosyal medyadır. Web2.0 teknolojisi ile birlikte yeni bir medya kavramı ortaya çıkmış, internet ve bilgisayar dünyası eskisinden çok daha çekici hale gelmiştir. Bireyler sosyal medyanın gücünü keşfetmiş, kendileri bu güce farkında olarak veya olmayarak daha büyük bir güç katmıştır. Sosyal medya bu yüzden son yılların en popüler araştırma konuları arasında yer almaktadır. Sosyal medyanın bu gücü metin madenciliği için çok önemli bir veri kaynağı olmuştur.

Duygular, insanların karar verme sürecinde bazen mantık kadar bazen de mantıktan bile ön planda gelmektedir. Bu yönüyle duyguların tüm boyutlarıyla

incelenmesi bilim insanları için önemli bir çalışma alanıdır. Duyguların ölçülebilmesi, analiz edilebilmesi; insan davranışlarının, kararlarının ölçülebilmesine, anlaşılabilmesine de yardımcı olacaktır. Bu çerçevede duygu analizi son yılların önemli araştırma alanlarından biri haline gelmiştir.

Duygu analizi alanında yapılan çalışmaların büyük bir kısmı İngilizce metinler üzerinedir. Bunun en önemli nedeni dil yapısının sadeliği ve bu alanda kullanılacak hazır veritabanlarının bulunmasıdır. Tez çalışması kapsamında Türkçe tweetler kullanılmıştır. Türkçe metinlerin analiz yapılabilmesi için hazır korpuslar olmadığı için üç farklı korpus hazırlanmıştır. Hazırlanan korpuslardan ilki 1293 adet Türkçe pozitif kelimeler, ikincisi 1980 adet Türkçe negatif kelimeler ve üçüncüsü de 621 adet Türkçe durak kelimelerden oluşmaktadır. Korpusun hazırlanmasında öncelikle İngilizce korpus baz alınıp bu kelimelerin tamamı Türkçe'ye çevrilmiştir. Tekrar eden terimler ayıklanmış TDK'nın Türkçe sözlüğü taranarak korpusta eksik olan kelimeler eklenmeye çalışılmıştır. Bu korpuslar diğer araştırmacıların da faydalanabilmeleri için kizilkayamurat.com ismiyle oluşturulan web sitesine yüklenmiştir. Bu sayede literatürde yer alan önemli bir açık ortadan kaldırılmıştır.

Literatürde duygu analizi alanında yapılan çalışmaların önemli bir kısmı fen bilimleri alanında özellikle de bilgisayar yazılımı geliştirme üzerinedir. Sosyal bilimler alanında ise duygu analizi çalışması çok az olmakla birlikte yapılan çalışmalarda sözcük tabanlı yöntemler kullanılmıştır. Bu çalışma da makine öğrenmesi ile duygu analizi yapılmış, bu amaçla karar ağaçları algoritmaları ile olasılığa dayalı algoritmalar kullanılmıştır.

Geleceğe yönelik sağlıklı tahminler yapılabilmesi bilim dünyası için önemlidir ve bu alanda pek çok yöntemler geliştirilmiş, geliştirilmeye devam etmektedir. Seçim sonuçlarının da önceden doğru bir şekilde tahmin edilebilmesi amacıyla çeşitli çalışmalar yapılmaktadır. Bu çalışmaların çok büyük bir kısmı geleneksel yöntemlere dayanmakta, çıkarımsal istatistikle tahminler sürdürülmeye çalışılmaktadır.

Bu tezin uygulama kısmı 24 Haziran 2018'de yapılan Cumhurbaşkanı ve Milletvekili Genel Seçimlerine ilişkin sosyal medya verilerinden hareketle duygu analizi yapmaktır. Veri seti olarak aday listelerinin YSK tarafından kesinleştirildiği 30

Mayıs 2018 den itibaren seçimlerin yapıldığı 24 Haziran 2018 gününe kadar en yüksek oy potansiyeline sahip adaylar ve partiler hakkında atılan tweetler ele alınmıştır. Tweetler hergün düzenli bir şekilde çekilerek birleştirilmiş, ön işleme aşamasından sonra duygu analizi yapılmıştır.

Verilerin elde edilmesinde, ön işleme aşamasında ve yapılan tüm analizlerde açık kaynak kodlu olan R yazılımı kullanılmıştır. R yazılımı temel istatistikten makine öğrenmesine dek pek çok alanda kullanıma sahip olan ve son yıllarda popülaritesi gittikçe artan önemli bir araçtır. Yazılım en büyük özelliklerinden biri açık kaynak kodlu yani ücretsiz olmasıdır. Açık kaynak kodlu olması sayesinde tüm dünyadan insanlar bu yazılımın gelişmesine katkıda bulunabilmektedir. Tez çalışmasının en büyük dezavantajı R’de kullanılan twitteR paketinin kısıtıdır. R ile yapılan tüm twitter analizlerinde kullanılması zorunlu olan paket sadece 140 karaktere kadar analiz yapabilmektedir. Bunun temel nedeni paketin geliştirildiği Ağustos 2016’da Twitter’da sadece 140 karakterlik mesaj yollanmasına müsaade edilmesidir. Analizin yapıldığı 2018 yılında ise Twitter’da 280 karaktere kadar mesaj atılabilmektedir.

Analize başlamadan titiz bir ön işleme aşaması gerçekleştirilmiştir. Bu aşamada sürekli tekrar eden tweetler ayıklanmıştır. Trol hesap denilen ve genellikle algı operasyonu gerçekleştirme amacı güden hesaplar birkaç dakika içinde aynı mesajı defalarca farklı isimler, farklı hesaplar gibi tekrar tekrar atmaktadır. Twitter bu durumun ciddiyetinin farkında olup önlem almak için güvenlik çalışmalarını sürdürmektedir. Ön işleme aşamasında tekrar eden tweetlerin temizlenmesi önemli bir aşamadır. Bu aşama sonucunda ilk başta 80.000 adet çekilmiş olan tweeter kimi aday veya partiler için yaklaşık 5’te birine kadar düşmüştür. Tekrarlanan tweetler ayıklandıktan sonra tweetler temizlenmiştir. Bu aşamada R’nin tm paketi kullanılmış, tüm semboller, rakamlar, Twitter’da özel bir anlamı olan ancak analizde hiçbir işe yaramayan @, RT gibi kavramlar ayıklanmıştır. Ön işleme aşamasında karşılaşılan en önemli problem tweet atılırken kullanılan fotoğraf, video gibi ortamların ayıklanmasıdır. R tweetleri çekerken bu tür görselleri çok uzun bitişik çeşitli karakterler halinde görmektedir. Bu karakterleri tek tek ayıklayarak silmek neredeyse imkansızdır. Bu sorunu aşabilmek için bir for döngüsü oluşturulmuş, döngü ekte sunulmuştur. Döngü sayesinde karşılaşılan bu sorun aşılabilmektedir. Bahsedilen uzun ve karmaşık karakterler

R yazılımının terim doküman matrisi (TDM) oluşturmasını engellerken, temizlik aşamasından sonra bu matris oluşturulabilmiştir. Analizin sağlıklı olabilmesi için TDM'de 30 dan az bulunan terimler ayıklanmıştır. Ön işleme aşamasında durak kelimelerin ayıklanması için oluşturmuş olduğumuz Türkçe durak kelimeler korpusu kullanılmıştır, bu sayede edat, bağlaç gibi analiz için gereksiz terimler ayıklanmıştır. Ön işleme aşamasının son işlemi ise tüm terimlerin küçük harfe çevrilmesidir.

Analiz iki aşamadan oluşmuştur, önce duygu skoru hesaplanmıştır. Hesaplama için tez kapsamında geliştirmiş olan Türkçe pozitif ve negatif duygu kelimeleri korpusları kullanılmıştır. Tüm adaylar ve partiler için tek tek tweetlerin duygu skorları hesaplanmış, ortalamaları alınmış ve karşılaştırılarak analiz edilmiştir. İkinci aşama da ise çeşitli makine öğrenmesi algoritmaları ile duygu analizi yapılmış, bulunan sonuçlar tablolastırılarak analiz edilmiştir.

Elde edilen sonuçlar Cumhurbaşkanı adayları için ve siyasi parti adayları için ayrı yorumlanmalıdır. Cumhurbaşkanı adayları için bulunan duygu skorları ortalaması sıralaması ile seçim sonuçlarında adayların aldıkları oyların oranları kıyaslandığında, ilk iki sıradaki adayların yerleri hariç bir örtüşme söz konusudur. Seçim sonucunda ikinci sırada yer alan adayın puan ortalaması tüm aday ve partilerin ortalamalarından çok daha yüksek çıkmıştır. Bunun temel nedeni adayın, seçim kampanyası sürecinde yakalamış olduğu olumlu rüzgar olarak gösterilebilir. Seçimi ikinci olarak tamamlayan adayın kendi aldığı oy ile partisinin aldığı oy arasındaki büyük farkta bu rüzgara bağlıdır ve duygu skorunda bunu açık şekilde göstermektedir. Bu aday sayesinde muhalefet yıllardan beri ilk kez coşkulu ve heyecanlı bir seçim süreci geçirmiştir ve cumhurbaşkanlığı seçiminin ikinci tura kalacağı beklentisine girmiştir. Diğer adaylarınsa aldıkları oy oranlarıyla hesaplanan duygu skorları birbirleriyle uyumludur. Siyasi partilerin duygu skorlarında ise seçimden birinci ve dördüncü çıkan partiler hariç, seçim sonuçlarıyla örtüşmeyen bir tablo ortaya çıkmıştır. Sonucun bu denli farklı çıkmasındaki temel nedenler belli partiler üzerinde sosyal medyada yaratılmaya çalışılan olumsuz algıdır. Atılan tweetler incelendiğinde, seçime katılan siyasi partilerden birinin terör örgütünün partisi olmakla, diğerinin ise terör örgütünün siyasi partisine destek vermekle itham edildiği görülmüştür. Seçimi ikinci olarak tamamlayan siyasi partinin duygu skorunun çok düşük çıkmasının nedenlerinden biri de kendi seçmenleri tarafından Twitter'da

kıyasıya eleştirilmesi, politikalarının ve çıkardığı adayların beğenilmeyerek bu durumun sosyal medyada dile getirilmesidir. Makine öğrenmesi sonuçlarında ise veri yapısından dolayı algoritmaların performansları aşırı yüksek olmasa da tatmin edici sonuçlar ortaya çıkmıştır.

İleride yapılması planlanan çalışmalar, bir proje kapsamında hazırlanmış olan korpusların geliştirilerek etkinliklerinin artırılması, sözcük tabanlı analizlerle makine öğrenmesi analizlerinin birleştirilebilmesi için hem metinlerin kök ayırma işlemleri hem de korpustaki kelimelerin kök ayırma işlemleri yapılarak duygu skorlarının hesaplanması planlanmaktadır.

Gerek geleneksel yöntemlerle seçim sonucunu tahmin etmekte gerekse sosyal medya verileri üzerinden hareketle seçim sonuçlarını tahmin etmede yüzde yüz başarılı bir yöntem henüz geliştirilememiştir. Bireylerin sosyo ekonomik yapılarıyla oy tercihleri arasındaki ilişkiler, sosyo ekonomik yapılarıyla internet kullanım alışkanlıkları arasındaki ilişki ve tüm seçmenlerin sosyal medya kullanmamaları gibi nedenler dolayı sosyal medya üzerinden doğru tahminler yürütmeyi zorlaştırmaktadır.

Sonuç olarak tüm dezavantajlarına karşın geliştirilecek yeni teknikler sayesinde sosyal medya verileri üzerinden çok daha doğru tahminler elde edilebilir. Sosyal medya üzerinden tahminler yapmak geleneksel yöntemlerle tahmin yapmaktan çok daha ekonomik ve kullanışlı olabilir. Bu amaçla gerek duygu analizi teknikleri gerekse diğer teknikler geliştirilerek kullanılmaya devam edilmelidir. Her ne kadar duygu analizi ile seçim sonuçlarını çok iyi tahmin etmek mümkün olmamışsa da farklı alanlarda duygu analizi kullanılarak elde edilen sonuçlar karar mekanizmasında etkin bir şekilde kullanılabilir. Siyasal tercihler, içinde pek çok parametreyi barındıran karmaşık bir süreçken satın alma güdüsü gibi daha az karmaşık yapıya sahip olan davranışlarda duygu analizi sonuçları çok daha etkili olacaktır.

## KAYNAKÇA

- ADALI Eşref, ‘Doğal Dil İşleme’, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5 (2012)
- AKAR Özlem ve GÜNGÖR Oğuz, ‘Rastgele Orman Algoritması Kullanılarak Çok Bantlı Görüntülerin Sınıflandırılması’, Jeodezi ve Jeoinformasyon Dergisi , 2013
- AKCAPINAR SEZER Ebru, BOZKIR A. Selman, YAĞIZ Saffet ve GÖKÇEOĞLU Candan, ‘Karar Ağacı Derinliğinin CART Algoritmasında Kestirim Kapasitesine Etkisi : Bir Tünel Açma Makinesinin İlerleme Hızı Üzerinde Uygulama’ Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Kayseri, 2010
- AKÇAY Ekrem Yaşar, ‘5 Temmuz Darbe Girişimi Çerçevesinde Dijital Diplomasinin Kullanımı’ 27 Mayıs 1960'dan 15 Temmuz 2016'ya Darbeler, Geçmişten Günümüze Darbe Olgusu Ve Millet Egemenliği Kültürü" Uluslararası Sempozyumu, Adnan Menderes Üniversitesi, 2017
- AKGÜL Eyüp Sercan, ERTANO Caner ve DİRİ Banu, ‘Twitter Verileri Ile Duygu Analizi’, Pamukkale Univ Muh Bilim Derg, 22 (2016), 106–10
- AKIN Mehmet Dündar ve AKIN Ahmet Afşin, ‘Türk Dilleri İçin Açık Kaynaklı Doğal Dil İşleme Kütüphanesi: Zembek’, Elektrik Mühendisliği, 2007
- AKKAYA TALİH Duygu, ‘Sosyal Medya Reklamlarında Tüketici Algılarının Tutum, Davranış Ve Satın Alma Niyeti Üzerine Etkisi’, Trakya Üniversitesi Sosyal Bilimler Enstitüsü, 2013
- AKKÜÇÜK Ulaş, Veri Madenciliği Kümeleme Ve Sınıflama Algoritmaları (İstanbul: Yalın Yayıncılık, 2011)
- AKMAN Muhammet, ‘Veri Madenciliğine Genel Bakış Ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama’, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü 2010
- ALAN Mehmet Ali, ‘Karar Ağaçlarıyla Öğrenci Verilerinin Sınıflandırılması’, Atatürk

Üniversitesi İktisadi ve İdari Bilimler Dergisi, Cilt 28, Sayı 4, 2014, 101–13

AKPINAR Haldun, ‘Veri Tabanlarında Bilgi Keşfi Veri Madenciliği’, İstanbul Üniversitesi İşletme Fakültesi Dergisi, 29 (2000), 1–22

ALBAYRAK Ali Sait ve YILMAZ Şebnem Koltan, ‘Veri Madenciliği: Karar Ağacı Algoritmaları Ve İmkb Verileri Üzerine Bir Uygulama’, Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 14 (2009), 31–52

ALİKILIÇ Özlem ve ONAT Ferah, ‘Bir Halkla İlişkiler Aracı Olarak Kurumsal Bloglar’ Yaşar Üniversitesi Dergisi, Cilt2 Sayı 8, 2007

ALPAYDIN Ethem, Introduction to Machine Learning Second Edition E-book, The MIT Press, Cambridge, Massachusetts, England, 2010

AYDIN Sinan ve ÖZKUL Ali Ekrem, ‘Veri Madenciliği Ve Anadolu Üniversitesi Açıköğretim Sisteminde Bir Uygulama’, Eğitim ve Öğretim Araştırmaları Dergisi, 4 (2015)

AYDİN SEVİM Bilgen, ‘Postkapitalist Dönemde Küresel Markaların Yeni Pazarlama Stratejisi: Kitlelerin Yeni Yaşam Koçu Olarak Coca Cola’, Uluslararası İktisadi ve İdari İncelemeler Dergisi, 15, 2015

BAFANDEH Sadegh, AND Imandoust ve BOLANDRAFTAR Mohammad, ‘Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background’, Journal of Engineering Research and Applications Wwww.ijera.com, 3, 605–10

BARSKY Eugene ve PURDON Michelle, ‘Introducing Web 2.0: Social Networking and Social Bookmarking for Health Librarians’, JCHLA / JABSC, 27 (2006)

BORGES Bernie, MARKETING 2.0 Bridging The Gap Between Seller and Buyer Through Social Media Marketing, 2009

BOYD Danah M. ve ELLİSON Nicole B., ‘Social Network Sites: Definition, History, and Scholarship’, International Review of Research in Open and Distance Learning, 2011

- BRUNS Axel, HİGHFIELD Tim ve BURGESS Jean, ‘The Arab Spring and Social Media Audiences: English and Arabic Twitter Users and Their Networks’, American Behavioral Scientist, 2013
- CELİKYİLMAZ Asli, HAKKANİ-TÜR Dilek ve FENG Junlan, ‘Probabilistic Model-Based Sentiment Analysis of Twitter Messages’, in 2010 IEEE Workshop on Spoken Language Technology, SLT 2010 - Proceedings, 2010
- CHAPELLE Olivier, WESTON Jason, ve SCHOLKOPF Bernhard, ‘Cluster Kernels for Semi-Supervised Learning’ NIPS'02 Proceedings of The 15th International Conference on Neural Information Processing Systems, 2002
- CHAUHAN Harvinder ve CHAUHAN Anu, ‘Implementation of Decision Tree Algorithm c4.5’, International Journal of Scientific and Research Publications, 3 (2013), 2250–3153
- CHOY Murphy, CHEONG Michelle L. F., LAİK Ma Nang ve SHUNG Koo Ping, ‘A Sentiment Analysis of Singapore Presidential Election 2011 Using Twitter Data with Census Correction’, 2011
- ÇALIŞ Aslı, KAYAPINAR Sema ve ÇETİNYOKUŞ Tahsin, ‘Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama’, Endüstri Mühendisliği Dergisi, 25 (2014), 2–19
- ÇELİKYAY Emine Kübra, ‘Metin Madenciliği Yöntemiyle Türkçe’de En Sık Kullanılan Ve Birbirini Takip Eden Harflerin Analizi Ve Birliktelik Kuralları’ Beykent Üniversitesi Fen Bilimleri Enstitüsü, 2010
- ÇETİN Mahmut ve AMASYALI Fatih, ‘Eğiticili ve Geleneksel Terim Ağırlıklandırma Yöntemleriyle Duygu Analizi’ 21. Sinyal İşleme ve İletişim Uygulamaları, IEEE KKTC, 2013
- ÇİNİCİOĞLU Esmâ Nur, ATALAY Muhammet ve YORULMAZ Harun, ‘Trafik Kazaları Analizi İçin Bayes Ağları Modeli’, Bilişim Teknolojileri Dergisi, 6 (2013)
- ÇİNİCİOĞLU Esmâ Nur, EKİCİ Şule Önsel ve ÜLENGİN Füsün, ‘Bayes Ağ Yapısının Oluşturulmasında Farklı Yaklaşımlar: Nedensel Bayes Ağları Ve Veriden Ağ



Öğrenme' Sn. Prof. Dr. Halil Sariaslan'a Armağan Kitabı, Siyasal Kitabevi, Ankara, 2015

DAI Wenyuan, XUE Gui-Rong, YANG Qiang ve YU Yong, 'Transferring Naive Bayes Classifiers for Text Classification' AAAI'07 Proceedings of the 22nd National Conference On Artificial Intelligence, Canada, 2007 s.540-545

DEMİREL Aykut, 'Metin Madenciliği Yöntemleri İle Sosyal Medyadan Toplanan Fotoğraflı Paylaşımların, Metin – Fotoğraf Eşleşmesinin İncelenmesi' Beykent Üniversitesi Fen Bilimleri Enstitüsü, 2015

DELEN Dursun ve CROSSLAND Martin D., 'Seeding the Survey and Analysis of Research Literature with Text Mining', Expert Systems with Applications, 2008

DERELİ Burak, 'Bayes Ağları İle Gelecek Odaklı Konumlandırma Analizi: Oyun Konsolları Pazarında Bir Uygulama', İTÜ Fen Bilimleri Enstitüsü, 2012

DOĞAN Sevinç, Konda Seçmen Kümeleri, CHP Seçmenleri, Mayıs 2018, s.18-19

DOĞAN Sibel, 'Türkçe Dokümanlar İçin N-Gram Tabanlı Sınıflandırma: Yazar, Tür Ve Cinsiyet' (Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, 2006)

DOLGUN M. Özgür, GÜZEL ÖZDEMİR Tülin ve DORUK Oğuz, 'Veri Madenciliğinde Yapsal Olmayan Verinin Analizi: Metin ve Web Madenciliği', İstatistikçiler Dergisi, 2 (2009), 48–58

EMEL Gül Gökay ve TAŞKIN Çağatan, 'Veri Madenciliğinde Karar Ağaçları Ve Bir Satış Analizi Uygulaması', Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 6 (2005)

ERGÜN Kadriye, 'Metin Madenciliği Yöntemleri İle Ürün Yorumlarının Otomatik Değerlendirilmesi' Sakarya Üniversitesi Fen Bilimler Enstitüsü, 2012

ESTELLÉS Enrique, ESTHER Del Moral ve GONZÁLEZ Fernando, 'Social Bookmarking Tools as Facilitators of Learning and Research Collaborative Processes: The Diigo Case', Interdisciplinary Journal of E-Learning and Learning Objects, 6 (2010)

- ERTEN Filiz, ‘Metin Madenciliği Tabanlı Bir Web Sitesi Sınıflandırma Aracı Tasarımı’, Maltepe Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi 2015
- FAN Weiguo, WALLACE Linda, RICH Stephanie ve ZHANG Zhongju, ‘Tapping into the Power of Text Mining’, 2005
- FARHADLOO Mohsen, ‘Statistical Models for Aspect-Level Sentiment Analysis’ University of California, Merced, 2015
- FELDMAN Ronen, ve DAGAN Ido, ‘Knowledge Discovery in Textual Databases (KDT).’, International Conference on Knowledge Discovery and Data Mining (Kdd), 1995, 112–17
- FELDMAN Ronen ve SANGER James, The Text Mining Handbook, E-book, Cambridge University Press, 2006
- FİĞLALİ Nilgün, FİĞLALİ Alpaslan ve UZUNDURUGAN Elçin, ‘Kansei Mühendisliği Ve Uygulamaları’, Doğu Üniversitesi Dergisi, 5, 2002, 85–96
- FUSHİKİ Tadayoshi, ‘Estimation of Prediction Error by Using K-Fold Cross-Validation’, Statistics and Computing, 2011
- Gezi Parkı Sosyal Analiz, Gezi Parkı İnfografik, Ynklabs, 2013, <http://www.ynklabs.com/assets/img/white-papers/gezipark-infografik.pdf>
- GHAG Kranti ve SHAH Ketan, ‘Comparative Analysis of the Techniques for Sentiment Analysis’, in ICATE 2013 (ICATE 2013, 2013)
- GÖÇOĞLU Volkan ve AYDIN Mehmet Devrim, ‘Kamu Politikası Ve Sosyal Medya İlişkisinin Toplumsal Hareketler Bağlamında İncelenmesi’, Uluslararası Sosyal Araştırmalar Dergisi, 8 (2015)
- GÖKGÜL Ahmet Neca, ‘Sosyal Medyanın Siyasal İletişim Aracı Olarak Kullanımı: 7 Haziran 2015 Genel Seçimleri Twitter Örneği’ İstanbul Yeni Yüzyıl Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 2016
- GÜVEN Ahmet, ‘Türkçe Belgelerin Anlam Tabanlı Yöntemlerle Madenciliği’ Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, 2007

- HALTAŞ Ahmet, ALKAN Ahmet ve KARABULUT Mustafa, ‘Metin Sınıflandırmada Sezgisel Arama Algoritmalarının Performans Analizi’, Journal of the Faculty of Engineering and Architecture of Gazi University Cilt, 30 (2015), 417–27
- HAŞILOĞLU Selçuk Burak, ‘Algı Haritalarının Değerlendirilmesi İçin Kümeleme Algoritmalarına Dayalı Yeni Bir Model Geliştirilmesi’ Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, 2017
- HEARST Marti A, ‘Untangling Text Data Mining’, in 99 Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics Stroudsburg, PA, USA, 1999, pp. 3–10
- HSSINA Badr, MERBOUHA Abdelkarim, EZZIKOURI Hanane ve ERRITALI Mohammed, ‘A Comparative Study of Decision Tree ID3 and C4.5’, International Journal of Advanced Computer Science and Applications, 2014
- JADHAV Sayali D ve CHANNE H. P, ‘Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques’, International Journal of Science and Research (IJSR) ISSN Online Index Copernicus Value Impact Factor, 14611 (2013),
- JOVANOVIĆ Jelena, ‘Introduction To Text Mining’  
<https://pdfs.semanticscholar.org/presentation/88e2/17bc3905ac3f66a1c074b00045237eb21a37.pdf>
- KAPLAN Andreas M. ve HAENLEIN Michael, ‘Users of the World, Unite! The Challenges and Opportunities of Social Media’, Business Horizons, 53 (2010), 59–68
- KARACA Mehmet Fatih, ‘Metin Madenciliği Yöntemi İle Haber Sitelerindeki Köşe Yazılarının Sınıflandırılması’ Karabük Üniversitesi Fen Bilimleri Enstitüsü, 2012
- KARTAL Elif, ‘Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri Ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulama’ İstanbul Üniversitesi Fen Bilimleri Enstitüsü, 2015
- KASABA Evrim ve YILDIZTEPE Engin, ‘Destek Vektör Makinesi Yöntemi ile Bir

Duygu Çözümlemesi' Akademik Bilişim, 2016

KILIÇ Selim, 'Kappa Testi', Journal of Mood Disorders, 5 (2015), 142–44

KUZUCU Kemal, 'Müşteri Memnuniyeti Belirlemek İçin Metin Madenciliği Tabanlı Bir Yazılım Aracı' Maltepe Üniversitesi Fen Bilimleri Enstitüsü, 2015

LAHTINEN Timo, 'Automatic Indexing: An Approach Using an Index Term Corpus and Combining Linguistic and Statistical Methods', 2000

LOH Wei-Yin ve SHIH Yu-Shan, 'Split Selection Methods For Classification Trees', Statistica Sinica, 7 (1997), 815–40

MANNING Christopher D., Foundations of Statistical Natural Language Processing London: The MIT Press Cambridge, Massachusetts, 1999

MATHUR Nishant, KUMAR Sumit, KUMAR Santosh ve JINDAL Rajni, 'The Base Strategy for ID3 Algorithm of Data Mining Using Havrda and Charvat Entropy Based on Decision Tree', International Journal of Information and Electronics Engineering, 2 (2012)

MAYER John D., DI PAOLO Maria, ve SALOVEY Peter, 'Perceiving Affective Content in Ambiguous Visual Stimuli: A Component of Emotional Intelligence', Journal of Personality Assessment, 54 (1990)

MAYFIELD Antony, What Is Social Media? E-book, Icrossing Press, 2008

MEDHAT Walaa, HASSAN Ahmed ve KORASHY Hoda, 'Sentiment Analysis Algorithms and Applications: A Survey', Ain Shams Engineering Journal, 5 (2014), 1093–1113

MELEK Cemile, 'Metin Madenciliği Teknikleri İle Şirketlerin Vizyon İfadelerinin Analizi', Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü 2012

MICHIE D, SPIEGELHALTER D J ve TAYLOR C C, Machine Learning, Neural and Statistical Classification, 1994

NAGAMACHI Mitsuo, 'Kansei Engineering: A New Ergonomic Consumer-Oriented

- Technology for Product Development’, International Journal of Industrial Ergonomics, 15 (1995), 3–11
- NAGASAWA Shinya, ‘Present State of Kansei Engineering in Japan’, in Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 2004, pp. 333–38
- NEETHU M S ve RAJESREE R, ‘Sentiment Analysis in Twitter Using Machine Learning Techniques’, in 4th ICCCNT (Tiruchengode, India, 2013)
- NİZAM Hatice ve AKIN Sıla Saliha, ‘Sosyal Medyada Makine Öğrenmesi Ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması’, XIX. Türkiye’de İnternet Konferansı Bildirileri 27 - 29 Kasım 2014 Yaşar Üniversitesi, İzmir, 77–84
- OĞUZLAR Ayşe, Temel Metin Madenciliği (Bursa: Dora Yayıncılık, 2011)
- OĞUZLAR Ayşe, ‘Cart Analizi İle Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi’, Atatürk Üniversitesi, İktisadi ve İdari Bilimler Dergisi, 18, 2004
- OĞUZLAR Ayşe ve KIZILKAYA Yusuf Murat, ‘Is Donald Trump’s Presidential Election Victory A Surprise According to Twitter Data’, in 12th International Conference on Social Sciences (EUSER EUROPEAN CENTER FOR SCIENCE EDUCATION AND RESEARCH, 2017), p. 568
- OSHOİRİBHOR Emmanuel O., JOHN OtumuAdetokunbo M., and OJİEABU Clement E., ‘An Implementation of ID3 Decision Tree Learning Algorithm for Tax Fraud Control and Prevention System’, Journal of Computer Science and Information Technology, 4 (2016), 73–93
- ÖZEKES Serhat, ‘Veri Madenciliği Modelleri Ve Uygulama Alanları’, İ İstanbul Ticaret Üniversitesi Dergisi, 3 (2003), 65–82
- PAK Alexander ve PAROUBEK Patrick, ‘Twitter as a Corpus for Sentiment Analysis and Opinion Mining’, IJARCCCE, 2016
- PANG Bo, LEE Lillian ve VAİTHYANATHAN Shivakumar, ‘Thumbs up? Sentiment

Classification Using Machine Learning Techniques' (Philadelphia: Association for Computational Linguistics, 2002), pp. 79–86

PARSON Vanessa, REDDY Peter, WOOD Jon ve SENIOR Carl, 'Educating an iPod Generation: Undergraduate Attitudes, Experiences and Understanding of Vodcast and Podcast Use', Learning, Media and Technology, 34 (2009), 215–228

PERİKOS Isidoros ve HATZİLYGEROUDİS Ioannis, 'Recognizing Emotions in Text Using Ensemble of Classifiers', Engineering Applications of Artificial Intelligence, 51 (2016), 191–201

PİLAVCILAR İsmail, 'Metin Madenciliği İle Metin Sınıflandırma 1', 2007 <http://www.csharpnedir.com/articles/read/?id=731>, erişim tarihi 20.09.2017

SAYGILI Ahmet, 'Veri Madenciliği İle Mühendislik Fakültesi Öğrencilerinin Okul Analizlerinin Başarıları' (Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, 2013)

SCHÜTTE Simon, 'Engineering Emotional Values in Product Design - Kansei Engineering in Development' (UniTryck, Linköping Linköpings Universitet Department of Mechanical Engineering, 2005)

SEBASTIANI Fabrizio, 'Machine Learning in Automated Text Categorization', ACM Computing Surveys, 34 (2002)

SEBASTIANI Paola, 'A Tutorial on Probability Theory', <http://www.sci.utah.edu/~gerig/CS6640-F2010/prob-Tut.pdf>

SEÇKİN Keziban, 'Metin Madenciliğinde Kullanılan Yöntemlerin Karşılaştırılması: Siyasi Parti Liderlerinin Grup Genel Toplantı Konuşmaları İle Bir Uygulama' (Sakarya Üniversitesi Sosyal Bilimler Enstitüsü, 2011)

SEVİNDİ İbrahim Burak, 'Türkçe Metinlerde Denetimli Ve Sözlük Tabanlı Duygu Analizi Yaklaşımlarının Karşılaştırılması', Gezi Üniversitesi Fen Bilimleri Enstitüsü, 2013

SHARDA Ramesh, DELEN Dursun, ve TURBAN Efraim, Business Intelligence and Analytics Systems for Decision Support, 10th Edition, 10 th, U.S.A: Pearson, 2015

- SMOLA Alex, ve VISHWANATHAN S. V. N., Introduction To Machine Learning E-Book Cambridge Universty Press, 2008
- SUZUKI Kenji, Artificial Neural Networks Methodological Advances And Biomedical Applications (intechweb.org, 2011)
- TAHİROĞLU B. Tahir, ‘Derlem Dil Bilimi ve Güncel Gelişmeler’ Uluslar arası Türk Dil Kurultayı Bildirileri, TDK Yayınları, 2008
- TANTUĞ A. Cüneyd, ‘Metin Sınıflandırma’, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5 (2012)
- TELLİ Azime, ‘Mısır Devrimi’nde Sosyal Medyanın Rolü’, Bilge Strateji, 4 (2012)
- TONG Simon, ve KOLLER Daphne, ‘Support Vector Machine Active Learning with Applications to Text Classification’, Journal of Machine Learning Research, 2001, 45–66
- TOPAÇAN Ümit, ‘Sosyal Medya Paylaşımlarında Duygu Analizi : Makine Öğrenimi Yaklaşımı Üzerine Bir Araştırma’ Marmara Üniversitesi Sosyal Bilimler Enstitüsü, 2016
- TUMASJAN Andranik, SPRENGER Timm O, SANDNER Philipp G ve WELPE Isabell M, ‘Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment’, in Proceedings of the Fourth International AAI Conference on Weblogs and Social Media, 2010
- UNCU Baran Alp, Konda Seçmen Kümeleri, Ak Parti Seçmenleri, Mayıs 2018, s.24
- VAPNİK Vladimir N, ‘An Overview of Statistical Learning Theory’, IEEE Transactions On Neural Networks, 10 (1999)
- WANG Wenbo, CHEN Lu, THIRUNARAYAN Krishnaprasad, ve SHETH Amit P., ‘Harnessing Twitter “Big Data” for Automatic Emotion Identification’, in 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012

ZARELLE Dan, Social Media Marketing Book, O'Reilly Books, 2010

<http://www.digitalinformationworld.com/2014/10/what-happens-in-a-twitter-minute-social-media-infographic.html>, 24.10.2016

<http://www.csharpnedir.com/articles/read/?id=731>, 08.11.2016

Navie Bayesian, [http://www.saedsayad.com/naive\\_bayesian.htm](http://www.saedsayad.com/naive_bayesian.htm), 10.11.2016

<http://www.emrealadag.com/makine-ogrenmesi-nedir.html>, 19.01.2017

[http://www.tdk.gov.tr/index.php?option=com\\_gts&arama=gts&guid=TDK.GTS.5891b60fb4f438.66021402](http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.5891b60fb4f438.66021402), 01.02.2017

Burak Ataalp, Kansei Mühendisliği, <http://www.endustrimuhendisligim.com/kansei-muhendisligi>, 06.02.2017

[https://tr.wikipedia.org/wiki/Yapay\\_sinir\\_a%C4%9Flar%C4%B1](https://tr.wikipedia.org/wiki/Yapay_sinir_a%C4%9Flar%C4%B1), 13.03.2017

Simon Kemp, Digital In 2017 Global Overview, <https://www.slideshare.net/wearesocials/digital-in-2017-global-overview>, 27.11.2017

<https://wearesocial-net.s3.amazonaws.com/uk/wp-content/uploads/sites/2/2017/01/Slide048.png>, 28.11.2017

Podcast, Wikipedia, <https://tr.wikipedia.org/wiki/Podcast>

Twitter Top 100 Most Followers, <https://twittercounter.com/pages/100>, Erişim tarihi, 22.12.2017

<http://blog.milliyet.com.tr/----ne-demek-hic-dusundunuz-mu-/Blog/?BlogNo=20467>

[http://chem-eng.utoronto.ca/~datamining/dmc/model\\_evaluation.htm](http://chem-eng.utoronto.ca/~datamining/dmc/model_evaluation.htm)

M. Ali Akcayol Web Mining, <http://w3.gazi.edu.tr/~akcayol/files/WML5ClassifierEvaluation.pdf>, erişim: 03.02.2018

Michael W Berry, Zlatko Drm ve Elizabeth R Jessup, "Matrices, Vector Spaces, And Information Retrieval",

<https://www.cs.colorado.edu/~jessup/SUBPAGES/PS/matrices.pdf>, 09.11.2017, s.4.



Makine Öğrenmesi 5 <https://showmethenotes.blogspot.com.tr/2017/10/mkina-ogrenmesi-5.html>, erişim tarihi 03.02.2018

<http://www.dw.com/tr/sosyal-medyada-yalan-haber-daha-%C3%A7ok-tutuluyor/a-42902833> erişim tarihi 05.03.2018

<https://ipfs.io/ipns/tr.wikipedia-on-ipfs.org/wiki/NUTS.html> erişim tarihi 06.06.2018

MEERMAN Scott D., Pazarlamanın ve İletişim Yeni Kuralları, Mediacat Kitapları, 2009

<http://t24.com.tr/haber/gezici-secmen-konusmuyor-18-anket-icin-40-kapiya-giderdik-simdi-120-bu-dip-dalgasi,638931> erişim tarihi 06.06.2018

<https://www.tech-worm.com/troll-nedir-troll-hesap-ne-demektir/> erişim tarihi; 25.02.2018

<https://www.r-bloggers.com/sentiment-analysis-of-the-lord-of-the-rings-with-tidytext/> erişim tarihi 10.03.2018

<https://www.ntv.com.tr/turkiye/ysk-kesin-olmayan-secim-sonuclarini-acikladi,9y2BK2Y8kyM0-4-3rZeZA> erişim tarihi; 27.06.2018

<https://cran.r-project.org/web/packages/twitteR/twitteR.pdf>, erişim tarihi; 26.06.2018

<http://t24.com.tr/haber/cnn-turk-haberturk-ve-ntv-hangi-partiye-ne-kadar-sure-ayirdi,648829>, erişim tarihi; 29.06.2018

<https://odatv.com/hdpnin-pkky-mesafe-koymasi-lazim-0209151200.html>, erişim tarihi; 29.06.2018

<http://www.mynet.com/haber/teknoloji/twitter-dan-onemli-aciklama-troll-hesaplara-savas-acti-4123334-1>

## EKLER

### Ek 1 Tezde Kullanılan Kodlar

#### Paketlerin İndirilip Kurulması

```
install.packages("twitteR")
```

```
install.packages("ROAuth")
```

```
install.packages("tm")
```

```
install.packages("wordcloud")
```

```
install.packages("ggplot2")
```

```
install.packages("stringr")
```

```
install.packages("plyr")
```

```
install.packages("e1071")
```

```
install.packages("RColorBrewer")
```

```
install.packages("readxl")
```

```
install.packages("SnowballC")
```

```
install.packages("rpart.plot")
```

```
install.packages("randomForest")
```

```
install.packages("caTools")
```

```
library(twitteR)
```

```
library(ROAuth)
```

```
library(tm)
```

```
library(wordcloud)
```

```
library(ggplot2)
```

```
library(e1071)
```

```
library(stringr)
```

```
library(plyr)
```

```
library(RColorBrewer)
```

```
library(readxl)
```

```
library(SnowballC)
```

```
library(rpart.plot)
```

```
library(randomForest)
```

```
library(caTools)
```

Twitter'a Bağlanıp Tweetlerin Çekilmesi

```
setup_twitter_oauth("consumerKey","consumerSecret","Access_token","token_secret")
```

```
tweets = searchTwitter('Aranacak Kelime veya Kelimelerin' = ÇekilecekTweetSayısı)
```

Pozitif Kelimelerin, Negatif Kelimleri ve Durak Kelimelerinin Yüklenmesi

```
setwd("path/kelimeler")
```

```
neg.words = scan("negatifkelimeler.txt", what="character", comment.char=";")
```

```
pos.words = scan("pozitifkelimeler.txt", what="character", comment.char=";")
```

```
durak_kelimeler = scan("durakkelimeler.txt", what="character", comment.char=";")
```

```
pos.words = sapply(pos.words,function(x) x)
```

```
neg.words = sapply(neg.words,function(x) x)
```

```
durak.words= sapply(durak.words,function(x) x)
```

```
pos.words = sapply(pos.words,function(x) x)
```

```
neg.words = sapply(neg.words,function(x) x)
```

```
durak.words= sapply(durak.words,function(x) x)
```

Tweetlerin Temizlenip Duygu Skorlarının Hesaplanması

```
score.sentiment = function(tweet, pos.words, neg.words)
```

```
{
```

```
  require(plyr)
```

```
  require(stringr)
```

```
  scores = sapply(tweet, function(tweet, pos.words, neg.words) {
```

```
    silinecekKelimeler=c("<U","RT ","http", "<ed", "\r","@",'")
```

```
    for(a in 1:length(silinecekKelimeler)){
```

```
      BaslayaniSil= silinecekKelimeler[a]
```

```
      silinecekUzunluk=nchar(BaslayaniSil)
```

```
      for(a in 1:length(tweet)){
```

```
        harfSayisi=nchar(tweet[a])
```

```
          for(b in 1:harfSayisi){
```

```
            if(substr(tweet[a],b,b+silinecekUzunluk-1)==toString(BaslayaniSil))
```

```
              {
```

```
                oncesi=substr(tweet[a],1,b-1)
```

```
                sonKelimemi="evet"
```

```
                for(c in b:harfSayisi){
```

```

        if(substr(tweet[a],c,c)==" "){
            sonrasi=substr(tweet[a],c+1,harfSayisi)

            gecici <- c(oncesi,sonrasi)

            tweet[a]=toString(gecici)

            remove(gecici)

            sonKelimemi="hayir"

            break
        }
    }
    if(sonKelimemi=="evet"){tweet[a]=oncesi}
}
}

tweet=gsub("[[:punct:]]", "",tweet)

tweet=gsub("[0-9]", "",tweet)

tweet=gsub(" ", " ",tweet)

tweet=gsub(" ", " ",tweet)

tweet=tolower(tweet)

tweet = gsub(durak.words,"",tweet)

tweet = gsub('parti/aday ismi',"", tweet)

word.list = str_split(tweet, "\\s+")

```

```

words = unlist(word.list)

pos.matches = match(words, pos.words)

neg.matches = match(words, neg.words)

pos.matches = !is.na(pos.matches)

neg.matches = !is.na(neg.matches)

score = sum(pos.matches) - sum(neg.matches)

return(score)
}, pos.words, neg.words )

scores.df = data.frame(score=scores, text=tweet)

return(scores.df)

}

```

### Analiz Sonuçları

```

analysis = score.sentiment(tweet, pos.words, neg.words)

mean(analysis$score)

table(analysis$score)

```

### Kelime Bulutunun Oluşturulması

```

wordcloud(tweet,min.freq=5,scale=c(3,0.5),colors=brewer.pal(8,"Dark2"),random.color
=TRUE, random.order=FALSE, max.words=220)

```

### Makine Öğrenmesi Aşaması

```

scores <- score.sentiment(tweet, pos.words, neg.words)

```

```

scores$Candidate = factor(rep(c("SutunAdi"), nrow(scores)))

scores$very.pos = as.numeric(scores$score >= 2)

scores$very.neg = as.numeric(scores$score <= -2)

numpos <- sum(scores$very.pos)

numneg <- sum(scores$very.neg)

global_score = paste0(round(100 * numpos / (numpos + numneg)), "% ")

```

### Verilerin Temizlenmesi

```

silinecekKelimeler=c("<U", "RT ", "http", "<ed", "\r", "@", "'")

for(a in 1:length(silinecekKelimeler)){

BaslayaniSil= silinecekKelimeler[a]

silinecekUzunluk=nchar(BaslayaniSil)

for(a in 1:length(tweet)){

harfSayisi=nchar(tweet[a])

    for(b in 1:harfSayisi){

        if(substr(tweet[a],b,b+silinecekUzunluk-1)==toString(BaslayaniSil))

        {

            oncesi=substr(tweet[a],1,b-1)

            sonKelimemi="evet"

            for(c in b:harfSayisi){

                if(substr(tweet[a],c,c)== " "){

                    sonrasi=substr(tweet[a],c+1,harfSayisi)

```

```

        gecici <- c(oncesi,sonrasi)

        tweet[a]=toString(gecici)

        remove(gecici)

        sonKelimemi="hayir"

        break

    }

}

if(sonKelimemi=="evet"){tweet[a]=oncesi }
}
}
}
}

```

```
tweet=gsub("[[:punct:]]", "",tweet)
```

```
tweet=gsub("[0-9]", "",tweet)
```

```
tweet=gsub(" ", " ",tweet)
```

```
tweet=gsub(" ", " ",tweet)
```

```
tweet=tolower(tweet)
```

Analiz

```
tweetCorpus <- Corpus(VectorSource(tweet))
```

```
tweetsCorpus <- tm_map(tweetCorpus, removePunctuation)
```

```
tweetsCorpus <- tm_map(tweetsCorpus, stripWhitespace)
```

```
terms <- DocumentTermMatrix(tweetsCorpus)
```



```

terms

length(findFreqTerms(terms, lowfreq=30))

sparseTerms <- removeSparseTerms(terms, 0.995)

sparseTerms

dataframe <- as.data.frame(as.matrix(sparseTerms))

colnames(dataframe) <- make.names(colnames(dataframe))

dataframe <- as.data.frame(as.matrix(sparseTerms))

colnames(dataframe) <- make.names(colnames(dataframe))

dataframe$Positive <- as.factor(scores$score >=1)

dataframe$score <- NULL

dataframe$Score <-NULL

set.seed(100)

split <- sample.split(dataframe$Positive, SplitRatio=0.7)

trainData <- subset(dataframe, split==TRUE)

testData <- subset(dataframe, split==FALSE)

modelCART <- rpart(Positive ~., data=trainData, method="class")

predictCART <- predict(modelCART, newdata = testData, type="class")

table(testData$Negative, predictCART)

modelNavie <- naiveBayes(Positive ., data = trainData, SplitRatio=0.7)

predictNavie <- predict(modelNavie, newdata = testData)

table(testData$Positive, predictNavie)

```

```
modelForest <- randomForest(Positive ~ ., data = trainData, nodesize = 25, ntrees = 200)
```

```
predictForest <- predict(modelForest, newdata=testData) table(testData$Negative, predictForest)
```

## Ek 2: Gruplarda İlk 6 Tweetin Duygu Puanı Ortalaması

Recep Tayyip ERDOĞAN'a ilişkin tweetlerin puanlanmış hali;

1 muharrem ince tayyipin diplomasını göstermesini istemişti yandaş yalaka televizyon kanalında aile boyu iktidar yala (-2)

2 tayyip tura inceyi istiyor inceye tüm medyalar sonuna kadar açık yıldır sola laf ede ede seçim kaybetmedi tek um (1)

3 recep tayyip erdoğan afrinden sonra diğer terör bataklıkların kurutmak gerekli adımları atmaya başladık (0)

4 tayyip emeklilere gençlere eskiyi anlatın demişti akşeneri dinlerken akla lardaki siyasi üslup ne nezaket geldi merala (0)

5 tayyip erdoğanın bu ülkeye en büyük hizmeti abdestsiz gezmeyen maarrem kazandırmış olmasıdır adam naapsın naapsın (1)

6 sayın erbakanfatih böyle bir zamanda hakkın yanında yer alarak recep tayyip erdoğan ı desteklediğinizi açıklamanız türk m (1)

Muharrem İNCE'ye ilişkin tweetlerin puanlanmış hali;

1 ben anket anlamam mersin yenişehir ilçesini az kez dolaşıp el sıkılmış bir siyasetçi olarak söylüyorum (0)

2 bir adet sevgili bir adet muharremince ince barışacağız (2)

3 yakışıklılığa andama bakar mısınız yav muharremince (1)

4 itfaiyeciler ekgoestergehakkıistiyor muharremince kemalkilictaroglu meralaksener temelkaramollaoğlu (0)

5 muharremincenin seçim vaadleri ince eskitürkiye refahgöstergeleri çarşamba hatırla türkiyedeğiştiriyor vakittü (0)

6 hayatım boyunca chpye oy vermedim sempati duymadım bırak muharremince yi tanımayı uzaktan gören biriyle (-1)

Meral AKŞENER'e ilişkin tweetlerin puanlanmış hali;

- 1 lütfen utanmadan maskaralık yapmaya devam ediniz ayrıca sizin birinin torununa ahlaklı bir öğretmesi beklenmezdi (-1)
- 2 böyle bunlar meral aksener (0)
- 3 insan sinde neyse inde odur kelime haznesi kelimedden oluşan meral aksener vatan evlatlarına ahmak dedi si (-1)
- 4 yürürken birkaç çocuğun yanından geçiyordum öp öldür evlen oynuyorlardı isimler muharrem ince recep tayyip erdo (0)
- 5 parti meral aksener üzerine ölü toprağı serpilmiş uyuyan siyaseti ay bir sürede uyandırdırması nedeniy (-1)
- 6 parti genel başkanı müstakbel cumhurbaşkanımız meral akşener seçilir seçilmez suriye ilişkileri düzeltip t (1)

Selahattin DEMİRTAŞ'a ilişkin tweetlerin puanlanmış hali;

- 1 selahattin demirtaş seçmene güzel hitap eden bir aday seçmen tarafından çok sevilen bir isim dediğim gibi sol (2)
- 2 selahattin demirtaş miti kapatacağım demedi doğu perinçek tikayı kapatacağım demedi ilginç bir vaka karşı karşıyayız (0)
- 3 kılıçdaroğlunasoruyorum enis berberoğlu selahattin demirtaş siyasi esir neden hiçbir zaman selahattin d (-1)
- 4 der horror in der türkei hat noch gar nicht begonnen dlfkultur berichtet über die solidaritätslesung für hdpdemirtas (0)
- 5 selahattin demirtaş ekim olaylarında halkı sokağa çağırdığının ispatı vakittürkiyevak (0)
- 6 mutluluk paylaşınca güzel seçim öncesi güzel haber barzaniye yakın kaynaklar tsk kandile yaptığı hava saldırı (2)

Ak Parti'ye ilişkin tweetlerin puanlanmış hali;

- 1 bakanımız abdulhamitgul nurdağı belediyemizi ak parti nurdağı ilçe merkezimizi ziyaret ettik nurdağı teş (0)
- 2 ak parti iktidarında aile kadın yapılan düzenlemeleri sizlerle paylaşacağım anayasada

kadın erkek eşitliği ilkesi (1)

3 ak partinin diyarbakırda olduğu gibi batmanda işi çok zor hdp dışında karşısında muhafazakâr kesimden hüda p (-1)

4 uyarı hdp barajı geçemiyor ak partiye mv daha geliyor durumu gören cıa fetö harekete geçiyor kendine milli (-2)

5 bu ülkede yaş üstü seçmenin neden akparti oy veriyor biliyormusunuz onlar evlatlarının geleceklerini d (0)

6 ak parti gençlik kolları olarak cumhurbaşkanımız recep tayyip erdoğanın talimatıyla kapı kapı dolaşmaya teşkilatımız il (1)

CHP'ye ilişkin tweetlerin puanlanmış hali;

1 şiddet politikasını savunanlar pkk özelinde hdp fetö özelinde ip her iki örgütle içli dışlı olan chp (-3)

2 saadetin chpnin diğer partilerimizin yer alacağı yeni mecliste erdemli vicdanlı yöneticilerin idare edeceği yeni (3)

3 chp twitter hesabından yaptığı anketle cumhur ittifakı dışında bir siyasi partiye oy veren vatandaşları terörist ilan (-1)

4 chpliyim atatürkçüyüm pkk iş tutan kılıçdaroğlu hdpye oy isteyen ince artık işim yok bunlara inat m (-3)

5 diyorlar sen bir akademisyensin neden partizanlık yapıyorsun partizanlık dedikleri özünde erdoğan ak partiy (0)

6 mersinden bir gazeteci kardeşim aradı chpnin cumhurbaşkanı adayı vekilince nin konuştuğu miting alanının tüm siya (0)

HDP'ye ilişkin tweetlerin puanlanmış hali;

1 dikkat hdpnin baraja garantilediğini iddia eden kişi anket haber akp propagandasıdır böylece hdpnin baraj (0)

2 erdoğanın öyle büyük oyun planları falan yok elinde en büyük hamlesi hdpnin barajın altında kalmasını sağlamak hdpnin (1)

3 muhalefetin anayasa değişikliği yapabilmesi toplam kaç vekil çıkarması gerekir millet ittifakı hdpnin vekil sayı (0)

4 pkkyya bađlı yayın yapan yeni 6zg6r politika gazetesinin dađda ziyaret ettiđi hdpnin cumhurbaşkanı adayını selahattin demir (1)

5 demirtaş halk erdoğan akpden bıktı seçimlerde artık yeter diyecektir (0)

6 haziran itibarıyla sonara göre partilerin adayların son oy oranları r t erdoğan akp m ince chp (0)

#### MHP'ye ilişkin tweetlerin puanlanmış hali;

1 mhp kahramanmaraş milletvekili adayımız zuhal karakoç doranın düzenlediđi çifköfte etkinliğine katıldık saha eki (0)

2 numan kurtulmuş af yok dedi devlet bahçelinin sözleri kendi düşüncesidir diyor neyin desteđini verdi o zaman mhp (0)

3 kayseri şehidimizin yavrusu mhp kayseri il başkanımıza emanet kin yuttuk kan kusturacağız (-2)

4 meral akşener kerim çoraklık 6lkücü bir kardeşimiz ten sonra mhp çatısı altındayken bahçelinin sözlerini tve (0)

5 mahkumların yerinde taş olsa çatlardı af verin artık sabredemiyoruz rt erdoğan akparti mhp bilgi dbdevletbahceli (0)

6 mhpbecmece erdemaydin mhp sosyal istanbulacak kazimerdiaktunc mhptbmmgrubu emin haber (0)

#### İyi Parti'ye ilişkin tweetlerin puanlanmış hali;

1 farkında mısın kadınlar gençler onunla yürüdüler (1)

2 adresinde form doldurarak parti üyesi olmadan gönüllü olabilir sandık başındaki (1)

3 partili sandık görevlileriyle birlikte and içtik milli irade hırsızlarına fırsat vermeyeceđiz sandıkları koruyac (0)

4 adalet sonviraj araba takla atmadan adalet gelsin yeter ynskp magduriyetle rtr akparti parti chpeu (1)

5 erzurumda milletimizleyiz yüzünü güneşedöntürkiye (0)

## ÖZGEÇMİŞ

Adı Soyadı : Yusuf Murat KIZILKAYA

Doğum Yeri ve Tarihi : Ardahan / 1980

Bildiği Yabancı Dil : İngilizce

Eğitim Durumu / Mezuniyet

Lise : Ardahan Lisesi / 1998

Lisans : Uludağ Üniversitesi İ.İ.B.F. Ekonometri 2003

Yüksek Lisans : Kafkas Üniversitesi Sos. Bil. Ens. İktisat A.B.D. 2014

Doktora : Uludağ Üniversitesi Sos. Bil. Ens. Ekonometri A.B.D.  
İstatistik Bilim Dalı

Çalıştığı Kurumlar / Yıllar : Kafkas Üniversitesi 2004 – 2008

Ardahan Üniversitesi 2008 – Devam

Katıldığı Bilimsel Toplantılarda Sunduğu Bildiriler:

- OĞUZLAR Ayşe, KIZILKAYA Yusuf Murat, With R Programming, Comparison of Performance of Different Machine Learning Algorithms,14th International Conference on Social Sciences,02.03.2018 03.03.2018, Frankfurt, Almanya, 02.03.2018
- KIZILKAYA Yusuf Murat, Sosyal Medya Modern Dünyanın Haykırma Ortamı, Uluslararası Eğitim Bilimleri Ve Sosyal Bilimler Sempozyumu, Bandırma, Balıkesir 03.11.2017

- KIZILKAYA Yusuf Murat, OĞUZLAR Ayşe, Önemli Turizm Destinasyonlarımızdan Olan Antayla İline İlişkin Duygu Analizi, XVIII. Uluslararası Ekonometri Yöneylem ve İstatistik Sempozyumu, Trabzon, 05.10.2017 07.10.2017
- OĞUZLAR Ayşe, KIZILKAYA Yusuf Murat, Is Donald Trump's Presidential Election Victory A Surprise According to Twitter Data? ICSS XII, 12th International Conference on Social Sciences, Amsterdam, Hollanda, 19.05.2017 20.05.2017
- KIZILKAYA Yusuf Murat, Sosyal Ağ Analizi ve NodeXL Pro ile Twitter Verileri Üzerinde Görselleştirme, ICITS 2017, 11 Uluslararası Bilgisayar ve Öğretim Teknolojileri Sempozyumu, Malatya, 24.05.2017 26.05.2017
- KIZILKAYA Yusuf Murat, Meslek Yüksekokullarına Sınavsız Geçişle Gelen Öğrencilerle Sınavla Gelen Öğrencilerin Başarı Durumlarının Karşılaştırılması, Ardahan Üniversitesi Örneği, 3.Uluslararası Meslek Yüksekokulları Sempozyumu, Ardahan, 02.10.2013 04.10.2013
- KIZILKAYA Yusuf Murat, Büro Yönetimi ve Yönetici Asistanlığı Programı Öğrencilerinin Profili ve Beklentileri: Ardahan Sosyal Bilimler Meslek Yüksekokulu Örneği, 1.Uluslararası Meslek Yüksekokulları Sosyal ve Teknik Bilimler Kongresi, Çaycuma, Zonguldak, 22.05.2014 23.05.2015
- KIZILKAYA Yusuf Murat, Türkiye'de İç Göçün Sosyo-Ekonomik Nedenleri Üzerine İstatistiksel Bir Çalışma; Ardahan Örneği,1.Uluslararası Göç ve Kültür Sempozyumu, Amasya, 01.12.2016 03.12.2016

Yayınlanmış Makaleleri :

- KIZILKAYA Yusuf Murat, OĞUZLAR Ayşe, Bazı Denetimli Öğrenme Algoritmalarının R Programlama Dili İle Kıyaslanması, Karadeniz Uluslararası Bilimsel Dergi, 2018, 37, 90-98
- KIZILKAYA Yusuf Murat, OĞUZLAR Ayşe, Tweetlerin Etkinliğinin Sosyal Ağlarile İrdelenmesi, Uludag University International Journal Of Social Inquiry, 2017, 10, 135-150
- BAYRAKTAR Tülin, KIZILKAYA Yusuf Murat, Büro Yönetimi Ve Yönetici Asistanlığı Programı Öğrencilerinin Profili Ve Kariyer Planlaması: Batı Ve Doğu Karşılaştırması, İstanbul Aydın Üniversitesi, 2013, 5, 13-22



