



**T.C**

**BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ**  
**SOSYAL BİLİMLER ENSTİTÜSÜ**  
**YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI**

**VERİ MADENCİLİĞİ TEKNİKLERİ İLE HAVA YOLU**  
**FİRMALARININ TWEETLERİ ÜZERİNDEN DUYGU ANALİZİ**

**Fatih AYKUL**

**YÜKSEK LİSANS TEZİ**

**DANIŞMAN**

**Dr. Öğr. Üyesi Melike ŞİŞECİ ÇEŞMELİ**

**BURDUR-2019**

İç Kapak



**T.C.**

**BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ**

**SOSYAL BİLİMLER ENSTİTÜSÜ**

**YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI**

**VERİ MADENCİLİĞİ TEKNİKLERİ İLE HAVA YOLU  
FİRMALARININ TWEETLERİ ÜZERİNDEN DUYGU ANALİZİ**

**FATİH AYKUL**

**YÜKSEK LİSANS TEZİ**

**DANIŞMAN: Dr. Öğr. Üyesi Melike ŞİŞECİ ÇEŞMELİ**

**JÜRİ ÜYESİ: Dr. Öğr. Üyesi İhsan PENÇE**

**JÜRİ ÜYESİ: Dr. Öğr. Üyesi Sinan UĞUZ**

**BURDUR-2019**

Tez Onay Sayfası

**T.C. MEHMET AKİF ERSOY ÜNİVERSİTESİ SOSYAL BİLİMLER  
ENSTİTÜSÜ**

Fatih AYKUL tarafından hazırlanan “Veri Madenciliği Teknikleri ile Hava Yolu Firmalarının Tweetleri Üzerinden Duygu Analizi” başlıklı bu çalışma 29.08.2019 tarihinde *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Lisansüstü Eğitim ve Öğretim Yönetmeliği*’nin ilgili maddesi uyarınca yapılan **Tez Savunma Sınavı** sonucunda başarılı bulunarak, jürimiz tarafından **Yönetim Bilişim Sistemleri Ana Bilim Dalı**na yüksek lisans tezi olarak kabul edilmiştir.

Bölüm/Anabilim Dalı

İmza

Danışman: Dr. Öğr. Üyesi Melike ŞİŞECİ ÇEŞMELİ

Üye: Dr. Öğr. Üyesi İhsan PENÇE

Üye: Dr. Öğr. Üyesi Sinan UĞUZ

Enstitü Müdürü

İMZA

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

## T.C

# BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ

## ETİK BEYAN

Mehmet Akif Ersoy Üniversitesi Sosyal bilimler Enstitüsü Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğine göre hazırlamış olduğum “Veri Madenciliği Teknikleri ile Hava Yolu Firmalarının Tweetleri Üzerinden Duygu Analizi” adlı tezin hazırlanması sürecinde akademik etik ilkeleri ihlal etmediğimi taahhüt eder, tezimin kâğıt ve elektronik kopyalarının Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü arşivinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım.

Sosyal Bilimler Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğinin ilgili maddeleri uyarınca gereğinin yapılmasını arz ederim.

- Tezimin tamamı her yerden erişime açılabilir.
- Tezim sadece Mehmet Akif Ersoy Üniversitesi yerleşkelerinde erişime açılabilir.
- Tezimin 3 yıl süreyle erişimi açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

Fatih Aykul

29.08.2019

## ÖNSÖZ

Yüksek Lisans eğitimimin başından tez savunma aşamama kadar geçen süreçte benden hiçbir desteğini esirgemeyen, engin bilgi tecrübeleriyle beni aydınlatan saygıdeğer danışman hocam Dr. Öğr. Üyesi Melike ŞİŞECİ ÇEŞMELİ yoğun çalışma temposu içerisinde benden bilgi birikimini esirgemeyen Dr. Öğr. Üyesi İhsan Pençe'ye ve Öğr. Görevlisi Ömer AYDOĞAN'a sonsuz teşekkürlerimi sunarım. Aynı zamanda bu yoğun süreçte yardım ve fikirlerini benden esirgemeyen her daim yanımda olan sevgili arkadaşlarım Kübra UĞUR ve Recep ŞERİT'e teşekkür ederim.

Ayrıca eğitim hayatımın her aşamasında yanımda olan, benden maddi-manevi desteğini esirgemeyen, bana sonsuz güvenen sevgili aileme teşekkürü borç bilirim.



(AYKUL, Fatih, Veri Madenciliği Teknikleri ile Hava Yolu Firmalarının Tweetleri Üzerinden Duygu Analizi, Yüksek Lisans Tezi, Burdur, 2019)

## ÖZET

Teknoloji ve internetin durmadan gelişmesinin doğal bir sonucu olarak bunların hayatımızın ayrılmaz birer parçaları haline gelmesi, beraberinde sosyal medya platformlarının artmasına sebep olmuştur. İnsanların düşüncelerini paylaşabildiği sosyal medya platformlarında büyük miktarda yapısal olmayan veri birikmektedir. Bu büyük verilerin çok fazla yer kaplaması ve yapısal olmaması sebepleriyle analiz edilmesi zorlaşmaktadır. Yapısal olmayan bu tür metin verileri analizi için duygu analizi kavramı ortaya çıkmıştır. Duygu analizi, verilerdeki duygu, düşünce ve o konu hakkındaki duygunun analiz edilerek olumlu, olumsuz veya nötr olarak sınıflandırılmasını hedeflemektedir.

Günümüzde firmaların müşteri memnuniyetini yakından takip etmeleri, müşterinin durumunu değerlendirmeleri avantajdan çok mecburiyet halini almıştır. Müşterinin durumunu değerlendirebilmek, reklam kampanyalarını takip edebilmek, yeni ürünlerin piyasadaki yorumlarını değerlendirip sonuç almak için önceden sıkça uygulanan anket çalışmaları uzun süreç gerektirmekte ve kesin sonuçlar verememektedir. Bundan dolayı müşterilerin kendilerini samimi ve açık olarak ifade edebildikleri sosyal medya platformlarındaki büyük veriler duygu analizi için daha iyi ve güncel sonuçlar vermektedir.

Bu tez çalışmasında sosyal medya platformu Twitter'da, iki hava yolu firması için belirli tarihler arasında atılan Türkçe tweetler elde edilip incelenmiştir. Bu işlemler yapılırken Twitter'daki verilerin genellikle imlâ kurallarına uyulmamasından dolayı, tweetler ön işleminden geçirilmiş ve düzenlenmiştir. Düzeltilen tweetler olumlu, olumsuz ve nötr olarak; klasik sınıflandırma yöntemlerinden Yapay Sinir Ağı, K En Yakın Komşu, Bayes, Destek Vektör Makinaları ve yeni popüler yöntem olan Derin Öğrenme kullanılarak otomatik sınıflandırılmıştır. Klasik sınıflandırma yöntemleri ile son yıllarda popüler olan derin öğrenme yönteminin analiz sonuçları karşılaştırılmıştır.

**Anahtar Kelimeler:** *Sosyal Medya, Twitter, Metin Madenciliği, Duygu Analizi, Kelime Gömme, Derin Öğrenme*

(AYKUL, Fatih, *Sentiment Analysis Through Tweet Of Airway Companies With Data Mining Techniques, Burdur, 2019*)

## ABSTRACT

As a result of the continuous development of the technology and internet, which are the inseparable parts of our lives, social media platforms have increased. There is vast amount of non-structural data in these social media platforms where people express their thoughts. Since this big data is non-structural and takes up a lot of space, it is impossible to analyze it, therefore; the notion of sentiment analysis has shown up. Sentiment analysis aims to negatively, positively or neutrally classify the emotions, thoughts and the feelings about the subject in data by analyzing them.

Today, keeping monitoring customer satisfaction has become an obligation rather than being an advantage. Previously applied survey studies used for being able to evaluate the situation of customers, keeping up with advertisement campaigns, evaluating new products market values to get results requires a long process and not always accurate. For this reason, social media platforms that make customers express themselves sincere and clear gives better and more current results for large data and emotion analyses.

At this thesis study, Turkish tweets written for two airline company is analyzed for certain dates at Twitter. During these processes, tweets were pre-treated and edited due to usual spelling errors. Edited tweets were classified as positive, negative, neutral; with classical classification methods; Artificial Nerve Network, Closest Neighbor, Bayes, Support Vector Machines and newest, the most popular method; Deep Learning. Deep Learning, which has gain popularity in the last years, is compared with classical classification methods in terms of analyzing results.

**Key words:** *Social Media, Twitter, Text Mining, Sentiment Analysis, Word Embedding, Deep Learning*

## İÇİNDEKİLER

İÇ KAPAK.....	i
TEZ ONAY SAYFASI.....	ii
ETİK BEYAN .....	iii
ÖNSÖZ.....	iv
ÖZET.....	v
ABSTRACT .....	vi
İÇİNDEKİLER.....	vii
KISALTMALAR DİZİNİ.....	x
TABLolar DİZİNİ .....	xi
ŞEKİLLER DİZİNİ.....	xii
GİRİŞ .....	1

### BİRİNCİ BÖLÜM

#### LİTERATÜR ARAŞTIRMASI VE KAVRAMSAL ÇERÇEVE

1.1. Sosyal Medya Analizi Literatür Taraması.....	5
1.2. Havayolu Firmalarında Müşteri Memnuniyeti Literatür Taraması .....	8
1.3. Müşteri Memnuniyeti.....	10
1.3.1.Havayolu Firmalarının Müşteri Memnuniyeti ve Müşteri Şikâyetleri .....	11
1.4.Sosyal Medya Platformu Twitter .....	12
1.4.1.Sosyal Medya Platformu Twitter’da Geçen Terimler .....	13
1.4.2.Twitter API.....	14

### İKİNCİ BÖLÜM

#### BİLİMSEL ARKA PLAN

2.1.Makine Öğrenimi Yöntemleri.....	16
2.1.1.Gözetimli Öğrenme .....	16
2.1.2.Gözetimsiz Öğrenme .....	18
2.1.3.Takviyeli Öğrenme .....	19
2.2.Metin Madenciliği.....	19
2.2.1.Metin Madenciliği Adımları .....	20
2.2.1.1. Metin Ön İşleme.....	20
2.2.1.2.Metin Dönüştürme .....	21
2.2.1.3.Özellik Seçimi.....	21
2.2.1.3.1.Temel Bileşen Analizi (Principal Component Analysis – PCA) .....	21



2.2.1.4. Verilerde Kayıp Değer .....	22
2.2.1.4.1. Beklenti Maksimizasyonu (Expectation Maximization-EM) Algoritması Yöntemi.....	23
2.2.1.5. Veri Madenciliği Yöntemlerinin Kullanımı .....	24
2.2.1.6. Yorumlama Değerlendirme .....	24
2.2.2. Bilgi Çıkarımı (Information Extraction).....	24
2.2.3. Özetleme .....	25
2.2.4. Kategorizasyon.....	25
2.2.5. Kümeleme .....	26
2.2.6. Kavram Bağlantılama .....	26
2.2.7. Bilgi Görselleştirme.....	27
2.2.8. Soru Cevaplama.....	27
2.3. Kelime Torbası .....	27
2.4. Vektör Uzay Modeli.....	28
2.5. Sınıflandırma Algoritmaları .....	29
2.5.1. Karar Ağaçları .....	30
2.5.2. Yapay Sinir Ağları.....	31
2.5.3. Destek Vektör Makinesi .....	32
2.5.4. Naive Bayes .....	33
2.6. Derin Öğrenme .....	35
2.6.1. Konvolüsyon Katmanı.....	37
2.6.2. Aktivasyon Fonksiyonu.....	38
2.7. Kelime Gömme (Word Embedding) .....	40

## ÜÇÜNCÜ BÖLÜM

### TWITTER'DAKİ PAYLAŞIMLARA GÖRE İKİ FARKLI UÇUŞ FİRMASINA AİT DUYGU ANALİZİ

3.1. Araştırmanın Amacı .....	42
3.2. Araştırma Sorunları .....	42
3.3. Araştırmada Evren ve Örneklem .....	42
3.4. Sosyal Medyada Metin Madenciliği .....	42
3.5. Veri Toplama Araçları ve Veri Analizi Araçları .....	43
3.5.1. C# Programı .....	43
3.5.2. Sql Server .....	44
3.6. Verilerin Elde Edilmesi ve Analizi .....	44

3.6.1. Örnek Bir Data Contract .....	47
3.6.2. Linq2Twitter ile Tweet Çekmek .....	49
3.6.3. Çekilen Tweet Verilerini Süzmek ve Kaydetmek .....	51
3.6.4. Çekilen Verileri Saydırma İşlemi .....	65
3.7.Bulgular ve Değerlendirme .....	69
3.7.1.Klasik Sınıflandırma Analizi ve Değerlendirmesi .....	69
3.7.2.Derin Öğrenme Tabanlı Kelime Gömmme Analiz Sonuçları .....	74
<b>SONUÇ</b> .....	<b>80</b>
<b>KAYNAKÇA</b> .....	<b>83</b>
<b>ÖZGEÇMİŞ</b> .....	<b>92</b>



## KISALTMALAR DİZİNİ

Vd: Ve Diğerleri

Vb: Ve Benzeri

RT: Retweet

DM: Direkt Mesaj

TT: Trending Topic

API: Application Programming Interface

HTML: Hyper Text Markup Language

XML: Extensible Markup Language

TF-IDF: Term Frequency-Inverse Document Frequency

ARFF: Attribute Relationship File Format

SQL: Structured Query Language

JSON: Java Script Object Notation

PCA: Principal Component Analysis

RELU: Rectified Linear Unit

KNN: K Nearest Neighborhood

DVM: Destek Vektör Makinaları

YSA: Yapay Sinir Ağı

SCG: Scaled Conjugate Gradient

LM: Levenberg-Marquardt

BR: Bayesian Regulation

## TABLÖLAR DİZİNİ

Tablo 1: Twitter Ortamından Çekilen Örnek Tweetler.....	50
Tablo 2: Twitter Ortamından Çekilen Örnek Tweetlerin Temizlenmiş Şekli.....	53
Tablo 3:Twitter Ortamından Çekilen Örnek Tweetlerin Sınıflara Ayrılması .....	58
Tablo 4:Klasik Sınıflandırıcılar Analiz Sonuçları .....	70
Tablo 5:YSA Standart Analiz Sonuçları .....	71
Tablo 6:YSA Normalize Analiz Sonuçları.....	72
Tablo 7:YSA PCA Analiz Sonuçları .....	73
Tablo 8:Tokenization Sonucu.....	75
Tablo 9:Derin Öğrenme Ağına Ait Katmanlar.....	77

## ŞEKİLLER DİZİNİ

Şekil 1: Makine Öğrenimi Gözetimsiz Öğrenme Öbekleme Örneği.....	18
Şekil 2: Metin Madenciliği Adımları.....	20
Şekil 3: Vektör Uzay Modeli.....	29
Şekil 4: Yapay Sinir Ağının Genel Görünümü.....	36
Şekil 5:Vektörler Arasındaki Yoğunluk Gösterimi (Demirtaş, 2018: 14). ....	41
Şekil 6:API Widget.....	45
Şekil 7: Data Members Değerler .....	46
Şekil 8:Data Contract.....	47
Şekil 9: Oluşturulan Örnek Methodlar.....	49
Şekil 10:Cloud Server .....	64
Şekil 11:Windows Servisi .....	65
Şekil 12: Kelime Bulutu.....	75
Şekil 13: Confusion Matris .....	77
Şekil 14: ROC Eğrisi .....	78
Şekil 15: Kelimelerin Yakınlık Uzaklık Görünümü.....	79

## GİRİŞ

Duygular neredeyse tüm insan faaliyetlerinin merkezinde yer almakta olup davranışlarımızın, yaşantılarımızın önemli unsurlarındandır. Bu nedenle duygular insan hayatında önemli bir yer kaplamaktadır. İnsan hayatında önemli bir yeri olan duygular, doğru olduğuna inandığımız düşüncelerimiz ile yaptığımız seçimler, kişilerin dünyayı nasıl gördüğü ve değerlendirdiği ile büyük ölçüde bağlıdır. Bu nedenle bir şeye karar vermemiz gerektiğinde başkalarının görüşlerini dinlemek isteriz. Bu olaylar sadece bireyler için değil organizasyonlar için de geçerlidir (Liu, 2012: 5).

Bir konu hakkında karar verilirken başkalarının görüşleri bilinmek istenir. Dünyada, firmalar ve kuruluşlar her zaman ürün ve hizmetleri hakkında tüketici veya kamuoyununa ulaşmak istemektedirler. Bireysel tüketiciler, bir ürünün satın alınmadan önceki mevcut kullanıcılarının görüşlerini ve benzer şekilde siyasi seçimlerle ilgili bir oy vermeden önce siyasi adaylarla ilgili başkalarının görüşlerini bilmek istemektedirler. Geçmişte, bir kişinin görüşe ihtiyacı olduğunda çevresindeki insanlardan fikir alırdı. Benzer şekilde bir organizasyon veya bir işin kamusal ya da tüketici görüşlerine ihtiyaç duyulduğunda anketler düzenlenmekte. Çünkü buradan elde edilecek bilgi ile organizasyon kendisini birkaç adım öne taşıyabilmektedir. Bu sebeple kamuoyu ve tüketici görüşlerini edinmek için birçok farklı alanda firma ortaya çıkmış; pazarlama, halkla ilişkiler, anket uygulayıcı ve siyasi kampanya firmaları için büyük bir ticaret fırsatı oluşmuştur (Liu, 2012: 8).

Depolanacak teknolojinin gelişmesi ve ucuzlaması ile internette, forum tartışmaları, bloglar, mikro bloglar, yorumlar ve sosyal paylaşım sitelerinde yayınlanan mesajlar giderek artmaktadır. Bireyler ve organizasyonların karar vermesi için bu platformdaki içerikler içinde bilgi gizli haldedir. Bu bilgi birey ve organizasyonlar tarafından kullanılmaktadır.

Günümüzdeki tüketiciler, bir ürünü satın almak istiyorsa, genellikle çevresindeki insanlardan fikir almak isteyecektir. Günümüzde bunun için internette herkese açık forumlarda birçok kullanıcı incelemesi ve tartışması bulunmaktadır. Bu kaynaklardan bilgi elde edebilmek varken firma, insanların ürün ile ilgili görüşlerine anketler veya kurulmuş olan odak gruplar ile yürütmek istememektedir. Fakat teknolojideki değişimin etkisiyle görüş sitelerinin artması, bireylerin ve firma sahiplerinin ürün ya da konu

hakkında fikir edinmede zorluk çekmesine neden olmaktadır. Birey ürün ya da konu ile ilgili siteleri tespit etmekte ve bunlar ile ilgili görüşleri özetlemekte zorluk çekebilmektedir. Bu nedenle duygu analizi sistemlerine ihtiyaç duyulmaktadır (Liu, 2012: 8-9).

Sosyal medya üzerinde oluşan veriler çok hızlı bir şekilde büyüme göstermektedir. Sosyal medya sayesinde son kullanıcı ile organizasyon ve popüler kişiler doğrudan bir araya gelmektedir. Örneğin geçmişte bir üründeki hatayı, şikâyet etmek için direk olarak o kuruma iletmekte zorluk yaşanırken, günümüzde sosyal medya sayesinde ürün hakkında paylaştığımız şikâyetler kurumlara daha kolay iletilebilmektedir. Bu şikâyetler diğer kullanıcılar tarafından da görüldüğü için daha çok farkındalık yaratmakta ve kurumların bu şikâyetler karşısında çözüm süresi oldukça kısalmaktadır. Sosyal medya sayesinde çıkarılan ürün ya da organizasyon hakkında çok hızlı geri dönüş alınabilmektedir. İnsanlar fikirlerini paylaşmaktan çekinmedikleri için bu da geri bildirimlerin, şikâyetlerin analiz edilmesi gereksinimini kaçınılmaz hale getirmektedir. Kurumlar ve popüler kullanıcılar bir sosyal medya uzmanıyla çalışmakta ve sosyal medyayı aktif bir şekilde kullanmaya özen göstermektedir. Bu alandaki yüksek kullanımın bir sonucu olarak, verilerin büyümesi hızlanmakta ve bu verilerin analizi için çok büyük kaynaklara ihtiyaç duyulmaktadır. (Özgirgin, 2016: 2)

Havayolu hizmetlerinde yaşanan yoğun rekabet ortamında firmaların piyasada varlıklarını sürdürebilmeleri için müşterilerini sadık müşteriye dönüştürmeleri gerekmektedir (Kazançoğlu, 2011: 137). Sadık müşteri için ürün ve hizmet kalitesinin yanında, sorunlara hızlı çözüm bulmak ya da rakip kuruluşların durumlarını incelemek oldukça önemlidir.

Şikâyetler işletmeler açısından birer fırsat ya da tehdit olarak görülmelidir. Memnun kalmayan müşteri şikâyette bulunmaz ise işletme müşterinin sorununu çözüp müşterilerini elde tutma şansını kaybedebilir (Ekiz ve Kökler, 2010: 2859). Şikâyetlerin hızlı bir şekilde çözülmesi ile birçok sektörde olduğu gibi, havayolu işletmeleri de müşterilerinin devamlılığını yeniden kazanma şansı bulurlar. Şikâyetlerin ortadan kaldırılmaması müşteri kayıplarına sebep olabilmektedir. Şikâyetler, havayolu işletmelerinde oluşan sorunları ortadan kaldırmak, problemleri çözmek, hizmet kalitesizliğini gidermek, yeni inovasyon fikirleri sunmak, yani hizmet kalitesini

geliştirebilmek için işletmeye yeni fırsatlar sunmaktadır. Birçok işletme gibi havayolu işletmelerinin kendilerini geliştirebilmeleri, kazancı artırabilmeleri ve devamlılığı sağlayabilmeleri için müşteri şikâyetlerini sözlü, yazılı veya sosyal medya ortamından dikkate almaları gerekmektedir (Ekiz ve Köker, 2010: 2859-2873).

Tim Berners-Lee tarafından 25 yıl önce “World Wide Web” halka açık hale getirilmiştir. Bunun kaçınılmaz bir neticesi olarak internet, günümüzde dünya nüfusunun çoğunun günlük hayatının ayrılmaz bir parçası haline gelmiştir. Dijital pazarlama firması olan We Are Social ve Hootsuite’in hazırlanmış olduğu “Digital in 2017 Global Overview” raporunda, 238 ülkeden toplanan internet ve dijitalin gelişimi ile ilgili güncel bilgiler, istatistikler ve trendler incelenmiştir. Global Digital raporuna göre hızlı bir şekilde büyüyen internet değil interneti kullananlardır. We Are Social’ın araştırmalarına göre (Kemp, 2018):

- Dünyanın yarısından fazlası en az bir tane akıllı telefon kullanıyor.
- Dünya nüfusunun neredeyse üçte ikisi en az bir tane cep telefonu sahibi.
- Dünya genelinde web trafiğinin yarısından fazlası akıllı telefonlardan geliyor.
- Dünyanın her yerindeki mobil bağlantıların yarısından çoğu “genişbant”.
- Dünya nüfusunun beşte birinden fazlası son 30 gün içinde alışveriş yaptı.

İstatistiklere göre günümüzde 4.38 milyarı bulan dünya çapında internet kullanıcı sayısı %56’lık bir penetrasyona eşittir. Bu nüfusun 3.48 milyarı aktif olarak sosyal medya platformlarını kullanmaktadır. Kullanıcıların 3.25 milyarı sosyal medyayı mobilden bağlanırken, toplam mobil kullanıcı sayısı ise 5.11 milyardır (Kemp, 2018).

Sosyal medya platformları arasında Facebook açık ara önde, ikinci ve üçüncü sırada Youtube ve Whatsapp uygulamaları ve dördüncü sırada Messenger yer almaktadır. 2018 yılında da sosyal medya platformları giderek artmakta olduğu bilinmektedir.

Türkiye’de ise aktif sosyal medya kullanıcı sayısı 52 milyon olarak belirtilmiştir. Fakat bu sayının tekil kullanım olmadığı çoklu kullanımların da yer aldığı raporda gösterilmiştir. Türkiye’deki mobil kullanıcı sayısı 71 milyon iken, sosyal medyaya bağlanan kullanıcı sayısı 42 milyondur. Aktif sosyal medya kullanıcı sayısının %14 artış ile 6 milyon olduğu görülmektedir. Ülkemizde sosyal medyayı mobilden kullanan kişi sayısı son bir yılda %17 oranında artmıştır.



Türkiye’de en çok kullanılan sosyal medya kanalları olarak birinci sırada Youtube bulunmaktadır ve onu ikinci sırada instagram takip etmektedir. Diğer popüler sosyal medya kanalları ise Facebook ve Twitter’dır (Kemp, 2018).

Twitter’daki verilerin büyük olması ve paylaşılan verilerin herkese açık olması sebebiyle bu çalışmada hedef platform olarak Twitter verileri kullanılmıştır. Kullanılan veriler Türkiye’de faaliyet gösteren iki büyük hava yolu firması adına atılan “#” tweetlerdir. 23.01.2018 ile 21.04.2018 tarihleri arasında “#” olarak atılan tweetler C# dilinde geliştirilen bir program sayesinde saat başı çalışarak bir veri tabanına kayıt edilmiştir. Birinci firma adına 1569335 toplam veri kayıt edilmiş olup ham veri toplamı 36464, retweet edilenlerin toplamı 54703, pozitif tweetlerin toplamı 612720, negatif tweetlerin toplamı 553644, nötr tweetlerin toplamı 402971 tanedir. İkinci firma adına 933036 toplam veri kaydedilmiş, ham veri toplamı 14075, sadece retweet edilenlerin toplamı 6894, pozitif tweetlerin toplamı 196765, negatif tweetlerin toplamı 434861, nötr tweetlerin toplamı 301410 tanedir. Çok fazla veri olduğu için iki hava yolu firmasına atılan tweetler arasından rastgele 30352 tweet seçilmiştir. Elde edilen tweetler daha sonra müşteri memnuniyetine göre olumlu olumsuz nötr sınıflarına ayrılmıştır. Daha sonra bu üç sınıf niteliğine göre tweetler üzerinde otomatik sınıflandırılma ve derin öğrenme algoritmaları uygulanmıştır.

## BİRİNCİ BÖLÜM

### LİTERATÜR ARAŞTIRMASI VE KAVRAMSAL ÇERÇEVE

#### 1.1. Sosyal Medya Analizi Literatür Taraması

Teknoloji çağının yaşandığı 21. yy da iletişim ve haberleşmenin yanında anlamlandırma ve ifade etme platformu sosyal medya aracılığıyla gerçekleştirilir. Farklı pek çok mecra üzerinden yansımaları görülen bu durum kişilerin bireyleşme süreçlerini, toplumsal anlamda kabul görme arzu ve kaygılarını işaretler. Dolayısıyla sosyal medya ve burada yapılan paylaşımlar duyuşsal tahliller açısından önem kazanır. Yurt içinde ve yurt dışında bu hususta pek çok araştırma yapılmıştır.

Jiang ve Zhai (2007) tarafından yapılan çalışmada duygu analizi kapsamında alan adaptasyonu probleminin genel bir örnek ağırlıklandırması üzerine çalışılmıştır. Çalışmada uyarılma için birçok farklı stratejiyi desteklemek için esnekleştirilmiştir. İki aşamalı yaklaşımda genelleştirme aşaması ile alanlar arasında ortak kullanılabilen öznitelikler çıkarılmakta, adaptasyon aşamasında ise yarı gözetimli öğrenme kullanılarak hedef alana özgü öznitelikler elde edilmiştir.

Türkçe duygu analizi konusunda ilk çalışanlardan Eroğul (2009), İngilizce alanındaki çalışmalarını Türkçe verilerde uygulanabileceğini araştırmış ve Türkçe çalışmalar için yeni yöntemler sunmuştur. Farklı yöntemler uygulanarak Türkçe ve İngilizce sonuçlar karşılaştırılmıştır. Türkçe film yorumları için duygu analizi üzerine doğal dil işleme yöntemi ile çalışma yapılmış, verileri beyazperde.com sitesinden elde ettiğini belirtmiştir. Veriler N-gram model kullanarak olumlu olumsuz kriterlerde sınıflandırılıp, %85 başarı elde edilmiştir.

Albayrak (2011) çalışmasında, Türkçe'nin kullanımı ile psikolojik durum arasındaki ilişkiyi araştırmıştır. Depresyonlu, depresyonsuz, anksiyeteli, anksiyetesiz kişilerden Türkçe yazılan veriler toplanmış olup, veriler her bir kelime, her bir tanı grubu tarafından en çok kullanılan kelime grupları, kipler, kişi zamirleri, fiiller ve isimler bu özelliklerin kullanımına göre analiz edilmiştir. Çalışma sonucunda Türkçe'de kelime kullanımının psikolojik durum hakkında birçok ipucu verdiği görülmüştür.

Bollegala ve ark. (2011) çalışmalarında, birden fazla veri setinden etiketli ve etiketsiz veri, hedef bölgesinde ise sadece etiketsiz veri kullanarak duygu analizi

sözlüğü oluşturmuş ve farklı alanlardaki benzer duygu belirten kelimeler arasındaki ilişkileri çıkartmışlardır.

Akbaş (2012) tarafından yapılan çalışmada, konu temelli duygu çıkarımı yapan bir sistem geliştirmiştir. Türkçe tweetler üzerindeki başlıklar ve alt konuları çıkararak, konulara göre gruplandırma yapmıştır. Düşünce analizi problemi üzerine yaptığı çalışmada kelimelerin otomatik olarak duygusunu belirlemiştir.

Uytun (2013) tarafından, duygu madenciliği ve doğal dil işleme yöntemleri kullanılarak kullanıcı verilerinden işlevsel ve anlamlı çıkarımlar konusunda bir program geliştirme hedeflenmiştir. E-ticaret sitesindeki kullanıcı yorumları ve anketleri veri olarak kullanılmıştır.

Sevindi (2013) çalışmasında, Türkçe film yorumları için duygu analizi üzerinde çalışmıştır. Bu amaçla film yorumlarının duygu kutupları, makine öğrenmesi ve sözlük tabanlı yaklaşım yöntemleri kullanılarak belirlenmeye çalışılmış ve bu yöntemlerle karşılaştırılmıştır. Yapılan çalışma sonucunda makine öğrenimi yaklaşımında destek vektör makinesi sınıflandırıcısıyla 0.8258 F-skor değeri, sözlük tabanlı yaklaşımda ise 0.5969 F-skor değeri elde edilmiştir. Çalışmanın sonuçları karşılaştırıldığında makine öğrenmesi yaklaşımlarıyla daha sağlıklı sonuçlar elde edildiği görülmüştür.

Zhou ve arkadaşları (2013) tarafından yapılan çalışmada aktif derin ağ olarak adlandırılan yeni bir yarı denetimli öğrenme algoritması sunulmuştur. Duygu sınıflandırma yönteminde beş duygu içeren veri seti üzerinden yapılan çalışmada, aktif derin ağın klasik yarı denetimli öğrenme algoritmalarından daha başarılı sonuçlar verdiği görülmüştür.

Akba (2014) tarafından yapılan çalışmada, duygu analizi problemlerinin başarı oranları karşılaştırılarak inceleme işlemleri yapıp, insan gücünden daha az yararlanılacak bir sistem geliştirilmesi hedeflenmiştir. Çalışmada kullanılan veriler internet sitesindeki kullanıcıların Türkçe film yorumlarını ve puan sistemini ele almıştır. Kullanılan sistemde olumlu ve olumsuz kriter yorumları karşılaştırmasında %83.9 başarı değeri bulunmuştur. Olumlu, olumsuz ve nötr kriter yorumları karşılaştırmasında ise %63.3 başarı değeri elde edilmiştir.

Tang ve arkadaşları (2014) çalışmalarında, Twitter verileri üzerinde duygu sınıflandırması için kelime gömmeyi öğrenen bir yöntem geliştirmişlerdir. Kelimelerin

temsillerindeki duygu bilgileri devamlı kodlanmış, duyguya özel kelime gömme işlemleri yapılmıştır. Denetlenmiş metinlerin duygu polaritelerini kayıp fonksiyonlarına etkin bir şekilde eklemek için sinir ağları sistemi geliştirilmiştir.

Pak (2015) yaptığı çalışmada, alan sınıflandırma temelli duygu sınıflandırma yaklaşımı uygulanmıştır. Uygulanan yaklaşım sayesinde Türkçe ve İngilizce veriler içeren iki ayrı veri kümesi kullanılmıştır. İki bölümden oluşan çalışmanın ilk bölümünde, alanı belli olmayan metinlerde etiket verisi olan mevcut bölümlerden hangisine kabul edileceği bulunmaktadır. İkinci bölümde ise metinleri kabul edilen bölümlere ait duygu sınıflandırıcı kullanılarak sınıflandırılmaktadır.

Yurt (2015), Türkçe metinlerde doğal dil işleme algoritmaları kullanarak başarılarının test edilmesini hedeflemiştir. Türkçe veriler internet ortamından elde edilip, ön işlemlerden geçirilerek veri madenciliği ve makine öğrenmesi algoritmaları ile analiz yapılmıştır.

Özgirgin (2016) çalışmasında, sosyal medya verilerinin sınıflandırılmasını ve sınıflandırma işleminden sonraki işlemlerde gerçek problemlerin ve niteliklerin belirlenmesini amaçlamıştır. Elde edilen sosyal medya verileri bir ön işlemde geçirilerek metinleri düzeltme işlemleri yapıp, makine öğrenmesi teknikleri kullanılarak olumlu olumsuz sınıflar için analizler gerçekleştirilmiştir.

Çoban (2016), Twitter'dan elde edilen Türkçe veriler üzerinde iki farklı duygu analizi çalışması yapmıştır. Twitter'dan elde edilen verilerin duyguların otomatik olarak tespitinde makine öğrenmesi yöntemleri kullanılmıştır. Etiketleme yöntemi Multinom Basit Bayes yöntemi ile %92.5 başarı sağlanmıştır. Çalışmada Türkçe ve İngilizce verileri işleyebilen bir yazılım örneği geliştirilmiştir.

Kaplan (2016) çalışmasında, sosyal medya sitesi Twitter kullanıcılarının paylaştığı Tweetleri analiz etmiştir. Elde edilen verilerin duyguları dört farklı grup olarak sınıflandırılmıştır. Bu sınıflar 'Şaşkınlık', 'Mutluluk', 'Kızgınlık', 'Üzüntü'dür. Verilerin sınıflandırılması için yazım hatalarının giderilmesinde 'Zemberek' programı kullanılmıştır. Hataları giderilen veriler karar ağacı analizinde %85.372, bulanık kural öğrenme analizinde %83.608 başarı sonuçlarına ulaşmıştır.

Dündar (2016), iki farklı veri seti üzerinde çalışma yaparak, iki farklı uygulama geliştirmeyi hedeflemiştir. Bu çalışmada ürün varlıklarını öğrenmeye ve

sınıflandırmaya, düşünce ifadelerini otomatik olarak ürün yorumlarından çıkarmayı ve özetlemeyi hedeflemiştir.

Amanet (2017) tarafından yapılan çalışmada, sosyal medyadaki Twitter metin verileri üzerinde duygu analizi yapılması hedeflenmiştir. Bu çalışma için belirlenen duygular ise, “Mutlu”, “Öfke”, “Takdir Etmek”, “Gurur Duyma”, “Beklenti”, “Hayal Kırıklığı”, “Tavsiye”, “Merak”, “Güvenmek” olarak belirlenmiştir. Sosyal medya metinleri sınıflandırılmış ve her duygu için öznitelikler belirlenmiştir. Çalışmadaki veriler duygu analizi için metin madenciliği sınıflandırma yöntemlerinden yalın bayes, k-en yakın komşu, destek vektör makineleri ve karar ağaçları kullanılmıştır. Çalışmanın sonucunda takdir duygusu %65, mutluluk duygusu %93.19, öfke duygusu %84.51 oranında doğru sınıflandırılmıştır.

Demirtaş (2018) çalışmasında, yenilenen sinir ağları ve kelime gömme yöntemlerini birleştirerek geliştirdiği modelde farklı aktivasyon fonksiyonlarının başarımlarını karşılaştırmıştır. Klasik sınıflandırma algoritmaları olarak destek vektör makineleri, çoklu nominal naive bayes, rastgele orman, lojistik regresyon ve k-en yakın komşu algoritmaları kullanılmış ve hem kendi aralarında hem de derin öğrenme yöntemi ile karşılaştırılmıştır. Türkçe ve İngilizce veri setlerinde derin öğrenme yöntemi bazı aktivasyon fonksiyonları ile klasik sınıflandırma yöntemlerine göre daha başarılı sonuçlara ulaşmıştır.

Yabancı kaynaklarda daha çok duygu analizine yönelik adaptasyon sorunları üzerine çalışmalar yapılmıştır. Türkçe yapılan çalışmalarda ise daha çok internet ortamında bulunan Twitter, film siteleri gibi sitelerden elde edilen verilerin duygularının ölçülmesi hedeflenmiştir.

## **1.2. Havayolu Firmalarında Müşteri Memnuniyeti Literatür Taraması**

Hava yolu taşımacılığının önem kazanması küresel piyasada etkin konuma erişmek isteyen hava yolu firmalarını hizmet standartlarını arttırma zorunluluğuna sevk etmektedir. Rekabet gücünü gözetmek ve kalite düzeylerini yükseltmek amacı taşıyan firmalar hemen her dönem müşteri memnuniyeti araştırmaları yapmaktadırlar. Söz konusu araştırmalarda yurt içi ve yurt dışında yapılan bazı çalışmalar aşağıda listelenmiştir.

Tsaur vd. (2002) yaptığı araştırmanın amacını havayollarında algılanan hizmet kalitesinin ortaya konulması oluşturmaktadır. Tayvan'da havayolu firmasının birinde seyahat eden 211 müşteri üzerinde gerçekleştirilen çalışmanın sonucunda, havayollarında algılanan hizmet kalitesinin en üst düzeyde fiziksel özellik boyutunun etkilediği sonucu ortaya çıkmıştır.

Park, Robertson ve Wu (2005) tarafından yapılan çalışmada, ekonomi sınıfında yer alan müşterilere servqual analizi kullanılmıştır. Ekonomi ve diğer sınıflar arasında farklılık olup olmadığına bakılmıştır. Analiz sonrasında elde edilen sonuçlarda havayolu hizmet kalite boyutları ile havayolu işletmesi imajı ve müşterilerin algısı arasında istatistiksel olarak anlamlı ve pozitif yönde ilişki olduğu sonucuna varılmıştır.

Bozorgi'nin (2006) çalışmasında, İran havayolları Grönross firması temel alınmış ve hizmet kalitesi değerlendirilmiştir. Araştırmada, müşterilerin kalite boyutlarından sırasıyla teknik kalite, fiziki unsurlar, güvenilirlik, isteklilik, imaj ve duygudaşlık boyutlarına önem verildiği anlaşılmıştır.

Okumuş ve Asil (2007) tarafından yapılan çalışmada, hizmet kalitesi algılanmasında havayolu müşterilerinin genel düzeyde memnuniyetlerine olan etkisi araştırılmıştır. Fiziki unsurlar dışındaki dört faktörün genel düzeyde memnuniyetleri üzerinde olumlu ve anlamlı etkisi olduğu araştırma sonucunda görülmüştür.

Yıldız ve Erdil (2013), havayolu firmaları müşterileri üzerinde yaptıkları çalışmada hizmet kalitesini ölçen iki farklı modeli karşılaştırmışlardır. Çalışmada ağırlıklı servperf ölçeğinin algılanan hizmet kalitesini, ağırlıklı servqual'dan daha çok açıkladığı ortaya çıkmıştır.

Pekkaya ve Akıllı (2013) tarafından yapılan çalışmada, Türkiye'de bulunan 8 havayolu firmasının müşterilerinin hizmet kalitesini hangi ölçüde değerlendirdiği ve hangi hizmet kalitesi boyutlarına daha çok önem verdiği araştırılmıştır. Müşteri memnuniyetsizliğinin belirgin bir şekilde heveslik boyutunda gözlemlendiği, diğer boyutlarda ise müşterilerin memnun olduğu çalışma sonucunda ortaya çıkmıştır.

Hatipoğlu ve Işık (2015) çalışmalarında, Anadolujet havayolu firmasının iç hat yolcu müşteri profilini ortaya çıkarmayı ve hizmet kalitesini ölçebilmeyi amaçlayarak servqual analizi ile anket çalışması gerçekleştirmişlerdir. Çalışma sonucunda elde edilen

sonuçlarda firmanın müşteri beklentilerini büyük oranda karşıladığını göstermiştir. Heveslilik durumu algı-beklenti farkını en az -0.45 olarak tespit etmişlerdir.

Hava yolu müşteri memnuniyeti konularında yapılan yabancı kaynaklarda hizmet kalitesinin ölçülmesi hedeflenmiştir. Bu hizmet kaliteleri fiziki unsurlar, güvenilirlik gibi konular üzerinde yoğunlaşmıştır. Türkçe kaynaklarda ise müşteri memnuniyet düzeyinin, müşteri profillerinin ortaya çıkarılarak ölçülmesi hedeflenmiştir.

### **1.3. Müşteri Memnuniyeti**

Günümüz yönetim anlayışında tüketiciler ve müşteri memnuniyeti firmaların en önemli konuları olarak görülmektedir. Hizmet firmalarının içerisinde yer alan havayolu firmaları da günümüzde yüksek kalite ve kusursuz hizmet sunma çabasıdadır. Bu nedenle müşteriler havayolları tarafından sunulan hizmetlerden memnun kalıp önerilerde bulunabilirler ya da memnun kalmayıp şikâyetlerde bulunabilmektedirler.

Müşteri memnuniyeti; tüketicilerin satın alma öncesi beklenti ve istekleri ile satın alma sonrası performans algılamalarını karşılaştırmaları sonucunda oluşan etkenlerdir. Müşteriler bu karşılaştırmaları devamlı karşılaştıkları durum veya arada karşılaştıkları duruma göre ve tüm konuları bir arada değerlendirerek yaparlar. Hizmet işletmelerinde de müşteriler karşılaştıkları her durum için memnuniyet değerlendirmesi yaparlar. Aynı zamanda işletmenin hepsinden oluşan genel memnuniyeti değerlendirirler (Tüzün ve Devrani, 2008, 15).

Hizmet firmaları; müşterilerinin öneri, şikâyet ve beklentilerini daha iyi anlayıp çözebilmek için ürünlerin ve hizmetlerin kalitesinin müşteriler tarafından değerlendirilmesi için araştırmalar yapmaktadır. Çok fazla rekabetin yaşandığı pazar ortamında firmaların başarılı olması, tüketiciye daha yakın durabilmesine ve müşteri isteklerini tatmin edebilmesine dayanmaktadır. Bu nedenlerden dolayı müşteri memnuniyetinin ölçülmesi günümüzde firmaların en çok uyguladığı pazarlama araştırmalarından biridir (Ofir ve Simonson, 2001).

Müşteri tatmini bir sonuç veya bir süreç olarak iki temel biçimde tanımlanmaktadır. Sonuç olduğunu ileri süren yaklaşımlarda, tatminin tüketim deneyimlerinden dolayı kaynaklanan nihai bir durum olduğu düşünülmektedir. Bu yaklaşımın tersine göre tatmin, bir de süreç olarak değerlendirilmektedir. Tatmin

seviyesine katkıda bulunan düşüncelerde, değerlendirmeler ve psikolojik süreçler üzerinde durulmuştur (Vavra, 1999, 18).

Günümüzde hizmet sektöründe, müşteri ile ileriye yönelik uzun dönemli ilişkiler kurabilen firmaların pazarda rekabet etme ve başarı elde etme şansları yüksektir. Bundan dolayı, uzun dönemli müşteri ilişkilerini oluşturabilmek için ortaya konulan hizmetten müşteriler, beklentilerinin de üstünde memnun kalmalıdır. Müşteriler, günümüz rekabet piyasasında satılan mal ve hizmetlerin teknolojik gelişmişliği ve farklı ürünler sebebiyle daha önceki piyasaya göre seçici davranmaktadır. Eskisi gibi kolay şekilde memnun olmamakta ve en küçük bir problemde dahi ürününü ya da hizmetini kullandığı firmayı değiştirebilmektedirler. Bu sebeplerden dolayı müşterinin şuanki ve gelecekteki ihtiyaç, beklenti ve isteklerini tahmin eden ve bu istek ve ihtiyaçlarını yerine getirebilmek için hizmet ve mal geliştirme, çeşitlendirme yaparak müşterilerini memnun etmeyi başarabilen firmaların rekabet üstünlüğü her zaman yükselmektedir. Bundan dolayı firmalar, müşterilerin hizmet ve maldan beklentilerini doğru tahmin etmek, beklentilerinin ne kadarlık kısmını karşılayabildiklerini bilmek ve müşterilerinin memnun olmadığı durumları belirleyip tatmin seviyesini arttırmak için yeni çalışmalar yapmak durumundadırlar (Aksu, 2012, 71-72).

### **1.3.1.Havayolu Firmalarının Müşteri Memnuniyeti ve Müşteri Şikâyetleri**

Havayolu firmalarında yaşanan yoğun rekabet ortamında, müşteri memnuniyetinin sağlanmasında firma imajı ve algılanan hizmet kalitesi iki önemli unsurdur. Günümüzde havayolu firmalarının fiyat odaklı yaklaşımla pazar payının arttırılmaya yönelik çalışmaları, orta ve uzun dönemde kârlılığa kötü etki yaparak, firma imajını olumsuz anlamda etkilemektedir. Bundan dolayı fiyata göre üstün kaliteyle farklılaşmanın sağlanması, müşteri memnuniyetini elde etmede daha etkili bir durumdur (Kazançoğlu, 2011: 137).

Havayolu firmalarındaki hizmet kalitesi, havayolu firmalarının müşteri ve havayolu personelleri arasındaki farklı etkileşimler sırasında mükemmel hizmet standartlarını sunarken, aynı zamanda havayolu firmalarının müşterileri gitmek istedikleri yerlere taşınması ile de ilgilidir (Suki, 2014: 26).

Havayolu firmalarının ayakta kalması için müşterilere yüksek kalitede hizmet sağlamaları en önemli husustur. Havayolu hizmet memnuniyetinin; müşteri



memnuniyeti, müşteri sadakati ve müşterilere yönelik havayolu seçimi açısından önemli bir güç olduğu bilinmektedir (Singh, 2015: 109).

Havayolu firmalarında fiyat, müşterileri elde etmek için ilk yol olarak daha fazla kullanılıyor olsa da; bazı havayolu firmaları, hizmetlerini farklılaştırarak rekabet üstünlüğü elde etmek için hizmet kalitesine daha çok önem vermektedir. Bunun sebebi diğer firmaların fiyat değişikliklerine nispeten cevap verebiliyor olmalarıdır. Havayolu firmalarında asıl rekabet üstünlüğü, müşteriler tarafından algılanan hizmet kalitesindedir. Bundan dolayı kaliteli hizmetin sağlanması, havayolu işletmelerinin ayakta kalmasının en önemli unsurudur (Namukasa, 2013: 522).

Havayolu firmaları, rekabetçi yolculuk elde etmelerine yardımcı olacak yolcuları ile uzun süreli bir ilişki sağlamak için, daha çok yolculuk programları gibi müşterilere katma değerli hizmetler sunabilmektedir. Müşterilere kaliteli hizmetler sunmak, müşterilerin sürekliliklerini sağlamak için yeterli olmayabilir, çünkü müşteriler aldıkları hizmetlerde ücret ve hizmet kalitesi ilişkisini önemsemektedir (Singh, 2015:110).

Hava yolu firmaları için müşteri şikâyetlerinin ortadan kaldırılması, müşteri memnuniyetini sağlaması ve buna ek olarak da müşteriyi elde tutması ekonomik olarak önem arz etmektedir. Müşteri şikâyetlerinin dikkate alınması ve çözüme ulaştırılması, müşterilerin çevrelerine sosyal medyadan ulaştıkları için yüz yüze olumsuz haberleri azaltarak firmanın imajının zedelenmesini ortadan kaldıracaktır (Kozak, 2007: 139).

Hizmet kalitesine önem veren havayolu firmaları, müşterilerin beklentileri uzantısında hizmetlerinde farklılıklar yapmakta ve müşterilere hizmet sunarken samimi davranmaya önem göstermektedir (Okumuş ve Asil, 2007:11). Bundan dolayı müşteri şikâyetlerini önemseyen firmalar bu şikâyetler sayesinde müşterilerini daha iyi anlayabilmekte ve hizmet kalitesini düzeltebilmektedir. Havayolu firmaları için müşterileri kendine bağlama önemli bir durumdur. Müşteri istek ve ihtiyaçlarını en iyi durumda karşılayabilmeli, en iyi hizmeti alabilmeli ve böylece devamlılığı sağlanabilmelidir.

#### **1.4.Sosyal Medya Platformu Twitter**

Twitter, Jack Dorsey tarafından 2006 yılında kurulmuştur. Twitter, kullanıcılarına 140 karakterle sınırlı metin, fotoğraf ve video paylaşabilme imkânı sağlayan sosyal medya platformudur. Twitter kullanıcıları akıllı telefon, tablet ve

bilgisayar araçlarıyla internet bağlantısının olduğu herhangi bir yerden bağlanarak, paylaşımında bulunabilir ve diğer kullanıcılarla iletişime geçebilirler (Amanet, 2017: 2).

Türkiye’de Twitter, oldukça önemli bir kullanıcı sayısına sahiptir. 25 Nisan 2011 tarihinden itibaren Türkçe dil desteği ile ülkemizde servis vermeye başlamıştır. Bunun doğal neticesi olarak kısa sürede milyonlarca Türk üyeye ulaşmıştır. Ülkemizde akıllı telefon ve tabletlerin satışıyla birlikte, birçok sosyal paylaşım sitesi gibi Twitter da önemli bir yere sahip olmuştur ve kullanıcılar 7/24 Twitter hesaplarına giriş yaparak tweet atabilmektedir (Çağıl, 2018).

Gün geçtikçe hızlı olması ve kısa bilgi paylaşımları içermesi sebebiyle Twitter birçok kişinin hayatının doğal bir parçası haline gelmiş ve kendisine özgü kısaltmaları da bütün dünya da kullanıcılar tarafından benimsenmiştir.

#### **1.4.1.Sosyal Medya Platformu Twitter’da Geçen Terimler**

- Twitter’da her yazılan mesaja “Tweet” adı verilmektedir. Metin mesajları, canlı yayınlar, video ve fotoğraf paylaşımı yapılan her mesaja “Tweet” denir.
- Retweet yani ‘RT’, diğer kullanıcılar tarafından gönderilen bir Tweet’in, alıntı yapılarak kendi kullanıcı hesabımızdaki takip ettiğimiz kullanıcılara aynı şekilde paylaşılmasıdır. RT işlemi başka kullanıcıların tweetlerinin altında yer alan ‘Retweet’ butonuna basılarak yapılır. RT yapılan tweetler RT yapan kullanıcının sayfasında asıl Tweeti paylaşan kullanıcının adı ve Retweetledi başlığı ile paylaşılır.
- Direkt mesaj ‘DM’, Twitter kullanıcıları arasında sadece birbirlerine özel mesaj yoluyla iletişime geçme işlemine denir.
- Yanıtlama, paylaşılan Tweet metninde başka bir kullanıcının “@kullanıcıadi” şeklinde geçen bir mesajdır. Bir Tweet de ‘@’ sembolü kullanıldığında o mesaj sadece o kullanıcı veya kullanıcılara paylaşılmaktadır.
- Trending Topic ‘TT’, o anki gündemde popüler olan konu üzerine kullanıcıların en çok paylaşımında buldukları başlıktır. Twitter’in ana sayfasında sol kısımda yer alan listede ilk 10 konudur. TT’ler ülkelere göre özel olarak değişim gösterdiği gibi yer konumları aracılığıyla bulunan bölgelere veya şehirlere özel olarak da yer almaktadır.

- Etiket '#', Twitter'da bir paylaşımda '#' ifadesi konularak yazılan kelimelere etiket denir. Kullanıcılar ya da kuruluşlar paylaştıkları Tweetleri kategorize ederek ayırırlar ve Twitter arama bölümünde tweetleri daha kolay gösterilmesine yardımcı olur. Tweetlerindeki anahtar kelimenin başına '#' ifadesi koyulmaktadır.

Duygu analizi kapsamında birçok çalışma yapılmaktadır. Bu kapsamda en çok kullanıma sahip sosyal medya platformu olan Twitter verileri kullanılabilir. Twitter uygulama ara yüzleri sayesinde Twitter sistemindeki verileri elde etme imkânı sunmaktadır. Twitter kullanıcılarının profilinde yer alan kullanıcı adı, paylaşılan Tweet sayısı, takipçi sayısı ve takip edilen kullanıcı sayılarına ulaşma imkânı vermektedir (Amanet, 2017: 2-3).

Ülkemizde sosyal medya kullanımının artması ile birlikte, sosyal medya kullanıcıları aynı anda birden fazla kullanıcıyla iletişime geçme şansı bulmaktadır. Bu durumda başka ihtiyaçlar ortaya çıkabilmektedir. Milyonlarca kullanıcısı olan Twitter ağında saniyede binlerce Tweet paylaşmaktadır. Büyük miktarda verinin bulunduğu Twitter ortamında, paylaşılan Tweetlerin duygu ve düşüncüyü analiz etmek için kullanılması büyük önem arz etmektedir (Amanet, 2017: 3).

#### **1.4.2. Twitter API**

Twitter uygulama geliştirme ara yüzü (API: Application Programming Interface), bünyesinde farklı özellikler barındırmaktadır. Bu özellikler Rest, Search ve Stream kütüphaneleridir. Program geliştiricileri tarafından sunulan ve bu kütüphaneleri kullanan diğer yazılımlarda mevcuttur. Geliştirilen bu yazılımlar sayesinde sosyal medya platformu Twitter'da paylaşılan verileri yani tweetleri elde etmek kolaylaşmıştır. Diğer işlemler ise; Twitter mesaj paylaşımı, Twitter kullanıcılarının gönderdiği mesajları sıralayabilme, farklı parametrelere (konum, dil, tarih vb.) bağlanarak sorgu gönderme gibi sıralanabilmektedir. Bu gönderilen sorgular terim, sözcük veya '#' sembolü içerebilmektedir. Twitter'ın gizlilik ve güvenlik şartları nedeniyle herkese açık olan mesajları elde etmek mümkün olmaktadır (Çoban, 2016: 30).

Twitter API'si Twitter verilerine programatik olarak okuma ve yazma işlemlerinin yapılmasını sağlar. Twitter API sayesinde programcılar Twitter'ın verilerine kolayca ulaşım sağlayabilir. Diğer sosyal medya platformlarında olduğu gibi

Twitter API'si kullanılarak iletiřim saęlayabilmek iin yetkilendirme amacıyla bir anahtar (key) alınması gerekir.

Bu alıřmada duygu analizinde kullanılan veriler, yukarıdaki nedenlerden dolayı sosyal medya platformu olan Twitter üzerinden Twitter API kullanılarak elde edilmiřtir.



## İKİNCİ BÖLÜM

### BİLİMSEL ARKA PLAN

#### 2.1.Makine Öğrenimi Yöntemleri

Bir sorunu bilgisayar ortamında çözmek için algoritmaya ihtiyaç duyulur. Algoritma; herhangi bir sonuç elde etmek için girilen parametreleri daha önceden tanımlanmış bir takım işlemlerden geçirerek sonuç çıkaran veya üreten işlemlerin bütünü olarak tanımlanmaktadır. Örnek olarak bir sıralama algoritması girilen sayılar kümesini, tanımlanan bir takım işlemlerden geçirdikten sonra azalan ya da artan bir biçimde sıralanmış olarak üretim çıktısı oluşturur. Fakat durmadan spam olarak gelen e-postaların temizlenmesi gibi işlemler için bulunan bir algoritma yoktur. En kolay haliyle basit bir e-posta dosyasının yazı karakterinden oluşan algoritmaya girdi olarak gönderileceği ve bu e-postanın spam olup olmadığını bulan “evet” veya “hayır” gibi bir durumun üretileceği bilinmemektedir. Bazen hangi e-postanın spam olduğu zamanla değişebilmekle birlikte alıcı ve gönderen kişiye göre de farklılık gösterebilmektedir. Spam olduğu bilinen milyonlarca e-posta kullanılarak spam e-postaların genel özellikleri çıkarılabilir. Bilgisayar sayesinde saniyede milyonlarca işlem yapabilen bir makine öğrenimi algoritması ile bu tarz işlemler için otomatik algoritmalar geliştirilebilmektedir. Makine öğreniminin temel konusu gözlenmiş bir örneklem kümesinden çıkarım yapmaktır. Bunun için model oluşturmada istatistik alanı kullanılır (Alpaydın, 2010: 1).

Metin sınıflandırma ve duygu analizi araştırmalarında kullanılan yöntemlerden biri makine öğrenimidir. Metin sınıflandırma araştırmalarında,  $D=\{X_1, X_2, \dots, X_n\}$ , eğitim veri kümesi ile öğretilmiş bir algoritma sayesinde sınıflandırılmaktadır. Bu çalışmalardaki yöntemler genel olarak gözetimli ve gözetimsiz öğrenme olarak ikiye ayrılmaktadır (Özgür, 2002: 1).

#### 2.1.1.Gözetimli Öğrenme

Gözetimli öğrenme yöntemlerinde daha önceden belirlenip etiketlenmiş eğitim verileri kullanılmaktadır. Algoritmanın amacı bu eğitim verilerine dayanarak girdiler ve çıktılar arasında eşleşme yapan bir fonksiyon elde etmektir. Gözetimli öğrenme yöntemleri sınıflandırma araştırmalarının ana temelini oluşturur (Topaçan, 2016: 57).

Gözetimli öğrenmenin temel amacı, girdileri ve girdilerden elde edilen sonuçları algoritmaya aktararak, algoritmanın bu verilere göre bir fonksiyon oluşturmasıdır. Fonksiyonun, girdi ( $x$ ) ve çıktısı ( $y$ ) daha önceden belirlenir. Asıl amaç, elde bulunan girdiler kullanılarak çıktının nasıl oluştuğunun bulunmasıdır.

$$y = g(x|\theta) \quad (1)$$

Denklem (1) de  $g$  oluşturulan modeli,  $\theta$  parametreleri,  $y$  ise sonucu ifade etmektedir. Girdi ve çıktı arasındaki bağlantının tümevarım yaklaşımıyla bulunma süreci öğrenme olarak isimlendirilir. Öğrenme evresinde ortaya çıkan model ise gelecekteki sonuçların tahmin edilmesi için kullanılır (Alpaydın, 2010: 9).

Gözetimli öğrenme yöntemlerinde veri kümesi önışlemeden geçirilip kullanılabilir hale getirildikten sonra, eğitim ve test işlemlerinde kullanılmak üzere birden fazla alt kümeye bölünür. Eğitim sürecinde algoritmalar değişik parametrelerle test edilip birbirleri ile kıyaslanarak en iyi sonucu sağlayan nihai sınıflandırıcı belirlenir. Eğitim veri kümesi ile eğitilen algoritma test kümesi ile karşılaştırılarak ihtiyacı ne kadar karşıladığı bulunur. Sonuçların tahminlerin altında kalması durumunda algoritma parametreleri değiştirilerek performansın artırılması hedeflenir. Bu işleme “parametre tuning” denilmektedir (Topaçan, 2016: 59).

Gözetimli öğrenme alanında en önemli etken algoritmaların test veri kümesi ile çalıştırılmasıdır. Test veri kümesi ile çalıştırılan algoritmalar arasında en iyi performansı gösteren seçilmektedir. Bu seçim işleminde en çok kullanılan metot doğru tahmin sayısının toplam sayıya bölünmesi ile elde edilen tahmin doğruluğu metodudur. Sınıflandırıcının gücünü ölçmek için tahmin doğruluğunu hesaplamada farklı metotlar bulunmaktadır. Bu metotlardan birincisi, veri kümesinin bir kısmını eğitim diğer kısmını ise test için ayırmaktır. En fazla kullanılan metot %66 ile sınıflandırıcıyı eğitip kalan %33 ile test etmektir. İkinci metot ise,  $k$  katlamalı çapraz doğrulama ( $k$  fold cross validation) olarak isimlendirilir.  $k$  katlamalı çapraz doğrulama metodun da ilk olarak bir  $k$  değeri seçilir ve veri kümesi  $k$  kadar alt kümeye ayrılır. Metot  $k$  defa çalıştırılarak her adımda veri kümesinin  $1/k$  kadar kısmı test için kullanılır. Geride kalan kısmı ise eğitim için kullanılır. Tüm alt kümeler test için kullanıldıktan sonra her bir adımda elde edilen tahmin doğruluklarının ortalaması alınır (Topaçan, 2016: 59-60).

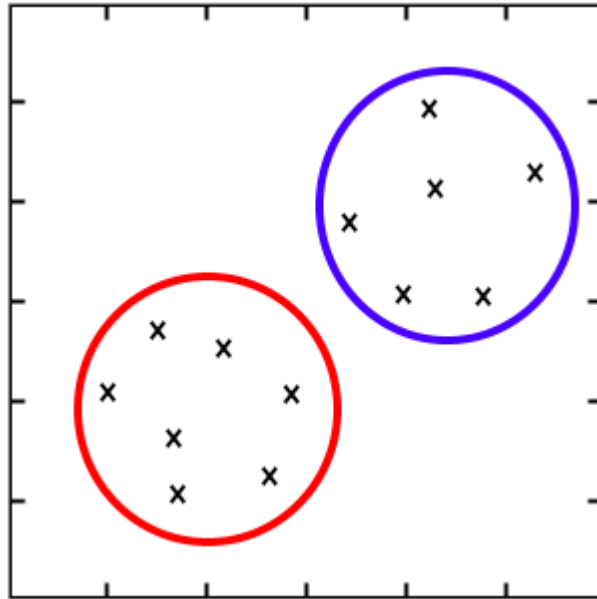
Örnek olarak; k değeri 3 olarak kabul edilirse, veri kümesi 3 parçaya ayrılır. Bu durumda sınıflandırma algoritması 3 defa çalışır ve ilk çalıştırmada 1. ve 2. grup eğitim, 3. grup test verisi olur. İkinci çalıştırmada 1. ve 3. grup eğitim 2. grup test verisi olur. Üçüncü çalıştırmada 2. ve 3. grup eğitim, 1. grup test için kullanarak her bir adım için tahmin doğruluğu hesaplanır ve bunların ortalaması alınır. K katlamalı çapraz doğrulama yönteminde en yaygın tercih edilen k değeri 10'dur.

### 2.1.2. Gözetimsiz Öğrenme

Gözetimsiz öğrenme yöntemi; etiket yapılmamış asıl verideki gözle görülmeyen bağlantıların ortaya çıkarılması işlemleridir. Gözetimli öğrenme yönteminde olduğu gibi verilerin etiketleri ya da sınıfları belli değildir. Veriler benzer özelliklerine ve birbirlerine yakınlık durumuna göre gruplandırılmaktadır.

Gözetimsiz öğrenmenin en yaygın kullanıldığı yöntem öbikleme yöntemidir. Öbikleme yönteminde yakın özneliklere sahip veriler aynı kümelerle gelecek şekilde gruplara ayrılmaktadır. Örnek olarak bir firma müşterilerinin demografik bilgilerini elde tutarak müşterilerin profillerini görmek isteyebilir. Böyle bir durumda, öbikleme yöntemi birbirine benzeyen müşterileri aynı öbeklere aktararak müşteri kümeleri oluşturabilir. Şekil 1'de iki farklı gruptan oluşan öbikleme örneği yer almaktadır.

**Şekil 1: Makine Öğrenimi Gözetimsiz Öğrenme Öbikleme Örneği**



### 2.1.3. Takviyeli Öğrenme

Takviyeli öğrenme, kendi ortamında hareket ve algılama yapan bağımsız bir durumun, hedefe ulaşması için en doğru hareketleri öğrenip uygulayacağını gösteren bir öğrenme yapısıdır. Takviyeli öğrenmede ceza ve ödül durumları vardır. Bir ortamda bulunan etmenin ortamdaki her bir hareketine karşılık ortaya çıkan durum için ceza ya da ödül verilir. Etmenin bu durumdaki amacı en yüksek ödülü ortaya koyan hareketlerin sırasını öğrenmesidir. Örnek olarak bir oyun içindeki etmen için, oyun kazanıldığı zamanlarda iyi bir ödül, kaybettiği zaman kötü bir ödül yani ceza diğer durumlarda ise ne ödül ne de ceza alabilir (Karadoğan, 2014: 13).

Takviyeli öğrenme, sayısal bir durum sinyalinin en yüksek düzeye çıkarmak için ne yapmak, eylemlerle durumları nasıl eşleştirmek gerektiğini gösteren bir öğrenmedir. Öğrenen duruma, makine öğrenmesinin birçok yöntemindeki gibi hangi hareketleri yapması gerektiği söylenmez, bu durum yerine hangi hareketlerin en fazla ödülü kazandıracaklarını, durumu deneyerek bulmasını söyler. Birçok farklı ve değişken durumlarda hareketler sadece anlık ödülleri değil, aynı zamanda bir sonraki durumları ve böylece sonraki tüm ödülleri etkileyebilir (Sutton ve Barto, 2012:1).

Takviyeli öğrenme algoritmaları işletmelerdeki işlemlerin optimisasyonu, robotik uygulamaları, oyun uygulamaları gibi birçok çalışma alanındaki sorunları çözmek için kullanılabilir.

### 2.2. Metin Madenciliği

Metin madenciliği; temel olarak yapısal olmayan metinlerden veri içeren yapısal metinleri geliştirme işlemi olarak adlandırılabilir. Metinlerin işlenerek, anlamlı bilgilerin elde edilmesi için özellik çıkartımı ve ön işleme gibi adımların gerçekleştirilmesi gerekmektedir. Bu uygulamalardan sonra yapısal olmayan veriler, metin madenciliğinin kullanılacağı ve bilgisayarlar tarafından işlenen yapısal bir biçime dönüştürülebilir. Bu sayede büyük veriler içerisinde bulunan önemli bilgilere ulaşılmış olunur. Elde edilen bilgiler kullanılarak kurumların yararlanacağı önemli sonuçlara ulaşılabilir. Metin madenciliğinin temelinde istatistiksel ve matematiksel bulgular bulunmaktadır. Metin madenciliği, duygu analizi, anahtar kelime çıkartma, yazar tanıma, metin sınıflandırma, fikir madenciliği, başlık çıkartımı gibi farklı alanlarda kullanılmaktadır (Kılınç vd., 2016: 90).



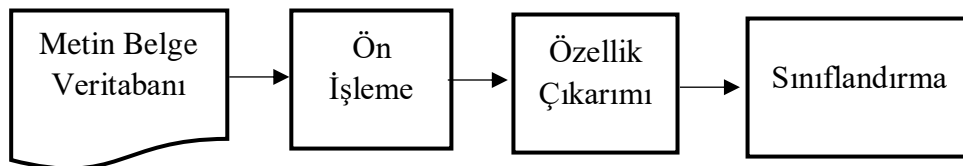
Metin sistemindeki verilerde kelime ve cümleler her zaman doğal sistematik sıralanışında değildir. Bu yüzden metin madenciliğinde kullanılacak istatistiksel uygulamaların doğru sonuçlar verebilmesi için doğal dil işleme yöntemleriyle kelime ve cümlelerin ön işleminin dikkatli bir şekilde yapılması gerekmektedir (Amanet, 2017: 15).

Metin madenciliği uygulama alanları; kendi alt başlıklarında bilgi çıkarımı, konu takibi, özetleme, kategorizasyon, kümeleme, kavram bağlantısı, bilginin görselleştirilmesi ve soru cevaplama şeklinde sekize ayrılmaktadır (Vijayarani vd., 2015).

### 2.2.1. Metin Madenciliği Adımları

Metin madenciliğin adımları ilk olarak metin belgelerinin elde edilip toplanması ile başlar ve daha sonra ön işleme için bir metin madenciliği aracı uygulanır. Ön işleme yöntemi veriyi temizler ve şekillendirir; ayrıca, bu verilerden anlamlı özellikler çıkarmakla görevlidir. Daha sonraki aşamada ise metinlerde bilgi keşfi için kümeleme veya sınıflandırma algoritmalarından uygun olanı seçilip uygulanmaktadır. Metin madenciliği adımları Şekil 2’de yer almaktadır (Periyasamy ve Kumar, 2014: 257).

**Şekil 2: Metin Madenciliği Adımları**



Kaynak: Periyasamy ve Kumar, 2014: 257.

#### 2.2.1.1. Metin Ön İşleme

Metin ön işleme 3 ana unsurdan oluşmaktadır:

- Singelere ayırma; metin verileri bir ifadeler bütünüdür. Singelere ayırma tam metni ara boşluklar, virgüller noktalama işaretleri vb. bırakmak suretiyle kelime parçalarına ayırmaktadır.
- Gereksiz kelime çıkarma: Bu unsurda internet sayfalarından HTML, XML gibi etiketlerin çıkarılmasını ele almaktadır. Daha sonra “bir”, “şey”, “ve”, “bu” gibi gereksiz kelimelerin yok etme işlemi gerçekleştirilebilmektedir.

- Kökleştirme: Bazı kelimelerin kökünü saptama işlemidir. Temelde iki çeşit kökleştirme bulunmaktadır. Birincisi çekimsel ikincisi yapımsaldır. En fazla kullanılan algoritma Porter kökleştirme algoritmasıdır (Sumathy ve Chidambaran, 2013: 30).

### 2.2.1.2. Metin Dönüştürme

Metin belgesi bulundurduğu kelimeler ve bu kelimelerin görünme sıklığı ile temsil edilmektedir. Belge temsili için kullanılan iki tür yaklaşım ele alınmaktadır (Sumathy ve Chidambaran, 2013: 30):

- Kelime havuzu.
- Vektör uzayları.

### 2.2.1.3. Özellik Seçimi

Model yaratımında kullanılmak üzere önemli öznitelikleri seçme işlemine özellik seçimi denir. Bu aşamada ilgisiz ve gereksiz öznitelikler bulunarak kaldırılmaktadır (Sumathy ve Chidambaran, 2013:30)

Özellik seçimi yöntemi, doğru öngörücü modeller oluşturma konusunda destek sağlamaktadır. İyi veya gelişkin ve daha çok kesinlik veren, daha az veri gerektiren özniteliklerin seçimi daha yararlı olmaktadır. Özellik seçimi yöntemleri, öngörücü modellerin doğruluğuna yardımcı olmayan veya modellerin doğruluğunu düşürebilen ilgisiz ve gereksiz özniteliklerin veri içinde saptanmasında ve kaldırılmasında en önemli yöntem olarak kullanılır (Rose, 2002: 29).

#### 2.2.1.3.1. Temel Bileşen Analizi (Principal Component Analysis – PCA)

Temel bileşen analizi (PCA) boyut küçültme teknikleri arasında yer alan eğitimcisz sınıflandırıcılarındandır. PCA kendisine verilen bilgi yığını temsil edecek temel bileşenler olarak adlandırılan yeni ortogonal değişkenler saptar. PCA elde olan veride en temel öznitelikleri çıkararak, verinin temsil edileceği öznitelik sayısını azaltabilmek amacıyla var olan değişken ve gözlem bilgilerinin yapısını incelemektedir. Herhangi bir  $d$  boyutu uzaydan  $k < d$  boyutlu bir uzayda eşleşme elde etmek amacıyla  $x$  vektörüne ait iz düşüm hesaplaması yapılmaktadır (Abdi and Williams, 2010: 433-459).

Normalizasyon işlemi ise; metinlerdeki yazı karakter setlerinin Unicode Transformation Format 8 kodlama standardına dönüştürme işlemidir.

#### 2.2.1.4. Verilerde Kayıp Değer

Araştırmacılar ne kadar çok çalışsalar ve eksiksiz veri elde etme çabasında olsalar da üzerinde çalışılmak istenen veriler bazı durumlarda istenildiği gibi eksiksiz bir şekilde toplanamayabilir. Genellikle büyük veriler üzerinde yürütülen çalışmalarda eksiksiz veri setlerine ulaşılması neredeyse imkânsızdır (Cool, 2002: 4).

Katılımcılardan kaynaklı veri kaybı; kişilerin bilinçli ya da bilinçsiz bir şekilde bazı soruları cevapsız bırakması, belirlenen süre içerisinde bazı durumlara ulaşamaması veya farklı nedenlerle veri toplama sürecinin bazı aşamalarında bulunamaması gibi sebeplerle oluşabilmektedir. Katılımcılardan kaynaklı olmayan veri kayıpları ise araştırmada kullanılan veri toplama araçlarının teknik özelliklerinin yetersizliği, kullanılan aracın uygulama koşullarının elverişli olmaması, araştırmacının veri girişi durumunda dikkatsizliği gibi sebeplerden dolayı ortaya çıkmaktadır. Veri setlerindeki bu eksiklikler kayıp veriler olarak tanımlanmaktadır. Kayıp veriler analizler için kullanılacak olan istatistiksel yöntemlerin hepsi için önemli bir sorun ortaya çıkarmakta olup bundan dolayı tüm yöntemler veri setinin eksiksiz olduğunu varsayımı altında gerçekleştirilmektedir (Osborne, 2013: 2).

Kayıp veriler dikkate alınmadan yapılacak olan analizler sonucunda yanıltıcı sonuçlar ortaya çıkabilmektedir. Ayrıca bu sonuçların araştırma evrenine genelleme noktasında hata miktarının artacağı da söylenebilir. Bundan dolayı, toplanan veriler üzerinden yürütülecek olan ileri düzey analizlere geçilmeden önce kayıp verilerin miktarı ve yapısıyla ilgili incelemelerin yapılması ve karşılaşılabilecek sıkıntıların ortadan kaldırılması için gerekli önlemlerin alınması araştırmacılar için bir zorunluluk haline gelmektedir (Byrne, 2000: 48).

Kayıp veriler değişkenliklerine göre üç farklı yapısal özellik göstermektedir. Bunlar tamamıyla rastlantısal olan kayıp veriler, rastlantısal olan kayıp veriler ve ihmal edilemez kayıp verilerdir.

Tamamıyla rastlantısal olan kayıp veriler, değişkenin düşük veya yüksek değerler almasıyla alakalı olmayan ve başka bir değişkenin etkisiyle ortaya çıkarılmayan, veri seti içerisine tamamen rastlantısal olarak dağılmış verilerdir. Rastlantısal olan kayıp veriler, kayıp verilerin başka değişkenlerle alakalı olduğunu, fakat değişkenin kendisindeki değişimle alakalı olmadığı anlamına gelmektedir. İhmal

edilemez kayıp veriler, değişkenlerdeki veri kaybı olasılığı hem diğer değişkenlerle alakalı hem de değişkenin kendi değerleriyle alakalıdır. İhmal edilemez kayıp verilerle başa çıkmak diğer kayıp verilere göre daha çok uzmanlık gerektiren bir çalışmadır (Çüm vd., 2018:231-232).

#### **2.2.1.4.1.Beklenti Maksimizasyonu (Expectation Maximization-EM) Algoritması Yöntemi**

Beklenti maksimizasyonu iki aşamadan oluşan, interaktif şekilde durmadan tekrar eden beklenti ve maksimizasyon durumlarından atama işlemleri yapan bir yöntemdir (Sayın, vd., 2017: 493).

İlk aşama olarak beklenti durumunda, gözlenen durumlar kullanılarak bir dizi regresyon denklemi oluşturulur. Elde edilen kayıp verilere başlangıç durumunda ilk değerleri eklenir. Maksimizasyon aşamasında, atanan değerler kullanılıp yeni durumdaki değerlere atama işlemi yapılarak ikinci aşama işlemi yapılmaktadır. Bu durumlar birçok defa tekrar edilmektedir. Beklenti ve maksimizasyon aşamalarından her adımda yeni değer durumları hesaplanmaktadır. Ardışık maksimizasyon aşamasında basamaklar arasında değerler değişme gösterdiğinde maksimum olasılık tahmini gerçekleşmektedir. Beklenti maksimizasyonu algoritmasının dezavantajı ise elde edilen standart hataların gerçek standart hatalarla tutarlı olmaması durumudur (İşikoğlu, 2017: 13-14).

Kovaryans ve ortalama matrisinin tahmin durumları beklenti maksimizasyonu algoritmasında üç adımdan oluşmaktadır. İlk adımda, ortalama ve kovaryans matrisinin tahminleri kullanılarak eksik değerleri içeren verilerin regresyon parametreleri hesaplanmaktadır. İkinci aşamada ise eksik değerler, hesaplanan parametreler kullanılarak tamamlanmaktadır. Tamamlama işleminden sonra ortalama ve kovaryans matrisi tamamlanmış veri seti kullanılarak baştan tahmin gerçekleştirilmektedir. Bu aşama yakınsama sağlayıncaya kadar devam etmekte ve sonucunda eksik veriler için en iyi tahmini değerlere ulaşılmaktadır (Şahin, 2012: 19-20).

### **2.2.1.5. Veri Madenciliği Yöntemlerinin Kullanımı**

Kümeleme, sınıflandırma ve bilgiye ulaşma gibi veri madenciliği yöntemleri metin madenciliği için de kullanılmaktadır. Bu durumda metin madenciliği veri madenciliği olmaktadır (Sumathy ve Chidambaram, 2013:30).

Eldeki metin madenciliği uygulanacak veri ön işlemlerden geçirilip analiz edilebilecek forma sokulunca, verinin yapısına göre bir ya da birkaç veri madenciliği yöntemi seçilerek analiz başlatılır. Analiz için hazır paket programlar olduğu gibi bir takım hazır fonksiyonların kullanılıp kodun analizci tarafından yazılabileceği platformlarda mevcuttur. Analizi yapacak kişi bunlardan birini seçerek veri içindeki bilgiyi veri madenciliği algoritmaları ile elde etmeye çalışır.

### **2.2.1.6. Yorumlama Değerlendirme**

Metin madenciliğinde araştırmaların sonucunda uygulanabilirliği, doğruluğu, uygunluğu ve yeniliğinin ölçümü için göstergelerin kullanımı kimi zaman seçime bağlıdır. Fakat iş dünyasının metin madenciliği uygulamasından elde edilebilecek ilgililik ve bilgi etkisi ölçülerine olanak sağlaması gerekebilir. Ayrıca karar alan kişilerin müdahaleleri ve tecrübeleri de belirli ölçüler oluşturmak ve göstergeler üzerinde hesaplamalar yapmakla ilişkilidir. Metin madenciliği ile birlikte işletmeler; ekonomik, finans, planlama, pazar analizi ve müşteri profili analizi gibi konularda incelenen veri sonucunda elde edilen bilgiyi çeşitli araçlar ile görselleştirerek yorum ve değerlendirme yapabilmektedir (Sebastiani, 2002: 23).

### **2.2.2. Bilgi Çıkarımı (Information Extraction)**

Bilgi çıkarımı metodu metinlerin içindeki ana kelimeleri ve bu kelimelerin birbirleri ile olan ilişki durumlarını tanımlar. Bu durumlara desen eşleme adı verilir. Bu işlemi metinlerin içindeki daha önceden tanımlanan diziler arasında yapar (Vijayarani vd., 2015).

Standart veri madenciliği yöntemlerinde bilginin ilişkisel veri tabanı üzerinden çıkarıldığı ön görülmektedir. Ancak çok fazla uygulamada veri doğal dil formundadır ve bilgi çıkarımı bu formlar üzerinden yapılmaktadır. Bilgi çıkarımının veri tabanları üzerindeki ilişkisel metinler üzerinden yapılması işlemi, bilgi alımı olarak adlandırılmaktadır (Grupta ve Gurpreet, 2009).

Bilgi çıkarımı algoritması; verileri işlemek amacıyla metinleri bir ilişkiel veri tabanına dönüştürür. Bu durumda ayıklanan verilerden elde edilen kurallar durumunda görülmeyen diğer bilgiler çıkarılabilmektedir.

### 2.2.3. Özetleme

Metin verileri özetleme yöntemi, kullanıcıların bir metni okuduklarında elde edecekleri bilgiyi hemen kullanıcıların anlamasını hedefler. Büyük metinler özelinde metin özetleme programları, kullanıcının sadece ilk paragrafı okuyacakları süre içerisinde metnin direk özetlenmiş durumu olarak kullanıcılara verilir. Burada ki asıl konu, metnin içerisindeki bütün anahtar kelimelerin bulunabilmesidir. Anahtar kelimeler bulunurken karşılaşılan en kötü sorun ise; bir metin içerisindeki söz bilimsel özelliklere dayanarak önemli kişi, yer, konu, tespitini yapma konusunda insan beyni bile zorlanırken, bunu bir programa yaptırmaya çalışılmasıdır. Bulanık mantık çerçevesinde cümlelerin özet içerisinde olup olmayacaklarına karar verilirken uygulanacak adımlar şunlardır (Tekin, 2017:47):

- Metin girdisi alınır.
- Metin yapı analizi çıkarılır.
- Kelimeler segmente edilirken her kelimenin metin içinde kaç defa geçtiği verisi çıkarılır.
- Anahtar kelime çıkarımı yapılır.
- Kelime ve cümleler ağırlıklandırılır.
- Cümle seçimi gerçekleştirilir.
- Kaba özet oluşturma işlemi gerçekleştirilir.
- Yumuşatma işlemi gerçekleştirilir.
- Özet çıktısı alınır.

### 2.2.4. Kategorizasyon

Kategorizasyon yönteminde bilgi çıkarımında olduğu gibi anlatılmak istenen anlamaya çalışılmaz. Metin içerisindeki kelimelerin kaç defa geçtiğine bakılarak metnin ana konusu anlaşılmaya çalışılır. Genellikle kategorizasyon yönteminde önceden tanımlanmış gruplar ve bu gruplar için belirlenmiş kurallardan oluşan sözlük

bulunmaktadır. Özetleme yöntemi ile birlikte kategorizasyon, konu belirleme amacı ile kullanılmaktadır. Bu durumda bir kullanıcının hangi konular ile daha fazla ilgilendiği anlaşılmasına çalışılabilir. Kategorizasyon yöntemi, müşteri memnuniyetini hedefleyen birçok firma tarafından kullanılmaktadır. Bu durumda müşterilerin sorduğu sorular daha çabuk anlaşılıp, ihtiyaçları olan cevaplar daha efektif bir şekilde cevaplanabilmektedir. Kategorizasyon yöntemi ile bir metin daha önceden belirlenmiş birden fazla gruba ilişkilendirilebilir. İncelenen öğrenme algoritmaları sayesinde, daha önceden kategorileri belirlenmiş içerikler kullanılarak, kategorisi hala belli olmayan içeriklerin kategorisi tespit edilebilmektedir (Tekin, 2017:48-49).

### 2.2.5. Kümeleme

Metin analizinde kümeleme yöntemi, birbirine benzer metinlerin gruplanma işleminde kullanılmaktadır. Kategorizasyondan farkı ise daha önceden belirli olan bir gruba atama yöntemi yerine, canlı olarak benzer metinlerin gruplanmasına yardımcı olmasıdır. Kümeleme yöntemi sayesinde, birden fazla konu ile alakalı bir metin arama işlemi gerçekleştirilebilmektedir. Standart bir kümeleme algoritmasında her bir veri için bir vektör oluşturulur ve eldeki mevcut gruplara uygunluk durumuna bakılır. Birçok metinden oluşan bir veri tabanına sahip firma bu teknolojiye en uygun şekilde yararlanabilmektedir. Kümeleme yöntemi için önerilen birçok algoritma yer almaktadır (Tekin, 2017: 49-50).

Kümeleme yönteminde aşağıdaki adımlar uygulanır:

- Gereksiz kelimelerin kaldırılması.
- Kelime köklerine inilmesi.
- Filtreleme.

### 2.2.6. Kavram Bağlantılama

Kavram bağlantılama, kullanıcıların eski yöntemlerle yaptıkları aramalar sonucunda bulamadıkları fakat kavramsal olarak veriler arası bağlantı sayesinde ilişkiyi ortaya koymayı hedefler. Bu yöntem genellikle biyomedikal alanında, verilerin hepsinin okunmasının zor olduğu alanlarda kavramsal bağlantıyı ortaya çıkarmak için kullanılır (Weiguo vd., 2005).

Örneğin; A ve B birbirleri ile bağlantılı iki metinken, B ve C de birbirleri ile bağlantılı olduklarından dolayı direk olarak A ve C birbirleri ile bağlantılı olmaktadır.

Fakat bu bağlantının insan beyni ile anlaşılamayacağı zaman kavram bağlantılama yazılımları işe girerek sorunları çözmektedir.

### 2.2.7. Bilgi Görselleştirme

Bilgi görselleştirme veya görsel metin incelemesi, büyük metinsel verileri görsel bir hiyerarşiye göre sıralamak veya bir haritaya yerleştirmektedir (Weiguo vd.,2005). Bir kullanıcının büyük bir metin aralığını küçültmesi veya ilgili konuları keşfetmesi gerektiğinde, bilgi görselleştirmeden faydalanılabilir. Gözle tespit edilemeyecek durumlar arası bağlantılar bilgi görselleştirme ile çözülebilmektedir.

Bilgi görselleştirme adımları şu şekilde sıralanmaktadır:

- Görselleştirilecek verilerin toplanması.
- Veri indeksleme.
- Bilgi çıkarımı.
- Görselleştirilecek verilerin oluşturulması.
- Verilerin ara yüz üzerinde gösterilmesi.

### 2.2.8. Soru Cevaplama

Metin madenciliğinin diğer uygulama alanlarından biri ise kullanıcı tarafından sorulan soruya en doğru cevabı bulup, kullanıcıya bu cevabın en doğru şekilde ulaştırılmasıdır. Bu durumda, birçok internet sitesinde “Bilgisayara sor” şeklinde bir ara yüz yazılım kurulmuştur. Bu teknoloji arka planda çoğunlukla bilgi çıkarımı teknolojisiyle birlikte çalışır (Gupta ve Gurpreet, 2009). Kullanıcılar tarafından sorulan soru üzerindeki konum, isim, kişisel bilgiler tespit edilip sorular bilgilere göre cevaplandırılır. Bu durumun en yaygın kullanımı “Sıkça sorulan sorular” ara yüzü olarak günümüzde internet sitelerinde mevcuttur.

### 2.3. Kelime Torbası

Kelime torbası modeli, verileri içlerinde geçen kelimeler ve bu kelimelerin sıklıklarını kullanarak oluşturulan vektörler şeklinde göstermeye yarayan bir modeldir. Veri öbekleme konusunda birçok çalışmada kullanılmıştır. Bu modelde kelimelerin sıklıkları genelde TF-IDF (term frequency – inverse document frequency) isminde bir ağırlıklandırma yöntemi ile hesaplanmaktadır. TF-IDF kelimeleri buldukları verilerin içindeki sıklığı (term frequency) ve bütün verilerin içerisindeki sıklığı (inverse



document frequency) değerlendirilerek Denklem (2) uygulanır (Salton ve Buckley, 1988:513-523).

$$tfidf_{ij} = \frac{f_{ij}}{\sum_{j=1} f_{ij}} \times \log\left(\frac{|D|}{|\{d_i | t_j \in d_i \in D\}|}\right) \quad (2)$$

Denklem (2)'de  $d_i$  dokümanın içindeki  $t_j$  kelimesinin sıklığı  $f_{ij}$  ile gösterilmektedir.  $d_i$  içerisindeki bütün kelimelerin toplam sıklıklarına bölünür. Veri seti  $D$ 'nin içerisindeki toplam veri sayısı  $|D|$  ve  $t_j$  kelimesini içeren veri sayısı  $|\{d_i | t_j \in d_i \in D\}|$  şeklindedir.

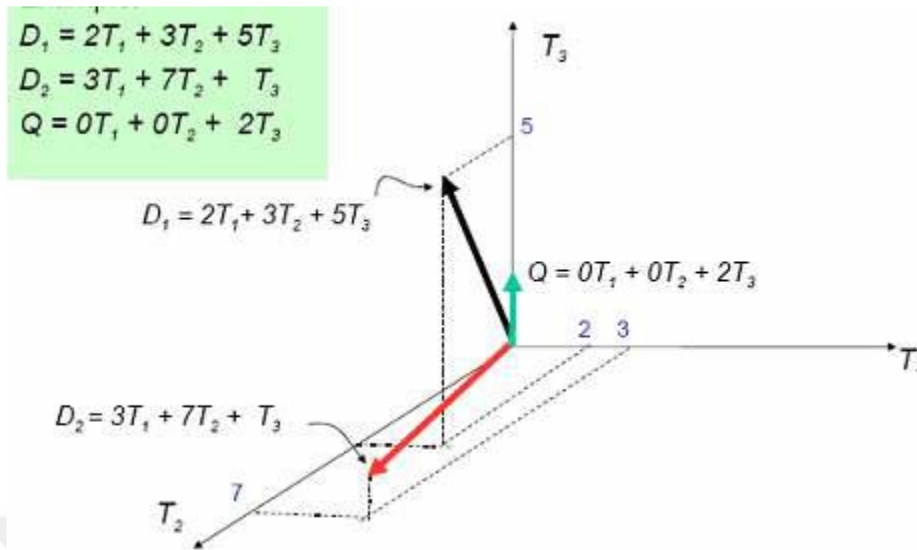
Kelime torbası yaklaşımları en çok bilinen yöntemlerden biri olmasına rağmen doküman öbeklemede eksik kalmaktadır. Bu yaklaşımda iki anlamlılık (ambiguity) ve eş anlamlılık (synonymy) sorunları vardır. Birden fazla anlama gelen bazı kelimeler aynı kelime olarak gözükmekteyken, aynı anlama gelen iki kelime ise yazılışları farklı olduğu için benzerlik hesabı olumsuz etkilemektedir. Bundan farklı olarak bir diğer problem ise kelimelerin verilerin içerisinde buldukları yerlerin benzerlik hesabına katılmamasıdır (Yücesan, 2016:14).

#### 2.4.Vektör Uzay Modeli

Metin şeklinde depolanan verilerin, sınıflandırılması veya üzerinde hesaplamaya dayanan işlemler yapılabilmesi için, bilgisayar ortamında incelenmesini sağlayacak farklı gösterim şekillerine dönüştürülmesi gerekmektedir (Adsız, 2006:25). Doğal dil kullanılarak yazılmış verilerin yapısal olmayan formdan sayısallaştırılarak yapısal forma dönüştürülmesi için kullanılan en yaygın yöntemlerden birisi de vektör uzay modelidir (İlhan vd., 2008:357).

Vektör uzay modeli bilgi çıkarımı, indeksleme, bilgi filtreleme gibi alanlarda kullanılan bir modeldir. Doğal dil verilerinin çok boyutlu uzayda özel bir anlamını ifade etmektedir (Pilavcılar, 2007:4).

Şekil 3: Vektör Uzay Modeli



Kaynak: Pilavcılar, 2007: 4.

Dokümanlar Şekil 3'te görüldüğü gibi kelimelerin vektörleri olarak ifade edilmektedir. Şekil 3'teki T'ler kelimeleri ifade etmektedir (Jun ve Houkuan, 2002'den akt. Pilavcılar, 2007:4)

Anahtar kelime araması yapılan verilerin ilişki durumları, veri benzerlik teorisindeki varsayımlar kullanılarak, yani her bir veri vektörü ile orijinal sorgu vektörü arasındaki açıların sapmalarını karşılaştırarak hesaplanır. Vektörler arasındaki gerçek açıların hesaplanması yerine, vektörler arasındaki açının kosinüsü hesaplanarak karşılaştırılmaktadır (Pilavcılar, 2007:4).

Vektör uzay modelinin, YxZ gibi büyük boyutlarda matrisler üzerinde çalışıldığında, vektörler arasındaki yakınlığın hesaplanması oldukça zorlaşır. Aynı zamanda matrislerin boyutunun artması daha fazla işlem gücü ve hafıza gerektirmektedir (Güven, 2007: 8).

## 2.5.Sınıflandırma Algoritmaları

Sınıflandırma veri madenciliği alanında en çok kullanılan yöntemlerden biridir. Elde edilen girdilerin özniteliklerine göre bir sınıflayıcı yardımıyla sınıflara atanması sürecine sınıflandırma denir. Elde edilen nesnelerin hangi sınıfa atanacağını ya da o sınıfa atanıp atanmayacağını belirleme işlemidir. Diğer bir ifade ile nesneler ya da durumlar için uygun sınıf tahmin edilmesidir. Sınıflandırma girdileri, her bir verinin bir

sınıf etiketi ile etiketlenerek gözlem ya da örneklerden oluşan bir eğitim kümesidir. Çıktı ise modelin her bir gözleme, niteliklerine bakılarak aktarıldığı sınıf etiketidir (Emel ve Taşkın, 2005: 224).

Bazı araştırmalara göre sınıflandırma tam sınıflandırma ve kısmi sınıflandırma olarak ikiye ayrılmaktadır. Tam sınıflandırma; veri içindeki tüm sınıflar ve örnekleri ele alan modellerle ilgilidir. Tam sınıflandırma yöntemleri olarak yapay sinir ağları, C&RT, CHAID, C4.5, C5.0 ve diğer karar ağaçları gibi akıllı teknikler ile diskriminant analizi gibi istatistiksel araçlar örnek verilebilir. Kısmi sınıflandırma ise tam sınıflandırmadaki gibi veri sınıflarının özellikleri gösterilmektedir. Fakat kısmi sınıflandırma modellerinde tüm sınıflar veya verilen sınıfın tüm örnekleri kapsam dâhilinde değildir. Kısmi sınıflandırma örneği olarak birliktelik kuralları verilebilir (Bloemer vd., 2003: 119-120).

### 2.5.1. Karar Ağaçları

Karar ağaçları algoritması, örneklemi sınıflandırmak için ağaç yapısı şeklini kullanmaktadır. Bir tane kök düğüm belirlenir ve buradan aşağı yaprak düğümlere doğru sıralanarak devam etmektedir. Yaprak düğümler nitelik durumlarını test ederek örnekleri ayırmaktadır. Yaprak düğümlerden elde edilen dallar da düğümlerden elde edilen test sonucuna göre ayırım işleminden sonra örnekleri temsil etmektedir.

Karar ağacı yönteminin temel algoritması ID3'tür. Algoritmaya başlamadan önce "Ağacın kökü hangi nitelik olmalıdır?" sorusu yöneltildikten sonra başlamaktadır. Asıl amaç ağacı kısa tutmak ve en hızlı şekilde sonuca ulaşmak olduğu için düğüm seçimleri oldukça önemlidir. Bu seçimi belirlemek için çeşitli istatistiksel yöntemler kullanılmaktadır. Bu yöntemlerden biriside bilgi kazancıdır. Bilgi kazancı yöntemini hesaplayabilmek için ilk olarak Entropy hesaplanmaktadır. Entropy, niteliklerin homojenliğini, öylesine yapılacak olan nitelik seçiminin kararlılığını ölçmektedir. Entropi formülü Denklem (4)'te verilmiştir (Mitchell, 1997).

$$Entropy(S) \equiv \sum_{i=1}^c -p_i \log_2 p_i \quad (4)$$

Bilgi kazancı Denklem (5) sayesinde hesaplanmaktadır. En yüksek kazanç Gain denir. En yüksek kazançta sahip nitelik, en iyi durum niteliği olarak seçilmektedir.

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (5)$$

ID3 algoritması ile ağaç algoritması oluşturulurken her düğüm seçiminde Denklem (5) kullanılmaktadır.

### 2.5.2.Yapay Sinir Ağları

Yapay sinir ağları algoritması; beynin çalışma durumundan esinlenerek modellenmiş bir öğrenme algoritmasıdır. Yapay sinir ağları algoritmasını anlayabilmek için ilk olarak biyolojik olarak düşünüp dendrit ve sinaps'a bakmak gerekmektedir. Dendritin görevi diğer sinir hücrelerinden ulaştırılan sinyalleri, sinir hücresinin çekirdeğine iletmektir. Sinaps'ın görevi de aksondan gelen toplam bilgiyi ön işlemde süzdükten sonra diğer sinir hücrelerinin dendritlerine ulaştırmaktır. Asıl önemli olan toplam bilginin direk olarak iletilmeyip bir işlemde geçirilmesidir. Bundan dolayı da öğrenme işlemi sinapslarda gerçekleşir. Yapay sinir ağı algoritması modelleri üzerinde "öğrenme" sinapslar ve dendritler arasında bulunan ağırlık katsayılarının güncellenmesi olarak algılanmaktadır (Haykin, 2008).

Yapay sinir ağları algoritmaları modeli ilk olarak giriş katmanı, gizli katman ve çıkış katmanı olarak üç katmandan oluşmaktadır. Aynı zamanda çoklu katman yapıları da bulunmaktadır. Giriş katmandan sinapslarla gelen ağırlıklı bilgiler gizli katmana, gizli katmandan da çıkış katmanına iletilmektedir (Kozan, 2017: 33).

$$x = w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k = \sum_{j=0}^k w_j a_j \quad (6)$$

Denklem (6)'da da görüldüğü gibi giriş katmanda bulunan özellikler ağırlıkları ile çarpılarak toplanmaktadır. Toplanan bu ağırlıklar aktivasyon fonksiyonundan geçirilerek çıkış katmanına sonuçlar ulaştırılsın mı ulaştırılmasın mı kararı verilmektedir. Öğrenme durumu için çıkış katmanındaki değerin doğru ya da yanlış olmasına göre ağırlık değiştirilerek işlemler tekrar edilmektedir (Kozan, 2017: 34).

Uygunabilir Moment Tahmini veya Adam optimizasyonu algoritması: Çok çeşitli derin öğrenme mimarileri için iyi çalışan algoritmalarından bir tanesidir. Adam optimizasyon algoritması, gradyan iniş momentum ve RMSprop algoritmalarının bir kombinasyonudur (Kingma and Ba, 2015).

Levenberg-Marquardt (LM) Algoritması: Çok katmanlı ağlarda küçük veya orta büyüklükteki desenlerde kullanılan en kullanışlı algoritmadır. Büyük verilerde ise aşırı bellek kullanımı ve çarpma işlemi ortaya çıkmakta, bu da verimliliğini

azaltmaktadır. Steepest descent ve Newton algoritmalarından türetilen bu algoritma geriye yayılım algoritmasından farklı olarak parametre güncelleme işlemleri için tüm giriş örnek değerleri için oluşturduğu hata vektörünü ve jacobian matrisini kullanarak yapmaktadır (Wılamowski and Yu, 2010: 930-937).

**Bayesian Regulation (BR) Algoritması:** Bayesian regülasyonu Levenberg-Marquardt optimizasyonuna göre ağırlık ve bias değerlerini güncellemektedir. Karasel hata ve ağırlıkların kombinasyonunu minimize eder ve ağı üretmek için doğru kombinasyonu belirlemektedir. Daha iyi güncelleştirme yeteneğine sahip bir ağı oluşturmak için Mackay regülasyon ile ağı parametrelerinin boyutunu sınırlayan yöntem tercih edilmelidir. Regülasyon teknikleri ağırlık değerlerinin daha küçük değerlerde kalması için ağı zorlamaktadır. Bu ağı cevabının daha yumuşak olmasına, ağı ezberleme olasılığının azalmasına ve gürültüyü yakalamasına neden olur (Mackay, 1992: 415-447).

**Scaled Conjugate Gradient (SCG) Algoritması:** Standart geri yayılım algoritması, büyük ölçekli problemlerde çok kötü sonuç vermektedir ve başarısı kullanıcı bağımlı parametreler, öğrenme hızı ve momentum sabitine bağlı olarak farklılık göstermektedir. Bunlardan bir kısmını ortadan kaldırmak için bazı danışmanlı öğrenme algoritmaları geliştirilmiştir. Moller tarafından geliştirilen Scaled Conjugate Gradient algoritması bunlardan bir tanesidir. Ölçekli eşlenik eğitim algoritması çoğu problemler üzerinde süper doğrusal yaklaşmayı gösteren Conjugate Gradient algoritması sınıfına girmektedir (Deperlioğlu, 2018: 95).

### **2.5.3. Destek Vektör Makinesi**

Destek vektör makinesi, Vladimir Vapnik ve Alexey Chervonenkis tarafından 1960'lı yıllarda geliştirilmiştir. Destek vektör makinesinin örüntü tanıma sınıflandırma problemlerinin çözümü için 1990'lı yıllarda başarılı çalışmalar sonucunda kullanılabileceğini kanıtlamışlardır (Joachims, 1998: 142).

Destek vektör makinesi, yapısal minimizasyon prensibine göre çok fazla etkili ve basit bir makine öğrenmesi algoritmasıdır. Sınıflandırma için bir uzay düzleminde bulunan iki grup arasında bir sınır çizgisi çizer. Uzay düzleminde bulunan noktalar her bir üyeyi göstermektedir. Gruplar arasına çizilecek sınırın yeri ise iki grubun üyelerine de en uzak yer olarak belirlenmelidir. Destek vektör makinesi farklı sınıftan örnekler

arasındaki uzaklığı en çok büyüten hiper düzlemi bulmak için kullanılmaktadır. Bu işlem yapılırken ilk olarak her iki sınıfa da yakın birbirine paralel iki sınır çizgisi belirlenir. Bundan sonra bu çizgiler birbirlerine yaklaştırılarak sınır çizgisi oluşturulur. Doğrusal olarak ayrılan sınıflar destek vektör makinesi için en basit durumları göstermektedir. İki sınıflı bir sorunda ayırım yapan hiperdüzleme ait eşitlik Denklem (7)'de verilmiştir (Topaçan, 2016: 101):

$$\begin{aligned} w \cdot x_i + b &\geq +1 \text{ her } y = +1 \text{ için} \\ w \cdot x_i + b &\leq -1 \text{ her } y = -1 \text{ için} \end{aligned} \quad (7)$$

Denklem (7)'de,  $x \in \mathbb{R}^N$  olup  $N$  boyutlu bir uzayı,  $y \in \{-1, +1\}$  sınıf etiketlerini,  $w$  hiperdüzleme ait ağırlık vektörünü ve  $b$  hiperdüzlemin orjinden olan uzaklığını göstermektedir.

#### 2.5.4. Naive Bayes

Naive Bayes sınıflandırıcısı, Thomas Bayes tarafından geliştirilen Bayes kuralına dayanmaktadır. Naive Bayes algoritması, bazı veri setlerindeki değerlerin frekanslarını ve değer birleşimlerini sayarak olasılık setini hesaplayan basit bir olasılık sınıflandırıcısıdır. Algoritmada, Naive Bayes teoremi kullanılır ve sınıf özniteliklerinin değeri göz önüne alındığında tüm özniteliklerin bağımsız olduğunu göstermektedir. Bu koşullu bağımsızlık varsayımı bazen gerçek dünyadaki uygulamalarda geçerlidir. Bundan dolayı niteleme Naive olarak geçmektedir. Ancak algoritma çeşitli denetlenen sınıflandırma problemlerinde iyi performans göstermektedir ve hızlı bir öğrenme sistemi gerçekleştirilmektedir (Tina ve Sherekar, 2013).

Naive Bayes sınıflandırıcısı, Bayes teoremine ve toplam olasılık teoremine dayanmaktadır. Naive Bayes'in kuralı en kolay haliyle Denklem (8)'deki gibidir:

$$\text{Sonsal Olasılık} = \frac{\text{Koşullu Olasılık} \cdot \text{Önsel Olasılık}}{\text{Evidence}} \quad (8)$$

Denklem (8)'de hesaplanmak istenilen sonsal olasılık; belirli bir nesnenin gözlenen özniteliklerinin “verilen bir sınıfa ait olma olasılığı nedir?” şeklindeki soruların istatistiksel cevabıdır. Metin sınıflandırma çalışmalarında verilen bir verinin hangi sınıfa ait olduğunu bulmak için kullanılmaktadır. Sonsal olasılık;  $P(c|d)$  şeklinde ifade edilmektedir, hesaplama formülü aşağıdaki gibidir.

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \quad (9)$$

Denklem (9)'da;  $d$  verilen dokümanın öznelik vektörüdür,  $c$  verilen dokümanın ait olabileceği sınıfı belirtmektedir. Sonsal olasılık hesaplamasına göre, sınıflandırmanın sonunda aşağıdaki örnek algoritma ile sonuç bulunmaktadır.

Eğer  $P(\text{pozitif}|d) > P(\text{negatif}|d)$

Doküman pozitif değilse

Doküman negatiftir.

Koşullu olasılık:  $P(d/c)$  veriyi oluşturan her bir özelliğin çarpımına eşittir.  $N_{w_i c_j}$ :  $w_i$  kelimesinin  $c_j$  sınıfında bulunan sayısını vermektedir.  $N_{w_i}$ :  $c_j$  sınıfındaki tüm kelimelerin sayısını vermektedir.

$$P(d_i|c_j) = P(w_1|c_j) \cdot P(w_2|c_j) \dots P(w_d|c_j) = \prod_{k=1}^d P(w_k|c_j)$$

$$P(w_i|c_j) = \frac{N_{w_i c_j}}{N_{w_i}} \quad (10)$$

Önsel olasılık:  $P(c)$  genel olarak belirli bir sınıfla karşılaşma durumudur. Metin sınıflandırma çalışmalarında  $P(\text{pozitif})$  verilen bir verinin pozitif olma olasılığı durumudur.  $N_{c_j}$ :  $c_j$  sınıfına ait örnek sayısını vermektedir.  $N_c$ : toplam örnek sayısını vermektedir.

$$P(c_j) = \frac{N_{c_j}}{N_c} \quad (11)$$

Evidence: Verilen bir veri setinin pozitif sınıfı için hesaplanan sonsal olasılığı negatif sınıf için hesaplanandan büyükse bu veri seti pozitif sınıfa aittir. Bu durum Denklem (12)'deki gibi formüle edilmektedir.

$$\frac{P(d_i|c_{\text{pozitif}})P(c_{\text{pozitif}})}{P(d_i)} > \frac{P(d_i|c_{\text{negatif}})P(c_{\text{negatif}})}{P(d_i)} \quad (12)$$

Denklem (12)'de, eşitsizliğin her iki yerinde de paydalar eşit olduğu için  $P(d_i)$  durumu formülden çıkarılabilir. Bu durumda sınıflandırma formülü Denklem (13)'teki gibi olur.

$$P(d_i|c_{\text{pozitif}})P(c_{\text{pozitif}}) > P(d_i|c_{\text{negatif}})P(c_{\text{negatif}}) \quad (13)$$

## 2.6.Derin Öğrenme

Derin öğrenme, yapay sinir ağlarının yapısal olarak daha karmaşık hali olarak düşünülebilir. Yapay sinir ağlarının birden fazla gizli katman içermesi ile oluşturulur. Yapay sinir ağları algoritmaları insandaki öğrenme durumundan hareketle geliştirilmiştir. Biyolojik sinir sisteminde bulunan nöronların birbirleri ile ilişki kurması gibi yapay sinir ağları sistemlerinde de nöron şeklinde tanımlanan yapılar birbirleri ile bağlantılı olacak durumda modellenmişlerdir. Bu şekilde öğrenme, hafızaya alma ve veriler arasındaki ilişkiyi ortaya koyma kapasitesine sahip olacağını düşünmüşlerdir. Derin öğrenmenin makine öğrenmesindeki var olan algoritmalarından ayrılan yönü, birçok katmandan oluşmasıdır. Bu katman sayısı yapısal olarak değiştirilebilir ve farklı yöntemler ile karma modeller oluşturulabilir (Polattimur, 2019: 25).

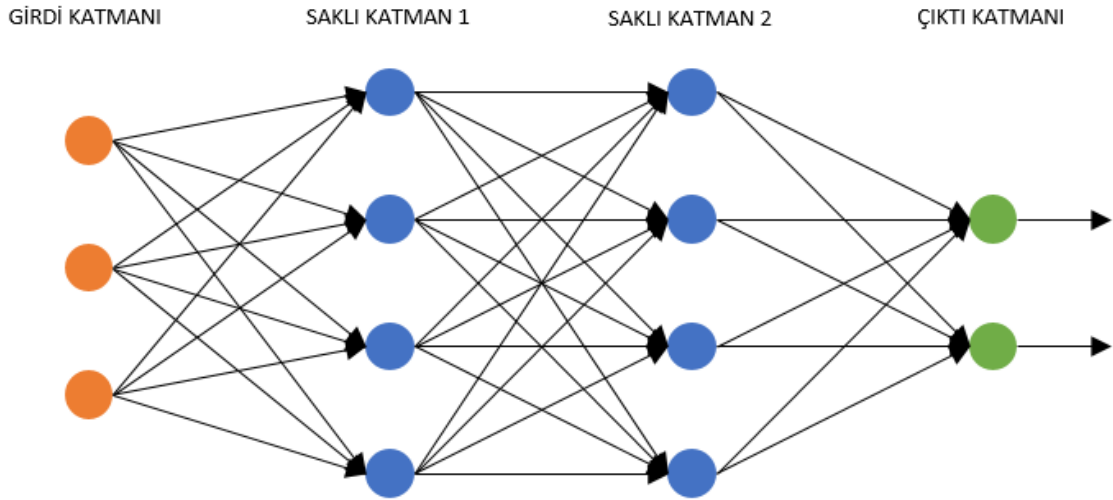
Derin öğrenme yöntemleri temel makine öğrenme yöntemlerinden farklı olarak daha çok nöron içerirler. Yapay sinir ağlarına göre hiyerarşik biçimde daha zor bir yapıdan oluşurlar ve eğitim sırasında daha çok güç tüketirler (Turgut, 2018: 58).

Derin öğrenme, konuşma veya nesne tanıma, video analizi yapma, tahminde bulunabilme gibi insanların işlerinin makinenin yapabilmesini sağlayan makine öğrenme yöntemidir. Makine öğrenme tekniklerinin hepsi kullanılabilir. Son yıllarda yapılan derin öğrenme yöntemi çalışmalarında başarı oranının yüksek, hata oranının az olması bu alana geçişi hızlandırmıştır. Yapılan çalışmalarda makinelerin insandan daha az hata oranıyla nesne tanıma yaptığı görülmektedir (Kavuncu, 2018: 58).

Derin öğrenme, öznitelikleri direk verilen veriler üzerinden öğrenmektedir. Sinir ağları ise diğer algoritmaların kolayca keşfedemeyeceği öznitelikleri ve aralarındaki ilişkileri kolayca öğrenebilir. Yapay sinir ağları matematiksel olarak veriyi dönüştüren nöron katmanlarını kullanarak karmaşık bir model öğrenirler. Normal bir sinir ağı modeli, girdi katmanı, çıktı katmanı ve ikisinin arasındaki bilgi akışını sağlayan saklı katmandan oluşmaktadır. Derin öğrenme, birçok saklı katmana sahip modeller için kullanılmaktadır (Pervan, 2019: 21).



**Şekil 4: Yapay Sinir Ağının Genel Görünümü**



Şekil 4’te her daire (düğüm) birer nöronu oluşturmaktadır. Her nöron ağırlık değeri (weight), yanlılık değeri (bias), ve aktivasyon fonksiyonuna sahip matematiksel bir fonksiyonu ihtiva eder. Nöronlar bir veya birden çok girdiye sahiptir. Bilgi, girdi katmanındaki saklı katmana aktarılır, saklı katmanlar işlem yapar ve bilgiyi çıktı katmanına gönderir (Pervan, 2019: 21).

Derin öğrenme altında birçok farklı teknikler bulunmaktadır bunlar:

- Derin Sinir Ağları.
- Derin İnanç Ağları.
- Derin Kısıtlı Boltzmann Makineleri.
- Derin Konvolüsyonel Ağlar.
- Derin Özdevinimli Otomatik Kodlayıcılar.
- Derin Yığılanmış Özdevinimli Otomatik Kodlayıcılar (Turgut, 2018: 59).

Nesne tanımanın en önemli kriterlerinden birisi de öznitelik çıkarım tekniğidir. Makine öğrenimi algoritmalarından önce gabor, hog, shift gibi yöntemlerle öznitelik çıkarımı işlemi yapılırken, derin öğrenmede eğitim esnasında gizli katmanlarda yapılmaktadır. Bu durum da ham veriden sınıflandırılmış veriye kadar olan sürecin insan müdahalesi olmaksızın gerçekleştiğini göstermektedir. Bu durum derin öğrenme yöntemini diğer öğrenme yöntemlerinden ayıran en önemli özelliğidir (Kavuncu, 2018: 58).

### 2.6.1.Konvolüsyon Katmanı

Öznitelik çıkarımı ve seçiminin yapıldığı katmana konvolüsyon katmanı denir. Giriş veriler veya bir önceki katmandan alınan veriler üzerinden ağırlık ismi verilen  $n \times n$  matrisindeki filtreler gezdirilerek konvolüsyon işlemi yapılmaktadır. Bu işlem sayesinde öznitelik çıkarma işlemi yapılmış olur. Öznitelik haritası elde edilerek bilgi çıkarımına ulaşılmış olur. Bilgi çıkarımının doğru olabilmesi için sınıflandırma işlemi önemlidir. Bu katmanda yapılan konvolüsyon işlemi derin öğrenme yöntemini diğer makine öğrenmesi yöntemlerinden ayırmaktadır. Genel olarak farklı bir uygulama yapmaksızın aynı sistem içerisinde hem öznitelik çıkarma işlemi hem de öznitelik seçme yöntemi gerçekleşmektedir. Derin öğrenme işleminde kurulan sistemin modeli geri yayımlı olarak çalışmaktadır. Ağın ileri yayılım aşamasında, konvolüsyon katmanında kullanılan parametre değerleri rastgele oluşturulabileceği gibi, istenilen değerlerde atanmaktadır. Bu parametreler ile eğitilen sistemin hata oranını düşürmek amacıyla çıkış biriminden giriş birimine doğru olan geri yayılım aşamasında parametreler genellikle gradyan inişi yöntemi ile değiştirilmektedir. Değişen parametreler hata oranı, ağırlık, öğrenme oranı ve bias değerlerinden oluşmaktadır (Kavuncu, 2018: 59-60).

Hata Fonksiyonu: Tahmin verisi üzerinden elde edilen sonuç ile istenilen sonuç arasındaki fark, hata oranını vermektedir. Evrimsel sinir ağı mimarisinde ortalama kare hatası olarak bilinen hata fonksiyonu kullanılmaktadır. Hata fonksiyonunun oranının sıfıra yakın çıkması beklenir (Kavuncu, 2018: 61).

İstenilen çıktılar için gerekli olan doğru ağırlıklar elde edilene kadar işlemler tekrar edilerek ağ eğitilir. Eğitim sürecinde bulunan uygun ağırlıklar test verilerine uygulanır ve istenilen sonuçların elde edilmesi beklenir. İstenilen sonuç elde edilirse ağ 'ağ eğitilmiş model ağ' olarak kabul edilmektedir. Ağırlıklar tek tek bakıldığında ne anlama geldikleri anlaşılabilir. Ağırlıklar girdiler ile ilgili analizin yapılmasına olanak sağlamaktadır. Ağın zekâsının ağırlıklarla yakından ilişkili olduğu görülmektedir. Ağırlıklar problem tiplerine göre ses, veri, görüntü gibi özelliklerle belirlenmektedir (Öztemel, 2012).

Bias değeri, verilere daha iyi uyum sağlamak için aktivasyon işlevinin sola veya sağa kaydırılmasına izin vermektedir. Bundan dolayı ağırlıklardaki değişiklikler sigmoid eğrisinin yüksekliğini değiştirirken, bias değeri onu dengeleyerek, eğriyi çözüm

için düzenli konuma getirmektedir. Bias değeri sadece çıkış değerlerinin etkilemektedir. Gerçek girdi verileriyle etkileşime girmez. Bias değeri fonksiyonun sonsuz sıfır döngüsüne girmesine izin vermez (Kavuncu, 2018:64).

Momentum katsayısı, ağı eğitilme hızını arttırmak için kullanılmaktadır. Ağ eğitilirken kullanılan parametrelerden elde edilen sonucun, bir sonraki eğitimde bir kısmının eklenmesi esasına dayanmaktadır. Ağı sürekli tekrarını önleyerek öğrenmesini sağlar. Öğrenme sırasında ağı yerel bir optimuma takılmaması için ağırlıktaki değişim değerinin bir kısmını bir sonrakine eklemektedir (Kavuncu, 2018:64).

Öğrenme oranı: ağ eğitilirken ağırlık değerleri doğru sonuca ulaşabilmek için değişiklik göstermektedir. Bu değişim miktarına öğrenme oranı denilmektedir. Öğrenme oranı düşükse, eğitim oranı daha güvenilirdir (Kavuncu, 2018: 64).

### **2.6.2. Aktivasyon Fonksiyonu**

Aktivasyon işleminden sonra elde edilen yapay sinir ağı modeli, basit bir doğrusal regresyon modelini oluşturmaktadır. Doğrusal regresyon modelinin çözümü kolay olmakla birlikte karmaşık problemler üzerinde etkili çözümler üretmeyebilir. Kısıtlı işlem gücüne sahip ve çoğu zaman istenilen performansı göstermeyen bir model oluşturmaktadır. Doğrusal bir modelin resim, video, ses gibi karmaşık problemleri öğrenip modellemesi oldukça zordur. Konvolüsyon işleminden sonra elde edilen çıktılar bir sonraki katmana iletilmeden önce belirli bir eşik değerine tabi tutulur. Eşik değeri nöronu aktif veya pasif edebilir. Aktivasyon fonksiyonu, algılayıcının doğru bilgileri elde edip etmediğini test ederek bir nöronun aktif olup olmayacağını belirlemektedir. Aktivasyon fonksiyonu ile öznitelik seçme işlemi gerçekleştirilir. Konvolüsyon işlemi ile oluşturulan öznitelik haritası içerisinde sınıflandırmayı etkileyen öznitelikler seçilmiş olmaktadır. Aktivasyon fonksiyonu, dil çevirileri ve görüntü sınıflandırmaları gibi karmaşık görevler üzerinde çalışmaktadır. ‘Başarılı, başarısız’, ‘evet, hayır’, ‘kadın, erkek’ gibi ikili sınıflandırıcılar için son katmandaki aktivasyon işlemleri kolay olmaktadır. Sınıflandırma işlemi için nöronun birinin aktif değerinin pasif olması beklenmektedir. Fakat birden fazla etiketli sınıflandırıcılar için işlem daha zor olmaktadır. Çünkü tek bir nöronun aktif değerinin pasif olması beklenirken, tüm nöronların aktif olması büyük bir sorun oluşturur. Örneğin %80 oranında aktif, %20

oranında pasif olması daha tercih edilebilir bir durumu oluşturmaktadır (Kavuncu, 2018: 65-66).

**Doğrusal Aktivasyon Fonksiyonu:** Doğrusal problemlerin çözümünde doğrusal bir aktivasyon fonksiyonu seçilmektedir. Konvülyasyon işleminde elde edilen sonuç ile sabit bir sayı çarpılarak aktivasyon işlemi gerçekleştirilmektedir. En yaygın kullanılanlardan Adım (Step) Aktivasyon Fonksiyonu, toplam fonksiyonundan gelen sonucu sıfırdan büyük olup olmamasına göre değerlendirilerek -1 veya 1 olmak üzere iki farklı çıktı vermektedir. Sonucun kontrolü yalnızca sıfır değer ile yapılmayabilir. Problem türüne göre uygun eşik değeri belirlenerek de aktivasyon işlemi sağlanabilir. Diğer sık tercih edilenlerden Sigmoid Aktivasyon Fonksiyonu ile sürekli türevlenebilir bir fonksiyonu oluşturulur. Değişim aralığı 0 ile 1 arasında olan fonksiyonlar verilmiştir. Sigmoid aktivasyon fonksiyonunun bazı sorunları vardır. Bunlar; gradyanın sıfıra yakınması sorunu, gradyan güncelleme sorunu, yavaş yakınsama sorunudur. Tanh Aktivasyon Fonksiyonu, sigmoid aktivasyon fonksiyonuna benzerdir. Tanh aktivasyon fonksiyonu sigmoid fonksiyonuna göre daha dik bir eğim göstermektedir. Gradyan inişinin sigmoid fonksiyonuna göre daha dik olması tanh fonksiyonunu avantajlı hale getirmektedir. Relu (Rectified Linear Unit) Aktivasyon Fonksiyonu, en çok kullanılan aktivasyon fonksiyonudur. Doğrusal fonksiyonlarda genellikle sıfır veya bire yaklaşan eğilimlerine göre sınıflandırma yapılmaktadır. Relu doğrusal bir fonksiyon olmadığından dolayı geri beslemeli ağlarda parametre değişimine olanak sağladığından başarılı sonuçlar elde etmektedir. Softmax Aktivasyon Fonksiyonunda ise Sigmoid fonksiyonunda olduğu gibi 0 ile 1 arasında değerler üretmektedir. Aynı zamanda tüm çıkışların toplamı 1 olacak şekilde bölünmesini sağlar ve sınıfların herhangi birinin doğru olma olasılığını belirler (Kavuncu, 2018: 66-73).

**Batch Boyutu:** Veri setinin boyutu makine öğrenmesi için çok önemlidir. Veri setinin aynı anda eğitimi işlemci, bellek ve zaman açısından maliyetli olmaktadır. Veri seti küçük gruplara ayrılarak eğitim ve öğrenme işlevi yapılmaktadır. Eğitim setinden tek seferde alınan veri sayısını oluşturmaktadır. Veri boyutu mevcut bellek ortamında eğitilebilecek kadar büyük, local optimuma takılmayacak kadar da küçük olmalıdır (Kavuncu, 2018: 83-84).

Epoch Sayısı: İleri ve geri yayılım eğitim sayısı epoch sayısını vermektedir. Epoch sayısının 1 olması demek 1 tam ileri ve geri yayılım eğitimi gerçekleştirmek anlamına gelmektedir. En yüksek doğruluk oranı elde edecek parametrelerin tespiti için en uygun epoch sayısı belirlenmesi çok önemlidir. Ağırlık, öğrenme oranı ve diğer parametrelerin güncellenerek hata oranını en düşüğe indiren değerlerin bulunması epoch sayısını fazlalaştırdıkça mümkün olmaktadır. Ancak epoch sayısının çok yüksek olması modelin ezberlemesine neden olabilmektedir (Kavuncu, 2018: 83-84).

İterasyon: Veri seti batch boyutlarına ayrılmış gruplar halinde eğitilmektedir. Bir epochta eğitilen veri boyutu sayısı iterasyonu vermektedir (Kavuncu, 2018: 83-84).

## 2.7. Kelime Gömme (Word Embedding)

Aynı anlama sahip farklı kelimelerin benzer bir vektör ile temsil edilmesi işlemine kelime gömme denir. Kelime gömme, bireysel sözcüklerin önceden tanımlanmış bir vektör uzayında gerçek değerli vektörler olarak temsil edildiği sınıflardır. Her kelime bir vektörle eşleştirilir ve vektör değerleri sinir ağına benzer bir durumda öğrenilir. Bundan dolayı teknik derin öğrenme alanında yer almaktadır. Vektör uzay boyutunda yer alan kelimeler arasındaki semantik özelliklerin bulunmasıdır. Kelime gömme işleminde birçok yöntem bulunmaktadır. Bunlardan bazıları; ikili kodlama, TF kodlaması, TF-IDF kodlaması, Latent semantik analiz kodlaması ve Word2Vec kodlamasıdır (Zhou, 2013).

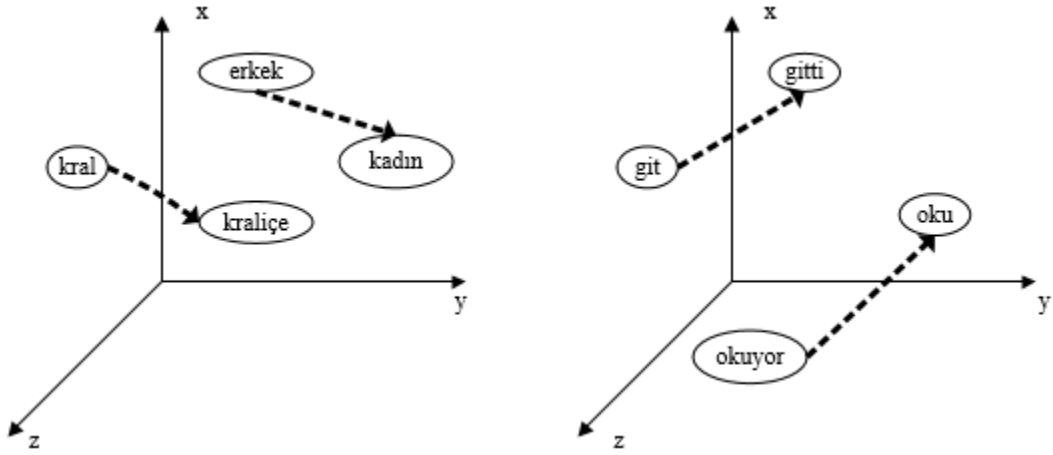
Makine öğrenimi algoritmaları metinlerin vektör değerlerini kullanmaktadır. Vektör değerlerinin kullanılması kelime gömme yönteminin farklı alanlarda kullanılmasında önemli bir etkidir (Demirtaş, 2018: 13).

Kelime gömme yönteminde vektör gösteriminin bazı önemli özellikleri:

- Boyut azaltma.
- Bağlamsal benzerlik.
- Benzer kelimelerin benzer vektörlere sahip olması.
- Gömme matrisinin daha az alan kullanması.
- Kelimeler arasındaki anlamsal ilişkilerin bulunması (Demirtaş, 2018: 13).

Kelime gömme yöntemi, anlamsal ayrıştırma ve doğal dil anlayışını kullanarak metinlerden anlam çıkarmaktadır. Bir metin modelinin metnin anlamını tahmin edebilmesi için kelimelerin bağlamsal benzerliğinden yararlanır. Kelime gömme yöntemi kelimelerin bağlamsal benzerliğini koruyan küçük boyutlu vektörel temsiller oluşturmaktadır (Demirtaş, 2018: 14).

**Şekil 5: Vektörler Arasındaki Yoğunluk Gösterimi (Demirtaş, 2018: 14).**



Sözcükler arasındaki benzerlikler Öklid mesafesi veya kosinüs benzerliği kullanılarak tanımlanmaktadır. Derin öğrenme yönteminde kullanılan yeni modeller klasik makine öğrenme yöntemindeki modellerine göre yaygın olarak kullanılan doğal dil işleme yöntemi uygulamalarıdır. Bu uygulamaların kullanıldığı bazı alanlar otomatik özetleme, makine çevirisi, adlandırılan varlık çözünürlüğü, duygu analizi, sohbet-bot, bilgi alma, konuşma tanıma ve soru cevaplama'dır. Bağlamsal benzerlikleri büyük olan resim veya metinler yakın, düşük olan değerler ise uzak olarak belirlenir. Bir karakter dizisi veya tanımlanmış bir belge birimi göz önüne alındığında tokenization işlemi sayesinde parçalara ayırma işlemi yapılır. (Demirtaş, 2018: 14).

## ÜÇÜNCÜ BÖLÜM

### TWITTER'DAKİ PAYLAŞIMLARA GÖRE İKİ FARKLI UÇUŞ FİRMASINA AİT DUYGU ANALİZİ

Bu bölümde tezde yapılan çalışmalar, kullanılan materyal ve yöntemler açıklanmıştır. Tez kapsamında C#, Matlap ve Python programlama dilleri kullanılmıştır.

#### 3.1. Araştırmanın Amacı

Bu çalışmanın amacında iki büyük hava yolu firmasının sosyal medya platformu Twitter'da “#” olarak atılan tweetleri c# dilinde geliştirilen program sayesinde saatte bir çalışarak veri tabanına verileri kayıt etme, elde edilen verilerin temizlenip hatalardan arındırılarak duygu analizi için hazırlanması ve daha sonra analiz yapılması hedeflenmiştir.

#### 3.2. Araştırma Sorunları

Sosyal medya platformu Twitter verilerinin elde edilebilmesi için kullanıcılara Twitter API'si sunulmaktadır. Fakat Twitter platformu istenildiği kadar veri elde toplamasına izin vermemektedir ve kısıtlama yapmaktadır. C# ortamında geliştirilen program sayesinde, 4 ay boyunca çalışan program ile veriler veri tabanına sorunsuz bir şekilde kayıt edilmiştir.

#### 3.3. Araştırmada Evren ve Örneklem

Bu çalışmanın evrenini sosyal medya platformu olan Twitter uygulamasına “#” olarak atılan iki büyük hava yolu firmasının tweetleri oluşturmaktadır. Çalışmanın örneklemini ise 4 ay boyunca bu firmalara atılan tweetler oluşturmuştur.

#### 3.4. Sosyal Medyada Metin Madenciliği

Sosyal medya, erişilebilir ve ölçülenebilir iletişim teknikleri kullanılarak sosyal etkileşim için kullanılan büyük bir ortamdır. Ayrıca bilgi ve deneyimleri diğer insanlarla daha etkili şekilde paylaşmak ve tartışmak için internet araçlarının kullanılması olarak tanımlanabilmektedir (Fushman vd., 2005: 9-13).

Televizyon, radyo ve gazete gibi geleneksel medya; iş ortamından tüketiciye tek bir dağıtım kanalını izlemektedir. Bilgi, medya ortamından veya reklam verenlerden ortaya konulur ve medya tüketicisine ulaştırılır. Bu geleneksel yöntemden farklı olarak

internet kullanımının artmasıyla birlikte bu durum tüketiciden tüketiciye hizmet şeklini almıştır. Bundan dolayı sosyal medya ortamları üreticilere pazar araştırması, kampanya takibi, müşteri memnuniyeti gibi önemli unsurlarda bilgi sağlayabilen bir kaynak haline almıştır (Aggarwal ve Zhai, 2012: 415-463).

Çalışmamızda duygu analizi için oluşturduğumuz sistem mimarisinin evreni sosyal medya platformu Twitter'dır. Örneklem ise Türkiye'de bulunan iki büyük hava yolu firması için atılan “#” tweetlerdir.

Twitter üzerinden elde edilen tweetler iki hava yolu firması için memnuniyet veya şikâyet gibi müşteri paylaşımlarından oluşmaktadır. Bu tweetler müşterinin duygularını ortaya koymaktadır. Atılan tweetlerden yola çıkarak hava yolu firmalarının müşteri memnuniyet ya da müşteri şikâyet düzeylerinin ölçülmesi hedeflenmiştir.

### **3.5. Veri Toplama Araçları ve Veri Analizi Araçları**

Veri toplama ve veri analizleri için kullanılan programlar bölümün alt başlıklarında incelenmiştir.

#### **3.5.1. C# Programı**

C#; C, C++ ve Java programlama dillerinden oluşan, güçlü, basit, esnek, güvenli, modern, gelişmiş ve Microsoft- .NET platformu için en baştan geliştirilmiş programlama dilidir. C#, programlama dillerinin arasında orta düzey diller grubunda yer almaktadır. Genellikle hem alt düzey hem üst düzey programlar yazılabilir (Algan, 2008: 17-18).

C# programlama dili, yazılım alanı içerisinde en çok kullanılan iki yazılım dili olan C ve C++ etkileşimi ile uyarlanmıştır. C# programı ortak platformlarda taşınabilir bir programlama dili olan Java ile çoğu özellik açısından benzerdir. En önemli özelliği ise .Net Framework platformu için hazırlanmış tamamen nesne yönelimli bir yazılım dilidir. Nesnelere önceden sınıflar halinde yazılıdır. Programcı nesneyi sürükleyerek ve sonrasında nesneyi amacına uygun çalıştırarak kod satırları yazar (Bozkurt, 2018).

Microsoft tarafından geliştirilen C#, C++ ve Visual Basic programlama dillerinde bulunan sorunları önlemek için geliştirilmiş bir programlama dilidir ve kısa sürede nesne yönelimli dillerin arasında en gelişmiş programlama dillerinden birisi olmuştur (Bozkurt, 2018).



C# programlama dili gelişmiş derleyicisi sayesinde hata olasılığını büyük oranda azaltmaktadır. Kodlanan program çalıştırıldıktan sonra derleyici tarafından istenilen sınıf ve söz dizimi hataları yazılımcıya farklı bir sekmede ayrıntılı bir şekilde gösterilmektedir. Yazılımcı hata sekmesindeki hataları kolayca bularak düzeltme işlemini yapabilmektedir (Bozkurt, 2018).

### **3.5.2. Sql Server**

Sql'in açılımı Structured Query Language (Yapısal Sorgu Dili) olan veri tabanı programı, temelde tablolarımızda yer alan verileri düzgün, sıralı bir şekilde elde edip görüntülemek için kullanılmaktadır. Bunun yanı sıra; tablo oluşturmak veya var olan tablo alanlarını modifiye etmek amacıyla da kullanılmaktadır (Köseoğlu, 2011:91).

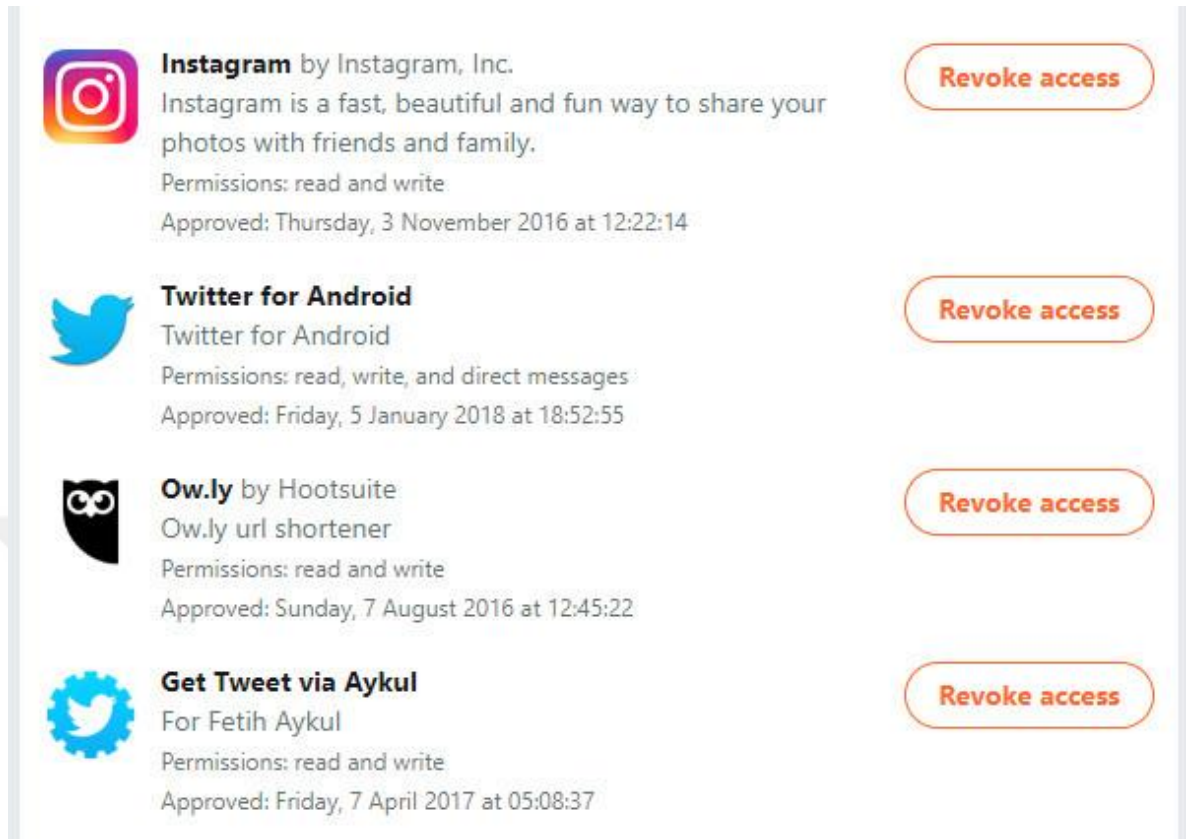
C# programı ile elde edilen tweetler sıralı bir şekilde, ön işlemlerden geçirerek sql server programına yani veri tabanına tablolar ve alt başlıklar halinde kayıt edilmektedir. Bir sonraki işlemde hazır hale gelen tweetler sql server veri tabanından çekilerek analiz edilmek üzere aktarılmaktadır.

### **3.6.Verilerin Elde Edilmesi ve Analizi**

Twitter verilerini elde etmek için .NET Framework ortamında C# programı ile program geliştirilmiştir. Katmanlı mimari yapısıyla Class Library ve Console Application yapısındadır. Elde edilen Twitter verileri veri tabanı olarak sql server programında tutulmuştur.

Twitter ortamından Twitter API aracılığıyla veri elde etmek için ilk olarak developer.twitter.com üzerinden standart API kullanılarak bir uygulama oluşturulmuştur. Bu uygulama üzerinden token (anahtar), Twitter API tarafından verilir. Twitter API sayesinde servis kullanılmak üzere .NET ortamında geliştirilir.

## Şekil 6:API Widjet

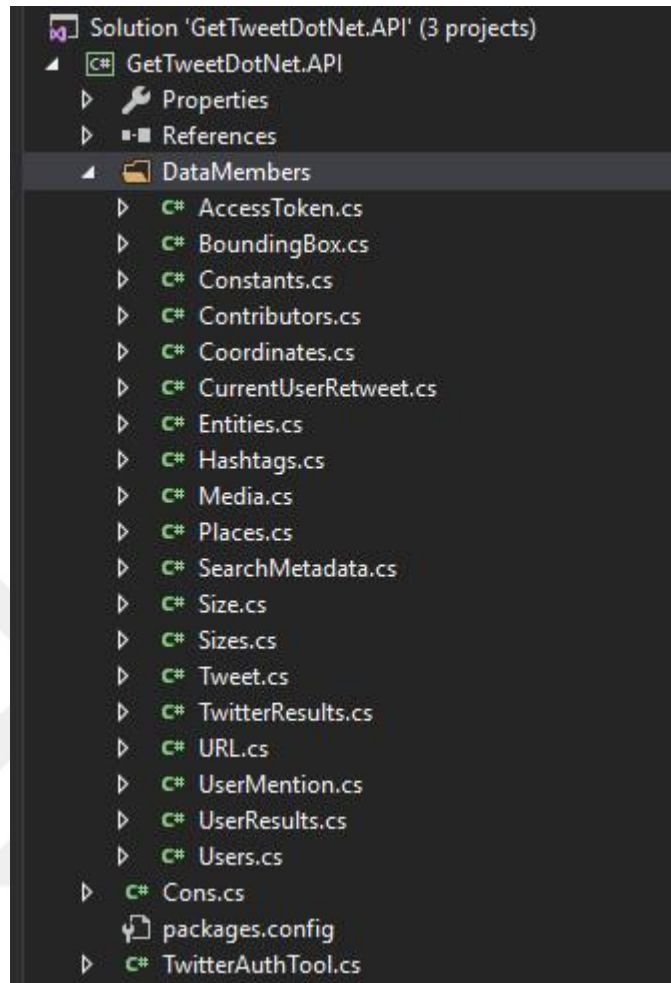


Şekil 6'da görüldüğü üzere Twitter API üzerinden birçok yöntem ile bağlantı sağlanabilir. NuGet Package Manager üzerinden hazır uygulamalar ile Twitter API özellikleri kullanılabilir. Bunlardan bir tanesi Linq2Twitter uygulamasıdır. Bu paket Linq yapısıyla C# ortamında verileri elde etmemize olanak sağlamaktadır. Bununla birlikte hazır paketten faydalanmak yerine çeşitli sınıflar ve metotlar oluşturularak veri elde edilebilir. Hazır paket üzerinden imkânlarımız ile kendi yazdığımız metotlardaki olanaklarımız aynı olmayacaktır. Bu sebeplerden dolayı daha iyi bir şekilde aramalara ve verilere erişmek için iki yöntemle de veri elde edilecektir.

Twitter API dokümantasyon, [developer.twitter.com/en/docs](https://developer.twitter.com/en/docs) internet adresinde yer almaktadır. Bu internet adresi üzerinden Twitter API'den döndürdüğü değerler görülebilmektedir. Dönen değerler ile ilgili sınıflar oluşturulmuştur.

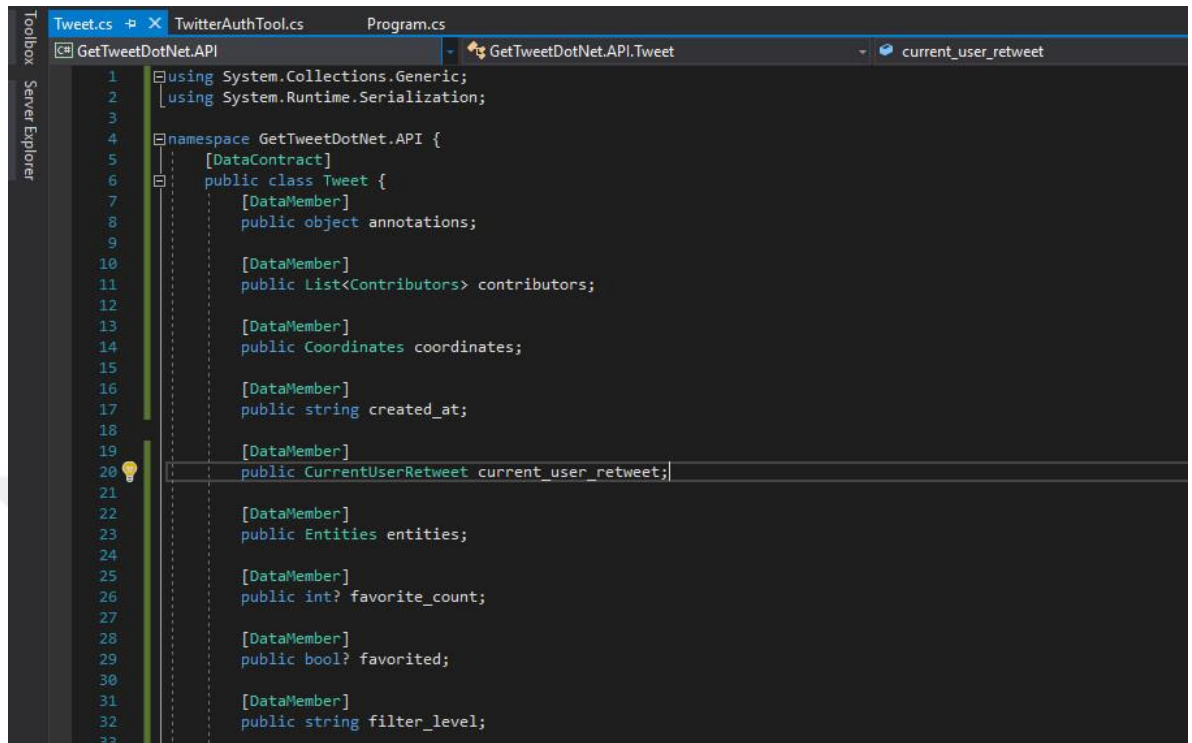
Uygulamamızda Twitter API ile ilgili olan her şey Twitter API katmanımızda yapılmaktadır. Twitter API'den dönen değerler, Data Member değerlerimizi oluşturmaktadır. Data Member barındıran sınıflar ise Data Contract olmaktadır. Bu değerler oluşturulduktan sonra Twitter API isteklerine başlanılır.

Şekil 7: Data Members Değerler



### 3.6.1. Örnek Bir Data Contract

Şekil 8:Data Contract



```

1  using System.Collections.Generic;
2  using System.Runtime.Serialization;
3
4  namespace GetTweetDotNet.API {
5      [DataContract]
6      public class Tweet {
7          [DataMember]
8          public object annotations;
9
10         [DataMember]
11         public List<Contributors> contributors;
12
13         [DataMember]
14         public Coordinates coordinates;
15
16         [DataMember]
17         public string created_at;
18
19         [DataMember]
20         public CurrentUserRetweet current_user_retweet;
21
22         [DataMember]
23         public Entities entities;
24
25         [DataMember]
26         public int? favorite_count;
27
28         [DataMember]
29         public bool? favorited;
30
31         [DataMember]
32         public string filter_level;
33

```

Twitter API üzerinden istekte bulunan sınıfımız TwitterAuthTool'dur. Bu sınıf Twitter API host tarafına istekte bulunur ve bununla bize JSON türünde değerler döndürür. Twitter üzerinden oluşturduğumuz uygulamadan aldığımız token bu sınıfa dâhil edilir. Bunlar consumer key ve secret key olarak ifade edilir. Consumer key ve secret key sayesinde Twitter API tarafından doğrulama sağlanır.

Twitter API'nin özelliklerine bağlı kalınmalıdır. Bunun için hangi aralıklarda, ne gibi özelliklerde veri elde edeceğimizi "standard search API" dökümanı açıklamaktadır.

Dönen değerler JSON olduğundan, serialize etmek için System.Runtime.Serialization.Json kütüphanesi kullanılarak DataContractJsonSerializer sınıfıyla veriler istenilen Data Contract sınıfına aktarılır. Bu aktarılanlar List türünde getirilmelidir. GetLatest metodu ile bu işlemler gerçekleştirilmektedir.

```

public List<Tweet> GetLatest(string Query, int Count = 100)
{
    TwitterResults _TwitterResults;

    List<Tweet> _ReturnValue = new List<Tweet>();

    DataContractJsonSerializer _JsonSerializer=new
DataContractJsonSerializer(typeof(TwitterResults));

    HttpWebRequest _Request =
WebRequest.Create(string.Format("{0}?q={1}&result_type=recent&count={2}",
Cons.SEARCH_URL, Query, Count)) as HttpWebRequest;

    _Request.Headers.Add("Authorization", string.Format("Bearer {0}",
Credentials.access_token));

    _Request.KeepAlive = false;

    _Request.Method = "GET";

    HttpWebResponse _Response = _Request.GetResponse() as
HttpWebResponse;

    _TwitterResults =
(TwitterResults)_JsonSerializer.ReadObject(_Response.GetResponseStream());

    _ReturnValue.AddRange(_TwitterResults.statuses);

    while
(!string.IsNullOrEmpty(_TwitterResults.search_metadata.next_results))
    {
        _Request = WebRequest.Create(string.Format("{0}{1}",
Cons.SEARCH_URL, _TwitterResults.search_metadata.next_results)) as
HttpWebRequest;

        _Request.Headers.Add("Authorization", string.Format("Bearer {0}",
Credentials.access_token));

        _Request.KeepAlive = false;

        _Request.Method = "GET";

        _Response = _Request.GetResponse() as HttpWebResponse;

        _TwitterResults =
(TwitterResults)_JsonSerializer.ReadObject(_Response.GetResponseStream());

        _ReturnValue.AddRange(_TwitterResults.statuses);

        return _ReturnValue;
    }
}

```

Benzer şekilde Twitter API'ye farklı parametrelerle de istekte bulunabiliriz. Maksimum ya da minimum değerler ile veri aratabiliriz. Bunlar için de metod oluşturulup test yapılabilmektedir.

### Şekil 9: Oluşturulan Örnek Methodlar

```
#region Direct Method
public TwitterAuthTool(string ConsumerKey, string ConsumerSecret)...
public List<Tweet> GetLatest(string Query, int Count = 100)...

public List<Tweet> GetLatestSince(string Query, long SinceId, int Count = 100)...

public List<Tweet> GetMixedSinceMaxId(string Query, long SinceId, long MaxId, int Count = 100)...

public List<Tweet> GetTweetsSinceUntil(string Query, long SinceId, string Until, int Count = 100)...

public void Dispose()...
#endregion
```

### 3.6.2. Linq2Twitter ile Tweet Çekmek

Bu hazır kütüphane sayesinde Linq yapısıyla rahatlıkla Twitter'dan Tweet elde edilebilir. Bunun için GetSearchData isiminde bir metod hazırlanır. Bu metod üzerinde doğrulama sağlayarak Context aracılığıyla verileri istediğimiz sınıfa getirebiliriz.

```
public List<Search> GetSearchData(string Word, int Count)
{
    var statusResponse = new List<Search>();
    Console.WriteLine("API isteğine başlandı");
    try
    {
        var auth = new SingleUserAuthorizer
        {
            CredentialStore = new InMemoryCredentialStore()
            {
                ConsumerKey = Cons.ConsumerKey,
                ConsumerSecret = Cons.SecretKey,
                OAuthToken = Cons.AccessToken,
                OAuthTokenSecret = Cons.AccessTokenKey
            }
        };
        var twitterCtx = new TwitterContext(auth);
        statusResponse = (from tweet in twitterCtx.Search
            where tweet.Query == Word
            && tweet.Type == SearchType.Search
            && tweet.SearchLanguage == "tr"
            && tweet.Count == Count
```

```

select tweet).ToList();

Thread.Sleep(90000); // API'den ikinci isteği hemen yapamadığı
için biraz bekletiyoruz

return statusResponse;

```

**Tablo 1: Twitter Ortamından Çekilen Örnek Tweetler**

TWEET METNİ
Vincin ucundaki A uçağı içimi acıttı <a href="https://t.co/BtOQEBNHIX">https://t.co/BtOQEBNHIX</a>
Bilet al,2 saat sonra iade et,kullanılmayan biletin kesintisi 524tl. İste #A iade sistemi. Sakın iade sistemi... <a href="https://t.co/N1vNThCzqJ">https://t.co/N1vNThCzqJ</a>
Sevgili A yetkilileri ölümden döndüm, hastalıktan kurtuldum derken, denize inmek gibi bi atraksiyonla karşıla... <a href="https://t.co/k7JuyeA9YQ">https://t.co/k7JuyeA9YQ</a>
#A #A ile ucun herkese bedava nefes alma sansi sunuyorlar diger hersey euro ya kulum (biletleri ucuz ols... <a href="https://t.co/g0mDAU4uDT">https://t.co/g0mDAU4uDT</a>
@A Teşekkür ederim verdiğiniz bilgi için... ben de yazdığımı söyleyim bari...
Galatasaray- Osmanlı maçına 2 adet kuzey alt (A alt) 1 adet doğu üst bilet vardır ilgilenenler DM
@Orcglc @A Los Angelas"a gittiysen iyi fiyat..
@sevalsoydas @elvanhira Aç kaldı bu kız sizin yüzünüzden! İnsanların psikolojisini bozdunuz, 2 saatte bu kızın evde... <a href="https://t.co/UProgwsEjJ">https://t.co/UProgwsEjJ</a>
Sen mi ? Ben mi ? Hangimiz rahat 😊 @B
3 saat teknik arızadan ötürü rötar varmışmış bu dalga geçmek olsa gerek ülkede uçak bitti çünkü suyu çıktı yedek y... <a href="https://t.co/H2yVaIYnVg">https://t.co/H2yVaIYnVg</a>
Sizinle başladığım @A yolculuğu, #B ile bitirdim çok mutluyum.Bilin istedim;))) <a href="https://t.co/rzq41RbR3R">https://t.co/rzq41RbR3R</a>
Çatır çatır çalışan canıım internet sitenizi değiştirerek halt ettiniz yap online check in yapabilirsen
Teşekkürler 😊 @B <a href="https://t.co/4yZkzu6PQ6">https://t.co/4yZkzu6PQ6</a>

RT @BrsEyn: B hava yolları reklamı mükemmel olmuş. Yine bir harikasın B..
RT @cnnturk: B'nin Paris seferine grev rötarı <a href="https://t.co/33UZid4WU6">https://t.co/33UZid4WU6</a> <a href="https://t.co/9L1NQHqgQY">https://t.co/9L1NQHqgQY</a>
Ya Rabbi sabır Dün sabah İzmir'den 7:35 de kalkması gereken uçak 9 a doğru kalktı Bugün S. Gökçen'den 21.40 da kalk... <a href="https://t.co/I4UDC39m6n">https://t.co/I4UDC39m6n</a>
70 dk rötâr mı olur yani pes @A ☐
@A Canlı destek üzerinden kısa sürede çözüm sundu . Bu güzel uygulamanız için Tebrik ederim <a href="https://t.co/OoG1SiTOFb">https://t.co/OoG1SiTOFb</a>
RT @Override1903: @A Almadığım hizmetten dolayı haksız olarak benden almış olduğunuz 40 lirayı hala iade etmediniz. Ayrıca yalan y...
RT @yenialanya: B'den Gazipaşa-Alanya Havalimanı'na uçuş müjdesi <a href="https://t.co/CRiolqkgHW">https://t.co/CRiolqkgHW</a> <a href="https://t.co/lprVWurUHa">https://t.co/lprVWurUHa</a>

### 3.6.3. Çekilen Tweet Verilerini Süzme ve Kaydetme

Alınan List türündeki değerin veri tabanına aktarılması gerekmektedir. Elde edilen toplam Tweetlerin hangi istemci üzerinden gönderildiği, hangi kullanıcıların gönderdiği gibi işlemler için süzme işlemi gerekmektedir.

İlk işlem parse işlemidir. Bu işlem sayesinde elde edilen Tweet'te farklı olan karakterler ortadan kaldırılıp Tweet daha temiz ve anlaşılabilir hale getirilmiş olur.

```
private string ParseTweet(string rawTweet)
{
    Regex link = new Regex(@"http(s)?://([\w+?\.\w+])+([a-zA-Z0-9\~\!\@\#\$\%\^\&\*\(\)_\-\=\+\|\|\\/\?\.:\;\'\,\,]*)?");
    Regex screenName = new Regex(@"@\w+");
    Regex hashTag = new Regex(@"#\w+");
    string formattedTweet = link.Replace(rawTweet, delegate (Match m)
    {
        string val = m.Value;
        return "<a href='" + val + "'>" + val + "</a>";
    });
};
```



```

        formattedTweet = screenName.Replace(formattedTweet, delegate (Match
m)
        {
            string val = m.Value.Trim('@');
            return string.Format("<a href='http://twitter.com/{0}'>{1}</a>",
val, val);
        });
        formattedTweet = hashTag.Replace(formattedTweet, delegate (Match m)
        {
            string val = m.Value;
            return string.Format("<a
href='http://twitter.com/#search?q=%23{0}'>{1}</a>", val, val);
        });
        return formattedTweet;
    }

```

Parse işlemi binlerce veri için yeterli olmasına rağmen önlem amacıyla daha basite indirgenmiş bir metot ile kontrol edilmelidir. Bu işlem için de RenderTweet metodu kullanılmalıdır.

```

public static string RenderTweet(string text)
    {
        try
        {
            text = text.ToLower();
            StringBuilder s = new StringBuilder();
            if (text.ToLower().Contains("rt"))
                text = text.Substring(2);
            var Words = text.Split(' ');
            foreach (var item in Words)
            {
                string result = string.Empty;
                if (!item.StartsWith("@") && !item.StartsWith("http://") &&
!item.StartsWith("https://") && !item.StartsWith("#"))
                {
                    if (item.Contains("http://"))
                    {
                        result = item.Substring(0, item.IndexOf("http://") -
1);
                        s.AppendFormat("{0} ", result);
                    }
                    else if (item.Contains("https://"))
                    {
                        result = item.Substring(0, item.IndexOf("https://") - 1);
                        s.AppendFormat("{0} ", result);
                    }
                    else if (item.Contains("#"))
                    {
                        result = item.Substring(0,
item.IndexOf("#") - 1);

```

```

        s.AppendFormat("{0} ", result);
    }
    else
        s.AppendFormat("{0} ", item);
    return s.ToString().Trim();
}
catch (Exception Ex)

```

**Tablo 2: Twitter Ortamından Çekilen Örnek Tweetlerin Temizlenmiş Şekli**

TWEET METNİ
vincin ucundaki A ucagi icimi acitti
bilet al2 saat sonra iade et,kullanılmayan biletin kesintisi 524tl iste iade sistemi. sakın iade sistemi
vgili A yetkilileri olumden dondum, hastaliktan kurtuldum derken, denize inmek gibi bi atraksiyonla karşıla
ile ucun herkese bedava nefes alma sansi sunuyorlar diger hersey euro ya kulum (biletleri ucuz ols
tesekkur ederim verdiginiz bilgi için ben de yazdığımı sileyim bari
galatasaray osmanli macına 2 adet kuzey alt (A alt) 1 adet dogu ust bilet vardır ilgilenenler dm
los angelasaa gittiysen iyi fiyat
ac kaldı bu kız sizin yuzunuzden insanların psikolojisini bozdunuz, 2 saatte bu kızın evde
sen mi ben mi hangimiz rahat 😊
3 saat teknik arızadan oturu rotar varmışmıs bu dalga gecmek olsa gerek ulkede ucak bitti cunku suyu çıktı yedek
sizinle başladığım ile bitirdim çok mutluyum bilin istedim
catır catır çalışan canıım internet sitenizi degistirerek halt ettiniz yap online check in yapabilirsen
tesekkurler 😊
B hava yolları reklamı mukemmel olmus yine bir harikasın B
B'nin paris seferine grev rotarı
ya rabbi sabır dun sabah izmirden 7:35 de kalkması gereken ucak 9 a dogru kalktı bugün s gökcenden 21.40 da kalk

70 dk rotar mı olur yani pes
canlı destek üzerinden kısa surede cozum sundu bu guzel uygulamaniz icin tebrik ederim
almadigim hizmetten dolayi haksiz olarak benden almıs oldugunuz 40 lirayı hala iade etmediniz ayrıca yalan y
B'den gazipasaalanya havalimanına ucus mujdesi

Tweet'in hangi istemci üzerinden atıldığını bildirmek için tweet içerisinde HTML türünde değer dönmektedir. İstemciyi bulmak için ise GetPlatform metodundan yararlanılmaktadır.

```
public static string GetPlatform(string text)
{
    try
    {
        StringBuilder s = new StringBuilder();
        text = text.ToLower();
        text = text.Substring(text.IndexOf("w\>")
3).Replace("</a>", "").Replace("twitter for ", "").Replace("client", "").Trim();
        s.AppendFormat("{0} ", text);
        return s.ToString().Trim();
    }
    catch (Exception Ex)
    {
        throw;
    }
}
```

İstemci tespitinden sonra atılan tweet'in durumu tespit edilmektedir. Tweet'in olumlu mu? Olumsuz mu? Yoksa nötr bir duygu mu? Belirttiğini bilmemiz gerekmektedir. Bunlar için SetState metodu kullanılmaktadır. Atılan Tweetler içerisinde olumlu, olumsuz ve nötr kelimeler barındırmaktadır. Bunları tespit edip, seçerek Tweet'in durumunu açıklamış oluruz. Bu metot bu işlemleri yapmamıza yardımcı olmaktadır.

```
public static string SetState(string text)
{
    try
    {
        string state = string.Empty;
        text = text.ToLower();
    }
}
```

```

#region Positive

    if (text.Contains("harika") || text.Contains("iyi") ||
text.Contains("çok güzel") || text.Contains("muhteşem") || text.Contains("süper")
|| text.Contains("😊") || text.Contains("😁") || text.Contains("👏") ||
text.Contains("🎵") || text.Contains("🎧") || text.Contains("😄") ||
text.Contains("👍") || text.Contains("😁") || text.Contains("😄") ||
text.Contains("❤️") || text.Contains("😺") || text.Contains("👊") ||
text.Contains("👊") || text.Contains("😄") || text.Contains("👊") ||
text.Contains("🏆") || text.Contains("bahar") || text.Contains("şenlik") ||
text.Contains("teşekkür") || text.Contains("bahar") || text.Contains("😄") ||
text.Contains("yarış") || text.Contains("rüzgarı") || text.Contains("love") ||
text.Contains("onur") || text.Contains("örnek") || text.Contains("tebrikler") ||
text.Contains("mutlu") || text.Contains("hayırlı") || text.Contains("tercih
ederdim") || text.Contains("işletecek") || text.Contains("gözlerim yaşlı") ||
text.Contains("yeni uçak") || text.Contains("bravo") || text.Contains("😄") ||
text.Contains("👊") || text.Contains("turkcell") || text.Contains("bedava") ||
text.Contains("hızlı") || text.Contains("çok iyi") || text.Contains("hızlı") ||
text.Contains("👍") || text.Contains("😄") || text.Contains("kral") ||
text.Contains("😄") || text.Contains("👊") || text.Contains("birinci") ||
text.Contains("😄") || text.Contains("gururlandım") || text.Contains("👍") ||
text.Contains("en büyük") || text.Contains("yönetici zirvesi") ||
text.Contains("başarılı") || text.Contains("doluluk oranı") ||
text.Contains("👊") || text.Contains("evrensel") || text.Contains("güvenli") ||
text.Contains("havayolu halkın yolu") || text.Contains("👍") || text.Contains("en
büyük") || text.Contains("👊") || text.Contains("milli bayrak") ||
text.Contains("memleketin her yeri") || text.Contains("güzel") ||
text.Contains("rekor") || text.Contains("keyif") || text.Contains("maşallah") ||
text.Contains("👍") || text.Contains("guzel") || text.Contains("gönlüne girmiş")
|| text.Contains("hayran") || text.Contains("büyüleyici") ||
text.Contains("müjde") || text.Contains("kampanya") || text.Contains("indirim")
|| text.Contains("gözünü sevdiğimin") || text.Contains("sevindirici") ||
text.Contains("geçmiş olsun") || text.Contains("yaban kazı") ||
text.Contains(":") || text.Contains("😄") || text.Contains("gezeyim") ||
text.Contains("mecerali") || text.Contains("kuruluşu") || text.Contains("milli")
|| text.Contains("aktarmasız") || text.Contains("😄") ||
text.Contains("ilgilerinden ötürü"))

    state = "Olumlu";

```

```

#endregion

#region Negative

    else if (text.Contains("çok kötü") || text.Contains("berbat") ||
text.Contains("kötü") || text.Contains("güzel değil") || text.Contains("iğrenç")
|| text.Contains("☹") || text.Contains("iğrenç") || text.Contains("çapsız") ||
text.Contains(":/") || text.Contains("çıkın") || text.Contains("leş") ||
text.Contains("çapsız") || text.Contains("köppekk") || text.Contains("yavas") ||
text.Contains("gidemiyorum") || text.Contains("allahın cezaları") ||
text.Contains("uçakta bekliyoruz") || text.Contains("rezalet") ||
text.Contains("doğru zamanlama") || text.Contains("rötar") ||
text.Contains("zamanında kalkmadılar") || text.Contains("kalkmadılar") ||
text.Contains("kalkmadılar") || text.Contains("boynuzu") ||
text.Contains("rezil") || text.Contains("ayıp") || text.Contains("pişmanlıktır")
|| text.Contains("pişman") || text.Contains("geç kaldım") || text.Contains("uçan
beygir") || text.Contains("duyarsız") || text.Contains("skandal") ||
text.Contains("☹") || text.Contains("haksız") || text.Contains("yuh") ||
text.Contains("mağduriyet") || text.Contains("mağdur") ||
text.Contains("tecrübesiz") || text.Contains("tercih etmeyeceğim") ||
text.Contains("tercih etmeyeceğim") || text.Contains("saçma") ||
text.Contains("keriz") || text.Contains("uçamadık") || text.Contains("belanı") ||
text.Contains("şrefsiz") || text.Contains("daha gelmedi") ||
text.Contains("dolandırıcılık") || text.Contains("küfür") ||
text.Contains("bavulum gelmedi") || text.Contains("bavulumu alamadım") ||
text.Contains("uçmamak") || text.Contains("☹") || text.Contains("iğrenç") ||
text.Contains("gecikmenizi") || text.Contains("allahtan korkun") ||
text.Contains("Allah'tan korkun") || text.Contains("sinirliyim") ||
text.Contains("unutur") || text.Contains("beynin bile yok") ||
text.Contains("utanmadan") || text.Contains("yetmezmiş") ||
text.Contains("yetmedi") || text.Contains("gecikmeli") ||
text.Contains("operasyonel nedenler") || text.Contains("şaşarım") ||
text.Contains("zor") || text.Contains("adliye") || text.Contains("kansert") ||
text.Contains("alagavat") || text.Contains("sömürü") || text.Contains("tıkanma")
|| text.Contains("sahtekar") || text.Contains("kabus") || text.Contains("eğitmek
gerekıyor") || text.Contains("çaresiz") || text.Contains("yazık") ||
text.Contains("şikayet") || text.Contains("batma") || text.Contains("yordunuz")
|| text.Contains("cayma") || text.Contains("vahim") ||
text.Contains("batarsınız") || text.Contains("mefta") || text.Contains("en
kötüsünüz") || text.Contains("perişan") || text.Contains("dolandıran") ||
text.Contains("kırdığınız valiz") || text.Contains("valizim yok") ||

```

```

text.Contains("kayboldu") || text.Contains("kayıp") || text.Contains("hırsızlar")
|| text.Contains("dolandırma") || text.Contains("intihar") ||
text.Contains("başarısız") || text.Contains("yaralı") ||
text.Contains("dolandırıcı") || text.Contains("👊") ||
text.Contains("güvenmiyorum") || text.Contains("😬") || text.Contains("günah") ||
text.Contains("garip") || text.Contains("tepki") || text.Contains("ağ") ||
text.Contains("fiyat uçmuş") || text.Contains("çınılıyor") || text.Contains("😡")
|| text.Contains("yaktınız") || text.Contains("igrencsin") ||
text.Contains("zalımsın") || text.Contains("rotar") ||
text.Contains("boyluyormuş") || text.Contains("çingeneler") || text.Contains("can
taşıyorsunuz") || text.Contains("metro") || text.Contains("beter") ||
text.Contains("kalkmadı") || text.Contains("yakışmıyor") || text.Contains("aç
firma") || text.Contains("isyan") || text.Contains("kullanmayacağım") ||
text.Contains("cimriler") || text.Contains("allah cezanı versin") ||
text.Contains("dehşet") || text.Contains("bagirdi") || text.Contains("ertelenme")
|| text.Contains("tartışma") || text.Contains("kaza") || text.Contains(":(") ||
text.Contains("kırdınız") || text.Contains("halt") || text.Contains("bomba") ||
text.Contains("gereksiz") || text.Contains("namussuz") ||
text.Contains("huzursuz") || text.Contains("yediniz") || text.Contains("isyanım")
|| text.Contains("kandırıp") || text.Contains("rotar") ||
text.Contains("arttirin") || text.Contains("agniyor") || text.Contains("😬") ||
text.Contains("gecikme") || text.Contains("yutturdu") ||
text.Contains("hayırdır") || text.Contains("fahiş") ||
text.Contains("oyalamaktan"))

        state = "Olumsuz";

        #endregion

        else state = "Nötr";

        return state;

    } catch (Exception Ex) {

        throw;

```

**Tablo 3: Twitter Ortamından Çekilen Örnek Tweetlerin Sınıflara Ayrılması**

TWEET METNİ	DUYGU NİTELİĞİ
vincin ucundaki A ucağı icimi acıttı	Olumsuz
bilet al2 saat sonra iade etkullanılmayan biletin kesintisi 524tl ıste iade sistemi sakın iade sistemi	Olumsuz
vgili A yetkilileri olumden dondum hastalıktan kurtuldum darken denize inmek gibi bi atraksiyonla karşıla	Olumsuz
ile ucun herkese bedava nefes alma sansi sunuyorlar diger hersey euro ya kulum biletleri ucuz ols	Olumsuz
tesekkür ederim verdiginiz bilgi için ben de yazdığımı söyleyim bari	Olumlu
galatasarayosmanli macına 2 adet kuzey alt (A alt) 1 adet dogu ust bilet vardır ilgilenenler dm	Nötr
los angelasa gittiysen iyi fiyat	Olumlu
ac kaldı bu kız sizin yuzunuzden insanların psikolojisini bozdunuz 2 saatte bu kızın evde	Olumsuz
sen mi ben mi hangimiz rahat 😊	Nötr
3 saat teknik arızadan oturu rotar varmışmıs bu dalga gecmek olsa gerek ulkede ucağ bitti cunku suyu çıktı yedek y	Olumsuz
sizinle başladığım ile bitirdim çok mutluyum bilin istedim;)))	Olumlu
catır catır çalışan caniiim internet sitenizi değiştirerek halt ettiniz yap online check in yapabilirsen	Olumsuz
tesekkürler 😊	Olumlu
A hava yolları reklamı mükemmel olmuş. yine bir harikası A	Olumlu
Anın paris seferine grev rotarı	Olumsuz

ya rabbi sabir dun sabah izmirden 7:35 de kalkması gereken ucak 9 a dogru kalktı bugün s. gökçenden 21.40 da kalk	Olumsuz
70 dk rotar mı olur yani pes ☐	Olumsuz
canlı destek uzereinden kısa surede cozum sundu bu guzel uygulamanız icin tebrik ederim	Olumlu
almadığım hizmetten dolayı haksız olarak benden almıs oldugunuz 40 lirayı hala iade etmediniz ayrıca yalan y	Olumsuz
Bden gazipasaalanya havalimanina ucus mujdesi	Olumlu

Elde edilen Tweet'ler ve duygu nitelikleri test edilebilmeleri için veri tabanına kayıt edilmektedir. Access ortamına kaydetme işlemini ise SaveDatabase metodu ile gerçekleştirilmektedir.

```
public void SaveDatabase(List<Search> statusTweets)
{
    Console.WriteLine("Veritabanı işlemlerine başlandı");
    try
    {
        string Query = string.Empty;
        string TweetRaw = string.Empty;
        string State = string.Empty;
        string Tweet = string.Empty;
        string ScreenName = string.Empty;
        string Platform = string.Empty;
        string Lang = string.Empty;
        DateTime Date;
        bool IsRetweet;
        if (Conn.State == System.Data.ConnectionState.Closed)
            Conn.Open();
        int i;
        if (statusTweets.Count > 0)
```



```

        {
            for (i = 0; i <
statusTweets[0].Statuses.Count; i++)
        {
            Query = statusTweets[0].Query;
            Date = statusTweets[0].Statuses[i].CreatedAt;
            Lang = statusTweets[0].Statuses[i].Lang;
            TweetRaw =
statusTweets[0].Statuses[i].Text.Trim().Replace("'", "\'");
            Tweet =
RenderTweet(statusTweets[0].Statuses[i].Text.Trim().Replace("'", "\'"));
            ScreenName =
statusTweets[0].Statuses[i].User.ScreenNameResponse;
            Platform =
GetPlatform(statusTweets[0].Statuses[i].Source);
            State = SetState(TweetRaw);
            IsRetweet =
statusTweets[0].Statuses[i].Text.StartsWith("RT") ? true : false;
            // Add
            OleDbCommand Com = new OleDbCommand("Insert Into
Tweets(Tweet, TweetRaw, ScreenName, Platform, ValueKey, State, IsRetweeted,
CreatedAt, Query) Values ('" + Tweet + "', '" + TweetRaw + "', '" + ScreenName +
"', '" + Platform + "', '" + State + "', '" + IsRetweet + "', '" + Date + "',
'" + Query + "')", Conn);
            Com.ExecuteNonQuery();
        }
        Console.WriteLine("Veritabanı işlemleri tamamlandı");
    }
    catch (Exception Ex)
    {

```

Console App oluşturarak bir test metodu yapılmaktadır. Oluşturduğumuz test ile verileri elde edilip veri tabanına kayıt ettirme işlemi yapılmaktadır.

```

public static string ConsumerKey = Cons.ConsumerKey;
        public static string SecretKey =Cons.SecretKey;

```

```
public static TwitterAuthTool Auth = new TwitterAuthTool(ConsumerKey,
SecretKey);

static void Main(string[] args)
{
    //GetTweet();
    CountTweetData();
    Console.ReadKey();    }
static void GetTweet()    {
    try        {
        #region A Firması
        var Result5 = Auth.GetSearchData("A Firması", 100);
        Auth.SaveDatabase(Result5);
        var Result6 = Auth.GetSearchData("@A Firması", 100);
        Auth.SaveDatabase(Result6);
        var Result7 = Auth.GetSearchData("#A Firması", 100);
        Auth.SaveDatabase(Result7);
        var Result8 = Auth.GetSearchData("#A Firması", 100);
        Auth.SaveDatabase(Result8);
        var Result9 = Auth.GetSearchData("@A Firması", 100);
        Auth.SaveDatabase(Result9);
        var Result10 = Auth.GetSearchData("#hadiozaman", 100);
        Auth.SaveDatabase(Result10);
        #endregion
        #region B Firması
        var Result11 = Auth.GetSearchData("B Firması", 100);
        Auth.SaveDatabase(Result11);
        var Result12 = Auth.GetSearchData("@B Firması", 100);
        Auth.SaveDatabase(Result12);

        var Result13 = Auth.GetSearchData("#ŞimdiNeGüzelDir", 100);
        Auth.SaveDatabase(Result13);
```

```

var Result14 = Auth.GetSearchData("#B Firması", 100);
Auth.SaveDatabase(Result14);

var Result15 = Auth.GetSearchData("B Firması", 100);
Auth.SaveDatabase(Result15);

var Result17 = Auth.GetSearchData("#B Firması", 100);
Auth.SaveDatabase(Result17);

var Result18 = Auth.GetSearchData("@B Firması", 100);
Auth.SaveDatabase(Result18);

#endregion

```

Windows servisleriyle sunucu üzerinde belirli aralıklarda uygulama çalışarak Tweetler çekilmektedir. Her çalıştığında elde edilen Tweet'ler veri tabanımıza kayıt edilmektedir. Bunun için bir service app oluşturulmalı ve bir servis sunucuya kurulmalıdır.

Servisi oluşturmak için Windows Service türünde proje oluşturulmuştur. İsteklerin zamanını ayarlamak için Timer üzerinden değil Quartz kütüphanesinden konfigürasyon sağlanmaktadır. Bu kütüphane istediğimiz metodu belirli zaman periyotlarında trigger işlemi ile çalıştırmaktadır.

```

using System;
using System.ServiceProcess;
namespace GetTweetDotNet.Service
{
    public partial class Service : ServiceBase
    {
        JobScheduler Scheduler;

        public Service()
        {
            InitializeComponent();
        }

        protected override void OnStart(string[] args)
        {
            Scheduler = new JobScheduler();
            Scheduler.Start();
        }
    }
}

```

```

    }    public void Search()    {
        try    {
            Scheduler = new JobScheduler();
            Scheduler.Start();    }
        catch (Exception Ex)
            protected override void OnStop()    {
            if (Scheduler != null)
                Scheduler.Stop();    }
    }

```

Trigger işlemi için kütüphane içerisinde bir job ataması yapılmaktadır.

```

using Quartz;
using Quartz.Impl;
namespace GetTweetDotNet.Service
{
    public class Job : IJob
    {
        Get _Get = new Get();    public void
        Execute(IJobExecutionContext context)
        {
            _Get.GetTweet();    }    }
    public class JobScheduler    {    public void Start()    {
        IScheduler Scheduler = StdSchedulerFactory.GetDefaultScheduler();
        Scheduler.Start();
        IJobDetail Job = JobBuilder.Create<Job>().Build();
        ITrigger Trigger = TriggerBuilder.Create()
            .WithSimpleSchedule(a    =>
a.WithIntervalInHours(1).RepeatForever())    .Build();
        Scheduler.ScheduleJob(Job, Trigger);
    }    public void Stop()
    IScheduler scheduler = StdSchedulerFactory.GetDefaultScheduler();
    scheduler.Shutdown();
}

```

Get sınıfı ile API kütüphanesinden veri isteğinde bulunmaktadır.

```
using GetTweetDotNet.API;
using System;
namespace GetTweetDotNet.Service
{
    public class Get
    {
        public static string ConsumerKey = Cons.ConsumerKey;
        public static string SecretKey = Cons.SecretKey;
        public static TwitterAuthTool Auth = new TwitterAuthTool(ConsumerKey,
        SecretKey);
        public void GetTweet()
        {
            try
            {
                A Firması
                B Firması}
            catch (Exception Ex)
            {
```

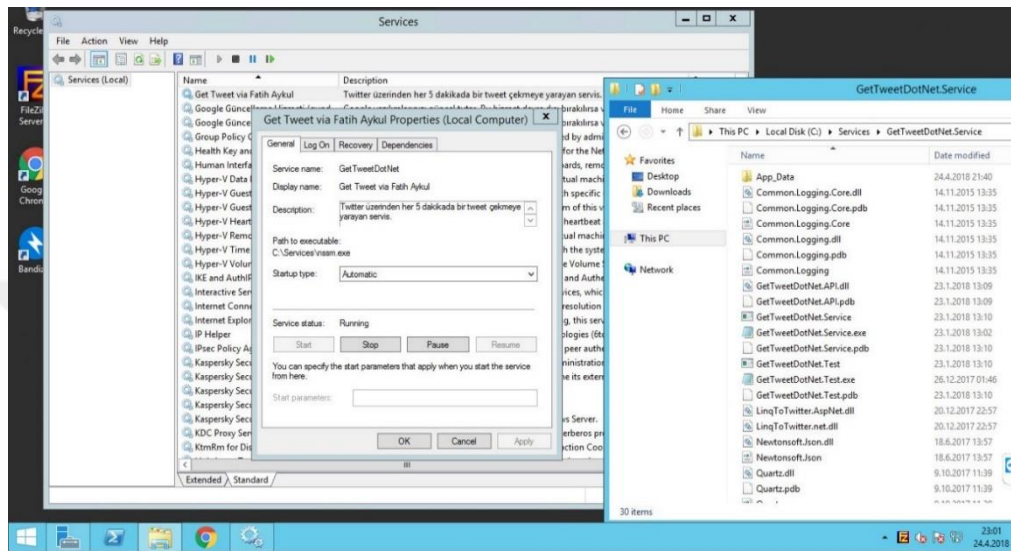
Servis hazırlama işlemi bittikten sonra sunucu işlemine geçilmektedir. Bu işlem Aruba Cloud üzerinden sağlanmaktadır. Aruba Cloud'tan VMware altyapılı cloud server kiralanmıştır.

### Şekil 10:Cloud Server

The screenshot shows the Aruba Cloud VPS website. The header includes 'aruba CLOUD' and navigation links like 'HOSTING', 'CLOUD', 'PRODUCTS & SOLUTIONS', 'STARTUP', 'INFRASTRUCTURE', 'WHY CHOOSE ARUBA', 'COMPANY', and 'USE YOUR VOUCHER'. The main content area has a blue background with a server rack image. The text reads 'Cloud VPS' and 'All the flexibility of Cloud Technology in 4 simple sizes'. A green banner on the right says 'more than 20% off'. Logos for CISPE, Intel Xeon, and VMware are present. The footer includes a sidebar with 'CLOUD VPS', 'FEATURES', and 'TECH SPECS' links, and a main section titled 'Cloud VPS - VPS Hosting' with a description: 'Our Cloud VPS can be activated quickly and easily within our 100% cloud infrastructure which guarantees a VPS hosting service with high standards of performance and reliability. The virtual private server solution features servers that are preconfigured, available in 4 sizes with various technical components, all based on VMware technology with built-in SSD storage.'

İşlemin yönetimi sağlandıktan sonra, servis uygulaması sunucuya yükledikten sonra kurulması gerekmektedir. Aynı zamanda elde edilen verilerin kaydedilmesi için Access Database Engine programı kurulmalıdır. Bu işlemlerden sonra servis çalışmaya başlamaktadır.

**Şekil 11: Windows Servisi**



### 3.6.4. Çekilen Verileri Saydırma İşlemi

Servisimiz belirli bir süre veri elde ettikten sonra veri tabanındaki RT veriler çıkartılarak, sadece ham olan Tweet verileri getirme işlemleri için yeni bir metod oluşturulmuştur.

```
static void CountTweetData()
{
    Console.WriteLine("Veritabanına bağlantı kuruluyor");

    OleDbConnection Conn = new OleDbConnection("Provider=Microsoft.ACE.OLEDB.12.0;Data Source=" + AppDomain.CurrentDomain.BaseDirectory + "App_Data\\Database.accdb;Persist Security Info=False;");

    OleDbConnection ConnNew = new OleDbConnection("Provider=Microsoft.ACE.OLEDB.12.0;Data Source=" + AppDomain.CurrentDomain.BaseDirectory + "App_Data\\DatabaseNew.accdb;Persist Security Info=False;");

    List<TweetData> _List = new List<TweetData>();

    List<TweetData> _RetweetedList = new List<TweetData>();
}
```

```

        _RetweetedList.Add(new TweetData { TweetRaw = "", ScreenName = "",
State = "" });

        int TotalCount = 0;

        int IsRetweetedCount = 0;

        int PositiveCount = 0;

        double PositiveCountPercent;

        int NegativeCount = 0;

        double NegativeCountPercent;

        int NotrCount = 0;

        double NotrCountPercent;

        try
        {
            Conn.Open();
            ConnNew.Open();

            OleDbDataAdapter Da = new OleDbDataAdapter("Select * from Tweets
Where (Query = 'A Firması' OR Query = '@A Firması' OR Query = 'A Firması.com/tr
')", Conn);

            DataTable Dt = new DataTable();

            Da.Fill(Dt);

            TotalCount = Dt.Rows.Count;

            Console.WriteLine("Veriler düzenleniyor");

            for (int i = 0; i < Dt.Rows.Count; i++)
            {
                TweetData _Data = new TweetData();

                _Data.Query = Dt.Rows[i]["Query"].ToString();

                _Data.ScreenName = Dt.Rows[i]["ScreenName"].ToString();

                _Data.TweetRaw = Dt.Rows[i]["TweetRaw"].ToString();

                _Data.State = Dt.Rows[i]["State"].ToString();

                _Data.Platform = Dt.Rows[i]["Platform"].ToString();

                _Data.IsRetweeted
                =
                Convert.ToBoolean(Dt.Rows[i]["IsRetweeted"]);

                _Data.Tweet = Dt.Rows[i]["Tweet"].ToString();

```

```

        _Data.CreatedAt =
Convert.ToDateTime(Dt.Rows[i]["CreatedAt"]);

        _List.Add(_Data);
    }

    Conn.Close();

    Console.WriteLine("İstatistikler oluşturuluyor");

    TotalCount = _List.Where(x => x.IsRetweeted == false).Select(x =>
x.TweetRaw).Distinct().Count();

    var RawTotalLst = _List.Where(x => x.IsRetweeted ==
false).GroupBy(x => x.TweetRaw).Select(x => x.FirstOrDefault()).ToList();

    foreach (var item in RawTotalLst)
    {
        OleDbCommand _Com = new OleDbCommand("Insert Into
Tweets(Query, Tweet, TweetRaw, ScreenName, State, Platform, IsRetweeted,
CreatedAt) Values('" + item.Query + "', '" + item.Tweet + "', '" + item.TweetRaw
+ "', '" + item.ScreenName + "', '" + item.State + "', '" + item.Platform + "', "
+ item.IsRetweeted + "', '" + item.CreatedAt + "')", ConnNew);

        _Com.ExecuteNonQuery();
    }

    var _IsRetweetedCount = _List.Where(x => x.IsRetweeted ==
true).GroupBy(x => x.TweetRaw).Select(x => x.FirstOrDefault()).ToList();

    foreach (var item in _IsRetweetedCount)
    {
        OleDbCommand _Com = new OleDbCommand("Insert Into
Tweets(Query, Tweet, TweetRaw, ScreenName, State, Platform, IsRetweeted,
CreatedAt) Values('" + item.Query + "', '" + item.Tweet + "', '" + item.TweetRaw
+ "', '" + item.ScreenName + "', '" + item.State + "', '" + item.Platform + "', "
+ item.IsRetweeted + "', '" + item.CreatedAt + "')", ConnNew);

        _Com.ExecuteNonQuery();
    }

    for (int i = 0; i < _IsRetweetedCount.Count; i++)
    {

        string _ScreenName = _IsRetweetedCount[i].ScreenName;

        string _TweetRaw = _IsRetweetedCount[i].TweetRaw;

        var _jitCount = _RetweetedList.Where(x => x.ScreenName ==
_ScreenName && x.TweetRaw == _TweetRaw).FirstOrDefault();

        if (_jitCount == null)

            _RetweetedList.Add(_IsRetweetedCount[i]);
    }
}

```



```

IsRetweetedCount = _RetweetedList.Count();

PositiveCount = _List.Where(x => x.State == "Olumlu").Count();
NegativeCount = _List.Where(x => x.State == "Olumsuz").Count();
NotrCount = _List.Where(x => x.State == "Nötr").Count();

PositiveCountPercent = (100 * PositiveCount) / _List.Count;
NegativeCountPercent = (100 * NegativeCount) / _List.Count;
NotrCountPercent = (100 * NotrCount) / _List.Count;

#endregion

Console.WriteLine("");
Console.WriteLine("");
Console.WriteLine("Toplanan veri: " + _List.Count);
Console.WriteLine("Ham veri toplamı: " + TotalCount);
Console.WriteLine("Sadece retweet edilmişlerin toplamı: " +
IsRetweetedCount);
Console.WriteLine("Pozitif tweetlerin toplamı: " +
PositiveCount);
Console.WriteLine("Negatif tweetlerin toplamı: " +
NegativeCount);

Console.WriteLine("Nötr tweetlerin toplamı: " + NotrCount);

Console.WriteLine("Toplam %" + PositiveCountPercent + " olumlu,
%" + NegativeCountPercent + " olumsuz, %" + NotrCountPercent + " nötr veri
bulundu");
}

```

Bu işlemlerden sonra elde edilen verileri saydırma işlemi tamamlanmıştır.

### 3.7.Bulgular ve Değerlendirme

Duygu analizi konusundaki çalışmalarda kullanılabilir Türkçe metin mesajlardan oluşmuş, herkese açık, ulaşılabilir bir Twitter Türkçe veri seti mevcut değildir. Bu sebeple öncelikle Twitter API kullanılarak Türkçe metin mesajları içeren bir veri seti oluşturulmuştur. Twitter API'ye gönderilen her sorgu için maksimum 100 Türkçe metin mesajı alınmasına izin verdiği için, geliştirilen C# uygulaması ile daha fazla sayıda Türkçe metin mesajlarına ulaşılması sağlanmıştır. Türkçe metin mesajlarının içerdiği duygu ve his bulunduğu gruba göre olumlu, olumsuz veya nötr olarak etiketlenmiştir.

İki büyük hava yolu firması için atılan “#”li Türkçe Twittler 4 ay boyunca bir sunucu sistem sayesinde sisteme olumlu, olumsuz veya nötr şeklinde ayrılarak kayıt edilmiştir. Oluşturulan Türkçe metin veri seti ön işlemlerden geçirildikten sonra 30328 adet kayıttan oluşmaktadır. Bu kayıtlardan olumlu tweet sayısı 10092, olumsuz tweet sayısı 10100 ve nötr tweet sayısı ise 10136 adettir. Ön işlem aşamasında his simgeleri, hashtag, URL ve kullanıcı adları temizlenmiştir. Temizleme işleminden sonra metinlere ait duyguların otomatik tespitinde klasik sınıflandırma algoritmaları ve son yıllarda popüler olan derin öğrenme yöntemi ile birlikte kelime gömme (Word Embedding) analizleri uygulanmıştır.

Bu çalışmada klasik sınıflandırma işlemlerinde matlab ve python programlama dilleri kullanılmıştır. Kullanılan modellerin karşılaştırmasını yapmak için eğitim sonuçlarını gözlemleyerek farklı sinir ağı mimarilerini ve hiper parametrelerini değiştirmek ve optimize etmek için TensorBoard kullanılmıştır. TensorFlow ve Keras'ta ağ eğitimi yapılırken özel bir log yolu tanımlanarak aktivasyon metrikleri için TensorBoard eklenmiştir. Bu işlemlerin yapılabilmesi için “Keras callbacks” olarak ifade edilen eğitim süresince çalışacak olan özel bir fonksiyon sınıfı kullanılmıştır.

#### 3.7.1.Klasik Sınıflandırma Analizi ve Değerlendirmesi

Metinlerin duygularını belirlemek için literatürde popüler olan klasik sınıflandırma yöntemlerinden YSA, KNN, Bayes ve SVM kullanılmış olup en iyi sınıflandırma başarısını belirlemek için bu yöntemler farklı parametreler ile 10-kat çapraz doğrulama yöntemi kullanılarak test edilmiştir. Bu parametrelerden iterasyonun sonuca etkisini incelemek için 100 ve 300 değerleri kullanılırken, YSA için gizli

katmanlardaki nöron sayısı olarak da bir dögü ile [20,10,5]' in kombinasyonları denenmiştir. Buların yanında öğrenme algoritmalarının da sınıflandırma sonucuna etkisi olabileceği düşünülerek YSA için Scg, Lm, ve Br olmak üzere üç farklı eğitim algoritmasından yararlanılmıştır. SVM için ise lineer ve radyal taban fonksiyonu kernelleri kullanılmıştır.

Farklı parametre ve öğrenme algoritmaları kullanılarak DVM, Bayes ve KNN yöntemleri ile gerçekleştirilen sınıflandırma sonuçları Tablo 4'de verilmiştir.

**Tablo 4: Klasik Sınıflandırıcılar Analiz Sonuçları**

Sınıflandırma Algoritmaları	Ön işlem Durumu	Parametre	Test Sonucu	Eğitim Sonucu
Bayes	Standart	-	39.28	39.83
	Normalize	-	39.28	39.83
	PCA	-	39.79	40.47
KNN	Standart	-	47.21	-
	Normalize	-	<b>47.21</b>	-
	PCA	-	46.85	-
DVM	Standart	Lineer	31.79	32.10
	Standart	RBF	44.86	<b>99.88</b>
	Normalize	Lineer	33.27	33.59
	Normalize	RBF	44.36	64.80
	PCA	Lineer	30.54	30.78
	PCA	RBF	44.86	<b>99.88</b>

Tablo 4'de elde edilen değerlere göre KNN sınıflandırma algoritması ile normalize test sonucunda 47.21 en yüksek değer elde edilmiştir. Eğitim değerlendirme sonucunda ise DVM'da PCA ve Standart RBF'de 99.88 eğitim sonucu ile en yüksek değerler elde edilmiştir. Bayes sınıflandırma algoritmasında ise en yüksek test sonucu 39.79 eğitim sonucunda ise 40.47 bulunmuştur. DVM sınıflandırma algoritmasında ise standart RBF ve PCA RBF de en yüksek test sonucu 44.86 elde edilmiştir.

Farklı parametre ve öğrenme algoritmaları kullanılarak YSA ile gerçekleştirilen sınıflandırma sonuçları ise Tablo 5, Tablo 6 ve Tablo 7 da görülmektedir.

YSA ile normalize edilmemiş standart veriler ile analiz sonuçları tablo 5'de verilmiştir.

**Tablo 5: YSA Standart Analiz Sonuçları**

İterasyon Sayısı	Nöron Sayıları	Eğitim Algoritması	Test Sonucu	Eğitim Sonucu
300	20 nöron - 20 nöron	SCG	39.70	40.92
		BR	40.51	43.09
		LM	<b>42.16</b>	<b>49.45</b>
	20 nöron- 10 nöron	SCG	37.37	38.45
		BR	40.28	43.63
		LM	41.75	46.65
	20 nöron -5 nöron	SCG	39.01	39.87
		BR	40.39	41.91
		LM	41.60	43.69
	10 nöron- 10 nöron	SCG	37.49	38.08
		BR	41.33	43.29
		LM	41.60	43.69
	10 nöron- 5 nöron	SCG	37.27	37.69
		BR	40.17	41.48
		LM	41.02	43.25
	5 nöron - 5 nöron	SCG	35.20	35.09
		BR	38.57	39.43
		LM	41.08	41.66
100	20 nöron- 20 nöron	SCG	38.31	38.80
		BR	41.11	44.25
		LM	<b>42.32</b>	<b>49.14</b>
	20 nöron- 10 nöron	SCG	37.97	38.81
		BR	40.78	43.11
		LM	41.79	45.93
	20 nöron -5 nöron	SCG	37.00	37.78
		BR	41.02	43.05
		LM	41.15	45.53
	10 nöron - 10 nöron	SCG	37.46	37.59
		BR	41.18	43.21
		LM	41.50	44.11
	10 nöron - 5 nöron	SCG	36.83	36.85
		BR	40.52	42.14
		LM	41.33	43.51
	5 nöron - 5 nöron	SCG	34.99	35.23
		BR	39.73	40.55
		LM	40.95	41.91

Tablo 5'de de görüldüğü gibi YSA standart analiz sonucunda 300 iterasyon sayısında 20 nöron 20 nöronunda LM eğitim algoritması ile test sonucu 42.16, eğitim sonucu 49.45 olarak en yüksek değerler elde edilmiştir. 100 iterasyon sayısında 20

nöron 20 nöronda LM eğitim algoritmasında test sonucu 42.32, eğitim sonucu 49.14 olarak yüksek değerlere ulaşılmıştır.

YSA'nın normalize edilmiş veri seti analiz sonuçları Tablo 6'da verilmiştir.

**Tablo 6: YSA Normalize Analiz Sonuçları**

İterasyon Sayısı	Nöron Sayıları	Eğitim Algoritması	Test Sonucu	Eğitim Sonucu
300	20 nöron - 20 nöron	SCG	40.19	41.90
		BR	<b>42.75</b>	<b>50.66</b>
		LM	41.93	50.10
	20 nöron- 10 nöron	SCG	40.46	41.93
		BR	42.26	48.82
		LM	41.70	47.87
	20 nöron -5 nöron	SCG	41.35	42.60
		BR	42.39	47.10
		LM	41.91	46.62
	10 nöron- 10 nöron	SCG	40.71	41.66
		BR	42.28	45.68
		LM	41.56	44.87
	10 nöron- 5 nöron	SCG	40.52	41.07
		BR	41.79	44.54
		LM	41.37	44.20
	5 nöron - 5 nöron	SCG	40.23	40.67
		BR	40.58	42.92
		LM	40.73	42.68
100	20 nöron- 20 nöron	SCG	40.31	40.75
		BR	42.52	<b>50.27</b>
		LM	42.24	50.06
	20 nöron- 10 nöron	SCG	38.74	39.3
		BR	<b>42.57</b>	48.33
		LM	41.77	47.90
	20 nöron -5 nöron	SCG	38.80	39.08
		BR	41.64	46.77
		LM	41.37	46.16
	10 nöron - 10 nöron	SCG	38.91	39.58
		BR	42.03	45.26
		LM	41.38	44.78
	10 nöron - 5 nöron	SCG	38.52	39.03
		BR	41.55	44.50
		LM	41.74	43.87
	5 nöron - 5 nöron	SCG	39.71	39.70
		BR	40.82	42.73
		LM	40.50	42.33

Tablo 6’da da görüldüğü gibi YSA normalize analiz sonucunda 300 iterasyon sayısında 20 nöron 20 nöronda BR eğitim algoritmasında test sonucu 42.75, eğitim sonucu 50.66 olarak en yüksek değerler elde edilmiştir. 100 iterasyon sayısında 20 nöron 10 nöronda BR eğitim algoritmasında test sonucu 42.57, 100 iterasyon 20 nöron 20 nöronda BR eğitim algoritmasında eğitim sonucu 50.27 olarak en yüksek değerler elde edilmiştir.

YSA’nın veri setinin %80’nini açıklama kabiliyeti bulunan ve öznelik sayısı 10’a düşürülmüş PCA analiz sonuçları tablo 7’de verilmiştir.

**Tablo 7: YSA PCA Analiz Sonuçları**

İterasyon Sayısı	Nöron Sayıları	Eğitim Algoritması	Test Sonucu	Eğitim Sonucu
300	20 nöron - 20 nöron	SCG	39.52	40.73
		BR	41.02	43.56
		LM	<b>42.45</b>	<b>48.52</b>
	20 nöron- 10 nöron	SCG	38.65	39.22
		BR	40.44	43.03
		LM	41.98	44.94
	20 nöron -5 nöron	SCG	36.43	36.76
		BR	40.09	41.64
		LM	41.10	43.16
	10 nöron- 10 nöron	SCG	35.82	36.20
		BR	40.57	41.99
		LM	41.74	43.98
	10 nöron- 5 nöron	SCG	34.85	35.24
		BR	39.99	40.65
		LM	40.96	42.99
	5 nöron - 5 nöron	SCG	34.66	34.81
		BR	38.73	39.26
		LM	41.14	41.69
100	20 nöron- 20 nöron	SCG	37.62	37.95
		BR	40.32	42.08
		LM	41.70	<b>46.34</b>
	20 nöron- 10 nöron	SCG	38.05	38.65
		BR	40.72	42.89
		LM	<b>41.94</b>	44.69
	20 nöron -5 nöron	SCG	36.71	37.61
		BR	40.69	41.86
		LM	41.51	43.08
	10 nöron - 10 nöron	SCG	35.26	35.28
		BR	40.35	41.66
		LM	41.52	43.61

10 nöron - 5 nöron	SCG	35.89	36.16
	BR	39.67	40.55
	LM	40.86	42.33
5 nöron - 5 nöron	SCG	35.33	35.24
	BR	38.23	38.79
	LM	40.63	41.39

Tablo 7’de görüldüğü gibi YSA PCA analiz sonucunda 300 iterasyon 20 nöron 20 nöronda LM eğitim algoritmasında test sonucu 42.45, eğitim sonucu 48.52 olarak en yüksek değerler elde edilmiştir. 100 iterasyon sayısında 20 nöron 10 nöronda LM eğitim algoritması test sonucu 41.94, 100 iterasyon 20 nöron 20 nöronda LM eğitim algoritmasında eğitim sonucu 46.34 olarak en yüksek değerler elde edilmiştir.

Elde edilen klasik sınıflandırma algoritmalarına ait sonuçlar incelendiğinde en iyi sonucu KNN’nin test setinde %47.21 başarı verdiği görülmüştür. Bu başarı oranının sınıflandırma için yetersiz olduğu görülmüş olup, metin madenciliğinde son yıllarda başarılı sonuçlar veren derin öğrenme tabanlı kelime gömme işlemi de çalışmaya dâhil edilmiştir

### 3.7.2.Derin Öğrenme Tabanlı Kelime Gömme Analiz Sonuçları

Kelime gömme metin madenciliği yöntemlerinde son zamanlarda sıklıkla kullanılmaya başlanmıştır. Metinler üzerinde sınıflandırma veya mevcut veri kullanılarak otomatik metin oluşturma gibi örnekleri görülmektedir. Bu çalışmada da twitter verilerin sınıflandırılmasındaki başarıyı arttırmak adına kelime gömme katmanı derin öğrenmeye eklenmiştir. Metinler içerisinde sıklıkla geçen kelimeler sınıflandırmada önemli özellikleri oluşturmakta olup 30328 adet tweet’e ilişkin sıklıkla geçen kelimeleri ifade eden kelime bulutu Şekil 12’de görülmektedir.

**Şekil 12: Kelime Bulutu**

Sınıflandırma işleminde öznitelikleri oluşturan ve metinlere ait tokenization işlemi sonucunda 60000 kelimedenden numaralandırılarak elde edilen 20 adet kelime kullanılarak derin öğrenme ağı eğitilmiştir. Tokenization işlemi sonucu elde edilen ve derin öğrenme ağının girişi için öznitelikleri oluşturan veri seti Tablo 8'de görülmektedir.

**Tablo 8: Tokenization Sonucu**

Metin	Etiket	Tokenization Sonucu
vincin ucundaki A uçağı içimi açtı	Olumsuz	[5387 1739 44133 36747 45675 48345 0 0 0 0 0 0 0 0 0 0 0 0]
A uçak görünmesin diye önüne branda çekmiş yanlardan hiç görmedik zaten 😞😞	Olumsuz	[44133 13550 59112 20704 4751 31998 22810 3800 57533 31804 20446 0 0 0 0 0 0 0 0]
özel olarak geliştirilmiş havacık kanalları sayesinde ayaklar nefes alır. zoom air iç ve dış yastıklama sistemiyle...	Olumsuz	[13931 50005 9642 59428 32675 43860 34295 32437 48265 30578 2602 59095 13991 43603 24208 54610 0 0 0 0]
daha uçuş gününe girmeden rötat mesajı atan rötatlı saatimiz gelip havaalanına geldiğimizde şova...	Olumsuz	[51265 45593 43117 50697 43650 45415 4556 43069 36849 16155 31059 27274 41331 0 0 0 0 0 0 0]
önce 25 şimdi de 15 dakika daha gecikme,yazıklar olsun !	Olumsuz	[41062 42383 44267 101 24182 9384 51265 50942 34766 54952 0 0 0 0 0 0 0 0 0]
piste inen uçak bileti 150.tl pist	Olumsuz	[26703 18049 13550 6121 53235 30652



dışına inen uçak bileti 20.tl		54826 18049 13550 6121 56563 0 0 0 0 0 0 0 0 0]
çağrı merkezinde çalışan arkadaşları tekrar bir eğitimden geçirmenizi tavsiye ederim. çok basit...	Olumsuz	[38002 19435 41736 6025 47898 32397 36944 14422 59528 23979 27610 29747 0 0 0 0 0 0 0]
biletim iptal ettirdim ama iade yapıp yapılmadığını kontrol edemiyorum bakıp bilgi verir misiniz	Nötr	[50556 29303 52802 20461 46919 9911 14522 43578 11286 45458 22715 42396 59370 0 0 0 0 0 0 0]
sadece 2 haftalık bir uygulama sonra normale dönecektir.	Nötr	[55043 13232 52599 32397 56388 38842 54524 22177 0 0 0 0 0 0 0 0 0 0 0]
mobil uygulama şu anda tüm uçuşlar için "yer yok" veriyor ilginizi rica ederim	Nötr	[51920 56388 38388 33671 45237 40145 20535 20680 28762 2035 45935 25838 17250 0 0 0 0 0 0 0]
gideliimmmmmm ㄹ ㄹ	Nötr	[43784 19267 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
uçacağı kaçırmak değilde şu check-in olayı ile ilgili bi ilgilenir misiniz?	Nötr	[36747 30922 26391 38388 49910 2167 56719 50927 32507 32454 36244 0 0 0 0 0 0 0]
son 4 koltuk var biletin tükenmesine. benim sabitlediğimde satılırmı onu öğrenmem gerek.	Nötr	[29072 26716 36052 706 8782 42793 53487 51203 57808 40080 735 8217 0 0 0 0 0 0 0]
peki aktarmalı yolcular ne olacak. bazen ara çok kısa olabiliyor <input type="checkbox"/>	Nötr	[8484 56754 1359 57712 1102 2371 27712 27610 2032 8151 13653 0 0 0 0 0 0 0]
bir gün gezisi teşekkürler ucuz bilet mühteşem tatil	Olumlu	[32397 12716 41581 33116 7347 29165 54197 14191 0 0 0 0 0 0 0 0 0 0 0]
boyle daha gercekci ve guzel oldu	Olumlu	[13062 51265 12388 13991 3392 42395 0 0 0 0 0 0 0 0 0 0 0 0]
ruhumu teslim ettim, teşekkürler	Olumlu	[35196 45547 22620 33116 0 0 0 0 0 0 0 0 0 0 0 0 0]
talebimi olumlu yanıtlamışsınız. 😊 çok teşekkür ederim. emeği geçen herkesin eline sağlık....	Olumlu	[12486 30775 25291 10668 27610 54438 23979 4851 20143 9658 49431 37230 0 0 0 0 0 0 0]
bir de şu kutsal beldelelere bilet fiyatlarınız makul hale getirseniz çok güzel olacak.	Olumlu	[32397 101 38388 45090 37924 29165 31464 53075 53140 18482 27610 57040 1102 0 0 0 0 0 0 0]
ver şöyle 2 kişilik bilet hediye. keşfedelim en güzelinden.	Olumlu	[48411 6962 13232 40971 29165 37867 22229 2997 38178 0 0 0 0 0 0 0 0 0 0]
hayırlı olsun seferler bol fiyatlar ucuz olsun inşallah	Olumlu	[12406 34766 1142 59391 10168 7347 34766 55964 0 0 0 0 0 0 0 0 0 0 0]

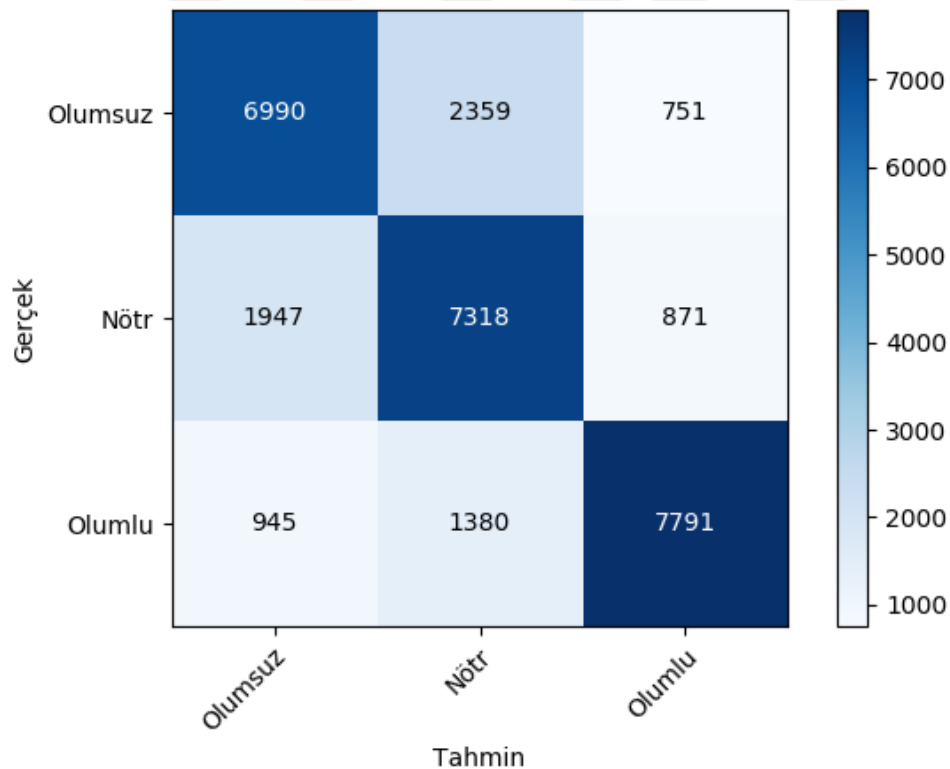
Derin öğrenme ağının sınıflandırılmasında 10-kat çapraz doğrulama yöntemi kullanılmış olup ağın eğitimi için Adam algoritması ile eğitim 10 epoch boyunca gerçekleştirilmiştir. Eğitilen ağa ilişkin katmanlar Tablo 9’da görülmektedir.

**Tablo 9: Derin Öğrenme Ağına Ait Katmanlar**

Katman	Parametreler
Embedding	Giriş: 20 Kelime, 60000 Kelimelik Sözlük, Çıktı: 250 Ağırlıklandırılmış Kelime
Flatten	-
Dense	3 Nöron
Classification	Softmax Aktivasyon Fonksiyonu

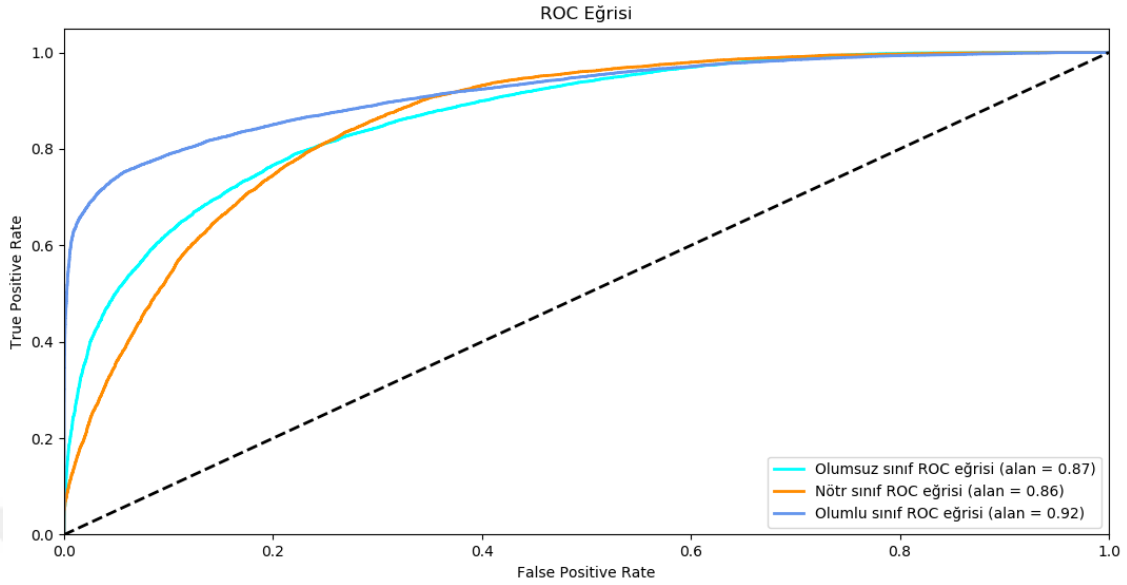
Sınıflandırma sonucunda eğitim setinde  $99.84 \pm 0.02$  başarı oranı elde edilirken test setinde ise  $72.81 \pm 0.95$  başarı oranı elde edilmiştir. Sınıflandırma sonucuna ait çarpıklık matrisi Şekil 13’de görülmektedir.

**Şekil 13: Confusion Matris**



Sınıflandırma sonucuna ilişkin ROC eğrisini ve bunun altında kalan alanı gösteren sonuçlar ise Şekil 14’de görülmektedir.

Şekil 14: ROC Eğrisi

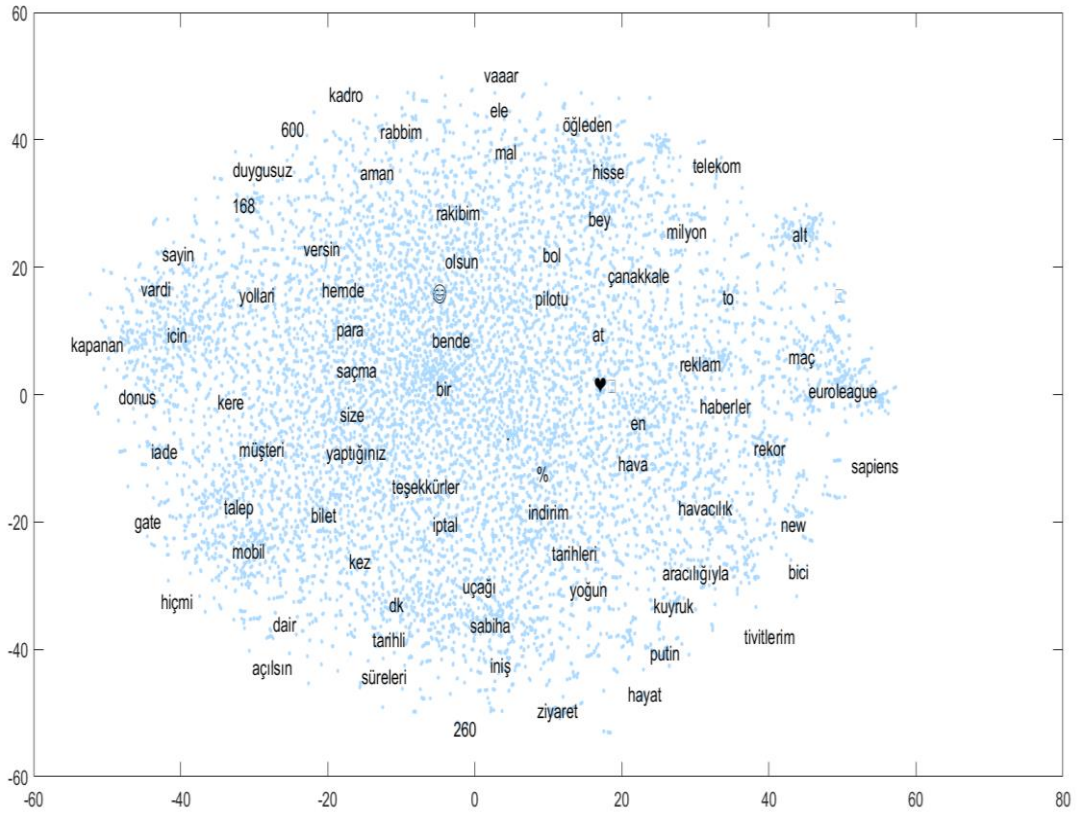


Şekil 14’de görüldüğü gibi sınıflara ait ROC eğrisi altında kalan alanlar yüksek olup yeterli düzeyde sınıflandırma başarısı elde edildiği görülmektedir.

Kelime gömme işleminin metinlerin sınıflandırılmasında klasik yöntemlere göre daha iyi sonuçlar verdiği görülmüş olup %72.81 başarı oranı ile twitter verileri duygu analizi için sınıflandırılmıştır. Elde edilen başarı düzeyi yeterli düzeyde olup bu sayede hava yolu firmalarının geliştirecekleri yapay zekâ sistemleri ile kendilerine gelen tweetleri analiz edip müşteri memnuniyetlerine ait hızlı bir şekilde durum değerlendirmesi yapabilmelerinin önü açılmıştır.

Kelime gömme işlemi sınıflandırma başarısını arttırmış olup ilgili katman ile veriler arasındaki semantik ilişkiler kurulmaktadır. Kelimeler arasında vektörel olarak ağırlıklandırma işlemleri ileri ve geri besleme yöntemleri kullanılarak sağlanmıştır. Birbirine benzer kelimelerin vektörel değerleri yakın olurken, birbirinden alakasız kelimelerin değerleri vektörel olarak uzak olmaktadır. Analizde kullanılan verilerin kelime gömme katmanı sonucu vektörel uzaklıkları Şekil 15’teki gibidir.

**Şekil 15: Kelimelerin Yakınlık-Uzaklık Görünümü**



Şekil 15 incelendiğinde tweetlere ilişkin sınıfları belirten seçici kelimelerin birbirine vektörel olarak daha yakın oldukları görülmektedir.

## SONUÇ

Hızla gelişen internet ortamında yeralan verilerin boyutların büyümesiyle, elverişli verinin diğer verilerden ayırt edilebilmesi son zamanlarda önemli bir çalışma alanı haline gelmiştir. Literatür incelemesi sonucunda özellikle duygu analizi alanında derin öğrenme yöntemleriyle yapılan çalışmalara katkı sağlamak için bu tez çalışmasında sosyal medya platformu twitter ortamından elde edilen iki hava yolu firmasına ait Türkçe veri setleri üzerinde otomatik olarak duygu sınıflandıran modeller geliştirilmiştir. Bu amaç doğrultusunda, duygu analizi ve metin sınıflandırma alanında birçok çalışma incelenmiştir. Duygu analizi alanında derin öğrenme yöntemleriyle yapılan çalışmalara katkı sağlamak ve sınıflandırma başarısını arttırmak için kelime gömme yönteminden faydalanılmıştır.

Sosyal medya kullanımının dünya genelinde artmasıyla birlikte, sosyal medyanın kendine özgü yazım dili oluşmuştur. Genellikle Twitter ortamında karakter sınırlamasının olduğu sosyal medya siteleri ile birlikte kısaltmalar, sesli harflerin kullanılmaması, yanyana bitişik yazım gibi yazım tarzları ortaya çıkmıştır. Sosyal medyada Türkçe yazım kuralı yapısından çok uzak metinlerin olması sebebiyle, bu metinler üzerinde metin madenciliği ve duygu analizi uygulamaları oldukça zor olmaktadır. Veri setleri metin tabanlı olduğu için duygu analizi ve metin sınıflandırma işlemleri yapılabilmesi için doğal dil işleme ve veriyi sınıflara ayırmak için sınıflandırma algoritmaları ve aktivasyonlar kullanılmıştır.

Sosyal medya platformlarında bir olayın bir haberin yayılma hızı göz önüne alındığında, paylaşılan mesajın kurumun imajını kötü etkilemesi çok hızlı olmakta ve sonucunda kurum, müşteri kaybetme, kötü izlenim gibi birçok maddi manevi zarara uğrayabilmektedir. Verilerin duygu durumları belirlendikten sonra duygu kriterlerinin birbiriyle olan ilişkisi incelenip, kullanıcıların neden o tarzda paylaşım yaptıkları öğrenilmeye çalışılabilir. Böylece kurumun imajının kötülenmesinin önüne geçilmiş olup, müşteri memnuniyeti ve kurumun durumu korunmuş olur. Sosyal medya kullanıcılarının paylaştığı mesajların tutarlılığını tespit etmek için paylaşımların ayrı ayrı duygu durumlarını tespit etmek gerekmektedir.

Derin öğrenmede kelime gömme katmanı için öncelikle veri setinde ön işlem aşamaları uygulanmıştır. Tokenizing işlemi ile metin içersindeki kelimeler ayrıştırılıp

60000 kelimeden oluşan sözlük ile 20-öznitelikten oluşacak şekilde numaralandırılmıştır. Bu öznitelikler ile gerçekleştirilen klasik ve derin öğrenme sınıflandırma sonuçları birbirleri ile kıyaslanmıştır. Yapılan çalışmada geliştirilen derin öğrenme yöntemi ile %72.81 başarı oranı elde edilmiş olup, bu sonuçların klasik sınıflandırıcılardan daha iyi olduğu gözlemlenmiştir. Derin öğrenmede, kelime gömme başarılı bir sınıflandırıcı olarak ortaya çıkmıştır.

Havayolu firmaları ve ürün reklamlarında sosyal medyadan elde edilen verilerle metin madenciliği adımları uyguladıktan sonra, duygu analizi işlemleri yapılarak konu ile ilgili doğru bilgiler elde edilmiştir. Elde edilen sonuçlara göre havayolu firmaları ve pazara yeni girmek isteyenler, müşteri memnuniyeti ve anında reklam stratejileri geliştirebilecek, kullanıcı yorumlarına bakılarak doğru kararlar alabileceklerdir.

Aynı zamanda ulaşılan sonuçlarla havayolu firmaları hakkında müşteri yorumları otomatik olarak analiz edilebilecek ve bunun sayesinde analiz maliyetleri düşürülebilecektir. Havayolu firmaları, müşteriler veya ürünlerin niteliklerine otomatik olarak anında ulaşılabilecektir. Bundan dolayı müşteri memnuniyeti ve havayolu firmalarının marka stratejisinin daha hızlı ve daha düşük maliyet ile gerçekleştirilebilmesine olanak sağlanabileceği düşünülmektedir.

Duygu analizi sonucunda elde edilen diğer sonuçlarda, havayolu firmalarının günlük yapılan kampanyalar veya reklamlar doğrultusunda hızlı bir geri bildirim alabilmesidir. Sosyal medyada paylaşılan yorumların analizi sonucu, ilgili firmanın veya reklamlarının çok hızlı bir zamanda nasıl bir izlenim bıraktığının görülebilmesini sağlayabilmektedir.

Elde edilen sonuçlarla havayolu firmalarının politikalarının ve sunulan hizmetin kalitesinin eksikliği olup olmadığını hızlı bir şekilde değerlendirebilmektedir. Bundan dolayı yapılan yeniliklerin de geri dönüşümü çok hızlı bir şekilde alınarak, ilgili yeniliğin faydalı olup olmadığı hızlı bir şekilde görülüp bilgi sahibi olunabilmektedir. Bu durum sayesinde gereksiz yapılan yatırımların önüne geçilebilecektir. Çalışmanın havayolu firmaları için oldukça önemli bir müşteri ilişkileri yönetimi aracı olabileceği düşünülmektedir.

Çalışmanın kısıtları arasında sadece iki farklı hava yolu firmasına ait tweetlerin değerlendirilmesi gösterilebilir.

İleride yapılacak çalışmalar arasında belirli bir zaman diliminde ardı ardına olumlu veya olumsuz özellikte tweetlerin atılması durumunda bu süreci otomatik yönetecek bir sistemin geliştirilmesi düşünülmektedir.



## KAYNAKÇA

- Abdi, H., and Williams, L. J., (2010), "Principal Component Analysis", *Wiley Interdiscip. Rev. Comput.*, Vol. 2, Issue.4, (433-459).
- Adsız, A., (2006), "Metin Madenciliği", Ahmet Yesevi Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, Kazakistan.
- Aggarwal, C. C., ve Zhai, C., (2012), *Mining Text Data*, Edt: C. Z., Charu ve C. Aggarwal, Springer, New York.
- Akba, F., (2014), "Duygu Analizinde Öznitelik Seçme Metriklerinin Değerlendirilmesi: Türkçe Film Eleştirileri", Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Ankara.
- Akbaş, E., (2012), "Türkçe Tweetlerde Konu Bazlı Düşünce Analizi", Bilkent Üniversitesi Mühendislik ve Bilimler Enstitüsü Yayınlanmış Yüksek Lisans Tezi, İstanbul.
- Aksu, M., (2012), "Hizmet Kalitesinin Bir Unsuru Olarak Atmosferin Müşteri Sadakati Üzerine Etkisi: Bozcaada'daki Otellerde Konaklayan Yerli Turistler Üzerine Bir Araştırma", Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü, Yayınlanmış Doktora Tezi, Balıkesir.
- Albayrak, A. S., ve Yılmaz, Ş., (2009), "Veri Madenciliği: Karar Ağacı Algoritmaları ve İmkb Verileri Üzerine Bir Uygulama", *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, C. 14, N. 1, (31-52).
- Algan, S., (2008), *Her Yönüyle C#*, Pusula Yayıncılık İletişim Yazılım İmalat San. Tic. Ltd. Şti. İstanbul.
- Alpaydın, E., (2010), *Introduction to Machine Learning*, The MIT Press Cambridge, Massachusetts London, England.
- Amanet, H., (2017), "Türkçe Sosyal Medya Metinlerinde Duygu Analizi", Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Trabzon.
- Bloemer, M. M. J., and Brijs, T., and Vanhoof, K., and Swinnen, G., (2003), "Comparing Complete And Partial Classification For Identifying Customers At



Risk”, *International Journal Of Ressearch In Marketing* , Vol.20, Issue.2, (117-131).

Bollegala, D., and Maehara, T., and Kawarabayashi k., (2011), “Unsupervised Cross-Domain Word Representation Learning”, *Kawarabayashi Large Graph Project*, Vol.7, (730-740).

Bozkurt, S., (2011), <http://teknokoliker.com/2011/11/c-nedir-c-temelleri-nelerdir.html> (Erişim Tarihi: 19.03.2018).

Byrne, B. M., (2000), *Structural Equation: Modeling With AMOS: Basic Concepts, Applications, and Programming*, Routledge Taylor & Francis Group New York, London.

Cool, A. L., (2002), “A Review Of Methods For Dealing With Missing Data”, *Annual Meeting Of The Southwest Educational Resarch Association*, Dallas.

Çağıl, A., (2015), <https://www.beyaztarih.com/makale/twitterin-tarihcesi> (Erişim Tarihi: 17.02.2018).

Çoban, Ö., (2016), “Metin Sınıflandırma Teknikleri İle Türkçe Twitter Duygu Analizi”, Atatürk Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Erzurum.

Çüm, S. ve Demir, E. K. ve Gelbal, S. ve Kışla, T., (2018), “Kayıp Veriler Yerine Yaklaşık Değer Atamak İçin Kullanılan Gelişmiş Yöntemlerin Farklı Koşullar Altında Karşılaştırılması”, *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, S.45, (230-249).

Demirtaş, M. A., (2018), “Derin Öğrenme Yöntemleri ile Duygu Analizi ve Aktivasyon Fonksiyonlarının Karşılaştırılması”, Atatürk Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Erzurum.

Deperlioğlu, Ö., (2018), “Farklı Eğitim Algoritmalarının Tıbbi Veritabanlarının Yapay Sinir Ağı Kullanılarak Sınıflandırılmasına Etkileri”, *European Conference on Science, Art & Culture (ECSAC)*, (91-99).

- Dindar, B., (2016), “Hizmet Sektörü Veri Setlerinde Dilsel Özetleme Ve Duygu Analizi Uygulamaları”, Gazi Üniversitesi Bilişim Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Ankara.
- Ekiz, E. H. ve Köker, N. E., (2010), “Şikayetlerin Kısıtlayıcı Faktörleri: Turistlerin Belirgin Şikayet Etme Davranışları”, *Journal Of Yaşar University*, C.17, S.5, (2859-2873).
- Emel, G. ve Taşkın, Ç., (2005), “Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması” *Eskişehir Osman Gazi Üniversitesi Sosyal Bilimler Dergisi*, C.6, S.2, (221-239).
- Eroğul, U., (2009), “Türkçe Metinlerde Düşünce Çözümleme”, Orta Doğu Teknik Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Ankara.
- Fushman, D. D. And Lin, J., (2005), “Knowledge Extraction For Clinical Question Answering: Preliminary Results”, *Proceedings Of The AAAI-05 Workshop On Question Answering In Restricted Domains*, (9-13).
- Gupta, V. And Gurpreet, S. L., (2009), “A Survey Of Text Mining Techniques And Applications”, *Journal Of Emerging Technologies In Web Intelligence*, Vol.1, Issue.1, (60-76).
- Güven, A. ve Bozkurt, Ö. Ö. ve Kalıpsız, O., (2007), “Veri Madenciliğinin Geleceği”, *Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Dergisi*, İstanbul.
- Hatipoğlu, S. ve Işık, E. S., (2015), “Havayolu Ulaşımında Hizmet Kalitesinin Ölçülmesi: İç Hatlarda Bir Uygulama”, *Kahramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Dergisi*, C. 12, S. 2, (293-312).
- Haykin, S. O., (2008), *Neural Networks And Learning Machines*, Pearson Prentice Hall, New York.
- Işıkoğlu, M. A., (2017), “Kayıp Veri İle Baş Etme Yöntemlerinin Ölçme Değişmezliğine Etkisi Açısından Karşılaştırılması”, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Ankara.

- İlhan, S. ve Duru, N. ve Karagöz, Ş. ve Sağır, M., (2008), “Metin Madenciliği ile Soru Cevaplama Sistemi”, *Kocaeli Üniversitesi Mühendislik Fakültesi Dergisi*.
- Jiang, J. and Zhai, C., (2007), “Instance Weighting For Domain Adaptation in NLP”, *Association For Computational Linguistics*, (264-271).
- Joachims, T., (1998), “Text Categorization With Support Vector Machines: Learning With Many Relevant Features”, *ECML '98 Proceedings of the 10th European Conference on Machine Learning London, UK: SpringerKVerlag*, (137-142).
- Kaplan, B. A., (2016), “Twitter Üzerindeki Türkçe Mesajlarda Veri Madenciliğiyle Duygu Analizi”, Beykent Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, İstanbul.
- Karadoğan, A., (2014), “Takviyeli Öğrenme İçin Yapay Atom Algoritması (A<sup>3</sup>) Kullanımı”, İnönü Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Malatya.
- Kavuncu, S. K., (2018), “Makine Öğrenmesi ve Derin Öğrenme: Nesne Tanıma Uygulaması”, Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Kırıkkale.
- Kazançoğlu, İ., (2011), “Havayolu Firmalarında Müşteri Sadakatinin Yaratılmasında Kurum imajının ve Algılanan Hizmet Kalitesinin Etkisi”, *Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, C.21 (130-158).
- Kemp, S., (2017), <https://wearesocial.com/special-reports/digital-in-2017-global-overview> (Erişim Tarihi: 01.16.2018).
- Kingma, D. P., and Ba, J. L., (2015), “Adam: A Method For Stochastic Optimization”, *Published as a Conference Paper at ICLR*, (2-3).
- Kılınç, D. ve Borandağ, E. ve Yucalar, F. ve Tunalı V., (2016), “KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi”, *Marmara Üniversitesi Fen Bilimleri Dergisi*, C.3, (89-94).
- Kozak, M., (2007), “Turizm Sektöründe Tüketicilerin Şikâyetlerini Bildirme Eğilimleri”, *Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, C.14, S.1, (137-151).

- Kozan, T. O., (2017), “Başvuru Yapan Adaylardan İş Bulanların Tahmin Edilmesi”, Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, İstanbul.
- Köseoğlu, K., (2011), *Veri Tabanı Mantığı*, Pusula Yayıncılık İletişim Yazılım İmalat San. Tic. Ltd. Şti. İstanbul.
- Liu, B., (2012), *Sentiment Analysis and Opinion Mining*, Morgan&Claypool Publishers.
- Bozorgi, M. M., (2006), “Measuring Service Quality İn The Airline Using Servqual Model”, Lulea University Of Technology Master Thesis, Sweden.
- Mackay, D. J. C., (1992), “Bayesian Interpolation”, *Neural Computation*, Vol. 4, (415-447).
- Mitchell, T. M., (1997), *Machine Learning, Mcgraw-Hill Series İn Computer Science*. McGraw-Hill New York.
- Namukasa, J., (2013), “The Influence Of Airline Service Quality On Passenger Satisfaction And Loyalty: The Case Of Uganda Airline Industry”, *The TQM Journal*, Vol.25, Issue.5, (520-532).
- Ofir, C. And Simonson, I., (2001), “In Search Of Negative Customer Feedback: The Effect Of Expecting To Evaluate On Satisfaction Evaluations”, *Journal Of Marketing Research*, Vol.38, Issue.2, (170-182).
- Okumuş, A. ve Asil, H., (2007), “Hizmet Kalitesi Algılamasının Havayolu Yolcularının Genel Memnuniyet Düzeylerine Olan Etkisinin İncelenmesi”, *İstanbul Üniversitesi İşletme Fakültesi İşletme Dergisi*, C.36, S.2, (07-29).
- Osborne, J. W., (2013), *Best Practices In Data Cleaning*, Sage Publication Inc, California.
- Özgür, A., (2002), “Supervised And Unsupervised Machine Learning Techbiques For Text Document Categorization”, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, İstanbul.
- Pak, M. Y., (2015), “Metinlerde Duygu Analizi Ve Sınıflandırma İçin Yeni Yöntemler”, Anadolu Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Eskişehir.

- Park, J. and Robertson, R. and Wu, C., (2005), "Investigating The Effects Of Airline Service Quality On Airline Image And Passengers' future Behavioural Intentions: Findings From Australian International Air Passengers", *The Journal Of Tourism Studies*, Vol.16, Issue.1, (2-11).
- Pekkaya, M ve Akıllı, F., (2013), "Hava Yolu Hizmet Kalitesinin SERVPERF-SERVQUAL Ölçeği İle Değerlendirilmesi Ve İstatiksel Analizi" *Ekonomik Ve Sosyal Araştırmalar Dergisi*, C.9, S.1, (75-96).
- Periyasamy, T. and Senthil, K., (2014), "An Efficient Clustering Algorithm For Text Mining Using Greedy Approach", *International Journal Of Advanced Research In Computer Science & Technology*, Vol.2, Issue.3, (257-263).
- Pervan, N., (2019), "Derin Öğrenme Yaklaşımları Kullanarak Türkçe Metinlerden Anlamsal Çıkarım Yapma", Ankara Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Ankara.
- Pilavcılar, İ. F., (2007), "Metin Madenciliği İle Metin Sınıflandırma", Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, İstanbul.
- Polattimur, R., (2019), "Derin Öğrenme Yöntemi ile Köpek Davranışlarının Analizi ve Sınıflandırılması", Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Bilecik.
- Rose, T., (2002), "The Reuters Corpus From Yesterday's News To Today's Language Resources", *Language Resources and Evaluation*, (29-31).
- Salton, G. And Buckley, C. (1988), "Term-Weighting Approaches In Automatic Text Retrieval", *In: Information Processing & Management*, Vol.24, Issue.5, (513-523).
- Sebastiani, F., (2002), "Machine Learning In Automated Text Categorization", *ACM Computing Surveys*, Vol.34, Issue.1, (1-47).
- Sevindi, B. İ., (2013), "Türkçe metinlerde denetimli ve sözlük tabanlı duygu analizi yaklaşımlarının karşılaştırılması", Gazi Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, Ankara.

- Singh, A. K., (2015), "Modeling Passengers' Future Behavioral Intentions In Airline Industry Using SEM", *Journal Of Advances In Management Research*, Vol.12, Issue.2, (107-127).
- Sukı, N. M. (2014), "Passenger Satisfaction With Airline Service Quality In Malaysia: A Structural Equation Modeling Approach", *Research In Transportation Business & Management*, Vol.10, (26-32).
- Sumanthy, K. L. And Chidambaram, M., (2013)., "Text Mining: Concepts, Applications, Tools And Issues: An Overview" *International Journal Of Computer Applications*, Vol.80, Issue.4, (28-30).
- Sutton, R. S. and Barto, A. G., (2012), *Reinforcement Learning An Introduction*, Second Edition The MIT Press, London.
- Şahin, Ş., (2012), "Büyük Menderes Nehri Üzerindeki Akım Gözlem İstasyonlarında Eksik Verilerinin Tamamlanması", Pamukkale Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Denizli.
- Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., and Qin B., (2014), "Learning Sentiment-Specific Word Embedding For Twitter Sentiment Classification. *In Proceedings Of The 52nd Annual Meeting Of The Association For Computational Linguistics*, Vol.1. (1555).
- Tekin, A. T., (2017), "Veri Analitiğinin Ekonomik Etkilerini : İnternet Yayıncılığında Metin Tabanlı Öneri Motoru Örneği", İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, İstanbul.
- Tina, R. P. and Sherekar, S. S., (2013), "Performance Analysis Of Naive Bayes And J48 Classification Algorithm For Data Classification", *International Journal Of Computer Science And Applications*, Vol.6, Issue.2, Amravati.
- Topaçan, Ü., (2016), "Sosyal Medya Paylaşımında Duygu Analizi: Makine Öğrenimi Yaklaşımı Üzerine Bir Araştırma", Marmara Üniversitesi Sosyal Bilimler Enstitüsü Yayınlanmış Doktora Tezi, İstanbul.
- Tsaur, S. and Chang, T. and Yen, C., (2002), "The Evaluation Of Airline Service Quality By Fuzzy MCDM", *Tourism Management*, Vol.23, (107-115).

- Turgut, Z., (2018), “Nesnelerin İnterneti İçin Hareketlilik Yönetimi”, İstanbul Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Doktora Tezi, İstanbul.
- Tüzün, İ. K. ve Devrani, T. K., (2008), “Müşteri Memnuniyeti ve Müşteri-Çalışan Etkileşimi Üzerine Bir Araştırma”, *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, C.3, S.2, (13-24).
- Uytun, M. B., (2013), “Duygu Analizi Ve E-Ticaret Sistemleri İçin Kullanıcı Hareketlerinden Veri Toplanması”, Çankaya Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Ankara.
- Vavra, T. G., (1999), “Müşteri Tatmini Ölçümlerinizi Geliştirmenin Yolları”, Edt. Günhan Günay, Kalder Yayınları, İstanbul.
- Vijayarani, S. and Ilamathi, J. and Nithya, M., (2015), “Preprocessing Techniques For Text Mining – An Overview”, *In International Journal Of Computer Science & Communication Networks*, Vol.5, Issue.1, (7-16).
- Özgirgin, T., (2016), “Sosyal Medyada Duygu Analizi Ve Nitelik Çıkarımı”, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, İstanbul.
- Öztemel, E., (2012), *Yapay Sinir Ağları*, Papatya Yayıncılık, İstanbul.
- Yıldız, O. ve Erdil T. S., (2013), “Türkiye Havayolu Yolcu Taşımacılığı Sektöründe Hizmet Kalitesinin Karşılaştırmalı Ölçülenmesi”, *Marmara Üniversitesi Sosyal Bilimler Enstitüsü Öneri Dergisi*, C.10, (89-100).
- Yıldız, O. ve Bal, A. ve Gülsecen, S., (2015), “Statistical And Clustering Based Rules Extraction Approaches For Fuzzy Model To Estimate Academic Performance In Distance Education”, *Eurasia Journal Of Mathematics Science & Technology Education*, Vol.11, Issue.2, (391-404).
- Yurt, E. A., (2015), “Türkçe Metinlerde Duygu Analizi”, Maltepe Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, İstanbul.
- Yücesan, M. M., (2016), “Bağlı Veri Kaynakları Ve İlişkileri Kullanılarak Haberlerin Öbeklendirilmesi”, Toob Ekonomi ve Teknoloji Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmış Yüksek Lisans Tezi, Ankara.

Zhou, S. and Chen, Q. And Wang, X., (2013), “Active Deep Learning Method For Semi-Supervised Sentiment Classification”, *Neurocomputing*, Vol. 120, (536-546).

Weiguo, F. and Wallace, L. and Rich, S. and Zhang, Z., (2006), “Tapping Into The Power Of Text Mining”, *Journal Of ACM Blacksburg*, Vol.49, Issue.9 (77-82).

Wilamowski, B. M. and Yu, H., (2010), “Improved Computation for Levenberg-Marquardt Training”, *Transactions on Neural Networks*, Vol. 21, Issue.6, (930-937).





## ÖZGEÇMİŞ

### **Kişisel Bilgiler** :

Adı Soyadı : Fatih AYKUL

Doğum Yeri : Bucak/Burdur

Medeni Hali : Bekar

### **Eğitim Durumu** :

Lisans Öğrenimi : Mehmet Akif Ersoy Üniversitesi Bucak Zeliha Tolunay  
Uygulamalı Teknoloji ve İşletmecilik Yüksek  
Okulu/Yönetim Bilişim Sistemleri

### **Yabancı Diller ve Düzeyi** :

İngilizce Orta Düzey

### **İş Denevimi** :

Yıldız Holding Bilgi İşlem Sistem Destek Uzmanı