

## COVID-19 SÜRECİNDE TWITTER MESAJLARININ DUYGU ANALİZİ

### SENTIMENT ANALYSIS OF TWITTER MESSAGES IN COVID-19 PROCESS

**Güncel SARIMAN** 

Dr. Öğr. Üyesi, Muğla Sıtkı Koçman Üniversitesi Teknoloji Fakültesi Bilişim Sistemleri Mühendisliği  
Bölümü, Muğla/Türkiye

**Emre MUTAF** 

Yüksek Lisans Öğrencisi, Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü Bilişim Sistemleri  
Mühendisliği Anabilim Dalı, İzmir/Türkiye

*Geliş Tarihi / Received: 17.05.2020*  
*Kabul Tarihi / Accepted: 14.07.2020*

*Araştırma Makalesi/Research Article*  
*DOI: 10.38065/euroasiaorg.149*

#### ÖZET

İnsanların gündelik yaşamlarındaki değişikliklerin sosyal medya mecraları üzerinden paylaşılması geride değerli bilgilerin birikmesine ve bu bilgilerle önemli çıkarımların yapılabilmesini sağlamaktadır. Sosyal medya üzerinden herhangi bir bilgiye veya paylaşımına verilen tepkinin dozu da kişiler hakkında detaylı bilgilere rahatlıkla ulaşılabilmesini ve karar verilebilmesini sağlayabilmektedir. İnsanların görüşlerini, değerlendirmelerini, tutumlarını ve duygularını yazdığı dilden analiz eden, görüntü incelemesi olarak da nitelendirilen çalışma alanı duygu analizi olarak tanımlanmaktadır. Önemli toplumsal olaylarda da kişilerin görüşleri önem arz etmektedir. 2020 Ocak ayından itibaren Dünyayı etkileyen Covid-19 sürecinde de kişilerin duygu durumları değişiklik göstermektedir. Bu çalışmada, Korona virüs sürecinde alınan önlemlerin ve hizmete sunulan uygulamaların duygu analizi yöntemiyle insanlar üzerinde bıraktığı etki ölçülmeye çalışılmıştır. 11 Mart 2020 tarihinden bu yana Türkiye’de kişilerin ve resmi kuruluşların önemli duyurularını metin veya video olarak yayınladığı Twitter sosyal medya mecrası aracılığı ile korona virüs için konuşulan önemli başlıklar çıkarılmıştır. Duygu analizi yöntemine göre 5 başlık altında toplanan konular için olumlu ve olumsuz yorumlar sınıflandırılarak genel bakış çıkarılmış, sonrasında ise haftalık süreçte bu konular hakkında gözle görülebilir bir değişiklik olup olmadığı analiz edilmiştir. Twitter verileri üzerinde olumlu ve olumsuz olarak gruplanan yaklaşımlar için yaklaşık 2.000.000 tweet üzerinde literatürde kabul gören ve hızlı sonuç alınabilen lojistik regresyon analizi yöntemi kullanılmıştır.

**Anahtar Kelimeler:** Duygu Analizi, Covid-19, Makine Öğrenmesi, Regresyon Analizi

#### ABSTRACT

Sharing changes in people's daily lives through social media channels ensures that valuable information is accumulated and important inferences can be made with this information. The dose of the reaction to any information or sharing via social media, ensure to easily access and decide detailed information about people. Analyzing people's views, evaluations, attitudes, and emotions from the language in which they write, the workspace, which is also described as image analysis, is defined as sentiment analysis. Opinions of people are also important in important social events. The emotional states of individuals also vary during the Covid-19 process, which has influenced the whole world since January 2020. In this study, it was tried to measure the effects of applications on the people with emotion analysis method in the corona virus process. Since 11 March 2020, highlights spoken to the corona virus in Turkey, which publishes announcements of people and official institutions as text or video via the Twitter social media channels have been identified. According to the sentiment analysis method, an overview was made by classifying the positive and negative comments for the topics collected under 5 headings, and then, whether there was a visible change in the weekly period regarding these issues was analyzed. For the opinions grouped as positive and negative on Twitter data, logistic regression analysis method, which is accepted in the literature and which can be obtained fast results, has been used on approximately 2,000,000 tweets.

**Keywords:** Sentiment Analysis, Covid-19, Machine Learning, Regression Analysis

## 1. GİRİŞ

İnsanların bilgiye eskisinden daha hızlı ulaşma çabası beraberinde teknolojik gelişmelerin de aynı hızda ilerlemesini sağlamıştır. Teknolojik gelişmelerin artması da daha çeşitli bilgilerin elde edilmesini sağlamaktadır. Yeni ve farklı olan her dönemde dikkat çekmiştir. Son dönemlerde de internetin neredeyse her ortamda kullanılabilir olması, işe yarar bilgileri arttırmaktadır. İnsanların gündelik yaşamlarındaki değişikliklerin sosyal medya mecraları üzerinden paylaşılması da geride değerli bilgilerin birikmesine ve bu bilgilerle önemli çıkarımların yapılabilmesini sağlamaktadır. Sosyal medya üzerinden herhangi bir bilgiye veya paylaşıma verilen tepkinin dozu da kişiler hakkında detaylı bilgilere rahatlıkla ulaşılabilmesini ve karar verilebilmesini sağlayabilmektedir. İnsanların görüşlerini, değerlendirmelerini, tutumlarını ve duygularını yazdığı dilden analiz eden, görüntü incelemesi olarak da nitelendirilen çalışma alanı duygu analizi olarak tanımlanmaktadır. Duygu analizi ile kişilerin durum hakkındaki fikirleri, onlarla fikir alışverişi yapmadan da olumlu, olumsuz ve tarafsız olarak kategorize edilebilmektedir. Sosyal medya son yıllarda, sadece bir iletişim aracı olmak yerine, belirli bir ürün veya konu hakkında, insanların görüşlerini paylaştığı önemli bir bilgi kaynağı haline gelmiştir (Onan ve Korukoğlu, 2016). Sosyal medya platformları sayesinde konuşulan, uygulamaya konan, karar alınan birçok konu hakkında insanlar, saygı sınırları çerçevesinde dilediğince yorum yapabilmekte ve geride anlamlandırılacak büyük bir veri bırakmaktadırlar. Müşteri profilini daha iyi analiz ederek, onlara bireyselleştirilmiş hizmet sunabilmek için şirketlerin müşterileriyle ilgili çok sayıda bireysel bilgiyi saklaması gerekmektedir. Sağlık, hükümet kaynakları, üretim, bankacılık, sosyal medya / duygu analizi, Telekom, e-ticaret, medya, eğitim ve perakende satış gibi birçok alanda kullanılan büyük veri bu anlamda kolaylık sağlamaktadır. Birçok alan için önem kazanan Büyük Veri'nin analizi için özel şirketler ve kamu kuruluşları tarafından ciddi yatırımlar yapılmakta, yeni teknik ve yazılımlar geliştirilmektedir (Özdeş, 2017).

Salgınlar dünya tarihinde global etkileri olan sağlık olaylarıdır. Günümüze kadar farklı türlerde Dünyayı etkisi altına alan birçok salgın yaşanmıştır. Veba, kolera, tifüs, çiçek, ebola ve grip bilinen salgınlardan birkaçıdır. Son yıllara baktığımızda Corona virüs ailesi farklı zamanlarda mutasyona uğrayarak insanları etkilemiştir. En bilinenleri Sars-COV, Mers-Cov ve Sars-Cov 2 olarak bilinen Covid-19 dur. 7 Ocak 2020 tarihinde tespit edilen bu virüs Dünyada büyük bir yayılım göstermektedir. Çin de görülen ilk vaka sonrasında tüm Dünya ülkeleri giriş-çıkışları sınırlandırma, yakın temas kurmama gibi birçok önlemi almıştır. Fakat yayılım beklentinin aksine daha hızlı olmuş ve ölümlü vaka sayıları her geçen gün artmaktadır. Türkiye, Covid-19 sürecinde önlemleri sıkı bir şekilde alan nadir ülkeler arasına girmiştir. Türkiye'de, kaynağın bulunması, hastalığın bildirilmesi, kesin tanı, hastaların tedavisi, izolasyon, taşıyıcı araması gibi bulaşma yoluna yönelik ve sağlam kişiye yönelik önlemler alınmıştır ve alınmaya devam etmektedir.

(Özdeş, 2017) yapmış olduğu tez çalışmasında İngilizce şarkı sözleri üzerinde denetimli öğrenme metoduyla duygu analizi işlemi gerçekleştirilmiştir. Dönüştürülen şarkı sözleri, Naive Bayes, Logistic Regresyon ve Decision Tree algoritmalarına tabi tutularak, algoritmaların çalıştırılması sonucu elde edilen başarı oranları karşılaştırılmıştır. Veri, RStudio ile işlenmiş ve algoritmanın çalışması için geçen süre, Spark üzerinde geçen süreyle karşılaştırılmıştır. Spark'ın bu karşılaştırma sonucunda çok daha hızlı olduğu görülmüştür. (Powar, Kadam ve Salvi, 2018) çalışmalarında geçmiş tweet verilerine dayanarak, yeni tweetleri, kötü ve normal olarak kategorize edebilecek bir yapı geliştirmişlerdir. Verilerin durumuna göre lojistik regresyon ve naive bayes yöntemleri kullanılmıştır. Elde edilen sonuçlara göre yöntemlerin her ikisinde de sınıflandırmaların başarılı olduğu gözlemlenmiştir. (Das ve Chakraborty, 2018) çalışmalarında amazon ürün yorumları, IMDB film yorumları ve SMS spam veri setleri üzerinde duygu analizi gerçekleştirmişlerdir. Tf-Idf ile birlikte sonraki kelime negatifliği kullanarak duygu analizi sınıflandırması için yeni bir yöntem önermişlerdir. Kelime çantası modelinde 2-gram, Tf-Idf ve Tf-Idf ile birlikte sonraki kelime negatifliği ni karşılaştırmışlardır. LSVM (Lineer Destek Vektör Makinesi) ni önerilen modelle uygulanarak sonraki kelime negatifliği ve Tf-idf ikilisinin sınıflandırma sonucunda daha iyi sonuç verdiğini göstermişlerdir. (Türkmenoğlu, 2015) tez çalışmasında Türkçe metinlerde duygu analizini

çalışmıştır. Sözlük tabanlı duygu analizi ve makine öğrenmesi yöntemleri Türkçe metinlerde denenmiştir. Türkçenin yapısal özellikleri değerlendirilerek ön işlemler uygulanmıştır. Çalışma, twitter ve film yorumları veri kümesi üzerinde gerçekleştirilmiştir. Twitter mesajlarında sözlük tabanlı yöntem ile %75.2, makine öğrenmesi tabanlı duygu analizi metodu ile ise, Karar Destek Makineleri sınıflandırıcısı kullanılarak, %85 başarı elde edilmiştir. Film yorumları veri kümesine uygulandığında ise sözlük tabanlı duygu analizi metodu ile %79,5, MÖ tabanlı duygu analizi metodu ile KDM sınıflandırıcısı kullanılarak %89 başarı elde edilmiştir. (Akçayol ve Özyurt, 2018) çalışmalarında fikir madenciliği ve duygu analizi konusu detaylarıyla, uygulanan yöntemlerle birlikte anlatılmış, bu alanda yapılmış olan çalışmalar incelenmiş ve literatür taraması şeklinde sunulmuştur. Bu çalışma alanında yerine getirilmesi gereken duygu polaritesinin tespit edilmesi, özellik çıkarımı görevleri için günümüzde yaygın olarak kullanılan yöntemler tanıtılmıştır. Bu yöntemlerin birbirlerine karşı avantaj ve dezavantajları analiz edilmiştir. Literatürde fikir madenciliği ve duygu analizi ile ilgili 2010 yılından günümüze kadar yapılmış çalışmalar incelenmiş ve bunların arasından seçilmiş çok sayıda çalışma, Doküman Düzeyindeki Duygu Analizi Çalışmaları ve Özellik Tabanlı Düzeydeki Çalışmaları başlıkları altında tek tek özetlenmiştir. Ayrıca bu çalışmalar toplu olarak çizelge şeklinde de sunulmuştur. Literatürde farklı yöntemlere göre duygu analiz çalışmaları araştırmak isteyenler için bu literatür taramasının önemli bir kaynak olarak kullanılabileceği değerlendirilmektedir.

Bu çalışmada, Koronavirüs sürecinde alınan önlemlerin ve hizmete sunulan uygulamaların duygu analizi yöntemiyle insanlar üzerinde bıraktığı etki ölçülmeye çalışılmıştır. 11 Mart 2020 tarihinden bu yana Türkiye’de kişilerin ve resmi kuruluşların önemli duyurularını metin veya video olarak yayınladığı Twitter sosyal medya mecrası aracılığı ile korona virüs için konuşulan önemli başlıklar çıkarılmıştır. Bu başlıklara göre Türkiye’de vakanın ilk görüldüğü tarihten bu yana twitter üzerinden neler konuşulmuş, nasıl tepkiler verilmiş ve toplumun bu süreçle ilgili konuşulanlara nasıl yaklaştığı ölçülmüştür. Duygu analizi yöntemine göre 5 başlık altında toplanan konular için olumlu ve olumsuz yorumlar sınıflandırılarak genel bakış çıkarılmış, sonrasında ise haftalık süreçte bu konular hakkında gözle görülebilir bir değişiklik olup olmadığı analiz edilmiştir. Türkçe yorumlar üzerinde analiz gerçekleştirildiği için genel doğal dil işleme yöntemlerinin başarısız sonuç vermesi nedeniyle, makine öğrenmesi algoritmalarıyla sınıflandırma yapılarak sonuç elde edilmeye çalışılmıştır. Twitter verileri üzerinde olumlu ve olumsuz olarak gruplanan yaklaşımlar için yaklaşık 2.000.000 tweet üzerinde literatürde kabul gören ve hızlı sonuç alınabilen lojistik regresyon analizi yöntemi kullanılmıştır. Çalışmanın materyal ve metot bölümünde, Türkiye’de korona virüs ile mücadele, duygu analizi yöntemleri, makine öğrenmesi yöntemiyle veri analizi ve lojistik regresyon analizi, bulgular bölümünde ise sistemin genel yapısı, tweetler ile duygu analizi aşamaları, son bölümde ise elde edilen sonuçlar hakkında bilgiler verilmiştir.

## **2. MATERYAL VE METOT**

### **2.1. Covid-19**

7 Ocak 2020’ de tanımlanan Covid-19, SARS-COV ve MERS-COV gibi ciddi solunum yetmezliği oluşturan bir virüsdür. 31 Aralık 2019 tarihinde bir olgunun bugüne kadar hiç görülmemiş yeni bir korona virüs ile enfekte olduğu saptanmıştır (Türkiye Bilimler Akademisi, 2020). Günümüzde hastalık hızla insandan insana bulaşarak, tüm dünyayı etkisi altına almıştır ve Dünya Sağlık Örgütü(WHO) virüsü 17 Mart 2020 tarihinden itibaren pandemi ilan etmiştir.

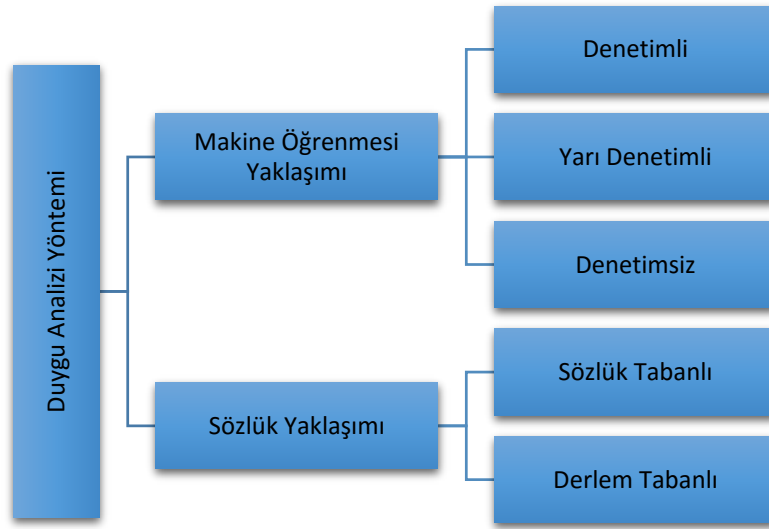
Avrupa ülkelerinin salgın durumu önemsememesi sonucunda virüs hızla yayılmış ve ölümler artmıştır, Mart ayında WHO tarafından virüsün yeni merkezi Avrupa gösterilmiştir. Dünya genelinde korona virüsüne karşı devletler sağlık, ekonomi, eğitim alanlarında bir takım tedbirler almışlardır. 2020 Mayıs ayı itibari ile dünya genelinde hasta sayısı 4 milyonu geçmiştir ve ölüm sayısı 300 binlere ulaşmıştır. Korona virüsünün (Covid-19’un) en ideal tanı yöntemi PCR yöntemidir. Türkiye, korona virüs ile ilgili dünyada en erken önlem alan ülkelerden birisi olmuştur. Türkiye’ye virüsün bulaşmasını engellemek için şubat ayı başından itibaren gelen yolcuların kontrolden geçirilmesi,

uçuşların durdurulması gibi bir takım tedbirler alınmıştır. 11 Mart tarihinde ilk vakanın görülmesi ile birlikte tedbirler sıkılaştırılmıştır. İlk önce toplu alanlara kısıtlamalar getirildi, ardından 16 Mart'ta tüm eğitim öğretime ara verildi, umreden ve yurt dışından gelenler 14 gün öğrenci yurtlarında karantinaya alındı ve 16 ülkeye uçuşlar durduruldu. Sağlık bakanı tarafından “evde kal” çağrısı yapıldı. 20-22 Mart arasında pandemi hastaneleri kuruldu, 65 yaş üzerinin sokağa çıkması yasaklandı ve 71 ülkeye daha uçuşlar durduruldu. Ülkede enfekte olan ve ölen sayısının sürekli artması üzerine daha sıkı tedbirler alınması yoluna gidildi. Bunlar, şehirlerarası yolculuklar sınırlandırılması, 20 yaş atına sokağa çıkma yasağı, 31 şehrin giriş çıkışlarının kapatılması, maske takma zorunluluğu ve hafta sonları sokağa çıkmanın yasaklanması gibi daha ağır tedbirlerdir. Bunların yanında halkın maddi olarak zarar görmemesi için bir takım önlemler alınmıştır. Bunların başında fiyatları aşırı zamlanan ve ulaşmanın güçleştiği maske satışının durdurulması ve devlet tarafından ücretsiz dağıtılması gelmektedir (Şenol ve Bilsel, 2020).

## **2.2. Duygu Analizi**

Fikir Madenciliği olarak da bilinen Duygu Analizi (Sentiment Analysis), bir varlık üzerinde insanların tutum, düşünce ve duygularının bilgisayar bilimleri kullanılarak ortaya çıkarılmasını amaçlayan bir araştırma alanıdır (Medhat, Hassan, ve Korashy, 2014). Psikoloji bilim dalı duygu analizini çokça işlemiş ve kişinin duygusal durumunun, kullandığı kelimeler ve bu kelimeleri kullanma şekilleriyle çok yakından ilgili olduğu tespit edilmiştir (Pennebaker, Mehl ve Niederhoffer, 2003). Toplumsal tepkilerin ve ayaklanmaların yoğun olduğu dönemlerde genelde ya da özelde belli konularda olumsuz bir dil kullanılır. Bu tepkiyi doğuran kişi veya kurum hakkındaki anket çalışması ise toplumun o konu hakkında duygusal durumunu gösterebilmektedir (Türkmenoğlu, 2015). Medya takibi yapan kişiler ve kurumlar, duygu analizine metinleri olumlu, olumsuz veya tarafsız olarak sınıflandırarak verimli ve verimsiz veriyi birbirinden ayırmak için ihtiyaç duymaktadırlar. Duygu analizi bir sınıflandırma işlemi olarak da düşünülebilir. Doküman seviyesinde, cümle seviyesinde, özellik temelli olmak üzere üç aşamada değerlendirilebilmektedir. Doküman seviyesinde sınıflandırma, bir konu üzerinde yazılmış olan dokümanı ele alır. Tek bir doküman ele alınarak, dokümanın olumlu ya da olumsuz olduğu tespit edilir. Cümle seviyesinde sınıflandırma, her cümlenin duygu analizini gerçekleştirmeyi amaçlar. İlk adım cümlenin objektif ya da sübjektif olup olmadığını belirlemektir. Eğer cümle sübjektif ise cümlenin olumlu veya olumsuz olduğu belirlenir. Özellik temelli duygu analizinde, diğer sınıflandırma seviyelerine göre yaklaşım biraz daha farklıdır. Özellik temelli (Aspect-level) duygu analizi varlığın tüm yönleriyle ele alınmasını amaçlar. Bir nesnenin belli özelliklerine göre sınıflandırma yapılır. Bir nesnenin hangi bakış açısına göre olumlu ya da olumsuz olduğu önemlidir.

Duygu analizi için makine öğrenmesine dayalı yöntemler ve sözlük tabanlı yöntemler kullanılmaktadır. Makine öğrenmesi algoritmalarının farklı alanlara kısmen kolay uygulanabilir olmasıyla birlikte istatistiksel yaklaşımların görece daha hızlı çalışması canlı veri üzerinde gerçek zamanlı olarak yapılan çalışmalarda tercih sebebi olmaktadır (Şeker, 2015). Sözlük tabanlı yaklaşımda doğal dil işleme yöntem ve araçları kullanılarak cümlelerin sentaktik analizine dayalı yöntemler kullanılmaktadır. Denetimli makine öğrenmesi yöntemlerindeki gibi etiketlenmiş eğitim verisine ihtiyaç yoktur. Doğal dil işleme araç ve yöntemleri ile cümleler analiz edilir, cümlelerdeki duygu terimleri tespit edilerek anlamsal çıkarımlar yapılır. Cümlelerdeki duygu ifadelerini tespit etmek için çoğunlukla duygu terimleri sözlüğü kullanılır. Sözlük tabanlı yöntemde, eş anlamlı ve zıt anlamlı kelimeler kullanılarak metnin görüş kutbu belirlenir. Derlem tabanlı yöntemde ise görüş kutbu belirlenirken istatistiksel ya da semantik yöntemler kullanılır. Şekil-1 de Duygu Analiz Yöntemlerine ait bir yapı gösterilmektedir.



Şekil 1. Duygu Analiz Yöntemleri

### 2.3. Makine Öğrenmesi ile Veri Analizi

Makine öğrenmesi ile veri analizi, verilerden sonuç çıkarmak ve karar vermeyi desteklemek amacıyla verileri incelemek, temizlemek, dönüştürmek ve modellemek için bir yöntemdir. Veri analizi, farklı iş, bilim ve sosyal bilim alanlarında çeşitli isimler altında çeşitli teknikleri kapsayan çok yönlü ve farklı yaklaşımlara sahiptir (Xia ve Gong, 2015). Veri analizinde önce veriler toplanır sonra da soruları cevaplamak, hipotezleri test etmek veya teorileri reddetmek için analiz gerçekleştirilir. Veri analizinin aşamaları ise; Veri Toplama, Veri İşleme, Veri Temizleme, Modelleme ve Algoritmalar, Uygulama ve Testler olarak belirlenmiştir. Gözetimli öğrenme, gözetimsiz öğrenme ve takviyeli öğrenme bilinen makine öğrenmesi yaklaşımlarıdır.

Gözetimli Öğrenme tekniğinde giriş değerleri ile istenen çıkış değerleri arasında eşleme yapan bir fonksiyon oluşturulur. Gözetimsiz öğrenme yönteminde işaretlenmemiş veri üzerinden bilinmeyen bir yapıyı tahmin etmek için bir algoritma kullanan makine öğrenmesi tekniğidir. Takviyeli öğrenme ise amaç odaklı bir yöntem olduğu için diğer iki öğrenme yöntemine göre biraz farklılıklar içermektedir. Her üç yaklaşım için en yaygın kullanılan makine öğrenmesi yöntemleri ise Naive Bayes Sınıflandırıcı Algoritması (Denetimli Öğrenme- Sınıflandırma), K-Means Algoritması (Denetimsiz Öğrenme- Kümeleme), Destek Vektör Makinesi Algoritması (Denetimli Öğrenme-Sınıflandırma), Doğrusal Regresyon (Denetimli Öğrenme - Regresyon), Lojistik Regresyon (Denetimli öğrenme – Sınıflandırma), Yapay Sinir Ağları (Takviye Öğrenimi) olarak bilinmektedir.

### 2.4. Lojistik Regresyon Analizi

Regresyon analizi, herhangi bir değişkenin bir veya birden fazla değişkenle arasındaki ilişkinin matematik bir fonksiyon şeklinde yazılmasıdır. Lojistik regresyon da bağımlı değişkenin iki sonucu olan, bağımsız değişkenlerin ise kategorik veya sürekli olduğu çoklu regresyondur. (Bayrak, 2013) Lojistik regresyonda, bağımlı değişken ikili yani yalnızca 1 (doğru, başarılı, pozitif vb.) veya 0 (yanlış, başarısız, negatif vb.) olarak kodlanmış verileri içeriyor. Lojistik regresyonun amacı, iki yönlü karakteristiği ile ilgili bir dizi bağımsız değişken arasındaki ilişkiyi tanımlamak için en uygun modeli bulmaktır. (Veri Bilimcisi, 2020) Aşağıdaki denklemlerde lojistik regresyon analizine ait formüller verilmiştir. Denklemlerdeki P, karakteristik özelliğinin var olma olasılığıdır.

$$\text{Logit}(p) = \log \log \left[ \frac{p}{1-p} \right] = \ln \left[ \frac{p}{1-p} \right] \quad (1)$$

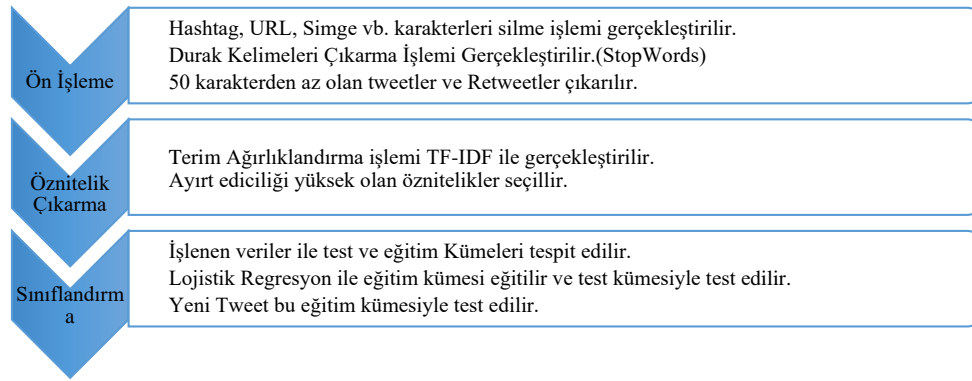
$$\text{Logit}[p(x)] = \log \log \left[ \frac{p(x)}{1-p(x)} \right] = a + b_1 * x_1 + b_2 * x_2 + \dots + b_i * x_i \quad (2)$$

$$P = e^{\log \log \left[ \frac{p(x)}{1-p(x)} \right] = a + b_1 * x_1 + b_2 * x_2 + \dots + b_i * x_i} / (1 + e^{\log \log \left[ \frac{p(x)}{1-p(x)} \right] = a + b_1 * x_1 + b_2 * x_2 + \dots + b_i * x_i}) \quad (3)$$

Regresyon formülüne baktığımızda olasılık hesaplarının modelin temelini oluşturduğu görülebilmektedir. Model, bir olayın gerçekleşme olasılığı ve gerçekleşmeme olasılığının birbirine bölünmesinin doğal logaritmasının alınması ile kurulmaktadır (Bayrak, 2013).

## 2.5. Geliştirilen Yöntem

Korona virüsüyle mücadele sürecinde twitter aracılığıyla yazılan yorumların analizinin yapıldığı sistemin modeli şekil-2 de verilmiştir. Twitterdan gelen veriler ön işlemeden geçirilir. İkinci aşamada kelime torbası yöntemiyle öznitelikler çıkarılır. Son aşamada öznitelikleri belirlenen kelimeler sınıflandırılarak olumlu ve olumsuz tweetler tespit edilir.



Şekil 2. Sistem Modeli

### Veri Seti

Korona virüsü (Covid-19) sürecinde twitter üzerindeki Türkçe verilerin barındığı toplu bir veri seti olmadığı için Twitter API kullanılarak 11 Mart 2020 tarihinden bu yana 5 başlık altında yaklaşık 2.000.000 adet Türkçe tweet toplanmıştır. Her kategori altındaki tweetler oluşturulan veritabanına kaydedilmiştir.

### Ön İşleme

Twitter mesajları içerisinde hashtag, http, simge gibi birçok anlam içermeyen karaktere sahiptir. Eğitim setinin doğru sınıflandırma işlemini tamamlayabilmesi için bu karakterler veri setinden çıkarılmıştır. Ayrıca retweet ve tekrar eden tweetler de silinmiştir. Türkçede sık geçen durak kelime havuzu oluşturularak (bazen, bazı vb.) değersiz kelimeler ayıklanmıştır. Son olarak da 280 karakter sınırı olan mesajlardan 50 karakterden az olanları metin analizinde anlamlı bir sonuç üretmeyeceği yapılan testlerde ortaya çıktıktan sonra veri setinden elenmiştir.

### Terim Ağırlıklandırma- Öz Nitelik Seçme

Terim ağırlıklandırma, sınıflandırma algoritmalarının ayırt edici gücünü daha belirgin şekilde ortaya çıkarmayı amaçlamaktadır. Bu çalışmada metin sınıflandırmada çok sık kullanılan TF-IDF terim ağırlıklandırma yöntemi kullanılmıştır. TF ilgili terimin dokümanda geçiş adedi iken, IDF ise denklem 4'de verilmiştir.

$$IDF = \log \frac{N}{a+b} \quad (4)$$

N, toplam doküman sayısını, a ilgili terimin pozitif kategorilerde geçtiği dokümanların sayısını, b ise ilgili terimin negatif kategorilerde geçtiği dokümanların sayısını ifade etmektedir. Terimin ağırlığı, TF ile IDF değerlerini çarparak bulunabilmektedir.

Klasik metin sınıflandırma yaklaşımlarında kelime çantası modeli tercih edilmektedir. Eğitim seti içerisinde yer alan bütün kelimeler kullanılarak doküman vektörleri tespit edilir. Öznitelik seçimi, modelde yer alan tüm terimleri kullanmak yerine, kelimeleri en iyi temsil eden terimlerin seçilerek, daha küçük boyutlu vektörlerin elde edilmesi işlemidir. (Şahin ve Kılıç, 2016). Bu çalışmada da kelime çantası modeli seçilmiştir.

### **Sınıflandırma**

Sınıflandırma işlemi, kategorisi bilinmeyen örneklerin, eğitim verisi kategorilerinden en uygun olan kategoriye atanmasıyla yapılmaktadır. Bu çalışmada doküman sınıflandırma çalışmalarında yaygın olarak kullanılan Lojistik Regresyon Analizi kullanılmıştır. Metinleri olumlu ve olumsuz olarak kategorize etmek ve sınıflandırma hızını arttırmak amacıyla bu analiz yöntemi seçilmiştir.

### **Performans Ölçeği**

Performans ölçeği, belgenin ilgili sınıfa ait olma doğruluğunu ölçmektedir. Veri setinde pozitif etiketli olan bir örnek, sınıflandırma sonucunda pozitif olarak sınıflandırılıyorsa True Positive (TP) olarak adlandırılır. Gerçekte negatif etiketli olan bir örnek, sınıflandırma sonucunda negatif olarak sınıflandırılıyorsa True Negative (TN) olarak adlandırılır. Gerçekte negatif olan bir örnek pozitif olarak sınıflandırılıyorsa False Positive (FP) ve gerçekte pozitif olan bir örnek negatif olarak sınıflandırılıyorsa False Negative (FN) olarak adlandırılır.

Doğruluk Oranı, sınıflayıcının ne sıklıkta doğru tahmin ettiğinin bir ölçüsüdür. Precision ise tüm sınıflardan, doğru olarak ne kadar tahmin edildiği tespit eder. Yaygınlık, tahminleme sonunda ne sıklıkta 1 değerinin bulunduğu yansıtır. F Puanı ise, gerçek pozitif değerlerin oranının (recall) ve hassasiyetin (precision) harmonik ortalamasıdır. Sınıflandırıcının ne kadar iyi performans gösterdiğinin bir ölçüsüdür ve sınıflandırıcıları karşılaştırmakta sıklıkla kullanılır. ROC eğrisi, sınıflandırıcının tüm olası değerler üzerinde performansını özetlemek için kullanılan bir grafikdir. Belirli bir sınıfa gözlem atanması eşiğini değiştirdiğinizde gerçek pozitif değerlerin oranına (Hassasiyet) (x eksen) karşı Yanlış Pozitif Değerlerin Oranını (Özgüllük) (y eksen) çizerek oluşturulur. ROC Eğrisi, Hassasiyet / Özgüllük (Sensitivity / Specificity) raporu oluşturmaya yarar. ROC eğrisinin altındaki alan (Area Under Curve (AUC)), bir parametrenin iki sınıf arasında ne kadar iyi ayırt edilebileceğinin bir ölçüsüdür.

Bu çalışmada temel alınan başarı ölçütü ise AUC değeri olmakla birlikte diğer başarımların bilgileri de bulgular bölümünde paylaşılmıştır.

## **3. BULGULAR**

Türkçe veri seti ön işleme aşamalarından geçirilmeden önce 1.657.173 adet, sonrasında ise 1.419.480 adete düşmüştür. Tablo 1’de anahtar kelimelerin ön işlem öncesi ve sonrasına ait rakamlar verilmiştir.

**Tablo 1.** Anahtar Kelime Sayıları

| Anahtar Kelimeler    | Ön İşlem Öncesi | Ön İşlem Sonrası |
|----------------------|-----------------|------------------|
| Maske                | 900.839         | 757.698          |
| Eba                  | 72.737          | 57.208           |
| Sokağa Çıkma Yasağı  | 662.261         | 584.921          |
| Kısa Çalışma Ödeneği | 13.876          | 12.488           |
| Devlet Desteği       | 7.460           | 7.165            |

Çalışma için python programlama dilinde pandas, numpy ve sklearn kütüphaneleri kullanılmıştır. Verileri anlamlandırmak ve yapısal olarak saklamak için Ms Sql Server da veritabanı geliştirilmiştir. Twitter API aracılığıyla çekilen tweetler için, veritabanına kaydedildikten sonra sırasıyla ön işleme,



terim ağırlıklandırma, kelime çantası (BOW) ile öz nitelik oluşturma aşamalarından geçildikten sonra sınıflandırma aşamasında eğitim ve test kümeleri belirlenmiştir. Nu aşamada 2 farklı yöntem denenmiştir. Sınıflandırma başarımlarına göre uygun model seçilmiştir.

1. yöntemde eğitim kümeleri her bir anahtar kelime için rastgele belirlenmiştir. Her bir anahtar kelimeye ait veri seti çeşitli denemeler sonucunda %25 ile % 50 oranında eğitim ve test olarak ikiye ayrılmıştır. Eğitim setinde de olumlu olumsuz veriler yarı yarıya rastgele alınmıştır. Bu denemeler sonucunda en iyi AUC değerleri aşağıdaki Tablo 2’de verilmiştir.

**Tablo 2.** Rastgele Oluşturulan Eğitim Setlerine Göre Başarım Oranları

| Anahtar Kelimeler | X_Train | Y_Train | Test  |        |        | AUC Değeri |
|-------------------|---------|---------|-------|--------|--------|------------|
|                   |         |         | Oranı | X_Test | Y_Test |            |
| Maske             | 300000  | 300000  | 0.25  | 100000 | 100000 | 0.7450     |
| Eba               | 28604   | 28604   | 0.5   | 28604  | 28604  | 0.8284     |
| Sokağa Çıkma      |         |         |       |        |        |            |
| Yasağı            | 292461  | 292461  | 0.5   | 292461 | 292461 | 0.8262     |
| Devlet Desteği    | 5231    | 5231    | 0.27  | 1934   | 1934   | 0.65       |
| Kısa Çalışma      |         |         |       |        |        |            |
| Ödeneği           | 9117    | 9117    | 0.27  | 3371   | 3371   | 0.8034     |

Başarımların istenen seviyede olmamasından dolayı eğitim kümeleri olumlu ve olumsuz kelime gruplarına göre belirlenmiştir. Buna göre olumlu tweetlerde aranan kelimeler; mükemmel, idare eder, aferin, doğru, olumsuz tweetlerde aranan kelimeler ise kötü, berbat, vefat, ölüm, fena, yanlış, olumsuz olarak belirlenmiştir. Kurallı eğitim kümelerine göre AUC değerleri Tablo 3’de verilmiştir.

**Tablo 3.** Kurallı Oluşturulan Eğitim Setlerine Göre Başarım Oranları

| Anahtar Kelime | X_Train | Y_Train | Test  |        |        | AUC Değeri         |
|----------------|---------|---------|-------|--------|--------|--------------------|
|                |         |         | Oranı | X_Test | Y_Test |                    |
| Maske          | 26376   | 26376   | 0.4   | 17583  | 17583  | 0.9728985876374309 |
| Eba            | 1916    | 1916    | 0.4   | 1277   | 1277   | 0.9473535067038976 |
| Sokağa Çıkma   |         |         |       |        |        |                    |
| Yasağı         | 26928   | 26928   | 0.4   | 17951  | 17951  | 0.9813433913127728 |
| Devlet Desteği | 251     | 251     | 0.4   | 166    | 166    | 0.8623946527172334 |
| Kısa Çalışma   |         |         |       |        |        |                    |
| Ödeneği        | 303     | 303     | 0.4   | 201    | 201    | 0.9164682539682539 |

Kurallı eğitim setlerine göre her bir anahtar kelime için performans ölçütleri aşağıda verilmiştir.



|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.98      | 0.96   | 0.97     | 7590    |
| 1            | 0.97      | 0.99   | 0.98     | 9994    |
| accuracy     |           |        | 0.97     | 17584   |
| macro avg    | 0.98      | 0.97   | 0.97     | 17584   |
| weighted avg | 0.97      | 0.97   | 0.97     | 17584   |

AUC: 0.9728985876374309

**Şekil 3.** Maske Performans Ölçekleri

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.91      | 0.95   | 0.93     | 491     |
| 1            | 0.97      | 0.94   | 0.96     | 787     |
| accuracy     |           |        | 0.95     | 1278    |
| macro avg    | 0.94      | 0.95   | 0.94     | 1278    |
| weighted avg | 0.95      | 0.95   | 0.95     | 1278    |

AUC: 0.9473535067038976

**Şekil 4.** Eba Performans Ölçekleri

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.99      | 0.98   | 0.99     | 11425   |
| 1            | 0.97      | 0.98   | 0.97     | 6527    |
| accuracy     |           |        | 0.98     | 17952   |
| macro avg    | 0.98      | 0.98   | 0.98     | 17952   |
| weighted avg | 0.98      | 0.98   | 0.98     | 17952   |

AUC: 0.9813433913127728

**Şekil 5.** Sokağa Çıkma Yasağı Performans Ölçekleri

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.86      | 0.91   | 0.89     | 93      |
| 1            | 0.88      | 0.81   | 0.85     | 74      |
| accuracy     |           |        | 0.87     | 167     |
| macro avg    | 0.87      | 0.86   | 0.87     | 167     |
| weighted avg | 0.87      | 0.87   | 0.87     | 167     |

AUC: 0.8623946527172334

**Şekil 6.** Devlet Desteği Performans Ölçekleri

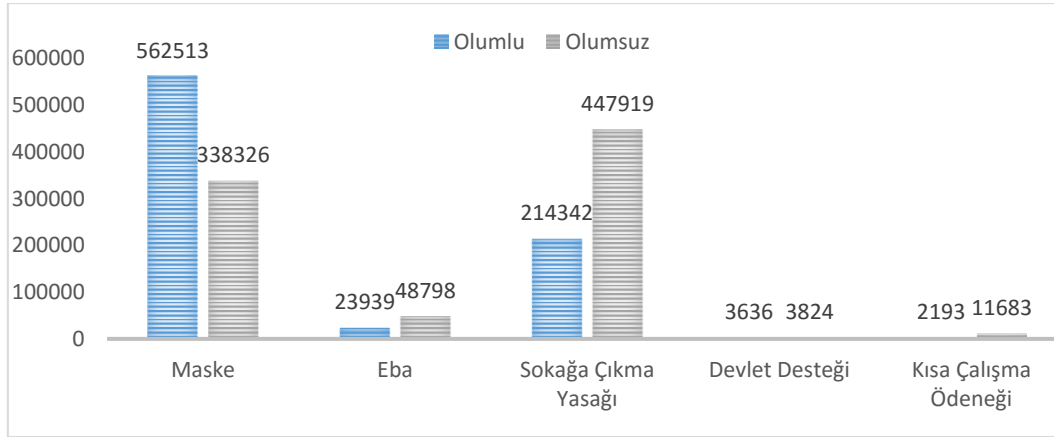
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.89      | 0.92   | 0.91     | 90      |
| 1            | 0.94      | 0.91   | 0.92     | 112     |
| accuracy     |           |        | 0.92     | 202     |
| macro avg    | 0.91      | 0.92   | 0.92     | 202     |
| weighted avg | 0.92      | 0.92   | 0.92     | 202     |

AUC: 0.9164682539682539

**Şekil 7.** Kısa Çalışma Ödeneği Performans Ölçekleri

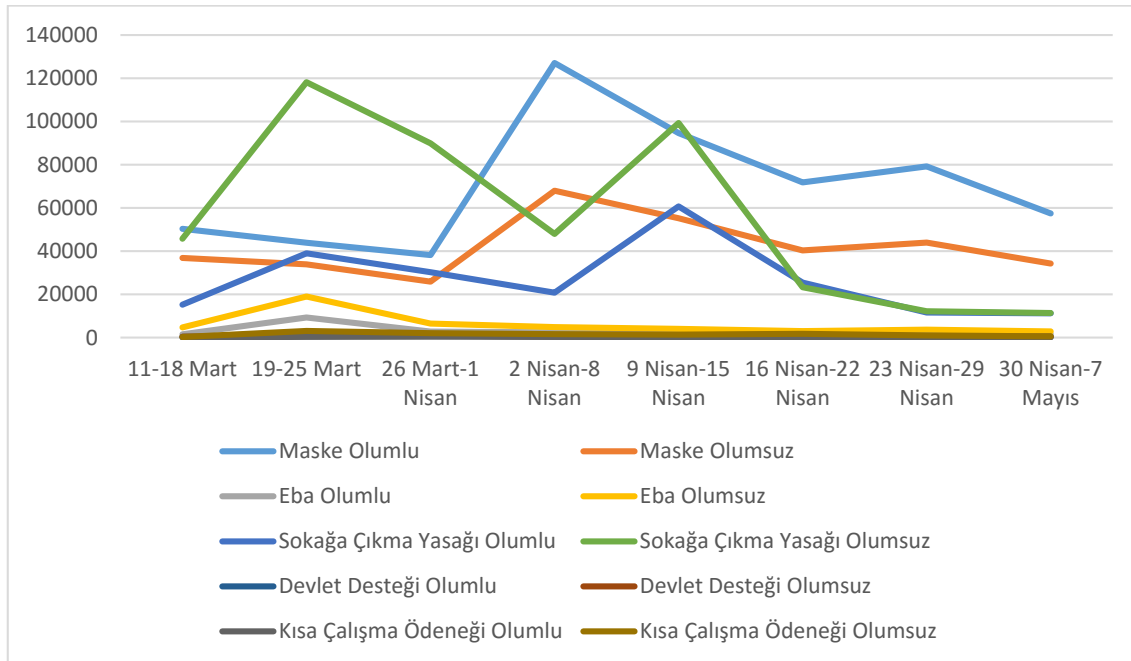
#### 4. SONUÇ VE TARTIŞMA

Anahtar kelimeler üzerine yapılan duygu analizlerinde 2 aylık süreç içerisinde insanların pozitif ve negatif düşünceleri sınıflandırma yöntemiyle tespit edilmiştir. Sosyal medya reaksiyonları, özellikle kurumların almış oldukları kararları daha iyiye götürme sürecinde belirleyici olmaktadır. Bu da alınan kararların süreç içerisinde iyileştirilmesine yol açabilmektedir. Buna göre 5 anahtar kelimedeki genel sınıflandırma sonuçları Şekil 8’de verilmiştir.



Şekil 8. Genel Sınıflandırma Sonuçları

Genel sonuçlar incelendiğinde insanların maske uygulamasının genelde olumlu olarak değerlendirildiği fakat diğer uygulamaların ise genelde olumsuz olarak değerlendirildiği gözlenmektedir. Fakat twitter vb. sosyal medya yorumlarında kişilerin ilk etapta olumsuz reaksiyon gösterdiği uygulamaların sonraki süreçte yapılan iyileştirmelerle olumlu yönde ilerlediği gözlenmiştir. Bu sebeple 11 Mart ile 07 Mayıs 2020 tarihleri arasında hafta hafta pozitif ve negatif yorumların izlenimi için Şekil 9’daki grafik çıkartılmıştır.



Şekil 9. Anahtar Kelimelerin Haftalık Duygu Analiz Sonuçları

Haftalık analizde ise uygulamaların genel olarak ilk hizmete alınma sürecinde olumsuz reaksiyon gösterdiği fakat süreç içerisinde yapılan iyileştirmeler ve hizmete alındıktan sonraki kullanıcı deneyimleri pozitif yorumların negatifleri yakalamasını sağlamıştır. Örneğin Eba anahtar kelimesi incelendiğinde ilk duyurulma aşamasındaki negatif düşünceler okulların uzaktan eğitime başlama sürecinden sonra düşüşe geçmiştir. Diğer anahtar kelimelerde de benzer durumlar gözlemlenmektedir. Haftalık grafiğe bakıldığında aynı zamanda toplumun uzun soluklu bu süreçte zaman içerisinde aynı konularda tweet atma oranlarının azaldığı da anlaşılmaktadır.

Çalışma sürecinde literatürde sıkça geçen genel dil işleme kütüphaneleriyle gerçek sonuçlar alınamamıştır. Çalışılan dilin Türkçe olması genel doğal dil işleme yöntemlerini başarısız kılmıştır. Sıkça kullanılan Textblob Kütüphanesi de bu çalışmada başarıyı yakalayamamıştır. Türkçe analiz için literatürde kabul gören makine öğrenmesi yöntemi olan sınıflandırma kullanılmıştır. Sınıflandırma, metin işleme ve büyük veri analizinde daha hızlı ve tutarlı yanıt vermektedir.

Yapılan bu çalışma korona virüsü ile mücadele sırasında toplumun en çok konuştuğu konular üzerinde duygu analizi ile konulara yaklaşımları ölçmektedir. Bununla birlikte duygu analizi aşamasında sınıflandırma yöntemi ve sınıflandırma işleminde iki farklı grubu en iyi sınıflandırabilen lojistik regresyon kullanılmıştır. Sınıflandırma başarımları bulgular bölümünde de verilmiştir. İlerleyen çalışmalarda farklı duygu sınıfları için (pozitif-nötr-negatif-vb.) farklı sınıflandırma algoritmaları test edilerek sisteme eklenmesi planlanmaktadır. Atılan tweetlerin ülke, bölge ve il bazında analizi gerçekleştirilerek aynı tarihler arası virüsün pozitif vaka ve ölüm oranları arasındaki ilişkinin tespit edilmesi de düşünülmektedir. Türkçe dil analizinin sözlük tabanlı modeli geliştirilerek Zemberek Kütüphanesi veya geliştirilecek olan derlem tabanlı doğal dil işleme modeli, sisteme eklenerek başarımın daha büyük metin veya paragraflar için analizi yapılabilecektir. Sistem mevcut yapıda kullanıcı arayüzü olmadan çalıştırılmıştır. Sonraki çalışmalarda web veya mobil arayüz geliştirilerek büyük kuruluşlar için tweet analiz aracı olarak kullanılması planlanmaktadır.

## KAYNAKÇA

Bayrak, B. (2013). *Lojistik Regresyon Teori ve SPSS*

*Çözümleri*. [https://www.academia.edu/11479607/Lojistik\\_Regresyon](https://www.academia.edu/11479607/Lojistik_Regresyon).

Das, B. ve Chakraborty, S. *An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation*. 1 Mayıs 2020 tarihinde, <https://arxiv.org/abs/1806.06407v1> adresinden erişildi.

Medhat, W., Hassan, A., ve Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.

Onan, A. ve Korukoğlu, S. (2016). A review of literature on the use of machine learning methods for opinion mining. *Pamukkale University Journal of Engineering Sciences*, 22(2), 111-122.

Özdeş, M. (2017). *Büyük Veri Araçlarını Kullanarak Duygu Analizi Gerçekleştirimi*. (Yüksek Lisans Tezi). Pamukkale Üniversitesi Fen Bilimleri Enstitüsü.

Özyurt, B. Ve Akçayol. M. A. (2018). Fikir Madenciliği Ve Duygu Analizi, Yaklaşımlar, Yöntemler Üzerine Bir Araştırma. *Selçuk Üniversitesi Mühendislik Bilim ve Teknoloji Dergisi*.6-4. 668-693.

Pennebaker, J.W., Mehl, M.R., ve Niederhoffer, K.G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*. 54(1), 547-577.

Powar, S., Kadam, U. ve Salvi, T. (2018). Twitter Sentiment Analysis for Classifying Hate Tweets and Normal Tweets Using Logistic Regression and Naive Bayes Algorithm. *Journal of Emerging Technologies and Innovative Research*.5-10. 371-373.



- Şahin, Ö. D. Ve Kılıç, E. (2016). Tıbbi Dokümanların Ayrıştırılmasında Kullanılan Sınıflandırma Algoritmalarının Karşılaştırılması. *International Conference on Computer Science and Engineering*. Tekirdağ/Türkiye.
- Şeker, S. E. (2015). Sosyal Ağlarda Veri Madenciliği (Data Mining on Social Networks). *YBS Ansiklopedi*. 2(2), 30-39.
- Şenol, C. ve Bilsel, A. (2020). Türkiye'de Covid-19 (Coronavirüs) Pandemisinin Durumu ve Algoritmik Hesaplamaya Göre Gelecekteki Durum Analizi, *Journal Of Social, Humanities and Administrative Sciences*, 6(25):535-546.
- Türkiye Bilimler Akademisi. (2010). *Covid-19 Pandemi Değerlendirme Raporu*. Türkiye Bilimler Akademisi Yayınları, TÜBA Raporları No: 34.
- Türkmenoğlu, C. (2015). *Türkçe Metinlerde Duygu Analizi*. (Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.
- Veri Bilimcisi. (2017). Lojistik Regresyon (Logistic Regression). <https://veribilimcisi.com/2017/07/18/lojistik-regresyon/>.
- Xia, B. S. ve Gong, P. (2015). Review of business intelligence through data analysis. *Benchmarking*, 21(2), 300-311. doi:10.1108/BIJ-08-2012-0050.