

**T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
YÖNETİM BİLİŞİM SİSTEMLERİ PROGRAMI
YÜKSEK LİSANS TEZİ**

**COVID 19 PANDEMİSİ İLE İLGİLİ EGE BÖLGESİ VE
GÜNEYDOĞU ANADOLU BÖLGESİ TWİTTER
VERİLERİNİN METİN MADENCİLİĞİ İLE ANALİZİ**

Medine ORAL

**Danışman
Doç. Dr. Çiğdem TARHAN**

İZMİR – 2022

TEZ ONAY SAYFASI



YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “COVID 19 Pandemisi ile İlgili Ege Bölgesi ve Güneydoğu Anadolu Bölgesi Twitter Verilerinin Metin Madenciliği ile Analizi” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

19/09/2022

Medine ORAL

ÖZET

Yüksek Lisans Tezi

COVID 19 Pandemisi ile İlgili Ege Bölgesi ve Güneydoğu Anadolu Bölgesi Twitter Verilerinin Metin Madenciliği ile Analizi

Medine ORAL

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Yönetim Bilişim Sistemleri Anabilim Dalı

Yönetim Bilişim Sistemleri Programı

Günümüz toplumu bilgi toplumu olarak adlandırılmaktadır. Tüm alanlarda ve sektörlerde bilgiyi keşfetmek ve üretmek önemini her geçen gün artmaktadır. Bilgiyi keşfetme amacına hizmet etmekte olan biri olan veri madenciliğinin değeri günümüzde anlaşılmıştır ve her gün veri madenciliğine verilen önemin arttığı görülmektedir. Bu çalışmada veri madenciliği ve veri madenciliğinin alt dallarından biri olan metin madenciliği üzerine gerçekleştirilmiştir. Sosyal medya ortamları, veri madenciliği ve metin madenciliği çalışmalarına hizmet etmekte ve edilen bilgiler araştırmacıları şaşırtmaktadır. Tez kapsamında veri madenciliği, sosyal medya ve metin madenciliği konularına kavramsal olarak değinilmiştir. Ardından 2020 yılı ile Dünya'yı etkisi altına alan Covid 19 konulu Twitter verileri kullanılarak bir veri madenciliği çalışması gerçekleştirilmiştir. Sanal dünya adı verilen sosyal medya ortamlarında üretilen içeriklerin gerçek yaşamda bireylerin takınmış olduğu davranışlara dair ipuçlarına yer verip vermediği sorusu sorgulanmış, bu amaçla Ege Bölgesi ve Güneydoğu Anadolu Bölgesinden seçilen temsili illerin Covid 19 konulu Tweetleri incelenmiş, çalışmada uygulamaya dair adım ve sonuçlara ilgili bölümler altında yer verilmiştir.

Anahtar Kelimeler: Metin Madenciliği, Sosyal Medya ve Metin madenciliği, Covid 19.

ABSTRACT

Master's Thesis

Analysis of Aegean Region and Southeastern Anatolia Region Twitter Data

Related to COVID 19 Pandemic by Text Mining

Medine Oral

Dokuz Eylül University

Graduate School of Social Sciences

Department of Management Information Systems

Management Information Systems Program

Today's society is called the information society. Discovering and producing knowledge in all fields and sectors is increasing its importance day by day. The value of data mining, which serves the purpose of discovering information, has been understood today and it is seen that the importance given to data mining is increasing every day. This study was carried out on data mining and text mining, which is one of the sub-branches of data mining. Social media environments serve mining and text mining studies, and the information provided surprises researchers. Within the scope of the thesis, data mining, social media and text mining are discussed conceptually. Then, a data mining study was carried out using Twitter data on Covid 19, which affected the world in 2020. The question was asked whether the content produced in the social media environments called virtual world gives clues about the behaviors of individuals in real life. placed under

Keywords: Text Mining, Social Media and Text mining, Covid 19.

**COVID 19 PANDEMİSİ İLE İLGİLİ EGE BÖLGESİ VE
GÜNEYDOĞU ANADOLU BÖLGESİ TWİTTER VERİLERİNİN
METİN MADENCİLİĞİ İLE ANALİZİ**

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
KISALTMALAR	viii
TABLOLAR LİSTESİ	ix
ŞEKİLLER LİSTESİ	x
GİRİŞ	1

**BİRİNCİ BÖLÜM
LİTERATÜR ARAŞTIRMASI**

1.1. LİTERATÜR ARAŞTIRMASI	3
----------------------------	---

**İKİNCİ BÖLÜM
VERİ MADENCİLİĞİ**

2.1. VERİ MADENCİLİĞİ KAVRAMI VE MODELLERİ	8
2.1.1. Sınıflama ve Regresyon	12
2.1.2. Kümeleme	13

2.1.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler	13
2.2. WEB 2.0 VE ARAÇLARI	14
2.2.1. Sosyal Ağlar	16
2.2.2. Twitter	17

ÜÇÜNCÜ BÖLÜM

UYGULAMA

3.1. PROBLEM TANIMI	23
3.2. VERİ MADENCİLİĞİ SÜRECİ	26
3.2.1. Veri Seti ve Özellikleri	26
3.2.2. Veri Ön İşleme	44
3.2.3. Bulgular	57
SONUÇ	68
KAYNAKÇA	71

KISALTMALAR

ABD	Amerika Birleşik Devletleri
API	Uygulama Programlama Arabirimi
ARPA	Gelişmiş Savunma Araştırmaları Projeleri Birimi
ARPANET	Gelişmiş Araştırma Projeleri Dairesi Ağı
BERT	Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri
CNN	Kablolu Haber Ağı
ENIAC	Elektronik Sayısal Entegreli Hesaplayıcı.
E-Posta	Elektronik Posta
E-Ticaret	Elektronik Ticaret
GPU	Grafik İşleme Ünitesi
HTML	İnternet Siteleri Metin İşaretleme Dili
IBM SPSS	Sosyal Bilimler İstatistik Paketi
KNIME	Konstanz Bilgi Madenci
KNN	K-En Yakın Komşu
M.Ö	Milattan Önce
MIT	Massachusetts Teknoloji Enstitüsü
MS	Microsoft
NLP	Doğal Dil İşleme
PHP	Köprü metni ön işlemcisi
s.	Sayfa
SMS	Kısa mesaj servisi.
Sn.	Sayın
SQL	Yapılandırılmış Sorgu Dili
TAGS	Etiket
vb.	Ve benzeri
VM	Veri Madenciliği
VTBK	Veri Tabanlarından Bilgi Keşfi.

TABLULAR LİSTESİ

Tablo 1.Türkiye Covid 19 Aşı Başlangıç Tarihleri

s.27



ŞEKİLLER LİSTESİ

Şekil 1. Yönetim Bilişim Sistemleri Piramidi	s.9
Şekil 2. Veri Tabanlarından Bilgini Keşfi ve Veri Madenciliği Süreci	s.10
Şekil 3. Veri Madenciliğinin Tarihi	s.11
Şekil 4. Sınıflandırma ve Regresyon	s.12
Şekil 5. Kümeleme ve Sınıflandırma	s.13
Şekil 6. ENIAC(İlk Bilgisayar-30 Ton Ağırlık, 167 m2 Alan)	s.14
Şekil 7. İlk Web Sitesi-Tim Berners-Lee Tarafından 1990 yılında kurulmuştur.	s.15
Şekil 8. We are social 2021 Raporu-Sosyal Medya Kullanımı	s.17
Şekil 9. We are social 2021 Raporu-Twitter Trafik	s.18
Şekil 10. We are social 2021 Raporu-Nüfusun Sosyal Medya Kullanma Oranı	s.19
Şekil 11. We are social 2021 Raporu-Nüfusun Sosyal Medya Kullanım Zamanı	s.19
Şekil 12. We are social 2021 Raporu- Twitter Kullanıcıları Yaş ve Cinsiyet Dağılımı	s.20
Şekil 13. We are social 2021 Raporu- Twitter Kullanıcıları	s.20
Şekil 14. We are social 2021 Raporu- Türkiye	s.21
Şekil 15. We are social 2021 Raporu- Türkiye Web Site Trafik	s.22
Şekil 16. We are social 2021 Raporu- Türkiye Sosyal Medya	s.22
Şekil 17. Türkiye’de İlk Covid 19 Vakası Açıklandı	s.24
Şekil 18. Türkiye Aşılma Oranları- 10 Ekim 2021	s.25
Şekil 19. Türkiye Covid 19 Tarihsel Sıralaması	s.27
Şekil 20. Microsoft Visual Code ve Python Kurulumu	s.30
Şekil 21. Microsoft Visual Code-Python Eklentisi Yükleme	s.30
Şekil 22. Snsrape Kütüphanesi Kurulumu	s.31
Şekil 23. Başlat Windows PowerShell Adımı	s.32
Şekil 24. Snsrape ile Veri Kazıma	s.32
Şekil 25. Snsrape İle Veri Kazıma	s.33
Şekil 26. WampServer Kurulumu	s.35
Şekil 27. WampServer Servisler Başarılı	s.36
Şekil 28. WampServer- Veri Tabanı Kurulumu	s.36
Şekil 29. Sütun Özellikleri	s.37

Şekil 30. Visual Studio ile MySQL Bağlantısı	s.39
Şekil 31. Visual Studio İle Veri Tabanı Bağlantısı Başarılı	s.39
Şekil 32. Microsoft Studio Code- SQL Server Client(mssql) Yükleme	s.40
Şekil 33. Veri Kazıma ve Veri Tabanına Aktarma	s.41
Şekil 34. Kazınan Tweetler' i Veri Tabanına Yükleme	s.42
Şekil 35. Güneydoğu Anadolu Bölgesi Veri Seti İl Dağılımı	s.43
Şekil 36. Ege Bölgesi Veri Seti İl Dağılımı	s.43
Şekil 37. Veri Seti Toplamı(bolge veri seti)	s.44
Şekil 38. Veri Ön İşleme Süreci Ham Veri	s.46
Şekil 39. Veri Ön İşleme Süreci Program Dışı Ön İleme Tamamlandı	s.47
Şekil 40. Veri Ön İşleme Süreci- Excel Deneniyor	s.48
Şekil 41. Veri Ön İşleme Süreci- Excel Deneniyor	s.49
Şekil 42. Veri Ön İşleme Süreci- Excel Deneniyor	s.50
Şekil 43. Veri Ön İşleme Google Colab Açma	s.51
Şekil 44. Veri Ön İşleme Google Colab ve Google Drive Bağlantı	s.51
Şekil 45. Veri Ön İşleme Google Colab CSV Bağlantısı	s.52
Şekil 46. Veri Ön İşleme Google Colab- Veriler Küçük Harfe Dönüştürülüyor	s.53
Şekil 47. Veri Ön İşleme Google Colab- Türkçe Karakterleri Değiştiriyoruz	s.53
Şekil 48. Veri Ön İşleme Google Colab- Noktalama İşaretlerini Çıkartıyoruz	s.54
Şekil 49. Veri Ön İşleme Google Colab- Hecelere(Tokenlere) Ayırma	s.55
Şekil 50. Veri Ön İşleme Google Colab- Sık Kullanılan Kelimelerin(Stop Words) Çıkarılması	s.56
Şekil 51. Veri Ön İşleme Google Colab- Sık Kullanılan Kelimelerin Köklerinin Bulunması	s.56
Şekil 52. Veri Ön İşleme Google Colab-Tamamlandı	s.57
Şekil 53. Google Colab-Kelime Çantası(Bag of Words)	s.58
Şekil 54. KNIME Install Ekranı Türkçe Metin İşleme Eklentisi Yükleme	s.59
Şekil 55. Ege Bölgesi Kelime Bulutu	s.59
Şekil 56. KNIME Zemberek Stemmer ile Kökler Bulundu.	s.60
Şekil 57. Ege Bölgesi ve Güneydoğu Anadolu Bölgesi Paylaşımları Kelime Bulutu	s.61
Şekil 58. Ege Bölgesi Paylaşımları Kelime Bulutu	s.62

Şekil 59. Güneydoğu Anadolu Bölgesi Kelime Bulutu	s.62
Şekil 60. KNIME Kökler Arası Birliktelik Analizi	s.63
Şekil 61. KNIME String Replacer	s.64
Şekil 62. KNIME GroupBy	s.65
Şekil 63. KNIME Cell Splitter	s.65
Şekil 64. KNIME Item Set Finder	s.66
Şekil 65. KNIME Ege Bölgesi Birliktelik Çıkarımları	s.67
Şekil 66. KNIME Güneydoğu Anadolu Bölgesi Birliktelik Çıkarımları	s.67



GİRİŞ

Bilişim teknolojilerinde yaşanan gelişmeler üretim sektörünü yakından etkileyerek endüstri alanında köklü değişikliğe neden olmuştur. Ancak bilişim teknolojilerinde yaşanan gelişmelerin neden olduğu değişiklikler yalnız endüstri alanında sınırlı kalmamış toplumsal ve sosyal yaşamda da değişiklikleri beraberinde getirmiştir. Bu değişiklikler 1990 sonrasında doğan ve dijital yerliler adı verilen kuşağın yaş alması ile günümüzde yaşanan toplumsal değişimler gözle görülür bir hal almış ve akademik çalışmalara konu olmuştur.

Sosyal medya adı verilen sitelerin kullanıcı dostu tasarımları kullanım kolaylığı imkânı verirken, içeriklerinin melez medya türlerinden meydana gelmesi dikkat çekiciliklerini artırmaktadır. Bu iki özellik zaten bilişim teknolojilerinin içerisine doğmuş ve sosyal medya kullanan dijital yerlileri içerik üretmeye teşvik ederken, 1990 öncesi doğmuş dijital göçmenleri de albenisi ile cezbetmektedir.

İstatistikler sosyal medya kullanım sıralamasında ülkemizin 32'inci sırada yer aldığını göstermektedir. Bu durum sosyal medya ortamlarında üretilen milyarlarca içeriği ifade ederken pazarlama, politika gibi alanlar için okunması ve yönetilmesi gerekli pazarlara dönüşmüştür. Bu durum ise sosyal medya içeriklerinin veri analistleri, veri madenciliği, metin madenciliği gibi dalların çalışma alanı içerisine alınması gerekliliğini doğurmuştur.

İnsanlar sosyal ağlarda içerikler üretirler ancak bu içeriklerin üreticilerinin gerçek tutum ve görüşlerini yansıtmadığı ya da gerçekleştireceği davranışlar hakkında bilgi verip vermediği henüz netleşmemiştir. Bu ilgi çekici sorunu sorgulamak amacı ile bu çalışma gerçekleştirilmiştir.

Aralık 2019 tarihinde Çin'in Wuhan eyaletinde Covid 19 virüsü ile enfekte olan hastalarda solunum rahatsızlığı görülmüş ve bu tarih itibari ile her gün daha hızlı bir şekilde vaka sayıları artarak küresel bir salgın etkisi oluşturmuştur. Covid 19 virüsünün tahribatına karşı ülkelerce aşular geliştirilmeye başlanmıştır. Faz 3 çalışmaları tamamlanan farklı ülkelere ait aşular devletlerce vatandaşlarına uygulanmaya başlanmıştır. Ülkemizde Covid 19 virüsüne karşı Biontech ve Sinovac olmak üzere iki aşı, belirlenen aşı gruplarına belirlenen tarihlerde ücretsiz olarak uygulanmaya başlanmıştır. 25 Haziran 2020 tarihinde 18 yaş grubu dâhil edilmek

üzere yedi aşılama grubuna aşı hakkı tanımlanmıştır. Sağlık Bakanlığı tarafından yapılan aşı oranları iller dağılımı yüzde ifadeler ile açıklanmış, Ege Bölgesi'nin aşılama oranının yüzde 80 üzerinde seyrederken, Güneydoğu Anadolu Bölgesi aşılama oranlarının yüzde 60 altında seyrettiği gözlenmiştir.

Tezde Ege Bölgesi ve Güneydoğu Anadolu Bölgesi arasındaki aşılama oranları ve Covid 19 konulu bölgelerde konumlu kullanıcıların tweetleri göz önüne alınarak aşılama oranlarına ait bölgelerin tutumlarını keşfetmek amaçlanmıştır.

Ege Bölgesi aşılama oranlarının yüksek, Güneydoğu Anadolu Bölgesi aşılama oranlarının düşük olması ile konumu bu bölgelerde olan Twitter kullanıcılarının Covid 19 konulu içerikleri arasında ilişki vardır ve yoktur olmak üzere H1 ve H0 hipotezleri oluşturulmuştur.

Tez kapsamında planlanan veri madenciliği uygulamasına yönelik kavramsal kısım ve hipotezleri ispat amacı için planlanan uygulama kısmı olmak üzere iki bölüm oluşturulmuştur.

Tezin birinci bölümünde literatür araştırmasına yer verilmiştir. İkinci bölümde veri madenciliği kavram modelleri ile Web 2.0 araçları konuları ele alınmıştır. Tezin üçüncü bölümünde gerçekleştirilen uygulama detaylandırılmış ve sonuç bölümünde bulgular ve gelecek çalışmalar hakkında bilgi aktarılmıştır.

BİRİNCİ BÖLÜM

LİTERATÜR ARAŞTIRMASI

1.1. LİTERATÜR ARAŞTIRMASI

Literatür Araştırması bölümü içerisinde üzerine çalışılan konu ile benzer alandaki çalışmalar incelenmiş ve bu çalışmalar hakkında bilgi verilmiştir.

Aydoğan ve Gündoğdu (2021), çalışmalarında Sağlık Bakanı Fahrettin Koca'nın Twitter paylaşımları ile etkileşim halinde olduğu takipçilerinin davranışlarına yön vermek amacıyla sergilediği dijital liderliğin önem derecesini ortaya koymayı amaçlamışlardır. Bu çalışma kapsamında Bakan Fahrettin Koca'ya ait 1.182 tweet ve bunlara ait retweet, beğeni ve yorum sayılarını dikkate alan bir veri tabanı oluşturulmuş ve veri tabanına içerik analizi ve sezgisel bulanık analitik hiyerarşi süreci uygulanmıştır. Çalışmada benimsenen yöntem sonucunda Bakan Fahrettin Koca'nın Twitter paylaşımlarında en önemli liderlik boyutunun, vizyon liderlik boyutu olduğu ve ilgili twitter paylaşımlarına ilişkin retweetlerin önem derecesinin sorgulanan diğer unsurlara oranla daha fazla olduğu bulgularına ulaşılmıştır.

Gün ve Kolukırcık (2022), çalışmalarında bir belirsizlik unsuru olarak Covid 19 algısında ortaya çıkan toplumsal eğilimleri ortaya çıkarmayı amaçlamışlardır. Tez kapsamında 11-16 Mart 2020 tarihleri arasında risk, tehlike ve belirsizlik kavramlarından herhangi birisini içeren tweetler içerik analizi tekniği kullanılarak incelenmiştir. Covid 19 salgını karşısında 5 farklı risk algısı ve sosyal tepkinin geliştiği bulgusuna ulaşılmıştır.

Sarıman ve Mutaf (2020), çalışmalarında Covid 19 sürecinde ülkemizde alınan önlemlerin ve hizmete sunulan uygulamaların insanlar üzerinde bıraktığı etkiyi ölçmeyi amaçlamışlardır. Olumlu ve olumsuz olmak üzere yaklaşık 2.000.000 veri üzerinde lojistik regresyon yöntemi uygulanmış, duygu analizi yöntemine göre beş başlık altında toplanan konularda haftalık süreçte gözle görülebilir bir değişikliğin varlığı sorgulanmıştır.

Bilgiç ve Akyüz (2020), çalışmalarında Sağlık Bakanı Fahrettin Koca'nın Twitter paylaşımlarını incelemiş ve mesajlarında en çok hangi noktalara vurgu yaptığı, dönem ve içerik olarak etkileşimlerde farklılaşma olup olmadığı Gündem Belirleme

Kuramı çerçevesinde araştırılmıştır. Araştırma kapsamında Fahrettin Koca'ya ait ülkemizde ilk vakanın görüldüğü 11 Mart 2020 ile ilk sokağa çıkma yasağının ilan edildiği 10 Nisan 2020 tarihleri arasındaki Twitter paylaşımları değerlendirilmeye alınmıştır. Salgın döneminde gündem belirlemede önemli bir aktör olan Fahrettin Koca'nın en çok sağlık, Koronavirüs ve tedbir kelimelerini kullanırken en az maske, pandemi, bulaşıcı ve sosyal mesafe kelimelerini kullandığı sonucuna ulaşılmıştır.

Karlı ve Dondurucu (2020), çalışmalarında ekonomik, politik ve ideolojik farklılıkların, haber kanallarının hükümetin pandemi ilgili çalışmalarının sunuş biçimini etkilediği savından hareketle CNN ve Fox News kanallarına ait Covid 19 dönemi tweetlerini incelemişlerdir. Mart- Nisan 2020 tarihleri arasında CNN ve Fox News' in hesaplarında Donald Trump ve Hükümetinin salgınla mücadele faaliyetlerini içeren, en yüksek etkileşim düzeyine sahip içerikler eleştirel söylem yöntemi ile çözümlenmiştir. Donald Trump ve Hükümetinin salgın yönetimi başarılı ve olumlu, CNN haberlerinde ise başarısız ve olumsuz olarak sunulduğu sonucuna ulaşılmıştır. Haberlerin söylemsel farklılığın kitle iletişim araçlarının ideolojik, iktisadi ve siyasal yapısı ile bağlantılı olduğu sonucuna ulaşılmıştır.

Bezirci ve Biçen Aras (2022), çalışmalarında bilgiye erişim hizmeti veren kütüphaneler üzerine odaklanmışlardır. Bilgiye erişimde devrim oluşturan bilişim teknolojilerinin kütüphanelerin çalışma yöntemlerini de değiştirdiği kabulü ile Covid 19 Pandemi döneminde tüm konularda bilgiye erişimin kişiler ve kurumlar ve kişiler arası iletişimin bağlamı olarak kullanılan sosyal medyanın kütüphaneler tarafından nasıl kullanıldığını incelemişlerdir. Bu inceleme için Dünya'dan ve ülkemizden beşer üniversite kütüphanesinin sosyal medya paylaşımları izlenmiştir. Kütüphaneler tarafından ağırlıklı olarak Twitter sosyal mecrasının kullanıldığı gözlenmiştir. Twitter'da kütüphaneler tarafından yapılan paylaşımların daha çok kütüphane çalışmalarının paylaşımı ve etkinliklerin duyurulması amaçladığı ve sosyal medyayı tek yönlü bir iletişim kanalı olarak kullandıkları sonucuna ulaşmışlardır. Bu sebeple kütüphanelerin sosyal medya hesaplarını kullananların sosyal medya, halkla ilişkiler ve pazarlama alanlarında eğitilmesi gerekliliğine vurgu yapmışlardır.

Artan Özoran (2020), çalışmasında ülkemizin on markasının Covid 19 pandemi sürecinin ilk iki ayında yapmış olduğu toplam 510 paylaşımı nitel ve nicel yöntem kullanarak incelemiştir. Markaların yapmış olduğu paylaşımlar paylaşımların

kategorileriler ve paylaşımların etkileşimine göre iki farklı değerlendirmeye alınmıştır. Yapılan değerlendirmeler markaların iki aylık süreçte en çok Covid 19 konulu paylaşımda bulunduğunu göstermektedir. Paylaşımlar beğeni ve yorum sayıları dikkate alınarak etkileşime göre değerlendirildiğinde ise markaların en çok çalışanlarına teşekkür, umut ve ekonomik destek mesajı veren paylaşımlarının en çok beğeni aldığı, yorum sayısının ise en çok ekonomik destek konulu paylaşımlarına yönelik olduğu ve ekonomik desteğin yetersiz olduğuna dair dönüş yapıldığı sonucuna ulaşılmıştır.

Temizhan ve Mendeş (2021), çalışmalarında halkın Covid 19 salgınına bakış açılarına ve düşüncelerine dair özetleyici bilgi üretmeyi amaçlamışlardır. Bu amaç ile Covid 19 konulu 40.000 adet tweeti regresyon analizi yöntemini kullanarak incelemiştirlerdir. Paylaşımların maske, Covid 19, Ankara Wuhan ve Bilim Kurulu konuları üzerinde şekillendiği sonucuna ulaşmışlardır. Yapılan duygu analizinde ise Covid 19 kelimesinin negatif bir duygu oluşturduğu ve bilim kurulu kararlarının dikkate değer bir etki bıraktığı bulgularına ulaşmışlardır.

Ateş ve Baran (2020), çalışmalarında ABD ve Türkiye devletleri Sağlık Bakanlıklarına ait resmi twitter adreslerini incelemiştirlerdir. Bu inceleme ile Covid 19 pandemi süreci boyunca hangi devlete ait sosyal medya hesabının kriz yönetiminde daha etkin kullanıldığı sorusuna yanıt aramışlardır. Çalışma kapsamında devletlerin ilk vaka görülme tarihi ile normalleşme sürecinin ilanı tarihi arasında Sağlık Bakanlıklarına ait resmi Twitter hesabından yapılan paylaşımlar toplanmış ve belirli başlıklar listesine göre her paylaşım var ya da yok etiketi ile etiketlenerek, IBM SPSS programında analiz edilmiştir. Türkiye Devleti Sağlık Bakanlığı paylaşımlarının sırasıyla bilgi sağlamak ve uyarmak, gözetim izleme ve durum farkındalığının sağlanması, risk ve kriz yönetimleri konusunda kamu bilincinin oluşturulması amacıyla gerçekleştirildiği ABD Sağlık Bakanlığının ise konu başlıklarının aynı sıralama ile olmasının yanında yapılan analiz sonrasında paylaşımlarının yetersiz ve etkisiz kaldığı sonucuna ulaşılmıştır.

Şeker Elmas (2019), çalışmasında Sivas konumlu halk tarafından tutulan en iyi 2.500 adet Twitter paylaşımını duygu analizi yöntemi ile incelemiştir. Bu paylaşımlar Knime programı ile Twitter API kullanılarak çekilmiş ve metin madenciliği ön işleme sürecine alınmıştır. Ardından Sivashlı gönüllü yüz kişi tarafından her biri mutlu, kızgın,

şaşkın, üzgün ve tarafsız olarak etiketlenmiştir. Yapılan işlemlerin ardından 2.500 adet paylaşımından 596 adeti Naive Bayes, Karar Ağaçları, Destek Vektör Makineleri algoritmaları kullanılarak duygu analizi için sınıflandırılmıştır. Veri analizi sonrası Sivas, Sivasspor kelimesi içeren paylaşımların mutlu olarak sınıflandırıldığı Türkçe 'de mutluluk duygusunun daha kesin ifade edilirken şaşkınlık, üzümlük ve kızgınlık ifadelerinin daha açık uçlu olduđu sonucuna ulaşılmıştır.

Demirok (2018), çalışmasında E-ticaret sitesindeki satış verilerini kullanarak ürünler arası birliktelik kurallarını incelemiştir. Bu inceleme sonucunda birlikte satılan ürün kümelerini keşfedilmesi bu ürün kümelerinin pazarlama bölümü tarafından kampanya faaliyetlerinde kullanılmasını hedeflemiştir. Çalışma kapsamında bir E-Ticaret sitesinin 7.114 farklı müşterisine ait 33.061 satış kaydı Knime ile değerlendirilmeye alınmıştır. Knime uygulaması üzerinde veri seti veri ön işleme sürecinin ardından Apriori algoritması ile Pazar sepet analizi gerçekleştirilmiştir. Ürünler arasında altılı, beşli, dördü, üçlü ve ikili birliktelik kuralları keşfedilmiş çalışma sonucunda bu ürünler ile ilgili pazarlama önerilerinde bulunulmuştur.

Bozkurt (2021), çalışmasında veri madenciliği yöntemlerini ile tam kan sayımı sonuçları ile kişilerin Covid 19 testi sonuçlarını tahmin etmeyi amaçlamıştır. Veri analizi aşamasında kan sayımları veri ön işleme alınmış ardından kişilere ait kan sayımları ve Covid 19 test sonuçları kullanılarak model kurulmuştur. Tahminleme için Rastgele Orman Algoritması, Naive Bayes ve Destek Vektör Makineleri algoritmaları kullanılmış olup en iyi tahmin Rasgele Orman Algoritması ile üretilmiştir.

Demirciođlu ve Eşiyok (2019), çalışmalarında aralarında ülkemizin de bulunduğu otuz altı ülkeyi Covid 19 ile mücadelede kümeleme analizi yöntemi ile sınıflandırmışlardır. Analiz için nüfus oranları, vaka oranları, test oranları, ölüm oranları, yaşlı nüfus oranı gibi on değişken kullanılmıştır. Analiz Weka uygulamasında KNN algoritması ile gerçekleştirilmiştir. Ülkeler ikili, üçlü ve dördü kümelere ayrılmıştır. Ülkemiz ikili kümelemede birinci kümede, üçlü kümelemede üçüncü kümede ve dördü kümelemede dördüncü kümede yer almaktadır. Her küme içerisindeki ülkelerin belirtilen değişkenlere aldığı değerler ile çalışmanın sonuç kısmında ayrı ayrı yorumlanmaktadır.

Kuş (2021), çalışmasında mültecilere yönelik dijital nefret söylemini tetikleyen sebepleri araştırmıştır. Çalışma kapsamında Covid 19 konulu haberlere

Youtube kanalında yapılan yorum ve beğenileri veri seti olarak kullanmış ve içerik analizi gerçekleştirmiştir. İçerik analizi 510 kullanıcı yorumu ve 3.636 kullanıcı etkileşimi üzerinde gerçekleştirilmiştir. Veri madenciliği süreci sonrasında kullanıcı yorumlarının yüzde yirmi dokuzunun mültecilere yönelik dijital nefret söylemi içerirken yorumlara atılan toplam beğenin yüzde kırk dokuzunun mültecilere yönelik dijital nefret söylemi yorumlarına ait olduğu bulgusuna ulaşılmıştır. Ayrıca ülke, amaç, kişisel tercih ve diğer olmak üzere dört kategoride mültecilere yönelik dijital nefret söylemini tetikleyici etkenlerin bulunduğu görülmüştür.

Tez ile benzer bazı çalışmalara dair bilgileri metnin üst kısmında yer verilmiş olup Twitter sosyal ağ paylaşımları ve Covid 19 konulu başka benzer çalışmalarda alanda mevcuttur. Sadece belli sayıda çalışmaya bu bölümde yer verilmiş olup çalışmalar diğerleri ile de Covid 19 konusu, Twitter üzerinde çalışma ve yöntem kısmı benzerlikler görülmektedir. Ancak ayrıntılarına çalışmanın devamında yer verilecek olan uygulama bölgesel çıkarımlar yapmaya hedeflemesi yönüyle alandaki benzer kaynaklardan farklılık göstermektedir.

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ

Veri Madenciliği bölümü altında veri madenciliği kavramı ve modellerini tanımları yer almaktadır. Alandaki çalışmalar incelenmiş, kaynaklardan hareketle veri madenciliği kavramı tanımlanmıştır. Ardından veri madenciliği modelleri açıklanmıştır.

2.1. VERİ MADENCİLİĞİ KAVRAMI VE MODELLERİ

Tez kapsamında alandaki veri madenciliği konulu kaynaklar incelenmiştir. Kaynaklardan edinilen bilgiler özümserenerek veri madenciliği kavramı ve modelleri özetlenmiş, bilgilerin veriliş sıralamasında veri madenciliği tanımına öncelik verilmiştir.

Gideon Gartner tarafından kurulan ve 1979 yılından itibaren faaliyette olan Gartner'e göre veri madenciliği süreci büyük veri depolarında bulunan büyük miktarda verinin, örüntü tanıma teknolojilerinin yanı sıra istatiksel ve matematiksel teknikler kullanılarak işlenmesi sürecidir. Bu süreç veriler arasındaki anlamlı korelasyon ve kalıpların keşfedilmesini amaçlamaktadır (Information Technology Gartner Glossary, 17.01.2022).

Bir başka kaynakta ise veri madenciliği, veri tabanlarında tutulan milyonlarca verinin yazılım sistemleri ile işlenerek aralarındaki ilişkilerin ortaya çıkarılması, ortaya çıkarılan ilişkilerin, örüntülerin ve trendlerin geleceğe yönelik doğru tahminlerin gerçekleştirilmesinde kullanılması süreci olarak tanımlanmıştır. Bu süreç kurumlardaki karar destek sistemlerini beslemekte ve sistemlerin başarısını etkilemektedir (5 Temel Soruda Veri Madenciliği (Data Mining) Nedir?, 18.01.2022).

Karar Destek Sistemleri ise, daha çok rutin olmayan durumlara yönelik kararların alınmasını desteklemekte bu kararların alınması için kurum içi veri kaynakları ile birlikte döviz kurları, devlet teşbih politikaları vb. gibi dış veri kaynaklarını kullanmaktadır (Laudon & Laudon, 1994).

Bir başka kaynakta ise veri madenciliği, veriden bilgi üretmek amacıyla geliştirilmiş yol ve yöntemlerden biri ve potansiyel bilgi barındıran, farklı özellikteki

verileri farklı, aynı özellikteki verileri aynı kümede olacak şekilde gruplandırmak olarak tanımlanmaktadır (Topal, 2021).

Tanım aynı özellikteki verileri aynı gruba dâhil etme ifadesi ile sonraki başlıklarda konu edinilecek olan sınıflama ve kümeleme kavramlarına atıfta bulunurken veri madenciliğinin yalnız sınıflama ve kümeleme kavramları ile ifade edilmesi doğru değildir. Ancak veri yığınları üzerinde çalışan veri madenciliğinin veriden bilgi üretme konusundaki başarısı alanda kabul görmektedir. Veriden bilgi üretme yöntemi olarak tanımlanan veri madenciliğinin bugün teknolojik altyapısını verimli ve etkili bir şekilde kurmayı başarmış kurumlar içerisinde yönetim bilişim sistemleri piramidinin başarısını doğrudan etkilediği söylenebilir. Şekil 1’de yönetim bilişim sistemleri piramidine yer verilmiştir.

Şekil 1. Yönetim Bilişim Sistemleri Piramidi

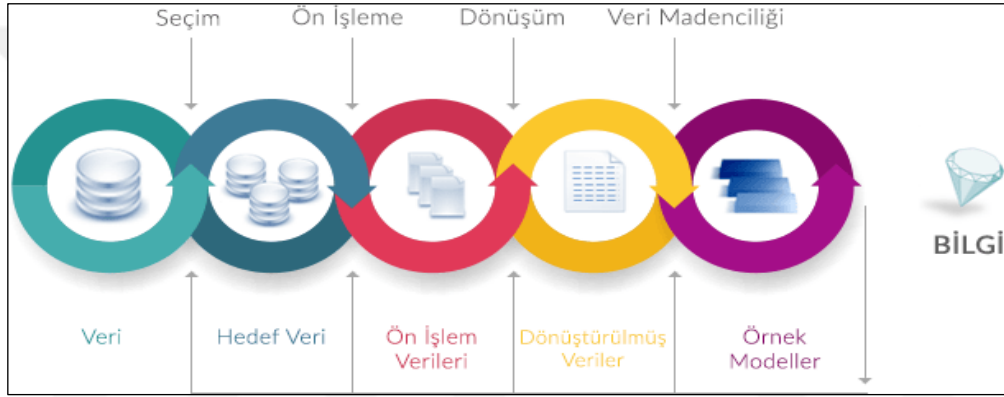


Kaynak: Tecim, V. “Yönetim Bilişim Sistemleri (YBS)”, <https://vahaptecim.com.tr/yonetim-bilisim-sistemleri/>, (20.01.2022)

Veri Madenciliğini tanımlamak adına yürütülen çalışmanın devamında alanda incelenen diğer kaynaklarda veri madenciliği, bilgi çıkarmak adına veri tabanlarında bulunan ham verinin işlenmesi olarak tanımlanmıştır. Bu tanım devamında Veri Tabanlarından Bilgi Keşfi (VTBK) ve Veri Madenciliği (VM) olmak üzere iki kavrama dikkat çekmektedir. Veri Tabanlarından Bilgi Keşfi, veri tabanlarından verinin seçimi, ön işleme, gerekli ise veri tipinin dönüştürülmesi süreçlerini ifade ederken,

veri madenciliği ise verinin değerlendirilmesi süreci olarak tanımlanmaktadır. Ancak günümüzde veri tabanlarından verinin toplanması-seçimi süreci ile başlayıp, sonuç kısmında bilgiye ulaşılan bu sürecin bütünü veri madenciliği olarak tanımlanmakta olup Veri Tabanından Bilginin Keşfi ve Veri Madenciliği kavramları aynı anlamda birbirini yerine kullanılmaktadır (Veli, 2020). Şekil 2’de veri madenciliği süreci şema ile gösterilmiş olup şemanın adımlarına dair ayrıntıya çalışmanın Üçüncü Bölüm, Uygulama başlığı altında yer verilmiştir.

Şekil 2. Veri Tabanlarından Bilgini Keşfi ve Veri Madenciliği Süreci



Kaynak: “5 Temel Soruda Veri Madenciliği (Data Mining) Nedir?”, 26.01.2019, 5 Temel Soruda Veri Madenciliği (Data Mining) Nedir? - Vizyoner Genç (vizyonergenc.com), (20.01.2022)

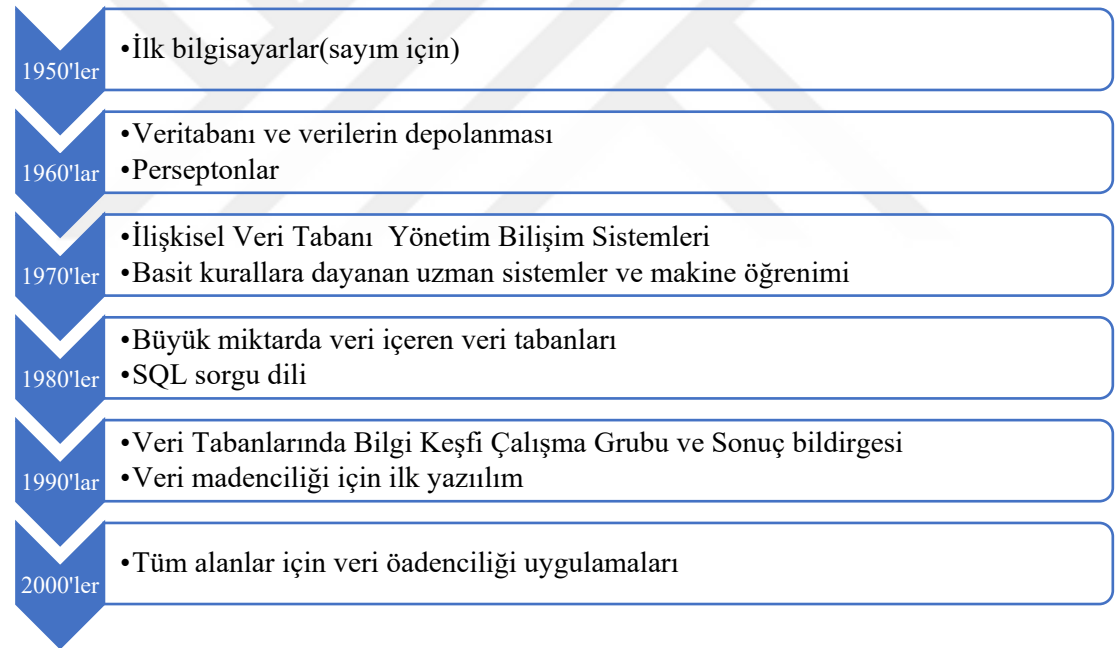
Veri madenciliği kavramı özetle, veri tabanlarında bulunan ham gerçekler yığını olarak tanımlanan (semboller, işaretler, sinyaller, kelimeler, sayılar vb.) verilerden bilgi üretmek amacı ile matematiksel ve istatistiksel yöntemlerin kullanıldığı, veriler arası anlamlı ilişkilerin keşfedilmeyi amaçlandığı süreçtir.

Veri madenciliği, veri tabanında veri saklama, verileri işleme, süreç sonunda ortaya çıkan bilgileri görselleştirme, kurum faaliyetlerinin, veri madenciliği sonucu ortaya çıkan bilgiye uygun yeniden düzenlenmesi gibi gereklilikleri beraberinde getirmektedir. Sayısı artırılabilir olan bu gereklilikler veri madenciliği sürecini teknolojik gelişmeler ile yakından ilişkili kılmaktadır. Bu yüzden teknolojik altyapısını verimli bir şekilde kurmuş ve etkin bir şekilde kullanan örgütler, alanında

veri madenciliği vb. gibi faaliyetlerle rekabet üstünlüğü sağlamakta ve devamlılığını sürdürülebilir kılmaktadır.

Bilgi teknolojilerinin gelişimi 1950'li yıllarda ilk bilgisayarların sayım için kullanılmasıyla başlamıştır. Sonraki yıllarda yaşanan gelişmelerin ardından 1980'li yıllara gelindiğinde veri tabanlarında biriken büyük veri kümelerinde saklı bilgilere ulaşmak için SQL sorgu dili yetersiz kalmıştır. 1990 yıllarda veri tabanlarında tutulan veriler üzerinden bilgi üretme çabasına hız verilmiş bu çalışmalar Veri Tabanlarından Bilgi Keşfi süreci olarak adlandırılmıştır. Çalışmalar sonucu veri madenciliği için ilk yazılım gerçekleştirilmiştir (Savaş, Topaloğlu, & Yılmaz, 11). Şekil 3'te veri madenciliğinin tarihi gösterilmektedir.

Şekil 3. Veri Madenciliğinin Tarihi



Kaynak: (Savaş, Topaloğlu, & Yılmaz, 11)

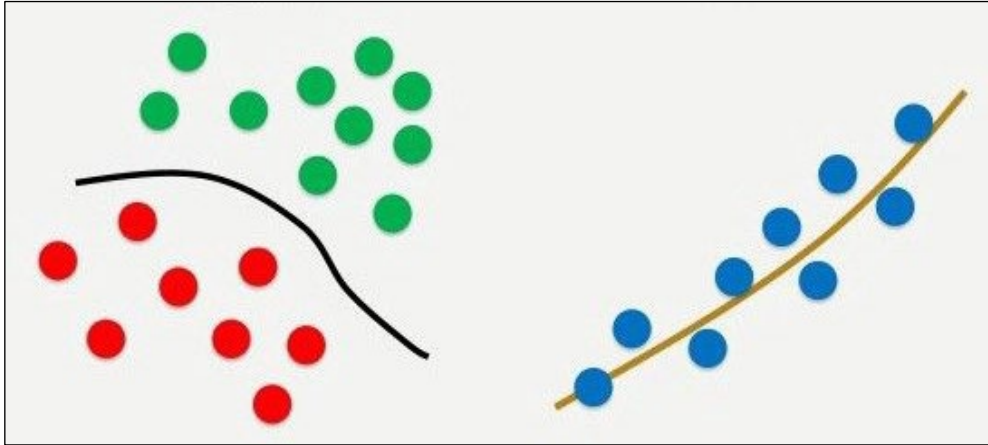
Veri madenciliğinde kullanılan modeller tahmin edici ve tanımlayıcı modeller olmak üzere iki başlık altında incelenmektedir. Tahmin Edici Modeller, sonuçları bilinen verilerden hareketle model geliştirilmesi ve geliştirilen modelden hareketle sonuçları bilinmeyen veri kümelerinin sonuç değerlerinin tahmin edilmesini amaçlamaktadır. Tanımlayıcı Modellerde ise sonuçları bilinen veriler söz konusu

olmayıp karar vermede kullanılmak üzere mevcut verilerdeki örüntülerin keşfedilmesi amaçlanmaktadır (Özekes, 2003).

2.1.1. Sınıflama ve Regresyon

Veri madenciliği modellerinden olan sınıflama ve regresyon, tahmin edici modeller olarak tanımlanmaktadır. Kategorik ve sürekli değişkenlerden yararlanarak sınıflandırmaya yönelik analizler yapmak amacı ile geliştirilmiş istatistiksel iki yöntemdirler. Tahmin edilen bağımlı değişkenin kategorik olması durumunda sınıflandırma, tahmin edilen bağımlı değişkenin sürekli olması durumunda regresyon olarak adlandırılmaktadırlar (Ceyhan, 2020). Ancak çok terimli lojistik regresyon gibi hem kategorik değerlerin hem de sürekli değişkenlerin tahmin edilmesine imkan veren teknikler sınıflandırma ve regresyon modellerini birbirine yaklaştırmakta ve tekniklerin birbiri yerine kullanılmasını mümkün kılmaktadır (Akpınar, 2000). Şekil 4'te solda sınıflandırma modeli, sağda ise regresyon modeli ile sınıflandırma gösterilmektedir.

Şekil 4. Sınıflandırma ve Regresyon



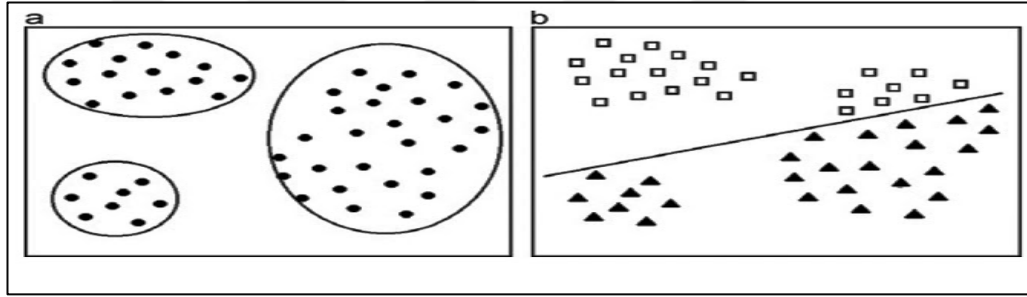
Kaynak: Bilgisayar Kavramları, “Makine Öğrenmesinde Kümeleme Algoritmaları”,23.04.2018, https://www.youtube.com/watch?v=CMbLMiJAx_g, (Erişim Tarihi: 02.02.2022)

2.1.2. Kümeleme

Veri madenciliği tanımlayıcı modellerinden olan kümeleme, verileri kümelere, ayırarak sınıflandırma işlemi yapmayı amaçlamaktadır. Bu yöntemde aynı kümedeki veriler birbirine benzerken, başka kümelerin üyeleri birbirinden farklılaşmaktadır (Akpınar, 2000).

Sınıflama tekniğinde sınıflar bilinmekte ve yeni gelen verinin sınıfını tahmin etmek amaçlanmaktadır. Kümelemede ise verilerin sınıfları bilinmemekte verilerin sınıfları tanımlanmaya çalışılmaktadır (Özekes, 2003). Şekil 5'te kümeleme ve sınıflandırma algoritmaları gösterilmektedir.

Şekil 5. Kümeleme ve Sınıflandırma



Kaynak: Bilgisayar Kavramları, “Makine Öğrenmesinde Kümeleme Algoritmaları”, 23.04.2018, https://www.youtube.com/watch?v=CMbLMiJAx_g, (Erişim Tarihi: 02.02.2022)

2.1.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

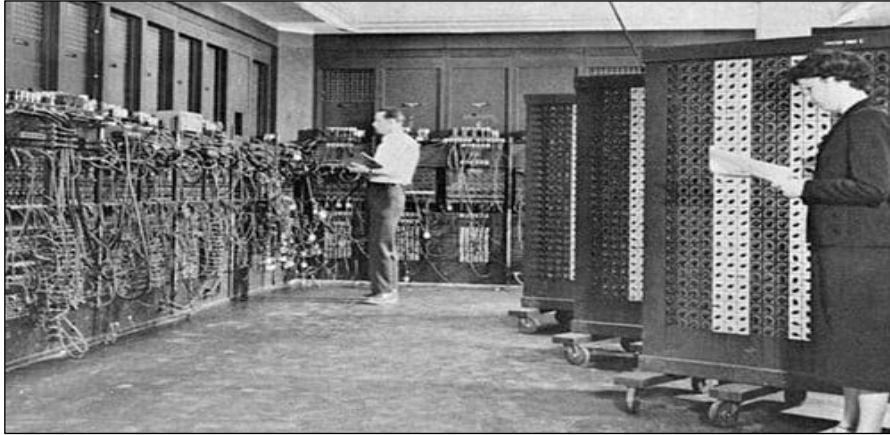
Veri madenciliği tanımlayıcı modellerinden olan birliktelik kuralları, verilen veri kümesi içerisinde sıkça görülen durumların tespiti amaçlamaktadır (KALIKOV, 2006). Veri kümesi içerisindeki nesnelere arasındaki ilişkilerin bulunmasını amaçlayan bu veri modelinde, tanımlayıcı modellerin ayırt edici özelliği olarak sonuç değerlerinin sisteme verilmediği bilinmektedir. Denetimsiz bir veri madenciliği yöntemi olup ilk ve en çok bilinen örnekleri Pazar sepet analizleri uygulamaları olarak görülmektedir (Ay, 2009). Birliktelik kuralları, eş zamanlı gerçekleşen ilişkilerin tanımlanmasında kullanılırken Ardışık Zamanlı Örüntüler ise birbiri ile ilişkisi olan ancak birbirini

izleyen dönemlerde gerçekleşen ilişkilerin keşfedilmesinde kullanılmaktadır (Akpınar, 2000). Birlikte kuralları, bir müşteri düşük yağlı peynir ve yağsız yoğurt aldığında %85 ihtimalle diyet süt de satın alır ifadesi ile örneklendirebilirken, Ardışık zamanlı örüntüler, X ameliyatı olan hasta, 15 gün içinde %45 ihtimalle Y enfeksiyonu yaşayacaktır ifadesi ile örneklendirilebilir (Akpınar, 2000).

2.2. WEB 2.0 VE ARAÇLARI

İnsanoğlunun birbiri ve çevresi ile iletişim kurma gayretinin çok eski tarihlere dayandığı bilinmektedir. İletişim kurmanın araçlarından biri olan yazı sisteminin tarihi M.Ö 3000 Mısır Hiyeroglif adı verilen resim yazılarına atfedilmektedir. Yazının bulunması ve kağıdın keşfi ile devam eden süreç 1946 yılında J.Eckert ve Mauchly ENIAC adlı ilk bilgisayarı icat etmesi ile yeni bir soluk kazanmıştır (Şekil 6). Askeri amaçlar ile icat edilen bu devasa alet, bilgi ve iletişim teknolojilerinin gelişimini başlatmıştır (Özmelek Taş, 2015).

Şekil 6. ENIAC (İlk Bilgisayar-30 Ton Ağırlık, 167 m² Alan)



Kaynak: Banu Cantekin, “Tarihin İlk Elektronik Bilgisayarı ENIAC”, 07 Ocak 2020, <https://www.iienstitu.com/blog/tarihin-ilk-elektronik-bilgisayari-eniac>, (Erişim Tarihi 08.02.2022)

1957 yılına gelindiğinde Sovyetler Birliği tarafından ilk uzay uydusu Sputnik projesi başlatılmıştır. Soğuk Savaş döneminde başlatılan bu projeye Amerika, İleri Araştırma Projeleri Ajansı -ARPA (Advanced Research Projects Agency)' yı kurarak tepki göstermiştir (RAHIMI, 2015). 1970 yılında Amerika'da üniversiteler arası iletişim kurmada kullanılmaya başlanan ARPANET isimli ağın ismi 1985 yılında INTERNET olarak değiştirilmiştir. 1990 yılında ise "World Wide Web" ismi WWW dünya gündemine girmiştir (Özmelek Taş, 2015).

Kaynaklarda World Wide Web -WWW Tim Berners-Lee tarafından internetin çok amaçlı kullanımı amacıyla geliştirilmiş yeni bir teknoloji olarak tanımlanır. World Wide Web -WWW aynı zamanda birinci nesil internet teknolojisi olarak tanımlanmakta ve Web 1.0 ismiyle adlandırılmaktadır.

Web 1.0 teknolojisi ile internet siteleri metin işaretleme dili (Hypertext Markup Language-HTML) geliştirilmektedir. Bu siteler iletişim amacı ile kullanılmakta olup sitelerin içerik üretimi kısıtlı sayıda ve bilgisayar teknolojileri konusunda uzman kişiler tarafından oluşturulmaktadır. İletişim sürecinin ulaşmayı hedeflediği kitle ise çok sayıda kişiyi ifade etmektedir (Özmelek Taş, 2015) (Aşman Alikılıç, 2011). Şekil 7'de ilk web sitesi gösterilmektedir. İlk web sitesi linklerden ve linklere bağlı metinlerden oluşmaktadır.

Şekil 7. İlk Web Sitesi-Tim Berners-Lee Tarafından 1990 yılında kurulmuştur.



Kaynak: Tim Berners-Lee, <http://info.cern.ch> - home of the first website, <http://info.cern.ch/>, (Erişim Tarihi 10.02.2022)

2.2.1. Sosyal Ağlar

Sosyal ağ, toplumsal ağ olarak da isimlendirilmekle birlikte sözlük anlamıyla birden fazla kişinin toplumsal ilişkiyle birbirine bağlanması olarak tanımlanmakta olup burada toplumsal bağı oluşturan ilişkilere akrabalık, arkadaşlık, iletişim ve otorite sayılabilmektedir. Tez kapsamında kullanılmış olan sosyal ağlar kelimesi ise sosyal ağ sitelerine karşılık gelmektedir. Sosyal ağ siteleri esnek ve kullanıcı dostu özellikleri ile milyonlarca kullanıcının üyelik yolu ile her türlü bilgiyi özgürce paylaşabildiği dijital ortam olarak tanımlanabilir. Sosyal ağ siteleri, bilişim teknolojilerinin kullanımı ile sosyal ağların sanal ortama taşınmasına, sanal cemaatlerin oluşmasına imkân vermektedir (Özmen, Aküzüm, Sünkür, & Baysal, 2012).

Web 2.0 teknolojileri ile birlikte iletişim, bilişim teknolojileri üzerinden çift taraflı kurulabilir özellik kazanmıştır. Sosyal ağ siteleri ya da daha yaygın ismi ile sosyal medya kullanıcıları zaman ve mekan sınırlaması olmaksızın arama yapma, içerik üretme paylaşma ve diğer kullanıcılara ait içeriklere erişim ve katkıda bulunma fırsatı vermekte olan bir iletişim ortamıdır (Sosyal Medya, 2022).

Sosyal medya; erişilebilirlik (internetin var olduğu her yârden erişim), katılım (içerikleri yönelik geri bildirim ve katkıda bulunabilir), konuşma ve topluluklar (hızlı ve çift yönlü iletişim kurulabilir), bağlantısallık (bağlantı-link paylaşılabilir), programlanabilirlik (Kullanıcı geliştirme ara yüzleri-API mevcuttur.), ölçeklenebilirlik (çok sayıda kitleye hizmet sağlanır), dinamiklik (sürekli içerik akımı) gibi özelliklere sahiptir (Sosyal Medya, 2022).

Sosyal ağ siteleri, sosyal medya gibi kavramlar ile yıllar içerisinde gösterdiği gelişim ile birlikte farklı isimlendirmeler ile adlandırılan yeni nesil iletişim sitelerinin ilk örnekleri 1995 yılında kurulan “classmates.com” ve 1997 yılında kurulan “sixdegrees.com” olarak kabul edilmektedir. Classmates.com özellikle geçmişteki sınıf arkadaşlarının bulunmasına yönelik bir tema sunarken, SixDegrees.com kullanıcılarına profil oluşturma ve arkadaşlarını listeleme olanağı sunmuştur (Özmen, Aküzüm, Sünkür, & Baysal, 2012).

Günümüze gelindiğinde ise 2021 yılı raporlarına göre Dünya’da 7,83 milyar internet kullanıcı bulunmakta olup 4,20 milyarı sosyal medya kullanıcıdır (Şekil 8).

Şekil 8. We are social 2021 Raporu-Sosyal Medya Kullanımı



Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

2.2.2. Twitter

Jack Dorsey, Noah Glass, Biz Stone, ve Evan Williams tarafından Mart 2006'da oluşturulup Temmuz 2006'da kullanıma açılan sosyal ağ hizmetidir. Kullanıcılarına “tweet” ismi ile adlandırılan içerikleri üretme, paylaşma ve okuma imkânı vermektedir. İnternetin sms olarak adlandırılmakta olup 25 Nisan 2011 tarihi ile Türkçe olarak kullanılabilir hale gelmiştir. Elon Musk tarafından 25 Nisan 2022 tarihinde satın alınmıştır. Kurucu Jack Dorsey tarafından yönetimi gerçekleştirilmektedir (Twitter, 2022) (Twitter Hakkında).

Şekil 9’da görüldüğü üzere we are social tarafından yayımlanan 2021 raporuna göre Twitter Dünya’da en çok trafik alan dördüncü web sitesidir.

Şekil 9. We are social 2021 Raporu-Twitter Trafik

#	WEBSITE	TOTAL VISITS	UNIQUE VISITS	TIME PER VISIT	PAGES PER VISIT	#	WEBSITE	TOTAL VISITS	UNIQUE VISITS	TIME PER VISIT	PAGES PER VISIT
01	GOOGLE.COM	92.21B	3,113M	10M 58S	8.3	11	PORNHUB.COM	3.24B	445M	8M 33S	7.2
02	YOUTUBE.COM	35.75B	1,926M	21M 10S	11.1	12	AMAZON.COM	3.10B	552M	7M 24S	9.6
03	FACEBOOK.COM	25.33B	2,003M	10M 36S	8.3	13	XNXX.COM	3.08B	382M	8M 27S	11.1
04	TWITTER.COM	6.54B	902M	10M 49S	11.9	14	WHATSAPP.COM	3.02B	457M	2M 42S	1.5
05	INSTAGRAM.COM	6.18B	1,009M	7M 45S	10.9	15	NETFLIX.COM	2.66B	261M	9M 54S	4.3
06	WIKIPEDIA.ORG	5.83B	1,148M	3M 55S	3.0	16	LIVE.COM	2.51B	293M	7M 25S	8.2
07	BAIDU.COM	5.70B	260M	6M 15S	8.1	17	YAHOO.CO.JP	2.44B	100M	9M 28S	6.7
08	YAHOO.COM	3.95B	517M	7M 35S	5.8	18	ZOOM.US	2.26B	462M	4M 09S	3.2
09	XVIDEOS.COM	3.75B	479M	10M 13S	8.9	19	VK.COM	1.81B	128M	16M 51S	19.8
10	YANDEX.RU	3.27B	183M	11M 06S	9.0	20	REDDIT.COM	1.74B	236M	9M 11S	6.3

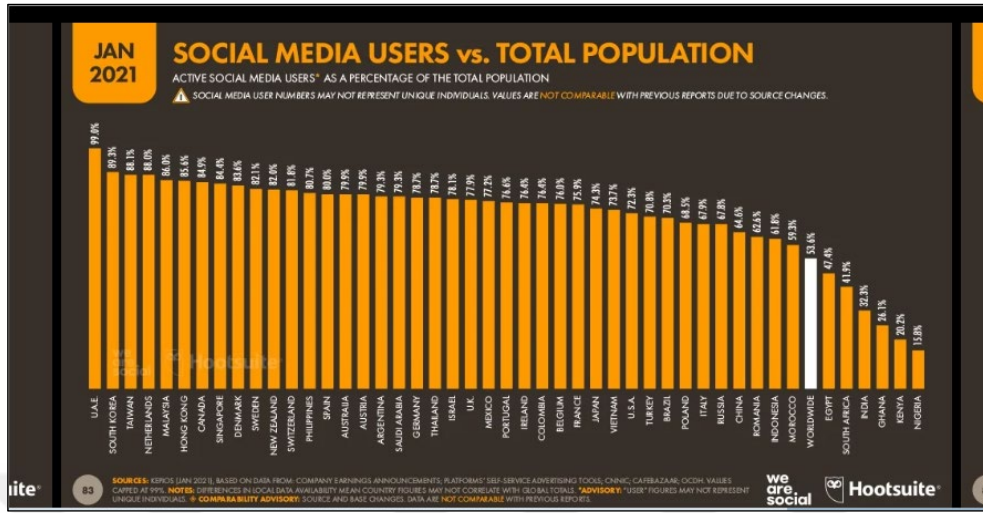
Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

Yayınlanan rapora göre Dünya’da 2021 yılı itibariyle toplam 4,20 milyar sosyal medya kullanıcı bulunmaktadır. Toplam sosyal medya kullanıcısının % 98’si mobil kanal ile sosyal ağ sitelerini kullanmakta olup toplam nüfusun % 99’u ile Birleşik Arap Emirlikleri en çok sosyal medya kullanılan devlettir.

Şekil 10’da devletlerin nüfusunun sosyal medya kullanma oranlarına yer verilmiştir. Twitter Dünya’da en çok zaman harcanan altıncı sosyal medya ortamıdır. Kullanıcılar ayda 5,6 saatini Twitter sosyal medya hesaplarında geçirmektedir.

Şekil 11’de sosyal medya uygulamalarında geçirilen aylık süreler yer verilmiştir. Twitter’e en çok Amerika Birleşik Devletleri, Japonya ve Hindistan’dan erişim sağlanmaktadır. Twitter’in en yoğun 35-49 yaş aralığında kullanıldı, bu yaş aralığındaki kullanıcıların toplam Twitter’in kullanıcılarının % 28,4 oluşturduğu bilinmektedir. 35-49 yaş aralığında kullanıcıların % 19,6’sını kadınlar oluştururken % 8,8’ini ise erkekler oluşturmaktadır.

Şekil 10. We are social 2021 Raporu-Nüfusun Sosyal Medya Kullanma Oranı



Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

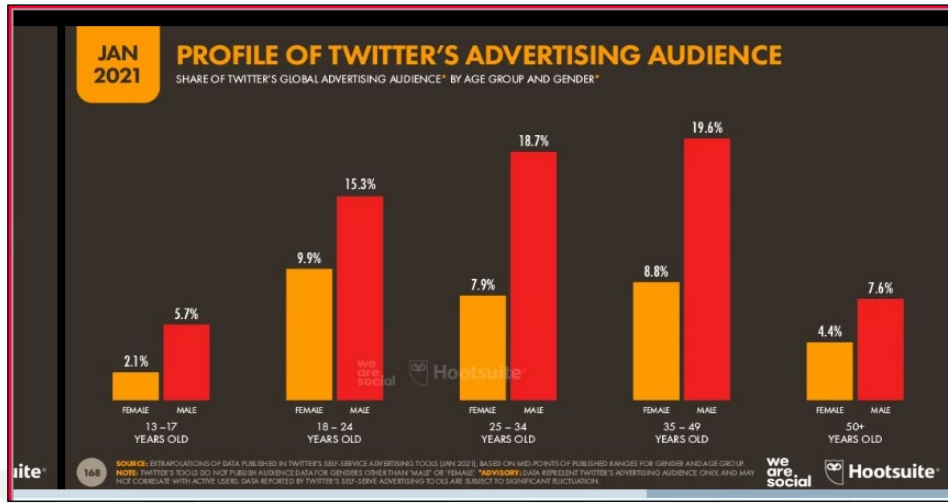
Şekil 11. We are social 2021 Raporu-Nüfusun Sosyal Medya Kullanım Zamanı

JAN 2021 TOP SOCIAL AND VIDEO STREAMING APPS BY TIME SPENT
RANKINGS OF THE TOP SOCIAL MEDIA AND VIDEO STREAMING MOBILE APPS ON ANDROID PHONES, BASED ON TOTAL CUMULATIVE TIME SPENT IN 2020

TOP SOCIAL MEDIA MOBILE APPS BY GLOBAL CUMULATIVE TIME SPENT*			TOP VIDEO STREAMING MOBILE APPS BY GLOBAL CUMULATIVE TIME SPENT*		
#	SOCIAL MEDIA APP	AVE. TIME PER USER	#	VIDEO STREAMING APP	AVE. TIME PER USER
01	FACEBOOK	19.5 HOURS / MONTH	01	YOUTUBE	23.2 HOURS / MONTH
02	WHATSAPP	19.4 HOURS / MONTH	02	MX PLAYER	7.6 HOURS / MONTH
03	INSTAGRAM	10.3 HOURS / MONTH	03	NETFLIX	7.0 HOURS / MONTH
04	TIKTOK	13.3 HOURS / MONTH	04	HOTSTAR	4.5 HOURS / MONTH
05	FACEBOOK MESSENGER	2.7 HOURS / MONTH	05	AMAZON PRIME VIDEO	3.7 HOURS / MONTH
06	TWITTER	5.6 HOURS / MONTH	06	YOUTUBE GO	9.5 HOURS / MONTH
07	LINE	10.6 HOURS / MONTH	07	TWITCH	5.1 HOURS / MONTH
08	TELEGRAM	2.9 HOURS / MONTH	08	JIO TV	2.5 HOURS / MONTH
09	VK	13.9 HOURS / MONTH	09	YOUTUBE KIDS	6.2 HOURS / MONTH
10	WHATSAPP BUSINESS	9.3 HOURS / MONTH	10	VOOT	4.2 HOURS / MONTH

Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

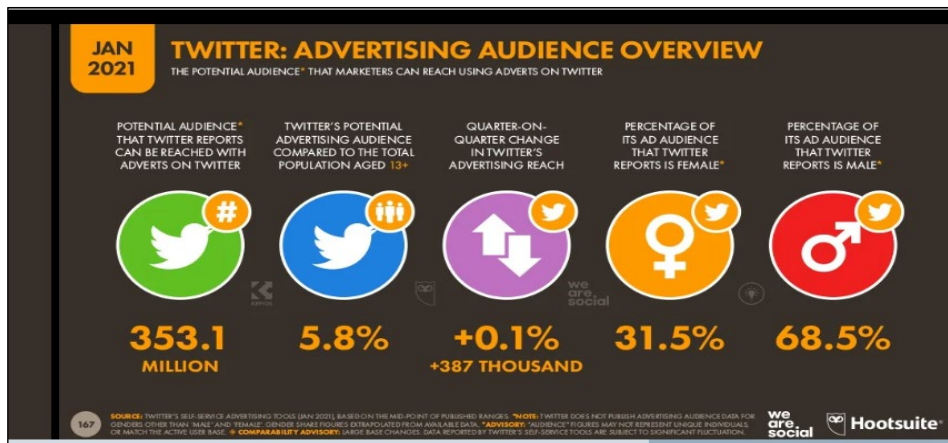
Şekil 12. We are social 2021 Raporu- Twitter Kullanıcıları Yaş ve Cinsiyet Dağılımı



Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

Dünya’da 353,1 milyar Twitter kullanıcı bulunmaktadır. Bu kullanıcıların % 68,5’ini kadınlar oluşturmaktadır. Ülkemiz nüfusu 84,69 milyon olup % 70,8’i ile 60 milyon kişi sosyal medya kullanıcısıdır.

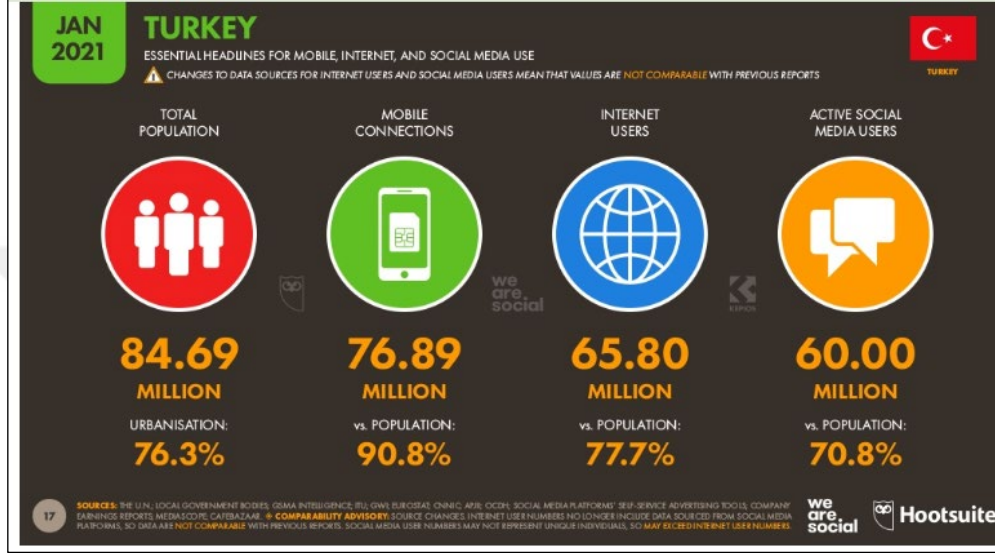
Şekil 13. We are social 2021 Raporu- Twitter Kullanıcıları



Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

Şekil 14'te ülkemiz internet kullanımı ile ilgili verilere yer verilmiştir. Ülkemizde en çok trafik alan web adresleri sıralamasında Twitter, 20'inci sırada yer almaktadır.

Şekil 14. We are social 2021 Raporu- Türkiye

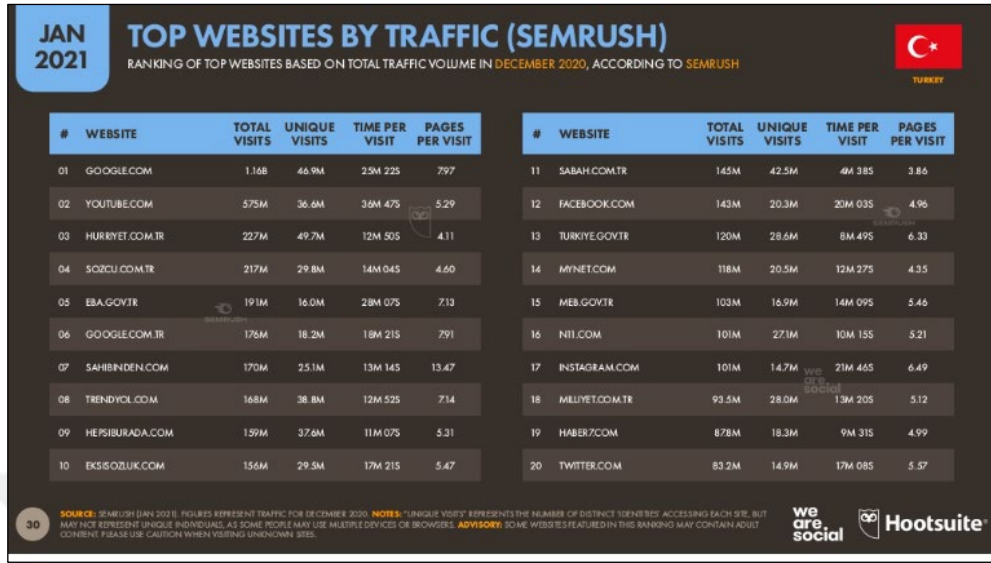


Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

Şekil 15'te Semrush'a göre ülkemizin ilk yirmi trafik sıralamasına yer verilmiştir. Ülkemizde en çok ziyaret edilen birinci website Google.com iken yirminci sırada ise Twitter.com yer almaktadır.

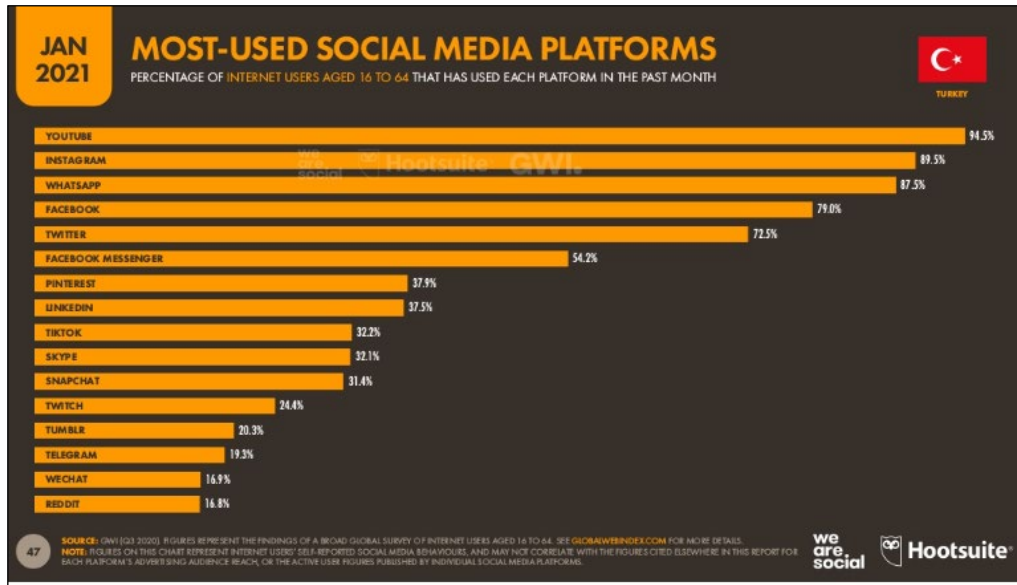
Şekil 16'da ülkemiz sosyal medya kullanım yüzdelere yer verilmiştir. Ülkemizde en çok % 94,5 ile Youtube kullanılırken Twitter'in dördüncü sırada yer aldığı görülmektedir.

Şekil 15. We are social 2021 Raporu- Türkiye Web Site Trafik



Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

Şekil 16. We are social 2021 Raporu- Türkiye Sosyal Medya



Kaynak: <https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>, Erişim Tarihi: Mayıs 2022

ÜÇÜNCÜ BÖLÜM

UYGULAMA

Tezin üçüncü bölümünde, ikinci bölümde kavramsal kısımda tanımlanmış olan metin madenciliği konusu, tez kapsamında gerçekleştirilen alanla ilgili uygulama ile örneklendirilmiştir. Bu uygulama sosyal medya ve metin madenciliği konulu olup konuya dair ayrıntılara ilgili bölüm başlıkları altında yer verilmiştir.

3.1. PROBLEM TANIMI

Beden ya da zihinde meydana gelen rahatsızlık, dert veya görev bozukluğu hastalık olarak tanımlanmaktadır (Hastalık, 2021). Bir toplumdaki hastalığın beklenen sayı üzerine çıkması salgın olarak tanımlanırken hastalığın çok sayıda insanı etkileyen birkaç ülkeye ya da kıtaya yayılması ise pandemi olarak tanımlanmaktadır (Kara, Sertçelik, & Çakır, 2020).

Aralık 2019 tarihinde Çin' in Vuhan eyaletinde ateş ve öksürük ile belirti gösteren bir solunum rahatsızlığı gözlemlenmiştir. 13 Ocak 2020 tarihinde hastalarda yapılan gözlemler sonucunda bu hastalığa sebep olan virüs Covid 19 olarak tanımlanmıştır (COVID-19 Nedir ?, tarih yok). İlerleyen günlerde Dünya'nın farklı bölgelerinde Covid 19 isimli virüs kaynaklı hastaların sayısında hızlı bir artış görülmüş 11 Mart 2020 tarihinde enfeksiyonun görüldüğü ülke sayısı 114, virüs kaynaklı toplam hasta sayısı ise 118.000' den fazla olarak açıklanmıştır. Dünya Sağlık Örgütü (WHO) mevcut durumu dikkate almış 11 Mart 2020 tarihi ile Covid 19 Salgınına pandemi ilan etmiştir (Pandemi, 2020), (WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020, 2020).

Covid 19 virüsü, Dünya Sağlık Örgütünün virüs kaynaklı salgını pandemi ilan ettiği 11 Mart 2020 tarihi ile ülkemizde görülmeye başlamış olup ilk vakanın tespiti dönemin sağlık bakanı Fahrettin Koca tarafından kamuoyuna duyurulmuştur.

Şekil 17. Türkiye’de İlk Covid 19 Vakası Açıklandı



Kaynak: Sağlık Bakanı Koca Türkiye’de ilk koronavirüs vakasının görüldüğünü açıkladı (aa.com.tr), Erişim Tarihi: Mayıs 2022

Ülke içinde salgının, dünyada ise pandeminin devam ettiği ileriki günlerde ülkeler birtakım önlemler planlamışlardır. Bu önlemler uluslararası kültür ve spor müsabakalarının ertelenmesi, eğitim faaliyetlerinin çevrimiçi ortamdaki gerçekleştirilmesi, olağanüstü hâl ilanları ve sokağa çıkma yasakları, ağız ve burun maskesi kullanma zorunluluğu gibi örnekler ile ifade edilebilir. Daha çok kişiler arası mesafe kuralı ile virüsün yayılmasını engellemeyi amaçlayan bu tedbirler insanlar arası iletişimde teknolojik ortam ve araçlara taşınması ve bu ortamlarda iletişimin yaygınlık kazanmasına neden olmuştur.

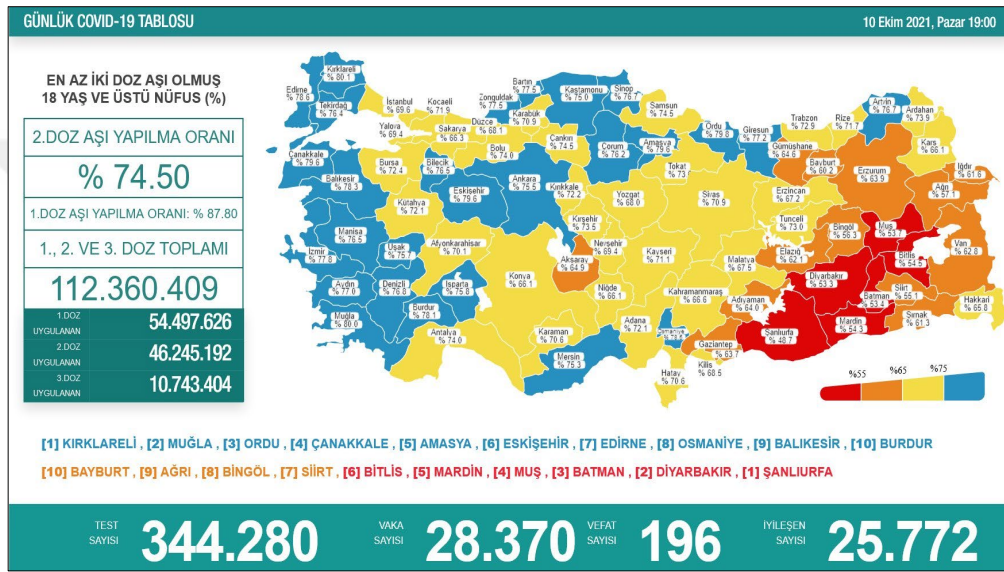
Bu çalışma önceki bölümlerde ele alınan sosyal ağlar ve veri madenciliği kavramları üzerine planlanmıştır. Sosyal ağlarda üretilen içerik üreticisinin tutum ve davranışları hakkında bilgi içerir mi sorusu çalışma kapsamında araştırılmıştır. Bu soru çalışmanın merkezine alınmış ve çalışmanın problemi şu şekilde tanımlanmıştır.

11 Mart 2020 tarihinden itibaren sosyal ağlarda konuları ülkemizin farklı bölgelerinde olan kişiler tarafından Covid 19 konulu birçok içerik üretilmiştir.

Yapılan aşı çalışmaları sonucunda farklı ülkeler tarafından faz 3 çalışmaları gerçekleştirilen iki adet aşı ülkemize getirilmiştir. Sağlık Bakanlığı tarafından virüs ile karşılaşma oranı yüksek meslek gruplarına ve virüs tarafından enfekte olma durumunda alınan hasar oranlarına dikkat edilerek aşı grupları ve aşılama tarihleri planlanmıştır.

Ülkemizde sosyal ağlara resmi kayıt yaşı olan 18 yaş aşırı grubuna aşılama hakkı devlet tarafından ücretsiz olarak tanındığında tarih 25 Haziran 2021 olarak kaydedilmektedir. Çalışmanın taslağının planlandığı 10 Ekim 2021 tarihi aşılama oranları dikkate alındığında Ege Bölgesi aşılama oranlarının yüksek Güneydoğu Anadolu Bölgesi aşılama oranlarının düşük kaydedildiği görülmektedir.

Şekil 18. Türkiye Aşılama Oranları- 10 Ekim 2021



Kaynak: <https://covid19asi.saglik.gov.tr/>, Erişim Tarihi: Mayıs 2022

Şekil 18’ da ülkemiz aşılama oranları gösterilmektedir. Ege Bölgesi aşılama oranlarının % 80 civarında gerçekleşmesi ve Güneydoğu Anadolu aşılama oranlarının % 60 civarında gerçekleşmesi ve bu konulardan sosyal ağlarda üretilen içeriklerin temaları arasında bir ilişkinin varlığı söz konusu mu sorusunu akıllara getirmektedir.

Bu soruya yanıt aramak amacı ile bu çalışma planlanmıştır. Tez kapsamında, yüksek aşılama oranlarının gözlemlendiği Ege Bölgesinden Muğla, İzmir ve Aydın illeri, düşük aşılama oranlarının gözlemlendiği Güneydoğu Anadolu Bölgesinden Şanlıurfa, Batman ve Diyarbakır illeri belirlenmiştir. Bu illerde konumlu kişiler tarafından belirlenen tarihler arasında sosyal ağlarda üretilen içerik ile illerin 10 Ekim 2021 tarihi aşılama oranları arasında herhangi bir birlikteliğin varlığı var mıdır sorusu

sorgulanmıştır. Bu soru ve sorgunun nasıl gerçekleştirileceği sorunu çalışmanın problem tanımını oluşturmaktadır.

3.2. VERİ MADENCİLİĞİ SÜRECİ

Tez kapsamında, 10 Ekim 2021 tarihinde yüksek aşılama oranlarının gözlemlendiği Ege Bölgesinden Muğla, İzmir ve Aydın illeri, düşük aşılama oranlarının gözlemlendiği Güneydoğu Anadolu Bölgesinden Şanlıurfa, Batman ve Diyarbakır illeri bölgeyi temsili iller olarak seçilmiştir. Sosyal ağlarda konumu bu illerden olup konusu Covid 19 olan içerikler veri madenciliğine tutulacak veri seti olarak planlanmıştır.

Verinin hangi sosyal ağ üzerinden toplanacağı, hangi tarihler arasındaki verinin dikkate alınacağı, veri setinin özellikleri ve verinin hangi ortam üzerinden toplanacağı gibi ayrıntılara veri seti ve özellikleri başlığı altında yer verilmiştir.

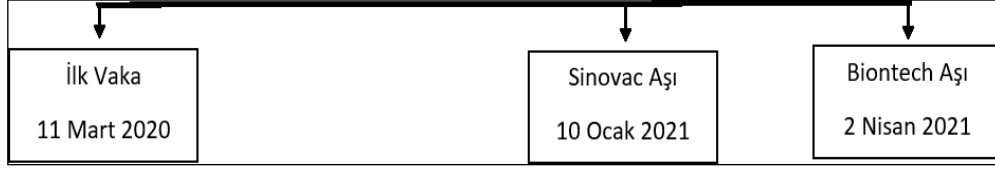
3.2.1. Veri Seti ve Özellikleri

Tez 10 Ekim 2021 Ege Bölgesi aşılama oranlarının yüksek, Güneydoğu Anadolu Bölgesinin aşılama oranlarının düşük olması ile bu bölgelerde konumlu kişiler tarafından sosyal ağlarda üretilen içerikler arası birliktelik vardır hipotezi üzerine kurulmuştur.

Yüksek aşılama oranlarının gözlemlendiği Ege Bölgesinden Muğla, İzmir ve Aydın illeri, düşük aşılama oranlarının gözlemlendiği Güneydoğu Anadolu Bölgesinden Şanlıurfa, Batman ve Diyarbakır illeri temsili iller olarak belirlenmiştir. İllerin belirlenmesinde Ege Bölgesinden en yüksek aşılama oranlarını gösteren üç il, Güneydoğu Anadolu Bölgesinden en düşük aşılama oranlarını gösteren üç il seçimi ölçüt olarak benimsenmiştir.

Veri setinin birinci seçim kriteri belirlenen illerde, konumlu kullanıcılar tarafından sosyal ağlarda üretilen içeriklerin olması gerekliliği olarak belirlenmiştir. Bir diğer aşama ise veri setini oluşturacak içeriklerin hangi tarih aralığı içerisinde olacağı sorusunu cevaplandırmak olmuştur.

Şekil 19. Türkiye Covid 19 Tarihsel Sıralaması



Kaynak: Yazar tarafından derlenmiştir.

Veri setini oluşturan içerikler için Şekil 19’da görüldüğü üzere ülkemizde Covid 19 vakasının ilk kez görüldüğü 11 Mart 2020 tarihi başlangıç tarihi olarak belirlenmiştir. Veri setini oluşturan içerikler için son tarihe karar vermek amacı ile Türkiye Covid 19 Aşı Başlangıç Tarihleri başlıklı Tablo 1 oluşturulmuştur.

Tablo 1. Türkiye Covid 19 Aşı Başlangıç Tarihleri

AŞI GRUPLARI	
GRUPLAR	TARİHLER
A Grubu	14 Ocak 2021
B Grubu	19 Ocak 2021
C Grubu	12 Şubat 2021
50-64 Yaş Grubu	31 Mayıs 2021
40-49 Yaş Grubu	9 Haziran 2021
30-39 Yaş Grubu	
18-39 Yaş Grubu	25 Haziran 2021

Kaynak: Yazar tarafından derlenmiştir.

Tablo 1’de görüldüğü üzere resmi olarak ülkemizde sosyal ağları kullanma hakkının doğduğu 18 yaş ve üzeri kullanıcılar 25 Haziran 2021 tarihi ile aşılma hakkını elde etmiştir. Bu sebeple uygulamaya konu olacak veri Covid 19 konulu ve temsili seçilen iller konumlu, 11 Mart 2020 ile 25 Haziran 2021 tarihleri arasında üretilen sanal içerikler olarak tekrar tanımlanmıştır.

Tez içerisinde uygulamanın bir diğer konusunu sanal içeriklerin hangi sosyal ağ üzerinden toplanacağı sorusu oluşturmaktadır. Bu soruya yanıt bulmak amacıyla çalışmanın önceki bölümlerinde bahsi geçen kaynak taraması gerçekleştirilmiş, Dünya’ da ve ülkemizde kullanılan sosyal ağlar sıralaması listesine ulaşılmıştır. Listeyi Facebook Whatsapp, Instagram ve Twitter gibi takip eden bir sıralamanın oluşturduğu görülmüştür. Ancak Whatsapp uygulamasının tam anlamıyla sosyal ağlar kavramına karşılık gelmemesi, Instagram’ın daha çok görsel içerik üretimi gibi nedenlerle Facebook ve Twitter sosyal mecraları üzerine yoğunlaşmıştır. Twitter’in kendine has yapısı, Covid 19 konusunun içerdiği ekonomik ve politik konuların daha rahat ifade edilmesine imkân vermektedir. Bu neden ile veri setine ulaşmak için Twitter sosyal mecrası üzerine odaklanılmıştır.

Uygulama için veri seti içerisindeki verinin sınırları tekrar tanımlanmıştır. Bu tanım Covid 19 konulu ve temsili seçilen iller konumlu, 11 Mart 2020 ile 25 Haziran 2021 tarihleri arasında Twitter sosyal mecrasında üretilen sanal içerikler olarak ifade edilebilir. Bu tanım ile uygulamada yeni bir aşamaya geçilmiştir. Bu aşama ise Twitter üzerinden belirlenen özellikteki içeriklerin toplama faaliyetlerini kapsamaktadır.

Twitter üzerinden veri çekebilmek amacı ile Twitter Developer hesabın açılması ve API alınması veri toplamada ilk yöntem olarak benimsenmiştir. Ancak yapılan başvurular ve yazışmalar sonuçsuz kalmış Twitter Developer hesabı aktifleştirilememiştir.

Veri toplamak amacı ile Twitter’den tweet çekme imkânı veren farklı ortamlar araştırılmıştır. Alanda yapılan çalışmalar incelendiğinde Workbench isimli çevrimiçi ortamında Tweet çekmek amacıyla kullanıldığı bilgisine ulaşılmıştır. Bu bilgi kapsamında Workbench ortamı incelendiğinde belirli kullanıcı, belirli konu ve listeden tweet çekmenin mümkün olduğu gözlenmiştir. Workbench kullanılarak Türkçe tweetlerin gelmesi amacıyla koronavirus konulu tweetler çekilmiş ancak Workbench

sadece son 1.000 adet tweet'i çekmesi nedeniyle geçmişe dönük tweetler'e ulaşamamıştır.

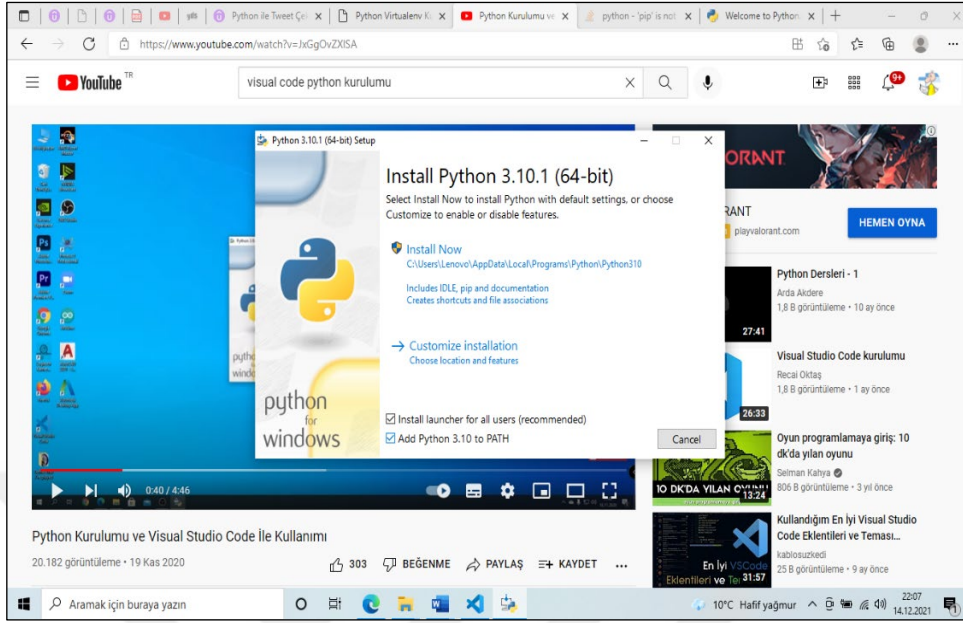
Tweet çekme imkânı veren farklı ortamların araştırılmasına devam edilmiş yine çevrimiçi bir ortam olarak Google ait TAGS uygulaması kullanılmıştır. Google E-Posta ve Drive ortamının kullanıldığı TAGS çevrimiçi uygulaması ise Twitter' den bir kullanıcı ya da konunun tweetlerini çekme imkanını vermektedir. Ancak Workbench ortamından adet ile fazla olmasına karşın bu ortamdan çekilebilecek Tweet sayısı da 18.000 adet ve son 7 gün ile sınırlandırılmaktadır. Bu sebeple uygulamanın kullanılması ile çekilen tweetlerin birçoğu Sn. Bakan Fahrettin Koca' ya aittir.

Uygulamanın geldiği bu aşama esnasında tez çalışmaları kapsamında Danışman hocam ve ekibiyle toplantı gerçekleştirilmiştir. Bu toplantı esnasında ekip arkadaşlarından biri aracıyla Python ile API'siz tweet çekme imkanından haberdar olunmuş ve veri toplamak amacıyla çalışma Python programı üzerine yoğunlaştırılmıştır.

Python programının kullanılması amacı ile Microsoft Visual Code programının kurulumu gerçekleştirilmiştir. Ardından Python uygulamasının kurulumu gerçekleştirilmiştir. Python programının kullanımı amacıyla Visual Code uygulamasına Python eklentisi kurulmuştur. Microsoft Visual Code uygulamasına Python eklentisinin kurulumunun ardından kodlama için gerekli kütüphanelerin kurulması aşamasına geçilmiştir.

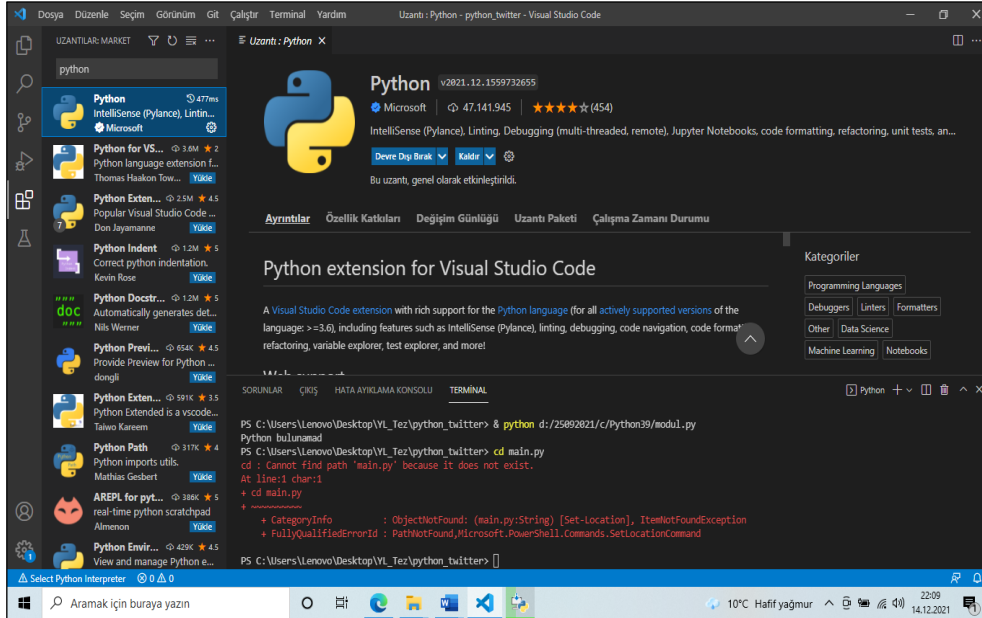
Şekil 20'de Microsoft Visual Code ve Python kurulumu ekran görüntüsüne yer verilmiştir. Şekil 21'de Microsoft Visual Code-Python Eklentisi Yükleme ekran görüntüsü gösterilmektedir. Şekil 22'de ise Snsrape Kütüphanesi kurulumu ekran görüntüsü bulunmaktadır.

Şekil 20. Microsoft Visual Code ve Python Kurulumu



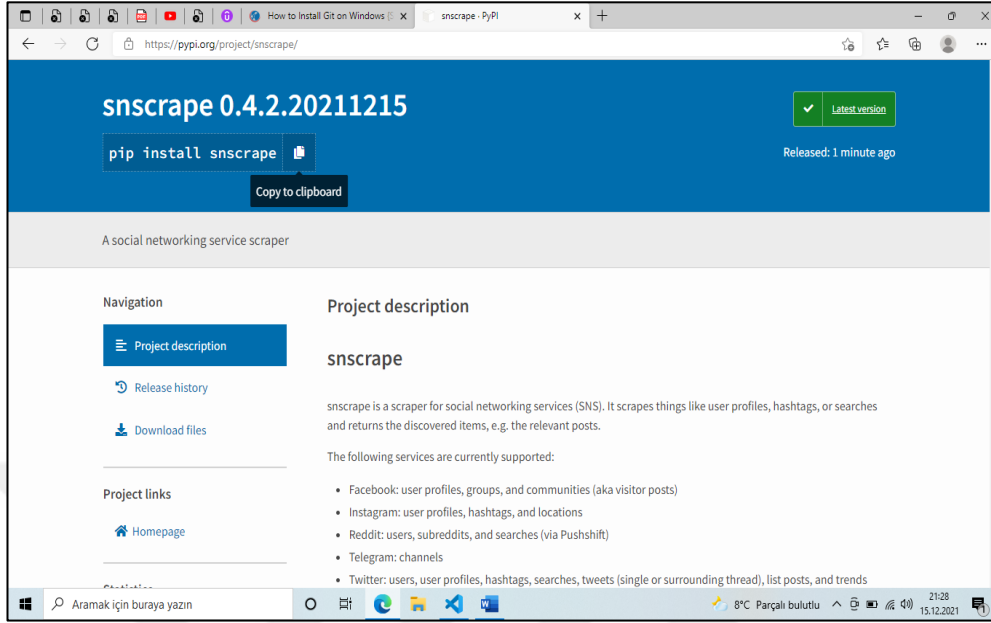
Kaynak: Yazar tarafından derlenmiştir.

Şekil 21. Microsoft Visual Code-Python Eklentisi Yükleme



Kaynak: Yazar tarafından derlenmiştir.

Şekil 22. Snsrape Kütüphanesi Kurulumu

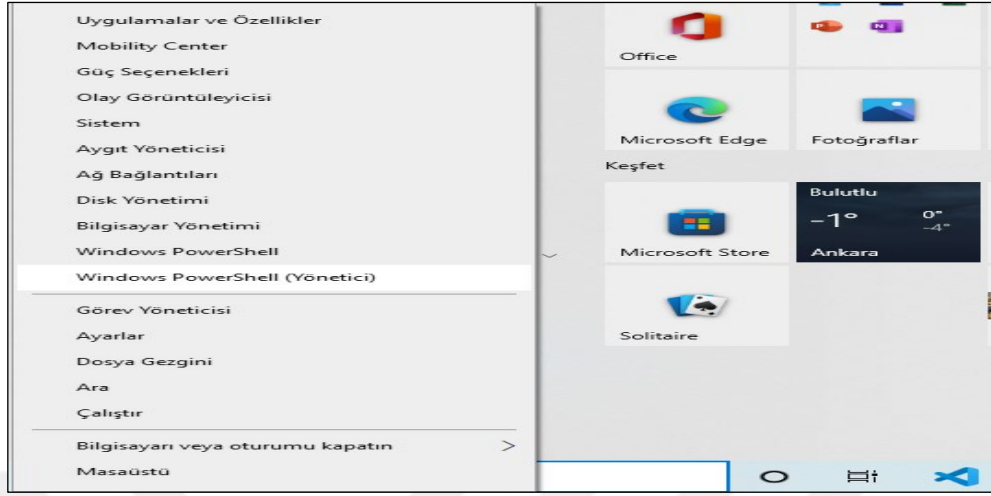


Kaynak: Yazar tarafından derlenmiştir.

Şekil 22' te görülen snsrape kütüphanesi, sosyal ağlar için bir kazıyıcı olarak tanımlanmaktadır. Kullanıcı profilleri, hashtag'ler veya aramaları kazıyıp, keşfedilen öğeleri döndürmektedir. Facebook, Instagram, Mastodon, Reddit, Telegram, VKontakte, Weibo (Sina Weibo) ve Twitter sosyal ağları üzerinden veri kazıma işleminde kullanılmakta ve sosyal ağlarda farklılaşan öğeler iade vermektedir. Twitter sosyal ağı üzerinden snsrape kullanıcılar, kullanıcı profilleri, hashtag'ler, aramalar, tweetler (tek veya çevresindeki konu başlığı), liste gönderileri ve trendler öğeleri veri kazıma işlemi sonrası kazanılmaktadır (snsrape 0.4.3.20220106, 2022).

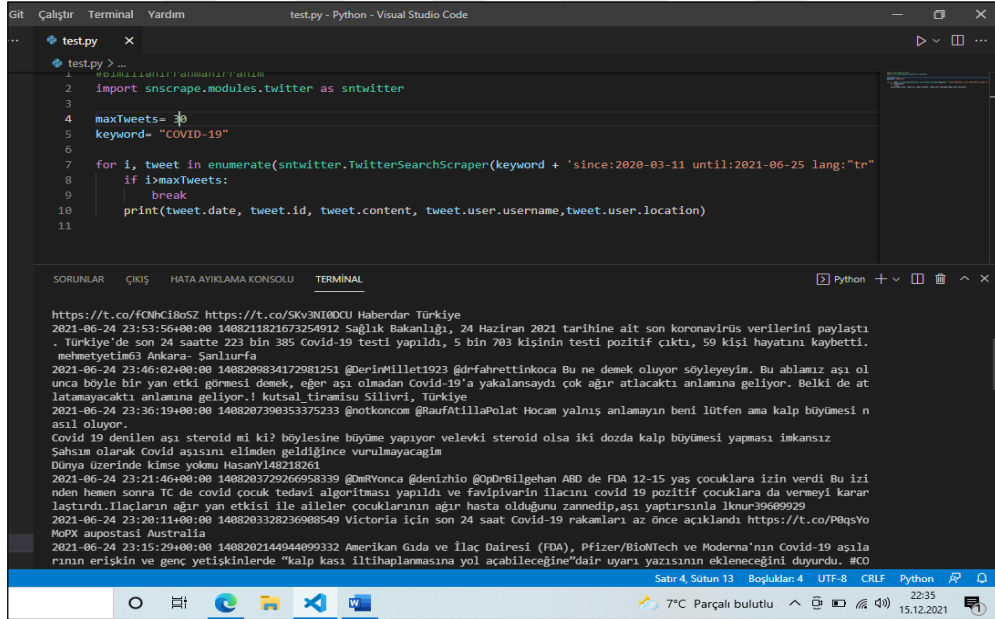
Python programı üzerine snsrape kütüphanesi kurulumu için Şekil 23'te görülen pypi.org adresine girilmiş snsrape kütüphanesi için paylaşılan pip install snsrape metni kopyalanmıştır. Kopyalanan metin başlat menüsü, sağ tık Windows PowerShell(Yönetici) seçeneğinin seçilmiş ve gelen PowerShell(Yönetici) ekranına yapıştırılmıştır. Kurulum bu adımlar sonrası gerçekleştirilmektedir. Şekil 23' ün ardından Şekil 24' teki adım gerçekleştirilmektedir. Şekil 24'te Python programında snsrape ile Twitter'den veri kazıma sonuçları gösterilmektedir.

Şekil 23. Başlat Windows PowerShell Adımı



Kaynak: Yazar tarafından derlenmiştir.

Şekil 24. Snsrape ile Veri Kazıma



Kaynak: Yazar tarafından derlenmiştir.

Bu adımdan itibaren çalışma iki kola ayrılmıştır. Birinci kol Snsrape kütüphanesinin kullanılarak veri kazınması işlemini içermektedir. İkinci kol ise snsrape kütüphanesi tarafından Covid 19 konulu ve temsili seçilen iller konumlu, 11 Mart 2020 ile 25 Haziran 2021 tarihleri arasında Twitter sosyal mecrasında üretilen sanal içerik nitelikleri ile iade edilen verinin seçilen veri yönetim sistemi programına aktarılmasıdır. İkinci kol ile ilgili ayrıntılara ve çalışmaya ait kodun bütüne metnin devamında yer verilecektir. Öncelikle veri kazıma için snsrape kütüphanesinin kullanıldığı kod parçasına değinilecektir.

Şekil 25. Snsrape İle Veri Kazıma

```
import snsrape.modules.twitter as sntwitter
maxTweets= 12
keyword= "COVID-19"
i=1
for i, tweet in enumerate(sntwitter.TwitterSearchScrapper(keyword +
'near:"Batman" within:10km since:2020-03-11 until:2020-03-31 lang:"tr"
').get_items()):
    if i>maxTweets:
        break
    print(tweet.content, tweet.likeCount, tweet.date, tweet.user.username,
tweet.user.location, tweet.id)
```

Kaynak: Yazar tarafından derlenmiştir.

Şekil 25 isimli tabloda metnin bir önceki kısmında yükleme işlemi adımlarına yer verilen snsrape kütüphanesinin kullanımı gösterilmektedir. Yükleme işlemi gerçekleştirilen snsrape kütüphanesi kullanılmak üzere test.py dosyası üzerine import edilmekte ve sntwitter ismi ile adlandırılmaktadır. Veri kazıma işlemi sonucunda iade edilmesi istenen en yüksek içerik sayısı maxTweets değişkeni ile 12 olarak belirlenmektedir. COVID-19 olarak tanımlanan anahtar sözcüğünü içeren Batman konumlu, 11 Mart 2020 ve 31 Mart 2021 tarihlerinde üretilen tweetler for döngüsü ile taranmaktadır. Ardından print komutu içerisinde sırasıyla tweet(içerik), tweet beğeni

sayısı, tweet tarihi, tweet'in üreticisi olan kullanıcı ismi ve tweet'in atıldığı konunun for döngüsünün sınırlandırıldığı en çok 12 sayısına ulaşıldığında geri döndürülmesi istenmektedir. Bu kod parçası Python ile API'siz tweet çekmeyi öğrenmek amacı ile tanımlanmış örnek bir koddur ve Şekil 25'te terminal kısmında görüldüğü üzere kod işlemi başarı ile gerçekleştirmiştir. Kod çalışması sorun tanımına uygun olarak tasarlanmış ve istenen veri seti kümesi için çalışmanın ileriki aşamalarında tekrar düzenlenmiştir.

Bu aşamadan sonra uygulamanın kodlama aşaması için planlanan ikinci kısma geçilmiştir. Bu çalışma çekilen tweetlerin seçilen veri tabanı yönetim sistemine aktarılması işleminden oluşmaktadır.

Python kodlama dili ile program üzerine çekilen verileri, veri tabanı üzerinde saklamak amacıyla MS SQL Server ve MySQL veri tabanı yönetim sistemleri değerlendirilmiş farklı sebepler nedeniyle MySQL veri tabanı yönetim sisteminin kullanılmasına karar verilmiştir.

MySQL, 1995 yılında kullanıma sürülen açık kaynaklı ilişkisel veri tabanı yönetim sistemidir. İsmi ana geliştiricilerinden Michael Widenius'un kızının adı My ve SQL kelimelerinin birleşiminden almaktadır. Web sunucularında en çok kullanılan veri tabanlarından biri olan MySQL, asp, php gibi web programlama dilleri ile birlikte kullanılabilir (MySQL Nedir?, 2020).

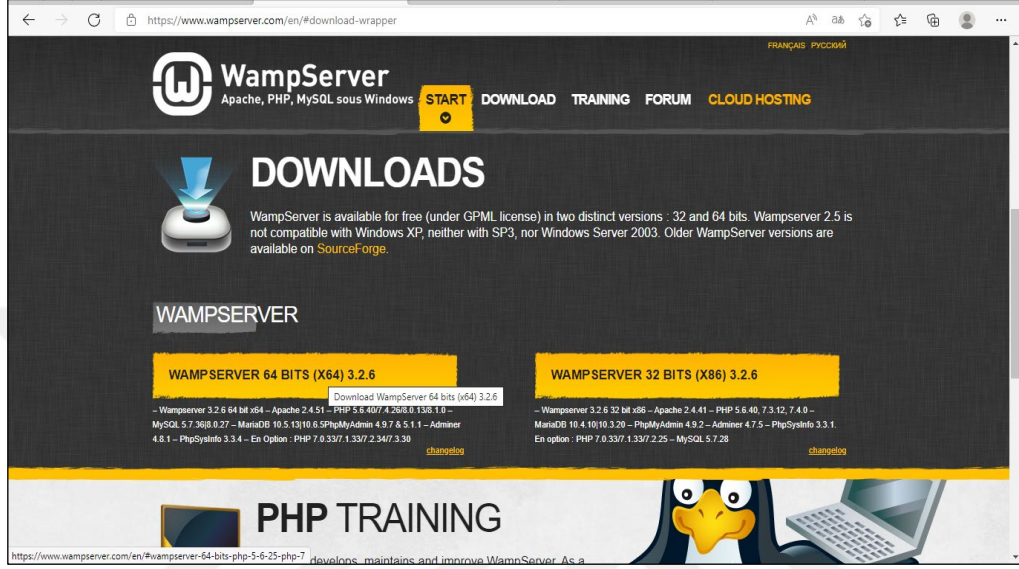
Tez kapsamında MySQL dilini web üzerinden kullanmak amacı ile phpMyAdmin kullanılmıştır. phpMyAdmin, MySQL' in Web üzerinden yönetimini ele almayı amaçlayan PHP ile yazılmış özgür bir yazılım aracı olarak tanımlanmaktadır. phpMyAdmin, MySQL ve MariaDB' de çok çeşitli işlemleri desteklemektedir. (Bringing MySQL to the web About)

Tezin bu kısmında PhpMyAdmin aracı kullanılarak localhost üzerinde bir veri tabanı kurulmuştur. Ardından kurulan veri tabanı ile MS Visual Studio programı arasında bağlantı gerçekleştirilmiş, python programlama dili ile çekilen tweetler bağlantısı gerçekleştirilen veri tabanının ilgili tablolarına aktarılmıştır. Öncelikle PhpMyAdmin' i kullanmak için WampServer programının kurulumu gerçekleştirilmiştir.

WampServer bir Windows geliştirme ortamı olarak tanımlanmakta, Apache2, PHP, MySQL veri tabanları ile web uygulamaları oluşmamıza olanak tanımakta ve

PhpMyAdmin, veri tabanları yönetimini kolaylaştırmaktadır (WampServer, a Windows web development environment.).

Şekil 26. WampServer Kurulumu

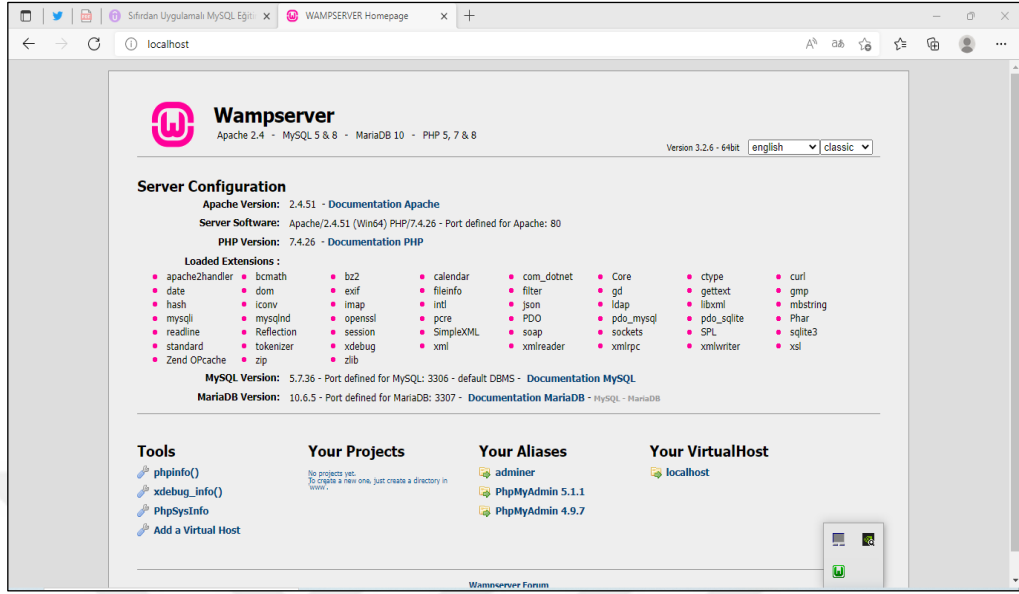


Kaynak: Yazar tarafından derlenmiştir.

WampServer resmi sayfasından işletim sistemine uygun olarak 64 bitlik WampServer kurulumu gerçekleştirilmiştir.

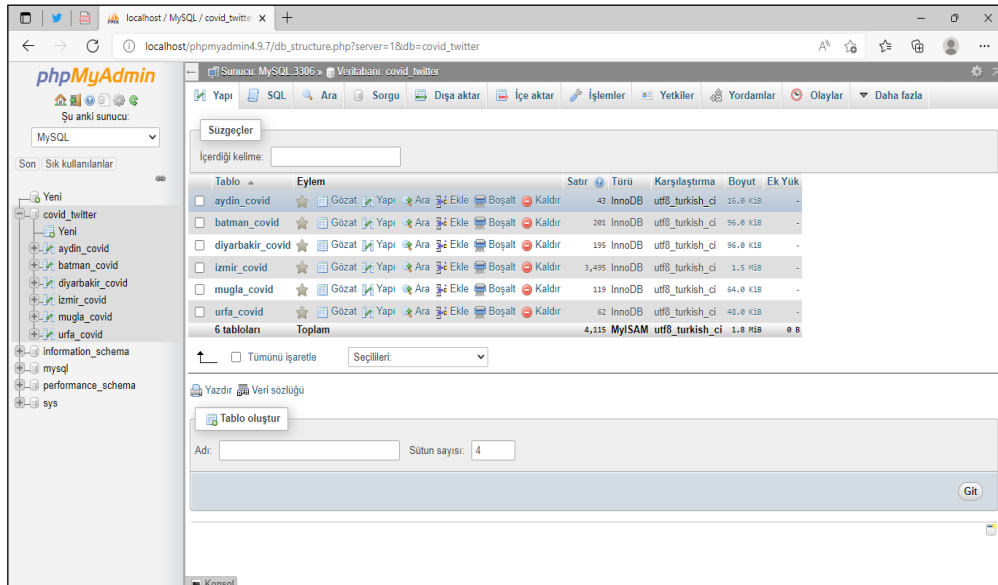
WampServer kurulumunun ardından WampServer başlatıldığında Şekil 27' de görüldüğü simge menüsündeki W işareti ile ifade edilen WampServer servisleri eksiksiz olarak başladığında yeşil renk almaktadır. Yeşil ikonun görülmesinin ardından tarayıcıdan localhost kısmına geçiş mümkündür. Kurulum esnasında değişikliğe gidilmemiş ise root varsayılan kullanıcı ismi ve şifresiz giriş ile phpMyAdmin kısmına ulaşılmaktadır.

Şekil 27. WampServer Servisler Başarılı



Kaynak: Yazar tarafından derlenmiştir.

Şekil 28. WampServer- Veri Tabanı Kurulumu

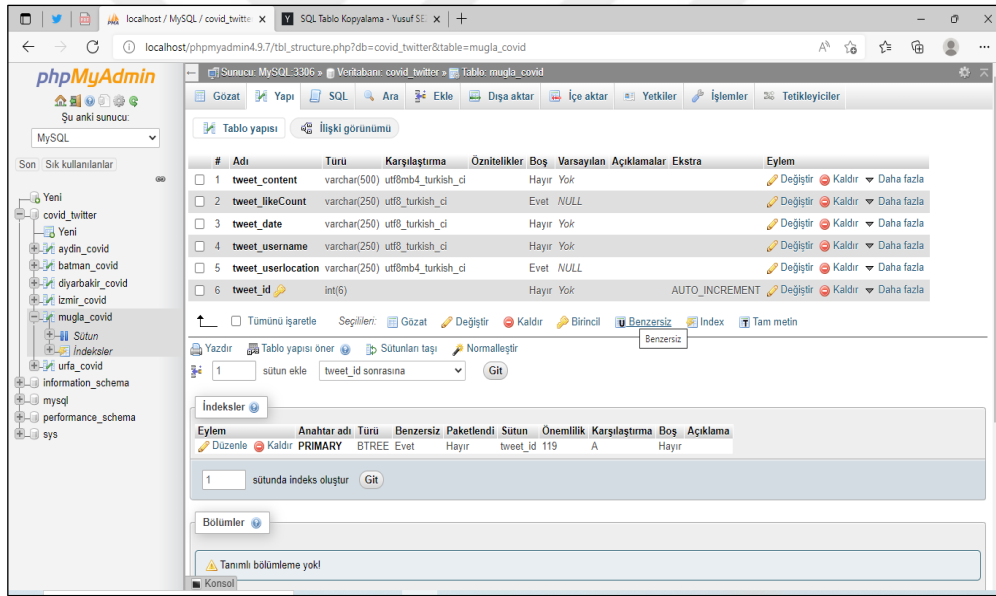


Kaynak: Yazar tarafından derlenmiştir.

Şekil 28’ de kurulan veri tabanı ve veri tabanına ait tablolar görülmektedir. Localhost üzerine covid_twitter isimli veri tabanı kurulmuştur. Problem tanımında Güneydoğu Anadolu Bölgesi ve Ege Bölgesi için seçilmiş temsili altı il ismi ile ilgili konumlardan çekilen Twitter verilerini kaydetmek için tablolar veri tabanına eklenmiştir.

İl tablolarının sütunları bire bir aynı olmakla birlikte tabloların üretilmesinde MySQL tablo çoğaltma yöntemi kullanılmıştır. Taslak mugla_covid isimli tablo ile oluşturulmuş ardından çoğaltılarak aydin_covid, izmir_covid, batman_covid, diyarbakir_covid ve urfa_covid isimleri ile adlandırılmıştır. Tablo sütunlarının özellikleri Şekil 29 ile gösterilmektedir.

Şekil 29. Sütun Özellikleri



#	Adı	Turu	Karşılaştırma	Özellikler	Boş	Varsayılan	Açıklamalar	Ekstra	Eylem
1	tweet_content	varchar(500)	utf8mb4_turkish_ci		Hayır	Yok			Değiştir Kaldır Daha fazla
2	tweet_likeCount	varchar(250)	utf8_turkish_ci		Evet	NULL			Değiştir Kaldır Daha fazla
3	tweet_date	varchar(250)	utf8_turkish_ci		Hayır	Yok			Değiştir Kaldır Daha fazla
4	tweet_username	varchar(250)	utf8_turkish_ci		Hayır	Yok			Değiştir Kaldır Daha fazla
5	tweet_userlocation	varchar(250)	utf8mb4_turkish_ci		Evet	NULL			Değiştir Kaldır Daha fazla
6	tweet_id	int(6)			Hayır	Yok		AUTO_INCREMENT	Değiştir Kaldır Daha fazla

İndeksler

Eylem	Anahtar adı	Turu	Benzersiz	Paketlendi	Sütun	Önemlilik	Karşılaştırma	Boş	Açıklama
Düzenle Kaldır	PRIMARY	BTREE	Evet	Hayır	tweet_id	119	A	Hayır	

Bölümler

Tanımlı bölümler yok!

Kaynak: Yazar tarafından derlenmiştir.

Şekil 29’da görüldüğü üzere tablolar içerikten oluşan tweet_content, tweet beğeni sayısını ifade eden tweet_likeCount, içeriğin paylaşım tarihini ifade eden tweet_date, daha sonraki süreçte etik kurallar gereği çalışmadan dışlanan kullanıcı ismi tweet_username, tweetlerin paylaşıldığı konumu ifade eden tweet_location ve veri tabanı yönetim sistemi tarafından otomatik olarak atanan artan sayıdan

oluşmaktadır. Python üzerinden verilerin aktarılması esnasında yaşanan sorun nedeniyle birincil anahtar olarak tanımlanan tweet_id hariç sütunların tamamı varchar olarak tanımlanmıştır. Karşılaştırma, Türkçe karakterlerimize uygunluk bakımından utf8_turkish_ci olarak belirlenirken boyut sorunu nedeniyle tweet_content ve tweet_userlocation utf8mb4_turkish_ci olarak belirlenmiştir. Tweet'in beğeni almaması durumu değerlendirilerek ve konunun paylaşılmaması durumu dikkate alınarak tweet_likeCount ve tweet_location boş geçilebilir kabul edilmiştir. Belirlenen özellikler ile tablolar oluşturulmuş ve çalışmanın veri tabanı kısmı tamamlanmıştır.

Localhost üzerinde oluşturulan veri tabanı ve tabloları Microsoft Visual Code üzerinden görebilmek amacı ile Microsoft Visual Code programı veri tabanı sekmesinden covid_twitter veri tabanı ile bağlantı gerçekleştirilmiştir. Şekil 30' da Yönetimini WampServer ile gerçekleştirdiğimiz veri tabanına Visual Studio bağlantısı gösterilmektedir.

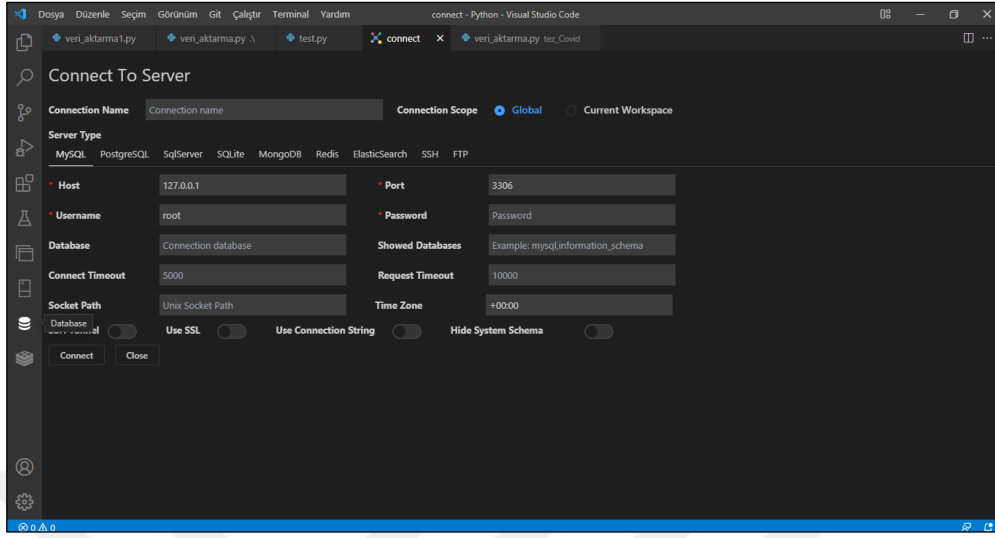
Şekil 31'de sol menüde covid_twitter veri tabanı tablo ve sütunları görüntülenmektedir.

Veri tabanı ile gerçekleştirilen bağlantının ardından veri tabanına veri aktarma gibi sql işlemlerinin yapılabilmesi için eklentiler kısmından Visual Studio Code programına Sql Server Client(mssql) eklentisi yüklenmiştir.

Şekil 32' de Sql Server Client eklentisinin Visual Studio Code' ye yüklenmesi gösterilmektedir.

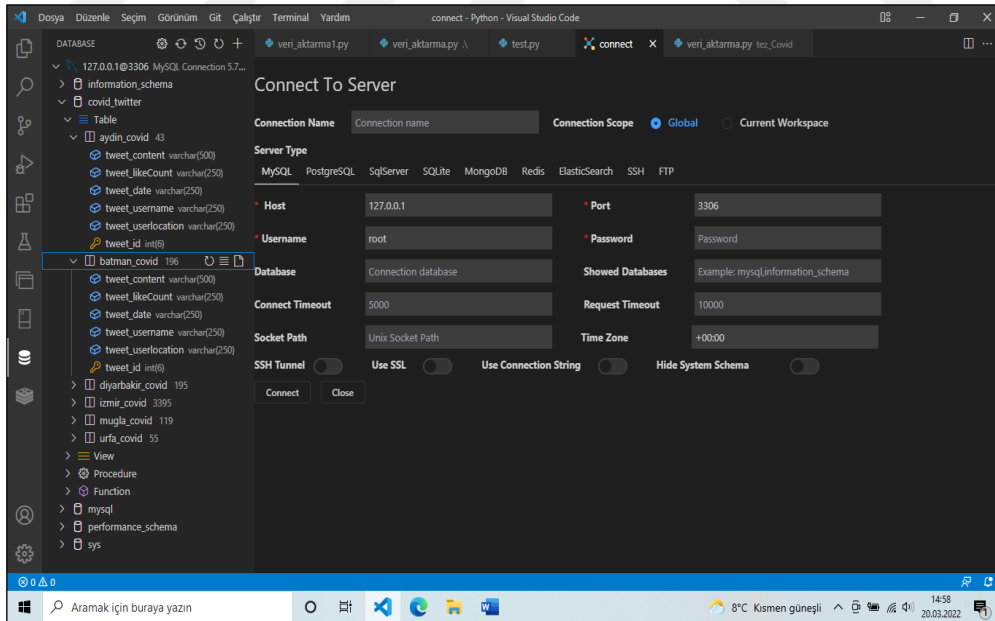
Python kodu tekrar düzenlenerek son haline getirilmiştir. Kod Şekil 33' de gösterilmekte ve ayrıntılarına yer verilmektedir.

Şekil 30. Visual Studio ile MySQL Bağlantısı



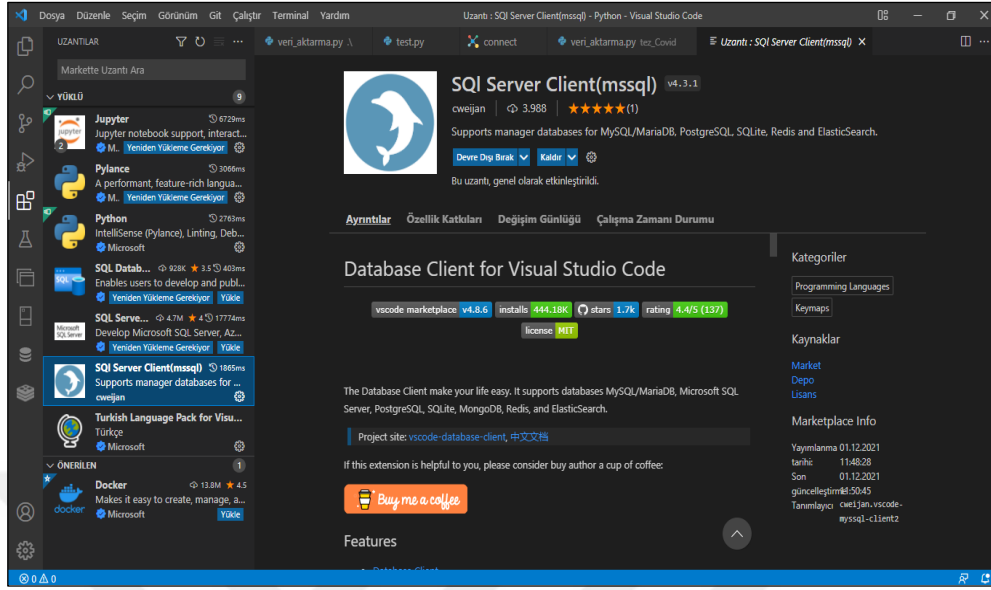
Kaynak: Yazar tarafından derlenmiştir.

Şekil 31. Visual Studio İle Veri Tabanı Bağlantısı Başarılı



Kaynak: Yazar tarafından derlenmiştir.

Şekil 32. Microsoft Studio Code- SQL Server Client(mssql) Yükleme



Kaynak: Yazar tarafından derlenmiştir.

Şekil 33' de Python ile API'siz tweet çekme ve MySQL veri yönetim sistemine aktarma işlemi için yazılan kodun son hali görülmektedir. Şekil 33 adlı tabloda verilmiş olan kod ile her il ve ay için for döngüsündeki ilgili kısım değiştirilmeye birlikte Batman, Şanlıurfa, Diyarbakır, Aydın, İzmir ve Muğla konumlu Covid 19 konulu 11 Mart 2020 ile 25 Haziran 2021 tarihleri arasında üretilen tweetleri çekmek amaçlanmıştır.

Tez için ilgili illerden olmak suretiyle 1.000.000 Covid 19 konulu tweeti kazımak hedeflenmiştir. Veri seti içerisinde adil dağılımı sağlamak ve tamsayıyı yakalamak amacıyla $968.832/6(\text{il})=161.472$ formülü kullanılmış ve her il için 11 Mart 2020 ile 25 Haziran 2021 tarihleri arasında 161.472 tweet hedeflenmiştir. $161.472/16(\text{Ay})=10.092$ formülü kullanılarak bir il için aylık 10.092 maxTweets sayısına ulaşılmıştır.

Db ve bağlantı değişkenleri ile covid_twitter isimli veri tabanı ile bağlantı gerçekleştirilmiştir. For döngüsü içerisinde veri kazıma işleminde dikkate alınacak anahtar kelime, konum, dil ve tarihler tanımlanmaktadır.

Şekil 33. Veri Kazıma ve Veri Tabanına Aktarma

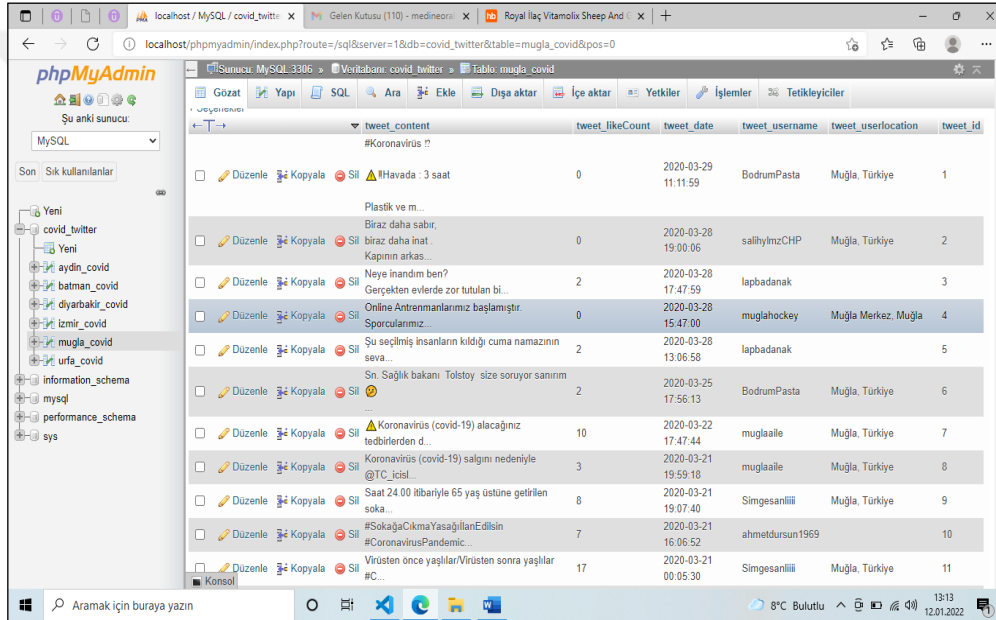
```
import sys
sys.path.insert(0,"C/python31/lib/site-packages")
import snsrape.modules.twitter as sntwitter
import mysql.connector
maxTweets= 10092
keyword= "COVID-19"
i=1
db = mysql.connector.connect(host='localhost',
database='covid_twitter', user='root', password='')
baglanti = db.cursor()
for i, tweet in enumerate(sntwitter.TwitterSearchScrapper(keyword
+ 'near:"Muğla" within:10km since:2021-06-01 until:2021-06-25 lang:"tr"
').get_items()):
    if i>maxTweets:
        print(str(i) + " kullanıcı eklendi")
        break
        print(tweet.content, tweet.likeCount, tweet.date,
tweet.user.username, tweet.user.location, tweet.id)
        d_content=str(tweet.content)
        sonuc = baglanti.execute('Insert Into
mugla_covid(tweet_content, tweet_likeCount, tweet_date,
tweet_username,
tweet_userlocation)VALUES(%s,%s,%s,%s,%s)',(d_content,
tweet.likeCount, tweet.date, tweet.user.username, tweet.user.location))
        print(str(i) + " kullanıcı eklendi")
        db.commit()
```

Kaynak: Yazar tarafından derlenmiştir.

Print kısmında, kazınan veriler kullanıcıya gösterilirken sonuç değişkeni ile çalıştırılan sql deyimi ile kazınan tweetler veri tabanına aktarılmaktadır.

Yukarıda bahsi geçen işlemlerin ardından veri kazıma işlemi tamamlanmıştır. Veri tabanına aktarılan veriler farklı türlerde dışarıya aktarılmıştır. Veriler Excel formatında da aktarılarak ve SPSS programında değerlendirilmiştir. Çalışmada kullanılmak üzere 1.000.000 veri kazımak hedeflenmiş olup konum kısıtı nedeniyle hedeflenen sayıya ulaşamamıştır.

Şekil 34. Kazınan Tweetler' i Veri Tabanına Yükleme



	tweet_likeCount	tweet_date	tweet_username	tweet_userlocation	tweet_id
#Koronavirüs !!	0	2020-03-29 11:11:59	BodrumPasta	Muğla, Türkiye	1
Plastik ve m...	0	2020-03-28 19:00:06	salihlymzCHP	Muğla, Türkiye	2
Biraz daha sabır, biraz daha inat Kapının arkas...	2	2020-03-28 17:47:59	lapbadanak		3
Neye inandım ben? Gerçekten evlerde zor tutulan bi...	0	2020-03-28 15:47:00	muğlahockey	Muğla Merkez, Muğla	4
Online Antrenmanlarımız başlamıştır. Sporcularımız...	2	2020-03-28 13:06:58	lapbadanak		5
Su seçilmiş insanların kıldığı cuma namazının seva...	2	2020-03-25 17:56:13	BodrumPasta	Muğla, Türkiye	6
Sn. Sağlık bakanı Tolstoy size soruyor sanırım	10	2020-03-22 17:47:44	muğlaaile	Muğla, Türkiye	7
Koronavirüs (covid-19) alacağınız tedbirlerden d...	3	2020-03-21 19:59:18	muğlaaile	Muğla, Türkiye	8
Koronavirüs (covid-19) salgını nedeniyle @TC_icsi...	8	2020-03-21 19:07:40	Simgesanlıi	Muğla, Türkiye	9
Saat 24.00 itibarıyla 65 yaş üstüne getirilen soka...	7	2020-03-21 16:06:52	ahmetdursun1969		10
#SokağaÇıkmaYasağilanEdilsin #CoronavirusPandemic...	17	2020-03-21 00:05:30	Simgesanlıi	Muğla, Türkiye	11

Kaynak: Yazar tarafından derlenmiştir.

Veri seti içerisindeki veriler bir sonraki bölümde anlatılmış olan bir takım ön işlemleri tabi tutulmuştur. Tweet_userlocation sütunu için ön işleme gerçekleştirildiğinde kazınan 458 veriden 100 tanesi konum sütununda problem tanımı ile yaşanan uyumsuz nedeniyle çalışmadan dışlanmıştır.

Güneydoğu Anadolu Bölgesi için eğerlendirmeye alınan 358 tweetten, 137 tanesi Batman, 161 tanesi Diyarbakır ve 60 tanesi Şanlıurfa konumlarından 11 Mart 2020 ile 25 Haziran 2021 tarihleri arasında üretilerek, Twitter sosyal ağında paylaşılmıştır.

Şekil 35. Güneydoğu Anadolu Bölgesi Veri Seti İl Dağılımı

	Frekans	Yüzde
Batman	137	38,3
Diyarbakır	161	45,0
Şanlıurfa	60	16,8
Total	358	100,0

Kaynak: Yazar tarafından derlenmiştir.

Şekil 36. Ege Bölgesi Veri Seti İl Dağılımı

	Frekans	Yüzde
İzmir	3501	95,6
Aydın	119	3,2
Muğla	43	1,2
Total	3663	100,0

Kaynak: Yazar tarafından derlenmiştir.

Ege bölgesi için seçilmiş olan temsili iller Python kodu içerisinde koşullandırıldığında Şekil 36' da görülmekte olan veri setine ulaşılmıştır. Twitter sosyal ağı üzerinden Python programlama dili ile kazınan Ege Bölgesi veri seti toplam 3.663 tweetten oluşmakta olup bu tweetlerin yüzde 95,6'sı İzmir il konumundan üretilmiştir.

Şekil 37. Veri Seti Toplamı(bolge veri seti)

	Frekans
Ege Bölgesi	3.663
Güneydoğu Anadolu Bölgesi	358
Kazınan toplam <u>tweet sayısı</u>	3.721

Kaynak: Yazar tarafından derlenmiştir.

Tez için Python kodlama dili ile Twitter sosyal ağı üzerinden problem tanımında koşulları ifade edilmiş olan veri seti, veri kazıma yöntemi ile tweetler kazınarak oluşturulmuştur. Tez içerisinde uygulama kullanılmak üzere toplanan verinin dağılımı Şekil 37’ de gösterilmektedir. Veri seti toplam 3.721 tweetten oluşmaktadır. 11 Mart 2020 ile 25 Haziran 2021 tarihleri arasında Covid 19 konulu Twitter içeriğinin Ege Bölgesi ve Güneydoğu Anadolu Bölgesinde belirlenen temsili illerden içerik üretimi 3.663 tweet ile Ege Bölgesi ve 358 tweet ile Güneydoğu Anadolu Bölgesi olarak dağılım göstermiştir.

3.2.2. Veri Ön İşleme

Bu bölümde bir önceki bölümde dağılımı hakkında bilgi verilmiş olan bölge veri setine uygulanmış olan veri ön işleme adımlarına ve Doğal Dil İşleme hakkında kısa bilgiye yer verilmiştir.

Doğal Dil İşleme, yapay zekânın alt dalı olan makine öğrenmesi alanının bir dalı olarak tanımlanmaktadır (Ballı, 2021). Alanda İngilizce Natural Language Processing (NLP) olarak ifade edilmektedir. Bilişim alanında iki farklı dil söz konusu olmakta olup bunlardan ilki makine dili diğeri ise doğal diller olarak ifade edilmektedir. İnsanların konuştuğu diller doğal diller olarak tanımlanmakta olup bu dillerin makineler tarafından işlenmesi ise doğal dil işleme olarak tanımlanmaktadır. Doğal dil işleme için metin ve ses olmak üzere iki ana çalışma unsuru bulunduğundan

söz edilir. Metin madenciliği çalışma alanına girmekte olan yazılı metinlerden anlamlı korelasyon, örüntü ve trendlerin keşfedilmesi sürecinin daha kolay olduğu ifade edilirken ses için metin madenciliğinde ise öncelikli işlemin sesi metne dönüştürme olduğu kaynaklarda ifade edilmektedir (Opuz, 2019).

Doğal Dil İşleme, bir metin ya da konuşmadan anlam çıkarmayı amaçlamakla birlikte kuralları çözülen bir dilin yapısının tekrar üretilmesini hedeflemektedir. Doğal işleme süreci, bir takım ön işleme süreci adımlarını içermektedir. Tokenization, işleme alınan metnin cümlelere, kelimelere ayrılma işlemini ifade etmektedir. Stop Words: gereksiz sık kullanılan kelimeler olarak ifade edilmekte olan ve her dil için değişkenlik gösteren kelimeleri ifade etmektedir. Türkçe için bazı Stop Words'ler "hangi, acaba, ve, böylece, madem, elbette, kadar, ise, henüz, hem". Kelimeleri ile örneklendirebilir. Stemming: Kelimelerin köklerine yönelik yapılan bir ön işlem uygulamasıdır. Aynı kökten türeyen kelimelerin farklı algılanmaması için stemming uygulaması gerçekleştirilmektedir (BASKIN, 2021) (Kılınç, 2018). Yapılan kaynak taramasının ardından bölümün devamında yukarı bahsi geçmekte veri ön işleme adımlarının çalışma kapsamında nasıl uygulandığına dair bilgilere yer verilmiştir.

Tez kapsamında daha önceki bölümlerde sebeplerine ayrıntılı olarak yer verilmekte olan 11 Mart 2020 ile 25 Haziran 2021 tarihleri arasında İzmir, Aydın, Muğla, Diyarbakır, Şanlıurfa ve Batman konumu ile üretilmiş Covid 19 konulu Twitter içerikleri veri seti olarak belirlenmiştir. Uygulamada bu illerden Aydın, Muğla ve İzmir konumlu tweetler Ege Bölgesini, Batman, Diyarbakır ve Şanlıurfa konumlu tweetler Güneydoğu Anadolu Bölgesini temsilen veri setinde kullanılmaktadır.

Veri setini oluşturan tweet içerikleri Veri Seti ve Özellikleri adlı başlık altında ayrıntılarına yer verilen yükleme ve kodlama işlemlerinin ardından MySQL veri tabanında saklanmış ardından veri tabanından veri ön işleme sürecine alınmak üzere Microsoft Excel dosyası olarak veri tabanından dışarıya alınmıştır.

Veri seti içeriği oluşturan tüm iller için Twitter içeriğini ifade eden tweet_content sütunu ile birlikte bolge_id ve tweet_id olmak üzere üç sütundan oluşan yeni bir Excel tablosu tasarlanmıştır. Sütunlardan bolge_id için Ege Bölgesi "0" rakamı, Güneydoğu Anadolu Bölgesi "1" rakamı ile ifade edilmiştir. Aydın, İzmir ve Muğla illeri "0" rakamı ile ifade edilirken, Batman, Şanlıurfa ve Diyarbakır illeri "1" rakamı ile ifade edilerek tweetlerin bölgeleri oluşturulmuştur. Sütunlardan tweet_id

ise otomatik artan numaralar (1-...3721) ile her tweet için ayırt edici nitelikte kullanılmaktadır. Uygulama kapsamında veri ön işleme süreci tweet_content sütunu için yürütülmüştür.

Çeşitli adımlardan oluşan veri ön işleme sürecinin yürütülmesi program dışı ve program içi olarak iki adım ile gerçekleştirilmiştir. Daha açık bir şekilde ifade edilir ise küfürlü ve argo içerikli tweetlerin düzenlenmesi, ilişkisiz tweetlerin veri setinden çıkarılması gibi anlamsal veri ön işleme süreci program kullanılmadan kullanıcının her tweeti göz ile süzmesi ve el ile düzenlemesi yoluyla gerçekleştirilmiştir. Tweetlerin tamamının küçük harfe dönüştürülmesi, Türkçe karakterin İngilizce alfabeyle uyarlanması gibi devam eden adımların her kelimenin kökünün bulunması ile sonuçlandırılmasından oluşan veri ön işleme süreci ise kodlama yolu ile gerçekleştirilmiştir.

Veri Ön işleme süreci ekran görüntüleri ve bilgilere ait ayrıntılara yazının devamında yer verilmiştir.

Şekil 38. Veri Ön İşleme Süreci Ham Veri

tweet_content
"kim var!" diye seslenilince, sağına ve soluna bakılmadan, fert fert "ben varım!" cevabını verici, her ferdi "benim olmadığım yerde kimse yoktur!" duygusuna sahip bir mill
Ev sahibim aradı geçen. Ben de bekliyorum "oğlum şu sıkıntı geçene kadar kirayı dert etme" diyecek diye.. Ay sonu kirayı yatacağın diti dedi. Bu arada kıymetli ev sahibin
Ülkemizin son hali 😊 ulaştırma bakanı kovulunca biz👉#ulastirmabakanı @GizemKaracaGifs #CumhurbaşkanıErdoğan #COVID19 https://t.co/OVgKCSBrc8
Beynimiz, üretme kapasitesine sahip olduğu hayalin ötesindeki sınırlarını henüz ortaya çıkaramadığımız bir gizem. Bu gizem çözüldüğünde, bana kalırsa dünyadaki en roman
Fransa'daki tüm kiliselerde, #COVID19 ile mücadelede halka moral vermek için 1 dakika boyunca çan çalındı. Bizde ise Camilerden dua edilince kuduran bir kitle var! Yazıklar
Siz Evde Kalın, Biz Buradayız! #COVID19 ile mücadele kapsamında devlet tedbirlerine riayet ederek, siz vatandaşlarımız evlerinizde kalın ihtiyaç durumunda; 17 ilçemizde eli
Siz Evde Kalın, Biz Buradayız! #COVID19 ile mücadele kapsamında devlet tedbirlerine riayet ederek, siz vatandaşlarımız evlerinizde kalın ihtiyaç durumunda; 17 ilçemizde eli
Cömert Sigorta için değerlisiniz! COVID-19 🚩 salgınını yavaşlatma sorumluluğumuz çerçevesinde Acenteliğimizi 21 Mart 2020 tarihinden itibaren geçici olarak kapatma kar
Dünyaca Ünlü Virolog Veteriner Hekim Ilaria CAPUA, Türk Basınına önemli açıklamalarda bulundu. @sağlikbakanligi @TCTarim @ntv @veterinerhekim #Covid2019TR #VETER
VETERİNER HEKİM VİROLOG PROF.DR. AYKUT ÖZDARENDELİ VE EKİBİ, COVID-19 AŞI GELİŞTİRME ÇALIŞMALARINDA İKİNCİ BÜYÜK AŞAMAYI BAŞARIYLA KAYDEDEREK ME
Almanya'da COVID-19 sürecini başarıyla koordine eden Robert Koch Enstitüsü Başkanı Lothar WİELER, bir Veteriner Hekim.#EvdeKal #veterinerdegilveterinerhekim #Sağlığı
COVID-19 kontrolünü başarıyla sağlayan Çin Hastalık Kontrol ve Önleme Merkezi Direktörü Dr. George GAO bir Veteriner Hekim.#EvdeKal #veterinerdegilveterinerhekim #S
#Repost dr.fericedavadzade with @get_repost · · · COVID-19 –Enfasyon PandemisiBen enfeksiyon uzmanı değilim, çocuk doktoruyum. Her doktor gibi doğru kayna
2 saat oldu otoban giselerindeyim ates kontrolu yapılıyor muhtemelen bi 2 saat daha beklicem sokağa çıkma yasağı kesinlikle şart... #COVID19
Durum halen çok ciddi. Tedbiri elden bırakmayalım.#COVID19
Canım Aydınım. 20 derece daha sıcak ol ve bitirelim şu COVID-19 virüsünü 😊 https://t.co/e0bfrwRep1
Dün sevgili amcaoğluma, bir konu ile ilgili Allahalaaa dedim. O da bana "Allah Allah da, kulları kul değil" dedi.Tırnak içine aldığım kadar var değil mi?#pazartesi #COVID19 #
19 O zaman geceye bir şarkı bırakıyorum. https://t.co/5GI68d8Ap7 #pazartesi #COVID19

Kaynak: Yazar tarafından derlenmiştir.

Şekil 38'de veri tabanı üzerinden Microsoft Excel dosyasına aktarılmış olan uygulama veri setinin ham hali görülmektedir. Veri ön işleme sürecine her tweet için küfür ve argo içerikli kelimelerin varlığının göz ile süzülmesi ile başlamıştır. "Layn,

Eee, sfsfssf, ...git “ gibi çalışmada devamına yer verilmeyecek olan argo açılımlı ya da anlamsız kelimeler tweetlerin içeriğinden el ile silinmiştir. “@...”içerikli retweetler, kullanıcı isimlerinin açık olarak verildiği tweetler tırnak içerisinde verilmiş olan ifadeden el ile arındırılmıştır. “https:/...” ifade tweetlerden çıkarılmıştır.,”#...” tweetin içeriği ile doğrudan bağlantılı olmayan başlık etiketleri(hashtag) içerikten el ile çıkarılmıştır. Veri kazımaya Covid 19 anahtar kelime ile yakalanmış ancak doğrudan konu ile bağlantısı olmayan satış ilanları, atama gibi farklı konuları içeren tweetler doğrudan veri setinden el ile çıkarılmıştır. Tweetler emoji içeriklerinden el ile arındırılmıştır. Bire bir tekrar eden tweetlerin tekrarları el ile veri setinden çıkarılmıştır. Yapılan işlemlerin ardından veri ön işleme sürecinin program dışı adımı tamamlanmıştır. İfade edilen sebepler nedeni ile toplam 50 tweet bolge veri setinden dışlanmıştır. Kodlama ile veri ön işleme sürecine 3671 tweet oluşan veri seti alınmıştır.

Şekil 39. Veri Ön İşleme Süreci Program Dışı Ön İleme Tamamlandı

A	B	C	D
tweet_content	bolge_id	tweet_id	
"kim var!" diye seslenince, sağına ve soluna bakmadan, fert fert "ben varım!" cevabını verici, her ferdi "benim olmadığım yerde kimse yoktur!" duygusuna sahip bir millet! Elhamdulillah!	0	1	
Ev sahibim aradı geçen. Ben de bekliyorum "oğlum şu sıkıntı geçene kadar kirayı dert etme" diyecek diye... Ay sonu kirayı yatacağın dımı dedi. Bu arada kıymetli ev sahibimin kirada 17 tane evi var Ülkemizin son hali ulaştırma bakanı kovulunca biz	0	2	
Beynimiz, üretme kapasitesine sahip olduğu hayalini ötesindeki sınırlarını henüz ortaya çıkarmadığımız bir gizem. Bu gizem çözüldüğünde, bana kalırsa dünyadaki en romantik hikâyeye ortaya çıkmış olacak. Fransa'daki tüm kiliselerde, #COVID19 ile mücadelede halka moral vermek için 1 dakika boyunca çan çaldı.	0	3	
Bizde ise Camilerden dua edilince kuduran bir kitle var! Yazıklar olsun!!! Siz Evde Kalın, Biz Buradayız!	0	4	
COVID19 ile mücadele kapsamında devlet tedbirlerine riayet ederek, siz vatandaşlarımız evlerinizde kalın ihtiyaç durumunda; 17 ilçemizde eliniz, gözünüz, kulağınız olacağız. BİZ BURADAYIZ Comert Sigorta için değerlisiniz!	0	5	
COVID-19 salgınına yavaşlatma sorumluluğumuz çerçevesinde Acentelğimizi 21 Mart 2020 tarihinden itibaren geçici olarak kapatma kararını aldık. Tekrar açılacağımız duyurusunu... Dünyaca Ünlü Virolog Veteriner Hekim Ilarita CAPUA, Türk Basınına önemli açıklamalarda bulundu. VETERİNER HEKİM VİROLOG PROF.DR. AYKUT ÖZDARENDELİ VE EKİBİ, COVID-19 AŞI GELİŞTİRME ÇALIŞMALARINDA İKİNCİ BÜYÜK AŞAMAYI BAŞARIYLA KAYDEDEREK MESLEK CAMIAMIZI 1. GURURLANDIRMIŞ, İNSANLIĞA YENİ BİR UMUT VERMİŞTİR. KENDİSİNİ VE EKİBİNİ YÜREKTEN KUTLUYORUZ... 2. Almanya'da COVID-19 sürecini başarıyla koordine eden Robert Koch Enstitüsü Başkanı Lothar WİELER, bir Veteriner Hekim. 3. COVID-19 kontrolünü başarıyla sağlayan Çin Hastalık Kontrol ve Önleme Merkezi Direktörü Dr. George GAO bir Veteriner Hekim. 4. 2 saat oldu otoban giselerindeyim ates kontrolü yapıyor muhtemelen bi 2 saat daha beklicem sokağa çıkma yasağı kesinlikle şart... #COVID19 5. Durum halen çok ciddi. Tedbiri elden bırakmavalım.	0	6	
	0	7	
	0	8	
	0	9	
	0	10	
	0	11	
	0	12	
	0	13	
	0	14	
	0	15	

Kaynak: Yazar tarafından derlenmiştir.

Bölge isimli veri seti anlamsal veri ön işleme sürecine alınmıştır. Ön işleme süreci sonrasında 50 tweet dışlanarak toplam 3.671 tweetten oluşan yeni bir seti oluşturulmuştur. Göz ve El ile gözetimli olarak gerçekleştirilen ön işleme süreci sonrasında oluşan bolge veri setinin yeni haline Şekil 39'da yer verilmiştir.

Veri ön işleme sürecinin ikinci adımının program ile gerçekleştirilmesi planlanmıştır. Öncülikle veri ön işleme sürecinin Microsofta Excel programı üzerinden

formüller ve programın kendine has bazı özelliklerinin kullanılması yolu ile gerçekleştirilmesi hedeflenmiştir.

Metni hecelere bölme işlemi Excel programı veri sekmesi, metni sütunlara dönüştür özelliği kullanılarak gerçekleştirilmiştir.

Şekil 39'da Excel ile metni sütunlara dönüştürme işlemi gösterilmektedir. Daha sonrasında hecelerine ayrılan metnin jeton olarak tanımlanan bu küçük parçaları Excel "KÜÇÜKHARF(değer)" formülü kullanılarak küçük harfe dönüştürülmüştür.

Şekil 40. Veri Ön İşleme Süreci- Excel Deneniyor

	A	B	C	D	E	F	G	H	I	J
1	tweet_content	tweet_cot	tweet_cot	tweet_cot	tweet_cot	tweet_cot	tweet_cot	tweet_cot	tweet_cot	tweet_cot
2	Biz gece gündüz demeden	biz gece	gündüz	demeden	covid-19	vakası	ile	savaşırker	bazılar	
3	Karantınadaki 15. Güne giri	karantina	15. Güne	girmiş	bulunmak	#14GünKu	da	uyguladığı	göre	
4	Covid-19 ile mücadelede ö	covid-19	ile	mücadele	önemli	gunlerder	geçiyoruz	vaka	sayısının	ARTM
5	#EvdKale #Covid_19'dan	#EvdKale	#Covid_15'dan	önce	ve	sonra		https://t.co/KiG		
6	#EvdKaleEvdKale Bu #Covi	#EvdKale	Bu	#Covid_15	bittiğinde	daha	bizim	#piskoloji	olacak	heme
7	#StayAtHome #türkiye	#Co	#stayata	#türkiye	#Covid_19					
8	Merhamet, zulmün merhem	merhame	zulmün	merhemi	olamaz	!	#huzursuz	#Covid_15	#HayatEve	https:
9	Hocam 10 dakika dşan	oka	hocam	10	dakika	dşan	gkarsak	karantina	bozulur	mu??
10	Yönetim kurulu toplantı	yonetim	kurulu	toplantı	#Covid19	kapsamını	alınan	tedbirler	gündemi	ile
11	Hane Atatürk, Çin'den gele	hane	Atatürk,	Çin'den	gelen	Covid-19	testi	paralarını	vermişti,	bunu
12	Covid-19 doğanın (ekolojik	covid-19	doğanın	(ekolojik	dengenin	bozulması	sonucu)	kendi	seleksiyo	mu
13	#EvdKale #CoronavirusPan	#EvdKale	#Coronav	Uyalım	bu	tabloya	!!	#Covid_19		
14	Senin virüsle dalga geçme	senin	virüsle	dalga	geçmeyi	maharet	sanman	nedenyle	her	gün
15	Koronavirüs riskine karşı	koronavir	riskine	karşın	14	kural	!			
16	Batman sağlıkplamı yanınd	batman	sağlıkpları	yanında						
17	#EvdKale Evinde uzaktaysa	#EvdKale	Evinde	uzaktaysa	#oteldeka	#cumarte	#Coronavi	#koronatu	#Covid_15	#EvdHay
18	#EvdKale Evinde Uzaktays	#EvdKale	Evinde	Uzaktaysa	#oteldeka	#koronatu	#otel	#hotel	#Coronavi	#kandil
19	İnsanı miraca götüren yol	insanı	miraca	götüren	yol	barştan,	sevğiden	ve	kardeşlikt	geçer.
20	Ruhumuzunda dezenfekte	ruhumuzu	dezenfekte	ihitiyacı	var..					
21	@ziyasetci bakınm da bi	@ziyasetc	bakınm	da	bizi	görece	mi			
22	Bizden kaynaklı olmayan	bi	bizden	kaynaklı	olmayan	bir	durum	dar	kaynaklı	mağdur
23	Mağdur edilen binlerce	ins	mağdur	edilen	binlerce	insan	var	evine	bu	edimemi
24	Birlikte her zaman daha	gü	birlikte	her	zaman	daha	güçlüyüz.			
25	Her eğitmen kadar bizim	d	Her	eğitmen	kadar	bizim	de	hakkımız	var..!	

Kaynak: Yazar tarafından derlenmiştir.

Şekil 42. Veri Ön İşleme Süreci- Excel Deneniyor

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi	tweet_coi
2	biz	gece	gunduz	demedem	covid-19	vakası	ile	savasirker	bazıları	bize	sıdöeti	on	goruyorsa	bunun	pesinin	birakilmag	gerekıyor	net	.	#mugeani	#korot
3	karantina	15.	gune	girmis	bulunmak	#14gunkui	da	uyguladig	gore	sayisinin	artmamas	2	yolu	var							
4	covid-19	ile	mucadele	onemli	gunlerde	geciyoruz	vaka	sayisinin	artmamas	2	yolu	var									
5	#evdekal	#covid_19	'dan	once	ve	sonra	☺	https://t.t													
6	#evdekal	bu	#covid_19	bittigind	de	daha	bizim	#piskoloji	nolacak	hemen	#covid19	bitti	demeyin								
7	#stayaata	#turkiye	#covid_19																		
8	merhame	zulmun	merhemi	olamaz	!	#huzursuz	#covid_19	#hayateve	https://t.t												
9	hocam	10	dakika	disari	cikarsak	karantina	bozulur	mu??													
10	yonetim	kurulu	toplantim	#covid19	kapsaminin	alinan	tedbirler	gundemi	ile	on-line	olarak	gercekle	https://t.t								
11	hane	ataturk,	cin'den	gelen	covid-19	testi	paralarını	vermist,	bunu	ataturk'e	baglayan	amik	kym,	lan	biz	parasını	vererek	almisiz.	serefsizle	neden	yanlis
12	covid-19	dogarin	(ekolojik	dengenin	bozulmasi	sonucu)	kendi	seleksiyor	mu	yoksa	kapitalizm	bir	oyunu	mu.	bence	uzerine	dusunulm	gerek	bir	konu.	zira
13	#evdekal	#coronavi	uyalim	bu	tabloya	!!	#covid_19														
14	senin	virusle	dalga	gecmeyi	maharet	sanman	nedeniyle	her	gun	bir	onceki	gun	vefat	eden	sayisinin	iki	kati	vefat	olacak.		
15	koronavir	risikine	karsin	14	kural	!															
16	batman	saglikliklar	yaninda																		
17	#evdekal	evinde	uzaktaysa	#oteldeka	#cumarter	#coronavi	#coronavi	#covid_19	#evdehay	#evdekal	#670vaka	#vaka690	#worlddd	#wirbleib	#taylortol	#عبدالله	الرحمن	https://t.t			
18	#evdekal	evinden	uzaktaysa	#oteldeka	#koronatu	#otel	#hotel	#Coronavi	#kandil	#mirac	#mirackar	#evdehay	#evdekal	#realkona	@drfahre!	@tckultur	@batman	@turob	@sahinhu	@cemkini	@tcha
19	insani	miraca	goturen	yol	baristan,	sevgidem	ve	kardeslikt	gecer.	birlik	ve	beraberlikt	icinde	mirac	kandiliniz	mubarek	olsun...	rabbim	tuttugunu	tutacagini	orucu
20	ruhumuz	dezenfekt	ihitiyaci	var..																	
21	@ziyaset	bakanim	da	bizi	gorecek	mi	☺														
22	bizden	kaynakli	olmayan	bir	durumdar	kaynakli	magdur	edilmemi	haksizlik	ozellikle	ulke	bu	durumday								
23	magdur	edilen	binlerce	insan	var	evine	bu	surede	bakamaya	binlerce	insandan	bahsedily	lutfem	gerek	magduriyy	giderilisin.					
24	birlikte	her	zaman	daha	guclu	uz.															
25	her	egitmen	kadar	bizim	de	hakkimiz	var..!														

Kaynak: Yazar tarafından derlenmiştir.

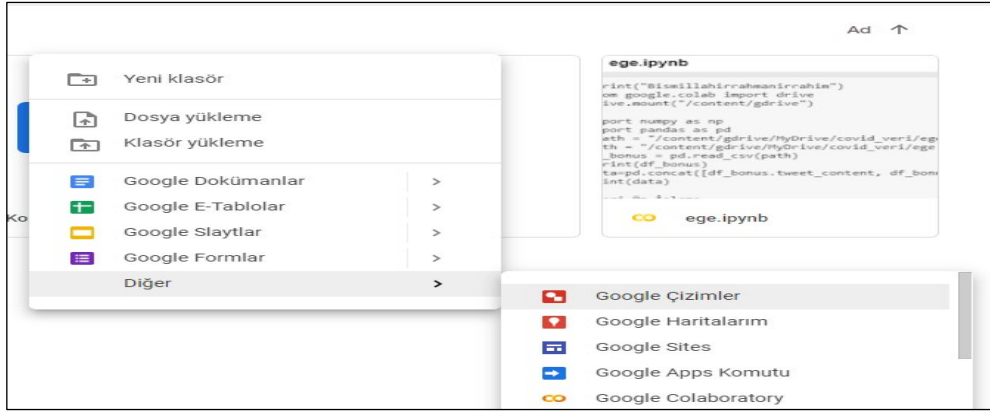
Şekil 42’de Türkçe karakterin İngilizce karşılıkları ile değiştirilmesine yer verilmektedir.

Microsoft Excel programı ile Doğal Dil İşleme için gerekli veri ön işleme adımlarının bir kısmını gerçekleştirmişti. Ancak kelime köklerinin bulunması aşamasında yetersiz kalınmıştır. Ardından bolge isimli veri setinin Doğal Dil İşleme için veri ön işleme sürecine Google Colab platformu üzerinden Python programlama dili ile devam edilmiştir (Şekil 43).

Google Colab, Colaboratory ismiyle bilinmekte olup tarayıcımızda Python kodu yazmamıza ve çalıştırmamıza imkan vermektedir. Hiç yapılandırma gerektirmemesi, GPU'lara ücretsiz erişim imkanı, kolay paylaşım imkanı gibi fırsatları sunmakta olan bir bulut bilişim hizmetidir (Welcome To Colaboratory - Google Research, 2017).

Google Elektronik Posta üyelik adresinden yeni bir klasör açılmalı ve sağ tık seçeneklerden Google Colaboratory seçeneğini seçilerek Colab Python sayfası açılmaktadır (Şekil 44).

Şekil 43. Veri Ön İşleme Google Colab Açma



Kaynak: Yazar tarafından derlenmiştir.

Python kodlama sayfasının açılmasının ardından Google Drive ile bağlantı kurulmaktadır.

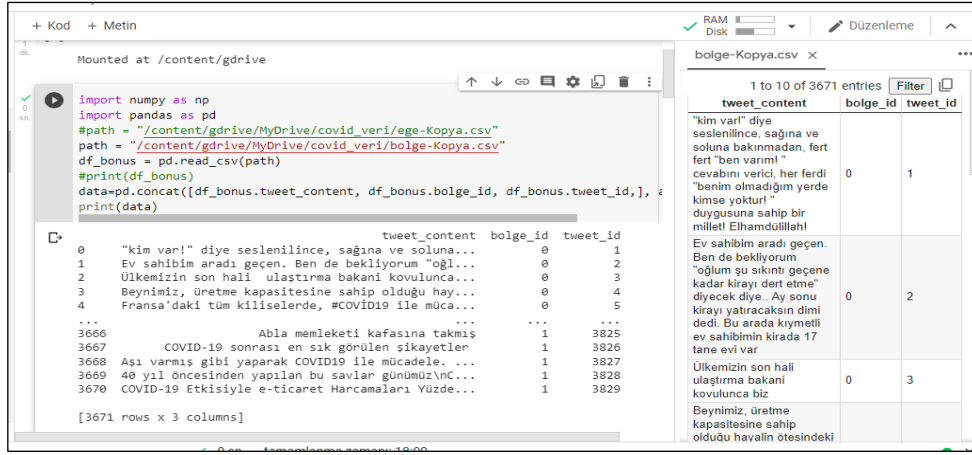
Şekil 44. Veri Ön İşleme Google Colab ve Google Drive Bağlantı



Kaynak: Yazar tarafından derlenmiştir.

Şekil 44'te Google Colab ve Google Drive bağlantı kodu görülmektedir. Kod, oynat butonundan çalıştırıldıktan sonra Şekil 20' de görülen "Mounted at/content/gdrive" çıktısı alındığın bağlantının başarı ile gerçekleştirildiği ifade edilmektedir.

Şekil 45. Veri Ön İşleme Google Colab CSV Bağlantısı



```
import numpy as np
import pandas as pd
#path = "/content/gdrive/MyDrive/covid_veri/ege-Kopya.csv"
path = "/content/gdrive/MyDrive/covid_veri/bolge-Kopya.csv"
df_bonus = pd.read_csv(path)
#print(df_bonus)
#pd.concat([df_bonus.tweet_content, df_bonus.bolge_id, df_bonus.tweet_id], axis=1)
data=pd.concat([df_bonus.tweet_content, df_bonus.bolge_id, df_bonus.tweet_id], axis=1)
print(data)
```

	tweet_content	bolge_id	tweet_id
0	"kim var!" diye seslenilince, sağına ve soluna...	0	1
1	Ev sahibim aradı geçen. Ben de bekliyorum "oğl...	0	2
2	Ülkemizin son halî ulaştırma bakanı kovulunca...	0	3
3	Beynimiz, üretme kapasitesine sahip olduğu hay...	0	4
4	Fransa'daki tüm kiliselerde, #COVID19 ile müca...	0	5
...
3666	Abla memleketi kafasına takmış	1	3825
3667	COVID-19 sonrası en sık görülen şikayetler	1	3826
3668	Ası varmış gibi yaparak COVID19 ile mücadele...	1	3827
3669	48 yıl öncesinden yapılan bu savlar günümüz\nc...	1	3828
3670	COVID-19 Etkisiyle e-ticaret Harcamaları Yüzd...	1	3829

Kaynak: Yazar tarafından derlenmiştir.

Şekil 45’de bolge veri setinin Colab’ta görüntülenmesi gösterilmektedir. Veri madenciliği için gerekli kütüphaneler Numpy ve Pandas yüklenmiştir. Path değişkeni içerisinde veri setinin bulunduğu csv dosyasına ait yol atılmıştır. Pandas kütüphanesi csv okuma kodu ile veri seti df_bonus değişkeni içerisinde yüklenmiştir. Concat komutu ile veri setinin sadece istenilen sütunları veri ön işleme süreci için data değişkeni içerisinde tanımlanmıştır. Ancak kodun yazılmasının ardından csv dosyasından diğer sütunların çıkarılması nedeniyle data değişkenine gerek kalmamış olup df_bonus değişkeni ile gerekli işlemler gerçekleştirilebilir.

Tüm veri seti için metnin devamında for döngüsü için veri ön işleme döngüsüne yer verilecek olup öncelikli olarak adımlara değinilecektir. Uygulama kapsamında veri ön işleme sürecine sadece tweet_content sütunu alınacaktır. Bu sebeple tweet_content sütunu metin değişkeni içerisinde tanımlanmıştır.

Şekil 46’da veri setinin ilk satırı üzerinde lower komutu çalışmakta olup tamamının küçük harf olarak döndüğü görülmektedir.

İkinci adımda ise dilimizi İngilizce dilinden ayıran “ı,ö,ü,ç,ş,ğ” harfler İngilizce dilindeki karşılıkları ile değiştirilmektedir.

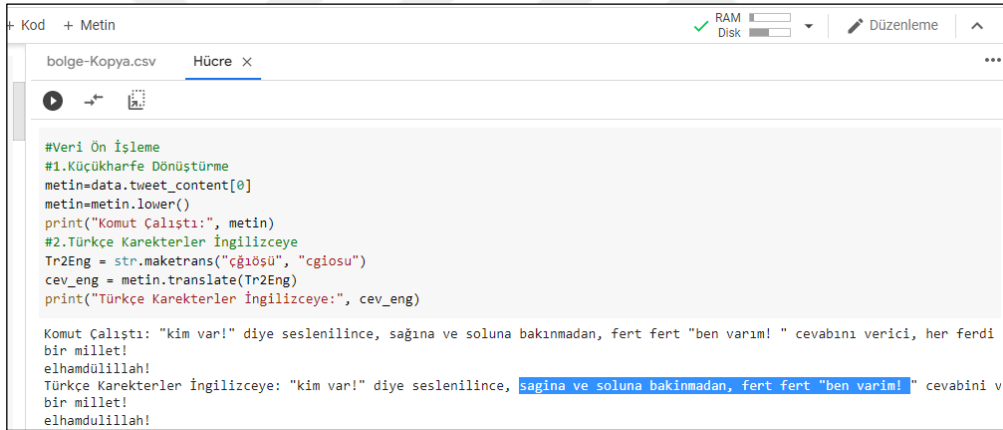
Şekil 46. Veri Ön İşleme Google Colab- Veriler Küçük Harfe Dönüştürülüyor



```
+ Kod + Metin
bolge-Kopya.csv Hücre x
#Veri Ön İşleme
#1.Küçükharfe Dönüştürme
metin=data.tweet_content[0]
metin=metin.lower()
print("Komut Çalıştı:", metin)
ye seslenilince, sağına ve soluna bakınmadan, fert fert "ben varım! " cevabını verici, her ferdi "benim olmadığım yerde kimse yok
```

Kaynak: Yazar tarafından derlenmiştir.

Şekil 47. Veri Ön İşleme Google Colab- Türkçe Karakterleri Değiştiriyoruz



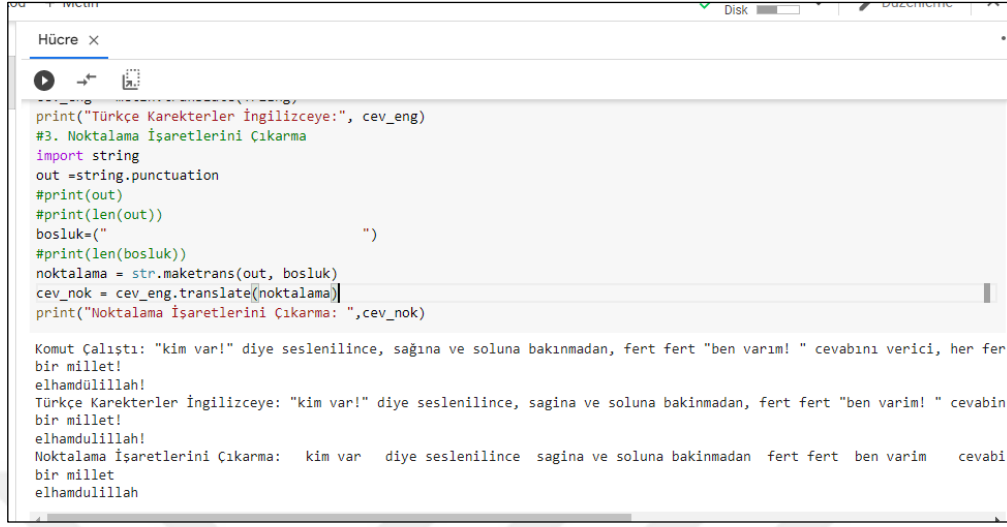
```
+ Kod + Metin
bolge-Kopya.csv Hücre x
#Veri Ön İşleme
#1.Küçükharfe Dönüştürme
metin=data.tweet_content[0]
metin=metin.lower()
print("Komut Çalıştı:", metin)
#2.Türkçe Karakterler İngilizceye
Tr2Eng = str.maketrans("çğıöüş", "cgiosu")
cev_eng = metin.translate(Tr2Eng)
print("Türkçe Karakterler İngilizceye:", cev_eng)
Komut Çalıştı: "kim var!" diye seslenilince, sağına ve soluna bakınmadan, fert fert "ben varım! " cevabını verici, her ferdi bir millet! elhamdülillah!
Türkçe Karakterler İngilizceye: "kim var!" diye seslenilince, sağına ve soluna bakınmadan, fert fert "ben varım! " cevabını v bir millet! elhamdülillah!
```

Kaynak: Yazar tarafından derlenmiştir.

Tr2Eng değişkeni içerisine string kütüphanesi özelliği tanımlanmakta olup sırasıyla ç,ğ,ı,ö,ü,ş harflerinin c,g,i,o,u,s harfleri ile değiştirilmesine dair komut tanımlanmıştır. Metin değişkeninde Tr2Eng kuralına göre değişikliği gidilmiş ve metinde yapılan değişiklik ile birlikte metin cev-eng değişkeni içerisine atılmıştır. Şekil 47’de kodun ekran çıktısında Türkçe karakterlerin tanımlanan şekilde değiştirildiği görülmektedir.

Veri ön işlemenin bir sonraki adımında ise metin içerisindeki noktalama işaretleri kaldırılmıştır.

Şekil 48. Veri Ön İşleme Google Colab- Noktalama İşaretlerini Çıkartıyoruz



```
Hücre x
print("Türkçe Karakterler İngilizceye:", cev_eng)
#3. Noktalama İşaretlerini Çıkarma
import string
out =string.punctuation
#print(out)
#print(len(out))
bosluk=" ")
#print(len(bosluk))
noktalama = str.maketrans(out, bosluk)
cev_nok = cev_eng.translate(noktalama)
print("Noktalama İşaretlerini Çıkarma: ",cev_nok)

Komut Çalıştı: "kim var!" diye seslenilince, sağına ve soluna bakınmadan, fert fert "ben varım! " cevabını verici, her fer
bir millet!
elhamdülillah!
Türkçe Karakterler İngilizceye: "kim var!" diye seslenilince, sağına ve soluna bakınmadan, fert fert "ben varım! " cevabın
bir millet!
elhamdülillah!
Noktalama İşaretlerini Çıkarma: kim var diye seslenilince sağına ve soluna bakınmadan fert fert ben varım cevabi
bir millet
elhamdülillah
```

Kaynak: Yazar tarafından derlenmiştir.

Şekil 48’de görüldüğü üzere metin içerisindeki noktalama işaretlerinin çıkarılması amacı ile öncelikle bir önceki işlemde de kullanılmış olan string kütüphanesi yüklenmiştir. Out isimli değişkene noktalama işaretleri tanımlanmış olup 32 adet noktalama işareti bulunmaktadır. Ardından boşluk değişkeni içerisinde Out değişkeni ile değiştirilmek üzere toplam 32 adet boşluk karakteri tanımlanmıştır. Noktalama isimli değişken içerisinde noktalama işaretleri boşluk işaretleri ile değiştirilmiş ve noktalama işaretlerinden arındırılmış yeni metin cev_nok değişkeni içerisine atanmıştır.

Şekil 48’de kodun çalışması sonrasında döndürülen metnin noktalama işaretlerinden arınmış olduğu görülmektedir. Bir sonraki aşamada en fazla 140 karakterden oluşan metin kelimelerine ayrılmıştır.

Şekil 49’da görüldüğü üzere hâlihazırda bir takım veri ön işleme adımların geçmiş bir dizi olan cev_nok değişkeni içeri split komutu kullanılarak boşluk karakterlerinden hecelerine ayrılarak kel_parca isimli yeni bir dizi değişkenine atanmıştır.

Şekil 49. Veri Ön İşleme Google Colab- Hecelere(Tokenlere) Ayırma

```
Hücre x
#print(out)
#print(len(out))
bosluk=" "
#print(len(bosluk))
noktalama = str.maketrans(out, bosluk)
cev_nok = cev_eng.translate(noktalama)
print("Noktalama İşaretlerini Çıkarma: ",cev_nok)
#4.Parçalara-Token Ayırma
kel_parca=cev_nok.split()
print("Kelimelere Ayrılmış:", kel_parca)

Komut Çalıştı: "kim var!" diye seslenilince, sağına ve soluna bakınmadan, fert fert "ben varım! " cevabını verici, her ferdi "b
bir millet!
elhamdulillah!
Türkçe Karakterler İngilizceye: "kim var!" diye seslenilince, sagina ve soluna bakınmadan, fert fert "ben varım! " cevabını ver
bir millet!
elhamdulillah!
Noktalama İşaretlerini Çıkarma: kim var diye seslenilince sagina ve soluna bakınmadan fert fert ben varım cevabını ve
bir millet
elhamdulillah
Kelimelere Ayrılmış: ['kim', 'var', 'diye', 'seslenilince', 'sagina', 've', 'soluna', 'bakınmadan', 'fert', 'fert', 'ben', 'var']
```

Kaynak: Yazar tarafından derlenmiştir.

Diller içerisinde çok sık kullanılan ancak arama motorları tarafından algoritmadan dışlanarak görmezden gelinen kelimeleri etkisiz kelimeler daha çok bilinen İngilizce karşılığı ile Stop Words adı verilmektedir. Arama motorlarının sonuçlarının isabet etmesi olumsuz etkilediğinden Doğal Dil İşlemede bir veri ön işleme adımı olarak çalışılmakta olup her dil için hazırlanmış farklı stop words listelerine kolaylıkla ulaşılabilmektedir. (Türkçe Etkisiz Kelimeler (Stop Words) Listesi 1.1, 2019)

Sık kullanılan kelimelerin metin içerisinden çıkarılması işleminde öncelikli olarak nltk kütüphanesi yüklenmekte ve ardından sık kullanılan kelimeler indirilmektedir. Türkçe için etkisiz kelimeler listesi stop_word_list değişkeni ile kodlamaya dâhil edilmekte ve metin içerisinde bulunan ancak stop_words_list bulunmayan kelimelerden durak_kel isimli değişken içerisinde yeni bir metin tanımlanmaktadır. Şekil 50’de kodun çıktısı gösterilmektedir.

Şekil 50. Veri Ön İşleme Google Colab- Sık Kullanılan Kelimelerin(Stop Words) Çıkarılması

```
Hücre x
print( 'kelimelere Ayrılmış: ', kel_parca)
#5.Gereksiz Kelimeleri(Stop_Words)Çıkarma
import nltk
nltk.download('stopwords')
stop_word_list = nltk.corpus.stopwords.words('turkish')
durak_kel = [kel_parca for kel_parca in kel_parca if kel_parca not in stop_word_list]
print("Anlamsız Kelimeleri(Stop_Words)Çıkarma:", durak_kel)

Komut Çalıştı: "kim var!" diye seslenilince, sağına ve soluna bakınmadan, fert fert "ben varım! " cevabını verici, her fi
bir millet!
elhamdulillah!
Türkçe Karakterler İngilizceye: "kim var!" diye seslenilince, sagina ve soluna bakınmadan, fert fert "ben varım! " cevabı
bir millet!
elhamdulillah!
Noktalama İşaretlerini Çıkarma: kim var diye seslenilince sagina ve soluna bakınmadan fert fert ben varım ceval
bir millet
elhamdulillah
Kelimelere Ayrılmış: ['kim', 'var', 'diye', 'seslenilince', 'sagina', 've', 'soluna', 'bakınmadan', 'fert', 'fert', 'ben
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
Anlamsız Kelimeleri(Stop_Words)Çıkarma: ['var', 'seslenilince', 'sagina', 'soluna', 'bakınmadan', 'fert', 'fert', 'ben',
```

Kaynak: Yazar tarafından derlenmiştir.

Şekil 51. Veri Ön İşleme Google Colab- Sık Kullanılan Kelimelerin Köklerinin Bulunması

```
Hücre x
durak_kel = [kel_parca for kel_parca in kel_parca if kel_parca not in stop_word_list]
print("Anlamsız Kelimeleri(Stop_Words)Çıkarma:", durak_kel)
#6.Kök Bulma
from snowballstemmer import TurkishStemmer
turkStem=TurkishStemmer()
print(turkStem.stemWord(durak_kel[2]))

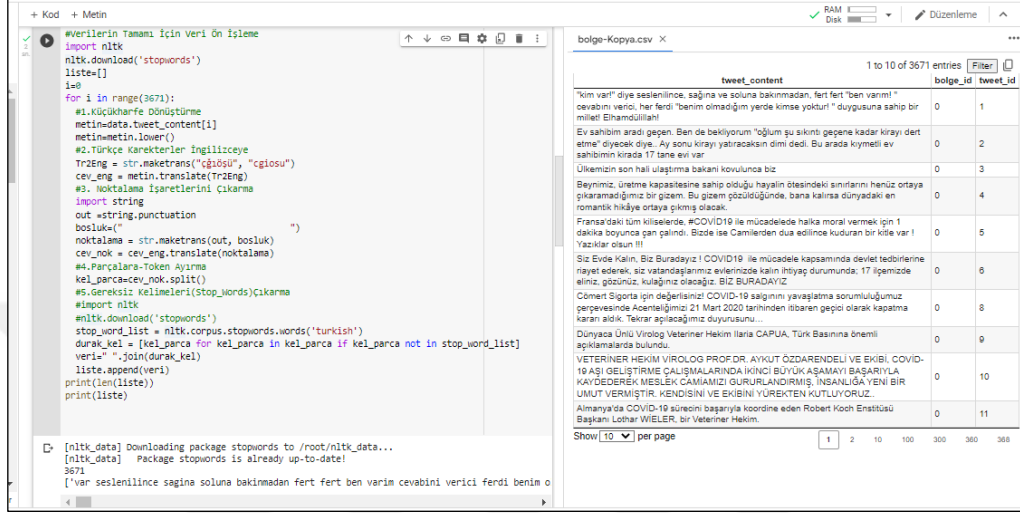
Komut Çalıştı: "kim var!" diye seslenilince, sağına ve soluna bakınmadan, fert fert "ben varım! " cevabını verici, her fi
bir millet!
elhamdulillah!
Türkçe Karakterler İngilizceye: "kim var!" diye seslenilince, sagina ve soluna bakınmadan, fert fert "ben varım! " cevabı
bir millet!
elhamdulillah!
Noktalama İşaretlerini Çıkarma: kim var diye seslenilince sagina ve soluna bakınmadan fert fert ben varım ceval
bir millet
elhamdulillah
Kelimelere Ayrılmış: ['kim', 'var', 'diye', 'seslenilince', 'sagina', 've', 'soluna', 'bakınmadan', 'fert', 'fert', 'ben'
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Anlamsız Kelimeleri(Stop_Words)Çıkarma: ['var', 'seslenilince', 'sagina', 'soluna', 'bakınmadan', 'fert', 'fert', 'ben',
sagi
```

Kaynak: Yazar tarafından derlenmiştir.

Metin Madenciliği için veri işleme sürecinin son adımı olan kök bulma işlemidir. Kaynaklarda dilimiz için en etkili kök bulma kütüphanesinin Zemberek Kütüphanesi olduğu ifade edilmekte olup farklı sebepler nedeni ile bu kütüphane kullanılmamış kaynaklarda bahsi geçen başka bir kütüphane Stemmer uygulamada

kullanılmıştır. Türkçe TurkishStemmer kütüphanesi ile birlikte snowballstemmer algoritması kullanılarak kelime kökleri tanımlanmıştır (Şekil 51).

Şekil 52. Veri Ön İşleme Google Colab-Tamamlandı



```
#Verilerin Tamamı İçin Veri Ön İşleme
import nltk
nltk.download('stopwords')
liste=[]
i=0
for i in range(3671):
    #1.Küçükharfe Dönüştürme
    metin=data.tweet_content[i]
    metin=metin.lower()
    #2.Türkçe Karakterler İngilizceye
    Tr2Eng = str.maketrans("çğıöşü", "cgiosu")
    cev_eng = metin.translate(Tr2Eng)
    #3. Noktalama İşaretlerini Çıkarma
    import string
    out =string.punctuation
    bosluk=" "
    noktalama = str.maketrans(out, bosluk)
    cev_nok = cev_eng.translate(noktalama)
    #4. Parçalara-Token Ayırma
    kel_parca=cev_nok.split()
    #5. Gereksiz Kelimeleri(Stop_words)Çıkarma
    #import nltk
    #nltk.download('stopwords')
    stop_word_list = nltk.corpus.stopwords.words('turkish')
    durak_kel = [kel_parca for kel_parca in kel_parca if kel_parca not in stop_word_list]
    veri=""
    liste.append(veri)
    print(len(liste))
    print(liste)
```

tweet_content	bolge_id	tweet_id
"kim var!" diye seslenilince, sağına ve soluna bakılmadan, fert fert "ben varım!" cevabını verici, her ferdi "benim olmadığım yerde kimse yoktur!" duygusuna sahip bir millet! Eihamedüillahi!	0	1
Ev sahibim aradı geçen. Ben de bekliyordum "oğlum şu sıkıntı gegene kadar kirayı dert etme!" diyecek diye. Ay sonu kirayı yatacağım dımı dedi. Bu arada kıymetli ev sahibimin kirada 17 tane evi var	0	2
Ülkemizin son hali ulaşırma bakani kovulunca biz	0	3
Beynimiz. Üretim kapasitesine sahip olduğu hayalin ötesindeki sınırlarını henüz ortaya çıkarmadığımız bir gözlem. Bu gözlem gözlemlüğünde, bana kalırsa dünyadaki en romantik hediye ortaya gilemiş olacak.	0	4
Fransa'daki tüm kiliselerde, #COVID19 ile mücadelede halka moral vermek için 1 dakika boyunca çan çalındı. Bizde ise Camilerden dua edilince kuduran bir kitle var! Yazıklar olsun!!!	0	5
Siz Evde Kalın. Biz Buradayız! COVID19 ile mücadele kapsamında devlet tedbirlerine riayet ederek, siz vatandaşlarımız evlerimizde kalın hıyayp durumunda, 17 işimizde eliniz, gözünüz, kulağınız olacaktır. BİZ BURADAYIZ.	0	6
Cömert Sigorta için değerlisiniz! COVID-19 salgınına yavaşlatma sorumluluğumuz perçevesinde Acenteliğimizi 21 Mart 2020 tarihinden itibaren geçici olarak kapatma kararını aldık. Tekrar açacağımız duyurumu...	0	8
Dünyaca Ünlü Virolog Veteriner Hekim İllaria CAPUA, Türk Basınına önemli açıklamalarda bulundu.	0	9
VETERİNER HEKİM VİROLOG PROF.DR. AYKUT ÖZDARENDİLİ VE EKİBİ, COVID-19 AŞI GELİŞTİRME ÇALIŞMALARINDA İKİNCİ BÜYÜK AŞAMAYI BAŞARIYLA KAVDİDEREK MESELEİN ÇAMAMIZI GURURLANDIRMIŞ. İNŞAĞIĞA YENİ BİR UMUT VERMİŞTİR. KENDİSİNİ VE EKİBİNİ YÜREKTEN KUTLUYORUZ.	0	10
Almanya'da COVID-19 sürecini başarıyla koordine eden Robert Koch Enstitüsü Başkanı Lothar WIELER, bir Veteriner Hekim.	0	11

Kaynak: Yazar tarafından derlenmiştir.

Şekil 52' de bölge veri seti içerisindeki tüm tweetle için kök bulma bölümü dışındaki veri ön işleme adımları gerçekleştirilmiş, tweetlerin beş adımlı veri ön işleme sonrasındaki yeni değerleri liste isimli değişkene atanmıştır.

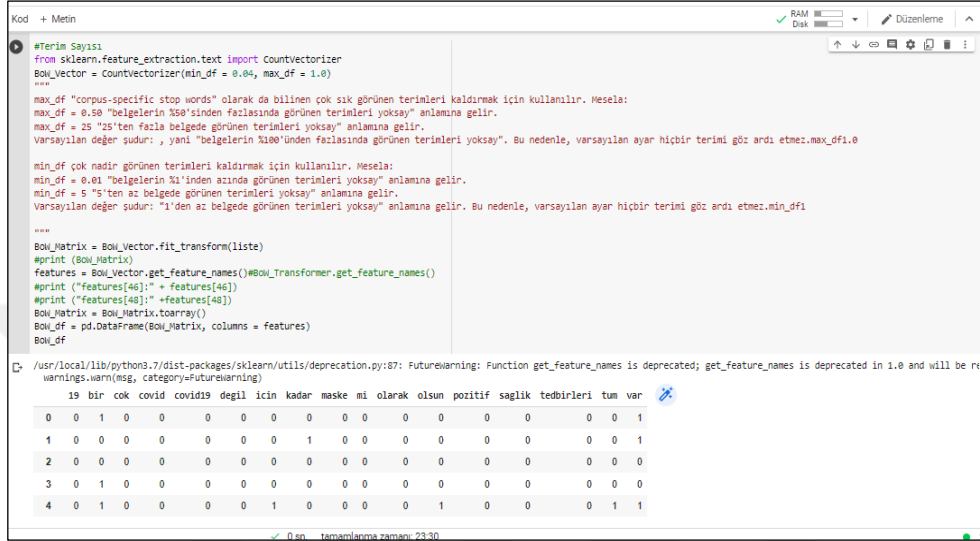
3.2.3. Bulgular

Yapılan uygulama sonrası elde edilen Şekil 53' de kelime çantası adı verilen matrise ulaşılmıştır. Tweetler'in içeriğini oluşturan her bir kelimenin 3.671 tweet içeriğinde var(1) ya da yok(0) şeklinde ifade edilmesinde oluşan matris inceliğinde kelime sayısının çok olması ve kelime köklerinin tam tahmin edilememesi nedeniyle bir karmaşıklığın söz konusu olduğu görülmekte ve matris uygulama hakkında doğru bilgi vermemektedir.

Ege Bölgesi temsili illerinden 3.531 tweet veri setine alınmış olup tweetler toplam 60.999 kelimededen oluşmaktadır. Ege bölge veri setini oluşturan 60.999

kelimeden 3.660 Covid 19 kelimesidir. Üretilen Twitter içeriklerine ait Ege Bölgesi kelime bulutu Şekil 53’de gösterilmektedir.

Şekil 53. Google Colab-Kelime Çantası(Bag of Words)



```
#Terim Sayısı
from sklearn.feature_extraction.text import CountVectorizer
BoW_Vector = CountVectorizer(min_df = 0.04, max_df = 1.0)
"""
max_df "corpus-specific stop words" olarak da bilinen çok sık görünen terimleri kaldırmak için kullanılır. Mesela:
max_df = 0.50 "belgelerin %50'sinden fazlasında görünen terimleri yoksay" anlamına gelir.
max_df = 25 "25'ten fazla belgede görünen terimleri yoksay" anlamına gelir.
Varsayılan değer şudur: , yani "belgelerin %100'ünden fazlasında görünen terimleri yoksay". Bu nedenle, varsayılan ayar hiçbir terimi göz ardı etmez.max_df:0

min_df çok nadir görünen terimleri kaldırmak için kullanılır. Mesela:
min_df = 0.01 "belgelerin %1'inden azında görünen terimleri yoksay" anlamına gelir.
min_df = 5 "5'ten az belgede görünen terimleri yoksay" anlamına gelir.
Varsayılan değer şudur: "1'den az belgede görünen terimleri yoksay" anlamına gelir. Bu nedenle, varsayılan ayar hiçbir terimi göz ardı etmez.min_df:1

"""
BoW_Matrix = BoW_Vector.fit_transform(liste)
#print (BoW_Matrix)
features = BoW_Vector.get_feature_names()#BoW_Transformer.get_feature_names()
#print ("features[46]:" + features[46])
#print ("features[48]:" + features[48])
BoW_Matrix = BoW_Matrix.toarray()
BoW_df = pd.DataFrame(BoW_Matrix, columns = features)
BoW_df
```

	19	bir	cok	covid	covid19	degil	icin	kadar	maske	mi	olarak	olsun	pozitif	saqlik	tedbirleri	tum	var	
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1

Kaynak: Yazar tarafından derlenmiştir.

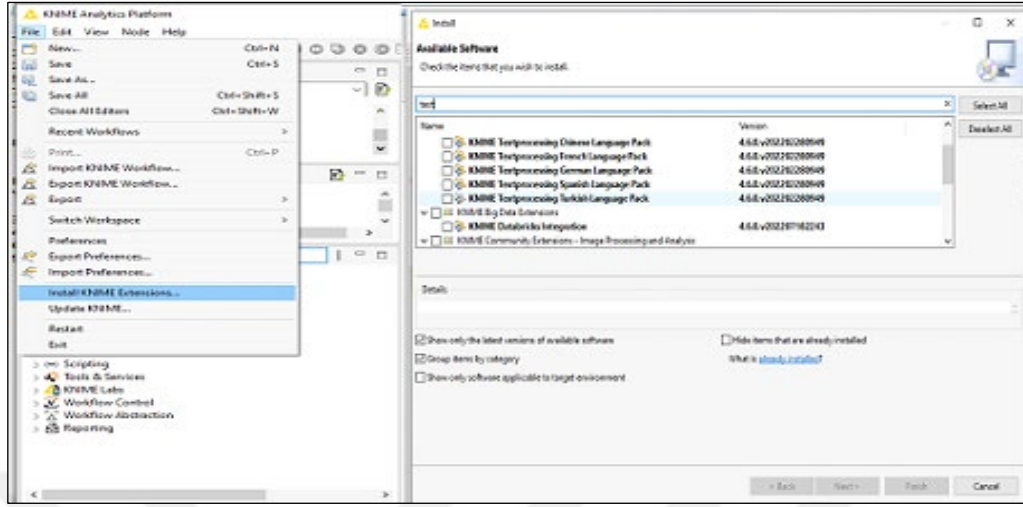
Google Colab üzerinde yapılan kök bulma çalışması aşamaları ve sonrasında elde edilen kelime çantasının doğru sonuçlar üretmemesi nedeniyle çalışma geliştirme aşamasına gidilmiş ve Knime uygulaması üzerine taşınmıştır.

KNIME, Java dili ile geliştirilmiş farklı uygulama ait uzantıları Node adı verilen düğümler ile uygulamaya dâhil etme imkanı sağlayan çok dilli, açık kodlu bir yazılım geliştirme ortamı olarak tanımlanmaktadır. (KNIME Open Source Story)

KNIME uygulamasına ait resmi web adresinden uygulama indirilmiş ve kurulumu gerçekleştirilmiştir. Daha sonra çalışmanın bir önceki bölümünde bahsi geçen kelime köklerinin daha doğru bulunabilmesi amacı ile KNIME metin işleme düğümleri kullanılmıştır.

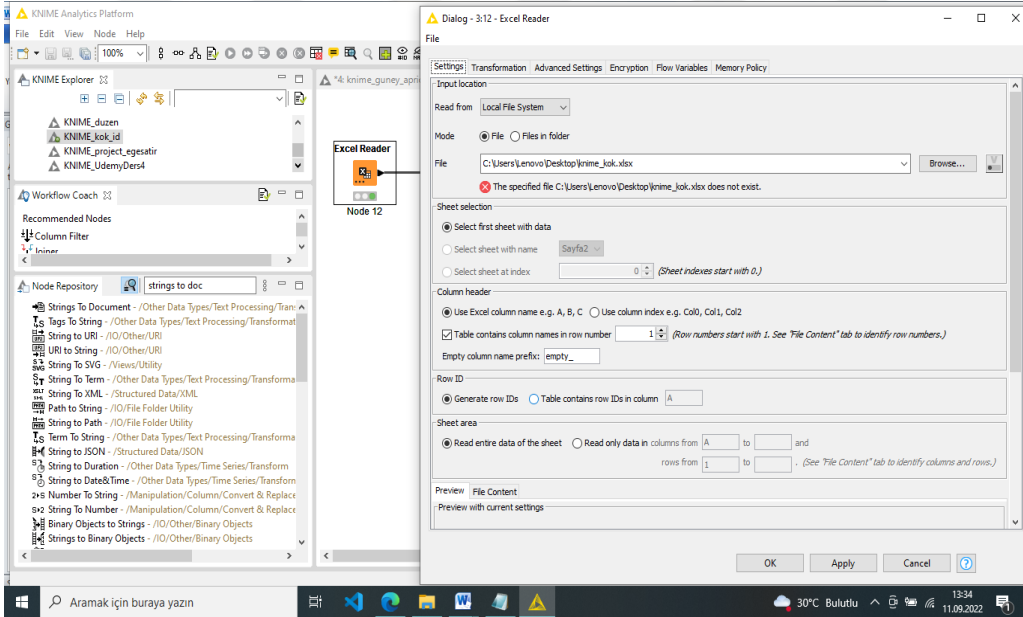
Şekil 54’ te görülen işlemin ardından KNIME uygulaması düğüm deposu sekmesinde Zembek Stemmer kullanılabilir olarak yer almıştır.

Şekil 54. KNIME Install Ekranı Türkçe Metin İşleme Eklentisi Yükleme



Kaynak: Yazar tarafından derlenmiştir.

Şekil 55. Ege Bölgesi Kelime Bulutu

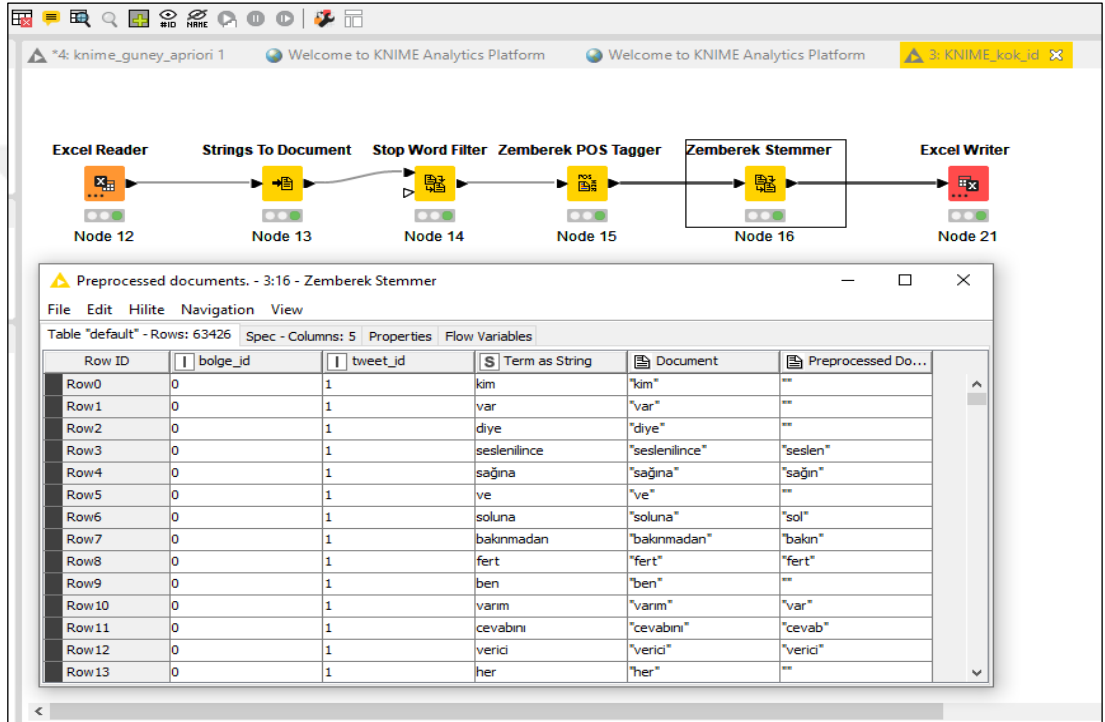


Kaynak: Yazar tarafından derlenmiştir.

Çalışmanın önceki bölümlerin veri setine uygulanan ön işleme adımlarına yer verilmiştir. Bu adımda KNIME uygulamasını kullanılarak Zemberek Stemmer

algoritması ile kelimelerin köklerinin daha doğru bulunması amaçlanmıştır. Bu amaç ile Knime uygulamasına Excel Reader düğümü kullanılarak veri setinin ön işlemden geçilerek kök bulmak için hazırlanmış Excel dosyası yüklenmiştir. Ardından Şekil 56' da gösterilmiş olan düğümler kullanılarak Zemberek Stemmer algoritması ile kelime kökleri bulunmuştur.

Şekil 56. KNIME Zemberek Stemmer ile Kökler Bulundu.



Row ID	bolge_id	tweet_id	Term as String	Document	Preprocessed Do...
Row0	0	1	kim	"kim"	""
Row1	0	1	var	"var"	""
Row2	0	1	diye	"diye"	""
Row3	0	1	seslenilince	"seslenilince"	"seslen"
Row4	0	1	sağına	"sağına"	"sağın"
Row5	0	1	ve	"ve"	""
Row6	0	1	soluna	"soluna"	"sol"
Row7	0	1	bakınmadan	"bakınmadan"	"bakın"
Row8	0	1	fert	"fert"	"fert"
Row9	0	1	ben	"ben"	""
Row10	0	1	varım	"varım"	"var"
Row11	0	1	cevabını	"cevabını"	"cevab"
Row12	0	1	verici	"verici"	"verici"
Row13	0	1	her	"her"	""

Kaynak: Yazar tarafından derlenmiştir.

Şekil 56' da gösterile düğümler ile bağlantılar kurulmuştur. Zemberek Stemmer düğümü sonuçları sağ tık menü Preprocessed documents ile incelenmiş ve Snowball Stemmer algoritmasından daha başarılı kök bulma gerçekleştiği görülmüştür. Elde edilen yeni kökler Excel Writer düğümü ile yerel dosya olarak kaydedilmiştir. Yeni kelime kökleri ile çevrimiçi wordart ortamı kullanılarak üç kelime bulutu elde edilmiştir.

Şekil 57. Ege Bölgesi ve Güneydoğu Anadolu Bölgesi Paylaşımları Kelime Bulutu



Kaynak: <https://wordart.com/create>

Şekil 57’de çalışmanın önceki bölümlerinde seçim sebeplerine değinilmiş olan iki bölgeyi temsili altı ile ait 11 Mart 2020 ile 25 Haziran 2020 tarihli Twitter paylaşımlarından elde edilmiş olan kelime bulutuna yer verilmiştir. Şekil 57 kabaca değerlendirildiğinde altı ilden yapılan paylaşımların en çok sağlık, covid, pozitif ve maske köklerini içerdiği görülmektedir.

Çalışma kapsamında veri kazıma yöntemi Ege Bölgesine ait 3.663, paylaşım toplanmıştır. İçerikler metin madenciliği ön işleme adımlarından geçirilmiş 3.531 tweet veri analizi alınmıştır. Google Colab üzerinde Snowball Stemmer algoritması ve stop wordsler uygulanmış toplam 60.999 kelime köküne ulaşılmıştır. Kelime kökleri gözetimli olarak değerlendirildiğinde durak kelimelere ve köklere ait bir takım hatalar gözlemlenmiştir. Çalışma KNIME uygulaması üzerinde Zemberek Stemmer algoritması ve stopwords’leri kullanılarak geliştirilmiştir.

Şekil 58’de görülen Ege Bölgesine ait kelime bulutu incelendiğinde bölgelere ait Şekil 57’de gösterilen kelime bulutu ile benzer olarak en çok Covid kökünün paylaşımlarda kullanıldığı görülmektedir. Bir, Çok, Tüm gibi köklerinde paylaşımlarda sıklıkla kullanıldığı görülmektedir.

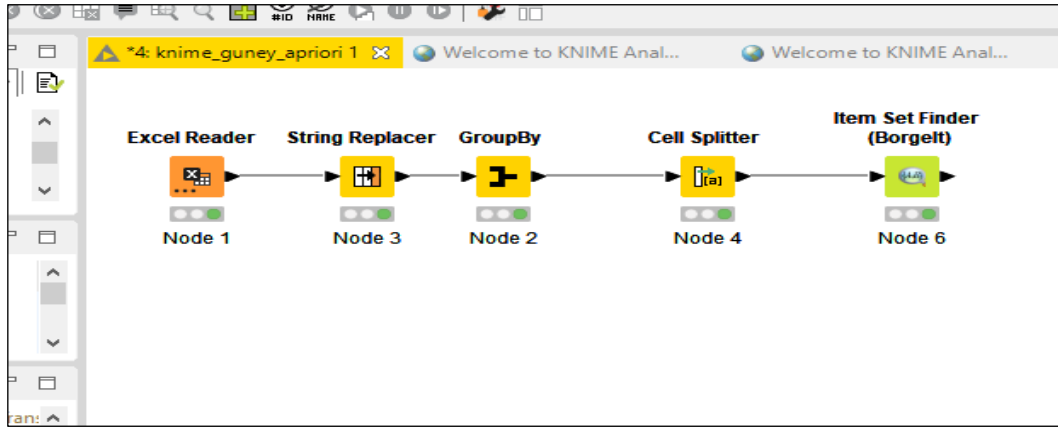
Çalışma kapsamında veri kazıma yöntemi Güneydoğu Anadolu Bölgesine ait 358 paylaşım toplanmıştır. İçerikler metin madenciliği ön işleme adımlarından geçirilmiş ve veri analizi alınmıştır. KNIME uygulaması üzerinde Zemberek Stemmer algoritması ve stopwords’leri kullanılarak 3.828 adet kelime köküne ulaşılmıştır.

Şekil 59’da görülen Güneydoğu Anadolu Bölgesine ait kelime bulutu incelendiğinde diğer iki kelime bulutu ile benzer olarak en çok Covid kökünün paylaşımlarda kullanıldığı görülmektedir. Bir, Batman, İçin, Çok ve Sağlık gibi köklerinde paylaşımlarda sıklıkla kullanıldığı görülmektedir.

Paylaşımlar üzerinde yapılan bir başka analiz ise paylaşımların bütünü oluşturarak kelime kökleri arasında var olan birliktelik kurallarını keşfetmeyi amaçlamaktadır. Birliktelik kurallarının analizi için her iki bölgeye ait paylaşımlar ayrı değerlendirilmiş, bölgelere ait paylaşımlar içerisindeki birliktelik kuralları sorgulanmıştır.

Birliktelik kurallarının analizi amacı ile Knime uygulaması üzerinde Şekil 60’da gösterilen düğümler ve akış kullanılmıştır.

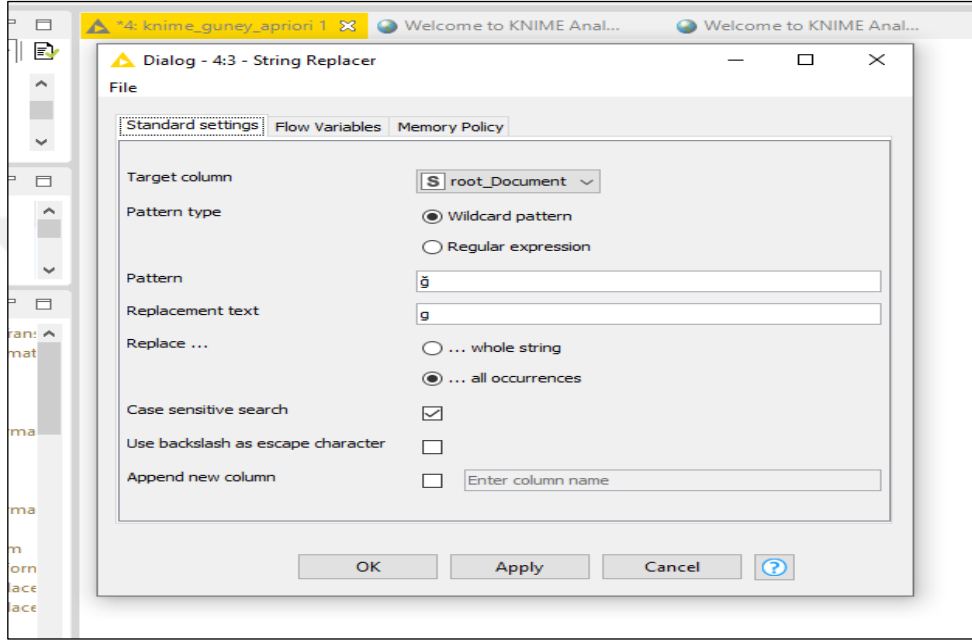
Şekil 60. KNIME Kökler Arası Birliktelik Analizi



Kaynak: Yazar tarafından derlenmiştir.

Şekil 60’da görülen düğümlerin kullanılabilir olması için Şekil 54’de gösterilen Knime Install kısmından Itemset Mining eklentisini kurulumu gerçekleştirilmiştir. String Replacer düğümü, veri kümesi içerisindeki Türkçe karakterleri İngilizceleri ile değiştirmektedir.

Şekil 61. KNIME String Replacer

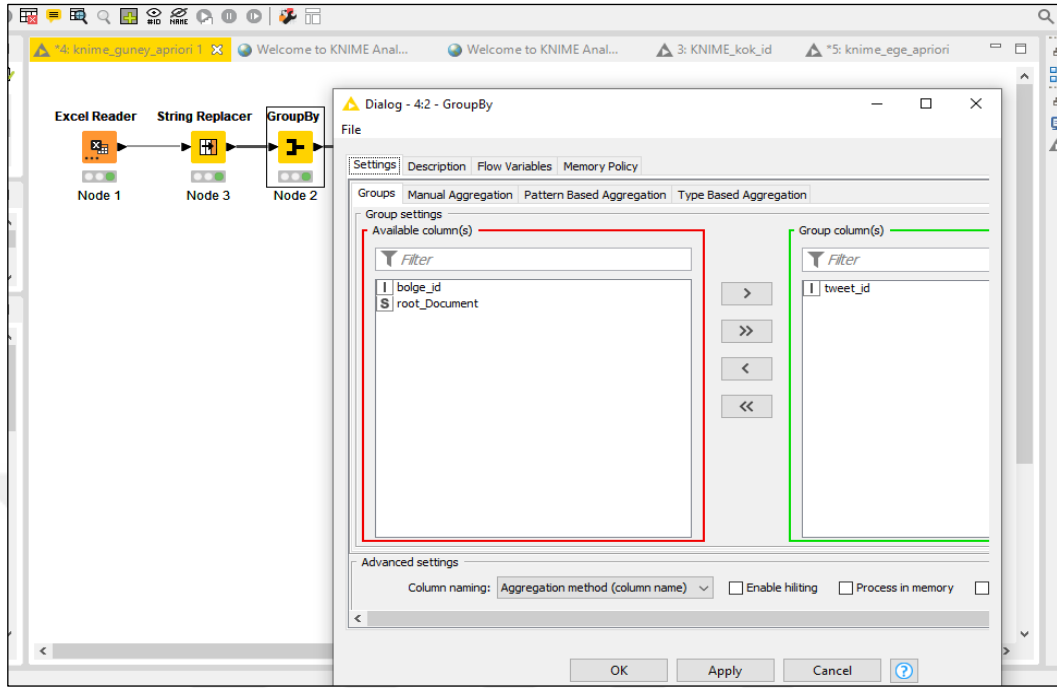


Kaynak: Yazar tarafından derlenmiştir.

Sağ tık menü configure menüsünde Şekil 61’de gösterildiği şekilde düzenlemeler gerçekleştirilir. Group By düğümünde ise metin ön işleme adımında her biri ayrı birer jeton haline getirilen kökler tweet_id göre gruplandırılır.

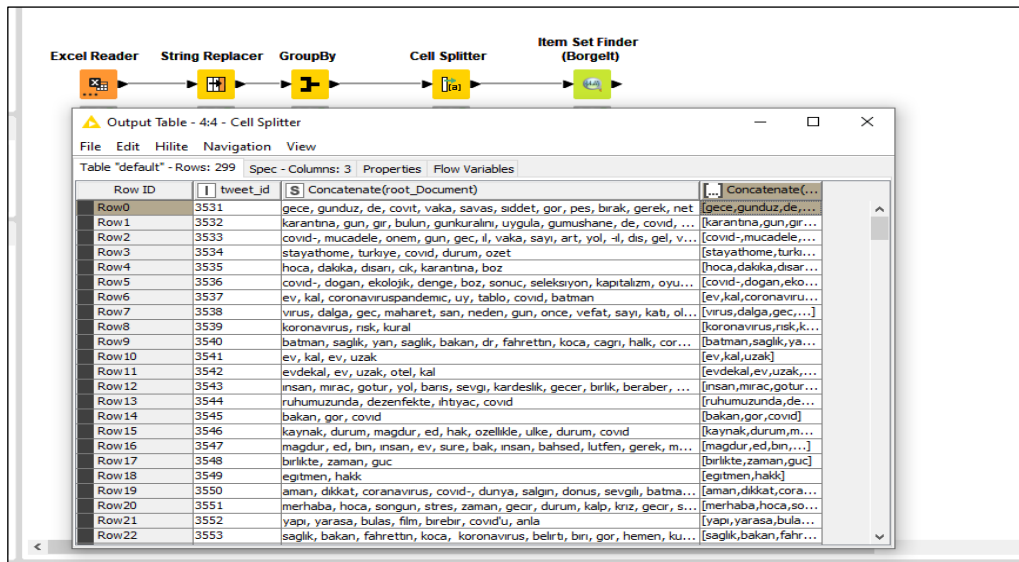
Cell Spiliter, düğümü ile her tweet farklı öğelerden oluşan bir dizi olarak ifade edilir. Bu bir sonraki adım için zorunlu bir ön işlemdir.

Şekil 62. KNIME GroupBy



Kaynak: Yazar tarafından derlenmiştir.

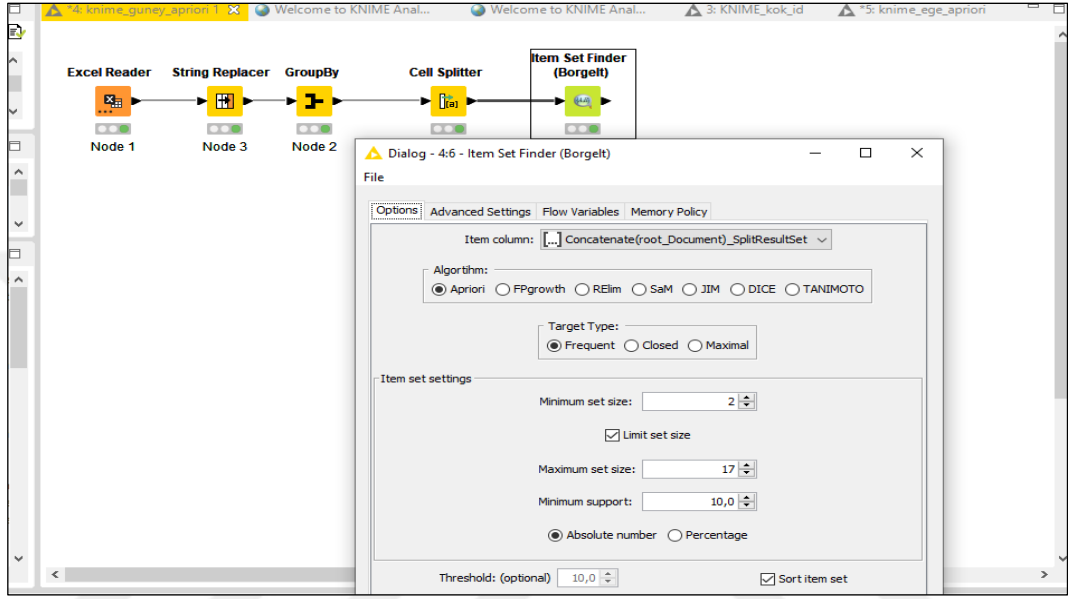
Şekil 63. KNIME Cell Splitter



Kaynak: Yazar tarafından derlenmiştir.

Itemset Finder, düğümü içerisinde bulunan farklı algoritmaların kullanılmasını sağlar Şekil 64'te gösterilen configure sekmesi işlemleri ile Apriori algoritması kullanılarak kökler arası birliktelik analiz gerçekleştirilmiştir.

Şekil 64. KNIME Item Set Finder



Kaynak: Yazar tarafından derlenmiştir.

Bu adım ile birlikte kökler içerisindeki birliktelik kurallarının analizi için model kurulumu tamamlanmıştır. Ege Bölgesi ve Güneydoğu Anadolu Bölgesine ait kök dosyaları modele çekilir ve model çalıştırılır.

Ege Bölgesine ait birliktelik analizi sonuçlarına Şekil 65'de Güneydoğu Anadolu Bölgesine ait birliktelik analizi sonuçlarına Şekil 66'da yer verilmiştir.

Ege Bölgesine ait paylaşımlarda kullanılan kökler arasında 17 adet köke kadar birliktelik varlığı sonucuna analiz sonucu ulaşılmıştır. Bahsi geçen 17 adet 11 paylaşımda birlikte kullanılmıştır. Sonuç ham veri üzerinde araştırıldığında o dönem birden fazla kişiye başarı belgesi dağıtıldığı, Milli Eğitime ait bir sınavın varlığı gibi konular nedeni ile paylaşımların bu kökler üzerinde yoğunlaştığı görülmüştür.

Şekil 65. KNIME Ege Bölgesi Birliktelik Çıkarımları

Row ID	ItemSet	ItemSetSize	ItemSetSupport
Row21020	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	17	11
Row45596	[ac,belge,bilgi,cals,covid-denetim,ed,hakk,kayıt,kontrol,mudur,tedbir,teslim,yeni,ilce,isle]	16	14
Row21021	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yeni,ilce,isle]	16	11
Row21022	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row21024	[ac,belge,bilgi,cals,hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row21028	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row21036	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row21052	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row21084	[ac,belge,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row21148	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row21276	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row21532	[ac,belge,bilgi,cals,covid-hakk,kayıt,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row22044	[ac,belge,bilgi,cals,covid-kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row23068	[belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row25116	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce]	16	11
Row25122	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row37404	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,yap,yeni,ilce,isle]	16	11
Row68124	[ac,belge,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,uyarı,yap,yeni,ilce,isle]	16	11
Row118434	[ac,bilgi,cals,covid-hakk,kayıt,kontrol,mudur,personel,tedbir,teslim,uyarı,yap,yeni,ilce,isle]	16	11
Row4616	[bilgi,cals,covid-denetim,ed,gercek,gorev,hakk,kontrol,mudur,rutn,tedbir,uyaret,ilce,isle]	15	16
Row45597	[ac,belge,bilgi,cals,covid-denetim,hakk,kayıt,kontrol,mudur,tedbir,teslim,yeni,ilce,isle]	15	14

Kaynak: Yazar tarafından derlenmiştir.

Şekil 66. KNIME Güneydoğu Anadolu Bölgesi Birliktelik Çıkarımları

Row ID	ItemSet	ItemSetSize	ItemSetSupport	RelativeItemSetSupport...
Row16	[asa,covid-]	2	27	9.03
Row19	[covid-salgın]	2	16	5.351
Row0	[covid-pandem]	2	14	4.682
Row14	[cik,pozitif]	2	14	4.682
Row1	[covid-turkiye]	2	13	4.348
Row9	[covid-saglik]	2	13	4.348
Row12	[covid-mucadele]	2	13	4.348
Row13	[covid-gele]	2	13	4.348
Row18	[covid-ck]	2	13	4.348
Row21	[covid-ol]	2	13	4.348
Row5	[pozitif,testi]	2	12	4.013
Row11	[cals,covid-]	2	12	4.013
Row4	[covid-konu]	2	11	3.679
Row6	[cik,pozitif,testi]	3	11	3.679
Row7	[cik,testi]	2	11	3.679
Row8	[covid-vaka]	2	11	3.679
Row10	[covid-yap]	2	11	3.679
Row15	[covid-pozitif]	2	11	3.679
Row2	[covid-ulke]	2	10	3.345
Row3	[covid-sifa]	2	10	3.345
Row17	[covid,ck]	2	10	3.345
Row20	[batman,covid-]	2	10	3.345

Kaynak: Yazar tarafından derlenmiştir.

Güneydoğu Anadolu Bölgesine ait paylaşımlarda kullanılan kökler arasında ikili ve üçlü birlikteliklerin varlığı sonucuna ulaşılmıştır. Birliktelik tamamı covid kökü ile şekillenmekte olup en çok tekrar eden ikili kök aşı ve covid kökleridir. Üçlü birliktelik kuralı ise çık, pozitif ve test kökleri arasında görülmüş olup destek sayısı on birdir.

SONUÇ

Bulgular değerlendirildiğinde Kelime Çantasını ifade etmekte olan matrisin büyük bir çoğunluğunun “0” değerini almasının nedeni kelime köklerinin çok çeşitlilik gösterdiği olarak ifade edilebilir ancak burada kelime kökleri incelendiğinde hecelerin kelime köklerine başarılı bir şekilde ayrılmadığı görülmüştür. Bu durumdan hareket ile Stremming kütüphanesi ve snowballstemming algoritmasının Türkçe dili üzerinde çok başarılı olmadığı kanaatine ulaşılmıştır. Yapılan kaynak taramasının desteklediği üzere Zemberek Kütüphanesinin Türkçe dili hece köklerini daha başarılı ayırdığına yer verilmesi nedeniyle uygulamanın Zemberek Kütüphanesi ile geliştirilmesi gerekliliğinin varlığı kabul edilmiştir.

Çalışma Zemberek algoritmasının kullanılması özelinde KNIME uygulamasına taşınmıştır. Açık kaynaklı kodlu bu yazılım farklı uygulamalara ait örneğin Wekaya ait Apriori vb. büyük veri algoritmalarından BERT gibi algoritmaları tek bir projede kullanma kolaylığı sağlamıştır. Aynı zamanda MIT App Inventor’ a benzeyen tasarımı düğüm yapısı ile proje geliştirmede kolaylık sağlamaktadır.

Google Colab sağlamış olduğu çok dilli yazılım geliştirme imkanı ve sanal makine sağlaması gibi fırsatları kullanıcılara vermektedir. Ancak bazı kütüphanelerin örneğin veri kazıma için snsrape, Türkçe dil destekli Zemberek gibi Google Colab üzerine yüklenmesi çok zor gerçekleştirilmektedir.

Çalışma kapsamında metin madenciliği ön işleme adımları için birçok uygulama ve çevrimiçi ortam kullanılmıştır. Veri kazıma konusunda en iyi başarı Visual Studio Code uygulaması ile gerçekleştirilmişken veri analizi en etkin KNIME uygulaması ile gerçekleştirilmiştir.

Metin madenciliği ön işleme adımları konusunda Microsoft Excel, Google Colab ve Knime deneyimlenmiştir. Her üç uygulamada jetonlara ayrılması aşamasına kadar metin madenciliği ön işleme adımlarını gerçekleştirebilmektedir. Kelime köklerinin bulunması aşamasını Microsoft Excel uygulamasında yazar tarafından gerçekleştirilememiştir. Kelime kökleri Google Colab üzerinde Snowball Stemming algoritması kullanılarak gerçekleştirilmiş olup 60.999 kök bulunmuştur. Kelime kök bulma işlemi Knime uygulamasında Zemberek Stemming algoritması kullanılarak gerçekleştirilmiş olup 45.515 kök bulunmuştur. İki algoritma karşılaştırıldığında 3.531

Türkçe paylaşım için Zemberek Stemming daha başarılı bulunmuştur. Çünkü algoritmaların çıktıları olan kökler gözetimli olarak ham halleri ile karşılaştırıldığında Snowball Stemming'de aynı kökten türeyen kelimelerin farklı kökler olarak döndürüldüğü görülmüştür.

Çalışma kapsamında üretilen kelime bulutları her iki bölge içinde benzer köklerden oluşmaktadır. Kelime bulutlarının odağında Covid 19 kökü bulunmaktadır. Ancak Ege Bölgesi birliktelik kuralları dikkate alındığında paylaşımların daha çok Covid 19 odağında alınan kurumsal tedbir ve bu tedbirlerin yürütülmesi, Covid 19 ortamında çalışanlara teşekkür gibi normalleşme konuları dikkat çekmektedir.

Güneydoğu Anadolu Bölgesi birliktelik kuralları dikkate alındığında paylaşımların daha çok Covid 19 odaklı test, ol, vaka, çık, aşı gibi kelime köklerinden oluştuğu görülmektedir.

Bölgelere ait veri setleri değerlendirildiğinde analize alınan veri setlerinin büyüklüğünün Ege Bölgesi 3.531, Güneydoğu Anadolu Bölgesi 358 adet ile adil dağılmadığı bilinmektedir. Bu sebeple hipotezlerin her ikisi içinde bir sonuca ulaşmak doğru olmayacaktır. Ancak sosyal medya paylaşım oranlarına bakıldığında ve Ege Bölgesi birliktelik kuralları çıkarımı kanaati göz önünde bulundurulduğunda Ege Bölgesi halkının normalleşme sürecine katılım için daha gönüllü olduğu çıkarımında bulunulabilir.

Yapılan çalışma veri madenciliği, metin madenciliği ve coğrafi bilgi sistemlerini içermekte olup öncelikli olarak yönetim bilişim sistemleri altında yer alan uzmanlıklar konusunda yazara deneyim ve beceri kazandırmıştır.

Yönetim Bilişim Sistemleri geçmiş verilerden bilgi çıkarımı ile geleceği planlamakta ve yönetim bilişim piramidini bilgelik düzeyine çıkarmaktadır. Yönetim Bilişim Sistemleri alanının ve uzmanlarının sorumluluklarının özü daha etkin süreçler daha az maliyetleri kapsamaktadır. Bu çalışmanın konusunu oluşturan Covid 19 virüsünü maliyetleri öncelikli olarak can kaybı, sağlık refahı ve ekonomik kayıp olarak sıralanmaktadır. Pandemi süreci ve aşılama devam etmektedir. Sosyal medyada yapılan paylaşımlardan insanların tutumları anlaşılması mümkün olur ise bu daha sağlıklı politikalar daha etkileyici adımlar ve her anlamda daha az maliyet olarak tanımlanabilir.

Çalışma bölgesel anlamda çıkarımlarda bulunmak amacı ile kurgulanmış bir sorun üzerine planlanmıştır. Çalışmanın amacına ve sorun tanımına uygun olarak iki bölgeden altı il seçilmiştir. Ancak çalışmanın ilgili bölümlerinde bahsi geçtiği üzere iki bölgeye ait veri seti büyüklüğünün uygunluğu sağlanamamıştır. Bu sebeple çalışma bölge ve seçilen temsili illerin nüfusu üzerinden tekrar planlanabilir. Bu planlama esnasında Güneydoğu Anadolu Bölgesi temsili il sayısı artırılarak bölgeye ait veri seti büyüklüğü genişletilebilir. Bu çalışma içerisinde veri ve metin madenciliği için birçok uygulama deneyimlenmiştir. Çalışma Knime programı üzerinde yoğunlaştırılabilir. Veri setlerini oluşturan kelime kökleri ve durak kelimeleri için iki yöntem denenmiştir. Kelime köklerinin bulunması ve durak kelimelerin veri setinden çıkarılmasında daha farklı algoritma ve yöntemler denebilir. Çalışma kapsamında gerçekleştirilen metin madenciliği süreci sonuçları bölgelerde gerçekleştirilen anketler ile sorgulanarak uygulama genişletilebilir.

KAYNAKÇA

03 - *String İşlemleri*. (tarih yok). Eğitim Bilgileri: <https://sites.google.com/site/egitimbilgileri/home/a---python---twisted---qt/03---string-islemleri> adresinden alınmıştır.

5 Temel Soruda Veri Madenciliği (Data Mining) Nedir? (26, 01 2019). 18 01, 2022 tarihinde vizyoner genç: <https://vizyonergenc.com/icerik/5-temel-soruda-veri-madenciligi-data-mining-nedir> adresinden alınmıştır.

Akpınar, Haldun. (2000, Nisan). Veri Tabanlarında Bilgi Keşfi Ve Veri Madenciliği. *İ.Ü İşletme Fakültesi Dergisi*, 29(1), 1-22.

Artan Özoran, B. (2020). Covid-19 Salgınında Markaların Twitter Kullanımları Üzerine Bir İnceleme. *Van Yüzüncü Yıl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*(Salgın Hastalıklar Özel Sayısı), 429-458.

Aşman Alikılıç, Ö. (2011). *Halkla İlişkiler 2.0 Sosyal Medyada Yeni Paydaşlar, Yeni Teknikler* (Cilt 1). Ankara: Efil Yayınevi.

Ateş, Nurbanu., & Baran, Sinan. (2020). Kriz İletişiminde Sosyal Medyanın Etkin Kullanımı: Covid-19 (Koronavirüs) Salgınına Yönelik Twitter Analizi. *kilad-Kocaeli Üniversitesi İletişim Fakültesi Araştırma Dergisi*, 66-99.

Ay, Derya. (2009, Şubat). Veri Madenciliği Ve Apriori Algoritması İle Süpermarket Analizi. *Yüksek Lisans Tezi*. Sakarya: Sakarya Üniversitesi Fen Bilimleri Enstitüsü.

Aydoğan, Sefer., & Kutlu Gündoğdu, Fatma. (2021). Sağlık Bakanı Fahrettin Koca'nın Covid-19 Salgınında Sergilediği Dijital Liderliğin Analitik Hiyerarşi Süreciyle İncelenmesi (Twitter Vaka Çalışması). *İşletme Araştırmaları Dergisi*, 13(2), 1737-1750.

Ballı, Çağla. (2021). Doğal Dil İşleme İle Türkçe İçerikli Paylaşımlardan Sosyal . *Yüksek Lisans Tezi*. Ankara: Ankara Üniversitesi Fen Bilimleri Enstitüsü.

Baskın, Hasan. (2021, 01 31). *Doğal Dil İşleme (Natural Language Processing)*. 05 2022 tarihinde {_Hasan BASKIN_}: <https://www.hasanbaskin.com/dogal-dil-isleme-natural-language-processing/#dogaldilisleme> adresinden alınmıştır.

Bektaş, Melike. (2019, 12 01). *Zemberek Kütüphanesi ile Türkçe Metinlerde Kelime Köklerinin Bulunması*. Mayıs 2022 tarihinde Melike Bektaş: <https://melikebektas95.medium.com/zemberek-k%C3%BCt%C3%BCphanesi-ile-t%C3%BCrk%C3%A7e-metinlerde-kelime-k%C3%B6klerinin-bulunmas%C4%B1-6ddd3a875d5f> adresinden alınmıştır.

Bezirci, Pervin., & Biçen Aras, Bahar. (2022). Covid-19 Ve Üniversite kütüphanelerinin Twitter Kullanımı. *Kütüphane, Arşiv ve Müze Araştırmaları Dergisi*, 3(1), 1-26.

Bilgiç, Alican., & Akyüz, Selman. Selim. (2020). Türkiye'de Covid-19 Pandemisi Döneminde Sağlık Bakanı Fahrettin Koca'nın Sosyal Medya Kullanımı: Twitter Paylaşımları İçerik Analizi. *Gaziantep University Journal of Social Sciences, Special Issue*, 230-243.

Bozkurt, Aybüke. (2021). Veri Madenciliği Yöntemleri İle Tam Kan Sayımı Sonuçlarından Covid-19 Test Sonuçlarının Tahmini. *Yüksek Lisans Tezi*. Konya: Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü.

Bringing MySQL to the web About. (tarih yok). 03 20, 2022 tarihinde phpMyAdmin: <https://www.phpmyadmin.net/> adresinden alındı

Ceyhan, Görkem. (2020, Temmuz). Sınıflandırmada Kullanılan Veri Madenciliği. *Doktora Tezi*. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü.

Covid-19 Nedir ? (tarih yok). 03 07, 2022 tarihinde T.C. Sağlık Bakanlığı Covid 19 Bilgilendirme Platformu: <https://covid19.saglik.gov.tr/TR-66300/covid-19-nedir.html> adresinden alınmıştır.

Demircioğlu, Mert., & Eşiyok, Sevgi. (2019). Covid–19 Salgını İle Mücadelede Kümeleme Analizi İle Ülkelerin Sınıflandırılması. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi, Covid-19 Sosyal Bilimler Özel Sayısı*(37), 369-389.

Demirok, Yavuz. (2018). Birliktelik Kuralı Yöntemleri ile E-Ticaret Satışlarının Analizi. *Yüksek Lisans Tezi*. İstanbul: Maltepe Üniversitesi Fen Bilimleri Enstitüsü.

Elif Gün, Suat. Kolukırık. (2022). Covid-19 Döneminde Toplumsal Risk Algısı, Refleksivite Ve Belirsizlik: Twitter Paylaşımları Üzerine Bir Analiz. *Sosyal Bilimler Dergisi Journal Of Social Sciences*, 32(1), 291-305.

Elmas, Şeker. Şerife. (2019). Sosyal Medya Mesajlarının Veri Madenciliği Yöntemi İle Duygu Analizi (Sivas İli Örneği). *Yüksek Lisans Tezi*. Sivas: Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü.

Güncel Sarıman, Emre. Mutaf. (2020). Covid-19 Sürecinde Twitter Mesajlarının Duygu Analizi. *Euroasia Journal Of Mathematics, Engineering, Natural & Medical Sciences International Indexed & Refereed*, 7(10), 137-148.

Hastalık. (2021, Ekim 11). 03 07, 2022 tarihinde Vikipedi Özgür Ansiklopedi: <https://tr.wikipedia.org/wiki/Hastal%C4%B1k> adresinden alınmıştır.

Information Technology Gartner Glossary. (1979). Ocak 17, 2022 tarihinde Gartner: <https://www.gartner.com/en/information-technology/glossary/data-mining> adresinden alınmıştır.

Johnston, Loundon. (2008, 10 05). *Noktalama işaretini Python'da bir dizeden çıkarmak için en iyi yol*. 05 2022 tarihinde web-gelistirme-sc.com: <https://www.web-gelistirme-sc.com>

gelistirme-sc.com/tr/python/noktalama-isaretini-pythonda-bir-dizeden-cikarmak-icin-en-iyi-yol/958542404/ adresinden alınmıştır.

Kalikov, Anarbek. (2006, Mayıs). Veri Madenciliği Ve Bir E-Ticaret Uygulaması. *Yüksek Lisans Tezi*. Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

Kara, Şeyma. Aliye., Sertçelik, Ahmet., & Çakır, Banu. (2020, 03 30). *Salgın ve Salgın ile İle İlgili Tanımlar Bilgilendirme Notu-HÜTF Halk Sağlığı AD Toplum İçin Bilgilendirme Serisi 2019/2020-41*. 03 07, 2022 tarihinde Hacettepe Üniversitesi Halk Sağlığı Anabilim Dalı: <http://www.halksagligi.hacettepe.edu.tr/duyurular/halkayonelik/salgin.html> adresinden alınmıştır.

Karlı, İhsan., & Dondurucu, Zenan. Benan. (2020). Eleştirel Söylem Çözümlemesi Bağlamında Fox News ve CNN'in COVID-19 Salgınında Twitter Kullanımı. *İnsan&İnsan*, 7(26), 163-186.

Kılınc, Deniz. (2018, 05 14). *Metin İşleme 1 — Eski Tarz Yöntemler (Bag of Words ve TFxIDF)*. 05 2022 tarihinde Published in Deep Learning Türkiye: <https://medium.com/deep-learning-turkiye/metin-i%CC%87%C5%9Fleme-1-eski-tarz-y%C3%B6ntemler-bag-of-words-ve-tfxidf-76d5a0cf1b29> adresinden alınmıştır.

KNIME Open Source Story. (tarih yok). 09 2022 tarihinde Open for Innovation KNIME. adresinden alınmıştır.

Kuş, Oğuz. (2021). Kovid-19 Salgını ve Mültecilere Yönelik Dijital Nefret Söylemi: Büyük Veri Perspektifinden Metin Madenciliği Tekniği ile Kullanıcı Kaynaklı İçeriklerin Analizi. *TRTakademi*, 6(11), 108-131.

Laudon, Kenneth. C., & Laudon, Jane. Price. (1994). *Management Information Systems Managing The Digital Firm* (Twelfth Edition B.). New Jersey: Pearson Prentice Hall.

Muslu Y. (2019, 10 10). *Python Türkçe karakterleri İngilizce karakterlere çevirme. maketrans*. 05 2022 tarihinde makdos blog: <https://makdos.blog/python/893/python-turkce-karakterleri-ingilizce-karakterlere-cevirme-maketrans> adresinden alınmıştır.

MySQL Nedir? (2020, Mart 01). 03 20, 2022 tarihinde biltek: www.biltektasarim.com/blog/mysql-nedir adresinden alınmıştır.

Opuz, Ceylan. (2019, 10 31). *Doğal Dil İşleme Nedir ve Uygulama Alanları Nelerdir?* 05 2022 tarihinde Veri Bilimi Okulu: <https://www.veribilimiokulu.com/dogal-dil-isleme-nedir-ve-uygulama-alanlari-nelerdir/#:~:text=Do%20dillerden%20kas%20insanlar%20konu%20Ftu%20Fu,zekas%20giderek%20parlamas%20da%20sa%20Flam%20> adresinden alınmıştır.

Özekes, Serhat. (2003). Veri Madenciliği Modelleri Ve. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 2(3), 65-82. <https://www.ticaret.edu.tr/uploads/yayin/dergi/d3/M00041.pdf> adresinden alınmıştır.

Özmelek Taş, Neslihan. (2015). Sosyal Medyada İzlenim Yönetimi. *Yüksek Lisans Tezi*. İzmir: Ege Üniversitesi Sosyal Bilimler Enstitüsü Halkla İlişkiler Ve Tanıtım Anabilim Dalı.

Özmen, Fatma., Aküzüm, Cemal., Sünkür, Meltem., & Baysal, Nigah. (2012). Sosyal Ağ Sitelerinin Eğitsel Ortamlardaki İşlevselliği. *E-Journal Of New World Sciences Academy*, 7(2), 497-506.

Pandemi. (2020, Temmuz 02). T.C. Sağlık Bakanlığı COVID-19 Bilgilendirme Platformu: <https://covid19.saglik.gov.tr/TR-66494/pandemi.html#:~:text=Bir%20hastal%20veya%20enfeksiyon%20etkeninin,DS%20taraf%20indan%20pandemi%20ilan%20edilmi%20tir.> adresinden alınmıştır.

RAHIMI, Zenab. (2015, Mart). Sanal Medyada Sosyal Ağlarda Reklam Uygulaması "Facebook" Örneği. *Yüksek Lisans Tezi*. Ankara: Gazi Üniversitesi Sosyal Bilimler Enstitüsü.

Savaş, Serkan., Topaloğlu, Nurettin., & Yılmaz, Mithat. (11). Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri. *Stambul Ticaret Üniversitesi Fen Bilimleri Dergisi, Bahar 2012*(21), 1-23.

snsrape 0.4.3.20220106. (2022, 01 06). 03 10, 2022 tarihinde Python Package Index: <https://pypi.org/project/snsrape/> adresinden alınmıştır.

Sosyal Medya. (2022, 04 14). Vikipedi Özgür Ansiklopedisi: https://tr.wikipedia.org/wiki/Sosyal_medya adresinden alınmıştır.

string.punctuation in Python. (2019, 10 15). geeksforgeeks: <https://www.geeksforgeeks.org/string-punctuation-in-python/> adresinden alınmıştır.

Tecim, Vahap. (tarih yok). *Yönetim Bilişim Sistemleri (YBS)*. Ocak 20, 2022 tarihinde Prof. Dr. Vahap Tecim: <https://vahaptecim.com.tr/yonetim-bilisim-sistemleri/> adresinden alınmıştır.

Temizhan, Ebru., & Mendeş, Mehmet. (2021). Covid-19 Pandemisi İle İlgili Twitter Mesajlarının Metin Madenciliği Tekniği İle Değerlendirilmesi. *Türkiye Klinikleri Biyoistatistik Dergisi*, 13(2), 185-200.

Topal, Sabriye. (2021, Ocak). Veri Madenciliği Tekniklerinin Bitcoin Piyasası. *Yüksek Lisans Tezi*. Giresun, Türkiye: T.C. Giresun Üniversitesi Sosyal Bilimler Enstitüsü.

Türkçe Etkisiz Kelimeler (Stop Words) Listesi 1.1. (2019, 01 01). 05 2022 tarihinde Türkçe Öğretimi Türkçe Öğrenimi Üzerine Her şey: <https://www.turkceogretimi.com/genel-konular/turkce-etkisiz-kelimeler-stop-words-listesi-11> adresinden alınmıştır.

Twitter. (2022, 05 01). 05 2022 tarihinde Vikipedi Özgür Ansiklopedi: <https://tr.wikipedia.org/wiki/Twitter> adresinden alınmıştır.

Twitter Hakkında. (tarih yok). 05 2022 tarihinde Twitter: <https://about.twitter.com/tr/who-we-are/our-company> adresinden alınmıştır.

Veli, Ahmet. (2020). Farklı Sektörlere Ait Büyük Verinin Birliktelik. *Yüksek Lisans Tezi*. Düzce: TC. Düzce Üniversitesi Fen Bilimleri Enstitüsü.

WampServer, a Windows web development environment. (tarih yok). 03 22, 2022 tarihinde WampServer: <https://www.wampserver.com/en/> adresinden alınmıştır.

Welcome To Colaboratory - Google Research. (2017). 05 2022 tarihinde Google Colab: https://colab.research.google.com/?utm_source=scs-index adresinden alınmıştır.

Welcome To Colaboratory - Google Research. (2017). 05 2022 tarihinde Google Colab: https://colab.research.google.com/?utm_source=scs-index#scrollTo=5fCEDCU_qrC0 adresinden alınmıştır.

WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. (2020, Mart 11). 03 07, 2022 tarihinde World Health Organization: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> adresinden alınmıştır.