

COVID-19 Twitter Veritabanı: Türkiye Örnekleme

Araştırmanın Amacı ve İçeriği

Bu veri seti çalışması, 2020 yılının ilk aylarından beri dünyada etkisini gösteren COVID-19 virüsünün Türkçe konuşan Twitter kullanıcılarının gündemine tesirini tespit etmek amacıyla kurulmuştur. Bu çalışma halihazırda birçok dilde hazırlanan kapsamlı bir Twitter veri setinin [#COVID-19: The First Public Coronavirus Twitter Dataset](#) tamamlayıcısı olarak görülebilir. Böyle bir çalışmanın şimdiye kadar yapılmamış olması bizi motive eden önemli bir nedendir. Yukarıda belirtilen çalışmada ise kelime havuzunun dar olması ve temalara göre ayrılması, Türkçe tweet sayısının az olması, Türkçe tweetleri ayıklamadaki zorluklar yeni bir veri seti çıkarmayı zorunlu hale getirmiştir.

“COVID-19 Twitter Veritabanı: Türkiye Örnekleme” adlı bu çalışmada tweetler hem temalara göre ayrı ayrı elde edilmiş hem de açıklanan ilk vakanın öncesi ve sonrasındaki gündemler ayrı ayrı tespit edilmiştir. Araştırma, aynı zamanda, Türkçe yazan Twitter kullanıcılarının Ortadoğu ve Avrupa’da vakaların artmasıyla beraber koronavirüsü gündemlerine aldıklarını ve Türkiye’de ilk vakanın açıklanmasıyla birlikte hem dini hem ekonomik hem politik hem de dış ülkeler hakkındaki gündemlerinde kırılmalar olduğunu varsaymaktadır. Bu raporda, bahsi geçen kırılmaları göstermek için en çok geçen kelimeler üzerinden bir ön analiz sunulacaktır.

Araştırmanın Dizaynı

Aşağıda bu araştırmanın dizaynı, içeriği ve kullanılan kodlar özetlenecektir. Aynı zamanda, bu veri setini kullanarak yürütülebilecek gelecek çalışmalara bazı tavsiyeler sıralanacaktır. İlk olarak, anahtar kelimeler temalara göre çıkartılmıştır. Daha sonra, tarih aralığı belirlenip tweetler çekilmiştir. Elde edilen tweetler, Twitter’ın [Terms of Service](#) ile uyumlu olarak ID’ler halinde herkese açık olarak depolanmıştır. En sonunda ise programlamalı metin analizi teknikleri kullanılarak tweetler üzerinden temel çıktılar alınıp takdim edilmiştir.

Anahtar kelimelerin çıkarılması ve temalara bölünmesi

Temalar, COVID-19 tartışmaları döneminde muhtemel olarak etkilenen veya etkilenebilecek alanlar olan politika, ekonomi, din, dezenformasyon, dış ülkeler hakkındaki algılar, ve virüsün kendisi olarak belirlenmiştir. Bu 6 başlık altındaki anahtar kelimeler rastgele tweet araması yapıp okuma yapılarak çıkarılmıştır:

- COVID-19: Covid-19, Covid 19, Kovid-19, korona, koronavirus, virüs, corona
- Politika: Bilim Kurulu, Bakan, Başkan, Sağlık Bakanlığı, Savaş, Seferberlik, Barış Pınarı, Fahreddin Koca, Cumhurbaşkanı
- Ülkeler: Çin, Avrupa, Amerika, İngiltere, İtalya, İspanya, Almanya, Fransa, Japonya, Güney Kore, İran, Türkiye, DSÖ (WHO), Nato, İsrail
- Dezenformasyon: Asılsız, provakatif, sahte, provokasyon, palavra, komplo, oyun, büyük oyun, siyonizm
- Din: Diyanet, cami, Cuma, muska, din, ezan, sela, sabır, müsibet, musibet,
- Ekonomi: Kredi, burs, işsiz, işsizlik, nakit, destek, dayanışma, gıda, mücadele, memur, şirket, gümrük

Verilerin Tarih Aralığının Belirlenmesi

- Yukarıda belirtilen kelime havuzları iki tarih aralığı belirlenerek ayrı ayrı depolanmıştır.
- Türkiye’de ilan edilen ilk vaka 11 Mart tarihindedir.
- Verileri önce ve sonra diye iki ayrı şekilde ayırmamızın amacı, gündemlerdeki süreklilik ve kırılımların belirlenmesini kolaylaştırmaktır.
- 11 Mart öncesi gündem, İran’da açıklanan ilk vaka tarihi olan 19 Şubat ile sınırlandırılmıştır. Buradaki amaç, Ortadoğu ve Avrupa bölgelerinde ilk vakaların açıklanmasını bir dönüm noktası olarak belirlemektir.
- Veriler 7 Nisan tarihine kadar çekilmiştir. Bu tarih Türkiye’de ilk vakanın açıklanmasından yaklaşık olarak bir ay sonrasına denk gelmektedir.
- Bu açılardan, bu dosyadaki veri seti ilk vakanın açıklanmasından önce ve sonraki gündem değişimlerini ve sürekliliklerini incelemeye imkan tanımaktadır.

Verilerin Depolanması ve ID'lerin Çıkarılması

- Twitter'ın [Terms of Service](#) ile uyumlu olarak toplanılan tweetlerin sadece Tweet ID'leri temalara bölünerek dosyalar halinde depolanmıştır. ID'ler çıkarılırken tekrar eden tweetler temizlenerek her tema için özgün tweetlerden oluşan veri setleri oluşturulmuştur.
- Bu ID'leri kullanarak çalışma yapmak isteyenler için [Hydrator](#) and [Twarc](#) gibi araçları kullanmayı tavsiye ediyoruz. Detaylı talimatlar için linklerdeki açıklamalara bakılabilir.

Kullanılan Kodlar

Bu çalışma, temel olarak üç kod setine dayanarak yapılmıştır.

- * İlk olarak Twitter'dan veri çekebilmek için Python dilinde yazılmış [Twint](#) paketinden yararlanılmıştır.
- * Daha sonra, elde edilen verileri düzenlemek ve ID'lerini çıkarıp kamusallaştırmak işlemleri gerekli [kodlar](#) kullanılarak gerçekleştirilmiştir.
- * En son olarak, içeriğe dair ilk fikirlerimizi elde etmek için [programlamalı metin analizi kodları](#) kullanılmıştır.

Temel Çıktılar

Bu bölümde ilk olarak tweetlerin temel istatistikleri sunulacaktır. Daha sonra, elde edilen tweetler üzerinden, ilk vakanın açıklanmasının öncesi ve sonrasında en çok geçen kelimeler veya anahtar kelimeler [tablolar](#) halinde görselleştirilecektir. Önemli görülen devamlılık ve kırılmalar kısaca not edilecektir. Daha detaylı analizler bir sonraki çalışmanın araştırma konusu olacaktır.

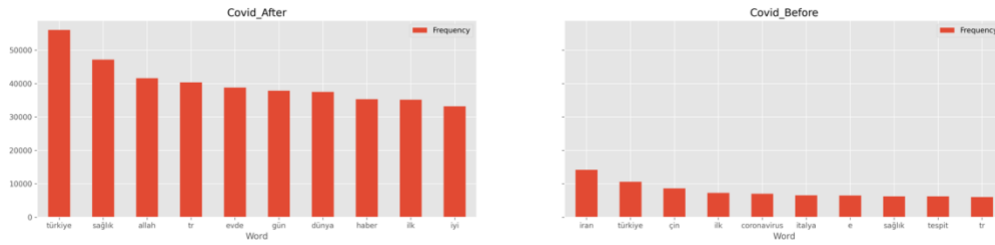
Temel İstatistikler

6 farklı temanın iki farklı zaman diliminden toplanan tweetlerin temel istatistikleri aşağıdaki gibidir:

Toplam Tweet Sayısı : **4,369,870**

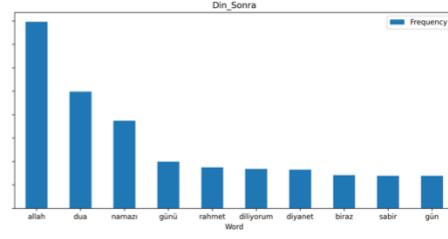
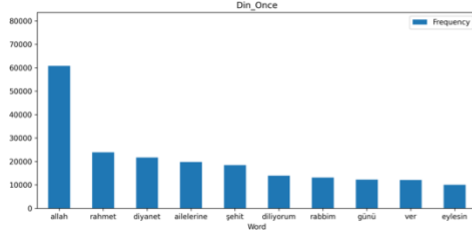
Tema	Önce	Sonra	Toplam
COVID	161,109	1,427,569	1,588,678
Politika	355,992	367,992	723,984
Dezenformasyon	51,521	88,007	139,528
Din	215,699	436,309	652,008
Ekonomi	19,532	74,682	94,214
Ülkeler	326,030	845,428	1,171,458
Toplam	1,129,883	3,239,987	4,369,870

Temel Görseller



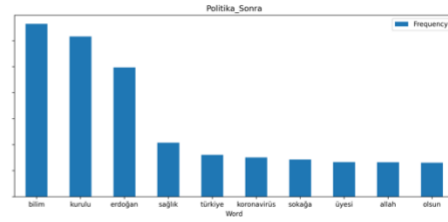
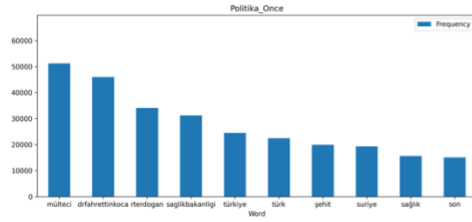
Tablo 1: COVID-19 hakkındaki en çok geçen kelimeler

- Öncesinde, İran, İtalya ve Çin gibi vakaların yoğunlukla görüldüğü yerlere sıkça referans verilmiştir. Bu bize, Türkiye’deki ilk vakanın açıklanmasından önce de dış ülkelere yönelik gündemin virüs yoğunluklu olduğunu göstermektedir. Vaka açıklandıktan sonra ise Allah ve sağlık gibi kelimeler öne çıkmaktadır. Bu durum bize Türkçe yazan Twitter kullanıcılarının süreci dini argümanlar ile açıkladığına dair bir fikir sunmaktadır.



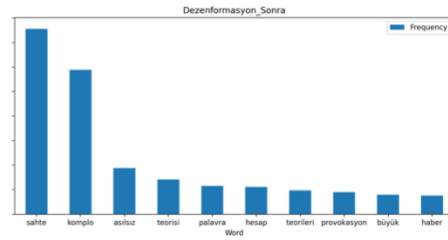
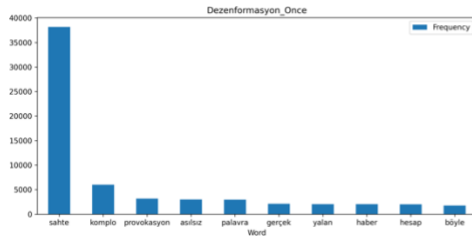
Tablo 2: Din hakkındaki en çok geçen kelimeler

- Daha önce şehit vurgusu yüksek iken, COVID-19 sonrası dönemde dua vurgusunun ön plana çıktığı görülmektedir.



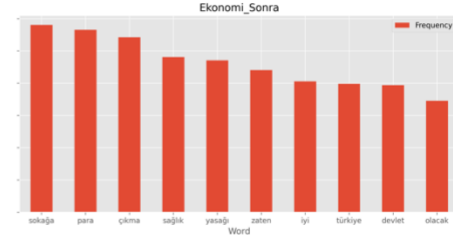
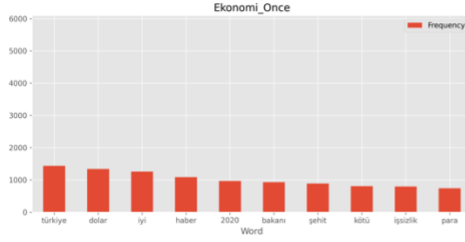
Tablo 3: Politika hakkındaki anahtar kelimeler

- Daha önce mülteci ve şehit vurguları öne çıkarken, sonrasında bilim ve sağlık vurgusu göze çarpmaktadır.



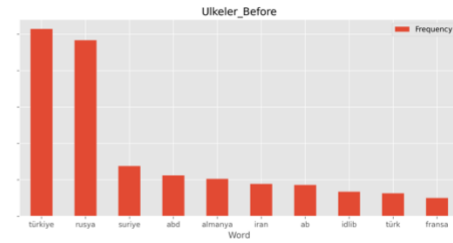
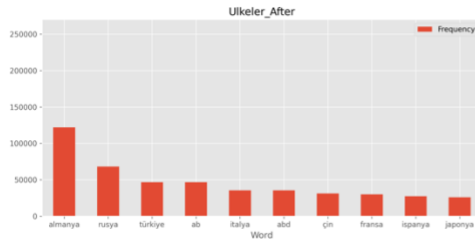
Tablo 4: Dezenformasyon hakkındaki anahtar kelimeler

- Komplo kelimesindeki bariz artış, Türkçe konuşan Twitter kamuoyunun olası dezenformasyonlara açık olduğunu göstermektedir.



Tablo 5: Ekonomi hakkındaki anahtar kelimeler

- Daha önceleri doların durumu ve işsizlik gibi temalar öne çıkarken, sonrasında ise sokağa çıkma yasağı ve devlet temaları öne çıkmaktadır. Devlet kelimesinin ön plana çıkması sorunun çözümünün devlet elinde görüldüğüne dair bir fikir verebilir.



Tablo 6: Ülkeler hakkındaki anahtar kelimeler

- İlk dönemde, Türkiye'nin son dönem dış politikasında önemli bir yer tutan Suriye, İdlib ve Rusya öne çıkmaktadır. Daha sonra ise, Almanya, AB, ABD ve İtalya gündemde öne çıkan kelimeler olarak göze çarpmaktadır. Batı'nın yoğunlukla ülke gündemine girmesi ayrıca çalışılması gerekli olan bir araştırma konusudur.

Gelecek Çalışmalar

Bu veri seti çalışması öncelikli olarak gelecek çalışmalara zemin hazırlamak için tasarlanmıştır. Vakit sınırlılığından dolayı sadece en çok geçen kelimeler karşılaştırmalı olarak sunulmuş ve bunlar üzerinden görülebilecek temel süreklilik ve kırılmalar not alınmıştır.

Bundan sonraki süreçte;

- Tweet içerikleri daha detaylı bir şekilde incelenerek Türkçe konuşan Twitter kullanıcılarının gündem değişiklikleri, kırılımları, ve devamlılıkları tespit edilebilir.
- ‘Topic Modelling’ gibi makine öğrenmesi metotları kullanılarak içerik hakkında farklı çalışmalar yapılabilir.

Detaylı Bilgi İçin

Bu veri seti hakkındaki teknik sorular ya da içeriğe dair sorularınız için aşağıdaki iletişim maillerine yazabilirsiniz.

* Burak Özturan burakoztrn@gmail.com

* Yunus Emre Tapan tapanyemre@gmail.com