

SOSYAL AĞ TABANLI VERİLERDEN FAYDALANARAK KORONA VİRÜS KONULU DUYGU ANALİZİ ÇALIŞMASI

Ebru KARAAHMETOĞLU^{1*}, Süleyman ERSÖZ², Osman KARAAHMETOĞLU³

¹ T. C. Ticaret Bakanlığı, ORCID No: <http://orcid.org/0000-0002-4482-9166>

² Kırıkkale Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü

ORCID No: <http://orcid.org/https://orcid.org/0000-0002-7534-6837>

³ Devlet Su İşleri Genel Müdürlüğü, ORCID No: <https://orcid.org/0000-0002-4482-9166>

Anahtar Kelimeler	Öz
Duygu Analizi Korona Ergonomi Sözlük Temelli Bert	<p>İletişim teknolojilerindeki gelişmeler sosyal medyayı toplum hayatının odak noktası haline getirmiştir. İnsanlar herhangi bir konu, ürün hakkındaki görüş düşünce ve yorumlarını, özgürce sosyal medyada paylaşmaktadırlar. Özellikle pandemi dönemindeki kişiler arası iletişimlere konan mesafe kuralı, sosyal medya üzerinde veri akışını yoğunlaştırmış, iş modellerini sosyal medya ortamına kaydırmış ve ergonomik değişiklikleri de gündeme getirmiştir. Günümüzün en önemli gündemi olan korona virüs salgınının toplum hayatı üzerindeki etkileri, kişilerin salgınla ilgili yoğunlaşan düşüncelerini oluşturmaktadır. Dolayısıyla pandemi sürecinin ergonomik koşullara ve sosyal hayata yansımaları üzerine bir analiz çalışması yapılmıştır. Yapılan çalışmada metin madenciliğinin duygu analizi teknikleri kullanılarak toplumun, korona virüs salgını konusundaki nabzı tutulmaya çalışılmıştır. Duygu analizine konu veriler kişilerin görüşleri olduğu için, sosyal medyada kişilerin düşünce ve içinde bulunduğu ruh hallerini özgürce paylaşabildiği tweet verileri çalışma veri kaynağı olarak seçilmiştir. Tweet verileri genellikle kısa uzunlukta olmakla birlikte, metinler dilbilimsel olarak hatalar içerebilmektedir. Buna ek olarak metinlerin ironiler ve emojiler içermesi metin analiz süreçlerini zorlaştırmaktadır. Metin analizlerinin doğru sonuçlar oluşturabilmesi için doğru veriye ihtiyaç duyması, veri analiz süreçlerinin analiz öncesi veriye uygulanmasını gerektirmektedir. Bu özellikleri ile tweet verileri doğal dil işleme teknikleri ile analiz çalışmalarında, önemli bir çalışma alanı olarak karşımıza çıkmaktadır. Bu nedenle çalışmaya konu olarak seçilen korona virüs ile ilgili kişilerin sosyal ağlarda paylaştığı tweet verileri, Phyton ile yazılan uygulama toplanmıştır. Bu veriler öncelikle, Türkçe'ye özgü çeşitli dilbilimsel tekniklerle temizlenmiş, daha sonra sözlük tabanlı, bert ve makine öğrenmesi duygu analizi teknikleri işlenmiştir. Elde edilen sonuçlar ve sonuçların duyarlılık analizleri karşılaştırmalı olarak sunulmuştur.</p>

EMOTION ANALYSIS STUDY ON CORONA VIRUS USING SOCIAL NETWORK BASED DATA

Keywords	Abstract
Sentiment Analysis Corona Ergonomic Dictionary Based Bert	<p>Developments in communication technologies have made social media the focal point of social life. People freely share their opinions and comments on any subject, product on social media. Especially, the distance rule for interpersonal communications during the pandemic period has intensified the flow of data on social media, shifted business models to the social media environment and bring up ergonomic conditions. The effects of the corona virus epidemic, which is the most important agenda of today, on public life constitute the people's intensified thoughts about the epidemic. Therefore, an analysis study was conducted on the reflection of the pandemic process on ergonomic conditions and social life. In this study, it was tried to keep the pulse of the society on the corona virus epidemic by using emotion analysis techniques of text mining. Since the data subject to sentiment analysis are the opinions of individuals, tweet data, in which people can freely share their thoughts and moods on social media, was selected as the study data source. While tweet data is usually short in length, texts can contain linguistic errors. In addition, the texts contain irony and emojis, making text analysis processes difficult. The fact that text analysis needs correct data in order to produce correct results requires the application of data analysis processes to the data before analysis. With these features, analysis of tweet data with natural language processing techniques emerges as an important field of study. For this reason, the tweet data shared by people related to the corona virus selected as the subject of the study on social networks were collected with an application written in Phyton. These data were first cleaned with various linguistic techniques specific to Turkish, then were processed by dictionary-based, bert and machine learning emotion analysis techniques. The results obtained and the sensitivity analysis of the results are presented comparatively.</p>
Araştırma Makalesi	Research Article
Başvuru Tarihi : 12.11.2020	Submission Date : 12.11.2020
Kabul Tarihi : 12.03.2021	Accepted Date : 12.03.2021

* Sorumlu yazar e-posta: eroglu@hotmail.com

1. Giriş

İnsanoğlu öğrenme sürecinde sadece kendi deneyimlerinden öğrenmez. Başkalarının deneyimlerini de harmanlayarak, kendi doğrularına üretir ve öğrenme sürecini daha etkin hale getirir. Bu noktada başkalarının herhangi bir konudaki görüş, düşünce ve deneyimleri, kişiler için önem arz etmektedir. Günümüzde, iletişim teknolojilerindeki gelişmeler, sosyal medyada toplumun görüş ve düşüncelerini paylaşacağı yeni ortamlar ortaya çıkarmıştır. Toplum olarak herhangi bir konu, satın alacağımız bir ürün, yaşanan bir olgu hakkında karar verme süreçlerinde, başkalarının düşüncelerini de hesaba katmak isteriz. Bu süreçte metinleri analiz ederek, duygu çıkarımında bulunan yöntemlerin kullanımı, daha geniş kitlelerin fikirlerine ulaşmamızı ve doğru sonuçlar elde etmemizi sağlayacaktır. Özellikle korona virüs gibi küresel bir salgınla ilgili duygu analiziyle toplumun nabızı tutulmak istenildiğinde, geniş kitlelerin düşüncelerine ihtiyaç duyulmaktadır.

Duygu analizinde, çalışmaya konu cümleler irdelenerek, cümlede ifade edilmek istenen duygu olumlu, olumsuz, tarafsız olarak etiketlenmeye çalışılır. Bu karar verme sürecinde kutuplaştırma, sözlük tabanlı yöntemler, öznitelik hesaplama gibi yöntemler kullanılır. Sözlük tabanlı yöntemlerde, cümlede geçen olumlu, olumsuz kelimelerin sayısına bakılarak cümle olumlu, olumsuz veya tarafsız olarak etiketlenir (Seker, 2016), (Liu, 2010) (Turkmenoglu, 2015).

(Wilson, vd., 2005), metinler üzerinde kutuplaştırma değeri hesaplayarak, duygu analizi çalışması gerçekleştirmiştir. (Agarwal vd., 2011), çalışmasında, twitter verileri üzerinde ağaç kernel ve özellik temelli modeller kullanarak duygu analizi yapmıştır. SemEval-2017 (Rosenthal vd., 2017) çalışmasında da tweet verileri üzerinde yapılan yedi yıllık analizin sonuçlarını paylaşmıştır. Sosyal medyada duygu bildiren tweet verileri normal metinlerden farklı olarak, kısa tek cümlelik metinler olduğu için, bu durumlara yönelik özel işlemler gerekebilmektedir (Kiritchenko vd., 2014).

İnternet teknolojilerinin yaygınlaşması, telefon dâhil kullandığımız küçük cihazların akıllanarak iletişim teknolojilerinde kullanılan cihazlar haline gelmesi, insanlar arasındaki mesafeleri azaltmış ve sosyal medya olgusunu tüm kitlelere yaymıştır. Sosyal medya üzerinde, satın alınan ürün yorumlarından, güncel gündem maddelerine kadar her konuda, görüş ve düşünceler paylaşılmaktadır. Toplumun paylaşmış olduğu bu düşünceler, alanlarına göre, araştırmacı, pazarlamacı vb. birçok uzmanın ilgisini çekmektedir. Bu yorumlar aracılığıyla, pazarlamacılar marka bilinirliği ve ürün kalitesi hakkında çıkarımlar yapabilirlerken, ihtiyaç analizleri yaparak yeni ürün stratejileri üretebilirler.

Politikacılar toplumun beğeni, zevk ve ihtiyaçlarını belirleyerek, plan, program ve demeçlerini bu analizlere göre belirleyebilirler. Araştırmacılar önemli gündem konuları ile ilgili analizler yaparak, toplumun moral seviyesini ölçebilirler. Bu noktada insanların görüş, düşünce ve söylemlerini, dilbilimsel tekniklerle analiz ederek, olumlu, olumsuz ve tarafsız olarak etiketleyen araştırma dalı, duygu analizi olarak adlandırılır (Onan vd., 2016).

Son yıllarda popüler bir çalışma alanı olan duygu analizi konusunda, şarkı sözlerinin duygu analizi (Özdeş, 2017), tweet verilerinin etiketleyecek yapıların oluşturulması (Powar vd., 2018), sözlük tabanlı duygu analizi ve makine öğrenmesi yöntemleriyle Türkçe metinlerin incelenmesi (Turkmenoglu, 2015). konulu çalışmalar yapılmıştır. Bu çalışmalara ek olarak, fikir madenciliği ve duygu analizi süreçlerini ayrıntılı olarak inceleyerek, literatürdeki çalışmaları karşılaştırmalı olarak analiz eden çalışmalarda bulunmaktadır (Özyurt vd., 2018).

Türkçe dilinin sondan eklemeli bir dil olması, olumlu olarak nitelenen bir kelimenin metinde birlikte geçtiği bir kelime ile olumsuz anlama dönüşmesi, duygu analizini zor bir süreç haline getirmektedir. Boğaziçi Üniversitesi tarafından Türkçe dilinde duygu analizi konusunda bir çatı oluşturacak çalışma yürütülmektedir (Boğaziçi Üniversitesi).

Duygu analizi konusunda yapılan çalışmalarda, baskınlık (Samuel vd., 2014) ve standartlaştırılmış olumlu ve olumsuz duygu puanlama yöntemleri (He vd., 2015) (Ravi ve Ravi 2015) kullanılmıştır.

Sariman ve Mutaf Covid 19 duygu analizi konulu çalışmasında, pandemiye yönelik önlem ve kısıtlamaları ölçeklendirerek twitter verileri ile duygu analizi yapmıştır (Sariman ve Mutaf, 2020).

Samuel vd. çalışmalarında pandemi süresince korku duygusunun ilerleyişini metin analitiği ve makine öğrenmesi yöntemleriyle analiz etmişlerdir (Samuel vd., 2020).

Chakraborty vd. çalışmalarında Gaussian bulanık kurallarını temel alan tekniklerle tweet verilerinde duygu analizi yapmışlardır. Çalışma salgına yönelik olarak başlangıçta olumlu olan düşüncelerin, bir şok halinde olumsuzla dönmelerini belirlemiş olması açısından değerlidir (Chakraborty, vd., 2020).

Bu çalışmada güncel twitter verileri toplanarak, korona virüs gibi özel bir konuyla ilgili duygu analizi yapılarak, toplumun salgınla ilgili nabızı tutulmaya çalışılmıştır.

2. Sosyal Medya ve Duygu Analizi

Sosyal medya üzerinde sosyal ağ analizleri, farklı konularla ilgili toplumun görüşleri, iletişimi ve

etkileşimlerini konu almaktadır. Günümüzde sosyal medya toplumsal hayatın odak noktasını oluşturmakta olup, toplumsal yaşamın büyük bir zamanı sosyal medyada geçmektedir. İnsanların çoğu düşüncelerini, twitter ve facebook gibi sosyal medya ortamlarında özgürce paylaşabilmektedir. Buna ek olarak, sosyal medya ortamları üzerinden iş modelleri oluşturulmaktadır. Bireyler olarak kişiler, özel sektör kuruluşları ve hatta kamu kurum ve kuruluşları, sosyal medyada kendi marka bilinirliklerini oluşturmak için yer almaktadır. Tam bu noktada, toplumun bir konu hakkındaki nabzını ölçecek çalışmalarda, sosyal medya verileri önem kazanmaktadır.

İstatistikçilere göre, twitter, toplumun birçok dalından kişilerin özgürce görüş ve düşüncelerini paylaştığı, milyonlarca üyesi olan bir sosyal medya ortamı olarak önümüze çıkmaktadır. Duygu analizinde kişilerin bir konu hakkındaki görüş ve düşüncelerinin analiz edilmesi, sosyal medyada düşünce paylaşma ortamlarını çalışma açısından zengin bir veri kaynağı olarak öne çıkarmaktadır.

Sosyal medya kullanıcılarının her geçen gün artması, sosyal medyada yoğun bir veri akışının olmasını sağlamaktadır. Özellikle, pandemi nedeniyle evlere kapanılan dönemlerde, sosyal medya üzerinden iletişim daha büyük bir yoğunluk kazanmaktadır. Sosyal medyada duygu, düşünce ve yorum bildiren sözlerin kısa olması, dilbilimsel olarak metinlerin ironiler, anlam kaymaları, kısaltmalar, emojiler içermesi, metin analizi süreçlerinde önemli zorluklar içermekle birlikte, doğal dil işleme teknikleri için bulunmaz bir araştırma sahası oluşturmaktadır.

Literatürdeki pandemiye yönelik çalışmalarda, twitter verileri yaygın olarak kriz yönetimi ve takibinde kullanılmıştır (Ye vd., 2016), (Fung vd., 2014), (Kim vd., 2016), (Kim vd., 2020). Nagar vd. acil servisler için grip ve benzeri hastalıkların analizinde twitter verilerinin gücünü analiz etmiştir (Nagar vd., 2014). Twitter verileri ile salgınları tahmin eden (Szomszor vd., 2010) ve kullanılan ilaçlar ve yan etkilerinin neler olabileceğini gösteren (Bian vd., 2012) modeller oluşturulmuştur. Widener ve Li, sağlıklı ve sağlıklı olmayan yiyeceklerle ilgili coğrafi olarak konumlanan tweet verilerinin, coğrafi olarak nasıl dağıldığını analiz eden duygu analizi çalışması gerçekleştirmiştir (Widener ve Li, 2014). De Choudhury vd., yeni doğum yapan annelerin doğum sonrası, davranış değişiklikleri ve durumlarını analiz etmiştir (De Choudhury vd., 2013). Twitter verileri ile yapılan diğer çalışmalarda, sosyal medya ile insanların algısının nasıl değiştiği analiz edilmiştir (Nguyen vd., 2012) ve turistik bir bölgeye gelen turistlerin tweetlerinden algı analizi yapılmıştır (Claster vd., 2010). Bu çalışmalara ek olarak, Akgül vd., Türkçe tweetler üzerinde, sözlük yöntemi ve ngram algoritmalarını kullanarak duygu analizi yapmıştır (Akgül vd., 2010).

Bu bilgilerin ışığında pandemiye yönelik duygu analizi çalışmamızda, tweet verilerini kullanarak toplumun moral seviyesini belirlemeye çalıştık.

3. Yöntem

Çalışmada, Phyton dilinde geliştirilen uygulama ile duygu analizi yapılmıştır. Şekil 1. de blok diyagram gösterimi verilen duygu analizi algoritmasının, analiz adımları aşağıda anlatılmıştır:

Veri Toplama: İçeriğinde "korona" kelimesi geçen korona virüsü ilgili twitter verileri geliştirilen uygulama aracılığıyla toplanmıştır.

Veri Temizleme: Metindeki tweet gönderen, Türkçe dili açısından konuya yönelik anlam içermeyen kelimeler, noktalama işaretleri vb. alfa karakterler, çeşitli yöntem ve algoritmalarla temizlenmiştir.

Analiz: Duygu analizi ile metnin kutuplaştırma ve öznellik değerleri hesaplanarak, düşünceler olumlu, olumsuz ve tarafsız olarak etiketlenmiştir. Bu algoritma ile tatmin edici sonuçlara ulaşılamamıştır.

Bu noktada sözlük tabanlı yöntem ile duygu analizi çalıştırılmıştır. Bu yöntemde özel uygulanan analiz süreçleri şöyledir:

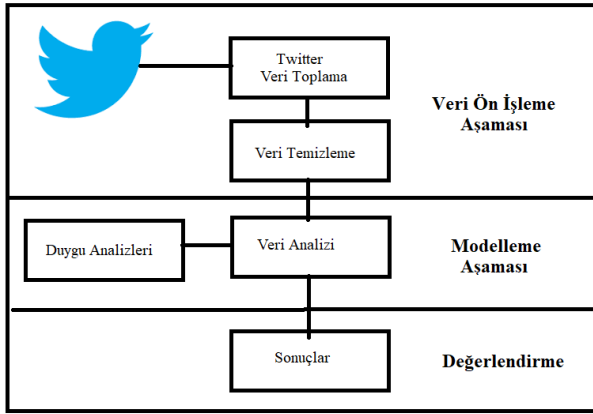
Bölütlere Ayırma: Bölütlere ayırma işleminde, analiz edeceğimiz metni oluşturan kelimeler bulunur.

Frekans Analizi: Frekans analizi aşamasında, bölüm dizisinde hangi kelimenin ne kadar geçtiği bilgisi bulunmaya çalışılır. Bölüm dizisi içerisinde daha fazla geçen kelimelerin, metnin konusu ile ilgili bilgileri içermesi beklenir.

Sözlük Tabanlı Duygu Analizi: Frekans analizi sonucunda elde edilen kelime dizisinde sayıca fazla geçen kelimeler, duygu sözlüğünde araştırılır. Kelimelerin duygu etiketlerine göre, hangi duygu etiketinde kelime sayısı fazla ise tweet metni o etiketle sınıflandırılır.

Sözlük tabanlı duygu analizinde, kelimelerin olumlu, olumsuz puanları içeren bir sözlüğe ihtiyaç duyulmaktadır. Çalışmamızda SentiTurkNet duygu kelime sözlüğü kullanılmıştır (Dehkharghani vd., 2015). Duygu kelime sözlüğündeki olumlu, olumsuz ve tarafsız kelime değerleri, cümlede geçen kelimeler için ayrı ayrı toplanarak hesaplanır. İşlem sonucunda elde edilen baskın değere göre cümle etiketlenir (Saygın).

$$\begin{aligned} p^+ &> p^- && \text{olumlu} \\ p^+ &< p^- && \text{olumsuz} \\ p^+ &= p^- && \text{tarafsız} \end{aligned}$$



Şekil 1. Duygu Analizi Algoritması

4. Duygu Analizi Çalışması ve Deneysel Sonuçlar

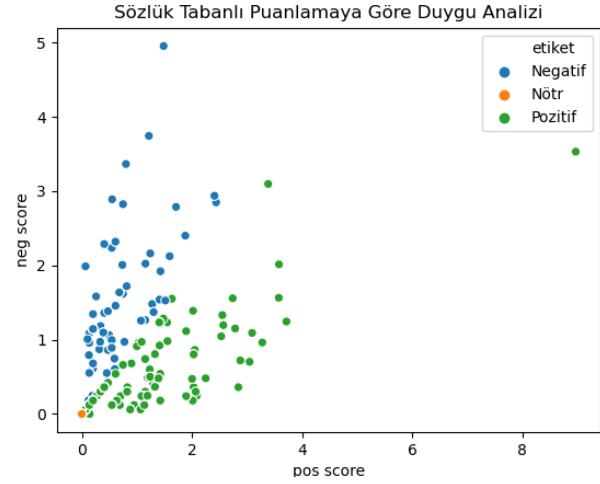
Çalışmada, Python ile yazılan uygulama ile toplanan twitter mesajları üzerinde, duygu analizi ile korona virüs salgınının neden olduğu yaşam koşullarında, toplumun moral seviyesi ölçülmeye çalışılmıştır.

Çalışmada öncelikle, sözlük tabanlı duygu analizi yöntemi kullanılmıştır. Sözlük tabanlı duygu analizinde SentiTurkNet (Dehkharghani vd., 2015) Türkçe kelime sözlüğü kullanılmıştır. SentiTurkNet kelimelerin olumlu, olumsuz ve tarafsız kelime puanlarını içeren bir sözlüktür.

Twitter üzerinden indirilen mesajlar içerisinde geçen kelimeler, bölütlere ayırma süreci ile bir kelime dizisine dönüştürülür. Kelime dizisi üzerinde işlenen veri temizleme süreçleri ile kelimelerin küçük harflere çevrilmesi, Türkçe açısından anlam içermeyen kelimeler ve noktalamaya işaretlerinin elenmesi işlemleri gerçekleştirilir.

Elde edilen kelime dizileri üzerinde sözlük tabanlı duygu analizi çalıştırılarak, tweet mesajları olumlu, olumsuz ve tarafsız olarak etiketlenir. Bu işlemde Türkçe’de geçen kelimelerle ilgili olumlu, olumsuz, tarafsız ağırlık değerlerini içeren bir duygu analizi sözlüğünden faydalanılmıştır. Sözlük temelli duygu analizinde, mesajın kelime dizisinde yer alan tüm sözcüklerin, ağırlık değerleri sözlükten bulunarak, olumlu ve olumsuz olmak üzere iki farklı değişkende, iki farklı duygu toplam değeri hesaplanır. Bu toplam değerlerinden, olumlu değer olumsuz değerden büyükse, mesaj olumlu, küçükse olumsuz, iki değer eşitse tarafsız olarak etiketlenir. Bu etiketleme işlemi ile elde edilen değerler, denetimli öğrenme aşamasında, öğrenme algoritmalarının eğitim ve test verilerinde, çıkış değeri olarak kullanılır. Sözlük temelli öğrenme algoritması ile elde edilen duygu analizi sonuçları, Şekil 2. deki dağılım grafiğinde gösterilmektedir. Toplam 9149 mesajın, 7073 adedi olumlu çıkarken, 1085 adedi olumsuz, 991 adedi tarafsız çıkmıştır. Grafikten de görüldüğü gibi, olumlu duygu sonuçları, olumsuz duygu sonuçlarının

7 katı kadar çıkmıştır. Korona virüs salgınının artış gösterdiği bir dönemde olumlu duygunun yüksek olması, salgın farkındalığının oluşmaması, durumu kanıksama, ileriye yönelik umut verici durumların farkındalığı ve yeni normale alışma olarak yorumlanabilir.



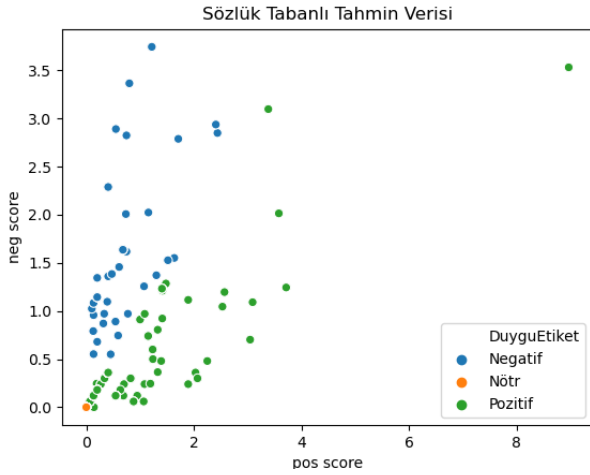
Şekil 2. Sözlük Tabanlı Puanlamaya Göre Duygu Analizi

Tablo 1. Sözlük Tabanlı Makine Öğrenmesi Modeli Değerlendirme Sonuçları

	duyarlılık	hatırlama	f1puan	destek
0	1.00	1.00	1.00	545
1	1.00	1.00	1.00	496
2	1.00	1.00	1.00	3534
doğruluk				4575
makro ortalama	1.00	1.00	1.00	4575
makro ortalama	1.00	1.00	1.00	4575

Analize konu veriler, öğrenme ve test verileri olarak ikiye bölünerek denetimli öğrenme algoritması ile tahminler üretilmiştir. Tahmin sonuçları olarak elde edilen duygu durumları Şekil 3. deki dağılım grafiğinde gösterilmiştir. Toplam 4575 mesajın, 3534 adedi olumlu çıkarken, 529 adedi olumsuz, 496 adedi tarafsız çıkmıştır. Olumlu ve olumsuz duygulu mesajlar arasındaki 7 katlık fark burada da karşımıza çıkmıştır. Bu da modelin tutarlılığını göstermektedir.

Genellikle sözlük tabanlı öğrenme modellerinde yüksek başarımlar sağlanamamasına rağmen, çalışmamızdaki modelde yüksek başarımla karşılaşmamız, test verilerinde y, duygu değerinin sözlük temelli olarak belirlenmesi olarak açıklanabilir. Modelin duyarlılık analizi sonuçları Tablo 1. den de görüldüğü gibi çok yüksek, tam 1 olarak hesaplanmıştır.

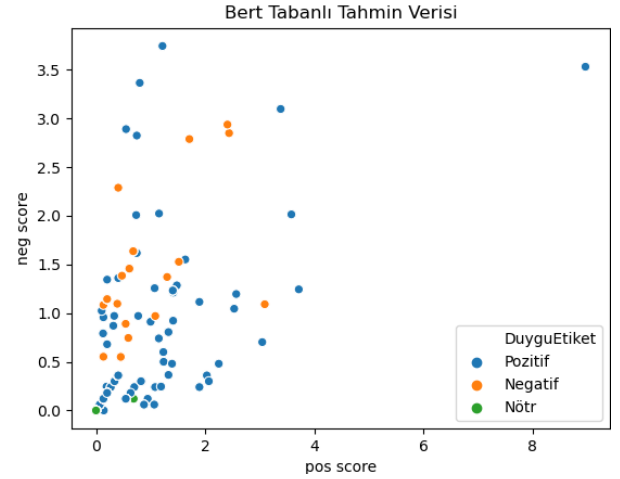


Şekil 3. Sözlük Tabanlı Tahmin Verisi

Sözlük temelli duygu analizi çoğu zaman doğru sonuçları üretmeyebilmektedir. Sözlük temelli duygu analizinin, kelimelerin olumlu, olumsuz puanlarına odaklanarak karar verme süreçlerini işletmesi, cümlede geçebilen ironileri yakalayamamasını neden olmaktadır. Bu nedenle daha tutarlı ve doğru sonuçlar üretebilen Bert bölütlere ayırma algoritması (Devlin vd., 2019) (Koksal, 2020), elde edilen eğitim ve test verilerine uygulanmıştır. Bert algoritmasının önemli özelliği, Türkçe’de kelimelerin bir cümlede hangi kelimelerle birlikte kullanıldığını, gösteren bir vektör dizisinden faydalanmasıdır. Çalışmada, Türkçe diline özel önceden eğitilmiş trmodel (Koksal, 2018) vektör dizisi kullanılmıştır.

Twitter mesajlarına Bert bölütlere ayırma algoritması uygulanarak, mesajlar olumlu, olumsuz olarak nitelendirilerek, Şekil 4. deki dağılım grafiğinde gösterilmiştir.

Grafikteki veriler incelendiğinde, korona virüs ile ilgili duygusal durum düzeyinde olumlu mesajların, olumsuz mesajlardan fazla olduğu görülmektedir. Toplam 4575 test mesajının, 3552 adedi olumlu çıkarken, 115 adedi olumsuz, 494 adedi tarafsız çıkmıştır. Mesajların, yaklaşık olarak %98’inin olumlu duygu içerdiği sonucuna ulaşılmıştır. Sağlık Bakanlığı verilerine göre, vakalarda yükseliş eğiliminin olduğu bir dönemde böyle bir sonuç, toplumun henüz salgının ciddiyetini kavrayamamış olması veya salgın olgusunu kanıksayarak, yeni normalde yaşama alışmaya başladığı olarak yorumlanabilir.



Şekil 4. Bert Tabanlı Tahmin Verisi

Öğrenme modelinin ürettiği sonuçlarla ilgili, doğruluk analizleri sonuçları Tablo 2. de görülmektedir. Modelin ürettiği değerlerin doğruluğunun 0.99 gibi, çok yüksek bir değer olduğu görülmektedir.

Tablo 1. Bert Tabanlı Makine Öğrenmesi Modeli Değerlendirme Sonuçları

	duyarlılık	hatırlama	f1puan	destek
0	0.99	0.96	0.97	545
1	1.00	0.99	1.00	496
2	0.99	1.00	1.00	3534
doğruluk			0.99	4575
makro ortalama	0.99	0.98	0.99	4575
makro ortalama	0.99	0.99	0.99	4575

Literatürdeki benzer çalışmalarda elde edilen sonuçların doğruluk oranları aşağıda verilmiştir:

Samuel vd. Covid 19 duygu analizi konulu çalışmalarında, uzun tweetlerde %91,43 gibi yüksek bir doğrulukta sınıflandırma başarımı sağlamıştır. Uzun tweetlerde bu oran %57’ye düşmüştür (Samuel vd., 2020).

Sariman ve Mutaf’ın çalışmasında çeşitli ölçeklerde %91 ile %97 doğruluk aralığında sonuçlar elde edilmiştir. Kısıtlamalar dâhilinde başlangıçta olumsuz karşılanan uygulamaların ilerleyen zamanlarda olumlu karşılanmaya başladığı ifade edilmektedir (Sariman ve Mutaf, 2020).

Chakraborty vd. çalışmalarında tweet verilerini yayınladıkları tarihlere göre iki farklı veri kümesinde toplamışlardır. Birinci veri kümesinde olumlu duygu sonuçlarına ulaşırken, ikinci veri kümesinde olumsuz duygu sonuçlarına ulaşmışlardır. Analiz sonuçlarının doğruluk oranları,

veri kümelerine göre en yüksek %81 ve %75 olarak elde edilmiştir (Chakraborty vd.,2020).

Çalışmamızda %99 doğruluk oranı ile daha iyi bir sonuç elde edilmiştir.

5. Sonuçlar ve Öneriler

Çalışmada, korona virüs ile ilgili "korona" arama metniyle aranan twitter verileri toplanarak, mesajlar üzerinde duygu analizleri çalıştırılmıştır. Sözlük tabanlı, bert algoritması ve makine öğrenmesi yöntemleri ile duygu analizleri çalıştırılmış ve sonuçlar sunulmuştur.

Hem sözlük tabanlı hem bert temelli algoritmalarla geliştirilen duygu analizlerinde, olumlu mesajların sayısı, olumsuz mesajların 7 katı düzeyinde olmuştur. Bu sonuçların ışığında korona virüs konusunda toplumun moral düzeyinin yüksek olduğu sonucuna varılabilir. Son günlerde korona virüs salgınında bir yükseliş eğiliminin olduğu, açıklanan verilerden aşikârdır. Bu durumda moral düzeyinin yüksek olması, salgın farkındalığının oluşmaması, yeni normale uyum sağlama, ileriye yönelik umut verici gelişmelerin olması veya durumu kanıksama olarak açıklanabilir.

Analizden elde edilen sonuçların güvenilirliği çeşitli değerlendirme yöntemleri ile irdelenmiş olup, sözlük tabanlı algoritma için 1, bert için 0,98 olmak üzere, çok yüksek güvenilirlik sonuçlarına ulaşılmıştır. Çalışma verilerinde öğrenme ve test verisi olarak kullanılan tweet verilerinde duygu değerleri (y) sözlük tabanlı duygu analizi algoritması ile belirlenmiştir. Bu nedenle, sözlük tabanlı algoritmanın güvenilirlik değerleri yüksek çıkmıştır. Bu duygu değerlerinin başka bir şekilde hesaplanması daha güvenilir sonuçlar ortaya çıkarabilir.

Yapılan çalışmada güncel tweet verileri incelenmiştir. Analize konu tweet verilerinin artırılması, örnek uzayını veri çeşitliliği ve kapsanan süre açısından zenginleştirecektir. Çeşitlilik açısından yüksek veri kümesi ile daha iyi sonuçlar elde edilebilecektir.

Çıkar Çatışması

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir.

Kaynaklar

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. (2011). Sentiment Analysis of Twitter Data. In Proceedings of the workshop on

Languages in Social Media, pp. 30-38. Association for Computational Linguistics.

Akgül E.S., Ertano C., Diri B. (2016). Twitter Verileri İle Duygu Analizi, Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, pp. 106-110, Volume 2, Issue 22.

Boğaziçi Üniversitesi, Türkçe için Kapsamlı Bir Duygu Analizi Çatısı Oluşturulması, <https://arastirma.boun.edu.tr/tr/proje/turkce-icin-kapsamli-bir-duygu-analizi-catisi-olusturulmasi>.

Bian J, Topaloglu U, Yu F. (2012). Towards Large-Scale Twitter Mining For Drug-Related Adverse Events. International Workshop on Smart Health and Wellbeing (SHB'12), Maui, Hawaii, USA.

Chakraborty K., Bhatia S., Bhattacharyya S., Platos J., and Bag R. (2020). Sentiment Analysis of COVID-19 Tweets by Deep Learning Classifiers - A Study to Show How Popularity is Affecting Accuracy in Social Media, Appl. Soft Comput. J., vol. 97, p. 106754, doi: 10.1016/j.asoc.2020.106754.

Cluster WB, Dinh H, Cooper M. (2010). Naive Bayes and Unsupervised Artificial Neural Nets For Cancun Tourism Social Media Data Analysis. 2nd World Congress on Nature and Biologically Inspired Computing (NaBIC). Kitakyushu, Fukuoka, Japan, 15-17.

Dehkharghani R., Saygin Y., Yanikoglu B., Oflazer K. (2015). SentiTurkNet: A Turkish Polarity Lexicon For Sentiment Analysis. Language Resources and Evaluation; 50 (3): 667-685. doi: 10.1007/s10579-015-9307-6

Devlin, J., Chang M.W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805v2.

De Choudhury, M., Counts, S. Horvitz, E. (2013). Predicting Postpartum Changes in Emotion and Behavior via Social Media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; pp. 3267–3276.

Fung, I. C.H. Yin, J. Pressley, K. D. Duke, C.H. Mo, C. Liang, H. Fu, K. W. Tse, Z. T. H. Hou, S. I. (2014).

- Pedagogical Demonstration of Twitter Data Analysis: A Case Study of World AIDS Day.
- He, W. Wu, H. Yan, G. Akula, V. Shen, J. (2015). A Novel Social Media Competitive Analytics Framework With Sentiment Benchmarks. *Inf. Manag.*, 52, 801–812.
- Kim, E. H. J. Jeong, Y. K. Kim, Y. Kang, K. Y. Song, M. (2016). Topic-based Content and Sentiment Analysis of Ebola Virus on Twitter and in The News. *J. Inf. Sci.*, 42, 763–781.
- Kiritchenko, S., Zhu, X., and Mohammad, S. M. (2014). Sentiment Analysis of Short Informal Texts. *Journal of Artificial Intelligence Research*, 50, 723-762.
- Koksal, A. (2018). Eğitilmiş Türkçe Word2Vec Modeli, <https://libraries.io/github/akoksal/Turkish-Word2Vec##TO-DO##>.
- Koksal, A. (2020). Bert Türkçe Duygu Analizi, <https://github.com/akoksal/BERT-Sentiment-Analysis-Turkish>.
- Liu, B. (2010). Sentiment Analysis and Subjectivity, *Handbook of Natural Language Processing*, 2, 627–666.
- Nguyen LE, Wu P, Chan W, Peng W, Zhang Y. (2012). Predicting Collective Sentiment Dynamics From Time-Series Social Media. *Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM '12)*, Beijing, China, 12.
- Onan, A. and Korukoğlu, S. (2016). A Review of Literature on The Use of Machine Learning Methods For Opinion Mining, *Pamukkale University Journal of Engineering Sciences*, 22(2), 111–122.
- Özdeş, M. (2017). Büyük Veri Araçlarını Kullanarak Duygu Analizi Gerçekleştirimi. (Yüksek Lisans Tezi). Pamukkale Üniversitesi Fen Bilimleri Enstitüsü.
- Özyurt, B. ve Akçayol. M. A. (2018). Fikir Madenciliği Ve Duygu Analizi, Yaklaşımlar, Yöntemler Üzerine Bir Araştırma. *Selçuk Üniversitesi Mühendislik Bilim ve Teknoloji Dergisi*.6-4. 668-693.
- Powar, S., Kadam, U. and Salvi, T. (2018). Twitter Sentiment Analysis for Classifying Hate Tweets and Normal Tweets Using Logistic Regression and Naive Bayes Algorithm. *Journal of Emerging Technologies and Innovative Research*.5-10. 371-373.
- Ravi, K.; Ravi, V. (2015). A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications. *Knowl. Based Syst.*, 89, 14–46.
- Rosenthal, S., Farra, N., and Nakov, P. (2017). SemEval-2017 Task 4: Sentiment Analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 502-518.
- Samuel, J. Ali, G. G. M. N. Rahman, M. M. Esawi, E. Samuel, Y. (2020). COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Information*, 11, 314. <https://doi.org/10.3390/info11060314>.
- Samuel, J. Holowczak, R. Benbunan-Fich, R. Levine, I. (2014). Automating Discovery of Dominance in Synchronous Computer-Mediated Communication. In *Proceedings of the 2014 47th IEEE Hawaii International Conference on System Sciences*, Waikoloa, HI, USA, 6–9; pp. 1804–1812.
- Sariman, G. ve Mutaf, E. (2020). Covid-19 Sürecinde Twitter Mesajlarının Duygu Analizi Sentiment Analysis Of Twitter Messages In Covis-19 Process. *Euroasia Journal of Mathematics Engineering Natural and Medical Sciences*. 7. 137-148.
- Saygın Y., *Metin Madenciliği: Temel Yöntemler ve Duygu Analizi* Sabancı Üniversitesi, http://akademikkamp.org/3_BAK/images/sunular/SIRINCE-YUCEL-SAYGIN.pdf.
- Seker, S. E (2016). Duygu Analizi (Sentimental Analysis), YBS Ansiklopedi www.YBSAnsiklopedi.com Cilt 3 Sayı 3, Eylül 2016, http://ybsansiklopedi.com/wp-content/uploads/2016/09/duygu_analizi.pdf.
- Szomszor M. N., Kostkova P., de Quincey, E. (2010). #Swineflu: Twitter predicts swine flu outbreak in 2009. *3rd International ICST Conference on Electronic Healthcare for the 21st Century (eHEALTH2010)*, Casablanca, Morocco, 13-15.

Türkmenoğlu, C. (2015). Türkçe Metinlerde Duygu Analizi. (Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

Ye, X. Li, S. Yang, X. Qin, C. (2016). Use of Social Media For The Detection and Analysis of Affective Diseases in China. ISPRS Int. J. Geo-Inf., 5, 156.

Widener, M.J. Li, W. (2014). Using Geolocated Twitter Data To Monitor The Prevalence of Healthy and Unhealthy Food References Across The US. Appl. Geogr., 54, 189-197.

Wilson, T., Wiebe, J. Hoffmann, Pmesa. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In Proceedings of The Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347-354. Association for Computational Linguistics.