



Mustafa Kemal Üniversitesi Eğitim Fakültesi Dergisi
Mustafa Kemal University Journal of the Faculty of Education
Yıl/Year: 2024 ♦ Cilt/Volume: 8 ♦ Sayı/Issue: 13, s. 16-29

COVID-19 SALGINI SÜRECİNDE UZAKTAN EĞİTİME YÖNELİK PAYLAŞILAN TÜRKÇE TWITTER MESAJLARININ DUYGU ANALİZİ

Öğr. Gör. Ali KÜSMÜŞ

Osmaniye Korkut Ata Üniversitesi, Bahçe Meslek Yüksekokulu, alikusmus@osmaniye.edu.tr
Orcid: 0000-0001-7067-4164

Özet

Günümüzde sosyal medya kullanımı gün geçtikçe artmaktadır. İnsanlar olaylara karşı tepkilerini sıklıkla bu mecralarda vermektedirler. Twitter da bu mecraların başında gelmektedir. Bilindiği gibi hem ülkemizde hem dünyada Covid-19 salgını sürecinde büyük ölçüde yüz yüze eğitime ara verilip uzaktan eğitim sistemleri kullanılmıştır. Bu sistemleri kullanan öğrencilerin Twitter' daki tepkileri araştırmanın kaynağını oluşturmaktadır. Bu çalışmada tüm dünyayı etkisi altına alan Covid-19 salgını sürecinde eğitim alanında uygulanan uzaktan eğitim sistemine insanların verdikleri tepkinin durumu tespit etmek için model geliştirilmiştir. Geliştirilen modelde kullanılan veri seti 01.09.2020-30.06.2021 tarihleri arasındaki uzaktan eğitimle ilgili yapılan paylaşımlardan elde edilmiştir. Veri seti geliştirilen ara yüz sayesinde duygu durumları manuel olarak elle etiketlenmiştir. Veri seti ön işleme aşamalarından geçirilerek duygu analizine uygun hale getirilmiştir. TF-IDF vektörizasyonunda geçirilen veri seti 5 farklı sınıflandırma algoritması ile sınıflandırılarak doğruluğu ölçülmüştür. Bu sınıflandırma algoritmalarından en başarılısı %76,6 ile Çok Katmanlı Algılayıcılar (MLP) algoritması olmuştur. Geliştirilen bu model devasa boyutlarda verinin üretildiği sosyal medya mecralarındaki verilerden anlam çıkarılmasına kolaylıklar sağlayacaktır.

Anahtar Kelimeler: Duygu Analizi, Covid-19, Uzaktan Eğitim, Çok Katmanlı Algılayıcılar Algoritması (MLP)

SENTIMENT ANALYSIS OF SHARED TURKISH TWITTER MESSAGES REGARDING DISTANCE EDUCATION DURING THE COVID-19 PANDEMIC

Abstract

Today, the use of social media is increasing day by day. People often give their reactions to events in these media. Tweeter is one of these channels. As it is known, during the Covid-19 outbreak both in our country and in the world, face-to-face education was largely interrupted in our country and in the world and distance education systems were used. The reactions of students using these systems on Twitter constitute the source of the research. In this study, a model has been developed to determine the emotional state of people's reaction to the distance education system applied in the field of education during the Covid-19 epidemic that affected the whole world. The data set used in the developed model was obtained from the shares made on distance education between 01.09.2020-30.06.2021. Thanks to the interface developed in the data set, the emotional states were manually labeled. The data set was made suitable for sentiment analysis by passing through the preprocessing stages. The data set passed in TF-IDF vectorization was classified with 5 different classification algorithms and its accuracy was measured. The most successful of these classification algorithms was the Multi-Layer Perceptron (MLP) algorithm with 76.6%. This developed model will make it easier to make sense of the data in social media channels, where huge amounts of data are produced.

Key Words: Sentiment Analysis, Covid-19, Distance Learning, Multilayer Perceptrons Algorithm (MLP)

Giriş

Geçmişten bugüne farklı kanallar aracılığı ile yapılabilen uzaktan eğitim artık günümüzde web yoluyla gelişmiş şekillerde yürütülebilmektedir. Yüz yüze eğitim olmadan sadece uzaktan eğitim kullanılarak etkili öğretim gerçekleştirilebilir olsa da yüz yüze eğitimin vazgeçilmez olduğu yönünde çeşitli görüşler vardır. Buna karşın deprem, salgın gibi doğal afetler nedeniyle yüz yüze eğitimin mümkün olmadığı dönemlerde web yoluyla uzaktan eğitim kaçınılmaz hale gelmiştir ve bu dönemde uzaktan eğitim günümüz koşulları için çok iyi bir alternatiftir.

Web tabanlı uzaktan eğitimde öğrenciler öğretmen tarafından paylaşılan materyalleri kullanarak asenkron eğitimi gerçekleştirebilirler. Bu uzaktan eğitim türünde öğrenci zaman ve mekan yönüyle bağımsızdır. Bununla birlikte canlı online dersler aracılığı ile de senkron eğitim yürütülebilir ki bu eğitimde de öğrenci ve öğretmen mekan bakımından bağımsız zaman bakımından bağımlıdır.

Uzaktan eğitim, öğretimi bireyselleştirerek öğretimin kontrolünü öğrenciye verir. Bu durum zaman yönetimi ve öz denetimi gelişmeyen veya yeterli olmayan öğrenciler için ciddi bir dezavantaj oluşturmaktadır. Bununla birlikte uzaktan eğitimde öğrencilerin derse katılımlarının kontrolü ve ölçme-değerlendirme süreçlerinin objektifliği de bu yöntem için önemli dezavantajlar arasındadır.

2023 yılı Haziran ayı verilerine göre ülkemizde bireyler günde ortalama 2 saat 54 dakika sosyal mecralarda zaman geçirirken bu süre dünya genelinde 2 saat 31 dakikadır. Twitter ise ülkemizde google ve youtube'dan sonra en fazla ziyaret edilen 3. adrestir (wearesocial.com, 2024). Bu veriye göre kullanılan sosyal medya sitelerinden en popüler olanının Twitter olduğu söylenebilir. Her bir mecranın kendine özgü özellikleri, sınırlamaları, kuralları, yapısı, çalışma şekli olabilir. Günümüzde X adıyla marka adını değiştiren Twitter'ın genellikle metin tabanlı olarak kısa mesaj paylaşmaya odaklandığı görülse de bu platformda ses, video gibi farklı unsurlar da yayınlanabilir. En fazla 280 karakterin paylaşılabilirdiği mesajlar tweet olarak anılmaktadır (Wikipedia, 2024).

Toplumun belli kesiminin belli konularda görüşlerini elde etmede sosyal mecralar kullanılabilir. Sosyal mecralarda kullanıcılar tarafından hızlı ve kolayca görüşlerin paylaşılabilmesi bu ortamları duyguların ifade edildiği büyük veri kütlelerine dönüştürmüştür. Bu derece fazla verilerin anlamlı bir bütün haline dönüştürmek için duygu analizlerine ihtiyaç duyulmaktadır (Pathak, Pandey ve Rautaray, 2021).

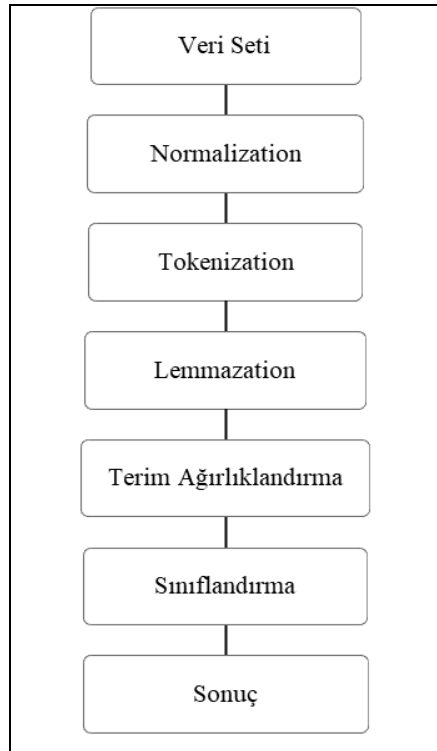
Başta facebook, instagram ve twitter olmak üzere sosyal medya, büyük boyutlarda verileri temsil eden "big data" ların temelinin oluşturmaktadır (Ceron, Curuni, Iacus, 2016). Big data lar Sosyal medya kullanıcılar tarafından metin, ses, video, resim gibi farklı unsurlar barındıran ham veriler içerir. Bu ham veriler duygu analizleri ile işlenerek daha anlamlı hale getirilebilir, genellemeler yapılabilir. Bundan dolayı toplumun belli kesiminin belli konularda görüşlerini elde etmede sosyal mecralar kullanılabilir. Sosyal mecralarda kullanıcılar tarafından hızlı ve kolayca görüşlerin paylaşılabilmesi bu ortamları duyguların ifade edildiği büyük veri kütlelerine dönüştürmüştür. Bu derece fazla verilerin anlamlı bir bütün haline dönüştürmek için duygu analizlerine ihtiyaç duyulmaktadır (Pathak, Pandey ve Rautaray, 2021). Sosyal mecralar bireylerin duygu ve düşüncelerini paylaştıkları, her türlü konuda iletişim kurdukları, hem içerik ürettikleri hem de tükettikleri dijital platformlardır (Sarigöz, Karakuş ve İrak, 2012). Günümüzde internet ve sosyal mecraların insanlar tarafından ne derece fazla kullanıldığı bilinmektedir ki bireyler bu ortamlarda ne kadar zaman geçirdiklerini bilemez hale gelmişlerdir (Karadağ ve Akçınar, 2019).

Günümüzde teknolojik gelişmeler ve internet kullanımının artması ile sosyal medya kullanımı da yoğun bir şekilde artmaktadır (Sarigöz, 2014). İnsanlar duygu ve düşüncelerini sosyal mecralarda paylaşarak kendilerini ifade etmektedirler. Twitter gerek anlık atılan tweet gerekse toplam kullanıcı sayısı ile en sık kullanılan sosyal medya mecralarından birisidir (Twitter, 2021).

Şahıslar, şirketler, gruplar, dernekler, devletler vb. kullanıcılar kendileri ile ilgili bilgilendirmeleri Twitter üzerinden duyurmaktadırlar. Hatta kullanıcılar bu mecrada oluşan algıya göre kendini şekillendirerek izleyeceği davranış ve politikaları belirlemektedir. Bu durum bu mecralarda oluşan verileri çok önemli hale getirmektedir. İşlenmemiş halde bulunan bu veriler gerekli doğal dil işleme teknikleri ile duygu analizi yapılarak anlamlı veriler haline getirilmektedir. Duygu analizi yapılan veriler olumlu, olumsuz ya da nötr olarak sınıflandırılmaktadır. Bu sayede firmalar, ürünler, kişiler, kuruluşlar gibi yapıların olumlu ve olumsuz yanları tespit edilebilmektedir. Pazar analizi, ürün memnuniyeti, film değerlendirmesi, müşteri memnuniyeti, siyasi icraatlara ilişkin durum değerlendirilmesi gibi birçok alanda duygu analizi kullanılmaktadır (Ravi ve Ravi, 2015).

Korona virüs hastalığı bulaşıcı bir hastalıktır. İlk görüldüğü 2019 yılından itibaren tüm dünyada hızla yayılmış ve milyonlarca insanın ölümüne neden olmuştur (Dünya Sağlık Örgütü, 2021). Ülkemizde ise ilk vaka Mart 2020 görülmüştür. İlk vaka görüldüğü tarihten itibaren çeşitli kısıtlamalara gidilerek hastalığın yayılması engellenmeye çalışılmıştır. Bu kapsamda restoranlar, kafeler, toplu organizasyonlar, spor müsabakaları gibi birçok toplu etkinliklerde kısıtlamalara gidilmiştir. Belirli zaman aralıklarında sokağa çıkma yasakları uygulanmıştır. Eğitim ve öğretimde ise ilk, orta, lise ve üniversiteler geçici bir süre kapatılarak uzaktan eğitime geçilmiştir (Wikipedia, 2021). Ülkemizde ilk defa milyonlarca öğrenci uzaktan eğitimle tanışmıştır. Bu uzaktan eğitim sürecinde teknolojik altyapı yetersizliği, bilgi, beceri eksikliği en büyük problemlerin başında gelmiş olsa da eğitim öğretim kesintiye uğramadan devam etmiştir.

Yapılan bu çalışmada korona virüs nedeni ile uygulanan kısıtlamalarda insanların sosyal medya üzerinden uzaktan eğitimle ilgili verdikleri olumlu ve olumsuz yorumları değerlendirerek bir duygu analizi modellemesi yapılmaya çalışılmıştır. Modelin oluşturulmasında izlenen adımlar Şekil 1'de verilmiştir.



Şekil 1. Model adımları

Sosyal mecralardaki paylaşımların artmasıyla metinlerde duygu analizi konusu da oldukça popüler konular arasına girmektedir (Liu, 2017). Bu konu ile ilgili son yıllarda birçok çalışma yapılmıştır.

Ramadhani ve Goo (2017) yapmış oldukları çalışmada sosyal medyanın çok popüler bir veri kaynağı olduğunu ve tahmin, duygu analizi gibi birçok amaçla kullanıldığına vurgu yapmışlardır. Derin öğrenme yöntemi kullanarak Twitter'dan alınan veriler üzerinde duygu analizi yapmışlardır.

Dubey'in (2020) yapmış olduğu çalışmada Covid-19 pandemi döneminde insanların atmış oldukları twitlerden duygu durumlarını analiz etmeye çalışmıştır. Analiz sonucunda insanların Covid-19 virüsüne karşı duygu durumları belirlenmiştir.

Yılmaz ve Orman (2021) yapmış oldukları çalışmada derin öğrenme yöntemlerinden Uzun Kısa Süreli Hafıza (LSTM) teknikleri ile Twitter üzerinden almış oldukları twitlerin duygu durumlarını olumlu ve olumsuz olarak sınıflandırmışlardır. Yaptıkları bu çalışmanın başarı doğruluk oranını ise %97 olarak elde etmişlerdir.

Chintalapudi ve ark. (2021) yapmış oldukları çalışmada Twitter'dan aldıkları veriler ile duygu analizi yapmışlardır. Yaptıkları analizde Bidirectional Encoder Representations (BERT) modelini kullanmışlardır. Kullandıkları modelle Lojistik Regresyon (LR), Destek Vektör Makineleri (SVM) ve Uzun kısa süreli bellek (LSTM) gibi diğer modelleri de karşılaştırarak her duygu durumu için ayrı bir hesaplama yapmışlardır. Bert modeli ile %89 başarı doğruluğu elde edilirken diğer karşılaştırdıkları modellerde ise LR %75, SVM %74,75 ve LSTM %65 başarı doğruluğu elde etmişlerdir.

Temizhan ve Mendeş (2021) yaptıkları çalışmada Covid-19 sürecinde atılan Twitlerden halkın bakış açısını duygu analizi yaparak tespit etmeye çalışmışlardır. 40.000 kadar twit üzerinde yapılan analizlerde atılan twitlerde en çok maske, Covid-19, Ankara Wuhan ve Bilim Kurulu kelimeleri kullanıldığı tespit edilmiştir. Tespit edilen bu kelimelerinde duygu analizleri yapılarak değerlendirilmiştir. Özellikle Covid-19 kelimesi üzerine olumsuz bir algı olduğu tespit edilmiştir. Aynı zamanda Bilim Kurulu kelimesinin de twit atanlar üzerinde dikkate değer bir etki bıraktığı da tespit edilmiştir.

Küçükartal (2020) yaptığı çalışmada sosyal medya mecralarından olan Twitter'da insanların tartıştığı konular hakkında metin madenciliği ve duygu analizi yöntemleri kullanarak twitlerin pozitif, negatif ve nötr olma durumlarını analiz etmiştir.

Akdeniz ve Cebeci (2021) yaptıkları çalışmada Twitter üzerinden Sakarya Belediyesinin yapmış olduğu hizmetlerle ilgili atılan twitler üzerinden Turkish Bert yöntemi ile duygu analiz yapılmıştır. Yapılan analiz sonucunda ilgili belediye ile ilgili olumlu ve olumsuz tutular tespit edilmiş ve değerlendirilmiştir.

Çoban ve ark. (2015) yapmış oldukları çalışmada sosyal medya platformu Twitter'dan elde ettikleri verilerle duygu analizi çalışması yapmışlardır. Duygu analizini beslemeli doküman sınıflandırma yöntemleri kullanarak veri setindeki twitleri olumlu ve olumsuz olarak analiz etmişlerdir. Çalışmada Destek Vektör Makineleri, Naive Bayes, Multinomial Naive Bayes ve K-en Yakın Komşu sınıflandırma algoritmalarını kullanmışlardır.

Yöntem

Kullanılan Araçlar ve Kütüphaneler

Bu çalışmada birçok teknoloji ve kütüphaneden yararlanılmıştır. Veri setinin eğitim amacı ile elle etiketleme kısmında verileri saklamak için MS Sql veri tabanı kullanılmıştır. Etiketleme arayüzü için ise C# programlama dilinde ASP.net teknolojisi kullanılmıştır. Veri setinin Twitter'dan indirilmesi, ön işleme yapılması ve modelleme kısımlarında ise Anaconda geliştirme ortamında Python programlama dili ile kullanılmıştır.

Veri ön işleme, modelleme, sınıflandırma aşamalarında python programlama dilinde bazı kütüphaneler kullanılmıştır. Bu kütüphaneler Tablo 1’de gösterilmiştir.

Tablo 1. Kullanılan python kütüphaneleri

Kütüphane	Kullanımı	Çalışmadaki kullanımı amacı
numpy	Matematiksel hesaplamalarda kullanılır.	Veri seti frame hesaplamalarında kullanılmıştır.
pandas	Veri işleme ve analizinde kullanılır.	Veri setinin okunmasında ve frame işlemlerinde kullanılmıştır.
re	Metni düzenleme ya da metinden alt parçalar edinmede kullanılır.	Veri seti ön işleme kısmında veri temizlemede kullanılmıştır.
NLTK	Doğal dil işlemede metinsel tokenizasyon, sınıflandırma, etiketleme gibi işlevleri gerçekleştirir	Veri setimizdeki gereksiz kelimeleri(stopwors) çıkarmak için kullanılmıştır.
jpype	Java dönüştürücüsü olarak kullanılır.	Zemberek kütüphanesinin çalıştırılmasında kullanılmıştır.
sklearn	Veri bilimi ve makine öğrenmesinde veri işleme, modelleme, regresyon, sınıflandırma ve kümeleme işlevlerini sağlamada kullanılır.	Veri ön işleme, metin vektör dönüşümü, TF-IDF, model sınıflandırma ve model başarı ölçümünde kullanılmıştır.

Veri Seti

Çalışmada kullanılan veri seti Twitter’den indirilmiştir. Twit indirmek için Twitter’ın kendi hazırlamış olduğu API’si veri büyüklüğü, zaman aralıkları vb. kısıtlamalar nedeni ile tercih edilmemiştir. Twitter API yerine python dilinde açık kaynak kodlu tüm ayarların yapabildiği Twitter Scraper Kütüphanesi kullanılarak twit indirme yazılımı geliştirilmiştir. Bu geliştirilen yazılım yardımı ile Türkçe dilinde atılmış “Uzaktan Eğitim” hashtagına sahip, 2020-2021 eğitim öğretim yılının uzaktan eğitim olarak yapıldığı zaman aralığının da ki 72.237 twit ‘csv’ formatında indirilmiştir. İndirilen twitler incelendiğinde bazı zaman aralıklarında twit yoğunlaştığı tespit edilmiştir. Bu dağılımı homojenleştirmek için 72.237 twitden belirli zaman aralıklarındaki Twitler silinerek seyreltmeler yapılmıştır. Yapılan seyreltmeler sonucunda veri setinde 2.948 adet Twit kalmıştır.

Veri setindeki twitler nötr, pozitif, negatif ve ilgisiz olarak elle etiketlenmiştir. 2.948 adet twitden elle etiketleme işlemi sonrasında nötr ve ilgisizler çıkarıldıktan sonra geriye 317 pozitif ve 889 negatif toplamda 1206 twit kalmıştır. Etiketleme işleminin sağlıklı bir şekilde yapılabilmesi için veriler öncelikle MS SQL veri tabanına alınmıştır. MS SQL veri tabanındaki veriler, web tabanlı olarak C# programlama dilinde geliştirilen ara yüzü Şekil 2’ de gösterilen program aracılığı ile dikkatli bir şekilde elle etiketlenmiştir.



Şekil 2. Veri seti etiketleme programı ara yüzü

Ön İşleme

Twitter'dan indirilen Tweetler yapısal olmayan formda bulunmaktadır. Bu Tweetleri yapısal forma dönüştürmek ve sonrasında duygu analizi yapabilmek için belirli ön işleme aşamalarından geçirilmesi gerekmektedir (García ve ark., 2016).

Veri setindeki yanlış yazılan kelimelerin doğruları ile düzeltilmesi gerekmektedir (Akın ve Akın, 2007). Bu düzeltme işlemlerinde zemberek kütüphanesinin metin normalizasyonu kullanılarak doğruları ile değiştirilmiştir. Daha sonra veri seti tokenize edilerek kelimelerine ayrılmıştır.

Veri setindeki etkisiz kelimeler duygu analizini olumsuz yönde etkilediği için veri setinden çıkarılmaları gerekmektedir (Qian, 2020). Bu çıkarma işleminde NLTK kütüphanesinin derlemiş olduğu etkisiz kelimeler (stopwords) listesi kullanılmıştır.

Ön işleme aşamalarından diğer bir önemli işlem ise gereksiz karakterlerin temizlenmesidir (İlhan ve Sağaltıcı, 2020). Veri setinde bulunan "#, @, 'http', 'https', '/n', '/r'" karakter ve kelimeleri, rakamlar, noktalama işaretleri temizlenmiştir. Ayrıca sosyal medya mecralarında sıklıkla kullanılan emojiler de temizlenmiştir. Veri seti tamamen küçük harfe dönüştürülmüştür. Veri setimizdeki kelimeler zemberek kütüphanesinin lemmatization fonksiyonu uygulanarak köklerine ayrılmıştır. Daha sonra veri setimizdeki kelimelerin belge içerisinde ne kadar önemli olduğunu belirlemek ve bu kelimeleri vektörize etmek için sklearn kütüphanesinden 'TfidfVectorizer' TF-IDF terim ağırlıklandırma metodu kullanılmıştır. IDF hesaplaması Denklem 1'de gösterildiği gibi yapılmaktadır.

$$IDF = \frac{\text{Döküman Sayısı}}{\text{Terimin Geçtiği Döküman Sayısı}}$$

Denklem 1. IDF Hesaplaması

Veri setine uygulanan ön işleme proseleri duygu analizinin doğru tahmin yapma ihtimalini arttırmaktadır.

Sınıflandırma

Veri setimizdeki Tweetlerin duygu durumlarını belirleyerek kategorize etmek için bazı sınıflandırma algoritmaları kullanılmıştır.

Kullanılan sınıflandırma algoritmaları:

- Naive Bayes (NB)
- Lojistik Regresyon (LR)
- Destek -Vektör Makineleri (SVM)
- Çok Katmanlı Algılayıcı (MLP)
- Ekstra Gradient Boosting (XGBoost)

Naive Bayes Algoritması

Naive Bayes algoritması olasılık tabanlı istatistiksel bir sınıflandırma algoritmasıdır. Bayes teorisi mantığı ile çalışarak koşul ve sınıf bağımsızlığı varsayımına dayanır. Sınıflandırma işlemlerinde başarı oranı yüksek ve kullanımı kolay olduğu için sıklıkla kullanılmaktadır (Han ve Kamber, 2006).

Destek Vektör Makineleri

Destek Vektör Makinelerinin temel amacı iki sınıfın verilerini birbirlerinden en uygun şekilde ayırarak sınıflandırmaktır. Bu işlem için hiper adı verilen düzlem sınırları belirler. Veri noktalarını ayırmak için birçok hiper düzlem bulunmaktadır. Bu hiper düzlemin farklı noktalarına düşen veriler farklı sınıflandırılarak bulunur (Shatkey ve Craven, 2012).

Lojistik Regresyon

Lojistik regresyon en temel sınıflandırma algoritmalarından biridir. Kategorik değişkenleri sınıflandırma da kullanılır. Eğitim setinde yer alan verilere göre sınıflandırma modeli oluşturur ve yeni gelen verilerin en yüksek olasılık değerine sahip sınıfa göre kategorize eder. Olasılık değerlerini ise Bayes teoremine göre hesaplayarak doğrudan parametreler üzerine tanımlar (Kaynar ve ark. 2016).

Çok Katmanlı Algılayıcı (MLP)

Çok katmanlı algılayıcılar yapa sinir ağı modellerinden biridir. Özellikle sınıflandırma işlemlerinde sık kullanılır. Katmandaki nöronların diğer nöronlara bağlanması ile oluşur. MLP 3 farklı katmandan oluşmaktadır. Bu katmalar girdi katmanı, gizli katman ve çıktı katmanıdır. Veriler girdi katmanında okunur. Sınıfların belirlendiği katman ise çıktı katmanıdır. Çıktı katmanı tek nöronda içerebilir, sınıf çeşidi kadar nöronda içerebilir. Bu durum oluşturulan model bağlıdır. Gizli katmanda ise girdi ile çıktı arasındaki işlemlerin gerçekleştiği kısımdır. Bu katmanda modele göre işlem yapıldığı için nöron sayıları net değildir (Taşcı ve Onan, 2016).

K-En Yakın Komşu

Örnek tabanlı bir sınıflandırma algoritması olan K-en yakın komşu algoritması oldukça yaygın olarak kullanılmaktadır. Eğitim setinde bulunan veriler modelin nasıl oluşturulacağını belirler. Örneklerin sınıflandırılmasında benzerlikler dikkate alınır. Veri setine katılacak yeni verinin, var olan verilere göre uzaklığı belirlenir, yakın komşuluğuna k sayısına göre bakılarak karar verilir. Öznitelik değerlerine göre sınıfa atanır (Towards Data Science, 2021).

Ekstra Gradient Boosting Algoritması (XGBoost)

Gradient Boosting algoritmasının geliştirilmesi ile ortaya çıkmıştır. Karar ağaçları algoritması tabanlı olarak geliştirilmiştir. Son yıllarda oldukça sık kullanılmaya başlanmıştır. Hız ve performans açısından diğer sınıflandırma algoritmalarına göre çok başarılı sonuçlar vermesi son yıllarda kullanım sıklığını arttırmıştır. XGboost yapılandırılmış veya tablo halindeki veri setlerinde sınıflandırma ve regresyon amaçlı kullanılmaktadır. Ayrıca bu algoritma ağaçları oluşturmak için CART algoritmasını kullanmaktadır (Towards Data Science, 2021).

Modelin Başarım Ölçütleri

Sınıflandırma algoritmalarının sonuçlarının kontrol edilerek başarı oranının değerlendirilmesi gerekmektedir. Değerlendirmede kullanılan en temel ölçüt doğru sınıflandırma oranıdır. Doğru sınıflandırma oranının (ACC) çalışma şekli ise doğru pozitifler (TP) ve doğru negatifler (TN) toplamının doğru pozitif (TP), yanlış pozitif (FP), yanlış negatif (FN) ve doğru negatiflerin (TN) toplanarak bölünmesi ile Denklem 2 deki gibi bulunur.

$$ACC = \frac{TN + TP}{TP + FP + FN + TN}$$

Denklem 2. Doğru Sınıflandırma Oranı Formülü

F-Ölçütü ise kesinlik (PRE) ve geri çağırma (REC) göre hesaplanır. 0 ile 1 arasında değerler alan F-Ölçütü Denklem (3, 4 ve 5) gibi hesaplanır.

$$F - \text{Ölçütü} = \frac{2 * PRE * REC}{PRE + REC}$$

Denklem 3. F – Ölçütü Formülü

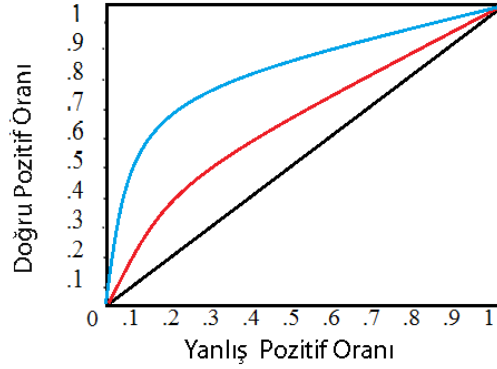
$$PRE = \frac{TP}{TP + FP}$$

Denklem 4. PRE Formülü

$$REC = \frac{TP}{TP + FN}$$

Denklem 5. REC Formülü

Şekil 3 'de gösterildiği gibi ROC eğrisinin altındaki alan (AUC) sınıflandırmanın başarısını ölçmede kullanılan bir yöntemdir. 1 ve 0 değerleri alınarak sınıflandırma yönteminin başarısını ölçer. Değer 1'e ne kadar yakınsa o kadar başarılı olarak kabul edilir.



Şekil 3. ROC Eğrisi

Bulgular ve Tartışma

Twitter'dan indirmiş olduğumuz veriler öncelikle elle etiketleme yapılarak her bir twitin duygu durumları olumlu, olumsuz, nötr ve ilgisiz olarak işaretlenmiştir. Daha sonra nötr ve ilgisizler veri setinden çıkarılmıştır. Geriye kalan 1206 twit ön işleme aşamalarından geçirilerek duygu analizine hazır hale getirilmiştir. Sonrasında ise veri setine duygu analizi yöntemleri uygulanarak sınıflandırma algoritmaları ile sınıflandırılmıştır. Sınıflandırma işlemleri ile oluşturulan modelin başarısı ölçülmüştür.

Veri setindeki tüm atılan twitler incelendiğinde Covid-19 virüsü ile ilgili olumsuz durumların yaşandığı bazı dönemde uzaktan eğitimle ilgili olarak olumsuz twitlerin yoğunlaştığı tespit edilmiştir. Belirli zamanlarda yaşanan bu durumun verilerin homojenliğini bozduğundan dolayı zaman bazlı verilerde seyreltme yapılmıştır. Kelime ağırlığı fazla olan cümleler incelendiğinde genel olarak atılan twitlerin uzaktan eğitime olumsuz yönde baktığı sonucu da çıkarılmaktadır.

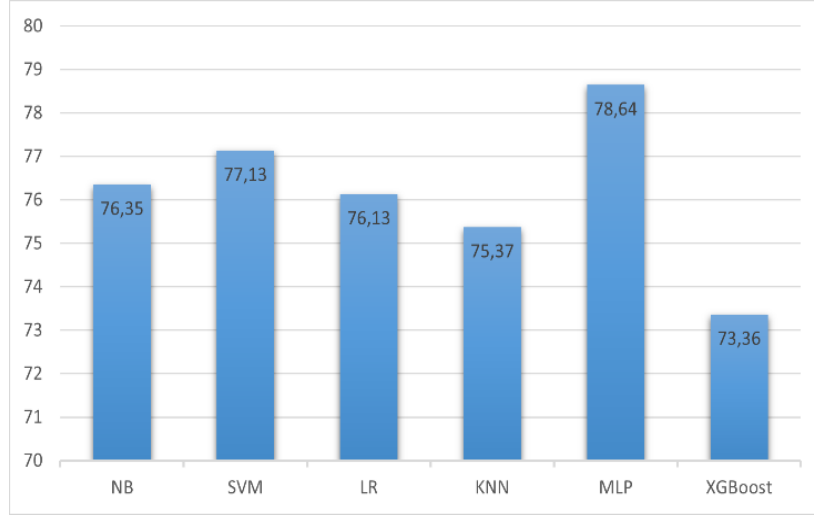
Veri setimizdeki kelimeler TF-IDF kelime ağırlıklandırılması Tablo 2'de gösterilmiştir. Kelimelerin ağırlıklarına göre ağırlık değerleri hesaplanmıştır. Veri setinin vektörleştirilmesi sonrasında veri kümesi %80 eğitim %20 test verisi olarak ayarlanmıştır. Daha doğru sonuçlar elde edebilmek için 10'lu çapraz doğrulama tekniği uygulanmıştır.

Tablo 2. TF-IDF Kelime Ağırlıklandırma

Kelime	Ağırlık
yeni	0.3970942437652462
öğretim	0.36539913190882006
süreci	0.32646785908303233
uzun	0.3001752785584245
yapmak	0.28054121285112704
döneminde	0.26499192954492734
durum	0.24135854328426465
dönemde	0.24135854328426465
geç	0.23200097541718012
olması	0.20397658247288808
geldi	0.19355824566443525
sinif	0.16406465640947934
bana	0.1513103229997521

devam 0.11146090588257186
yeni 0.3970942437652462

Modelin başarı oranları Şekil 4 'de görüldüğü gibi Naive Bayes(NB) %76.35, Destek Vektör Makineleri(SVM) %77.13, Lojistik Regresyon(LR) %76.13, K-En Yakın Komşu(KNN) Algoritması %75.37, XGBoost algoritması %73.36, Çok Katmanlı Algılayıcılar %78.64 başarı elde etmiştir. Görüldüğü üzere en başarılı algoritma MLP algoritması olmuştur. Genel itibari ile bakıldığında diğer sınıflandırma yöntemlerinin başarı değerlerinin de birbirlerine yakın olduğu gözlemlenmiştir.



Şekil 2. Modellerin başarı oranları(NB: Naive Bayes, SVM: Destek Vektör Makineleri, LR: Lojistik Regresyon, KNN: K En Yakın Komşu, MLP: Çok Katmanlı Algılayıcılar Algoritması)

Sonuç

Bu çalışmada, tüm dünyayı etkisi altına alan Covid-19 virüsünün etkisi ile ülkemizde başlatılan uzaktan eğitim hakkında insanların duygu durumlarını tahmin eden bir model geliştirilmiştir. Modelde kullanılan veri seti özel bir kod geliştirilerek belirli zaman aralılarını kapsayacak şekilde Twitter'dan indirilmiştir. Daha kolay ve doğru bir şekilde denetimli öğrenme yapılabilmesi için web tabanlı bir uygulama geliştirilerek veriler manuel olarak duygu durumları işaretlenmiştir. Çalışmada kullanılan veri setindeki veri miktarı artırılarak modelin daha başarılı sonuçlar elde edeceği ön görülmektedir. Ayrıca Türkçe dili için yeni geliştirilecek metin işleme kütüphaneleri modelde kullanılarak daha iyi sonuçlar elde edileceği düşünülmektedir.

Türkçe dilinde duygu analizi ile ilgili yapılan çalışmaların ve duygu durumları işaretlenmiş Türkçe veri setlerinin yetersizliği düşünüldüğünde bu çalışma gelecekte yapılacak çalışmalara bir iyi bir örnek teşkil edeceği düşünülmektedir.

Kaynakça

- Akdeniz, FNU., Cebeci, Hİ (2021). Belediye hizmetlerin değerlendirilmesinde duygu analizi yaklaşımı: Sakarya ili örneği. *Journal of Intelligent Systems: Theory and Applications*, 4(2), 127-135.
- Akın, MD., Akın, AA (2007). Türk dilleri için açık kaynaklı doğal dil işleme kütüphanesi: Zemberek. *Elektrik mühendisliği*, 431, 38-44.
- Ceron, A., Curini, L., & Iacus, S. M. (2016). iSA: A fast, scalable and accurate algorithm for sentiment analysis of social media content. *Information Sciences*, 367, 105-124.
- Chintalapudi N., Battineni G., Amenta F. (2020). COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. *Journal of Microbiology, Immunology and Infection*, 53(3): 396-403.
- Çoban Ö., Özyer B., Özyer GT (2015). Sentiment analysis for turkish twitter feeds. *23rd Signal Processing and Communications Applications Conference (SIU)*, 2388-2391.
- Dubey AD. (2020). Twitter sentiment analysis during COVID-19 outbreak. Available at SSRN 3572023.
- García S., Ramírez-Gallego S., Luengo J., Benítez, JM., Herrera F. (2016). Big data preprocessing: methods and prospects. *Big data analytics*, 1(1), 1-22.
- Han J., Kamber M.(2006). "Data mining: concepts and techniques", Morgan Kaufmann publishers, 340, 94104-3205.
- İlhan N., Sağaltıcı D. (2020). Twitter'da duygu analizi. *Harran Üniversitesi Mühendislik Dergisi*, 5(2): 146-156.
- Karadağ, A., & Akçınar, B. (2019). Üniversite öğrencilerinde sosyal medya bağımlılığı ve psikolojik semptomlar arasındaki ilişki. *Bağımlılık Dergisi*, 20(3), 154-166.
- Kaynar O., Yıldız M., Görmez Y., Albayrak A. (2016). Makine öğrenmesi yöntemleri ile duygu analizi. *International Artificial Intelligence and Data Processing Symposium*, 1-8, Malatya.
- Koronavirüs hastalığı (COVID-19), <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>. Erişim 01.12.2021.
- Küçükkartal HK. (2020). Twitter'daki verilere metin madenciliği yöntemlerinin uygulanması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 1(2): 10-13.
- Pathak, A. R., Pandey, M., & Rautaray, S. (2021). Topic-level sentiment analysis of social media data using deep learning. *Applied Soft Computing*, 108, 107440.
- Qian L., Hao P., Jianxin L., Congying X., Renyu Y., Lichao S., Philip SY., Lifang H. (2022). A survey on text classification: From Traditional to Deep Learning, 13(2): 1-41.
- Ramadhani AM., Goo HS (2017). Twitter sentiment analysis using deep learning methods. In 2017 7th International annual engineering seminar, 1-4.
- Ravi K., Ravi V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
- ROC Eğrisi, https://commons.wikimedia.org/wiki/File:Roc_curve.svg. Erişim 01.12.2021.
- Sarıgöz, O. (2014). Öğretim teknolojilerinin kullanımına yönelik öğretmen görüşlerinin değerlendirilmesi. *Uluslararası Beşeri ve Hakemli Akademik Bilimler Dergisi*, 5(10), 100–116.
- Sarıgöz, O., Karakuş, A., & İrak, K. (2012). Meslek yüksekokulu öğrencilerinin elektromanyetik kirlilik ile ilgili görüşlerinin değerlendirilmesi. *Electronic Journal of Vocational Colleges*, 2(2), 1-8.
- Shatkay H., Craven M. (2012). Mining the biomedical literature. MIT Press.

Taşcı E., Onan A. (2016). En yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. *Akademik Bilişim*, 1(1), 4-8.

Temizhan E., Mendeş M. (2021). COVID-19 pandemisi ile ilgili Twitter mesajlarının metin madenciliği tekniği ile değerlendirilmesi. *Türkiye Klinikleri Biyoistatistik*, 13(2), 185-200.

Towards Data Science, XGBoost Algoritması, <https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d>. Erişim 01.12.2021.

Twitter (2021). Twitter kurumsal web sitesi, <https://about.twitter.com/tr>, Erişim 01.12.2021.

Wearesocial.com (2024). wearesocial.com/wp-content/uploads/2023/03/Digital-2023-Global-Overview-Report.pdf, Erişim 01.04.2024.

Wikipedia (2021). Wikipedia, Türkiye'de COVID-19 Pandemisi, https://tr.wikipedia.org/wiki/T%C3%BCrkiye%27de_COVID-19_pandemisi. Erişim 01.12.2021.

Wikipedia (2024). <https://tr.wikipedia.org/wiki/Twitter#:~:text=Twitter'da%20kullan%C4%B1c%C4%B1lar%20%22tweet%22,yaln%C4%B1zca%20herkese%20a%C3%A7%C4%B1k%20tweetleri%20g%C3%B6r%C3%BCnt%C3%BCleyebilir>.

Yılmaz MC., Orman Z. (2021). LSTM Derin öğrenme yaklaşımı ile Covid-19 pandemi sürecinde twitter verilerinden duygu analizi, *Acta Infologica*, 5(2): 359-372.

Introduction

Distance education, which can be carried out through different channels from past to present, can now be carried out in advanced ways through the web. Although effective teaching can be carried out only by using distance education without face-to-face education, there are various opinions that face-to-face education is indispensable. However, in periods when face-to-face education is not possible due to natural disasters such as earthquakes and epidemics, distance education through the web has become inevitable and distance education is a very good alternative for today's conditions.

Distance education individualises teaching and gives the control of teaching to the student. This is a serious disadvantage for students whose time management and self-control are not developed or insufficient. In addition, the control of students' participation in the course and the objectivity of the measurement and evaluation processes in distance education are among the important disadvantages of this method.

Social media can be used to obtain the opinions of certain segments of the society on certain issues. The ability to share opinions quickly and easily by users on social media has turned these environments into large data masses where emotions are expressed. Sentiment analyses are needed to transform such a large amount of data into a meaningful whole (Pathak, Pandey, & Rautaray, 2021).

Method:

Many technologies and libraries were utilised in this study. MS Sql database was used to store the data in the manual labelling part of the dataset for training purposes. ASP.net technology in C# programming language was used for the labelling interface. Python programming language was used in Anaconda development environment for downloading the dataset from Twitter, pre-processing and modelling.

The data set used in the study was downloaded from Twitter. Twitter's own API for downloading tweets was not preferred due to limitations such as data size, time intervals, etc. Instead of Twitter API, tweet downloading software was developed in python language using the Twitter Scraper library, where all open source settings can be made. With the help of this developed software, 72,237 tweets with the hashtag "Uzaktan Eğitim" in Turkish language, in the time interval when the 2020-2021 academic year was held as distance education, were downloaded in 'csv' format. When the downloaded tweets were analysed, it was determined that there was a concentration of tweets in some time intervals. In order to homogenise this distribution, dilutions were made by deleting the tweets in certain time intervals from 72,237 tweets. As a result of the dilutions, 2,948 Twits remained in the data set.

The tweets in the data set were manually labelled as neutral, positive, negative and irrelevant. After the 2,948 tweets were manually labelled, after the neutral and irrelevant ones were removed, 317 positive and 889 negative tweets remained, leaving a total of 1206 tweets. In order to perform the labelling process in a healthy way, the data was first imported into the MS SQL database. The data in the MS SQL database were carefully labelled manually by means of a web-based program developed in C# programming language.

Results:

The data we downloaded from Twitter were first manually labelled and the emotional states of each tweet were marked as positive, negative, neutral and indifferent. Then, neutral and indifferent ones were removed from the data set. The remaining 1206 tweets were preprocessed and made ready for sentiment analysis. Afterwards, sentiment analysis methods were applied to the data set and classified with classification algorithms. The success of the model created with classification processes was measured.

When all the tweets in the data set were examined, it was determined that negative tweets about distance education were intensified in some periods when negative situations related to the Covid-19 virus were experienced. Since this situation experienced at certain times disrupted the homogeneity of the data, time-based data were diluted. When the sentences with high word weight are analysed, it is also concluded that the tweets generally look at distance education negatively.

Discussion and Conclusion:

In this study, a model that predicts people's moods about distance education, which was initiated in our country with the effect of the Covid-19 virus that has affected the whole world, has been developed. The data set used in the model was downloaded from Twitter to cover certain time intervals by developing a special code. In order to make supervised learning easier and more accurate, a web-based application was developed and the data were manually labelled with moods. It is predicted that the model will obtain more successful results by increasing the amount of data in the data set used in the study. In addition, it is thought that better results will be obtained by using newly developed text processing libraries for Turkish language in the model.

Considering the lack of studies on sentiment analysis in the Turkish language and the insufficiency of sentiment labelled Turkish data sets, this study is considered to be a good example for future studies.