

Sosyal medya paylaşımları hisse senedi değerleri üzerinde etkili mi?*

Are social media posts effective on stock values?

Demet Erdoğan Şengel¹ 

Naciye Tuba Yılmaz² 

¹ İş Zekası ve Data Analitiği Uzmanı, MARUBENİ Dağıtım ve Servis, Türkiye, e-mail: eng.demeterdogan@gmail.com

² Dr. Öğr. Üyesi, Marmara Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü, Türkiye, e-mail: tuba.yilmaz@marmara.edu.tr

Öz

Son yıllarda internet kullanımının yaygınlaşmasıyla birlikte bireyler bir mal veya hizmet konusunda karar verme süreçlerinde başka insanların yorumlarını da dikkate almaktadır. Benzer şekilde şirketler de müşterilerine daha iyi hizmet verebilmek ve kendi güçlü ve zayıf yanlarını görebilmek adına internetteki verileri kullanmaktadır. Bu süreçte duygu analizi, büyük verilerin yorumlanması için son yıllarda en çok tercih edilen yöntemlerden biri olmuştur. Bu çalışmanın amacı, yaygın olarak kullanılan sosyal medya platformlarından biri olan Twitter'da yapılan yazılı paylaşımların hisse senedi fiyatları üzerinde etkisinin olup olmadığını incelemektir. Bu anlamda, 1 Ocak 2016 ile 10 Nisan 2021 tarih aralığında #TSLA etiketi ile Twitter platformu üzerinde paylaşılan yazılı ifadelerin derin öğrenme yöntemlerinden biri olan BERT modeli ile duygu analizi gerçekleştirilmiş ve paylaşımların pozitif veya negatif anlama sahip olup olmadığı tespit edilmiştir. Paylaşım yapılan günün ardından gelen günde yorumların olumlu veya olumsuz olması durumuna göre TSLA hisse senedi değerinin azalma veya artış gösterip göstermediğinin tahmin edilmesi Neural Prophet ile gerçekleştirilmiştir. Sonuç olarak sosyal medya paylaşımlarının bireylerin hisse senetlerini alıp ve satıp kararları üzerinde etkili olduğu sonucuna varılmıştır.

Anahtar kelimeler: TSLA, Sosyal Medya, Twitter, BERT, Neural Prophet

JEL kodları: C45, D53, G17

* Bu çalışma birinci yazarın yüksek lisans tezinden üretilmiştir.

Citation/Attf: ERDOĞAN ŞENGEL, D. & YILMAZ, N. T. (2023). Sosyal medya paylaşımları hisse senedi değerleri üzerinde etkili mi? *Journal of Life Economics*. 10(4): 191-200, DOI: 10.15637/jlecon.2145

Corresponding Author/ Sorumlu Yazar:
Demet Erdoğan Şengel
E-mail: eng.demeterdogan@gmail.com



Bu çalışma, Creative Commons Atif 4.0 Uluslararası Lisansı ile lisanslanmıştır.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Abstract

With the widespread use of the internet in recent years, individuals take into account other people's comments in their decision-making processes about a good or service. Similarly, companies use online data to provide better service to their customers and to see their own strengths and weaknesses. In this process, sentiment analysis has become one of the most preferred methods for interpreting big data in recent years. The purpose of this study is to examine whether written posts on Twitter, one of the widely used social media platforms, have an impact on stock prices. In this sense, sentiment analysis of the written expressions shared on the Twitter platform with the hashtag #TSLA between January 1, 2016 and April 10, 2021 was performed with the BERT model, one of the deep learning methods, and it was determined whether the posts had positive or negative meanings. The prediction of whether the TSLA stock value decreases or increases depending on whether the comments are positive or negative on the day following the day of the post was performed with Neural Prophet. As a result, it is concluded that social media posts have an impact on individuals' decisions to buy and sell stocks.

Keywords: TSLA, Social Media, Twitter, BERT, Neural Prophet

JEL codes: C45, D53, G17

1. GİRİŞ

Her ne kadar karar almanın temel prensiplerinden biri rasyonellik olsa da bireyler karar alırken tüm alternatifleri çeşitli kriterler altında değerlendirirler ancak değerlendirme sonuçlarında her zaman rasyonel kararlar almamakta ve hatta duyguları da aldıkları kararlarda etkili olmaktadır. Almak istedikleri ürün ile alternatifleri fiyat, kalite, satış sırasında ve satış sonrasındaki destek, kullanım ömrü gibi objektif değerlendirmelerle ve ölçülebilen özelliklerini değerlendirmekle birlikte, ürünü daha önce kullanan kullanıcıların yaptıkları yorumlar nihai kararın verilmesinde etkili olabilmektedir. Aynı şekilde finans sektöründe de yatırımcılar yatırımlarını değerlendirebilmek için hisse senedi piyasasına yöneldiklerinde senedin açıklık değeri, hissenin yüksek ve düşük değeri, gün sonu kapanış değeri ve işlem hacmi, volatilitesi, senedin dönemsel seyri gibi kriterleri dikkate almakla birlikte diğer yatırımcıların yorumları da satın alma kararlarında etkili olabilmektedir.

Klasik medyada tek taraflı bir iletişim söz konusu iken, sosyal medyada karşılıklı bilgi paylaşımının daha kolay ve hızlı olduğu, katılımcı ve kullanıcının da aktif olarak konuşmaya dahil olabildiği bir ortam söz konusudur. Sosyal medya; katılımcılar, açıklık, toplum, konuşma ve bağlantılılık gibi özelliklere sahiptir (Solmaz vd. 2013). Akıllı telefon kullanımının da artmasıyla birlikte zaman ve mekân fark etmeksizin fotoğ-

raf, video ve metin gibi birçok bilgi ve belgenin diğer kullanıcılarla paylaşılabilme imkânı oluşmakta ve bu sayede topluluklar arasında hızlı ve etkili bir iletişim imkânı sunulmaktadır. Dolayısıyla sosyal medya ile insanların düşüncelerini, görüşlerini ve fikirlerini internet üzerinden diğer insanlarla etkileşimli olarak paylaştıkları aynı zamanda karar almalarında etkili bir ortam yaratılmış olmaktadır. Ayrıca sosyal medya platformları oylama, paylaşılan bilgi için yorum ve anket yapılabilmesi gibi yeni özellikleriyle de katılımcılara açıklık ve geribildirim imkânı sunarak kullanıcıların aktif bir şekilde etkileşimlerine olanak sağlamaktadır.

Sosyal medya uygulamalarından biri olan Twitter; 2006 yılında Jack Dorsey tarafından geliştirilmiştir ve sınırlı mikroblog altyapısına sahip ve günümüzde en aktif kullanılanıdır. Twitter'ın 2006 yılında ilk çıktığında ana ekranında yazan "what are you doing?" yazısı 2009 yılında "what's happening?" şeklinde değiştirilip bireysel kullanımdan çıkarılarak kişilerin hayatın her anı, alanı ve durumuyla ilgili gördüğü, yaşadığı, hissettiği şeyleri başkaları ile paylaşma konusunda daha teşvikçi bir yaklaşım izlenmiştir (Çalışkan ve Mencik 2015). Dünyada gündemde en çok konuşulan konulara erişilebilmekte, merak edilen konular hakkında arama yapılabilmekte ve hatta tüm olaylar hakkında klasik medyadan bile daha hızlı haber sahibi olmak Twitter ile mümkün olabilmektedir.

Twitter son yıllarda hem piyasalar için önemli bilgiler içeren finans haberlerinin paylaşıldığı hem de alanında uzman ve ilgili kişiler tarafından yatırım tavsiyelerinin verildiği bir platform ortamı da sunmaktadır. Yatırımcılar, yatırım kararlarını etkileyecek haberleri, tavsiyeleri ve yorumları Twitter'dan takip etmekte ve hızlı bir biçimde yatırım kararlarına yön verebilmektedirler. Bu paylaşımları dikkate alarak yapılan yatırım hamleleri ile finansal piyasaların davranış şekli de etkilenmektedir. Bu anlamda değerlendirildiğinde, Twitter paylaşımlarının analiz edilmesi gerekliliği öne çıkmaktadır. Son yıllarda sosyal medyada yapılan paylaşımları analiz etmek için sıklıkla duygu analizinden faydalanılmaktadır.

Duygu analizi, çeşitli algoritmalar aracılığı ile metinler üzerinde inceleme yapmaya olanak sağlayan, ifadelerin olumlu/olumsuz/tarafsız olarak değerlendirilme imkânı tanıyan bir alandır. Literatür incelendiğinde, sinema yorumlarının film seçimine etkisi, kitap yorumlarının satın alma kararına etkisi, futbol karşılaşmaları esnasında verilen tepkilerin sosyal medya ile ilişkisi, bir ürünle ilgili yapılan yorumların ürün satın alma kararı üzerindeki etkisi gibi konularda yapay sinir ağları, destek vektör makineleri, Naive Bayes gibi klasik algoritmalar kullanılarak duygu analizi gerçekleştiren çalışmalara rastlanmaktadır (Kaynar vd., 2016; Tuzcu, 2020; Korkusuz ve Carus, 2020; Seveli ve Kemaloğlu, 2020; Göçgün ve Onan, 2021). Büyük veri, makine öğrenmesi derin öğrenme yöntemlerinin gelişmesi sonucunda günümüzde BERT, DistilBERT, Electra vb. yeni algoritmaların daha yüksek başarı ile biyomedikal, sosyal medya gönderilerinden cinsiyet tespiti, bitcoin fiyatlarının tahmini, hisse senedi fiyatlarının tahmini gibi duygu analizinin etkisini gerçekleştirdikleri görülmektedir (Yıldırım ve Yüksel, 2017; Lee vd., 2019; Sel ve Hanbay, 2021; Khayyat vd., 2021; Borneklint 2021; Nagesh 2021; Akdağ ve Bozma 2021; Koca, 2021; Köksal vd., 2021; Yusufoglu vd., 2021).

Bu çalışmada amaç, Twitter sosyal platformu üzerinden yapılan yorumların bir hisse senedinin değeri üzerinde etkisi olup olmadığının incelenmesidir. Bu amaç doğrultusunda Twitter'da #TSLA etiketi ile paylaşılan yorumların derin öğrenme yöntemlerinden BERT modeli ile duygu

analizine bakılmış ve ertesi günkü TSLA hisse senedinin market değerinin artış veya azalışının tahmin edilmesinde etkili olup olmadığı incelenmiştir.

2. YÖNTEM

Teknoloji ile birlikte artık bilgisayar ve insan arasında da iletişim için bir dil doğmuştur. Bu bir yönüyle insan ve bilgisayar arasındaki etkileşim olarak da ifade edilebilmektedir. (Preece, 1994). Fakat insanlar kendi aralarındaki iletişim için doğal dil kullanırken bilgisayarların bu dili anlaması için işlemek gerekmektedir. Dolayısıyla bu iletişimin kurulabilmesi için makine öğrenmesi yöntemleri geliştirilmiştir. Makine öğrenmesinde kullanılacak veri boyutlarının büyümesi ve model eğitimlerinin bu boyut ile yetersiz kalması, modellerin daha fazla alanda kullanılmak istenmesi ve modelden daha fazla başarı beklentisi gibi durumlar derin öğrenme yöntemlerinin popüleritesini arttırmıştır. Video, film, müzik, grafik gibi ham veri seti ya da veri kaynağının artık bilgisayarın normal işletimcisinde işlenmesi ve makineye bir olgu öğretilerek değerli bilgiye dönüştürülmesi çok güç ve maliyetli olmasından dolayı büyük veri kaynağı ile çalışmak güçlüğü derin öğrenme yöntemlerinin geliştirilmesini önemli hale getirmiştir. Derin öğrenme yöntemleri, Convolutional Neural Network (CNN) olarak adlandırılan evrişimli sinir ağlarında yapay sinir ağına verilen girdi verisi çıktı seviyesine kadar tek yönlü iletilmektedir. Yapısında en az bir düğüm bulundurduğu ve girdi verilerinin ağırlıklar ile çarpıldığı ve ardından aktivasyon fonksiyonuna aktarıldığı bu tarz ileri beslemeli ağlara algılayıcı ağlar (perceptron) adı verilmektedir (Mikolov vd., 2013). İleri beslemeli ağlarda minimum bir girdi, bir veya daha fazla gizli girdi ve bir çıktı katmanının olması durumunda çok katmanlı algılayıcı ağlardan (multi layer perceptron – MLP) bahsedilmektedir (Konakoğlu, 2020). Tek katmanlı yapılar daha çok doğrusal problemlerin çözümünde kullanılabilirken, çok katmanlı algılayıcılar doğrusal olmayan, karmaşık problem çözümlerinde de kullanılabilir. Geleneksel sinir ağlarında bir önceki cümle, metin hafızasında kaydedilmediği için neden sonuç ilişkisi ölçümlendirilemez. Dolayısıyla klasik yapay sinir ağları her girdiyi tek başına ve

diğerleriyle ilişkisine bakmaksızın değerlendirir. Ancak, cümle öğelerinin sıralaması önemlidir ve tam ve doğru anlam bu sıra bir bütün olarak değerlendirildiğinde ortaya çıkmaktadır. Bu nedenle hafızası olan ve bağlantılılığı ölçeklendirilen sinir ağlarına ihtiyaç doğmuştur. Video, kelime dizisi, ve birbirine bağlı arka arkaya gelmiş veriler tekrarlamalı sinir ağları yani RNN (Recurrent Neural Network) mimarisini oluşturmaktadır. RNN isminden de anlaşılacağı üzere, mimarisinde metindeki her öğe için önceki çıktılarla bağlantılı olarak yeniden ve yeniden çalışan bir sistem kurulmuştur. RNN mimarisi gereği birçok alanda kullanım kolaylığı sağlamaktadır. Özellikle önceki katman bilgisi ve yeni katman bilgisi birleştirildiği için kelime tahmininde yüksek başarı elde edilmektedir. Long Short Term Memory (LSTM), giriş çıkış kapılarına ek olarak bünyesinde özel unutkan kapıları bulunduran özel bir tekrarlamalı sinir ağlarıdır (RNN) (Işık ve Yağcı, 2020). Bu yapı RNN'de oluşan bazı dezavantajları kaldırmak için bellek hücrelerinin eklenerek oluşturulmuş özel mimaridir. Birbirine bağlı ve birbirinin devamı niteliğinde olan bilgiler LSTM'in bellekli mimarisi sayesinde avantajlı bir şekilde işlenebilmektedir. Bu durum örneğin bir metinde genel bağlamın çıkarılmasına olanak sunmaktadır. LSTM tek yönlü ve çift yönlü çeşitleri bulunmaktadır. Çift yönlü (bidirectional) LSTM iki yönlü işlem yapabilmesinden dolayı metindeki anlamı çıkarma, soru cevaplama, sınıflandırma, öğrenme başarısı gibi alanlarda daha başarılıdır. Herhangi bir yön bazlı yani soldan sağa ya da sağdan sola şeklinde incelemek yerine aynı anda çift yönlü incelemektedir. Bu yönden evrişimli sinir ağları (CNN) ve tekrarlayan sinir ağlarına (RNN) göre başarısı daha yüksektir. Bu yapı bilhassa büyük boyutlu girdinin (metin) sınıflandırmasına, ön eğitime, özetlemesine oldukça uygundur ve bu alanlarda başarılıdır (Yang vd., 2019). Son yıllarda bu yöntemlerle birlikte doğal dil işleme alanında sıklıkla kullanılan ve başarılı sonuçlar vermesi bakımından dikkat çeken modeller arasında A lite Bert, Electra, DistilBert ve BERT modelleri yer almaktadır.

2.1. Bert Modeli

Google tarafından 2018 yılında bulunan BERT, Bidirectional Encoder Representations form Transformers kelimelerinin baş harflerinden oluşmaktadır. Bert modelini diğer klasik doğal dil işleme modellerinden ayıran özelliklerinden birisi de cümleyi öncelikle soldan sağa işlemesi ve sonrasında sağdan sola yeniden işlemesidir. Bert yalnızca cümle çevirisi için değil, doğal dil işleminin yapılabileceği tüm alanlarda kullanılmak üzere geliştirilmiştir. Soru yanıtlama, dil tercüme, bilginin bulunması ve getirilmesi, kelime işleme, arama motorları, metin ön işleme, argüman birleştirme, yazılı metnin sesli metne, sesli metnin yazılı metne çevrilmesi, hedeflenmiş reklamcılık, otomatik arama düzeltme ve tamamlama alanları, duygu analizi, yazı özetleme ve benzeri alanlar için probleme göre farklı katmanlar eklenerek Bert ile yüksek başarı ile çözümlenebilmektedir. Bert genel kapsamda iki ana bölümden oluşmaktadır. İlk aşamasında, yapının dilini ve bağlamı hakkında bilgi edinmeyi hedeflerken; ikinci aşamasında özel problemler için modelin nasıl kullanılacağına cevabını hedeflemektedir.

Diğer gözetimli öğrenen modeller gibi Bert'te de ilk aşama pre-train yani ön öğrenme aşamasıdır. Bu aşama dilin ve içeriğin ilk öğrenilmeye başlandığı süreçtir.

Bert, kelime dizinlerini bir bütün olarak almadan önce literatürde Masked Language Modeling (MLM) şeklinde geçen ve Türkçe karşılığı olarak Maskelenmiş Dil Modeli (MDM) olarak söylenebilecek ve %15'e karşılık gelen bir değerle cümledeki kelimeleri maskelenmektedir. Maskelenen kelimeleri dahil etmeden cümlenin geri kalanından anlam çıkararak bu boşta kalmış alanları tahmin etmeye çalışır. %15'e denk gelen rastgele seçilen kelimelerin %80'i [Mask] kelime parçacığı (token) ile değiştirilmekte, %10 herhangi rastgele atanan bir kelime ile değiştirilmekte ve geri kalan %10'luk kısım ise aynen orijinal kelime olarak kalmaktadır (Devlin vd, 2018). Bir cümledeki kelimelerle ilgili çalışılan bu aşamada kelimelerin birbiri ile ilişkisi incelenmiş olmaktadır.

Yöntemin ikinci aşamasında Sonraki Cümle Tahmini (Next Sentence Prediction (NSP) için keli-

melerin birbiri ile ilişkisi yerine cümlelerin birbirleriyle ilişkisine bakılır. Bu ilişkide cümlelerin birbirinin devamı niteliğinde olup olmadığı incelenir. Ön eğitimin bu görevinde her seferinde rastgele alınan iki cümlede öncelikle birinci cümlelerin %50'si ve ikinci cümlelerin %50'si seçilir. Önceki modellerde sadece cümle yerleştirmeleri (embeddings) alt görevlere aktarılırken Bert son görev (task) parametresini başlatmak için tüm yerleştirmeleri aktarır. Bert gibi ön eğitim modellerine modelin ilk eğitim süreçlerinin manuel yapılması çok masraflı ve zaman alan kısıtlayıcı bir durumdur. Problemin çeşidine göre bazı katmanlar eklenerek son eğitim aşaması son kullanıcı tarafından yapılabilir. Öğrenme sürecindeki maskelenmiş dil modeli ve sonraki cümle tahmini aşamaları eş zamanlı yapılmaktadır. Bu durum birbirlerinden etkilenmesini engellediği için başarıyı arttırmakta ve zaman kaybı oluşturmamaktadır. Bert modelinin en önemli avantajlarından birisi de kelimeyi konumuna göre değerlendirmesidir. Cümlelerin hangi bölümünde hangi kelimedenden önce ya da sonra geldiğine göre dikkati değişkenlik göstermektedir.

2.2. Neural Prophet Modeli

Neural prophet zaman serisi verilerini modellemek için üretilmiş sinir ağlarına dayalı gözetimli öğrenme modelidir. Zaman serisi düzenli periyotlar halinde tekrar eden eylemlerin gözlemlenmesi ve gözlemler sonucu elde edilen bilginin kaydedilmesi ile elde edilir. Bu periyotlar videolardaki gibi salise boyutundan gözlemden başlayarak saniye, dakika, saat, gün, ay, yıl şeklinde süreçler de olabilmektedir. Başarısı yüksek tahminler için deneyim, temiz ve kullanışlı veri kaynağı olmazsa olmazlardır. Prophet hala Facebook tarafından kullanılan ve ürünlerin karar verilme aşamasında, ürünlerin özellikleri aşamasında, karmaşıklığı giderme konusunda hala başarılı ve geliştirilerek devam eden model olma özelliği taşımaktadır. Neural Prophet, veri setinde aykırı değer veya eksik veri olması durumunda da kullanılabilen özelliklere sahip bir modeldir (Letham ve Taylor, 2017).

Neural Prophet modelinin dört ana birleşene sahip eklemeli regresyon modeli olduğu söylenebilir (Letham ve Taylor, 2017);

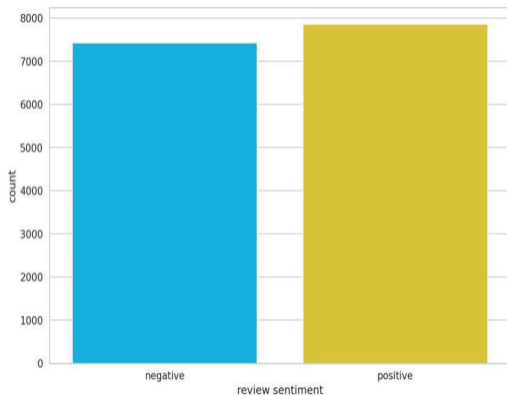
- Doğrusal veya lojistik büyüme eğrisinden kendisine göre noktalar seçerek eğilimdeki trend değişikliklerini inceleyip bunları otomatikleştirmesi,
- Fourier serisi kullanarak yıllık mevsimsel birleşen modellemesi,
- Kukla (dummy) değişkeni kullanarak haftanın mevsimsel bileşen etkisini belirlemesi,
- Kullanıcı tarafından belirlenecek önemli tarihin sisteme girilmesine olanak sağlaması.

3. ANALİZ VE BULGULAR

İnternet kullanımının ve sosyal medya platformlarının hızla yaygınlaşması ile yapılan görsel veya yazılı her türlü paylaşım anında, hesabı takip etmeyen bireylere bile ulaşabilmekte ve herhangi bir konu ile ilgili birçok kişinin fikir sahibi olmasına hatta kararlarına yön vermesine neden olmaktadır. Sosyal medyanın kattığı olumlu tarafların yanında beklenen ya da beklenmeyen olumsuz yanları da olmaktadır. Örneğin taraflı bir yaklaşımdaki olumsuz yorum restoranların, kitapların, filmlerin veya herhangi bir düşüncenin diğer insanlar üzerindeki etkisini de olumsuz etkileyebilmektedir. Yapılan birçok araştırmada sosyal medyanın birçok alanda etkisinin olduğunu kanıtlamıştır. Bu etkiden yola çıkarak bu çalışma kapsamında ise Twitter sosyal platformu üzerinde TSLA etiketi kullanılarak yapılan olumlu ve olumsuz yorumların ertesi gün markette TSLA hisse senedi değerinde artış veya azalış yaratıp yaratmadığı incelenmek istenmiştir. Twitter verilerinin probleme uyarlanabilmesi için ön eğitim süreci sinema, finans yorumları ve duygu etiketlerini içeren veri seti ile duygu analizi BERT yöntemi kullanılarak gerçekleştirilmiş, ardından twitler hisse senedi değerleri ile birleştirilerek Neural Prophet modeline aktarılmış ve gelecek tahmini için kullanılmıştır. Çalışma kapsamındaki verilere 1 Ocak 2016 tarihi ile 10 Nisan 2021 tarihleri arasında günlük 100 farklı #TSLA etiketini kullanarak yapılan İngilizce paylaşımlar, Twitter'ın Twint¹ aracı kullanılarak toplam 2432 adet twit çekilmiştir. Ayrıca finansal değerler için Yahoo Finans² Finance API³ kaynağı üzerinden TSLA hisse senedine ait hisselerin açılış değeri, hisselerin yüksek ve düşük değeri, gün sonu kapanış

değeri ve işlem hacmi verileri elde edilmiştir. Doğal dil işlemede Bert modeli gözetimli yani öğreticiyle öğrenen bir modeldir. Gözetimli öğrenmesi için daha önceden hazırlanmış ve analiz edilmiş yani duyguların etiketlenmiş halini modele tanıtmak ve onun öğrenmesini sağlamak gerekmektedir. Bert modelinin ilk adımdaki eğitiminde kullanılmak üzere çalışma kapsamında Kaggle'dan etiketlenmiş veri setleri alınmıştır⁴. Orijinal veri setinde duygular; negatif, aşağı yukarı negatif, nötr, aşağı yukarı pozitif ve pozitif duygu etiketleri bulunmaktadır. Bu nedenle veriler negatif yani olumsuz ve pozitif yani olumlu olmak üzere iki ana gruba indirgenmiş ve nötr yorumlar dahil edilmemiştir. Sadece sinema ya da film yorumları kullanımı modeldeki başarıyı etkileyeceğinden dolayı bu durum hisse senedi marketi hakkında etiketlenmiş veri seti ihtiyacını doğurmuştur. Bu nedenle bir diğer etiketlenmiş veri seti de yine Kaggle üzerinden⁵ elde edilmiştir. Bu veri setinde bir öncekinden farklı olarak olumlu ya da olumsuz yani negatif / olumsuz yorumda (-1) ve pozitif/ olumlu için (+1) kullanılmıştır. Veri setinde 5791 cümle ve bunların duygu analiz etiketleri vardır. Bu veri seti ile sinema yorumları veri seti birleştirilmiş ve -1 negatif yorumu +1 ise pozitif yorumu içeren üç veri setinin toplamı toplamda 14000 satır oluşturmuştur. Birleştirilen sinema yorum verileri ve hisse senedi yorum verilerinde negatif ve pozitif etiketlerde sayıca dengesizlik tespit edilmiş veri dengesi sağlanabilmesi için iyileştirme yoluna gidilmiştir. Bu iyileştirme işleminden sonra pozitif ve negatif etiketli verilerin toplamı 15270 satır olmuştur.

Şekil 1. Veri Setindeki Negatif ve Pozitif Etiketli Satırlar



Verinin modele hazır hale getirilmesinden sonra Colab platformu üzerinde BERT modeli işlenmiştir. Bert modeli, ön eğitilmiş probleme özel katman eklenerek problemle ilgili veri setinden sonraki etiketlenmiş öğrenme süreci çalışılmıştır. Önceden eğitilmiş model Hugging face API'sinden çekilmiştir. Bert'in de bunlardan birisi olduğu beş binden fazla büyük organizasyonun paylaştığı açık kaynaklı modeller bulunmaktadır. BERT modelinde Maskelenmiş Dil Modeli (MLM) ve örnek kelimelere ayırma (tokenization) işlemleri gerçekleştirilmiştir. Ön işlemlerden geçirilmiş 15270 satırlık, iki etiket sınıfından oluşan veri ile modele ön öğretim işlemi gerçekleştirilmiştir. Bu çalışma kapsamındaki modelin ön öğretimi çalışma yapıldığı tarihte 800 milyon kelime hazinesi olan Toronto Book Corpus ve 2,5 milyon kelimedenden oluşan İngilizce Wikipedia kaynakları ile yapılmıştır. Transformer'ın yapısında olan kodlayıcı (encoder) ve kod çözücü (decoder) sürecinin encoder süreci yani kelimenin kökeni, yapısı, modelin dili algılama süreci bu ön öğretim (pre-train) ile modele verilmiş olmaktadır. Bert modeli bu süreçle birlikte kod çözücü sürecini işleyebilecek duruma getirilmiştir.

Veri setinin kullanılabilir olması, sonucu direkt etkileyeceği için büyük önem taşımaktadır. Bu yüzden veri setinde eksik değer olup olmadığı kontrol edilmiş ve eksik değer olmadığı görülmüştür.

Şekil 2. Değer Şeması

```
[ ] df.info()

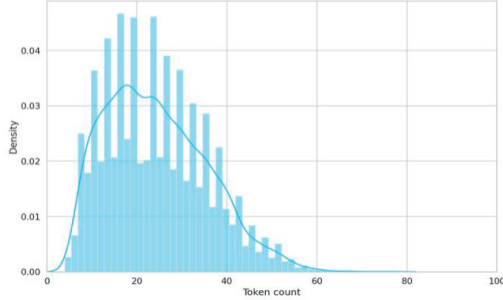
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15270 entries, 0 to 15269
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Text        15270 non-null  object
1   Sentiment   15270 non-null  int64
dtypes: int64(1), object(1)
memory usage: 238.7+ KB
```

Bert modelinin kasalı (cased) sürümü duygu analizinde daha iyi çalıştığı için tercih edilmiştir. Bu yöntemde olumsuz anlama sahip bir kelimenin küçük veya büyük harfle yazılması arasındaki duygu farkı da dikkate alınmaktadır.

Bölütleme (tokenization) işlemi metni bölütlere ve bölütleri de benzersiz tam sayılara (id) dönüş-

türür. Cümle sonlarında bitirildiğine dair cümle bitiş işareti kullanılır. Belirteçler yani bölütlenmiş özel id edinmiş benzersiz sayılar Tensor'lerde (Tensor) saklanır. Dikkat maskesi (attention mask) ismi verilen sabit uzunluk ise bu çalışma kapsamında 32 seçilmiştir.

Şekil 3. Bölüt Sayıları ve Ağırlıkları



Şekil 3'te görüldüğü üzere geneli 80'den az kelime içeren cümleler olduğu için maksimum uzunluk 80 olarak belirlenmiştir.

Sonraki aşamada makinanın öğrenmesi ve verinin hesaplanabilmesi için metin veri seti %90 eğitim (training), %5 doğrulama (validation) ve %5 test (testing) şeklinde üç parçaya ayrılmıştır. Tüm veri setini bölme işlemi sonucu eğitim için 13743 satır metin, doğrulama için 763 satır metin ve test için 764 satır metin elde edilmiştir. Eğitim, doğrulama ve test verisine ait bilgiler Şekil 4'te görüldüğü gibidir.

Şekil 4. Veri Bölütleme Sonucu

```
[ ] df_train.shape, df_val.shape, df_test.shape
((13743, 3), (763, 3), (764, 3))
```

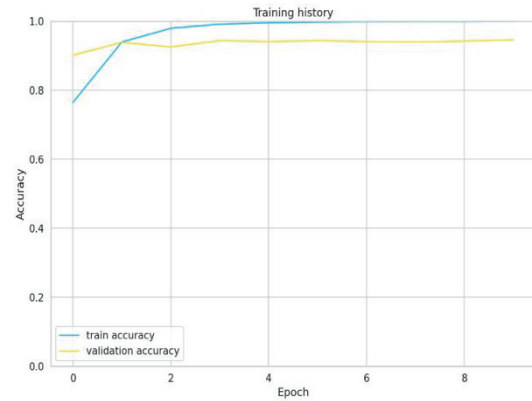
Sınıflandırma işlemi için olması gereken işaret cümlelerin en başına ve cümlelerin en sonuna eklenir. Metin değeri sabit değer belirtildiği için belirtilen maksimum uzunluktan kısa ise boşluklar doldurulur fakat eğer metin daha uzun ise o zaman belirtilen sabit değer kadar kelime ile ifade edilir. Bu işlem ile dikkat maskeleri oluşturulmuştur (attention mask) ve cümleler bu işlemler ile birlikte tensor objesi şeklinde geri döndürülmektedir. Bölütlerine ayrılmış kelimelerin id numaraları ve dikkat maskesi değerleri ile olasılıkların toplam değerinin 1 olduğu Şekil 5'te görülmektedir.

Şekil 5. Tensor ve Dikkat Maskesi Çıktısı

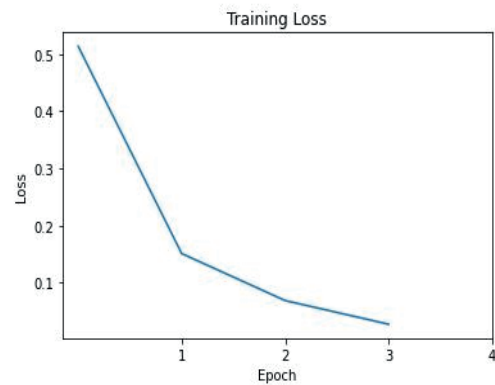
```
[ ] F.softmax(model(input_ids, attention_mask), dim=1)
tensor([[0.3311, 0.6689],
        [0.4530, 0.5470],
        [0.5488, 0.4512],
        [0.3166, 0.6834],
        [0.7348, 0.2652],
        [0.3665, 0.6335],
        [0.4566, 0.5434],
        [0.6227, 0.3773],
        [0.6790, 0.3210],
        [0.5921, 0.4079],
        [0.4571, 0.5429],
        [0.6397, 0.3603],
        [0.3888, 0.6112],
        [0.5508, 0.4492],
```

Modelle ilgili aşamalar tamamlandıktan sonra toplanan sinema yorumları, borsa yorumları ve duygu puanları (-1, +1) ile modelin eğitiminin gerçekleştirilmesi aşamasına geçilmektedir. Modelin devir tekrar sayısı ile öğrenme sonucunda elde edilen doğruluk oranı Şekil 6'da görülmektedir. Şekil 7'de işlem sayısı arttıkça kayıp değerlerinin düştüğü görülmektedir.

Şekil 6. Epoch Sayısı ve Doğruluk Orantısı



Şekil 7. İşlem Sayısı ve Kayıp Değeri



Elde edilen duygu verileri kullanarak öğrenim süreci tamamlanan BERT modeli ile #TSLA eti-

ketiyle toplanan twitlerin duygu analizi her yorum için 0 ile 1 arasında duygu puanlaması yapılmıştır.

Şekil 8. Bert Model ile Tweet Sentiment Analizi

date	Text	Sentiments
1/1/2016	#HBRLive: Is #Tesla's	0,9999610185623160
1/1/2016	TSLA: \$240.01 - Markı	0,9999639987945550
1/1/2016	CityFALCON score giv	0,9999347925186150
1/1/2016	Potential Tesla compet	0,288075170828,96
1/1/2016	TSLA: \$240.01 - Markı	0,9999639987945550
1/1/2016	Faraday Future aims tı	0,354088188032,619
1/1/2016	TSLA: \$240.01 - Markı	0,9999639987945550
1/1/2016	Panasonic to build \$40	0,9999576807022090
1/1/2016	A collection of testimor	0,9997859597206110
1/1/2016	Tesla Motors (TSLA) tı	0,43861677113,4547
1/1/2016	TSLA: \$240.01 - Markı	0,9999639987945550
1/1/2016	We calculated 63-day	0,9722539186477660
1/1/2016	3 Activity Ratios to Act	0,9998416900634760
1/1/2016	Looking for someone v	0,9999247789382930
1/1/2016	Will Faraday Future Rı	0,74341471190,5643
1/1/2016	Will Faraday Future Rı	0,545435868843,924
1/1/2016	TSLA https://t.co/JVy	0,12456854165066000
1/1/2016	TSLA: \$240.01 - Markı	0,9999635219573970
1/1/2016	Will Faraday Future Rı	0,49723406846,1693

Modelin çalıştırılması sonucunda her yorum için 0 ile 1 aralığında duygu analiz değeri belirlenmiştir. Elde edilen duygu analizi sonuçları hisse senedi değer tablosu ile birleştirilerek Neural Prophet modeli ile analiz gerçekleştirilmiş ve her gün için bir gün sonraki hisse senedi değerleri tahmin edilmiştir.

Şekil 9. TSLA Market Değeri ve Sentiment Değerleri

Date	Open	High	Low	Close	Volume	Sentiments
0 1/4/2016	46.144001	46.276001	43.799999	44.681999	34135500	0.622939
1 1/5/2016	45.271999	45.377998	44.000000	44.686001	15934000	0.371077
2 1/6/2016	44.000000	44.009998	43.195999	43.807999	18895500	0.704350
3 1/7/2016	42.838001	43.688000	42.734001	43.130001	17771500	0.687406
4 1/8/2016	43.571999	44.088001	42.153999	42.200001	18140500	0.758833
5 1/11/2016	42.801998	42.889999	40.599998	41.570000	20457000	0.900371
6 1/12/2016	42.320000	42.748001	41.062000	41.993999	15459500	0.751501
7 1/13/2016	42.402000	42.529999	40.000000	40.062000	20632000	0.705132
8 1/14/2016	40.442001	42.000000	38.675999	41.236000	32453500	0.800704
9 1/15/2016	39.793999	41.014000	39.450001	40.998001	27893000	0.689339

Tablo 1 TSLA hisse senetlerinin kapanış değerlerinin Neural Prophet modeli sonucunda elde edilen 50 günlük tahmin değerlerini göstermektedir. Şekil 10'da ise 1 ile kodlanan ifadeler hisse senedi değerinin bir sonraki gün arttığını, 0 ise azaldığını ifade etmektedir.

Tablo 1. Neural Prophet Sonucu Gelecek 50 Günlük Market Kapanış Tahmin Değerleri

GÜN	DEĞERLER				
1-5	626.517	606.082	575.260	602.863	607.866
6-10	603.175	603.233	613.689	610.822	622.121
11-15	600.752	607.577	622,273	628.494	621.699
16-20	628.002	661.537	687.510	674.535	695.553
21-25	682,493	686.881	679.769	688.591	661.085
26-30	645.815	656.012	657.872	690.813	671.961
31-35	658.040	651.664	646.108	647.995	660.790
36-40	656.936	651.280	644.040	661.900	642.073
41-45	649.074	683.722	718.701	709.933	716.653
46-50	715.937	702.022	716.075	710.610	708.802

Şekil 10. Gerçekleşen ve Tahmin Edilemeyen Değer Tablosu

Actual y	0	1
Predicted	1	1

Yalnızca doğruluk (accuracy) sonucuna bakılarak modelin değerlendirmesi yapılmamalıdır. Birden fazla metriğe bakmak model başarısı anlamında daha doğru yorum yapılmasını sağlamaktadır. Doğruluk değeri, doğru tahmin edilen değerlerin veri setindeki tüm değerlere oranı şeklinde açıklanabilmektedir. Kesinlik (precision), etiket verilen yani seçilen durumların gerçekte de doğru etiketi aldığı durumlardır. Duyarlılık (recall) değeri ise yanlış negatiflerin değeridir. F1 score değeri, kesinlik ve duyarlılık test değerlerinin ortalamasıdır.

Model için tüm kriter sonuçları incelenmiş ve BERT ve Neural Prophet modellerine ait değerler Şekil 11 ve 12'de gösterilmiştir.

Şekil 11. Bert Model Başarı Tablosu

	precision	recall	f1-score	support
negative	0.95	0.97	0.96	387
positive	0.97	0.95	0.96	377
accuracy			0.96	764
macro avg	0.96	0.96	0.96	764
weighted avg	0.96	0.96	0.96	764

Şekil 12. Neural Prophet Model Başarı Tablosu

```

Accuracy is 64.0
True positve 31
True negative 1
False positive 17
False negative 1
50

```

Şekil 11 incelendiğinde, tüm kriterlere göre Bert modelinin başarılı sonuçlar verdiği görülmektedir. Şekil 12 incelendiğinde ise, Neural Prophet modelinin kabul edilebilir nitelikte olduğu söylenebilir.

4. SONUÇ VE DEĞERLENDİRME

Teknolojik gelişmenin sonucu olarak internet kullanımındaki kolay erişim sayesinde sosyal medyanın da önemi her geçen gün artmaktadır. Sosyal medyadaki kullanım artışı anlık etkileşimi arttırmaktadır. Sosyal medya kullanımının artması ve etkileşimin de artmasından dolayı bu çalışma kapsamında sosyal medyadaki yorumların borsa üzerine etkisi ölçülmüştür. Bert modeli ile duygu analizi yapılarak günlük bir değer ortalaması tespit edilmiştir. Bu duygu etiketinin pozitif ise ertesi gün hisse değerinin artış göstermesi, negatif ise ertesi gün hisse senedi değerinin azalması beklenmiş ve bu test edilmiştir. Sonuç olarak Bert modeli başarısı %96 olarak bulunmuştur. Neural Prophet modelinin başarısı ise %64 olarak gerçekleşmiştir. Bu anlamda sosyal medyada yapılan olumlu veya olumsuz yorumların hisse senedinin değeri üzerinde etkisi olduğu söylenebilmektedir. Sosyal medya üzerinden alınan verinin sayısının arttırılması ile modelin başarısının artacağı öngörülmektedir. Büyük veri seti ve uzun süreli veri toplama ile yapılması başarıyı arttırabilecek faktörlerden olduğu düşünülmektedir.

Sonnot

¹<https://github.com/twintproject/twint>

²<https://finance.yahoo.com/quote/TSLA/history/?guccounter=1>

³<https://algotrading101.com/learn/yahoo-finance-api-guide/>

⁴<https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews/data>

⁵<https://www.kaggle.com/yash612/stockmarket-sentiment-dataset>

KAYNAKÇA

AKDAĞ, M. & BOZMA, G. (2021). Stok Akış Modeli ve Facebook Prophet Algoritması İle Bitcoin Fiyatı Tahmini. *Uluslararası Ekonomi İşletme ve Politika Dergisi*, 5(1), 16-30.

BORNEKLINT, N. (2021). *Forecasting prices of Bitcoin and Google stock with ARIMA vs Facebook Prophet*, Bachelor's Thesis, University West, School of Business, Economics and IT Division of Law, Economics and Politics.

ÇALIŞKAN, M. & MENCİK, Y. (2015). Değişen Dünyanın Yeni Yüzü: Sosyal Medya. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi*, (50), 254-277.

DEVLIN, J., CHANG, M. W., LEE, K. & TOUTANOVA, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

GÖÇGÜN, Ö. F. & AYTUĞ, O. (2021). Amazon Ürün Değerlendirmeleri Üzerinde Derin Öğrenme/Makine Öğrenmesi Tabanlı Duygu Analizi Yapılması. *Avrupa Bilim ve Teknoloji Dergisi*, (24), 445-448.

SEL, İ. & HANBAY, D. (2021). Ön Eğitimli Dil Modelleri Kullanarak Türkçe Tweetlerden Cinsiyet Tespiti. *Firat University Journal of Engineering*, 33(2), 675-684.

IŞIK, A. H. & YAĞCI, A.Ş. (2020). Sıradan Sıraya LSTM Modeli ile Telegram Bot Uygulaması. *Gazi Mühendislik Bilimleri Dergisi*, 6(1), 32-39.

KAYNAR, O., GÖRMEZ, Y. & YILDIZ, M., ALBAYRAK, A. (2016). Makine öğrenmesi yöntemleri ile Duygu Analizi. *International Artificial Intelligence and Data Processing Symposium (IDAP'16)* (Vol. 17).

KHAYYAT, M., LAABIDI, K., ALMALKI, N. & AL-ZAHRANI, M. (2021). Time Series Facebook Prophet Model and Python for COVID-19 Outbreak Prediction. *CMC-COMPUTERS MATERIALS & CONTINUUA*, 67(3), 3781-3793.

KOCA, G. (2021). Bitcoin Üzerine Twitter Verileri ile Duygu Analizi. *Anadolu Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, 22(4), 19-30.

KONAKOĞLU, B. (2020). Çok Katmanlı Algılayıcı Yapay Sinir Ağı ile Jeodezik Elipsoid Koordinatların (φ, λ, h) 3 Boyutlu Global Kartezyen Koordinatlara (X, Y, Z) Dönüşümü. *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 10(3), 702-710.

KORKUSUZ, R. & CARUS, A. (2020). Futbol Müsabakaları ile İlgili Tweetlerin Anlık Duygu Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, 386-396.

KÖKSAL, B. , ERDEM, G. , TÜRKELİ, C. & KAMIŞLI ÖZTÜRK, Z. (2021). Twitter’da Duygu Analizi Yöntemi Kullanılarak Bitcoin Değer Tahminlemesi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, Additional Issue, 280-297. doi: 10.29130/dubited.792909

LEE, J., YOON, W., KİM, S., KİM, D., KİM, S., SO, C. H. & KANG, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234-1240.

MIKOLOV, T., CHEN, K., CORRADO, G. & DEAN, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

NAGESH, P. (2021). Combination of Facebook Prophet and Attention-Based LSTM with Multi-Source data for Indian Stock Market Prediction. Dissertation, Technological University Dublin, doi:10.21427/ tzy-4r42

PREECE, J., ROGERS, Y., SHARP, H., BENYON, D., HOLLAND, H. & CAREY, T. (1994). *Human-Computer Interaction*. Pearson Education Limited: Essex, England.

TUZCU, S. (2020). Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 1(2), 1-5.

SEVLİ, O. & KEMALOĞLU, N. (2021). Olağandışı Olaylar Hakkındaki Tweet’lerin Gerçek ve Gerçek Dışı Olarak Google BERT Modeli ile Sınıflandırılması. *Veri Bilimi*, 4(1), 31-37.

SOLMAZ, B., TEKİN, G., HERZEM, Z. & DEMİR, M. (2013). İnternet ve sosyal medya kullanımı üzerine bir uygulama. *Selçuk İletişim*, 7(4), 23-32.

TAYLOR, S. J. & LETHAM, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37-45.

YANG, Z., DAI, Z., YANG, Y., CARBONELL, J., SALAKHUTDINOV, R. R. & LE, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*. 33rd Conference on Neural Information Processing Systems, Vancouver, Canada.

YILDIRIM, M. & YÜKSEL, C.A., (2017). Sosyal Medya ile Hisse Senedi Fiyatının Günlük Hareket Yönü Arasındaki İlişkinin İncelenmesi: Duygu Analizi Uygulaması. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 22. UPK Ahmet Hamdi İslamoğlu Özel Sayısı, 33-44.

YUSUFOĞLU, H., AYDIN, H. & ÇETİNKAYA, A. (2021). Twitter Üzerindeki Finansal Tweetlerin LSTM Sinir Ağı Algoritması ile Duygu Analizi. *Veri Bilimi*, 4(3), 28-43.