

Sosyal Medya ve Diğer Yatırım Aracı Verilerine Dayalı Hisse Senedi Değeri Tahmini

Forecasting Stock Value Based on Data from Social Media and Investment Instruments

Ömer Faruk Uyrun¹ , İbrahim Sabuncu² 



ÖZ

Bu çalışmada, farklı makine öğrenmesi teknikleriyle yatırım aracı verileri ile birlikte sosyal medya verileri kullanılarak hisse senedi tahminlenmesi amaçlanmıştır. Çalışma kapsamında, beş farklı havayolu firmasına ilişkin Ekim 2019 – Şubat 2020 dönemine ait 236 764 adet tweet ve söz konusu şirketlerin hisse senedi değeri ve işlem gördüğü borsanın günlük verileri, dolar kuru ve altın fiyatları ele alınmış olup, tweet'lerin analizinde duygu analizi gerçekleştirilmiştir. Çalışmada, Gradyan Destekli Ağaçlar (Gradient Boosted Trees) algoritmasının hisse senedi tahminlemesinde en düşük hata payına sahip tahmin modeli olduğu tespit edilmiş olup, şirketler hakkındaki net pozitif (pozitif-negatif) tweet sayılarının hisse senedi değeri tahminindeki en etkili faktörlerden birisi olduğu görülmüştür. Çalışma sonucunda, Gradyan Destekli Ağaçlar algoritmasının çalışma kapsamında kullanılan diğer algoritmalara göre hisse senedi tahminlemesinde etkin olduğu ve Twitter verisinin diğer yatırım verileri ile birlikte hisse senedi değeri tahmininde faydalanılabilecek bir veri kaynağı olduğu düşünülmektedir.

Anahtar kelimeler: Sosyal Medya Analitiği, Veri Madenciliği, Tahmin Modeli, Hisse Değeri, Twitter

ABSTRACT

This study aimed to predict stocks using different machine learning techniques with social media data and investment instrument data. Within the scope of the study, 236,764 tweets related to five different airline companies during the period October 2019 - February 2020, the stock value of those companies, the daily data of the stock market, dollar rate and gold prices were discussed. Additionally, sentiment analysis was carried out in the analysis of the tweets. In the study, it was determined that the Gradient Boosted Trees algorithm was the prediction model with the lowest margin of error in stock prediction, and it was seen that the number of net positives (positive-negative) tweets about companies was one of the most influential factors in forecasting stock value. As a result of the study, it is thought that the Gradient Boosted Trees algorithm is effective in stock prediction compared to the other algorithms used in the study, and that Twitter data is a data source that can be used in forecasting stock value together with other investment data.

Keywords: Social Media Analytics, Data Mining, Prediction Model, Stock Value, Twitter

¹ (Öğr. Gör.), Yalova Üniversitesi, Armutlu Meslek Yüksekokulu, Yönetim ve Organizasyon Bölümü, Yalova, Türkiye
² (Dr. Öğr. Üyesi), Yalova Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Yalova, Türkiye

ORCID: Ö.F.U. 0000-0002-4060-5069;
İ.S. 0000-0001-8625-9256

Corresponding author:

Ömer Faruk UYRUN
Yalova Üniversitesi, Armutlu Meslek Yüksekokulu, Yönetim ve Organizasyon Bölümü, Yalova, Türkiye
E-mail address: omer.uyrun@yalova.edu.tr

Submitted: 12.05.2021
Revision Requested: 26.07.2021
Last Revision Received: 15.11.2021
Accepted: 18.11.2021
Published Online: 13.12.2021

Citation: Uyrun, O. F. ve Sabuncu, I. (2021). Sosyal medya ve diğer yatırım aracı verilerine dayalı hisse senedi değeri tahmini. *Acta Infologica*, 5(2), 267-285.
<https://doi.org/10.26650/acin.934130>

1. GİRİŞ

İş analitiği, işletmelerin faaliyetlerinin izlemesine yardımcı olan ve gerçekçi kararlar verme süreçlerine katkı sağlanması amacıyla bilgi, istatistik, matematik ve bilgisayar tabanlı yöntemlerin kullanılmasıdır (Evans, 2016). Üretim, ar-ge, kalite, halkla ilişkiler, pazarlama ve finans gibi farklı departmanlarda, iş analitiği yöntemlerinden faydalanmak mümkündür (Charan, 2015). İş analitiği çalışmaları kapsamında, sosyal medyadan veri toplayıp, pazarlama veya finans ile ilgili konularda karar vermeye yardımcı olacak bilgilerin de üretilmesi mümkündür. Sosyal medya verilerinin toplanıp analiz edilmesi, iş analitiği çalışmalarının önemli bir alt dalı olarak sosyal medya analitiği şeklinde tanımlanmıştır (Holsapple, Hsiao, & Pakath, 2018).

Sosyal medya, bireylerin yeni nesil internet teknolojilerinin sağladığı iletişim hızı ve kolaylıkları ile istedikleri içeriği üretip yayınlamalarını sağlayan çevrimiçi bir platform ağıdır (Ulusoy, 2012). Günümüzde sosyal medya, bireylerin, işletmelerin ürün ve hizmetleri hakkında fikir paylaşımı yapabildiği ve diğer tüketicilerin bilgi edinmelerine imkân tanımaktadır (Murugesan, 2007). Bu tür içeriklerin sosyal medya platformlarında paylaşılması kişilere ve işletmelere birçok konuda fikir sahibi olmalarına olanak sağlamaktadır.

Sosyal medya analitiği çalışmaları ile sosyal medyadan elde edilen veriler doğrultusunda işletmeler satış, pazarlama, finansal vb. konularda kullanılmaktadır (Salo, 2017). Sosyal medya verileri, metin madenciliği yöntemleri analiz edilerek, faydalı bilgiler elde etmek mümkündür (Charan, 2015). Örneğin, potansiyel müşterilerin tespit edilmesi (Lee, 2018) (Syam & Sharma, 2018), marka imaj takibi yapılması, markayı destekleyici veya kötüleyici mesaj içeriklerini tespit edip, olumsuz görüşleri azaltmak için yeni stratejiler belirlenmesi (Özüpek & Diker, 2013) çalışmalarında sosyal medyadan yararlanılmaktadır. Finansal çalışmalarda, hisse senedi değeri tahmini gibi kararların verilmesinde de sosyal medyadan faydalanılabileceği iddia edilmektedir (Attigeri, M, Pai, & Nayak, 2015; Chen & Lazer, 2011; Vu, Chang, Quang, & Collier, 2012; Zhang, Fuehres, & Gloor, 2011). Bu kapsamda, hisse senedi değer tahmininin de sosyal medya verilerinin rolü araştırılmıştır. Ayrıca, diğer finansal göstergeler ve yatırım aracı verileri ile birlikte kullanıldığında sosyal medya verileri ile hisse senedi değeri tahminine etkisi analiz edilmiştir. Böyle bir tahminleme çalışması için kullanılabilir en uygun yöntem, en iyi tahmin algoritması sunulmuştur.

Bu çalışmada, öncelikle literatür taraması yapılmıştır. Tarama sonucu, sosyal medya verilerinin işletmelerin geliri ve hisse senedi değeri gibi finansal göstergeleri tahmin etmek amacıyla yapılan çeşitli makaleler tespit edilmiştir. Bu makaleler, literatür kısmında değerlendirilmiştir. Farklı ülkelerdeki çeşitli şirketlerin hisse senedi değerinin sosyal medya verilerine dayalı tahmin edilebileceğini gösteren çalışmaların olduğu görülmüştür (Attigeri vd., 2015; Chen & Lazer, 2011). Ancak, Türk şirketlerin hisse senedi değerleri ile sosyal medya verileri arasında korelasyon olduğunu belirtilen çalışmalara rağmen, hisse senedi değer tahmini yapabilen bir çalışmaya rastlanmamıştır (Eliaçık & Erdogan, 2015; Yıldırım & Yüksel, 2017). Ayrıca, Sosyal medya verileri ile hisse senedi değeri tahmini çalışmaları Telekom, bilişim gibi çeşitli sektörlerde yapılmış olsa da havayolu sektöründe bu tür bir çalışma tespit edilememiştir. Halbuki, hava yolu şirketleri küresel markalar olup, Twitter'ı etkin olarak kullanmaktadırlar. Bu nedenlerle, çalışma amacı doğrultusunda ve literatürdeki bu boşlukları gidermek için, havayolu sektörü ele alınmıştır. Türkiye'nin milli markası olan Türk Hava Yolları (THY) şirketi için sosyal medya verilerine dayalı hisse senedi değer tahmini çalışması yapılmıştır. Analizlerde karşılaştırmalar yapabilmek için, THY'ye ek olarak, 2019 Dünya Havayolu Firmaları listesinde (Skytrax, 2020) yer alan ilk 30 firma arasında yer alan ve Twitter etkinliği fazla olan dört farklı ülkenin uluslararası hava yolu firmaları da (LHY, QHY, FHY, SHY) çalışmaya dahil edilmiştir.

Belirlenen uluslararası havayolu firmaları hakkında, önemli bir sosyal medya platformu olan Twitter'daki paylaşımlar toplanıp, analizi edilmiştir. Toplanan tweetler ve diğer yatırım araçlarıyla ilgili göstergeler kullanılarak, bir gün sonraki hisse değerini tahmin edebilecek model geliştirmiştir. Çeşitli yatırım araçlarının hisse senedi fiyatları üzerine etkisi olduğu konusunda birçok çalışma yapılmıştır (Aktaş & Akdağ, 2013; Cingöz & Kendirli, 2019; Gençtürk, 2009). Ancak hem sosyal medya hem de farklı yatırım araçları veya finansal göstergeleri birlikte kullanan bir çalışma tespit edilememiştir. Tahminlemede, literatürdeki benzer çalışmalardan farklı olarak, tek bir algoritma değil, altı farklı makine öğrenmesi algoritması kullanılıp, algoritmaların başarı karşılaştırması da yapılmıştır. Böylece hisse senedi değer tahminlemesini en düşük hata oranı ile yapabilecek en uygun algoritmanın belirlenmesi hedeflenmiştir.

Çalışmanın birinci bölümünde konunun kapsamı ve içeriği ile hakkında bilgilere yer verilmiştir. Ardından literatürde yapılan benzer çalışmalardan ve içeriklerinden bahsedilmiştir. Üçüncü bölümde, çalışmada kullanılan veri seti ve yöntem

detaylandırılmıştır. Bu bölümde ayrıca tahmin modelinde kullanılan verilerin nasıl toplandığından, istatistiksel analizlerden ve kurulan tahmin modelinin güvenilirliğinden bahsedilmiştir. Dördüncü bölümde ise analizler sonucunda elde edilen bulgulara yer verilmiştir. Çalışmanın son bölümünde ise, elde edilen bulgulara dayalı olarak havayolu firmalarına ait hisse senetlerine yatırım kararlarıyla ilgili değerlendirme ve önerilerde bulunulmuştur.

2. LİTERATÜR

Günümüzde sosyal medya platformlarının kullanımı artması ile birlikte tüketiciler herhangi bir marka için görüşlerini, olumlu ve olumsuz deneyimlerini dile getirebilecekleri platformlardan biri olan Twitter popüler bir uygulama haline gelmiştir. Bu kapsamda Twitter verileri ile hisse senedi değeri arasındaki ilişkiyi araştıran çeşitli çalışmalara rastlanmıştır.

Bu kapsamda incelenen ilk çalışmada, Attigeri et al. (2015), işletmelerin hisse senedi değeri tahmininde sosyal medya verilerinin çok büyük etkisi olduğu tespit edilmiştir. Sosyal medyadan toplanan mesajların duygu analizleri sonucu elde edilen veriler ile işletmelerin hisse senedi değeri arasındaki korelasyon tespit edilmiştir. Bu yöntem ile sosyal medyadan toplanan mesajların duygu durumu ile hisse senedi değerinin tahmin edilebileceği ve sosyal medya verilerinin finansal değerlerin tahmini ile yakından ilişkili olduğu söylenebilmektedir. Asur ve Huberman (2010) ise, Twitter verilerine dayalı sinema sektöründeki gişe gelirleri tahmin edilmiştir. Verileri makine öğrenmesi tekniği ile lojistik regresyon modeli kurulmuştur. Elde edilen model, seçmiş oldukları filmlerle ilgili Twitter platformundaki anılma sayıları ile gişe gelirleri arasında güçlü korelasyon olduğunu tespit edilmiştir. Diğer bir çalışmada Chen ve Lazer (2011), Twitter verileri ile hisse senedi değerinin hareket yönü arasındaki ilişki incelemiştir. Duygu analizi yapılan veriler doğrusal regresyon kullanılarak modelleme yapılmıştır. Hisse senedinin gelecekteki fiyat değişikliklerini tahmin edebilmek için model oluşturmuşlardır. Oluşturdukları modelin başarılı olduğunu ifade edilmiştir.

Sosyal medya verilerine dayalı hisse senedi değer tahminiyle ilgili önemli bir çalışma Vu et al. (2012) tarafından gerçekleştirilmiştir. Tüketicilerin duygu durumları ile hisse değerleri arasındaki ilişki incelenmiştir. Twitter'dan elde edilen veriler makine öğrenmesi metodu ile karar ağacı algoritması kullanılarak tahmin modeli oluşturulmuştur. Bu kapsamda Nasdaq borsasında işlem gören dört teknoloji firmasına ait hisse değerlerini ele alınmıştır. Sosyal medya verilerinin günlük olarak artacağı veya azalacağını tahmin edecek model Apple (AAPL) için %92,93, Google (GOOG) için %80,49, Microsoft (MSFT) için %75,61 ve Amazon (AMZN) için ise %75 korelasyon olduğunu ortaya konulmuştur.

Borsa endeks tahminleri için de sosyal medyadan faydalanılmıştır. Zhang et al. (2011), Twitter platformundaki paylaşımları analiz ederek Dow Jones, Nasdaq, S&P 500 borsa göstergelerini önceden tahmin edebilmiştir. Twitter verilerinin duygu durumları yüzdesinin Dow Jones, Nasdaq ve S&P 500 endeksleri ile negatif yönlü korelasyon, VIX endeksi ile pozitif yönlü korelasyon olduğu sonucuna ulaşılmıştır. Bollen et al. (2011) tarafından, Twitter platformundaki pozitif ve negatif yorumların borsa piyasalarını nasıl etkilediği üzerine çalışma yapılmıştır. Bu çalışmada Granger nedensellik analizi ve Kendi Kendini Düzenleyen Bulanık Sinir Ağı kullanılarak Twitter'daki paylaşımların duygu durumları ile Dow Jones Industrial Average (DJIA) arasında %87,6 oranında yüksek korelasyon olduğunu saptanmıştır. Benzer bir çalışmada, Mittal ve Goel (2012), Kendi Kendini Düzenleyen Bulanık Sinir Ağı kullanılarak Twitter platformundaki paylaşımların duygu durumları ile bir sonraki günlük DJIA değeri ile %75,56 korelasyon tespit edilmiştir.

Bir diğer çalışmada ise Porshnev et al. (2013), Twitter verilerine dayalı borsa göstergeleri tahminlerinin doğruluğunu artırma olasılığı üzerine çalışma yürütülmüştür. Bu kapsamda 755 milyondan fazla Twitter mesajının duygu analizlerini yapılmıştır. Destek Vektör Makinesi ve Sinir Ağları algoritması kullanılarak DJIA hisse değerini destek vektör makine algoritması modelini kullanarak %64,10 doğruluk oranı ile tahmin edilmiştir. S&P 100 endeksinin tahmini için yapılan bir diğer çalışmada, Si et al. (2013), Twitter platformundan elde ettikleri verileri kullanmışlardır. S&P 100 endeksinin tahmini için kullandıkları Dirichlet Proses Karışım modelinin etkili performans gösterdiği ve faydalı olduğundan bahsedilmiştir.

Literatüre farklı bir yaklaşım sunan Bhardwaj et al. (2015), yatırımcıların hisse senedi piyasasına yatırım yapmadan önce çeşitli yöntemlerle tahminde bulunmalarının öneminden bahsedilmiştir. Bu amaçla Twitter platformundan Python programı ile elde ettikleri verileri Sensex ve Nifty hisse senetlerine ait değerleri tahmin etmek için doğal dil işleme yöntemleriyle duygu analizi yapılmıştır. Sonuç olarak yatırımcılara öngörülerinde yardımcı olacak faydalı bir tahmin modeli sunulmuştur.

Dickinson ve Hu (2015), Twitter platformundan topladıkları verileri duygu analizi yaparak sınıflandırmıştır. Elde ettikleri bu verileri hisse senedi tahmini için istatistiksel analizler yapılmıştır. Elde ettikleri sonuca göre Walmart ve Microsoft firmaları için pozitif yönlü yüksek korelasyon, Goldman Sachs ve Cisco firmaları için negatif yönlü yüksek korelasyon olduğunu tespit edilmiştir.

Yıldırım ve Yüksel (2017), Türkiye’de telekomünikasyon sektöründe hizmet veren bir firma için Twitter platformunda elde ettikleri verileri duygu durumlarına göre sınıflandırılmıştır. Bu verileri ile firmanın hisse senedi değerini tahmin edebilmek için istatistiksel analizler yapılmıştır. Analizler sonucunda Naive Bayes algoritması ile sınıflandırma gerçekleştirilmiştir. Bu yöntem ertesi günkü hisse değerinin hareket yönü tahmini için kullanılmıştır. Sonuç olarak pozitif tweet sayısının hisse senedinin ertesi günkü hareket yönü ile negatif yönlü orta düzeyde korelasyon olduğu saptanmıştır.

Das et al. (2017), hisse senedi tahmini için, Twitter platformundan Twitter API ve Apache Flume yazılımı kullanılarak geçmiş zamanlı verileri elde edilmiştir. Tweetlerin duygu analizi sonuçları ile zaman serisi analizi yapılmıştır. Bununla birlikte Tekrarlayan Sinir Ağı algoritması da kullanılmıştır. Kurdukları hibrit modelde duygu analizlerinin ertesi günkü hisse değerinin artacağı veya azalacağı tahminleri güvenilir sonuçlar verdiği elde edilmiştir.

Son olarak Yadav et al. (2019), anlık haber verilerinin duygu analizlerine dayalı vadeli işlem piyasalarında alış satış davranışları hakkında tahmin edilebilirliği araştırılmıştır. Duygu analizi ve metin sınıflandırma probleminin çözümü için, saf Bayes sınıflandırıcıları ve destek vektör makinesi kullanılmıştır. Piyasanın duygu durumu, alış satış davranışları ve net alım baskısı (NBP: Net Buying Pressure) verileri kullanılarak fiyat oluşumunu arz-talep aşamasında gecikmeli olarak etkilediğini saptanmıştır.

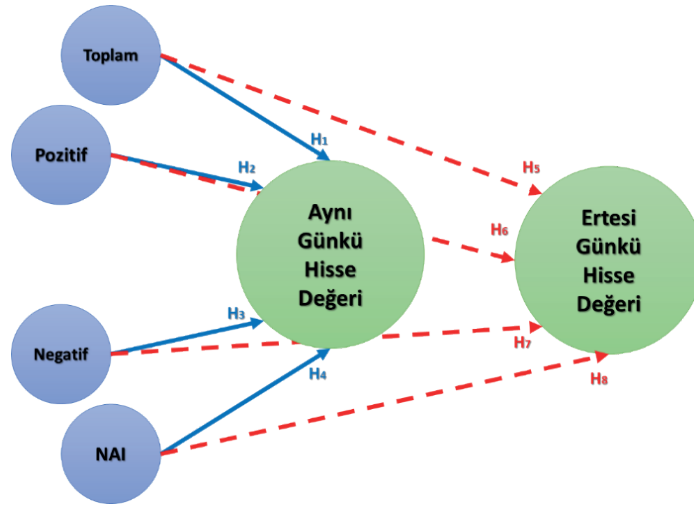
İncelenen literatür kaynaklarına göre sosyal medya verilerine dayalı finansal verilerin tahmini konusunda çeşitli çalışmalara rastlanmıştır. Bu kapsamda Türkiye’de yapılan Eliaçık ve Erdoğan’ ın (2015) çalışmalarında istatistiksel analizler sonucunda sosyal medya paylaşımlarının duygu durumları ile BİST-100 endeks değeri arasında pozitif yönlü orta düzeyde korelasyon tespit edilmiştir. Yine benzer bir çalışmada Yıldırım ve Yüksel’ in (2017) yaptıkları çalışmada telekomünikasyon firması hakkında Twitter platformundan elde edilen verilere dayalı ertesi günkü hisse değerinin hareket yönünü orta düzeyde korelasyon ile tahmin edilmiştir. Ancak, bu çalışmalarda hisse senedi değerinin ertesi günlerde ne olacağı konusunda tahmin modeli oluşturulmamıştır.

Literatürde tespit edilen çalışmalarda, sosyal medya (örneğin Twitter) verilerinin hisse senedi değerini tahmin etmedeki faktörü araştırılmıştır (Bhardwaj vd., 2015; S. Das & Behera, 2018; Dickinson & Hu, 2015; Yıldırım & Yüksel, 2017). Ayrıca, diğer yatırım aracı verilerine dayalı hisse senedi fiyatları üzerine etkisi konusunda da çalışmalar yapılmıştır (Aktaş & Akdağ, 2013; Cingöz & Kendirli, 2019; Gençtürk, 2009). Ancak hem diğer yatırım aracı verileri hem de sosyal medya verilerinin birlikte kullanılmasıyla hisse senedi değerinin tahmin edilebilirliğinin ele alındığı bir çalışmaya rastlanmamıştır. Bu çalışma, literatürdeki bu boşluğu gidermek için, Twitter sosyal medya verileri ile birlikte diğer yatırım aracı verileri hisse senedi değeri tahmininde birlikte kullanılmıştır. Bu faktörlerden hangisinin hisse senedi üzerinde daha etkili olduğu da araştırılmıştır. Bununla birlikte, literatürde genel de tek bir tahmin algoritması ile çalışmalar yapılırken, bu çalışmada altı farklı makine öğrenmesi tabanlı tahmin algoritması kullanılmıştır.

3. VERİSETİ VE YÖNTEM

Bu bölümde, çalışma sürecinde uluslararası havayolu firmalarına ait sosyal medya verilerine dayalı hisse senedi değeri tahmini için kullanılan metod ve analizlerden bahsedilmiştir. İlk olarak müşteriler tarafından oylanan ve Skytrax firması tarafından yayımlanan 2019 Dünya Havayolu Firmaları listesinde (Skytrax, 2020) yer alan ilk 30 firma arasından beş firma (THY, LHY, QHY, FHY, SHY) seçilmiştir. Çalışma kapsamında bu firmaların seçilme nedeni, en büyük firmalar arasında yer almaları ve Twitter etkinliği fazla olmasıdır. İncelenen firmalar ile ilgili ertesi hisse senedi değerlerini Twitter verileri ile birlikte diğer yatırım aracı verileri de dahil edilerek tahmin modeli oluşturulmuştur. Diğer yatırım aracı verileri, THY için BİST-100 endeks değeri (investing, 2020b), döviz olarak Dolar/TL endeksi (investing, 2020f) ve Altın(gr-TL) endeksi (investing, 2020a) verileri kullanılmıştır. SHY için ise firmaya ait hisse değeri SIAL’ ın (investing, 2020d) işlem gördüğü Singapur Borsası (SGXL) (investing, 2020e), döviz olarak ise Amerikan dolarının Singapur dolarındaki karşılığı USD/SGD

değeri (investing, 2020g) verileri kullanılmıştır. Bu çalışmada literatürde daha önce rastlanılmamış sosyal medya verileri ile birlikte diğer yatırım aracı verilerine dayalı hisse senedi değeri tahmini için Gradyan Destekli Ağaçlar (Gradient Boosted Trees - GBT) (Dolapçı, 2020) modeli kurulmuştur.



Şekil 1. Hipotez tasarımı

Şekil 1'de gösterildiği gibi araştırma kapsamında kurulan hipotezler:

Twitter verileri ile hisse senedi değerleri arasındaki ilişkinin korelasyon analizleri ile test edilmesi amacıyla, aşağıdaki hipotez oluşturulmuştur:

H_0 : (Pozitif/Negatif/Toplam/NAI) Tweet Sayıları ile (Aynı Günkü/Ertesi Günkü) hisse senedi değerleri arasında bir ilişki yoktur.

H_a : (Pozitif/Negatif/Toplam/NAI) Tweet Sayıları ile (Aynı Günkü/Ertesi Günkü) hisse senedi değerleri arasında bir ilişki vardır.

Belirtilen ana hipotezin alt hipotezlerinin ilgili oldukları değişkenler aşağıda listelenmiştir:

- H_1 : Toplam tweet sayısı – Aynı günlük hisse senedi değeri,
- H_2 : Pozitif tweet sayısı – Aynı günlük hisse senedi değeri,
- H_3 : Negatif tweet sayısı – Aynı günlük hisse senedi değeri,
- H_4 : Pozitif-Negatif farkı (NAI) tweet sayısı – Aynı günlük hisse senedi değeri,
- H_5 : Toplam tweet sayısı – Ertesi günlük hisse senedi değeri,
- H_6 : Pozitif tweet sayısı – Ertesi günlük hisse senedi değeri,
- H_7 : Negatif tweet sayısı – Ertesi günlük hisse senedi değeri,
- H_8 : Pozitif-Negatif farkı (NAI) tweet sayısı – Ertesi günlük hisse senedi değeri

Yukarıda bahsedilen parametreler arasındaki korelasyon IBM SPSS Statistics programı ile analiz edilmiştir. Kurulan hipotezleri test etmek amacıyla çalışma süreci bu bölümde anlatılmıştır.

3.1. Veri Toplama

Havayolu firmalarına ait hisse senedi değeri verileri ve diğer yatırım aracı verileri günlük olarak investing.com (2020) sitesinden toplanmıştır. Investing sitesinden ilgili firmalara ait hisse senedi değeri verileri ve diğer diğer yatırım aracına ait

veriler günlük olarak Excel dosyasında kaydedildi. Örnek olarak THY firmasına ait hisse senedi değeri verileri Tablo 1’de paylaşılmıştır.

Tablo 1

THY’ye ait hisse değeri verileri

Tarih	Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri
01.10.2019	11,780	11,840
02.10.2019	11,840	11,750
03.10.2019	11,750	11,870
.	.	.
.	.	.
.	.	.
04.02.2020	14,390	14,610
05.02.2020	14,610	14,740
06.02.2020	14,740	14,740

Hisse senedi verileri elde edildikten sonra, sosyal medyadan veri toplamak amacıyla yaygın olarak kullanılan 3 araç tespit edilmiştir. Bunlar, Python (Pagolu & Majhi, 2016), R Programlama dili (Arun, Srinagesh, & Ramesh, 2017) ve RapidMiner (Gazioğlu & Şeker, 2017) (Tripathi, Vishwakarma, & Lala, 2015) programlarıdır. RapidMiner metin madenciliği, duygu analizi ve makine öğrenmesine dayalı farklı tahmin araçlarını, kolay ve hızlı kullanım imkânı sunan bir veri madenciliği yazılımıdır. Bu nitelikleri nedeniyle, sosyal medya verilerinin toplanıp analiz edilmesi için, RapidMiner yazılımı, genel amaçlı R ve Python programlama dillerine tercih edilmiştir.

Çalışma kapsamında Twitter platformundan veri toplarken kullanılan anahtar kelime listesi ve indirilen tweetlerin dilleri Tablo 2’de gösterilmiştir.

Tablo 2

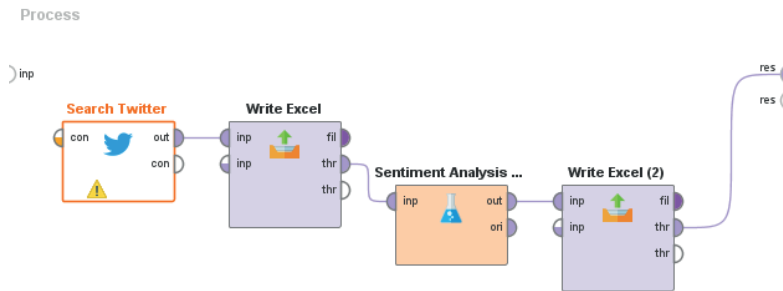
THY’ye ait hisse değeri verileri

Firma Adı	Anahtar Kelime	Dil/ler
Türk Hava Yolları	“@TurkishAirlines”	İngilizce
Singapur Hava Yolları	“@SingaporeAir”	İngilizce
Fransız Hava Yolları	“@airfrance”	İngilizce-Fransızca
Qantas Hava Yolları	“@qantas”	İngilizce
Lufthansa Hava Yolları	“@lufthansa”	İngilizce

Twitter’ın 15 dakika’da 10,000 Tweet limiti olmasına rağmen, ele alınan şirketler hakkında günlük tweet sayısı bile çoğu zaman 10,000 adeti geçmediği için ek bir veri toplama yönetimine ihtiyaç duyulmamıştır.

3.2. Rapid Miner Veri Toplama ve Analiz Modeli

Rapid Miner programı, veri madenciliği, metin madenciliği, makine öğrenmesi gibi özelliklere sahip olan Java tabanlı bir programdır (Gazioğlu & Şeker, 2017). Program içerisinde eklenti olarak kullanılan Search Twitter operatörü ve Meaning Cloud operatörü yardımıyla Twitter platformunda veriler indirilip duygu analizleri yapıldıktan sonra Excel dosyasına kaydedilmesi için Şekil 2’deki model kurulmuştur.



Şekil 2. RapidMiner veri indirme modeli

Şekil 2’de gösterilen modelde, Search Twitter operatörü ile Twitter platformuna bağlanması için gerekli tanımlamalar yapılmıştır. Ardından indirilmek istenen firmaya ait anahtar kelime, tweetlerin dili, indirilmesi istenilen adet gibi bilgiler tanımlanmıştır. Sentiment Analysis operatöründe ise Meaning Cloud sitesinde üyelik yapıldıktan sonra elde edilecek olan lisans anahtarı tanımlanır ve hangi dilde duygu analizi yapılmak isteniyorsa gerekli parametre ayarları yapılmıştır. Write Excel operatörleri ise yapılan her işlem sonucunda elde edilen verileri Excel dosyasına yazdırmak için kullanılmıştır. Twitter platformundan indirilen verilerin örneği Tablo 3’te gösterilmiştir.

Tweetlerin duygu analizi için Meaning Cloud (2021) metin madenciliği aracı kullanılmıştır. Meaning Cloud Duygu analizi (2021) aracı ile herhangi bir metindeki olumlu, olumsuz, tarafsız duygu kutuplarını tanımlamak mümkündür. Bunu yapmak için, metindeki farklı cümlelerin yerel kutupluluğu belirlenir ve aralarındaki ilişki değerlendirilir, sonuçta tüm metin için küresel bir kutupluluk değeri elde edilir.

Tablo 3

Excel dosyasına kaydedilen tweetlerin örneği

Oluşturma Tarihi	Kullanıcı	Dil	Kaynak	Tweet
2019-10-01 00:00:29	गौतम	en	Twitter Web App	RT @DrGPradhan: It is time to go offensive We all, who travel outside India should avoid taking @TurkishAirlines Turkey is a nation whic...
2019-10-01 00:00:46	Aimy	en	Twitter for Android	RT @iamShaniera: I request more people to use @TurkishAirlines as their preferred means of air travel. Not only is Turkey a beautiful count...
2019-10-01 00:00:49	Her.Amy.Dobby 💎	en	Twitter for Android	RT @DrGPradhan: It is time to go offensive We all, who travel outside India should avoid taking @TurkishAirlines Turkey is a nation whic...
2020-02-06 08:39:48	Harsha Vardhan Achari	en	Twitter for Android	RT @TurkishAirlines: 🌍 Earth is here, waiting to be reached! #StepOnEarth with the airline that flies to the most countries on the planet....
2020-02-06 08:41:52	잠수함	en	Twitter for Android	RT @TurkishAirlines: 🌍 Earth is here, waiting to be reached! #StepOnEarth with the airline that flies to the most countries on the planet....
2020-02-06 08:42:33	CAL??_fans	en	Twitter for Android	Coming soon! @TurkishAirlines First A350XWB is TC-LPA ??? https://t.co/q7IRqeuYLP

Meaning Cloud eklentisi ile her dilde duygu analizi yapılmasına izin verilmediği için tüm firmalar hakkında ortak dil olarak İngilizce, FHY firması için ise ek olarak Fransızca dilinde tweetler analize dahil edilmiştir.

4. BULGULAR

4.1. Tanımlayıcı İstatistikler

Araştırma kapsamında firmalara ait günlük tweet sayıları, diğer yatırım aracına ait veriler ve hisse senedi değeri verileri aşağıda gösterilmiştir.

Tablo 4

THY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NAI	Dolar	ALTIN/gr	BİST-100	Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri
1.10.2019	6152	1904	2232	997	1019	-328	5,7231	272,479	103876,56	11,780	11,840
2.10.2019	475	148	179	79	69	-31	5,7020	275,102	103509,22	11,840	11,750
3.10.2019	962	157	107	46	652	50	5,6947	276,232	102784,33	11,750	11,870
.
4.2.2020	217	106	60	10	41	46	5,9801	298,803	122136,17	14,390	14,610
5.2.2020	364	258	47	13	46	211	5,9827	299,297	122320,77	14,610	14,740
6.2.2020	236	179	21	9	27	158	5,9874	301,471	122283,87	14,740	14,740

Tablo 4'te gösterilen THY'ye ait toplam 56 064 adet tweet Excel tablosunda analiz edilmek üzere düzenlenmiştir.

SHY, QHY, LHY ve FHY firmaları için sırasıyla 32 232, 79 863, 42 837 ve 25 768 adet tweetlerin duygu analizler yapılmış ve EK-1 Tablo 1-4'te gösterilmiştir. Tüm firmalar için bağımlı değişken Ertesi Günkü Hisse Değeri parametresi iken, diğer tüm parametreler ise bağımsız değişken olarak tanımlanmıştır. Tanımlanan parametreler için korelasyon analizleri yapılmıştır.

4.2. Korelasyon Analizleri

Bu bölümdeki analizlerde bağımlı değişkenlerin, bağımsız değişken ile arasındaki ilişki incelenecektir. Ayrıca veriler ait istatistiksel analiz sonuçları paylaşılacaktır.

THY, SHY, LHY, QHY ve FHY firmalarına ait normallik testi sonuçlarına göre verilerin normal dağılım göstermediği ve bundan dolayı parametrik olmadığı tespit edilmiştir. Bu yüzden korelasyon analizi değişkenler arasındaki korelasyon incelenirken Spearman's rho değerleri dikkate alınmıştır. Firmalara ait korelasyon analizi sonuçları EK-2 Tablo 1-5'de gösterilmiştir. Bu sonuçlar incelendiğinde korelasyon değerleri yüksek olan SHY ve THY firmaları için tahmin modeli kurulmak istenmiştir.

4.3. RapidMiner ile Tahmin Modeli Bulguları

THY ve SHY firmalarına ait sosyal medya ve diğer yatırım aracı (THY için Dolar, Altın, BİST-100 değerleri, SHY için SGXL/SGD, USD/SGD değerleri) verileri hisse senetlerinin değerlerinin tahmini için kullanılmıştır. Tahmin çalışmaları RapidMiner veri madenciliği programıyla, programın bize sunduğu altı farklı tahmin modeli kullanılarak gerçekleştirilmiştir.

Gradyan Destekli Ağaçlar modeli ile tahmin edilen ertesi günkü hisse değeri verilerinin hata analizleri yapılmıştır. Öncelikle, tahmin edilen değerler $x(t)$, gerçekleşen değerler $f(t)$, arasındaki sapma değeri $e(t)$ aşağıdaki formülde gösterilmiştir (Karahan, 2011).

$$e(t) = x(t) - f(t)$$

Sonrasında, tahmin edilen değerlerin hata yüzdesi $p(t)$, aşağıdaki formül kullanılarak hesaplanmıştır.

$$p(t) = \frac{e(t)}{x(t)}$$

Son olarak, tahmin modelinin ertesi günkü hisse değeri tahmin sonuçlarının tutarlılığını ölçmede yaygın olarak kullanılan hata ölçütlerinden (Karahan, 2011); Ortalama Hata Kareleri (Mean Squared Error-MSE) ve Ortalama Mutlak Yüzde Hata (Mean Absolute Percent Error-MAPE) değerleri hesaplanmıştır.

$$MSE = \frac{1}{n} \sum [e(t)]^2$$

$$MAPE = \frac{1}{n} \sum |p(t)|$$

İki firmaya ait verilerden 100 günlük veri makine öğrenmesinde kullanılmıştır. Öğrenme verileri kullanılarak yapılan analizler sonucunda THY ve SHY firmasına ait tahmin modeli sonuçları Tablo 5'te gösterilmiştir.

Tablo 5

THY ve SHY için tahmin modeli analiz sonuçları

Tahmin Modeli	THY		SHY	
	Korelasyon	Mutlak Hata	Korelasyon	Mutlak Hata
Genelleştirilmiş Doğrusal Model (Generalized Linear Model)	0,970	0,219	0,825	0,051
Derin Öğrenme (Deep Learning)	0,929	0,347	0,743	0,077
Karar Ağacı (Decision Tree)	0,970	0,211	0,759	0,059
Rassal Orman (Random Forest)	0,985	0,197	0,810	0,057
Gradyan Destekli Ağaçlar* (Gradient Boosted Trees)	0,985	0,156	0,827	0,057
Destek Vektör Makinesi (Support Vector Machine)	0,957	0,234	0,753	0,077
* En iyi çözüm				

Öğrenme verileri kullanılarak yapılan analizler sonucunda THY ve SHY firmaları için tahmin modellerinden en iyi çözüm Gradyan Destekli Ağaçlar modeli tespit edilmiştir. Gradyan Destekli Ağaçlar Modeli (Rapid Miner, 2019), regresyon veya sınıflandırma ağacı modellerinin bir bütünüdür. Her ikisi de kademeli olarak geliştirilmiş tahminler yolu ile öngörücü sonuçlar elde eden ileri öğrenme yöntemleridir.

Analizler sonucunda kullanılacak olan tahmin modeli tespit edildikten sonra THY ve SHY firmaları için tahmin edilmesi istenilen değişken verileri kullanılarak hisse senedi değerleri tahmin edilmiştir. THY firması için 18 günlük, SHY firması için 14 günlük veri kullanılmıştır.

THY için Gradyan Destekli Ağaçlar tahmin modeli ile elde edilen 18 günlük tahminlerin hata testlerine ait sonuçlar Tablo 6'te gösterilmiştir.

Tablo 6

THY için tahminin hata testleri sonucu

Tarih	Tahmin Edilen Ertesi Günkü Hisse Değeri (f)	Gerçekleşen Ertesi Günkü Hisse Değeri (x)	Sapma (e)	Yüzde Hata (p)	Mutlak Yüzde Hata p
2020-01-15	14,636	15,080	0,444	0,0294	0,029
2020-01-20	14,244	14,880	0,636	0,0428	0,043
2020-01-21	14,286	14,720	0,434	0,0295	0,029
2020-01-22	14,242	14,240	-0,002	-0,0001	0,000
2020-01-23	14,271	14,540	0,269	0,0185	0,019
2020-01-24	14,205	14,540	0,335	0,0230	0,023
2020-01-25	14,379	14,540	0,161	0,0111	0,011
2020-01-26	14,424	14,100	-0,324	-0,0230	0,023
2020-01-27	14,250	14,080	-0,170	-0,0121	0,012
2020-01-29	14,216	13,830	-0,386	-0,0279	0,028
2020-01-30	14,279	13,680	-0,599	-0,0438	0,044
2020-01-31	14,255	13,680	-0,575	-0,0420	0,042
2020-02-01	14,402	13,680	-0,722	-0,0528	0,053
2020-02-02	14,406	13,840	-0,566	-0,0409	0,041
2020-02-03	14,252	14,390	0,138	0,0096	0,010
2020-02-04	14,287	14,610	0,323	0,0221	0,022
2020-02-05	14,285	14,740	0,455	0,0309	0,031
2020-02-06	14,419	14,470	0,051	0,0035	0,004

	MSE	MAPE
Tahmin Modeli	0,175	0,026

Tablo 7’te verilen tabloda tahmin sonuçlarının performansını ölçmede yaygın olarak kullanılan hata ölçütlerinden “Ortalama Hata Kareleri (MSE)” değerine göre iyi performans gösterdiği söylenebilir. Yine tahmin modelinin tutarlılığını ölçmede kullanılan hata ölçütlerinden “Ortalama Mutlak Yüzde Hata (MAPE)” değeri tahmin modelinin %2,6 hata oranı ile başarılı olduğu söylenebilir.

Ayrıca SHY firması için Gradyan Destekli Ağaçlar tahmin modeli tarafından 14 günlük tahminlerin hata analizleri Tablo 8’de gösterilmiştir.

Tarih	Tahmin Edilen Ertesi Günlük Hisse Değeri (f)	Gerçekleşen Ertesi Günlük Hisse Değeri (x)	Sapma (e)	Yüzde Hata (p)	Mutlak Yüzde Hata p
2020-01-22	9,1116	8,8500	0,2616	0,0287	0,0287
2020-01-23	9,1228	8,8200	0,3028	0,0332	0,0332
2020-01-24	9,0996	8,8200	0,2796	0,0307	0,0307
2020-01-25	9,1030	8,8200	0,2830	0,0311	0,0311
2020-01-26	9,0972	8,8200	0,2772	0,0305	0,0305
2020-01-27	9,1029	8,5700	0,5329	0,0585	0,0585
2020-01-29	9,1071	8,5400	0,5671	0,0623	0,0623
2020-01-30	9,1752	8,5500	0,6252	0,0681	0,0681
2020-01-31	9,1351	8,5500	0,5851	0,0641	0,0641
2020-02-01	9,1330	8,5500	0,5830	0,0638	0,0638
2020-02-02	9,0948	8,4900	0,6048	0,0665	0,0665
2020-02-03	9,1093	8,5500	0,5593	0,0614	0,0614
2020-02-04	9,0745	8,5900	0,4845	0,0534	0,0534
2020-02-05	9,0994	8,5800	0,5194	0,0571	0,0571

	MSE	MAPE
Tahmin Modeli	0,233	0,051

Tablo 9’da verilen tabloda tahmin sonuçlarının performansı için “Ortalama Hata Kareleri (MSE)” değerine göre iyi performans gösterdiği söylenebilir. Tahmin modelinin tutarlılığını ölçmede kullanılan hata ölçütlerinden “Ortalama Mutlak Yüzde Hata (MAPE)” değeri tahmin modelinin %5,1 hata oranı ile başarılı olduğu söylenebilir.

4.4. Hipotez Testi Bulguları

Firmalar için toplanan verilere ilk olarak korelasyon analizleri yapılmıştır. 5 hava yolu firmasına ait toplam 40 adet korelasyon sonuç tablosu Tablo 10’de gösterilmiştir.

Tablo 10

Korelasyon analizi sonuçları

Hipotez No:	Türk Hava Yolları	Singapore Hava Yolları	Lufthansa Hava Yolları	Qantas Hava Yolları	France Hava Yolları
1	-0,267**	0,318**	-0,197*		-0,185*
2	-0,284**	0,546**			-0,339**
3	-0,303**				
4		0,600**			-0,216*
5	-0,280**	0,334**	-0,181*		
6	-0,296**	0,545**			-0,245**
7	-0,296**				
8		0,611**			

** Korelasyon 0,01 düzeyinde anlamlıdır.

* Korelasyon 0,05 düzeyinde anlamlıdır.

Firmalar için kurulan hipotezler, korelasyon analizi sonuçlarına göre değerlendirilmiştir. Sırasıyla hipotezleri inceleyecek olursak;

1. Hipotez 'de Toplam tweet sayısı ile aynı günlük hisse değeri arasındaki ilişki incelenmiştir. THY, SHY, LHY ve FHY firmaları için anlamlı sonuç elde edilmiştir. THY, LHY ve FHY firmalarında negatif yönlü düşük düzeyde korelasyon tespit edilirken, SHY firmasında pozitif yönlü orta düzeyde korelasyon tespit edilmiştir.

2. Hipotez 'de Pozitif tweet sayısı ile aynı günlük hisse değeri arasındaki ilişki incelenmiştir. THY, SHY ve FHY firmaları için anlamlı sonuç elde edilmiştir. THY firmasında negatif yönlü düşük düzeyde korelasyon, FHY firmasında negatif yönlü orta düzeyde korelasyon tespit edilirken, SHY firmasında pozitif yönlü orta düzeyde korelasyon tespit edilmiştir.

3. Hipotez 'de Negatif tweet sayısı ile aynı günlük hisse değeri arasındaki ilişki incelenmiştir. THY firmasında negatif yönlü orta düzeyde korelasyon tespit edilmiştir.

4. Hipotez 'de NAI (Pozitif-Negatif farkı tweet sayısı) ile aynı günlük hisse değeri arasındaki ilişki incelenmiştir. SHY firmasında pozitif yönlü orta düzeyde korelasyon tespit edilirken, FHY firmasında negatif yönlü orta düzeyde korelasyon tespit edilmiştir.

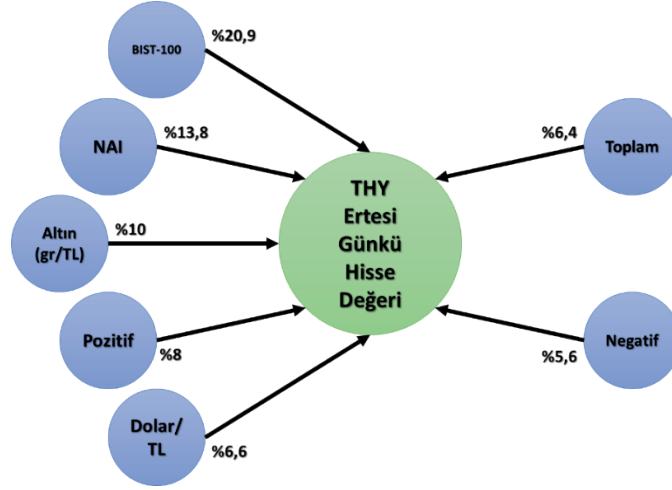
5. Hipotez 'de Toplam tweet sayısı ile ertesi günlük hisse değeri arasındaki ilişki incelenmiştir. THY ve LHY firmaları için negatif yönlü düşük düzeyde korelasyon tespit edilirken, SHY firmasında pozitif yönlü orta düzeyde korelasyon tespit edilmiştir.

6. Hipotez 'de Pozitif tweet sayısı ile ertesi günlük hisse değeri arasındaki ilişki incelenmiştir. THY ve FHY firmaları için negatif yönlü düşük düzeyde korelasyon tespit edilirken, SHY firmasında pozitif yönlü orta düzeyde korelasyon tespit edilmiştir.

7. Hipotez 'de Negatif tweet sayısı ile ertesi günlük hisse değeri arasındaki ilişki incelenmiştir. THY firmasında negatif yönlü orta düzeyde korelasyon tespit edilmiştir.

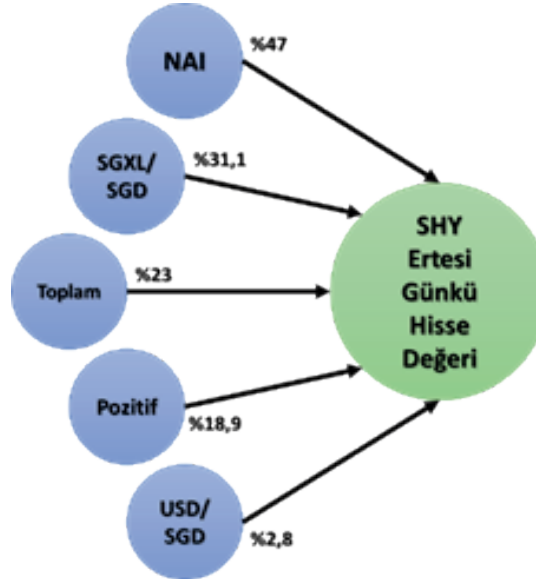
8. Hipotez 'de NAI (Pozitif-Negatif farkı tweet sayısı) ile ertesi günlük hisse değeri arasındaki ilişki incelenmiştir. SHY firmasında pozitif yönlü orta düzeyde korelasyon tespit edilmiştir.

Hipotezler test edilmesi için 40 adet korelasyon analizi incelenmiş ve bunlardan 22 tane anlamlı ilişki tespit edilmiştir. Korelasyon tespit edilemeyen analizler için günler ve haftalara göre tweet sayılarında yüksek farklılıklar olmasından ve duygu analizlerinin kısıtlı diller için yapılabilmesinden kaynaklanmaktadır. Bu sonuçlara dayanarak THY, LHY ve FHY firmalarına ait tweet sayıları azaldığında, hisse senedi değeri artacağı, SHY firmasına ait tweet sayısı arttığında ise hisse senedi değeri artacağı söylenebilir.



Şekil 3. THY değişkenlerinin ağırlık oranları

THY firması için Gradyan Destekli Ağaçlar tahmin modelinde %98,5 korelasyon ve %1,5 mutlak hata olduğu tespit edilmiştir. Tahmin edilecek günlere ait hisse senedi değeri ise %2,6 hata oranı ile tahmin modeli başarılı sonuç üretmiştir. Tahmin modelinde değişkenlerin ağırlıklı olarak hisse değerini ne düzeyde etkilediği Şekil 3'de gösterilmiştir. Görüleceği üzere sosyal medya verilerinden Net Destekçi Skoru(NAI) değişkeni Hisse değerinin tahmin edilmesinde en etkili 2. Faktör olduğu sonucuna varılmıştır.



Şekil 4. SHY değişkenlerinin ağırlık oranları

SHY firması için Gradyan Destekli Ağaçlar tahmin modelinde %82,7 korelasyon ve %5,7 mutlak hata olduğu tespit edilmiştir. Tahmin edilecek günlere ait hisse senedi değeri ise %5,1 hata oranı ile tahmin modeli başarılı sonuç üretmiştir. Tahmin modelinde değişkenlerin ağırlıklı olarak hisse değerini ne düzeyde etkilediği Şekil 4'de gösterilmiştir. Görüleceği üzere sosyal medya verilerinden Net Destekçi Skoru(NAI) değişkeni Hisse değerinin tahmin edilmesinde en etkili faktör olduğu sonucuna varılmıştır.

5. SONUÇ VE TARTIŞMA

Çalışmada üç farklı araştırma sorusu ele alınmıştır. Öncelikle sosyal medya verilerinin hava yolu sektöründe de diğer bazı sektörlerde olduğu gibi hisse senedi değerinin tahmin edilmesinde kullanılıp kullanılmayacağını araştırılması yapılmıştır.

Elde edilen bulgular havayolu sektöründe de literatürdeki çalışmalara benzer şekilde sosyal medya verilerinin hisse senedi değerini tahmininde kullanılabilir önemli bir veri kaynağı olduğunu göstermektedir. Ayrıca, hisse senedi değerini tahmin etmede diğer yatırım aracı verileri ile birlikte sosyal medya verilerinin kullanılabilir olduğu tespit edilmiştir. Ve sosyal medya verileri kullanılarak hesaplanan net destekçi skorunun tahminlemede önemli bir rolü olduğu görülmüştür.

Bu çalışmanın bir diğer amacı ise gerek diğer yatırım araçları verileri gerek sosyal medyadan elde edilen verilerle hisse senedi değeri arasında ilişki kurmak için faydalanabilecek makine öğrenmesi algoritmalarından en başarılısının hangisinin olabileceğini tespit etmektir. Bu amaçla çalışmada yapılan karşılaştırmalı analizler sonucunda elde edilen bilgilere göre Gradyan Destekli Ağaçlar algoritması diğer algoritmalara göre daha yüksek bir başarı elde edilmiştir.

Çalışma sonucunda elde edilen bulgular, yatırımcılar için önemli bir konu olan hisse senedi değer tahmininde yeni bir veri kaynağı olarak sosyal medyanın önemini göstermektedir. Bu sosyal medyadan elde edilen bilgilerin kullanımında diğer yatırım aracı verileriyle desteklenerek ve Gradyan Destekli Ağaçlar algoritması kullanarak tahmin çalışmalarının yürütülmesi, tahminin başarısını arttırıp ilgili karar vericilere önemli bir bilgi sağlayacağına inanılmaktadır.

Elde edilen olumlu ve başarılı sonuçlara rağmen, bazı hisse senedi verilerinde gerek korelasyon analizlerinde herhangi bir ilişki bulunamaması gerek tahmin algoritmalarında uygun sonuçlar elde edilememesi nedeniyle sosyal medya verilerine dayalı hisse senedi tahminleme yönteminin tüm sektörler veya tüm firmalar için uygun olduğu bilinmemektedir. Örneğin QHY, LHY, FHY hava yolu için korelasyon analizinde bir ilişki tespit edilememiştir. Bu nedenle sadece THY ve SHY havayolu şirketleri için sadece makine öğrenmesiyle hisse değeri tahmin yöntemi kullanılmıştır. Analizlerdeki bu durum bize şu mesajı da verebilmektedir: Sosyal medya verileri firmanın sosyal medyadaki varlıklarına, sosyal medya hesaplarına, aktif kullanımına, firmanın bulunduğu sektöre bağlı olarak değişkenlik göstermesi nedeniyle her firma için sosyal medya verilerine dayalı hisse senedi değer tahmini yapmak mümkün olmayabilir. Yapılacak analizlerde firmaların sektör farklılığı dikkate alınmalıdır.

Yukarıda bahsettiğimiz hususlardan dolayı, ileride de yapılabilecek çalışmalarda, sektörün genişletilmesi, farklı sektörlerden firmaların verilerinin toplanması, hem de veri toplama kaynağı olarak farklı sosyal medya mecralarının ve ek diğer yatırım aracı verilerine ait göstergelerin kullanılması önerilmektedir. Böylelikle, çalışma kapsamı, örneklem sayısı, bağımsız değişken sayısı genişletilerek, tahminleme modellerinin başarısı arttırılabilecektir. Bu alanda yapılacak yeni çalışmalarla, sosyal medyadaki büyük verilerden finans sektöründe de faydalanılması yaygınlaşabilecektir.

EK-1: Tanımlayıcı İstatistik Sonuçları

Tablo 1

SHY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NAI	SGXL/SGD	USD/SGD	Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri
1.10.2019	116	78	13	6	19	65	8,50	1,3850	9,24	9,08
2.10.2019	101	70	8	3	20	62	8,38	1,3841	9,08	9,02
3.10.2019	147	87	18	3	39	69	8,41	1,3809	9,02	9,01
.
.
3.2.2020	113	46	30	7	30	16	8,59	1,3689	8,49	8,55
4.2.2020	96	25	26	2	43	-1	8,68	1,3701	8,55	8,59
5.2.2020	93	31	17	11	34	14	8,64	1,3815	8,59	8,58

Tablo 2

QHY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NAI	Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri
1.10.2019	366	143	86	20	117	57	6,44	6,40
2.10.2019	253	110	44	11	88	66	6,40	6,30
3.10.2019	279	115	54	11	99	61	6,30	6,37
.
.
4.2.2020	174	75	52	19	28	23	6,51	6,56
5.2.2020	167	63	59	14	31	4	6,56	6,66
6.2.2020	103	33	28	15	27	5	6,66	6,48

Tablo 3

LHY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NAI	Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri
1.10.2019	653	170	169	148	166	1	14,680	14,395
2.10.2019	337	74	64	88	111	10	14,395	14,395
3.10.2019	302	78	73	74	77	5	14,395	13,935
.
.
.
4.2.2020	152	53	43	24	32	10	14,520	15,035
5.2.2020	241	67	61	52	61	6	15,035	15,000
6.2.2020	51	17	18	9	7	-1	15,000	14,690

Tablo 4

FHY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NAI	Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri
10.10.2019	200	73	39	13	75	34	9,970	10,010
11.10.2019	124	64	24	2	34	40	10,010	10,010
12.10.2019	109	44	23	9	33	21	10,010	10,010
.
.
.
4.2.2020	341	260	14	7	60	246	8,79	9,14
5.2.2020	266	138	37	8	83	101	9,14	9,16
6.2.2020	65	34	6	3	22	28	9,16	8,85

EK-2: Korelasyon Analizi Sonuçları

Tablo 1

THY'ye ait verilerin korelasyon analizi

		Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri	Toplam	Pozitif	Negatif	NAI	Dolar	Altın/gr	Bist-100	
Spearman's rho	Aynı Günkü Hisse Değeri	Korelasyon Katsayısı	1	,921**	-,267**	-,284**	-,303**	-,034	,474**	,349**	,927**
		Anlamlılık Düzeyi (Sig.)		,000	,003	,002	,001	,711	,000	,000	,000
		Veri Sayısı (N)	118	118	118	118	118	118	118	118	118
	Ertesi Günkü Hisse Değeri	Korelasyon Katsayısı	,921**	1	-,280**	-,296**	-,296**	-,058	,471**	,348**	,906**
		Anlamlılık Düzeyi (Sig.)	,000		,002	,001	,001	,535	,000	,000	,000
		Veri Sayısı (N)	118	118	118	118	118	118	118	118	118

** Korelasyon 0,01 düzeyinde anlamlıdır.

* Korelasyon 0,05 düzeyinde anlamlıdır.

Tablo 2

SHY'ye ait verilerin korelasyon analizi

		Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri	Toplam	Pozitif	Negatif	NAI	SGXL/SGD	USD/SGD	
Spearman's rho	Aynı Günkü Hisse Değeri	Korelasyon Katsayısı	1	,912**	,318**	,546**	-,117	,600**	,427**	,159
		Anlamlılık Düzeyi (Sig.)		,000	,001	,000	,215	,000	,000	,091
		Veri Sayısı (N)	114	114	114	114	114	114	114	114
	Ertesi Günkü Hisse Değeri	Korelasyon Katsayısı	,912**	1	,334**	,545**	-,111	,611**	,435**	,152
		Anlamlılık Düzeyi (Sig.)	,000		,000	,000	,241	,000	,000	,106
		Veri Sayısı (N)	114	114	114	114	114	114	114	114

** Korelasyon 0,01 düzeyinde anlamlıdır.

* Korelasyon 0,05 düzeyinde anlamlıdır.

Tablo 3

LHY'ye ait verilerin korelasyon analizi

		Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri	Toplam	Pozitif	Negatif	NAI	
Spearman's rho	Aynı Günkü Hisse Değeri	Korelasyon Katsayısı	1	,971**	-,197*	-,146	-,174	,020
		Anlamlılık Düzeyi (Sig.)		,000	,031	,111	,056	,826
		Veri Sayısı (N)	121	121	121	121	121	121
	Ertesi Günkü Hisse Değeri	Korelasyon Katsayısı	,971**	1	-,181*	-,119	-,149	-,044
		Anlamlılık Düzeyi (Sig.)	,000		,047	,195	,104	,631
		Veri Sayısı (N)	121	121	121	121	121	121

** Korelasyon 0,01 düzeyinde anlamlıdır.

* Korelasyon 0,05 düzeyinde anlamlıdır.

Tablo 4

QHY'ye ait verilerin korelasyon analizi

			Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri	Toplam	Pozitif	Negatif	NAI
Spearman's rho	Aynı	Korelasyon Katsayısı	1	,958**	-,096	-,016	-,100	-,100
	Günlük	Anlamlılık Düzeyi (Sig.)		,000	,313	,866	,293	,293
	Hisse	Veri Sayısı (N)	112	112	112	112	112	112
	Değeri	Korelasyon Katsayısı	,958**	1	-,065	-,002	-,057	-,061
	Ertesi	Anlamlılık Düzeyi (Sig.)	,000		,497	,986	,551	,524
	Günlük	Veri Sayısı (N)	112	112	112	112	112	112

** Korelasyon 0,01 düzeyinde anlamlıdır.

* Korelasyon 0,05 düzeyinde anlamlıdır.

Tablo 5

FHY'ye ait verilerin korelasyon analizi

			Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri	Toplam	Pozitif	Negatif	NAI
Spearman's rho	Aynı	Korelasyon Katsayısı	1	,932**	-,185*	-,339**	-,141	-,216*
	Günlük	Anlamlılık Düzeyi (Sig.)		,000	,050	,000	,137	,022
	Hisse	Veri Sayısı (N)	113	113	113	113	113	113
	Değeri	Korelasyon Katsayısı	,932**	1	-,104	-,245**	-,099	-,143
	Ertesi	Anlamlılık Düzeyi (Sig.)	,000		,271	,009	,295	,131
	Günlük	Veri Sayısı (N)	113	113	113	113	113	113

** Korelasyon 0,01 düzeyinde anlamlıdır.

* Korelasyon 0,05 düzeyinde anlamlıdır.

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazarlar çıkar çatışması etmemişlerdir.

Finansal Destek: Bu çalışma, Yalova Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından desteklenmiştir. (Proje No: 2019/YL/0001).

Yazar Katkıları: Çalışma Konsepti/Tasarım- İ.S.; Veri Toplama-Ö.F.U.; Veri Analizi/Yorumlama- Ö.F.U.; Yazı Taslağı- Ö.F.U.; İçerigin Eleştirel İncelemesi-İ.S.; Son Onay ve Sorumluluk- Ö.F.U., İ.S.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: This study was supported by Yalova University Scientific Research Projects Coordination Unit. (Project No: 2019/YL/0001).

Author Contributions: Conception/Design of Study- İ.S.; Data Acquisition- Ö.F.U.; Data Analysis/Interpretation- Ö.F.U.; Drafting Manuscript- Ö.F.U.; Critical Revision of Manuscript-İ.S.; Final Approval and Accountability- Ö.F.U., İ.S.

Kaynaklar/References

- Aktaş, M., & Akdağ, S. (2013). Türkiye’de Ekonomik Faktörlerin Hisse Senedi Fiyatları İle İlişkilerinin Araştırılması. *International Journal of Social Science Research (Las Vegas, Nev.)*, 2(1), 50–67.
- Arun, K., Srinagesh, A., & Ramesh, M. (2017). Twitter Sentiment Analysis on Demonetization tweets in India Using R language. *International Journal of Computer Engineering In Research Trends*, 4(6), 252–258. Tarihinde adresinden erişildi www.ijcert.org
- Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. *Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010, I*(March 2010), 492–499. <https://doi.org/10.1109/WI-IAT.2010.63>
- Attigeri, G. V, M, M. P. M., Pai, R. M., & Nayak, A. (2015). Stock market prediction: A big data approach. *TENCON 2015 - 2015 IEEE Region 10 Conference*, 1–5. <https://doi.org/10.1109/TENCON.2015.7373006>
- Bhardwaj, A., Narayan, Y., & Dutta, M. (2015). *Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty*. 70, 85–91. <https://doi.org/10.1016/j.procs.2015.10.043>
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Charan, A. (2015). *Marketing analytics: A Practitioner’s Guide to Marketing Analytics and Research Methods*. World Scientific Publishing Company.
- Chen, R., & Lazer, M. (2011). *Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement*. 1–5.
- Cingöz, F., & Kendirli, S. (2019). Altın Fiyatları, Döviz Kuru ve Borsa İstanbul Arasındaki İlişki. *Finans Ekonomi ve Sosyal Araştırmalar Dergisi*, 545–554. <https://doi.org/10.29106/fesa.649254>
- Das, S., & Behera, R. K. (2018). Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction. *Procedia Computer Science*, 132(Iccids), 956–964. <https://doi.org/10.1016/j.procs.2018.05.111>
- Das, S. R., Kim, S., & Kothari, B. (2017). Zero-Revelation RegTech: Detecting Risk through Linguistic Analysis of Corporate Emails and News. *SSRN Electronic Journal*, 0–32. <https://doi.org/10.2139/ssrn.2909380>
- Dickinson, B., & Hu, W. (2015). Sentiment Analysis of Investor Opinions on Twitter. *Social Networking*, 04(03), 62–71. <https://doi.org/10.4236/sn.2015.43008>
- Dolapçı, B. (2020). *APACHE SPARK KULLANILARAK BÜYÜK BOYUTLU GÖRÜNTÜLERİN ANALİZİ*. Karabük Üniversitesi.
- Eliacık, A. B., & Erdogan, N. (2015). Mikro Bloglardaki Finans Toplulukları için Kullanıcı Ağırlıklandırılmış Duygu Analizi Yöntemi. *Ulusal Yazılım Mühendisliği Sempozyumu*, 782–793.
- Evans, J. R. (2016). *Business Analytics* (2nd baskı). Boston: Pearson Education.
- Gazioğlu, K., & Şeker, Ş. E. (2017). *Veri Madenciliği Yöntemleri ile Twitter Üzerinden Girişimcilik Analizi*. 3–10.
- Gençtürk, M. (2009). Finansal Kriz Dönemlerinde Makroekonomik Faktörlerin Hisse Senedi Fiyatlarına Etkisi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14/1, 127–136.
- Holsapple, C. W., Hsiao, S.-H., & Pakath, R. (2018). Business social media analytics: Characterization and conceptual framework. *Decision Support Systems*. <https://doi.org/https://doi.org/10.1016/j.dss.2018.03.004>
- investing. (2020a). Altın-Try. Tarihinde 03 Ocak 2020, adresinden erişildi <https://tr.investing.com/currencies/gau-try>
- investing. (2020b). BİST-100 hisse değeri. Tarihinde 01 Mart 2020, adresinden erişildi <https://tr.investing.com/indices/ise-100>
- investing. (2020c). “investing.com” internet sitesi. Tarihinde 06 Şubat 2020, adresinden erişildi <https://tr.investing.com/>
- investing. (2020d). SIAL hisse değeri. Tarihinde 06 Şubat 2020, adresinden erişildi <https://tr.investing.com/equities/singapore-airlines>
- investing. (2020e). Singapore Exchange(SGXL/SGD). Tarihinde 06 Şubat 2020, adresinden erişildi <https://tr.investing.com/equities/singapore-exchange>
- investing. (2020f). Usd-Try. Tarihinde 01 Mart 2020, adresinden erişildi <https://tr.investing.com/currencies/usd-try-chart>
- investing. (2020g). USD/SGD değeri. Tarihinde 06 Şubat 2020, adresinden erişildi <https://tr.investing.com/currencies/usd-sgd>
- Karahan, M. (2011). İstatistiksel Tahmin Yöntemleri: Yapay Sinir Ağları Metodu ile Ürün Talep Tahmini Uygulaması. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı*, 112–113.
- Lee, I. (2018). Social media analytics for enterprises: Typology, methods, and processes. *Business Horizons*, 61(2), 199–210. <https://doi.org/https://doi.org/10.1016/j.bushor.2017.11.002>

- Mittal, A., & Goel, A. (2012). Stock Prediction Using Twitter Sentiment Analysis. <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>, (December). Tarihinde adresinden erişildi <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>
- Murugesan, S. (2007). Understanding Web 2.0. *IT Professional*, 9(4), 34–41. <https://doi.org/10.1109/MITP.2007.78>
- Özüpek, M. N., & Diker, E. (2013). İLETİŞİM FAKÜLTESİ ÖĞRENCİLERİNİN CEP TELEFONU MARKALARINA YÖNELİK İMAJ ALGISI: NOKIA VE SAMSUNG ÖRNEĞİ ÖZET. *New World Sciences Academy*, 369(1), 100–120. Tarihinde adresinden erişildi <http://dx.doi.org/10.1016/j.jsames.2011.03.003><https://doi.org/10.1016/j.gr.2017.08.001><http://dx.doi.org/10.1016/j.precamres.2014.12.018><http://dx.doi.org/10.1016/j.precamres.2011.08.005><http://dx.doi.org/10.1080/00206814.2014.902757><http://dx.doi.org/10.1016/j.jsames.2011.03.003>
- Pagolu, V. S., & Majhi, B. (2016). Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPE5)*, 1345–1350. <https://doi.org/10.1109/SCOPE5.2016.7955659>
- Porshnev, A., Redkin, I., & Shevchenko, A. (2013). Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis. *2013 IEEE 13th International Conference on Data Mining Workshops*, 440–444. <https://doi.org/10.1109/ICDMW.2013.111>
- Rapid Miner. (2019). Gradient Boosted Trees Model. Tarihinde 06 Şubat 2020, adresinden erişildi https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/gradient_boosted_trees.html
- Salo, J. (2017). Social media research in the industrial marketing field: Review of literature and future research directions. *Industrial Marketing Management*, 66, 115–129. <https://doi.org/https://doi.org/10.1016/j.indmarman.2017.07.013>
- Si, J., Mukherjee, A., Liu, B., & Li, Q. (2013). *Exploiting Topic based Twitter Sentiment for Stock Prediction*. (2011), 24–29.
- Skytrax. (2020). 2019 World Airline Ranking List. Tarihinde 10 Şubat 2020, adresinden erişildi <https://www.worldairlineawards.com/worlds-top-100-airlines-2019/>
- Syam, N., & Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management*, 69, 135–146. <https://doi.org/https://doi.org/10.1016/j.indmarman.2017.12.019>
- Tripathi, P., Vishwakarma, S. K., & Lala, A. (2015). Sentiment Analysis of English Tweets Using Rapid Miner. *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*.
- Ulusoy, N. (2012). Sözlüklerdeki Sinema Sevgisi: New York'ta Beş Minare ve Çoğunluğun İnternet Sözlüklerine Yansıması. *Beta Yayıncılık*, 195–211.
- Vu, T. T., Chang, S., Quang, T. H., & Collier, N. (2012). An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter. *Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data*, 3, 23–38.
- Yadav, R., Kumar, A. V., & Kumar, A. (2019). News-based supervised sentiment analysis for prediction of futures buying behaviour. *IIMB Management Review*, 31(2), 157–166. <https://doi.org/10.1016/j.iimb.2019.03.006>
- Yıldırım, M., & Yüksel, C. A. (2017). Sosyal Medya İle Hisse Senedi Fiyatının Günlük Hareket Yönü Arasında İlişkini İncelenmesi: Duygu Analizi Uygulaması. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*. <https://doi.org/10.18092/ulikidince.352414>
- Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter “ I hope it is not as bad as I fear ”. *Procedia - Social and Behavioral Sciences*, 26(2007), 55–62. <https://doi.org/10.1016/j.sbspro.2011.10.562>

