



T.C.
YALOVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**HAVAYOLU FİRMALARININ SOSYAL MEDYA VERİLERİNDEN
HİSSE DEĞERİNİN TAHMİN EDİLMESİ**

YÜKSEK LİSANS TEZİ

**Ömer Faruk UYRUN
175109009**

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi İbrahim SABUNCU

MAYIS 2020



T.C.
YALOVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**HAVAYOLU FİRMALARININ SOSYAL MEDYA VERİLERİNDEN
HİSSE DEĞERİNİN TAHMİN EDİLMESİ**

YÜKSEK LİSANS TEZİ

**Ömer Faruk UYRUN
175109009**

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi İbrahim SABUNCU

MAYIS 2020

YALOVA Üniversitesi Fen Bilimleri Enstitüsü'nün 175109009 numaralı Yüksek Lisans Öğrencisi **Ömer Faruk UYRUN**, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "**HAVAYOLU FİRMALARININ SOSYAL MEDYA VERİLERİNDEN HİSSE DEĞERİNİN TAHMİN EDİLMESİ**" başlıklı tezini aşağıda imzaları olan jüri önünde sunmuş ve oy birliği/oy çokluğu ile başarılı bulunmuştur.

İmza / Kanaati
(Kabul/Red)

Tez Danışmanı :

..... /.....

Jüri Üyeleri :

..... /.....

..... /.....

Teslim Tarihi : 06 Mayıs 2020

Savunma Tarihi : 15 Mayıs 2020

ÖNSÖZ

Tez konusunun belirlenmesinden tezin son bölümüne kadar her aşamada yol gösteren tez danışmanım Sayın Dr. Öğr. Üyesi İbrahim Sabuncu'ya, bu süreçte hiçbir zaman yardımını esirgemeyen Dr. Öğr Üyesi Eyüp Çalık ve Arş. Gör. Mahir Atmış'a ve beni en iyi şekilde yetiştiren, her zaman yanımda olan, dualarını esirgemeyen annem Süreyya Uyrun'a ve babam Nezir Uyrun'a teşekkürlerimi sunarım.

Mayıs 2020

Ömer Faruk UYRUN
Endüstri Mühendisi





İÇİNDEKİLER

	Sayfa
	No
KISALTMALAR	vii
ÇİZELGE LİSTESİ.....	ix
ŞEKİL LİSTESİ.....	xi
ÖZET.....	xiii
ABSTRACT.....	xv
1. GİRİŞ	1
2. LİTERATÜR	5
3. METODOLOJİ VE ANALİZLER.....	11
3.1 Veri Toplama	12
3.2 Twitter API	14
3.3 R Programlama Dili	16
3.4 Python Programlama Dili	17
3.5 Rapid Miner ve Meaning Cloud	20
3.6 Tanımlayıcı İstatistikler	24
3.7 Korelasyon Analizleri	27
3.8 Auto Model ile Tahmin Modeli Kurma	33
4. BULGULAR VE TARTIŞMA	47
5. SONUÇ	51
Destekleyen Kişi ve Kurum Bilgisi	53
KAYNAKLAR	55
ÖZGEÇMİŞ	61



KISALTMALAR

THY	: Türk Hava Yolları
SHY	: Singapore Hava Yolları
LHY	: Lufthansa Hava Yolları
QHY	: Qantas Hava Yolları
FHY	: France Hava Yolları
NPS	: Net Promoter Score (Pozitif-Negatif farkı)
USD	: Amerikan Doları
SGXL	: Singapur Borsası Hisse Deęeri
SGD	: Singapur Doları
BİST-100	: Borsa İstanbul -100 Endeksi Hisse Deęeri
THYAO	: Türk Hava Yolları'na ait Hisse Deęeri
SIAL	: Singapur Hava Yolları'na ait Hisse Deęeri
QAN	: Qantas Hava Yolları'na ait Hisse Deęeri
LHAG	: Lufthansa Hava Yolları'na ait Hisse Deęeri
AIRF	: France Hava Yolları'na ait Hisse Deęeri
SPSS	: Statistical Package for the Social Sciences
MSE	: Ortalama Hata Kareleri
MAPE	: Ortalama Mutlak Yüzde Hata



ÇİZELGE LİSTESİ

Sayfa No

Çizelge 3.1	THY'ye ait hisse değeri verileri	13
Çizelge 3.2	Excel dosyasına kaydedilen tweetlerin örneği.....	23
Çizelge 3.3	THY'ye ait tweet sayıları ve hisse değeri verileri	25
Çizelge 3.4	SHY'ye ait tweet sayıları ve hisse değeri verileri.....	25
Çizelge 3.5	QHY'ye ait tweet sayıları ve hisse değeri verileri	26
Çizelge 3.6	LHY'ye ait tweet sayıları ve hisse değeri verileri	26
Çizelge 3.7	FHY'ye ait tweet sayıları ve hisse değeri verileri.....	27
Çizelge 3.8	THY'ye ait verilerin normallik testi	28
Çizelge 3.9	THY'ye ait verilerin korelasyon analizi	28
Çizelge 3.10	SHY'ye ait verilerin normallik testi.....	29
Çizelge 3.11	SHY'ye ait verilerin korelasyon analizi.....	29
Çizelge 3.12	LHY'ye ait verilerin normallik testi	30
Çizelge 3.13	LHY'ye ait verilerin korelasyon analizi	30
Çizelge 3.14	QHY'ye ait verilerin normallik testi	31
Çizelge 3.15	QHY'ye ait verilerin korelasyon analizi	31
Çizelge 3.16	FHY'ye ait verilerin normallik testi.....	32
Çizelge 3.17	FHY'ye ait verilerin korelasyon analizi.....	32
Çizelge 3.18	THY için yüklenen örnek veriler	33
Çizelge 3.19	THY için tahmin modeli analiz sonuçları	38
Çizelge 3.20	Tahmin edilen ve gerçekleşen hisse değeri verileri (THY)	40
Çizelge 3.21	SHY için tahmin modeli analiz sonuçları	41
Çizelge 3.22	Tahmin edilen ve gerçekleşen hisse değeri (SHY)	42
Çizelge 3.23	THY için tahminin hata testleri sonucu	43
Çizelge 3.24	THY için tahmin hata sonuçları.....	44
Çizelge 3.25	SHY için tahminin hata testleri sonucu	44
Çizelge 3.26	SHY için tahmin hata sonuçları	45
Çizelge 4.1	Korelasyon analizi sonuçları.....	47

ŞEKİL LİSTESİ	Sayfa No
Şekil 2.1 Metodoloji.....	10
Şekil 3.1 Investing.com hisse değeri verileri	12
Şekil 3.2 Developer Twitter [49]	14
Şekil 3.3 Developer Twitter uygulaması.....	15
Şekil 3.4 Developer Twitter uygulama detayları	15
Şekil 3.5 Twitter API kodları	16
Şekil 3.6 Twitter API kodlarının R Studio'da kullanımı	17
Şekil 3.7 R programlama dilinde yazılan kodlar.....	17
Şekil 3.8 Twitter'dan veri çekme Python kodları	18
Şekil 3.9 Twitter API Python kodları.....	18
Şekil 3.10 JSON formatında kaydedilen tweetler	19
Şekil 3.11 Rapid Miner veri indirme modeli - 1	20
Şekil 3.12 Search Twitter parametre ayarları.....	21
Şekil 3.13 Twitter connection tanımlanması.....	22
Şekil 3.14 Rapid Miner veri indirme modeli – 2	23
Şekil 3.15 Meaning cloud excel eklentisi duygu analizi.....	24
Şekil 3.16 Rapid Miner'a verilerin yüklenmesi.....	34
Şekil 3.17 Yüklenecek verilerin seçilmesi	34
Şekil 3.18 Tahmin edilecek değişkenin rolünü değiştirme	35
Şekil 3.19 THY'ye ait veriler ve otomodel	35
Şekil 3.20 Değişkenlerin seçilmesi	36
Şekil 3.21 Tahmin modeli seçenekleri	37
Şekil 3.22 Tahmin modelinin kaydedilmesi	38
Şekil 3.23 Tahmin edilecek verinin seçilmesi.....	39
Şekil 3.24 Tahmin edilen veriler	39
Şekil 3.25 Tahmin edilen veriler	41
Şekil 4.1 SHY ağırlık oranları.....	49
Şekil 4.2 THY ağırlık oranları	50



HAVAYOLU FİRMALARININ SOSYAL MEDYA VERİLERİNDEN HİSSE DEĞERİNİN TAHMİN EDİLMESİ

ÖZET

Bu çalışmanın amacı Havayolu firmaları için sosyal medya verilerinden hisse değerinin tahmin edilip, bu bilgiye dayalı yatırım yapılabilirliğinin araştırılmasıdır. Bu amaçla, Twitter platformundan Türk Hava Yolları(THY), Singapore Hava Yolları(SHY), Lufthansa Hava Yolları(LHY), Qantas Hava Yolları(QHY) ve France Hava Yolları(FHY) firmalarına ait, 01.10.2019 - 06.02.2020 tarihleri arasında paylaşılan 236 764 adet tweet, Rapid Miner kullanılarak, toplanmıştır. Ayrıca günlük hisse değeri verileri ise investing.com sitesinden elde edilmiştir. Toplanan Twitter gönderileri, MeaningCloud uygulaması ile duygu analizleri yapılarak, pozitif, negatif, nötr, tanımlanmayan, şeklinde dört duygu kategorisinde sınıflandırılmıştır. Ardından, bu duygu kategorileri ile hesaplanan Net Destekçi Skoru (NPS:Net Promoter Score:pozitif-negatif tweet sayıları) ve toplam tweet sayıları ile ertesi günkü hisse değeri arasındaki ilişki SPSS istatistik programı ile korelasyon analizleri yapılarak araştırılmıştır. Korelasyon analizleri sonucunda hisse senedi değeri verileri ile Twitter platformundaki anılma sayısı arasında, THY, LHY ve FHY firmaları için negatif, SHY firması için ise pozitif yönlü orta düzeyde korelasyon olduğu tespit edilmiştir.

Hisse senedi değeri tahmin modeli oluşturmak için Rapid Miner programında Auto Model'den faydalanılmıştır. "Gradient Boosted Trees" modeli, THY ve SHY firmaları için en düşük hata değerine sahip tahmin modeli olmuştur. Ertesi günkü hisse değerinin artış veya azalışları THY için %55,6 , SHY için ise %57,1 oranda doğru tahmin edilmiştir. Model sayesinde, SHY'nin ertesi günkü hisse değeri %5,1 , THY'nin ise %2,6 hata oranı ile tahmin edilebilmiştir. Tahmin modelinde, NPS değerinin SHY için en yüksek, THY için ise BİST-100'den sonra ikinci en yüksek etkiye sahip faktör ağırlığına sahip olduğu görülmüştür. Sonuç olarak, Twitter anılma sayılarının yatırım kararları verirken faydalanılabilecek önemli bir veri kaynağı olduğu tespit edilmiştir.

Anahtar Kelimeler: Sosyal Medya Analizi; Hisse Değeri; Twitter; Korelasyon; Veri Madenciliği; Tahmin Modeli



FORECASTING OF STOCK MARKET VALUES OF AIRLINE COMPANIES BY USING SOCIAL MEDIA DATA

ABSTRACT

The main objective of this study is to predictive the stock market value of Airline companies by using social media data and to investigate the investment feasibility based on this information. For this purpose, 236 764 tweets shared between 01.10.2019-06.02.2020 are collected from Twitter for Turkish Airlines (THY), Singapore Airlines (SIA), Lufthansa Airlines (DLH), Qantas Airlines (QFA), and Air France Airlines (AFR) companies by using Rapid Miner. In addition, daily stock market value data is obtained from investing.com website. The collected Twitter posts are classified into four sentiment categories: positive, negative, neutral, and none by making sentiment analysis with the MeaningCloud application. Then, the relationship between positive-negative tweet counts Net Promoter Score (NPS) and total tweet counts calculated with these emotion categories is studied, and the stock market value of the next day was investigated by using correlation analysis with SPSS statistics program. As a result of the correlation analysis, it was determined that there was a negative between the stock market value data and the number of citations on the Twitter platform for THY, DLH, and AFR companies, and a positive mid-rate correlation for SIA company.

Finally, Auto-Model was used in the RapidMiner program to create a stock market value prediction model. “Gradient Boosted Trees” model has the lowest error prediction for THY and SIA companies. The increase or decrease of the stock market value of the next day is estimated correctly as 55,6% for THY and 57,1% for SIA. With this model, the stock market value of SIA on the next day is predicted with 5,1% and THY’s with a 2,6% error rate using the number of commemorations. In the prediction model, it was observed that the NPS value had the highest weight-factor for SIA and the second-highest effect for THY after BIST-100 value. Consequently, it has been determined that the number of Twitter citations is a crucial data source that can be utilized in making investment decisions.

Keywords: Social Media Analysis; Stock Market Value; Twitter; Correlation; Data Mining; Prediction Model.



1. GİRİŞ

İş analitiği, firmaların faaliyetlerinin durumlarını öğrenebilmesine yardımcı olmak ve daha gerçekçi kararlar vermelerini sağlamak amacıyla, bilgi, istatistik, matematik ve bilgisayar tabanlı yöntemlerin kullanılması olarak tanımlanmaktadır [1]. İş analitiği yöntemleri, üretim, ar-ge, finans, kalite, lojistik gibi firmanın tüm departmanlarında kullanılabilir. Ama en sık kullanılan departman pazarlama departmanı olduğu görülmektedir. Vaughan [2] Bain & Company araştırma firmasının yapmış olduğu bir anket çalışmasında büyük verilerin en büyük etkisinin % 38'inin Pazarlamada olduğu sonucuna varmıştır. Pazarlama dalındaki iş analitiği uygulamaları, müşteri analitiği veya analitik pazarlama [3] olarak isimlendirilmektedir.

Müşteri analitiği, firmaların muhasebe kayıtları, müşterilerinin sosyal medya hareketleri, web sayfaları vb. farklı kanallardan elde ettikleri işlem kayıtları verilerinin analiz edilip işlendikten sonra elde edilen bilgilere göre pazarlama kararları vermek için kullanılan bir yöntem olarak tanımlanmaktadır [4]. Analitik pazarlama alanında sosyal medya verilerinin kullanılması, iş analitiğinin önemli bir alt dalı olarak sosyal medya analitiği şeklinde tanımlanmıştır [5].

Sosyal medya, kişilerin yeni nesil internet ağı teknolojilerinin sağladığı kolaylık ve iletişim hızı ile elde ettiği içeriği yayınlamalarını ve paylaşmalarını sağlayan çevrimiçi bir platform ağıdır [6]. Günümüzde ise kişiler duygu ve düşüncelerini, fotoğraflarını, videolarını paylaşmak, yeni arkadaşlar edinmek, eski arkadaşlarını bulmak gibi nedenlerden dolayı kullandıkları çevrimiçi platformları sosyal medya olarak adlandırmışlardır [7]. Dewing [8] ise sosyal medya kavramını “kullanıcıların çevrimiçi olarak karşılıklı değişimlerde bulunduğu, kullanıcı tarafından oluşturulan içeriklere katkıda bulunmalarını veya çevrimiçi topluluklara katılmalarını sağlayan çeşitli şekillerde bulunan internet tabanlı hizmetler” olarak tanımlamıştır.

Sosyal medya kişilerin firmaların ürün ve hizmetleri hakkında paylaşım yapabildiği ve tüketicilerin bilgi edinmelerini sağlayacak bilgiler içerir [9]. Bilgilerin sosyal medya platformlarında üretilmesi kişilere ve firmalara birçok olanak sağlamaktadır. Sosyal medya sadece kişiler için değil aynı zamanda firmalar için de vazgeçilmez konuma sahiptir. Çünkü firmalar sosyal medyayı aktif kullanarak nihai tüketicilerin isteklerine, şikayetlerine ve problemlerine hızlı bir şekilde ulaşabilmektedir. Bu yönden bakınca sosyal medya firmalar için olumlu bir platform olsa da, nihai

tüketiciler ve firmalar için olumsuz sonuçlar doğurabilmektedir. Bir örnek üzerinden bakacak olursak nihai tüketiciler sosyal medya ortamında firmalar hakkında olumsuz görüş paylaştıklarında geniş kitlelere hızlı bir şekilde yayılabilmektedir. Sosyal Medyanın kullanımının artmasıyla firmalar ürünlerini sosyal medya aracılığıyla pazarlamak için bazı sosyal medya analiz tekniklerini kullanmayı tercih ederler. Bu teknikler, müşterilerinin çeşitli sosyal medya platformları aracılığıyla şirketlerin ürünleri ve hizmetleri hakkındaki öneri ve görüşleri hakkında bilgi edinmelerine olanak sağlar.

Sosyal medya analitiği kapsamında, sosyal medyadan elde edilen veriler satış, pazarlama ve fiyatlandırmada kullanılabilir [10]. Elde edilen bu veriler aracılığıyla, marka imajı takibi, net destekçi skoru (NPS: Net Promoter Score) ve net taraftar endeksi (NAI: Net Advocacy Index) hesaplanması [3], potansiyel müşterilerin belirlenebilmesi için doğal dil işleme (NLP: Natural Language Processing) ve yapay zeka algoritmaları kullanılmaktadır [11, 12]. Sosyal medyada yapılan paylaşımlar, pozitif, negatif, nötr olarak sınıflandırılabilir. Ayrıca metin madenciliği yöntemleri ile yapılan paylaşımlardaki öne çıkan konular tespit edilebilir [3]. Böylece firmaların marka imaj takibi yönetimi için markayı destekleyen veya kötüleyen mesaj içeriklerinin hangi konulardan kaynaklandığı, bölgesel olarak olumlu/olumsuz mesajların dağılımı tespit edilerek, yöneticiler markaları ile ilgili olumsuzlukları kaldırabilmek için proaktif kararlar verebilmelerine imkân sağlamaktadır.

Marka imajı takibi için, İletişim ve teknoloji tabanlı platformların hızlı gelişimi ile birlikte müşteriler, memnuniyetlerini/şikâyetlerini firmalara ulaştırma sürecinin ötesine giderek sosyal medya platformlarında tüketiciden tüketiciye (C2C) doğru yönelen bir memnuniyet/şikâyet iletme süreci içine girmektedirler [13]. Müşterilerin firmalar hakkında yaptıkları olumlu/olumsuz yorumlar firmaların marka imajını potansiyel müşterilerin zihninde olumlu veya olumsuz etkiler bırakabilmektedir [13]. Bu yüzden sosyal medya platformlarında müşteriler tarafından firmalar hakkında yaptıkları yorumlar, firmaları marka imajını etkilemektedir.

Firmalar marka imajı takibi yapabilmek için birçok yöntem kullanabilmektedir. Bunlardan biri sosyal medya analizi yöntemidir. Sosyal medya verileri analiz edilerek firmaların hisse değerini tahmin etmede kullanılmaktadır. Attigeri vd. [14],

sosyal medya verilerinin firmaların hisse deęeri tahmin edilmesinde ok byk etkisi olduęunu tespit etmiřlerdir. Sosyal medyadan toplanan yorumların duygu analizleri ile firmaların hisse deęeri arasındaki korelasyon analiz edilmiřtir. Bu yntem ile sosyal medyadaki duygu durumu ile hisse deęeri performansının tahmin edebileceęi ve sosyal medya verilerinin hisse deęeri performansı ile yakından iliřkili olduęu sylenbilir.

Sosyal medya analitięi kapsamında yakın tarihte yapılan alıřmalar ierisinde sosyal medya verilerinden hareketle řirketlerin geliri, hisse senedi deęeri gibi finansal deęerleri tahmini amacıyla da yapılan alıřmalar vardır. Bu alıřmada havayolu firmaları iin sosyal medya verileri ile hisse senedi deęeri arasındaki korelasyon arařtırılmıřtır.

Bu alıřmada Uluslararası Havayolu Firmaları (THY, LHY, QHY, FHY, SHY) hakkında sosyal medya platformu olan Twitter'dan kullanıcıların yapmıř olduęu yorumlar toplanarak ve duygu analizleri sonucunda hisse deęerleri ile iliřkinin var olup olmadıęının tespit edilmesi amalanmıřtır. alıřmamız sonucunda hisse deęeri tahmini iin kullanılan yntemlere gre ok dřk maliyetli bir yntem olan sosyal medya verileri analiz edilerek firmaların hisse deęeri tahmin edilebileceęi sonucuna varılmıřtır.

Sonuç olarak toplamda 236 764 adet tweet ile yapılan tm analizler sonucunda sosyal medya verileri ve makro ekonomik faktrler (THY Firması iin Dolar, Altın, BİST-100 deęerleri, SHY firması iin SGXL/SGD, USD/SGD deęerleri) ile birlikte ertesini gnk hisse deęerini tahmin eden model sunulmaya alıřılmıřtır. alıřma beř ana blmden oluřmaktadır. Birinci blmde alıřmanın amacından bahsedilmektedir. İkinici blmde ise, literatrde benzer olan alıřmalar ve ieriklerden kısaca bahsedilmektedir. alıřmanın nc blmnde kullanılan metodoloji ve analizlerden detaylı bir řekilde bahsedilmektedir. Bu blmde ayrıca, tahmin modeli iin gerekli verilerin nasıl toplandıęından, istatistiksel analizlerinden, kurulan hipotezlerden, tahmin modelinin nasıl oluřturulduęundan ve modelin gvenilirlięi gerek veriler ile karřılařtırılmasından bahsedilmektedir. Drdnc blmde ise analizler sonucunda elde edilen bulgulardan bahsedilmektedir. Son blmde ise tahmin modellerinin sonuları analiz edilip havayolu firmalarına ait hisselerle yatırım kararıyla ilgili tavsiyelerde bulunulmuřtur.



2. LİTERATÜR

Sosyal Medyanın kullanımının artmasıyla tüketicilerin marka bakış açısını, memnuniyetlerini ve memnuniyetsizliklerini dile getirebileceği sosyal medyada Twitter popüler bir platform haline gelmiştir. Twitter verileri ile hisse değerleri arasındaki ilişkiyi araştıran çok az çalışmaya rastlanmıştır.

Asur ve Huberman [15], Twitter'dan elde ettikleri veriler doğrultusunda sinema sektöründe gişe gelirlerinin tahmin edilmesi üzerinde çalışmışlardır. Böylece seçtikleri filmler ile ilgili Twitter'daki anılma sayılarıyla gişe geliri arasında güçlü bir korelasyon olduğu tespit etmişlerdir.

Başka bir çalışmada Chen ve Lazer [16], hisse değeri hareketleri ile Twitter verileri arasındaki ilişkiyi incelemişlerdir. Özellikle hisse senedi fiyatlarındaki gelecekteki değişiklikleri tahmin edebilmek için model oluşturulmuştur. Oluşturulan modelin başarılı olduğu saptanmıştır.

Benzer çalışma olarak Zhang ve diğerleri [17], Twitter paylaşımlarını analiz ederek Dow Jones, NASDAQ ve S&P 500 borsa göstergelerini erken tahmin etmeye çalışan araştırma yapmışlardır. Duygusal tweet yüzdesinin Dow Jones, NASDAQ ve S&P 500 ile negatif yönde korelasyon olduğu, VIX ile anlamlı pozitif korelasyon olduğu sonucuna varılmıştır.

Ek olarak Bollen ve diğerleri [18], Twitter'daki olumlu ve olumsuz yorumların borsa piyasasını nasıl etkilediği üzerinde çalışmışlardır. Bu çalışmada Twitter'da yapılan paylaşımların duygu durumları ile hisse değeri Dow Jones Industrial Average (DJIA) arasında %87,6 oranında yüksek ilişki olduğu gösterilmiştir.

Bir benzer çalışmada Mittal ve Goel [19], Twitter paylaşımlarından elde edilen duygu durumu verileri ile bir sonraki günkü DJIA hisse değeri ile %75,56 korelasyon olduğunu tespit etmişlerdir.

Başka bir çalışmada Vu ve diğerleri [20], tüketicilerin duygu durumları ile hisse değeri arasındaki ilişkiyi araştırmıştır. Bu araştırmada NASDAQ borsasında işlem gören 4 teknoloji firmasını (Apple, Google, Microsoft, Amazon) incelemişlerdir. Sosyal medya verilerinin günlük olarak artış ve azalış durumunu tahmin edecek model Apple (AAPL) için %82,93, Google (GOOG) için %80,49, Microsoft (MSFT)

için %75,61 ve Amazon (AMZN) için ise %75,00 anlamlı ilişki olduğunu ortaya koymuşlardır.

Farklı bir çalışmada ise Porshnev ve diğerleri [21], Twitter kullanıcılarının duygu durumları hakkındaki verileri kullanarak borsa göstergelerinin öngörülerinin doğruluğunu artırma olasılığı üzerinde çalışma yapmışlardır. Duygu durumlarını analiz etmek için 755 milyondan fazla tweet i analiz etmişlerdir. DJIA hisse değerini öngörebilecekleri destek vektör makinesi algoritması modeli kullanılarak %64,10 doğruluk oranı elde etmişlerdir.

Benzer bir çalışma yapan Si ve diğerleri [22], borsa tahminine de yardımcı olmak için Twitter'dan konuya dayalı düşüncelerden yararlanmaya yönelik bir teknik çalışma yapmışlardır. Günlük konu setini öğrenmek için öncelikle Dirichlet Proses Karışım modelini (sonsuz karışım modeli olarak da bilinir) kullanmışlardır. S&P100 hisse senedi değerine ilişkin yaptıkları deneyler, yaklaşımlarının etkili olduğunu ve mevcut son teknoloji konu dışı yöntemlerden daha iyi performans gösterdiği sonucuna varmışlardır.

Farklı bir ayrıntıya değinen Khadjehve diğerleri [23], çevrimiçi metin madenciliğine dayalı finansal tahmini ile ilgili çalışmaları gözden geçirerek yapılan çalışmaları sahip olduğu jenerik bileşenlerin resmini ortaya çıkarmışlardır. Her sistemi diğerleriyle karşılaştırır ve her birinin arasındaki teorik ve teknik temelleri anlatmıştır. Bu çalışmada, bu alanda yapılacak çalışmalara yardımcı olmak ve daha fazla araştırma gerektiren, özel öneme sahip olan yönleri belirtmişlerdir.

Konuya farklı bir yorum getiren Bhardwaj ve diğerleri [24], yatırımcıların borsaya para yatırmadan önce, borsa tahmininde bulunmalarının önemine vurgu yapmışlardır. Bu amaçla borsa yatırımcıları için önemli bir faktör olan borsa tahminleme üzerine çalışma yapmışlardır. Sensex ve Nifty hisse senedinin fiyatını tahmin etmek için doğal dil işleme (NLP: Natural Language Processing) yöntemleri ile duygu analizi yapmışlardır. Bu amaçla Python programlama dilini kullanmışlardır. Sonuç olarak yatırımcılara paranın ne kadarını yatırım yapacaklarına dair tahmin yapmak için yardımcı olacak, aynı zamanda hisse piyasanın ekonomik dengesinin korunmasına yardımcı olacağını tespit etmişlerdir.

İncelenen başka bir çalışmada Dickinson ve Hu [25], yaptığı bir araştırmada kişilerin firmalar hakkında Twitter’da yaptıkları yorumları duygu analizi ile pozitif, negatif ve nötr olarak analiz ettikten sonra firmaların hisse değeri arasındaki ilişki üzerinde çalışmıştır. Duygu analizi sonuçları ile hisse değeri arasında istatistiksel analizler yapılmış ve hisse değerlerindeki değişim ile anlamlı bir ilişki olduğu tespit edilmiştir. Yapılan araştırmalar sonucunda Walmart ve Microsoft’ta yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu tespit edilirken, Goldman Sachs ve Cisco firmalarında yüksek düzeyde, negatif ve anlamlı bir ilişki olduğu saptanmıştır [25].

Türkiye’de yapılan bir çalışmada Eliaçık ve Erdoğan [26], 2015 yılında yaptıkları çalışmada mikro bloklardaki finans topluluklarındaki Türkçe iletilerin pozitif ve negatif olmak üzere duygu durumlarını belirlemek için bilinen yöntemlerden farklı bir duygu analiz yöntemi geliştirmişlerdir. Bu yöntemi kişilerin bilgileriyle mikro bloglarda bulunan finans topluluklarının duydu polaritesini ölçmek için kullanmışlardır. Kullandıkları yöntemin, önceki yöntemlere göre %73,63 başarı oranı ile daha hassas hesaplama yapabildiğini görmüşlerdir. Finans topluluklarının duygu polariteleri verileri ile Borsa İstanbul 100 endeksinin hareketleri arasında korelasyon analizi yapılmış ve orta düzeyde, pozitif ve anlamlı bir ilişki olduğunu tespit etmişlerdir.

Diğer taraftan başka bir çalışmada Li ve diğerleri [27], yaptıkları çalışmada SMeDA-SA (Social Media Data Analyzer – Sentiment Analysis) olarak adlandırılan bir tekniği, Twitter verilerini duygu analizi yapmak için ve daha sonra şirketlerin hisse senedi hareketini tahmin etmek için kullanmışlardır. Deneme amacıyla NASDAQ ve New York borsasında listelenen 30 şirket için 200 Milyon tweet toplamışlardır. SMeDA-SA ile, birçok şirketin hisse senedi hareketinin %70 in üzerinde bir ortalama doğrulukla oldukça doğru şekilde doğru bir şekilde tahmin edilebileceğini keşfetmişlerdir.

İncelenen bir diğer çalışmada Oliviera ve diğerleri [28], mikroblog verilerinden (Twitter veri seti) ve anket endekslerinden (AAII, II, USMC ve Sentix) çıkarılan duygu gösterge verilerini birleştirmek için Kalman Filtresini kullanmışlardır. Twitter ve KF duygu analizi göstergeleri bazı anket duygu durumu göstergelerinin öngörülmesinde faydalı olduğunu tespit etmişlerdir. Bu sonuçlar, yatırımcılar için

mikroblog verilerinin kullanılabilirliğini teyit ettiği, borsa davranışını tahmin etmeyi sağlayacağı sonucuna varmışlardır.

Bir başka çalışma yapan Pagolu ve diğerleri [29], yaptıkları çalışmada Microsoft şirketine ait hisse senedi fiyatlarındaki değişimlerin, mikro bloklardaki tweet'lerde ifade edilen duygu durumları ile ne kadar ilişkili olduğunu gözlemlemişlerdir. Tweet'lerde duygu durumlarını analiz etmek için Word2vec ve Ngram adlı iki farklı metin gösterimi kullanılmıştır. Twitter'dan elde edilen tweetlere duygu analizi ve makine öğrenmesi teknikleri uygulayarak, bu şirketin borsa hareketleri ile tweetlerin duygu durumları arasındaki ilişkiyi analiz etmişlerdir. Yapılan çalışmanın sonucunda pozitif olarak sınıflandırılmış tweetlerin oranı ile hisse değerinin hareket yönü (artış veya azalış olması) arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu tespit edilmiştir.

Türkiye'de yapılan Yıldırım ve Yüksel [30], 2017 yılında yaptıkları çalışmada Borsa İstanbul'da işlem gören Telekomünikasyon sektöründeki bir firma için 15 Mart 2017-15 Mayıs 2017 tarihleri arasında işlem gören günlerdeki hisse değeri verileri toplanmıştır. Firmanın hangi marka olduğu belirtilmemiş, 15 yılı aşkın süredir Borsa İstanbul'da işlem gördüğü, toplum tarafından bilinen bir marka olduğu belirtilmiştir. Seçilen firma için günlük olarak ilgili hisse değerlerinin açılış ve kapanış değerleri arasındaki fark alınarak, hisse değerinin hareket yönü o gün için pozitif veya negatif olarak belirlenmiştir. İlgili günler ile Twitter'dan elde edilen tweetler üzerinden duygu analizi yapılarak pozitif ve negatif olarak sınıflandırılmış ve günlük duygu polaritesi elde edilmiştir. Bu veriler kullanılarak yapılan analizler sonucunda pozitif tweet sayısının hisse senedinin hareket yönü ile arasında negatif, orta düzeyde bir ilişki saptanmıştır.

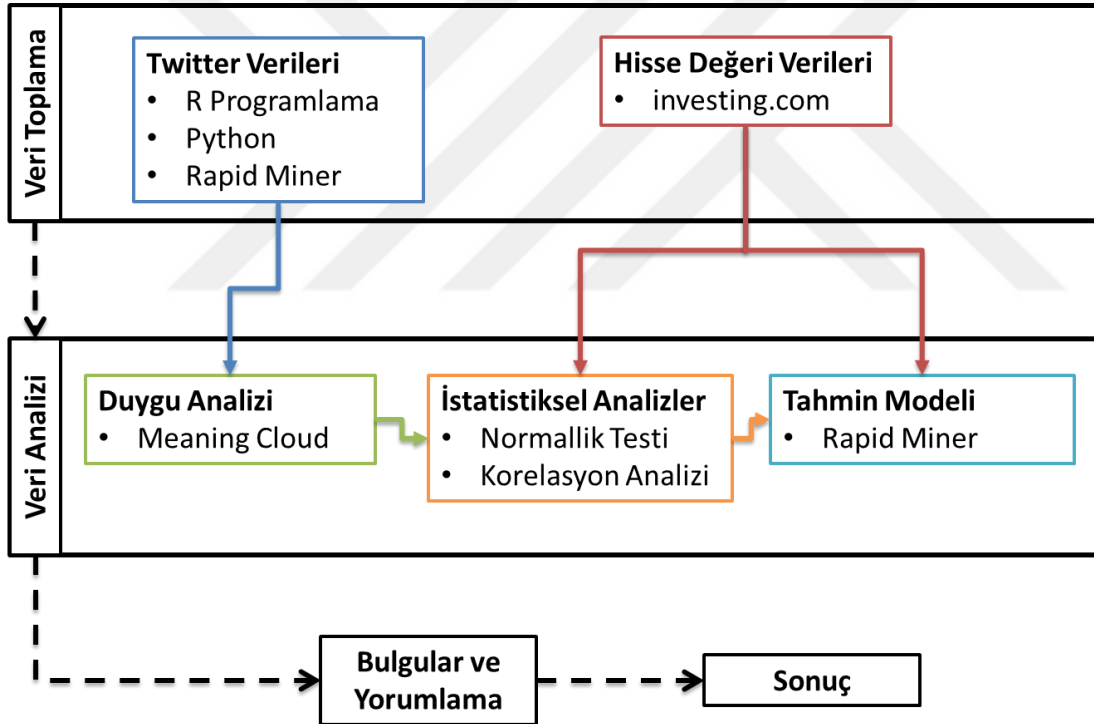
Ek olarak Das ve diğerleri [31], borsa tahmini gibi finansal kararların alınması için, bir şirketin hisselerinin potansiyel fiyatlarının tahmin edilmesi için çalışma yapmışlardır. Yaptıkları çalışmada Twitter üzerinden veri toplamışlardır. Twitter API ve Apache Flume veri toplama yazılımı kullanılarak geçmiş zamanlı verileri elde edip analiz etmişlerdir. Kurmuş oldukları hibrit modelde duygu analizlerinin borsa değerindeki yükseliş ve düşüşü için tahmini için güvenilir sonuçlar verdiği sonucuna varmışlardır.

Son olarak incelenen bir diğerk çalıřmada Yadav ve diğerkleri [32], gerçek zamanlı haber verilerinin, vadeli iřlem piyasasında yatırımcıların satın alma davranıřları üzerindeki tahmin edilebilirliđini duygu analizleri kullanarak incelemiřlerdir. Piyasanın duygu durumu ve yatırımcıların satın alma davranıřları, net alım baskısı (NBP: Net Buying Pressure) kullanılarak fiyat oluřumunun teklif-talep ařamasında yakalanmaktadır. NBP modellerinde herhangi bir önemli deđiřiklik “ilginç bir Pazar olayı” olarak tanımlanmaktadır. Gerçek zamanlı haber bařlıkları, piyasa bilgileri arasında bir gecikme olduđu varsayıp ve ilginç piyasa olayları kullanılarak satın alma davranıřlarını gecikmeli olarak otomatik etkileyeceđini saptamıřlardır. S&P 500 ve NIFTY endeksleri için piyasadaki satın alma davranıřını 5 dakikalık bir gecikme ile etkilediđi sonucuna varmıřlardır.

Yapılan literatür arařtırması sonucunda Sosyal medya verileri ile hisse senedi deđeri arasında bir iliřkinin varlıđı ile alakalı dünyada pek çok çalıřma yapılmaktadır. Türkiye’de Eliaçık ve Erdoğan’ın [26] yaptıđı çalıřmada mikrobloğlardaki paylařımların duygu polariteleri ile BİST-100 hisse senedi deđeri arasındaki korelasyon analizi sonucu pozitif yönlü, orta düzeyde iliřki olduđu saptanmıřtır. Yine aynı řekilde Türkiye’de Yıldırım ve Yüksel’in [30] yaptıđı çalıřmada ismi belirtilmeyen bir telekomünikasyon firması hakkında Twitter’dan elde ettikleri verileri duygu analizi yaparak sınıflandırmıřtır. Pozitif paylařım sayısı ile ilgili firma için hisse senedi deđerinin hareket yönü arasında negatif yönlü, orta düzeyde iliřki korelasyon analizi sonucuna göre saptanmıřtır. Fakat bu çalıřmalarda firmaların hisse senedi deđerinin sonraki günlerdeki ne olacađına dair tahmin modeli oluřturulmamıřtır.

Bizim yaptıđımız bu çalıřmada ise Türkiye’deki farklı bir sektör olan Havayolu tařımacılıđı sektöründeki **THY**’nin ve uluslararası boyutta rakibi olan **SHY**, **LHY**, **QHY**, **FHY** firmalarıyla ilgili hisse senedi deđerleri ile Twitter’daki paylařım sayıları arasındaki iliřki test edilmiř, korelasyon iliřkisi tespit edilen parametreler için bir tahmin modeli oluřturulmuřtur ve bu tahmin modelinde Rapid Miner programı aracılıđıyla Deep Learning kullanılarak, mevcut Twitter’daki paylařım sayılarından hareketle bir gün sonra hisse deđerinin artacađı veya azalacađıyla ilgili tahmin modeli geliřtirilmiřtir. THY ve SHY firması için aynı zamanda makro-ekonomik faktörlerde dikkate alınmıřtır. Bu makro-ekonomik faktörler, THY firması için

borsada THY Anonim Ortaklığı Hisse değeri (THYAO)'nin işlem gördüğü genel endeks değeri Borsa İstanbul 100 endeksi (BİST-100) [33], döviz olarak Dolar (USD/TRY) endeksi, kıymetli madenleri temsilen Altın(gr/TL) endeksi seçilmiştir. SHY firması için ise borsada SIAL [34]'ın işlem gördüğü genel endeks değeri Singapur Borsası(SGXL) [35], döviz olarak ise Dolar(USD/SGD) değerleri seçilmiştir. THY'ye ait hisse değeri tahmin modelinde Twitter verilerinden farklı olarak makro ekonomik faktörlerden BİST-100 [36], Dolar(USD/TRY) [37], Altın [38] endeks değerleri de kullanılmıştır. SHY firmasına ait hisse değeri tahmin modelinde ise Twitter verilerinden farklı olarak SGXL ve Dolar (USD/SGD) endeks değerleri de kullanılmıştır. Bu çalışma literatürdeki bu eksikliği gidermek için yapılmıştır. Bir sonraki bölümde çalışma metodolojisi Şekil 2.1'de gösterildiği gibi anlatılacaktır.



Şekil 2.1: Metodoloji

3. METODOLOJİ VE ANALİZLER

Bu çalışmada uluslararası havayolu firmaları hakkındaki sosyal medya verileri ile hisse değerleri arasındaki bir ilişki olup olmadığı test edilmiştir. Bu amaçla investing.com sitesinde hisse değeri verilerini takip edebileceğimiz ve Skytrax firması tarafından müşterilerin oyları ile 2019 Dünya Havayolu Firmaları listesinde bulunan ilk 30 firma arasından 5 hava yolu firması (THY, LHY, QHY, FHY, SHY) seçilmiştir. Sosyal medya verileri olarak bu firmalar hakkında Twitter sosyal medya platformundaki günlük bahsedilme (mention) sayıları toplanmıştır. Ayrıca firmaların günlük hisse senedi değerleri kaydedilmiştir. Ardından bu iki değişken arasında bir ilişki olup olmadığının tespit edilmesi amacıyla korelasyon analizi yapılmıştır. Hisse senedi değerleri ile Twitter’da bahsedilme sayıları arasında korelasyon bulunan firmalar için, Twitter bahsedilme sayılarından ertesi günkü hisse senedi değerini ve hareket yönünü (artacak mı azalacak mı) tahmin için makine öğrenmesi ile geliştirilen tahmin modeli kullanılmıştır.

Belirlenen hava yolu firmalarına ait değişkenlerin ilişkisini analiz etmek için aşağıdaki hipotezler geliştirilmiştir:

- Toplam Tweet Sayısı – Aynı Günlük Hisse Senedi Değeri,
- Pozitif Tweet Sayısı – Aynı Günlük Hisse Senedi Değeri,
- Negatif Tweet Sayısı – Aynı Günlük Hisse Senedi Değeri,
- Pozitif-Negatif farkı (Net Promoter Score-NPS) Tweet Sayısı – Aynı Günlük Hisse Senedi Değeri,
- Toplam Tweet Sayısı – Ertesi Günlük Hisse Senedi Değeri,
- Pozitif Tweet Sayısı – Ertesi Günlük Hisse Senedi Değeri,
- Negatif Tweet Sayısı – Ertesi Günlük Hisse Senedi Değeri,
- Pozitif-Negatif (Net Promoter Score-NPS) Tweet Sayısı – Ertesi Günlük Hisse Senedi Değeri,

her firma için IBM SPSS Statistics programı ile aralarındaki korelasyon incelenmiştir.

Bu hipotezleri test etmek için gerekli verilerin toplanma süreci aşağıda detaylı anlatılmıştır.

3.1 Veri Toplama

Hava yolu firmalarına ait hisse değeri verileri günlük olarak toplanmıştır. Bu süreçte tüm değerler anlık ve geçmişe yönelik bilgilere erişebileceğimiz investing.com [39] sitesinden alınmıştır. Investing sitesinden alınan verileri Şekil 3.1’de örnek olarak gösterilmiştir.



Şekil 3.2: Investing.com hisse değeri verileri

Analizleri yapılacak olan THY, LHY, QHY, FHY, SHY isimli uluslararası havayolu firmaları Skytrax tarafından müşterilerin oyları ile 2019 Dünya Havayolu Sıralamalarında bulunan ilk 30 firma arasından seçilmiştir [40]. Seçilen havayolu firmaları için Investing sitesinde günlük olarak takip edilebilen hisse senedi değeri verileri gün sonu hisse değerlerine göre excel dosyasına kaydedildi [34, 41-44]. Örnek olarak THYAO’ya ait gün sonu hisse değeri verileri Çizelge 3.1’de gösterilmiştir.

Çizelge 3.1: THY'ye ait hisse değeri verileri

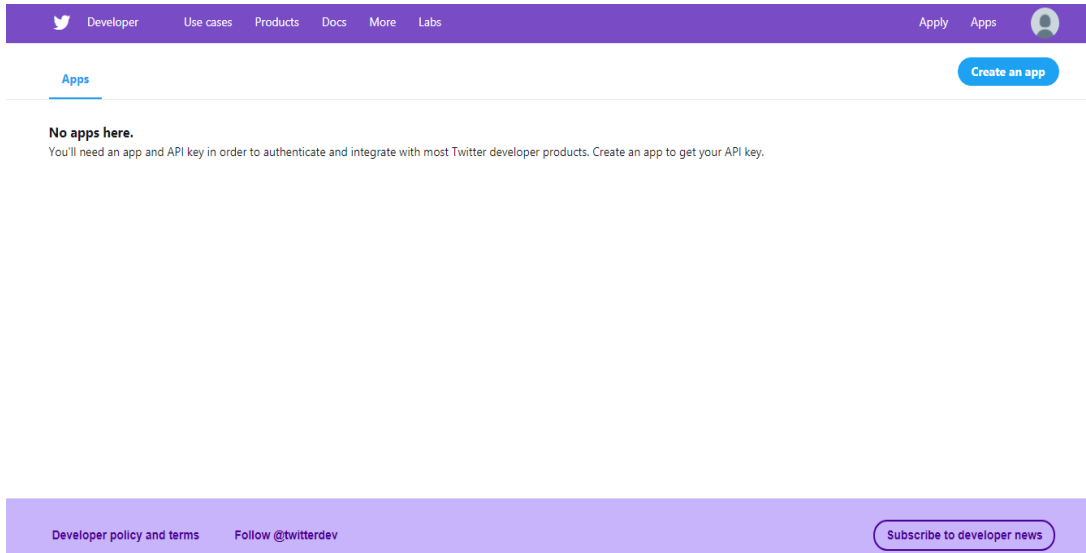
Tarih	Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri
01.10.2019	11,780	11,840
02.10.2019	11,840	11,750
03.10.2019	11,750	11,870
.	.	.
.	.	.
.	.	.
04.02.2020	14,390	14,610
05.02.2020	14,610	14,740
06.02.2020	14,740	14,740

Sosyal medya verilerini analiz etmek için ise Twitter platformundan firmalar hakkında, twitter kullacılarının paylaşmış olduğu duygu durumları, sosyal medya analitiği yöntemleri ile elde edilmiştir. Sosyal medyadan veri çekmek amacıyla yaygın olarak kullanılan 3 araç tespit edilmiştir. Bunlar, R programlama dili, Python programlama dili ve Rapid Miner veri madenciliği programlarıdır. Twitter'dan tarihsel verileri indirebilmek için developer.twitter'dan Premium:Full Archive (99\$/ay) üyeliği yapılarak Twitter API uygulaması oluşturulmuştur. Öncelikle R Programlama dili ile veri çekmeye çalışılmıştır. Fakat R programlama dilinde Twitter'dan veri çekmeye yarayan fonksiyonlarının ücretli Premium üyelikler için geçerli olmadığını tespit edilmiştir. Yani R programlama dili ile Premium: Full Archive üyeliği ile elde ettiğimiz API kodları ile tarihsel verileri elde edemeyeceğimiz anlaşılmıştır. Daha sonra Python programlama dili ile Premium üyelik ile elde edilen API kodları ile twitter verileri çekilmiştir. Python programlama dilinde Twitter'ın mevcut operatörleri kullanılmıştır [45]. Fakat her bir veri çekme isteği gönderme işleminde en fazla 500 adet tweet indirme sınırı twitter tarafından kısıtlandığı anlaşılmıştır [46]. Premium: Full Archive üyelikte aylık 100 istek gönderme hakkı tanındığı için en fazla 50 000 adet tweet indirebilecektik. Bu yöntem de bize hem yüksek maliyetli olacağı için hemde yeterli sayıda veri sağlayamayacağı için geçmiş verilere ulaşip analiz etmek yerine 4 aylık güncel verileri toplayarak 120 günlük toplam "tweet sayısı" tweet toplanması amaçlanmıştır. R programlama dili ve Python Programlama dili çalışmanın amacına uygun olarak bizlere yeterli sayıda veri sağlayamayacağı tespit edilmiştir. Rapid Miner veri madenciliği programı diğer programlama dillerine göre kullanıcı arayüzü daha etkin olduğu için ve duygu analizi yapmamıza imkan tanıyan Meaning Cloud eklentisi olduğu için Rapid Miner veri

madenciliği programı tercih edilmiştir. Rapid Miner veri madenciliği programı, developer twitter sitesinden Premium üyelik ile elde edilen Twitter API kodlarını kullanmaya ve twitter dan tarihsel veri çekme imkanı sunmadığı anlaşılmıştır. Bu nedenle başlangıçta Rapid Miner veri madenciliği programı tercih edilmemişti. Ama Premium üyeliğin çalışmanın amacına uygun olarak yeterli veri indirme imkanı sağlayamayacağından ve yüksek maliyetli olacağından dolayı Rapid Miner veri madenciliği programı kullanılmıştır. R programlama dili yerine Rapid Miner programını kullanma nedenimiz ise R programlama dili için kullanılan R Studio programında eklenti olarak bulunan Syuzhet duygu analizi eklentisi aracı, Rapid Miner veri madenciliği programı eklentisi olan Meaning Cloud'a göre çok etkin olmadığını gözlemlerimiz ile tespit edilmiştir. Yapılan araştırmalar ve gözlemler sonucunda Twitter'dan veri toplamak için Rapid Miner veri madenciliği programını kullanmaya devam edilmiştir.

3.2 Twitter API

Python ve R programlama dillerini kullanarak Twitter platformunda tweet indirebilmemiz için Twitter API kodları developer.twitter.com [47] sitesinden alınması gerekmektedir. Twitter API kodlarını temin etmek için öncelikle Twitter [48] sitesine üyelik yapılmıştır. Sonrasında ise geliştirici hesabı oluşturmak ve uygulama oluşturabilmemiz için developer.twitter.com [47] sitesine giriş yapılmıştır.



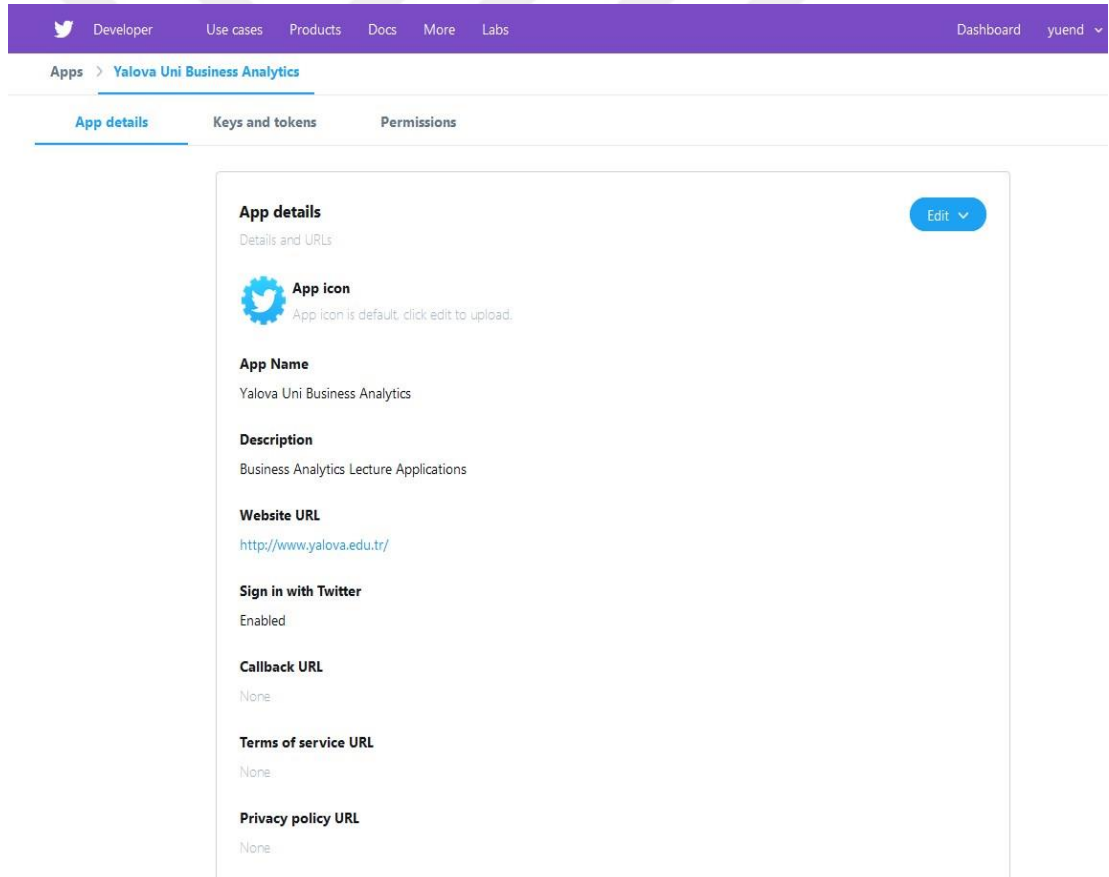
Şekil 3.3: Developer Twitter [49]

Twitter API kodu temin edebilmek için Şekil 3.2’de de gösterilen developer.twitter.com/en/apps siteden “Create an app” e giriş yapıp gerekli adımları tamamladıktan sonra uygulama oluşturulmuştur.



Şekil 3.4: Developer Twitter uygulaması

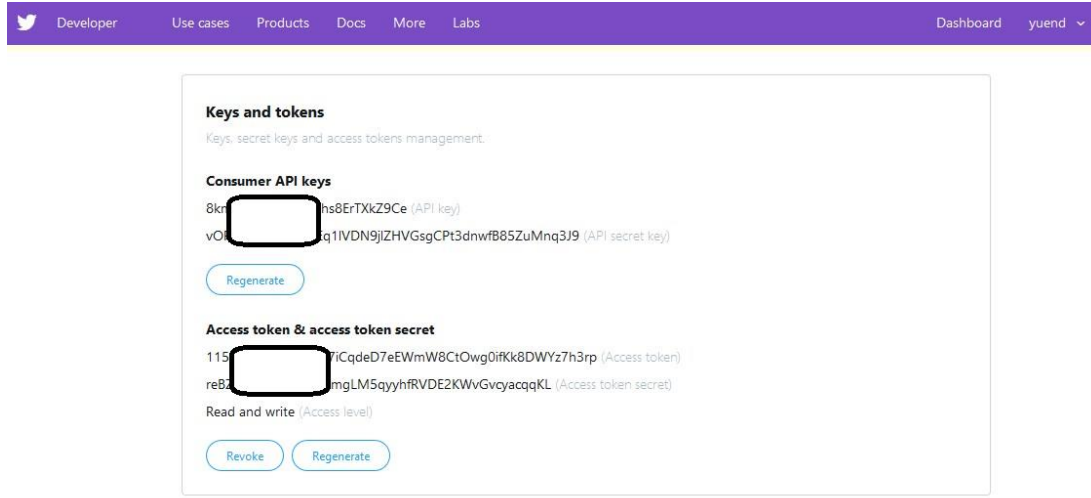
Oluşturulan uygulama Şekil 3.3’de gösterilmiştir. Uygulama oluşturulduktan sonra uygulama detaylarına giriş yapılarak Twitter API kodları Şekil 3.4 ve Şekil 3.5’da gösterilen “Key and Tokens” kısmından temin edilmiştir.



Şekil 3.5: Developer Twitter uygulama detayları

Developer Twitter sitesinden uygulama oluşturulduktan sonra Premium Full Archive üyeliği yapılmıştır. Premium: Full Archive üyeliğinin özelliği 99\$/ay ücret karşılığında özel API kodu vererek Twitter platformundan 2006 yılından itibaren

paylaşmış tüm twitter arşivini geliştirici hesabı oluşturan kullanıcılara ulaşma imkanı sunmaktadır.



Şekil 3.6: Twitter API kodları

3.3 R Programlama Dili

R programlama dili açık kaynaklı ve kullanıma hazır, istatistiksel analiz, grafik gösterimleri ve analizleri için kullanılan bir dil olarak tanımlanmaktadır. Beaujean [50], R programlamanın güçlü bir programlama dili olduğu, açık kaynaklı bir yazılım olduğunu belirtmiştir. Dinamik ve sürekli gelişime açık olan R yazılımı temel olarak istatistik yazılım geliştirmek için kullanılmaktadır [51]. Bununla birlikte diğer istatistiksel hesaplamaları yapabilen programların aksine R programlamanın arayüzü gelişmemiş, öğrenmesi ve kullanması diğer paket programlara göre uzun zaman almaktadır [51]. Bu özelliklerin yanı sıra R programlama dili ile Twitter platformunda veri indirebilme imkanı da sunmaktadır. Bu sebepten dolayı ilk olarak Twitter dan veri çekmek için R programlama kullanılmıştır.

R programlama dilini kullanabilmemiz için R-Studio programı bilgisayarımıza kurulmuştur. R-Studio programında, Şekil 3.6'de gösterildiği gibi developer Twitter sitesinden Premium:Full Archive üyeliği yapılarak elde edilen Twitter API kodları kullanılmıştır. Twitter platformundan R programlama dili ile yazılmış Şekil 3.7'de gösterilen kodlar kullanılarak Twitter platformundan tweetler indirilerek excel formatında kaydedilmiştir.

```

17 library(rworldmap)
18
19 dosya_adi<-"thytweets.csv"
20 data_dir <- "~/data"
21
22 consumer_key <- '8k...ErTXkZ9Ce'
23 consumer_secret <- 'vof...q11VDN9j1ZHVGSgCPT3dnwF885ZuMnq3J9'
24 access_token <- '11...7iCqdeD7eEwmw8CtOwg0ifkk8Dwyz7h3rp'
25 access_secret <- 're...jmgLM5qyyhfRVDE2KwvGvcyacqqKL'
26
27 setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
28
29

```

Şekil 3.7: Twitter API kodlarının R Studio’da kullanımı

```

30
31 listofNames <- c("@TurkishAirlines")
32
33 isFirst <- T
34 for (searchkeyword in listofNames){
35   print(sprintf("Searching for %s", searchkeyword))
36   allthy <- searchtwitter(searchkeyword, lang="en",
37                           resulttype="recent", n = 1000)
38   if(length(allthy) == 0){
39     print("Returned no results")r
40   }
41   else{
42     allthyDF <- twListtoDF(allthy)
43     print(sprintf("Returned %i results each having %i columns", nrow(allthyDF), ncol(allthyDF)))
44     ff <- file(paste0(data_dir,dosya_adi),open="at")
45     if(isFirst == T){
46       write.table(allthyDF, file=ff, quote=F, sep="," , col.names=T, row.names=F)
47       isFirst <- F
48     }else{
49       write.table(allthyDF, file=ff, quote=F, sep="," , col.names=F, row.names=F)
50     }
51     close(ff)
52   }
53   Sys.sleep(3) #sleep a while to prevent Twitter API rate limit error.
54 }
55
542 (Top Level)
R Script

```

Şekil 3.8: R programlama dilinde yazılan kodlar

R programlama dilinde yazılan kodlar ile Twitter API kodları kullanılarak yapılan veri indirme denemelerinde tarihsel olarak tweetler elde edilememiştir. Premium:Full Archive üyeliği ile elde edilen Twitter API kodları R Programlama kullanılarak istenilen tarihsel veri indirilemediğinden dolayı R Programlama dili kullanmaya devam edilmemiştir.

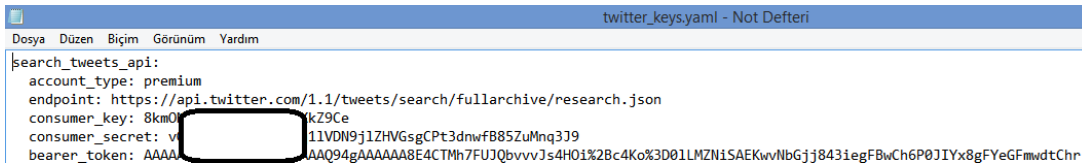
3.4 Python Programlama Dili

Python programlama dili, Hollandalı bir programcı olan Guido Van Rossum tarafından yazılan bir dildir. Python 1990 yılında geliştirilmeye başlanmasına rağmen, yani çok uzun geçmişi olmasa da bu dil öteki programlama dillerine göre kolay, ayrı bir derleyici program ihtiyacı olmadan kullanılabilmesinden dolayı son zamanlarda kullanımı yaygınlaşmaktadır. Python programlama dilinde metin, sayı, bilimsel veriler ve bilgisayarda kaydedilebilen diğer her şeyi işlemek için kullanılmaktadır. Python, Doğal Dil İşleme (NLP) algoritmalarını oluşturmak ve kullanmak için uygun ve kullanışlı bir programlama dilidir [52].

Çalışmamızda daha önceden temin edilen Twitter API kodları aracılığıyla Twitter platformundan belirli anahtar kelimeleri içeren tweetleri indirebilmek için Python programlama dilinde yazılmış kodlar kullanılmıştır. Öncelikle Python programlama dilini kullanabilmemiz için PyCharm editörü bilgisayarımıza kurulmuştur. Ardından çalışmanın amacına uygun Twitter platformundan veri indirebilmek için Search Twitter operatörleri [53] kullanılarak Python programlama dilinde yazılan kodlar ile veriler indirilerek JSON formatında kaydedilmiştir.

```
1 from searchtweets import gen_rule_payload, load_credentials, collect_results
2 import json
3
4 premium_search_args = load_credentials(filename="./twitter_keys.yaml",
5                                     yaml_key="search_tweets_api",
6                                     env_overwrite=False)
7
8 rule = gen_rule_payload("turkishairlines lang:en",
9                        results_per_call=500,
10                       from_date="2019-10-01", # UTC 2017-09-01 00:00
11                       to_date="2019-12-01", # UTC 2017-10-30 00:00
12                       )
13
14 tweets = collect_results(rule,
15                          max_results=5000,
16                          result_stream_args=premium_search_args)
17
18 with open('resultsthy8.json', 'w') as jsonFile:
19     json.dump(tweets, jsonFile)
```

Şekil 3.9: Twitter'dan veri çekme Python kodları



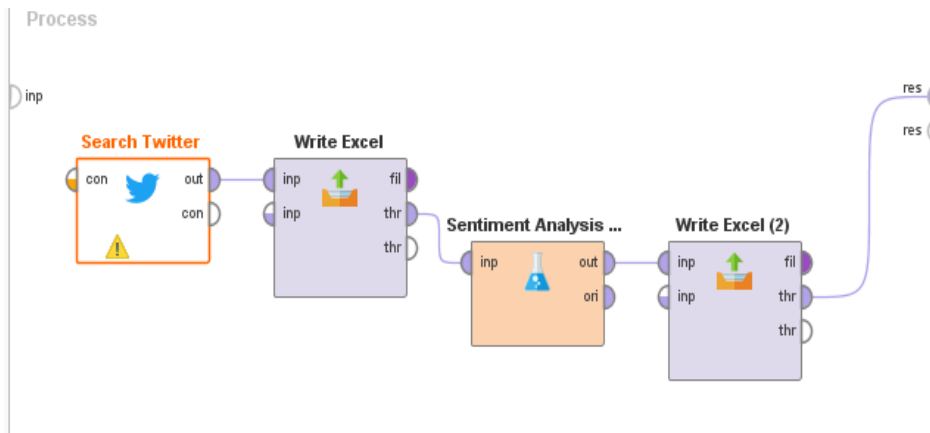
```
search_tweets_api:
  account_type: premium
  endpoint: https://api.twitter.com/1.1/tweets/search/fullarchive/research.json
  consumer_key: 8km0...kZ9Ce
  consumer_secret: v...11VDN9j1ZHVgsgCpt3dnwfb85ZuMnq3j9
  bearer_token: AAAA...AAQ94gAAAAAA8E4CTMh7FUJQbvuvvJs4HOi%2Bc4Ko%3D01LMZNiSAEKwvNbgjj843iegFBwCh6P0JIYx8gFYeGfmdtChr
```

Şekil 3.10: Twitter API Python kodları

3.5 Rapid Miner ve Meaning Cloud

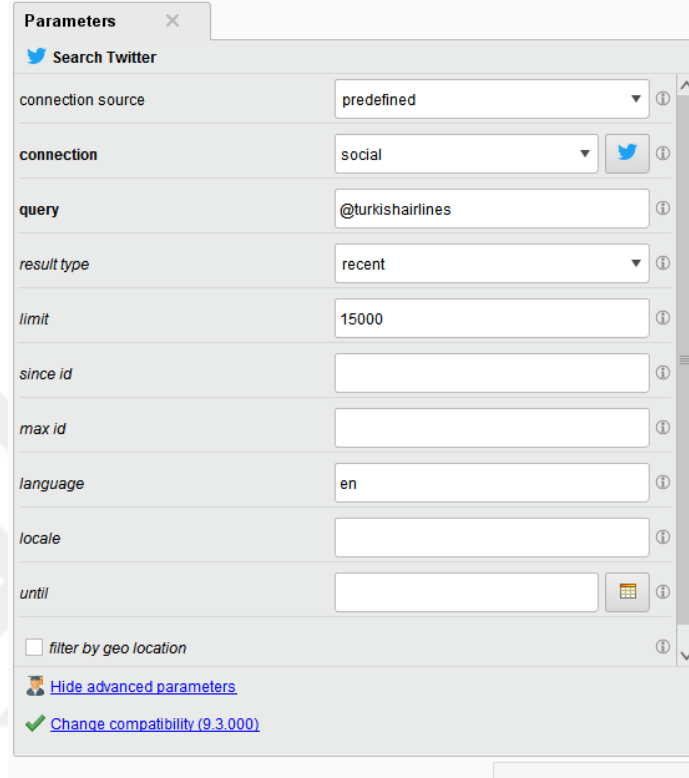
Rapid Miner veri madenciliği programı, makine öğrenmesi, metin madenciliği gibi konularda kullanıcıların ihtiyaçlarını karşılayabilecek Java tabanlı yazılım platformudur [54]. Tripathi ve diğerleri [55] gibi kişilerin Twitter platformunda paylaştığı fikirlere ilişkin duygu durumlarını analiz etmek için veri madenciliği programı Rapid Miner'ı kullanmışlardır. Yapmış olduğumuz bu çalışmada Twitter platformundan kullanıcıların paylaşımlarını indirmek için Rapid Miner programı kullanılmıştır.

Rapid Miner programını kullanabilmek için RapidMiner Studio Educational 9.4.001 [56] versiyonu bilgisayara kurulmuştur. Sonrasında Twitter'dan veri çekmemiz için Rapid Miner programına Search Twitter eklentisi entegre edilmiştir. İndirilen verilerin duygu analizleri yapılabilmesi için Meaning Cloud'un Sentiment Analysis eklentisi Rapid Miner programına entegre edilmiştir. Meaning Cloud sitesine üyelik yapıldıktan sonra lisans anahtarı duygu analizi yapılabilmesi için kullanılmıştır. Twitter'dan verilerin indirilmesi ve duygu analizlerinin yapılabilmesi için Şekil 3.11'de gösterildiği gibi çalışmanın modeli kurulmuştur. Şekil 3.11'de gösterilen modelde Search Twitter operatörü, Twitter platformundan kullanıcının belirlediği anahtar kelime, tweet sayısı ve arama yapılacak dil kriterlerine göre arama yapmasına olanak sağlamaktadır. Sentiment Analysis operatörü ise arama yapılan tweetlerin duygu analizlerini yapılabilmesine imkan sağlar. Write Excel operatörleri ise arama yapılan tweetlerin ve ardından duygu analizleri yapılan tweetlerin belirlenen excel dosyasına yazdırılması için kullanılmaktadır.



Şekil 3.12: Rapid Miner veri indirme modeli - 1

Rapid Miner programı ile Twitter platformundan veri indirme modelinin sağlıklı çalışması için Şekil 3.15’de Search Twitter operatörü için parametrelerin tanımlanması yapılmıştır. Şekil 3.16’da Search Twitter operatörü için Twitter platformuna bağlantı kurulması ve tanımlanması için kullanılmıştır.

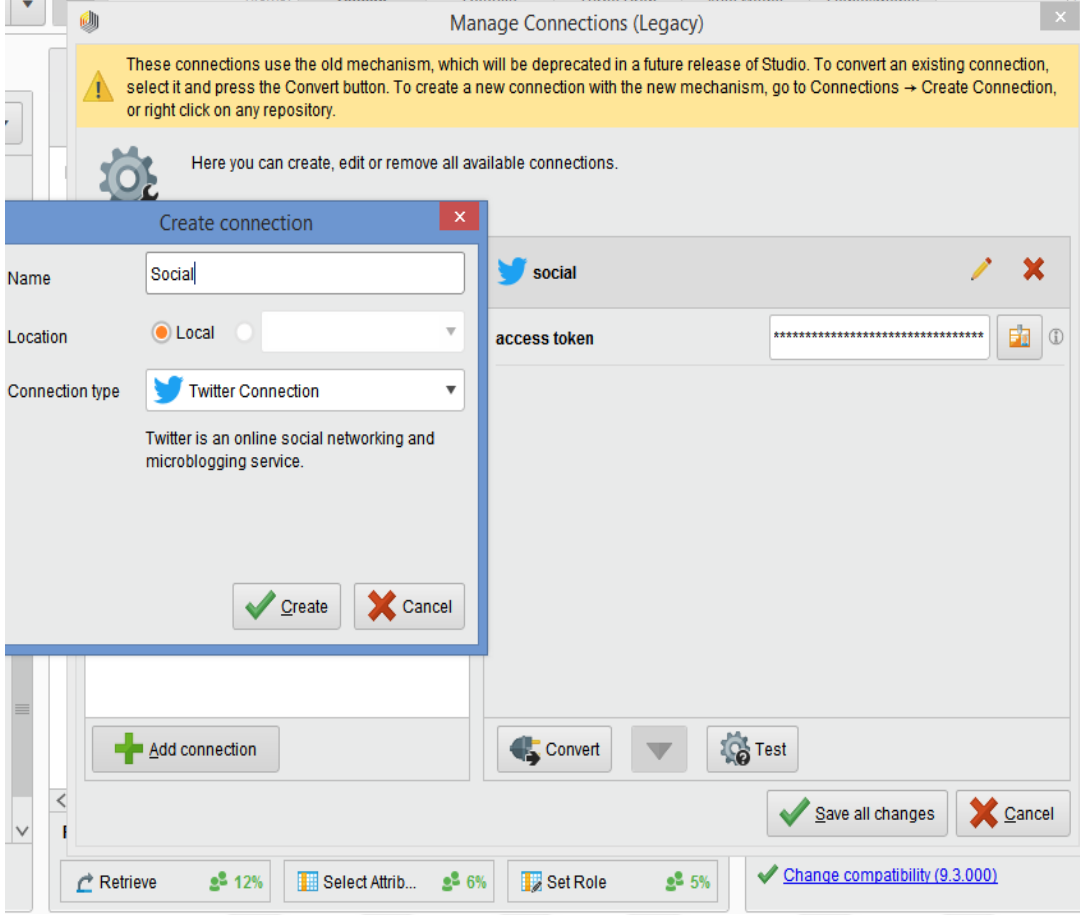


Parameter	Value
connection source	predefined
connection	social
query	@turkishairlines
result type	recent
limit	15000
since id	
max id	
language	en
locale	
until	
filter by geo location	<input type="checkbox"/>

[Hide advanced parameters](#)
[Change compatibility \(9.3.000\)](#)

Şekil 3.13: Search Twitter parametre ayarları

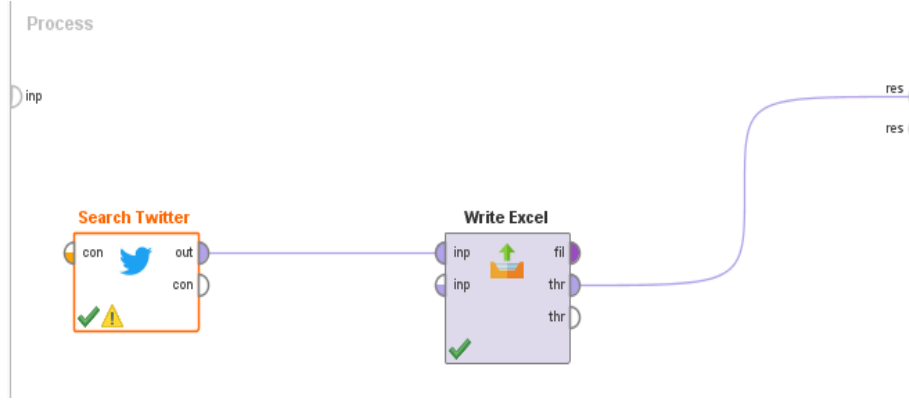
Şekil 3.12’de gösterilen Search Twitter Parametrelerinde “query” arama yapılacak tweetlerin içerisinde bulunan anahtar kelimeyi tanımlamak için, “result type” kriterinde “recent” seçeneği arama yapılan tarihten başlayarak eski tarihlere doğru arama yapılması için, “limit” kriteri ise arama yapılacak maksimum tweet sayısını kısıtlamamız için kullanılmıştır. “language” parametresi için İngilizce dilinde arama yapıldı ve bunun için “en” tanımlaması yapılmıştır.



Şekil 3.14: Twitter connection tanımlanması

Şekil 3.13’da gösterilen Twitter tanımlama işlemi için, Şekil 3.12’de gösterilen “connetcion” parametresindeki simgeye tıklandıktan sonra, Şekil 3.13’daki “add connection” ve ardından “create” butonuna tıklayıp ilgili adımları tamamladıktan sonra Twitter platformuna bağlantı yapılması sağlanmıştır.

Parametre ayarları tanımlandıktan sonra Şekil 3.11’de kurulan model her bir havayolu firması için ayrı ayrı çalıştırılarak veriler excel formatında kaydedilmiştir. Şekil 3.11’deki model ile indirilen tweetler duygu analizlerini de yaptığı için her bir veri indirme ve analiz işlemleri uzun sürdüğünden ve stabil olarak çalışmadığından dolayı Şekil 3.14’deki veri indirme modeli kullanılarak tweetler öncelikle excel dosyasında kaydedilmiştir.



Şekil 3.15: Rapid Miner veri indirme modeli – 2

Çizelge 3.2: Excel dosyasına kaydedilen tweetlerin örneği

Oluşturma Tarihi	Kullanıcı	Dil	Kaynak	Tweet
2019-10-01 00:00:29	गौयमा	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.Dobb y	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

Çizelge 3.2’de gösterilen Excel dosyasına kaydedilen tweetler, Şekil 3.15’de gösterildiği gibi Meaning Cloud eklentisi kullanılarak excel’de tweetlerin duygu analizleri yapılmıştır.

Text	Polarity	Agreement	Subjectivity	Confidence	Irony
Why does the blonde on the @TurkishAirlines advert say 'konnichi wa', the Japanese greeting, at the end of the tv advert, am I missing something?	N	AGREEMENT	SUBJECTIVE	100	NONIRONIC
@TurkishAirlines your customer support in North Macedonia is as hideous as trying to issue myself an invoice for a flight from December to no avail with in Skopje!			OBJECTIVE	100	NONIRONIC
RT @FinnPartners: Our FINN Travel team had an exciting & productive week representing clients @visitlauderdale, @...			OBJECTIVE	100	NONIRONIC
RT @MessiQuote: @TurkishAirlines: "Our commercial featuring Kobe Bryant the Decade." https://t.co/QMLr8...			OBJECTIVE	100	NONIRONIC
RT @MessiQuote: @TurkishAirlines: "Our commercial featuring Kobe Bryant the Decade." https://t.co/QMLr8...			OBJECTIVE	100	NONIRONIC
RT @timelessports : (2012) This Kobe Bryant and Messi commercial will forever be a classic. #WorldCupFinal2018 ??	NONE	AGREEMENT	OBJECTIVE	100	NONIRONIC
RT @loudbintfem: @TurkishAirlines I asked a member of your gate passport security team why passports had to be checked three times within...	NONE	AGREEMENT	OBJECTIVE	100	NONIRONIC
@TurkishAirlines I am trying to purchase ticket from Atlanta to Izmir with my ESAN dog, THY excepted as pet in the cabin but not as ESAN dog for same day and same flight. According USA federal regulation THY can not deny accepting Emotional Support Dog. Can someone give me better answer for this	P	DISAGREEMENT	SUBJECTIVE	86	NONIRONIC
RT @SecretFlying: Oslo, Norway to Bangkok, Thailand for only €378 roundtrip with @TurkishAirlines.					

Şekil 3.16: Meaning cloud excel eklentisi duygu analizi

THY, LHY, QHY, SHY firmaları için sadece İngilizce, FHY için ise İngilizce ve Fransızca dilinde yazılan tweetler indirilip, Excel’in Meaning Cloud eklentisi aracı ile duygu analizleri yapılabilmektedir. Bu dillerde analiz yapmamızın nedeni ise Meaning Cloud eklentisinin sınırlı dillerde duygu analizi yapabilmesinden dolayı kaynaklanmaktadır. Meaning Cloud eklentisi İspanyolca, İngilizce, Fransızca, İtalyanca, Portekizce, Katalonya dillerinde duygu analizleri yapılmasına imkân tanımaktadır. Bu sebepten dolayı firmalara ait tweetler indirilirken duygu analizi yapabileceğimiz diller tercih edilmiştir.

3.6 Tanımlayıcı İstatistikler

THY, LHY, QHY, FHY, SHY firmalarına ait her gün için tweet sayıları ve hisse değeri verileri Excel’de gösterilmiştir.

Çizelge 3.3: THY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NPS	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

Çizelge 3.3'te gösterilen THY'ye ait tweetler 01.10.2019-06.02.2020 tarihlerini kapsamaktadır. Bu tarihleri kapsayan günlerde toplam 56 064 adet tweet duygu analizleri yapılarak excel tablosunda analiz edilmek üzere düzenlenmiştir.

Çizelge 3.4: SHY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NPS	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

Çizelge 3.4'te gösterilen SHY'ye ait tweetler 01.10.2019-05.02.2020 tarihlerini kapsamaktadır. Bu tarihleri kapsayan günlerde toplam 32 232 adet tweet duygu analizleri yapılarak excel tablosunda analiz edilmek üzere düzenlenmiştir.

Çizelge 3.5: QHY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NPS	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

Çizelge 3.5'te gösterilen QHY'ye ait tweetler 01.10.2019-06.02.2020 tarihlerini kapsamaktadır. Bu tarihleri kapsayan günlerde toplam 79 863 adet tweet duygu analizleri yapılarak excel tablosunda analiz edilmek üzere düzenlenmiştir.

Çizelge 3.6: LHY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NPS	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

Çizelge 3.6'da gösterilen LHY'ye ait tweetler 01.10.2019-06.02.2020 tarihlerini kapsamaktadır. Bu tarihleri kapsayan günlerde toplam 42 837 adet tweet duygu analizleri yapılarak excel tablosunda analiz edilmek üzere düzenlenmiştir.

Çizelge 3.7: FHY'ye ait tweet sayıları ve hisse değeri verileri

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NPS	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

Çizelge 3.7'de gösterilen FHY'ye ait tweetler 10.10.2019-06.02.2020 tarihlerini kapsamaktadır. Bu tarihleri kapsayan günlerde toplam 25 768 adet tweet duygu analizleri yapılarak excel tablosunda analiz edilmek üzere düzenlenmiştir.

Sonuç olarak tüm firmalara ait toplam 236 764 adet tweet indirilip, duygu analizleri yapılmıştır. Elde edilen ve analiz edilecek olan bağımlı değişkenimiz (Ertesi Günkü Hisse Değeri) ve bağımsız değişkenlerimizin (Toplam, Pozitif, Negatif, Nötr, Tanımlanamayan, NPS tweet sayıları ile Dolar, Altın, Bist-100, SGXL/SGD, USD/SGD) metrik olarak tanımlanmıştır. Tüm firmalar için bir sonraki bölümde korelasyon testleri yapılmıştır.

3.7 Korelasyon Analizleri

THY firmasına ait Toplam, Pozitif, Negatif, NPS, USD/TRY, Altın, BİST-100, aynı günlük Hisse Değeri, ertesi günlük Hisse Değeri verileri için SPSS programında korelasyon analizi yapılmadan önce Çizelge 3.8'de gösterilen Normallik testi yapılmıştır.

Çizelge 3.8: THY'ye ait verilerin normallik testi

	Kolmogorov-Smirnov			Shapiro-Wilk		
	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)
Toplam	,347	118	,000	,445	118	,000
Pozitif	,371	118	,000	,438	118	,000
Negatif	,401	118	,000	,228	118	,000
NPS	,356	118	,000	,500	118	,000
Dolar	,153	118	,000	,908	118	,000
Altın/gr	,155	118	,000	,881	118	,000
Bist-100	,099	118	,007	,948	118	,000
Aynı Günkü Hisse Değeri	,232	118	,000	,883	118	,000
Ertesi Günkü Hisse Değeri	,237	118	,000	,882	118	,000

H₀: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlıdır.

H₁: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlı değildir.

Normallik testi sonucuna göre Sig. değerleri 0,05'ten küçük olduğu H₀ reddedilir yani değişkenler normal dağılmamıştır. Parametrik değildir(Nonparametric'tir). Elde edilen bu bilgiye göre değişkenler arasındaki korelasyon incelenirken Spearman's rho değerleri dikkate alınmıştır. Çizelge 3.9'de görüldüğü üzere THY firmasına ait korelasyonlar tespit edilmiştir.

Çizelge 3.9: THY'ye ait verilerin korelasyon analizi

		Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri	Toplam	Pozitif	Negatif	NPS	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.

SHY firmasına ait Toplam, Pozitif, Negatif, NPS, SGXL/SGD, USD/SGD, aynı günlük Hisse Değeri, ertesi günlük Hisse Değeri verileri için SPSS programında korelasyon analizi yapılmadan önce Çizelge 3.10'da gösterilen Normallik testi yapılmıştır. (SGXL/SGD: Singapur borsasına ait hisse değerinin Singapur Doları cinsinden değeri)

Çizelge 3.10: SHY'ye ait verilerin normallik testi

	Kolmogorov-Smirnov			Shapiro-Wilk		
	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)
Toplam	,411	114	,000	,170	114	,000
Pozitif	,429	114	,000	,133	114	,000
Negatif	,194	114	,000	,711	114	,000
NPS	,430	114	,000	,137	114	,000
SGXL/USD	,258	114	,000	,442	114	,000
USD/SGD	,117	114	,001	,959	114	,001
Aynı Günlük Hisse Değeri	,161	114	,000	,912	114	,000
Ertesi Günlük Hisse Değeri	,167	114	,000	,910	114	,000

H₀: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlıdır.

H₁: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlı değildir.

Normallik testi sonucuna göre Sig. değerleri 0,05'ten küçük olduğu H₀ reddedilir yani değişkenler normal dağılmamıştır. Parametrik değildir(Nonparametric'tir). Elde edilen bu bilgiye göre Korelasyonlar incelenirken Spearman's rho'nun korelasyon katsayısı değerlerine göre analizler yapılmıştır. Çizelge 3.11'de görüldüğü üzere SHY firmasına ait korelasyonlar tespit edilmiştir.

Çizelge 3.11: SHY'ye ait verilerin korelasyon analizi

		Aynı Günlük Hisse Değeri	Ertesi Günlük Hisse Değeri	Toplam	Pozitif	Negatif	NPS	SGXL/SGD	USD/SGD	
Spearman's rho	Aynı Günlük 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 Günlük 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.

LHY firmasına ait Toplam, Pozitif, Negatif, NPS, aynı günlük Hisse Değeri, ertesi günlük Hisse Değeri verileri için SPSS programında korelasyon analizi yapılmadan önce Çizelge 3.12'de gösterilen Normallik testi yapılmıştır.

Çizelge 3.12: LHY'ye ait verilerin normallik testi

	Kolmogorov-Smirnov			Shapiro-Wilk		
	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)
Toplam	,295	121	,000	,440	121	,000
Pozitif	,264	121	,000	,648	121	,000
Negatif	,255	121	,000	,643	121	,000
NPS	,109	121	,001	,913	121	,000
Aynı Günkü Hisse Değeri	,116	121	,000	,946	121	,000
Ertesi Günkü Hisse Değeri	,116	121	,000	,946	121	,000

H₀: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlıdır.

H₁: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlı değildir.

Normallik testi sonucuna göre Sig. değerleri 0,05'ten küçük olduğu H₀ reddedilir yani değişkenler normal dağılmamıştır. Parametrik değildir(Nonparametric'tir). Elde edilen bu bilgiye göre Korelasyonlar incelenirken Spearman's rho'nun korelasyon katsayısı değerlerine göre analizler yapılmıştır. Çizelge 3.13'te görüldüğü üzere LHY firmasına ait korelasyonlar tespit edilmiştir.

Çizelge 3.13: LHY'ye ait verilerin korelasyon analizi

		Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri	Toplam	Pozitif	Negatif	NPS	
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.

QHY firmasına ait Toplam, Pozitif, Negatif, NPS, aynı günkü Hisse Değeri, ertesi günkü Hisse Değeri verileri için SPSS programında korelasyon analizi yapılmadan önce Çizelge 3.14'te gösterilen Normallik testi yapılmıştır.

Çizelge 3.14: QHY'ye ait verilerin normallik testi

	Kolmogorov-Smirnov			Shapiro-Wilk		
	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)
Toplam	,284	112	,000	,550	112	,000
Pozitif	,267	112	,000	,609	112	,000
Negatif	,293	112	,000	,520	112	,000
NPS	,217	112	,000	,766	112	,000
Aynı Günkü Hisse Değeri	,138	112	,000	,913	112	,000
Ertesi Günkü Hisse Değeri	,138	112	,000	,914	112	,000

H₀: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlıdır.

H₁: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlı değildir.

Normallik testi sonucuna göre Sig. değerleri 0,05'ten küçük olduğu H₀ reddedilir yani değişkenler normal dağılmamıştır. Parametrik değildir(Nonparametric'tir). Elde edilen bu bilgiye göre Korelasyonlar incelenirken Spearman's rho'nun korelasyon katsayısı değerlerine göre analizler yapılmıştır. Çizelge 3.15'te görüldüğü üzere QHY firmasına ait korelasyonlar tespit edilmiştir.

Çizelge 3.15: QHY'ye ait verilerin korelasyon analizi

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

** Korelasyon 0,01 düzeyinde anlamlıdır.
* Korelasyon 0,05 düzeyinde anlamlıdır.

FHY firmasına ait Toplam, Pozitif, Negatif, NPS, aynı günkü Hisse Değeri, ertesi günkü Hisse Değeri verileri için SPSS programında korelasyon analizi yapılmadan önce Çizelge 3.16'da gösterilen Normallik testi yapılmıştır.

Çizelge 3.16: FHY'ye ait verilerin normallik testi

	Kolmogorov-Smirnov			Shapiro-Wilk		
	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)	İstatistik	Veri Sayısı	Anlamlılık Düzeyi (Sig.)
Toplam	,314	113	,000	,329	113	,000
Pozitif	,202	113	,000	,726	113	,000
Negatif	,378	113	,000	,203	113	,000
NPS	,335	113	,000	,371	113	,000
Aynı Günkü Hisse Değeri	,175	113	,000	,881	113	,000
Ertesi Günkü Hisse Değeri	,173	113	,000	,883	113	,000

H₀: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlıdır.

H₁: Değişkenlerin hepsi %95 güvenle veriler normal dağılımlı değildir.

Normallik testi sonucuna göre Sig. değerleri 0,05'ten küçük olduğu H₀ reddedilir yani değişkenler normal dağılmamıştır. Parametrik değildir(Nonparametric'tir). Elde edilen bu bilgiye göre Korelasyonlar incelenirken Spearman's rho'nun korelasyon katsayısı değerlerine göre analizler yapılmıştır. Çizelge 3.17'de görüldüğü üzere FHY firmasına ait korelasyonlar tespit edilmiştir.

Çizelge 3.17: FHY'ye ait verilerin korelasyon analizi

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

** Korelasyon 0,01 düzeyinde anlamlıdır.
* Korelasyon 0,05 düzeyinde anlamlıdır.

Tüm firmalara ait yapılan korelasyon analizleri sonucunda en yüksek korelasyon değerlerine ulaştığımız SHY firması için ve esas incelemek istenen Türkiye borsasına ait şirket olan THY firması için tahmin modelleri Rapid Miner programı kullanılarak kurulmuştur.

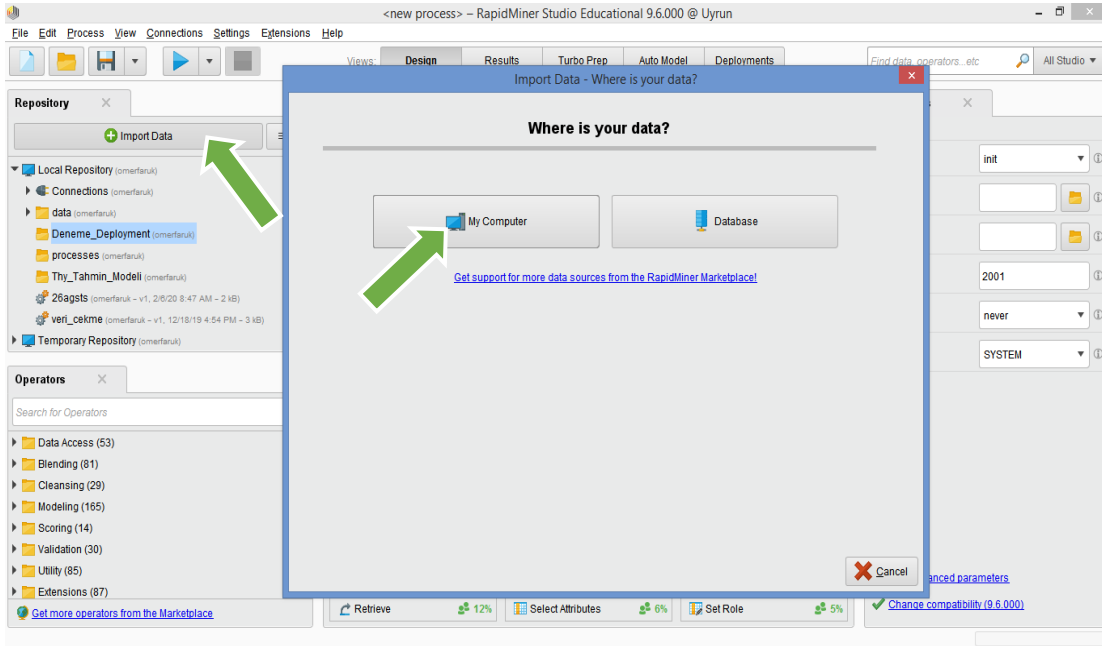
3.8 Auto Model ile Tahmin Modeli Kurma

THY ve SHY firmaları için Rapid Miner programında Auto Model özelliği kullanılarak tahmin modelleri kurulmuştur. Auto Model'e yüklenen verilerin makine öğrenmesi analizleri sonucunda programın bizlere sunduğu korelasyon ve hata değerlerine göre en başarılı tahmin modeli kullanılmıştır. Firmalara ait kurulan tahmin modelleri detaylı bir şekilde anlatılmıştır.

THY firması için Rapid Miner programında Auto Model özelliği kullanılarak tahmin modeli oluşturulmuştur. Tahmin modeli oluşturulurken Rapid Miner Auto Model kullanımını da anlatılmıştır. Başlangıç olarak Çizelge 3.18'de gösterilen 100 günlük veri, öğrenme verisi olarak Rapid Miner programına yüklenmiştir.

Çizelge 3.18: THY için yüklenen örnek veriler

	Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NPS	Dolar	ALTIN/gr	BİST-100	Aynı Günkü Hisse Değeri	Ertesi Günkü Hisse Değeri
1	1.10.2019	6152	1904	2232	997	1019	-328	5,7231	272,479	103876,56	11,780	11,840
2	2.10.2019	475	148	179	79	69	-31	5,7020	275,102	103509,22	11,840	11,750
3	3.10.2019	962	157	107	46	652	50	5,6947	276,232	102784,33	11,750	11,870
.
.
.
98	12.1.2020	64	28	14	6	16	14	43,75	21,88	9,38	25,00	5,8777
99	13.1.2020	66	30	14	3	19	16	45,45	21,21	4,55	28,79	5,8671
100	14.1.2020	104	60	19	2	23	41	57,69	18,27	1,92	22,12	5,8819



Şekil 3.17: Rapid Miner'a verilerin yüklenmesi

Şekil 3.16'de gösterildiği gibi programa yüklenecek excel dosyası bilgisayardan seçilerek yüklenmiştir. Gerekli adımlar tamamlandıktan sonra yüklenen verilerin özizlemesi Şekil 3.17'de gösterildiği gibidir.

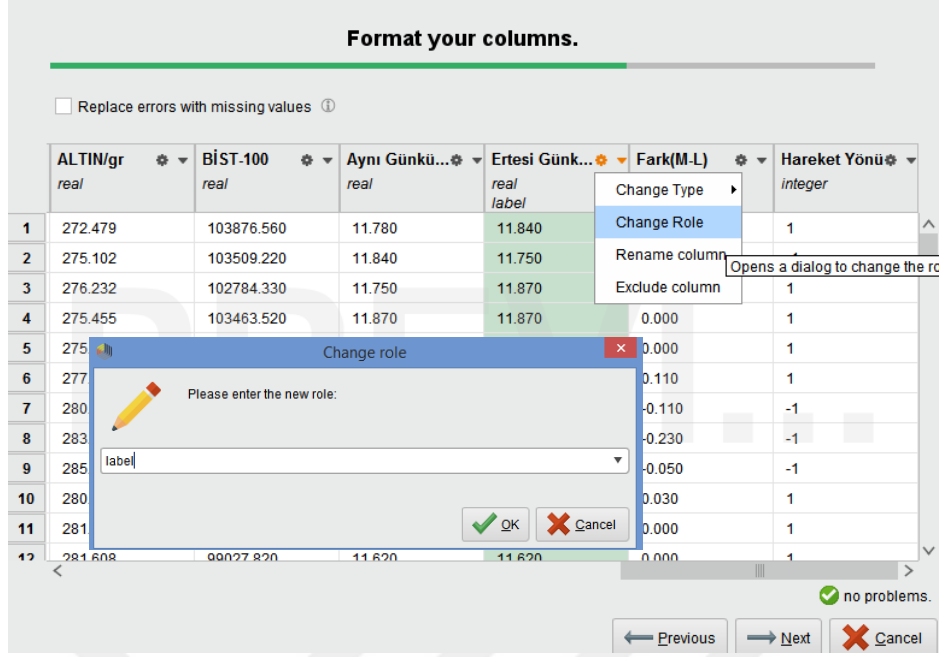
Select the cells to import.

Sheet: Sonraki Günü Tahmini Hareket Yö Cell range: :1:R101 Select All Define hea... 1

	A	B	C	D	E	F	G	H	I	J
1	Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımla...	NPS	Pozitif %	Negatif %	Nötr %
2	Oct 1, 20...	6152.000	1904.000	2232.000	997.000	1019.000	-328.000	30.949	36.281	16.206
3	Oct 2, 20...	475.000	148.000	179.000	79.000	69.000	-31.000	31.158	37.684	16.632
4	Oct 3, 20...	962.000	157.000	107.000	46.000	652.000	50.000	16.320	11.123	4.782
5	Oct 4, 20...	364.000	116.000	55.000	31.000	162.000	61.000	31.868	15.110	8.516
6	Oct 5, 20...	319.000	174.000	49.000	16.000	80.000	125.000	54.545	15.361	5.016
7	Oct 6, 20...	163.000	79.000	41.000	8.000	35.000	38.000	48.466	25.153	4.908
8	Oct 7, 20...	230.000	144.000	42.000	13.000	31.000	102.000	62.609	18.261	5.652
9	Oct 8, 20...	206.000	105.000	44.000	24.000	33.000	61.000	50.971	21.359	11.650
10	Oct 9, 20...	295.000	155.000	36.000	11.000	93.000	119.000	52.542	12.203	3.729
11	Oct 10, 2...	4995.000	1291.000	2905.000	361.000	438.000	-1614.000	25.846	58.158	7.227
12	Oct 11, 2...	2262.000	376.000	1439.000	157.000	290.000	-1063.000	16.622	63.616	6.941
13	Oct 12, 2...	569.000	116.000	332.000	20.000	101.000	-216.000	20.387	58.348	3.515

← Previous → Next ✖ Cancel

Şekil 3.18: Yüklenecek verilerin seçilmesi



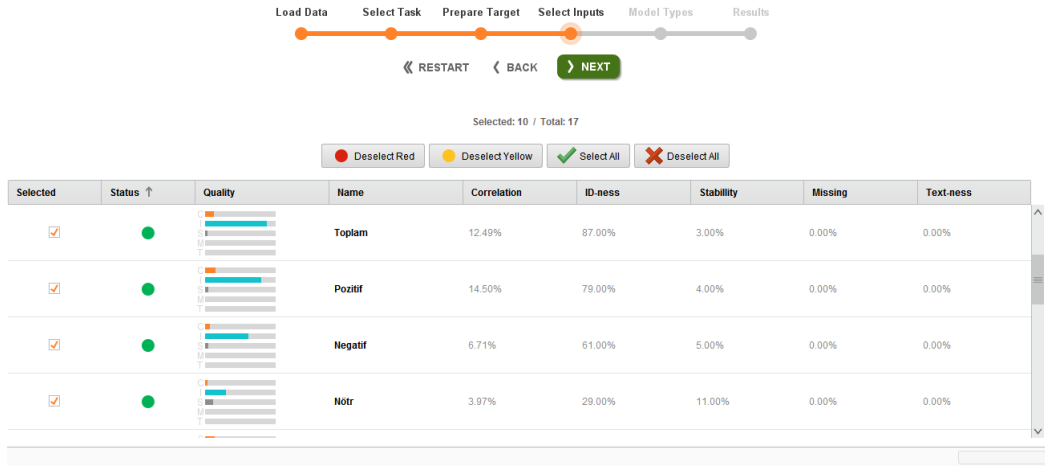
Şekil 3.19: Tahmin edilecek değişkenin rolünü değiştirme

Şekil 3.18’de tahmin etmek istediğimiz hisse değerinin ertesi günkü değişkeninin rolünü “label” olarak değiştirildi ve ilgili adımlar tamamlandıktan sonra Şekil 3.19’daki Auto Model seçeneği ile veriler makine öğrenmesi özelliğinde kullanıldı.

Row No.	Stocks of Th...	Date	Total Mention	Positive	Negative	Neutral	None
1	11.840	Oct 1, 2019	6152	1904	2232	997	1019
2	11.750	Oct 2, 2019	475	148	179	79	69
3	11.870	Oct 3, 2019	962	157	107	46	652
4	11.870	Oct 4, 2019	364	116	55	31	162
5	11.870	Oct 5, 2019	319	174	49	16	80
6	11.980	Oct 6, 2019	163	79	41	8	35
7	11.870	Oct 7, 2019	230	144	42	13	31
8	11.640	Oct 8, 2019	206	105	44	24	33
9	11.590	Oct 9, 2019	295	155	36	11	93
10	11.620	Oct 10, 2019	4995	1291	2905	361	438
11	11.620	Oct 11, 2019	2262	376	1439	157	290
12	11.620	Oct 12, 2019	569	116	332	20	101
13	10.860	Oct 13, 2019	1719	1359	252	26	82
14	11.060	Oct 14, 2019	873	195	129	26	523

Şekil 3.20: THY'ye ait veriler ve otomodel

Şekil 3.20’de tahmin edilecek ertesi günkü hisse değeri için kullanılacak değişkenler seçilmiştir. Bu değişkenler Toplam, Pozitif, Negatif, Nötr, Tanımlanmayan tweet sayıları ile Tarih, NPS, Dolar, Altın, Bist-100 değeri verileri kullanılmıştır.



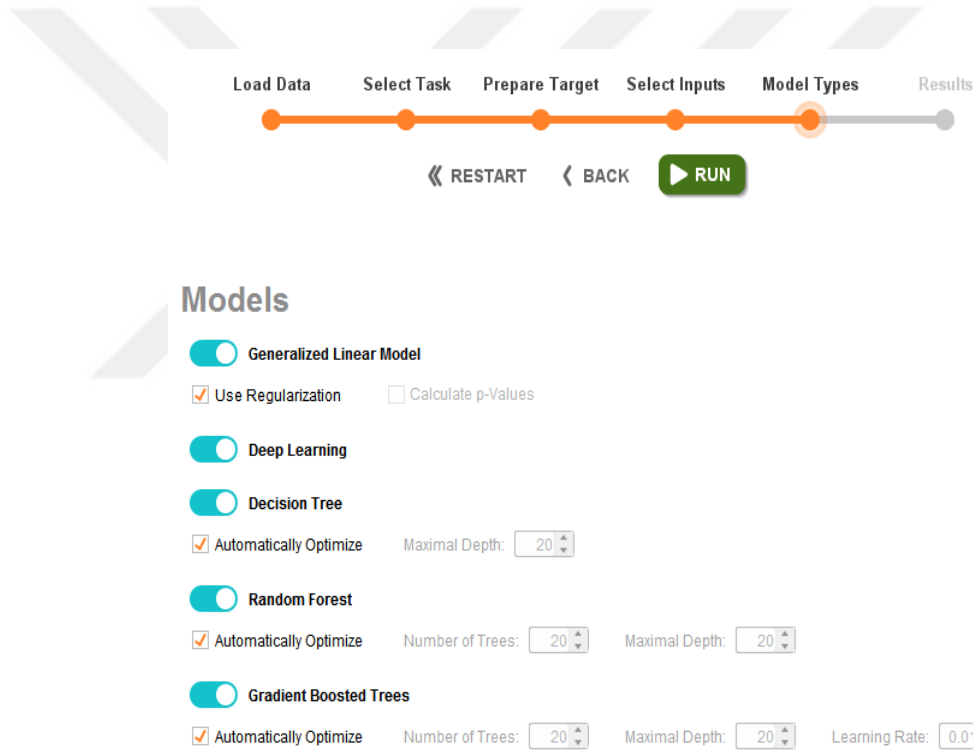
Şekil 3.21: Değişkenlerin seçilmesi

Şekil 3.21’de makine öğrenmesi ile çalışılması istenilen model tipleri seçilerek verilerin çalıştırılması sağlandı. Kullanıcıya sunulan Modelleri şu şekilde açıklayacak olursak;

- Generalized Linear Model [57], geleneksel doğrusal modellerin bir uzantısıdır. Bu algoritma, verilerin tahmin edilebilirliğini en üst düzeye çıkararak genelleştirilmiş doğrusal modellerine uyarlar. Model kurulması ve tahminlerin hesaplanması son derece hızlıdır.
- Deep Learning Model [58], geri yayılım kullanılarak stokastik eğitim inişi ile eğitilmiş çok katmanlı ileri beslemeli yapay sinir ağına dayanır. Her hesaplama düğümü, global model parametrelerinin bir kopyasını yerel veriler üzerinde çoklu iş parçacıklı (eşzamansız olarak) ile eğitir ve ağ genelinde model ortalaması yoluyla periyodik olarak küresel modele katkıda bulunur.
- Decision Tree Model [59], bir sınıfa bağlı değerleri veya sayısal bir hedef değer tahmini hakkında bir karar oluşturmayı amaçlayan düğümlerin toplanması gibidir. Her düğüm, belirli bir özellik için bölme kuralını temsil eder. Sınıflandırma için bu kural farklı sınıflara ait değerleri ayırır, regresyon için hatayı seçilen parametre kriteri için en uygun şekilde azaltmak amacıyla ayırır.
- Random Forest Model [60], ağaç sayısı parametresi tarafından belirtilen belirli sayıda rasgele ağaç grubudur. Bu ağaçlar, giriş portunda sağlanan SampleSet’in önyükleme alt kümeleri üzerinde eğitilir. Sınıflandırma için

kurali farklı sınıflara ait verileri ayırmak iken, regresyon tahmin tarafından yapılan hatayı azaltmak amacıyla bunları ayırır.

- Gradient Boosted Trees Model [61], regresyon veya sınıflandırma ağacı modellerinin bir grubudur. Her ikisi de, kademeli olarak geliştirilmiş tahminler yoluyla öngörücü sonuçlar elde eden ileri öğrenme yöntemleridir.
- Support Vector Machine Model [62], Stefan Rueping tarafından mySVM'nin Java uygulamasını kullanır. Bu öğrenme yöntemi hem regresyon hem de sınıflandırma için kullanılabilir. Birçok öğrenme görevi için hızlı bir algoritma ve iyi sonuçlar sağlar. mySVM doğrusal veya karesel ve hatta asimetrik kayıp işlevleriyle çalışır.



Şekil 3.22: Tahmin modeli seçenekleri

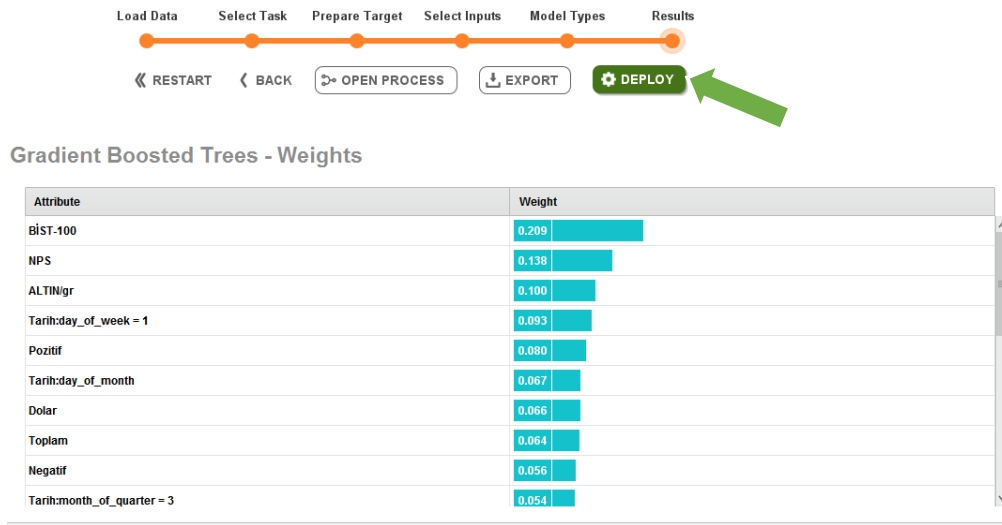
Öğrenme verileri kullanılarak çalıştırılan modellerin korelasyon ve hata değerleri ile ilgili analiz sonuçları Çizelde 3.19'da gösterilmiştir. Tahmin modellerinden yüksek korelasyon ve hata değeri en düşük olan “Gradient Boosted Trees” Modeli seçilerek tahmin modeli kurulmuştur.

Çizelge 3.19: THY için tahmin modeli analiz sonuçları

Tahmin Modeli	Korelasyon	Mutlak Hata
Generalized Linear Model	0,970	0,219
Deep Learning	0,929	0,347
Decision Tree	0,970	0,211
Random Forest	0,985	0,197
Gradient Boosted Trees*	0,985	0,156
Support Vector Machine	0,957	0,234

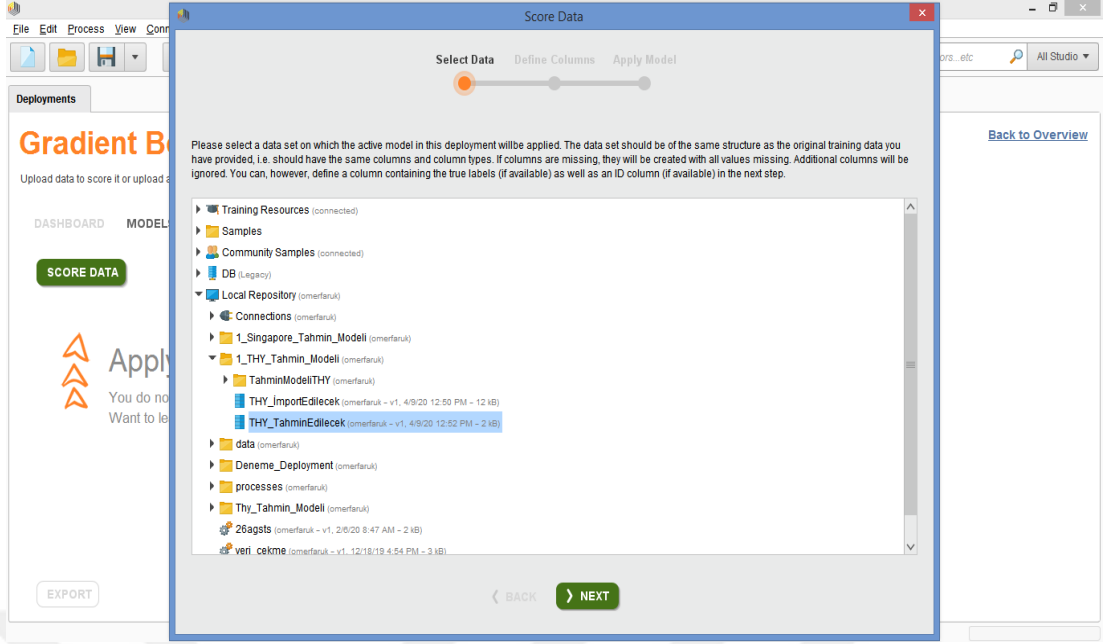
* En iyi çözüm

Seçilen “Gradient Boosted Trees” modeli Şekil 3.22’de gösterilen Deployment seçeneği ile tahmin modeli kaydedilmiştir. “Deploy” seçeneğine tıkladıktan sonra “name” ve “location” bilgileri otomatik olarak doldurulmaktadır. İlgili adımları ilerleterek tamamladıktan sonra kurulan tahmin modelini ekleyeceğimiz dosya oluşturulmuştur.



Şekil 3.23: Tahmin modelinin kaydedilmesi

“Deployment” kısmında gerekli adımlar ilerledikten sonra tahmin modelimiz kaydedilmiş olacaktır. Sonraki adımda ise tahmin edilecek 18 günlük veri Rapid Miner programına yüklenmiştir.



Şekil 3.24: Tahmin edilecek verinin seçilmesi

Şekil 3.23'te önceden yüklenen tahmin edilmesini istediğimiz 18 günlük veri "Scoring" özelliği kullanılmış ve tahmin denemesi tamamlanmıştır. Şekil 3.24'te tahmin edilen değerler ile gerçekleşen değerleri karşılaştırmalı olarak gösterilmiştir.

Thy Gradient Boosted Trees: Scoring [Back to Overview](#)

Upload data to score it or upload actual outcomes for performance calculations. ⓘ

DASHBOARD MODELS PERFORMANCE DRIFTS SIMULATOR SCORING ALERTS INTEGRATIONS

SCORE DATA

Row No.	Tarih	prediction(Ertesi Günlük Hisse Değeri)	Ertesi Günlük Hisse Değeri	Toplam	Pozitif	Negatif	Noir	Tanımlanma...	NPS	Dolar
1	Jan 15, 2020	14.636	15.080	76	34	19	2	21	15	5.876
2	Jan 20, 2020	14.244	14.880	2656	2475	89	27	65	2386	5.913
3	Jan 21, 2020	14.286	14.720	4073	3906	86	11	70	3820	5.938
4	Jan 22, 2020	14.242	14.240	1952	1854	40	5	53	1814	5.922
5	Jan 23, 2020	14.271	14.540	1853	1745	51	14	43	1694	5.934
6	Jan 24, 2020	14.205	14.540	1600	1508	40	10	42	1468	5.939
7	Jan 25, 2020	14.379	14.540	204	125	26	5	48	99	5.939
8	Jan 26, 2020	14.424	14.100	128	53	27	3	45	26	5.939

Şekil 3.25: Tahmin edilen veriler

Rapid Miner ile yapılan tahmin modeli oluşturulurken, 100 günlük veri programın tahmin edilecek değişkeni, tahmin edebilmesi için makine öğrenmesi yapmaktadır. Program verileri belirli makine öğrenmesi metodlarıyla öğrenip, en az hata ve yüksek kolerasyon olan modeli kullanıcıya sunmaktadır. THY firmasına ait verileri Rapid Miner programının önerdiği en başarılı model olan "Gradient Boosted Trees" Modeli ile 18 günlük Tweet sayıları ile ilgili veriler ve makro değişken verileri programa yüklenerek Çizelge 3.20'de gösterildiği gibi gerçekleşen ve tahmin edilen değerleri elde edilmiştir.

Çizelge 3.20: Tahmin edilen ve gerçekleşen hisse değer verileri (THY)

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NPS	Dolar	ALTIN/gr	BIST-100	Tahmin Edilen (Ertesi Günkü Hisse Değeri)	Gerçekleşen (Ertesi Günkü Hisse Değeri)	
1	2020-01-15	76	34	19	2	21	15	5,8758	294,1020	120938,78	14,64	15,08
2	2020-01-20	2656	2475	89	27	65	2386	5,9132	296,8890	122619,77	14,24	14,88
3	2020-01-21	4073	3906	86	11	70	3820	5,9376	296,9960	123556,10	14,29	14,72
4	2020-01-22	1952	1854	40	5	53	1814	5,9222	297,2990	122747,08	14,24	14,24
5	2020-01-23	1853	1745	51	14	43	1694	5,9344	297,6150	122540,22	14,27	14,54
6	2020-01-24	1600	1508	40	10	42	1468	5,9391	300,0830	122141,75	14,21	14,54
7	2020-01-25	204	125	26	5	48	99	5,9391	300,0830	122141,75	14,38	14,54
8	2020-01-26	128	53	27	3	45	26	5,9391	302,5910	122141,75	14,42	14,10
9	2020-01-27	286	70	36	5	175	34	5,9440	302,1340	120149,70	14,25	14,08
10	2020-01-29	220	131	38	8	43	93	5,9567	302,2450	119689,48	14,22	13,83
11	2020-01-30	394	235	74	13	72	161	5,9725	302,7500	119849,65	14,28	13,68
12	2020-01-31	322	209	36	5	72	173	5,9822	305,6120	119140,08	14,25	13,68
13	2020-02-01	170	107	24	2	37	83	5,9822	305,6120	119140,08	14,40	13,68
14	2020-02-02	149	90	24	4	31	66	5,9822	305,5460	119140,08	14,41	13,84
15	2020-02-03	924	547	116	19	242	431	5,9844	303,8070	119242,80	14,25	14,39
16	2020-02-04	217	106	60	10	41	46	5,9801	298,8030	122136,17	14,29	14,61
17	2020-02-05	364	258	47	13	46	211	5,9827	299,2970	122320,77	14,28	14,74
18	2020-02-06	236	179	21	9	27	158	5,9874	301,4710	122283,87	14,42	14,47

SHY firması için Rapid Miner programında makine öğrenmesi özelliği kullanılarak tahmin modeli oluşturulmuştur. Tahmin modeli oluşturulurken THY için anlatılan Tahmin Modeli oluşturma adımları SHY'ye ait 100 günlük veri öğrenme verisi olarak Rapid Miner programına import edilmiştir.

SHY için input edilecek değişkenler Toplam, Pozitif, Negatif, Nötr, Tanımlanmayan tweet sayısı, Tarih, NPS, SGXL/SGD, USD/SGD değeri verileri kullanılmıştır.

Çizelge 3.21: SHY için tahmin modeli analiz sonuçları

Tahmin Modeli	Korelasyon	Mutlak Hata
Generalized Linear Model	0,825	0,051
Deep Learning	0,743	0,077
Decision Tree	0,759	0,059
Random Forest	0,810	0,057
Gradient Boosted Trees*	0,827	0,057
Support Vector Machine	0,753	0,077

* En iyi çözüm

Öğrenme verileri kullanılarak çalıştırılan modellerin korelasyon ve hata değerleri ile ilgili analiz sonuçları Çizelge 3.21’de gösterilmiştir. Tahmin modellerinden en iyi çözüm olan “Gradient Boosted Trees” modeli seçilerek tahmin modeli kurulmuştur.

Önceden yüklenen ve tahmin edilmesini istediğimiz SHY firmasına ait 14 günlük veri için Şekil 3.25’de tahmin edilen değerler ile gerçekleşen değerleri karşılaştırmalı olarak gösterilmiştir.

Gradient Boosted Trees(Singapore): ... [Back to Overview](#)

Upload data to score it or upload actual outcomes for performance calculations. ⓘ

DASHBOARD MODELS PERFORMANCE DRIFTS SIMULATOR **SCORING** ALERTS INTEGRATIONS

SCORE DATA

Row No.	Tarih	prediction(E...	Ertesi Günk...	Toplam	Pozitif	Negatif	No-tr	Tanımlanma...	NPS	SGXL/SGD	USD/SGD	Date/day_...
1	Jan 22, 2020	9.112	8.850	93	42	12	6	33	30	8.780	1.349	0.150
2	Jan 23, 2020	9.123	8.820	154	44	69	9	32	-25	8.820	1.351	0.150
3	Jan 24, 2020	9.100	8.820	84	32	23	9	20	9	8.810	1.351	0.150
4	Jan 25, 2020	9.103	8.820	108	34	38	6	30	-4	8.810	1.351	0.150
5	Jan 26, 2020	9.097	8.820	67	28	18	1	20	10	8.810	1.351	0.150
6	Jan 27, 2020	9.103	8.570	104	35	32	9	28	3	8.810	1.358	0.150
7	Jan 29, 2020	9.107	8.540	105	41	29	7	28	12	8.730	1.361	0.150
8	Jan 30, 2020	9.175	8.550	160	89	21	13	37	68	8.700	1.362	0.150

Şekil 3.26: Tahmin edilen veriler

Rapid Miner ile tahmin modeli oluşturulurken, programın otomatik tahmin modelleri 100 günlük veriyi öğrenme verisi olarak kullanmaktadır. Program verileri belirli makine öğrenmesi metodlarıyla öğrenip, en düşük hata ve yüksek korelasyon olan modeli kullanıcıya sunmaktadır. SHY firmasına ait verileri Rapid Miner programının önerdiği en başarılı model olan “Gradient Boosted Trees” modeli ile 14 günlük Tweet sayıları ile ilgili veriler ve makro değişken verileri programa yüklenerek, Çizelge 3.22’de gösterildiği gibi gerçekleşen ve tahmin edilen değerleri elde edilmiştir.

Çizelge 3.22: Tahmin edilen ve gerçekleşen hisse değeri (SHY)

Tarih	Toplam	Pozitif	Negatif	Nötr	Tanımlanmayan	NPS	Dolar	ALTIN/gr	Tahmini Ertesi Günü (Ertesi Günü Hisse Değeri)	Gerçekleşen (Ertesi Günü Hisse Değeri)	
1	2020-01-22	93	42	12	6	33	30	8,7800	1,3486	9,1116	8,8500
2	2020-01-23	154	44	69	9	32	-25	8,8200	1,3514	9,1228	8,8200
3	2020-01-24	84	32	23	9	20	9	8,8100	1,3515	9,0996	8,8200
4	2020-01-25	108	34	38	6	30	-4	8,8100	1,3515	9,1030	8,8200
5	2020-01-26	67	28	18	1	20	10	8,8100	1,3515	9,0972	8,8200
6	2020-01-27	104	35	32	9	28	3	8,8100	1,3577	9,1029	8,5700
7	2020-01-29	105	41	29	7	28	12	8,7300	1,3607	9,1071	8,5400
8	2020-01-30	160	89	21	13	37	68	8,7000	1,3618	9,1752	8,5500
9	2020-01-31	142	64	32	12	34	32	8,7000	1,3651	9,1351	8,5500
10	2020-02-01	182	71	45	10	56	26	8,7000	1,3651	9,1330	8,5500
11	2020-02-02	105	45	23	4	33	22	8,7000	1,3651	9,0948	8,4900
12	2020-02-03	113	46	30	7	30	16	8,5900	1,3689	9,1093	8,5500
13	2020-02-04	96	25	26	2	43	-1	8,6800	1,3701	9,0745	8,5900
14	2020-02-05	93	31	17	11	34	14	8,6400	1,3815	9,0994	8,5800

Test edilmek üzere tahmin edilen ertesi günü hisse değeri verilerinin hata testleri yapılmıştır. Öncelikle, tahmin edilen değerler $x(t)$ ile gerçekleşen değerler $f(t)$ arasındaki sapma değeri, yani basit hata $e(t)$ aşağıdaki formülde gösterildiği gibi[63], tahmin edilen değer ile gerçekleşen değer arasındaki fark alınarak hesaplanmış ve Çizelge 3.23'te gösterilmiştir.

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

Sonrasında, tahmin edilen değerlerin hata yüzdesi $p(t)$, aşağıdaki formülde gösterildiği gibi[63] basit hatanın, tahmin değerlerine oranlanarak hesaplanmış ve Çizelge 3.23'te gösterilmiştir.

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

Son olarak, tahmin modelimizin ertesi günü hisse değeri tahmin sonuçlarının tutarlılığını ölçmede yaygın olarak kullanılan hata ölçütlerinden; Ortalama Hata

Kareleri(Mean Squared Error-MSE) ve Ortalama Mutlak Yüzde Hata(Mean Absolute Percent Error-MAPE) hesaplamaları yapılmış ve sonuçları Çizelge 3.23'te gösterilmiştir.

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

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

Aşağıdaki çizelgede THY için “Gradient Boosted Trees” modeli tarafından 18 günlük tahminlerin hata testleri MS Excel ortamında hesaplanarak gösterilmiştir. Çizelge 3.23'te, tahminin sapma (e) miktarı, yüzde hata (p), daha sonra ise mutlak yüzde hata (|p|), her biri ayrı sütunda olmak üzere hesaplanmıştır.

Çizelge 3.23: THY için tahminin hata testleri sonucu

Tarih	Tahmin Edilen Ertesi Günlük Hisse Değeri (f)	Gerçekleşen Ertesi Günlük Hisse Değeri (x)	Sapma (e)	Yüzde Hata (p)	Mutlak Yüzde Hata p
2020-01-15	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

Çizelge 3.24’te modelin açıklayıcılığını hata testi hesaplamaları yapılmıştır.

Çizelge 3.24: THY için tahmin hata sonuçları

	MSE	MAPE
Tahmin Modeli	0,175	0,026

Yukarıdaki tabloda tahmin sonuçlarının performansını ölçmede yaygın olarak kullanılan hata ölçütlerinden Ortalama Hata Kareleri (MSE) değeri 0,175 olması sifıra yakın olduğundan dolayı iyi performans gösterdiği söylenebilir. Yine tahmin modelinin tutarlılığını ölçmede kullanılan hata ölçütlerinden Ortalama Mutlak Yüzde Hata (MAPE) değeri 0,026 olarak hesaplanmıştır. Hesaplanan bu hata değeri tahmin modelimizin %2,6 oranında hata yaptığını dolayısıyla tahmin tutarlılığının yüksek olduğu göstermektedir.

Ayrıca SHY firması için “Gradient Boosted Trees” modeli tarafından 14 günlük tahminlerin hata testleri MS Excel ortamında hesaplanarak yapılmış olup Çizelge 3.25 ve 3.26’da gösterilmiştir.

Çizelge 3.25: SHY için tahminin hata testleri sonucu

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

Çizelge 3.26: SHY için tahmin hata sonuçları

	MSE	MAPE
Tahmin Modeli	0,233	0,051

Yukarıdaki tabloda tahmin sonuçlarının performansını ölçmede yaygın olarak kullanılan hata ölçütlerinden Ortalama Hata Kareleri (MSE) değeri 0,233 olması sifira yakın olduğundan dolayı iyi performans gösterdiği söylenebilir. Yine tahmin modelinin tutarlılığını ölçmede kullanılan hata ölçütlerinden Ortalama Mutlak Yüzde Hata (MAPE) değeri 0,051 olarak hesaplanmıştır. Hesaplanan bu hata değeri tahmin modelimizin %5,1 oranında hata yaptığını dolayısıyla tahmin tutarlılığının yüksek olduğu göstermektedir.





4. BULGULAR VE TARTIŞMA

SPSS programı kullanılarak elde edilen Korelasyon analizi sonuçlarına göre, 5 havayolu firması için toplam 40 adet korelasyon analiz sonucu incelenmiştir. Sonuç çizelgesi aşağıdaki gibidir;

Çizelge 4.1: 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	Anlamli ilişki yoktur.	-0,185
2	-0,284	0,546	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.	-0,339
3	-0,303	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.
4	Anlamli ilişki yoktur.	0,600	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.	-0,216
5	-0,280	0,334	-0,181	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.
6	-0,296	0,545	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.	-0,245
7	-0,296	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.
8	Anlamli ilişki yoktur.	0,611	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.	Anlamli ilişki yoktur.

5 havayolu firması için korelasyon analizi sonuçları ve uyguladığımız hipotezlerin sonuçları Çizelge 4.1’de gösterilmiştir. Hipotez sonuçlarına baktığımızda, QHY firmasına ait hipotezlerin anlamsız olduğu sonucuna varılmıştır. Sırasıyla kurulan hipotezleri incelersek;

1 nolu hipotez Toplam Tweet Sayısı ile Aynı Günkü Hisse Değeri arasındaki ilişkiyi göstermektedir. THY, SHY, LHY ve FHY firmalarında anlamlı sonuç elde edilmiştir. THY, LHY, FHY firmaları için negatif yönlü düşük düzeyde korelasyon, SHY firması için pozitif yönlü orta düzeyde korelasyon olduğu sonucu elde edilmiştir.

2 nolu hipotez Pozitif Tweet Sayısı ile Aynı Günü Hisse Değeri arasındaki ilişkiyi göstermektedir. THY için negatif yönlü düşük düzeyde korelasyon, FHY için negatif yönlü orta düzeyde korelasyon SHY firması için pozitif yönlü orta düzeyde korelasyon olduğu sonucu elde edilmiştir.

3 nolu hipotez Negatif Tweet Sayısı ile Aynı Günü Hisse Değeri arasındaki ilişkiyi göstermektedir. Sadece THY firması için negatif yönlü orta düzeyde korelasyon olduğu sonucu elde edilmiştir.

4 nolu hipotez Pozitif-Negatif farkı Tweet Sayısı(NPS) ile Aynı Günü Hisse Değeri arasındaki ilişkiyi göstermektedir. SHY firması için pozitif yönlü orta düzeyde, FHY için negatif yönlü düşük düzeyde korelasyon olduğu sonucu elde edilmiştir.

5 nolu hipotez Toplam Tweet Sayısı ile Ertesi Günü Hisse Değeri arasındaki ilişkiyi göstermektedir. THY ve LHY firmaları için negatif yönlü düşük düzeyde, SHY için pozitif yönlü orta düzeyde korelasyon olduğu sonucu elde edilmiştir.

6 nolu hipotez Pozitif Tweet Sayısı ile Ertesi Günü Hisse Değeri arasındaki ilişkiyi göstermektedir. THY ve FHY firmaları için negatif yönlü düşük düzeyde, SHY firması için pozitif yönlü orta düzeyde korelasyon olduğu sonucu elde edilmiştir.

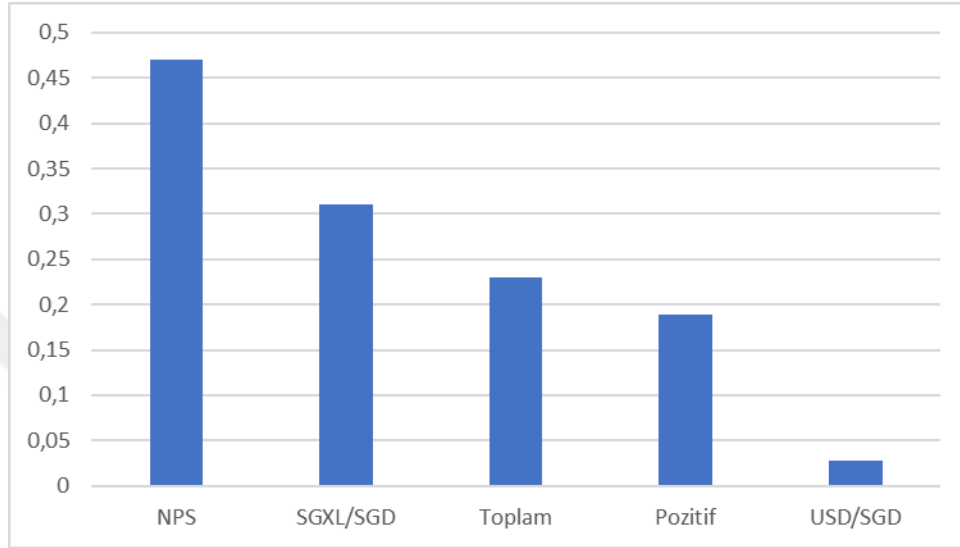
7 nolu hipotez Negatif Tweet Sayısı ile Ertesi Günü Hisse Değeri arasındaki ilişkiyi göstermektedir. Sadece THY firması için negatif yönlü düşük düzeyde korelasyon olduğu sonucu elde edilmiştir.

8 nolu hipotez Pozitif-Negatif farkı Tweet Sayısı(NPS) ile Ertesi Günü Hisse Değeri arasındaki ilişkiyi göstermektedir. Sadece SHY firması için pozitif yönlü orta düzeyde korelasyon olduğu sonucu elde edilmiştir.

SPSS programı ile yapılan analizler sonucunda, 40 adet korelasyon analizin 22 tanesinde anlamlı ilişki tespit edilememiştir. Yarısından fazla anlamsız korelasyon olmasının nedeni günlere ve haftalara göre tweet sayılarında yüksek farklar olmasından ve analiz edilen tweetlerin Meaning Cloud eklentisi ile duygu analizi yapılacak dil seçeneklerinin kısıtlı olmasından dolayı kaynaklanmaktadır. Bu sonuçları genel olarak yorumlarsak, THY, LHY ve FHY firmalarıyla ilgili Twitter platformundaki anılma sayısı azaldığında, bu firmaların hisse değerleri artmaktadır.

SHY firması için ise Twitter platformundaki anılma sayıları artıyorsa, hisse değerleri de artmaktadır.

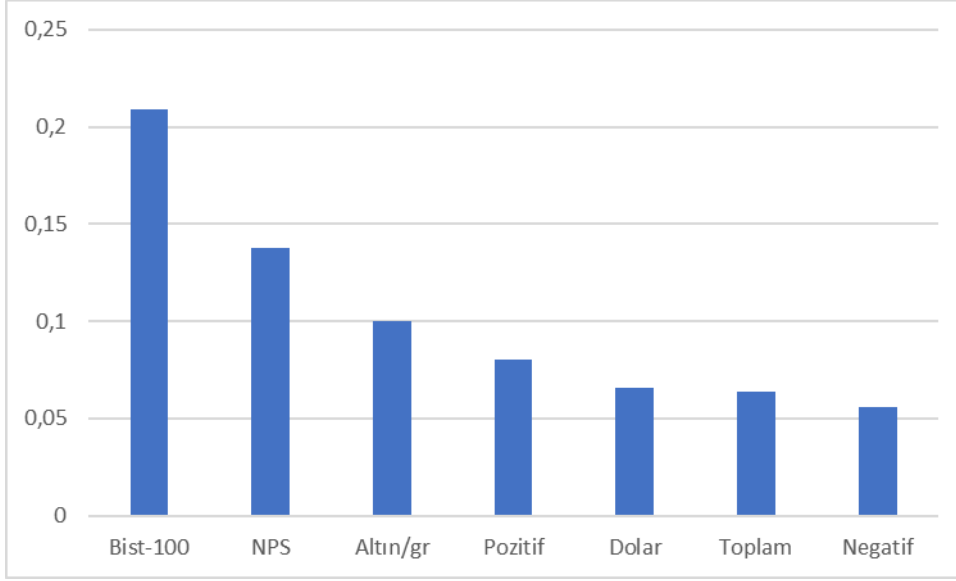
SHY için kurulan “Gradient Boosted Trees” modelinde öğrenme verileri için korelasyon %82,7 ve mutlak hata değeri ise %5,7 olduğu tespit edilmiştir. Elde edilen sonuç analiz sonuçlarına göre tahmin modeli olarak kullanılmıştır.



Şekil 4.27: SHY ağırlık oranları

SHY için kurulan tahmin modelindeki değişkenlerin tahmin sonucunu etkilediği ağırlık düzeyleri Şekil 4.1’de gösterildiği gibidir. Tahmin modelinde NPS, Toplam, Pozitif Tweet sayıları ile SGXL/SGD değişkenleri ağırlıklı olarak tahmin değerini etkilediği sonucu elde edilmiştir. Sosyal medyadan elde edilen veriler hisse değeri tahmininde önemli bir faktör olduğu tespit edilmiştir. Bu firma için 100 günlük öğrenme verisi kullanılmış olup, 14 günlük test verisi ile tahmin değeri verileri test edilmiştir. Tahmin edilen değerler, test edilen gerçek değerler ile karşılaştırıldığında %5,1 hata ile tahmin edilmiştir. Hisse değerinin ertesi gün için artış veya azalış olma durumu ise 14 günlük test için 8 gününde artış veya azalış olacağını başarılı şekilde tahmin ettiği gözlemlenmiştir.

THY için kurulan “Gradient Boosted Trees” modelinde öğrenme verileri için korelasyon %98,5 ve mutlak hata değerinin %1,2 olduğu tespit edilmiştir. Elde edilen sonuç analiz sonuçlarına göre tahmin modeli olarak kullanılmıştır.



Şekil 4.28: THY ağırlık oranları

THY firması için kurulan tahmin modelindeki değişkenlerin tahmin sonucunu etkilediği ağırlık düzeyleri Şekil 4.2’de gösterildiği gibidir. Tahmin modelinde NPS, Toplam, Pozitif, Negatif Tweet sayıları ile Bist-100, Altın, Dolar değişkenleri ağırlıklı olarak tahmin değerini etkilediği sonucu elde edilmiştir. Sosyal medyadan elde edilen veriler hisse değeri tahmininde önemli bir faktör olduğu tespit edilmiştir. Bu firma için 100 günlük öğrenme verisi kullanılmış olup, 18 günlük test verisi ile tahmin değeri verileri test edilmiştir. Tahmin edilen değerler, test edilen gerçek değerler ile karşılaştırıldığında %2,6 hata ile tahminler başarılı bir şekilde elde edilmiştir. Hisse değerinin ertesi gün için artış veya azalış olma durumu ise 18 günlük test için 9 gününde artış veya azalış olacağını başarılı şekilde tahmin ettiği gözlemlenmiştir.

5. SONUÇ

Elde edilen SPSS analizi sonuçlarına göre, THY, LHY ve FHY firmaları için negatif, SHY firması için ise pozitif yönlü orta düzeyde korelasyon tespit edilmiştir. Bu sonucun yanı sıra sosyal medya verileri ile birlikte makro ekonomik faktörlerde (Borsa, Döviz, Altın) hisse değeri tahmininde kullanılmıştır.

Tahmin modeli oluşturmak için THY firmasına ait Twitter verileri ile birlikte BİST-100, Dolar, Altın faktörleri de dahil edilerek Rapid Miner programında Auto Model kullanılmıştır. Auto Model ile kurulan “Gradient Boosted Trees” tahmin modeli ile 18 günlük veri ile tahmin denemeleri yapılmıştır. Ayrıca NPS değerinin BİST-100’den sonra ikinci en yüksek etkiye sahip olduğu tespit edilmiştir. Ertesi günkü hisse değerinin artış veya azalışları %55,6 oranda doğru tahmin edilmiştir. Ayrıca tahmin edilen değerler, test edilmek üzere gerçek değerler ile karşılaştırıldığında %2,6 hata oranı ile tahmin edilmiş ve hisselerle yatırım yapılırken önemli bilgi kaynağı olacağı sonucuna varılmıştır.

Bir diğer tahmin modeli SHY için Twitter verileri ile birlikte SGXL/SGD, USD/SGD faktörleri de dahil edilerek Rapid Miner programında Auto Model kullanılmıştır. Auto Model ile kurulan “Gradient Boosted Trees” tahmin modeli ile 14 günlük veri ile tahmin denemeleri yapılmıştır. Ayrıca, NPS değerinin en yüksek etkiye sahip olduğu tespit edilmiştir. Ertesi günkü hisse senedi değerinin artış veya azalışları %57,1 doğruluk oranı ile tahmin edilmiştir. Ayrıca tahmin edilen değerler, test edilmek üzere gerçek değerler ile karşılaştırıldığında %5,1 hata oranı ile tahmin edilmiş ve hisselerle yatırım yapılırken önemli bilgi kaynağı olacağı sonucuna varılmıştır.

Sosyal medyadan elde ettiğimiz tweet’lerin FHY için Fransızca ve İngilizce dilinde, diğer firmalar için ise sadece İngilizce dilinde toplanması çalışmamızı kısıtlamıştır. Firmalar için birden farklı dilde tweetler toplanıp analiz edilmesi sonucumuzu olumlu veya olumsuz etkileyeceği aşikardır. Elimizdeki imkanlar dahilinde ulusal ortak dil olarak İngilizce tweetler üzerinde çalışmamız tamamlanmıştır.

Sonuç olarak bu çalışma literatüre sosyal medya verileri ile hisse değeri arasındaki ilişkiyi tahmin modeli kurarak incelediğimiz için önemli bir katkı sunacaktır. Ayrıca sosyal medya verileri ve makro ekonomik faktörler ile yapılan tahminlerde, sosyal

medyadan elde edilen verilerin önemli bir faktör olduğu sonucuna varılmıştır. Yatırımcılar bu model üzerinden çalışmalarına ve yatırımlarına devam edebileceklerdir. Aynı zamanda modelimiz bu alandaki çalışmalara öncülük edecektir.

Yapılan çalışmanın yatırımcılara karar vermelerini etkileyen değişkenlerin analiz edilmesiyle yaptığı katkıları olmasına rağmen bazı sınırlamaları vardır. İlk olarak, veri analizi için maliyet ve zaman kısıtlamalarından dolayı 4 aylık süre ile toplanan Twitter verisi ile toplamda 236 764 adet tweet ile çalışma tamamlanmıştır. Veriler sosyal medya platformu Twitter ile sınırlandırılmıştır. Çalışmamızda sadece İngilizce dilinde tweetler ile çalışılmıştır ve FHY firması için ekstra Fransızca dilindeki tweetler de çalışmaya dahil edilmiştir. Bu sınırlamalar ortadan kaldırılarak Twitter dışındaki sosyal medya platformları ve mikrobloglardaki veriler ile çalışmanın kapsamı genişletilebilir. Ayrıca Türkçe, İngilizce, Almanca, Fransızca vb. diller ile çalışma için gerekli veri sayısı artırılabilir ve daha uzun süreli veri toplanması ile çalışma geliştirilebilir. Tahmin modeli kurulurken farklı makro ekonomik faktörler eklenebilir ve değişkenler geliştirilebilir.

Destekleyen Kiři ve Kurum Bilgisi

Bu tez alıřması, Yalova niversitesi Bilimsel Arařtırma Projeleri Koordinasyon Birimi tarafından, 2019/YL/0001 numaralı proje ile desteklenmiřtir.





KAYNAKLAR

- [1] J. R. Evans, *Business Analytics*, 2nd ed. Boston: Pearson Education, 2016.
- [2] R. J. Vaughan, "Examining the Data Analytics Skill Gap in Mid-Level Marketing Professionals , Driven by the Continuing Exponential Growth of Big Data," vol. 5, no. 3, 2017.
- [3] A. Charan, *Marketing analytics: A Practitioner's Guide to Marketing Analytics and Research Methods*. World Scientific Publishing Company, 2015.
- [4] R. Venkatesan, P. Farris, and R. T. Wilcox, *Cutting Edge Marketing Analytics: Real World Cases and Data Sets for Hands on Learning*. Pearson Education, 2014.
- [5] C. W. Holsapple, S.-H. Hsiao, and R. Pakath, "Business social media analytics: Characterization and conceptual framework," *Decis. Support Syst.*, 2018.
- [6] N. Ulusoy, "Sözlüklerdeki Sinema Sevgisi: New York'ta Beş Minare ve Çoğunluğun İnternet Sözlüklerine Yansıması," *Beta Yayıncılık*, İstanbul, pp. 195–211, 2012.
- [7] A. J. Kim and E. Ko, "Impacts of luxury fashion brand's social media marketing on customer relationship and purchase intention," *J. Glob. Fashion Mark.*, vol. 1, no. 3, pp. 164–171, 2010.
- [8] M. Dewing, "Social media: An Introduction," *Parliam. Inf. Res. Serv. Soc. Aff. Div. Libr. Parliam. Canada*, 2012.
- [9] S. Murugesan, "Understanding Web 2.0," *IT Prof.*, vol. 9, no. 4, pp. 34–41, 2007.
- [10] J. Salo, "Social media research in the industrial marketing field: Review of literature and future research directions," *Ind. Mark. Manag.*, vol. 66, pp. 115–129, 2017.
- [11] I. Lee, "Social media analytics for enterprises: Typology, methods, and processes," *Bus. Horiz.*, vol. 61, no. 2, pp. 199–210, 2018.
- [12] N. Syam and A. Sharma, "Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice," *Ind. Mark. Manag.*, vol. 69, pp. 135–146, 2018.
- [13] F. Ö. Güzel, "Marka İtibarını Korumada Şikâyet Takibi: Çevrimiçi Seyahat 2.0 Bilgi Kanallarında Bir Uygulama," vol. 5, no. 1, pp. 5–19, 2014.

- [14] G. V Attigeri, M. P. M. M, R. M. Pai, and A. Nayak, "Stock market prediction: A big data approach," *TENCON 2015 - 2015 IEEE Reg. 10 Conf.*, pp. 1–5, 2015.
- [15] S. Asur and B. A. Huberman, "Predicting the future with social media," *Proc. - 2010 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2010*, vol. 1, no. March 2010, pp. 492–499, 2010.
- [16] R. Chen and M. Lazer, "Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement," pp. 1–5, 2011.
- [17] X. Zhang, H. Fuehres, and P. A. Gloor, "Predicting Stock Market Indicators Through Twitter ' I hope it is not as bad as I fear ,'" *Procedia - Soc. Behav. Sci.*, vol. 26, no. 2007, pp. 55–62, 2011.
- [18] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011.
- [19] A. Mittal and A. Goel, "Stock Prediction Using Twitter Sentiment Analysis," <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>, no. December, 2012.
- [20] T. T. V, S. Chang, T. H. Quang, and N. Collier, "An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter," vol. 3, no. December 2012, pp. 23–38.
- [21] A. Porshnev, I. Redkin, and A. Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis .," *2013 IEEE 13th Int. Conf. Data Min. Work.*, pp. 440–444, 2013.
- [22] J. Si, A. Mukherjee, B. Liu, and Q. Li, "Exploiting Topic based Twitter Sentiment for Stock Prediction," no. 2011, pp. 24–29, 2013.
- [23] A. Khadjeh, S. Aghabozorgi, T. Ying, D. Chek, and L. Ngo, "Expert Systems with Applications Text mining for market prediction : A systematic review," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7653–7670, 2014.
- [24] A. Bhardwaj, Y. Narayan, and M. Dutta, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty," vol. 70, pp. 85–91, 2015.
- [25] B. Dickinson and W. Hu, "Sentiment Analysis of Investor Opinions on Twitter," *Soc. Netw.*, vol. 04, no. 03, pp. 62–71, 2015.
- [26] N. Eliaçık, A.B., & Erdogan, "Mikro Bloglardaki Finans Toplulukları için

- Kullanıcı Ağırlıklandırılmış Duygu Analizi Yöntemi,” *Ulus. Yazılım Mühendisliği Sempozyumu*, pp. 782–793, 2015.
- [27] B. Li, K. C. C. Chan, C. Ou, and S. Ruifeng, “Discovering public sentiment in social media for predicting stock movement of publicly listed companies,” pp. 81–92, 2017.
- [28] N. Oliveira, P. Cortez, and N. Areal, “The impact of microblogging data for stock market prediction : Using Twitter to predict returns , volatility , trading volume and survey sentiment indices,” *Expert Syst. Appl.*, vol. 73, pp. 125–144, 2017.
- [29] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, “Sentiment analysis of Twitter data for predicting stock market movements,” *Int. Conf. Signal Process. Commun. Power Embed. Syst. SCOPES 2016 - Proc.*, pp. 1345–1350, 2017.
- [30] M. YILDIRIM and C. A. YÜKSEL, “Sosyal Medya İle Hisse Senedi Fiyatının Günlük Hareket Yönü Arasında İlişkini İncelenmesi: Duygu Analizi Uygulaması,” *Uluslararası İktisadi ve İdari İncelemeler Derg.*, 2017.
- [31] S. Das and R. K. Behera, “Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction,” *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 956–964, 2018.
- [32] R. Yadav, A. V. Kumar, and A. Kumar, “News-based supervised sentiment analysis for prediction of futures buying behaviour,” *IIMB Manag. Rev.*, vol. 31, no. 2, pp. 157–166, 2019.
- [33] Kamu Aydınlatma Platformu, “BİST-100 Endeksinde İşlem Gören Şirketler.” [Online]. Available: <https://www.kap.org.tr/tr/Endeksler>. [Accessed: 06-Feb-2020].
- [34] investing.com, “SIAL.” [Online]. Available: <https://tr.investing.com/equities/singapore-airlines>. [Accessed: 06-Feb-2020].
- [35] investing.com, “Singapore Exchange(SGXL/SGD).” [Online]. Available: <https://tr.investing.com/equities/singapore-exchange>. [Accessed: 06-Feb-2020].
- [36] investing.com, “Investing/BİST-100.” [Online]. Available: <https://tr.investing.com/indices/ise-100>. [Accessed: 01-Mar-2020].
- [37] investing.com, “Investing/Usd-Try.” [Online]. Available: <https://tr.investing.com/currencies/usd-try-chart>. [Accessed: 01-Mar-2020].

- [38] investing.com, “Investing/Altın-Try.” [Online]. Available: <https://tr.investing.com/currencies/gau-try>. [Accessed: 03-Jan-2020].
- [39] investing.com, “Investing.com.” [Online]. Available: <https://tr.investing.com/>. [Accessed: 06-Feb-2020].
- [40] Skytrax, “2019 World Airline Ranking List.” [Online]. Available: <https://www.worldairlineawards.com/worlds-top-100-airlines-2019/>. [Accessed: 10-Feb-2020].
- [41] investing.com, “THYAO.” [Online]. Available: <https://tr.investing.com/equities/turk-hava-yollari>. [Accessed: 06-Feb-2020].
- [42] investing.com, “LHAG.” [Online]. Available: <https://tr.investing.com/equities/dt-lufthansa>. [Accessed: 06-Feb-2020].
- [43] investing.com, “QAN.” [Online]. Available: <https://tr.investing.com/equities/qantas-airways-limited>. [Accessed: 06-Feb-2020].
- [44] investing.com, “AIRF.” [Online]. Available: <https://tr.investing.com/equities/air-france---klm>. [Accessed: 06-Feb-2020].
- [45] Developer.twitter.com, “Twitter Premium Operators.” [Online]. Available: <https://developer.twitter.com/en/docs/tweets/rules-and-filtering/overview/premium-operators>. [Accessed: 14-Nov-2019].
- [46] Twitter, “Premium Search Tweets.” [Online]. Available: <https://developer.twitter.com/en/docs/tweets/search/overview/premium>. [Accessed: 14-Nov-2019].
- [47] Twitter, “Developer Twitter Apps.” [Online]. Available: <https://developer.twitter.com/en/apps>. [Accessed: 26-Nov-2019].
- [48] Twitter, “twitter.com.” [Online]. Available: <https://www.twitter.com/>. [Accessed: 06-Feb-2020].
- [49] Twitter, “Developer Twitter.” [Online]. Available: <https://developer.twitter.com/en/apps>. [Accessed: 26-Nov-2019].
- [50] A. A. Beaujean, “Factor Analysis Using R - Practical Assessment, Research & Evaluation,” *Pract. Assessment, Res. Eval.*, vol. 18, no. 4, 2011.
- [51] C. D. Doğan and M. Uluman, “İstatistiksel Veri Analizinde R Yazılımı ve Kullanımı,” *Elem. Educ. Online*, vol. 15, no. 2, pp. 615–634, 2016.
- [52] E. Çürük, “Sosyal Ağlardaki Siber Zorbalığın Yapay Zeka Algoritmaları İle Tespiti Ve Sınıflandırılması,” Mersin Üniversitesi, 2018.

- [53] Twitter, “Search Tweets Operators - Premium search APIs.” [Online]. Available:
<https://developer.twitter.com/en/docs/tweets/search/overview/premium>.
[Accessed: 02-Dec-2019].
- [54] K. Gazioğlu and Ş. E. Şeker, “Veri Madenciliği Yöntemleri ile Twitter Üzerinden Girişimcilik Analizi,” pp. 3–10, 2017.
- [55] P. Tripathi, S. K. Vishwakarma, and A. Lala, “Sentiment Analysis of English Tweets Using Rapid Miner,” *2015 Int. Conf. Comput. Intell. Commun. Networks*, 2015.
- [56] Rapid Miner, “RapidMiner Studio 9.4.” [Online]. Available:
<https://my.rapidminer.com/nexus/account/index.html#downloads>. [Accessed: 26-Aug-2019].
- [57] Rapid Miner, “Generalized Linear Model.” [Online]. Available:
https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/functions/generalized_linear_model.html. [Accessed: 06-Feb-2020].
- [58] Rapid Miner, “Deep Learning Model.” [Online]. Available:
https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/neural_nets/deep_learning.html. [Accessed: 06-Feb-2020].
- [59] Rapid Miner, “Decision Tree Model.” [Online]. Available:
https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/parallel_decision_tree.html. [Accessed: 06-Feb-2020].
- [60] Rapid Miner, “Random Forest Model.” [Online]. Available:
https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/parallel_random_forest.html. [Accessed: 06-Feb-2020].
- [61] Rapid Miner, “Gradient Boosted Trees Model.” [Online]. Available:
https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/gradient_boosted_trees.html. [Accessed: 06-Feb-2020].
- [62] Rapid Miner, “Support Vector Machine Model.” [Online]. Available:
https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/support_vector_machines/support_vector_machine.html. [Accessed: 06-Feb-2020].
- [63] M. Karahan, “İstatistiksel Tahmin Yöntemleri: Yapay Sinir Ağları Metodu ile Ürün Talep Tahmini Uygulaması,” *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı*, pp. 112–113, 2011.



ÖZGEÇMİŞ



Ad Soyad : Ömer Faruk UYRUN
Doğum Yeri ve Tarihi : SİİRT / 21.02.1992
E-Posta : omerfarukuyrun@gmail.com
Lisans : Yalova Üniversitesi
Mühendislik Fakültesi – Endüstri Mühendisliği

TEZDEN TÜRETİLEN YAYINLAR/SUNUMLAR

Ö. F. UYRUN and İ. SABUNCU, “Forecasting of Stock Market Values of Airline Companies by Using Social Media Data,” 2nd Int. Conf. Data Sci. Appl., vol. 1, pp. 169–170, 2019, Balıkesir, TURKEY.