

---

## SOSYAL MEDYA İLE HİSSE SENEDİ FİYATININ GÜNLÜK HAREKET YÖNÜ ARASINDAKİ İLİŞKİNİN İNCELENMESİ: DUYGU ANALİZİ UYGULAMASI

---

Metin YILDIRIM<sup>1</sup>

Cenk Arsun YÜKSEL<sup>2</sup>

### Öz

Elektronik ağızdan ağıza iletişim (eWOM), bireylerin çevrimiçi ortamda işletmeler, işletmelerin sunduğu ürün ve hizmetler hakkındaki duyguların ve düşüncelerin paylaşıldığı bir yapı olması nedeniyle önemli bir araştırma başlığı konumunu almıştır. Elektronik ağızdan ağıza iletişimin gerçekleştiği belli başlı platformlar arasında Twitter yer almaktadır. Ülkemizde ve dünyada yaygın olarak kullanılan mikro blog sistemleri arasında yer alan Twitter ile sermaye piyasalarının etkileşimi giderek artmaktadır. Pazarlama ve finansman konusunda yapılan araştırmaların arakesitinde yer alan başlık, araştırma konusu olarak seçilmiştir. Telekomünikasyon sektöründe yer alan, Borsa İstanbul'da hisseleri işlem gören firma üzerinden araştırma yürütülmüştür. Araştırma dönemi olarak 15.03.2017 ile 15.05.2017 tarihleri arasında Borsa İstanbul'da hisselerin işlem gördüğü günler seçilmiştir. Günlük olarak ilgili hissenin açılış ve kapanış değerleri elde edilmiş, açılış değeri kapanış değerlerinin farkı alınmış ve hisse hareket yönü o gün için negatif ya da pozitif olarak sınıflandırılmıştır. İlgili günler için twitter sayfasından bireylerin gönderdiği mesajlar üzerinden Duygu Analizi gerçekleştirilmiş, pozitif ve negatif olarak sınıflandırılmış, günlük duygu polaritesi değeri elde edilmiştir. Bu çalışmada duygu polarite değerleri ile hisse fiyatı hareketinin yönü arasında günlük bazda korelasyon analizi yapılmıştır. Elde edilen Spearman'ın sıra ilişkisi katsayısı değeri çerçevesinde iki veri seti arasında negatif ve orta düzeyde bir bağlantıya rastlanmıştır.

**Anahtar Kelimeler:** Duygu Analizi, Mikro Blog, Sermaye Piyasası, Elektronik Ağızdan Ağıza İletişim, eWOM  
**JEL Sınıflandırması:** M31,M39

---

## INVESTIGATION OF RELATIONSHIP BETWEEN SOCIAL MEDIA AND STOCK PRICE DAILY MOVEMENT DIRECTION: SENTIMENT ANALYSIS IMPLEMENTATION

---

### Abstract

People share their opinions and sentiments about companies, services and products on Internet. This makes Electronic Word of Mouth to be an important research topic in recent years. Twitter is one of the major platform for Electronic Word of Mouth. Interaction of capital markets with Twitter has been highly increasing, it has been selected as our research topic. Research has been carried out on the company, which has been operating in telecommunications industry and whose shares are traded in Borsa Istanbul. The research period is the dates between 15.03.2017 and 15.05.2017. Daily opening and closing prices of the selected stock have been collected, daily raise and fall of the stock price has been classified accordingly. Sentiment analysis has been conducted for the same period and daily polarity values have been obtained. In this study, correlation analysis between public sentiment analysis and stock price movement direction is conducted. Spearman's rank correlation test was used. The obtained Spearman's rank correlation coefficient result showed, a negative and moderate correlation exists between daily changes in the company's stock prices with the public sentiments in tweets.

**Keywords:** Sentiment Analysis, Micro Blog, Capital Market, eWOM (Electronic Word Of Mouth)  
**JEL Classification:** M31,M39

---

<sup>1</sup> İşletme Doktora Programı Öğrencisi, İKÜ Sosyal Bilimler Enstitüsü, metinyildirim@hotmail.com

<sup>2</sup> Doç.Dr.,T:C İstanbul Üniversitesi, İşletme Fakültesi Pazarlama Anabilim Dalı, cenka@istanbul.edu.tr

## 1. Giriş

Pazarlama alanında çalışan araştırmacıların ve bu alanda çalışan profesyonellerin elektronik ağızdan ağıza iletişim'in önemi, gücü ve diğer başlıklar üzerindeki etkisi hakkındaki ilgileri her geçen gün artmakta olup, bu alandaki araştırma ve çalışmalarını yoğunlaştırdıkları görülmektedir. Bu konudaki ilgi artışının en önemli nedeni olarak, elektronik ağızdan ağıza iletişim'in yoğun olarak metin tabanlı içerikten oluşması ve önemli nitelikte bilgiye sahip olması söylenebilir. En az bu konu kadar önemli bir başka etmen ise de, elektronik ağızdan ağıza iletişim'in diğer kişilerin firma, ürün ve hizmetler konusundaki düşüncelerini etkilemedeki gücün fark edilmiş olmasıdır (Moe ve Trusov, 2011:444-446). Elektronik ağızdan ağıza iletişim'in ürün satışları üzerindeki etkisi, diğer bireylerin algı ve niyet değişimindeki etkileri başta olmak üzere birçok konuda araştırma yapılmıştır. Araştırmaların ağırlıklı olarak tüketici davranışları araştırması üzerinde yoğunlaştığı görülmektedir. Pazarlama ve finansın ortak ilgi alanına giren Hisse Senedi Fiyatının Günlük Hareket Yönü ile elektronik ağızdan ağıza iletişimin etkileşimi yoğun çalışılmış bir başlık olarak yer almamaktadır.

Sermaye piyasasında işlem gören araçlar arasında getirisi ve riski en yüksek aracın hisse senetleri olduğu kabul görmektedir. Siyasi, ekonomik ve sosyal faktörlerin hızlı tepki verdiği ve etkilendiği kabul edilmektedir (Karaatlı vd., 2005:38). Hisse senetlerinin fiyatlarındaki değişim üzerinde yapılan ilk çalışmalar, geçmiş değerleri temel alan çalışmalar olmuştur. Takip eden çalışmalar hisse senedi fiyatlarındaki değişimin sadece geçmiş fiyatlar üzerine olduğu yaklaşımını çürütecek yönde yer almıştır (Pagolu vd, 2016:1). Etkin Piyasa Kuramı çerçevesinde incelendiğinde, var olan fiyatlanmanın gelecekteki beklentiler üzerinden oluştuğunu belirtmişlerdir. Bu noktadan hareketle, geçmişe ait verilerin gelecekteki fiyatlamaya etkisinin mümkün olmaması gerektiği düşüncesi ortaya atılmıştır (Albeni ve Demir, 2005:2). Etkin piyasa kuramı kısaca, hisse senedi fiyatlamasının geçmiş yada güncel fiyatlamadan daha ziyade, alınan yeni bilgiler çerçevesinde şekillendiğini kabul etmektedir.

Sosyal medya ve mikro blog siteleri çok geniş kapsama sahip olan bilgi kaynağı konumuna gelmiştir. Özellikle mikro bloglar kurguları gereği, kişilerin günlük yaşamda kullandıkları hizmet ve ürünler hakkında duygularını, şikayetlerini, düşüncelerini ve oluşan izlenimlerini gerçek zamanlı ve yoğun olarak aktardığı önemli bir mecra olarak yer almış olup, önemini ve etkinliğini her geçen gün artırmaktadır (Agarwal vd., 2011:30). Twitter üzerinden elde edilen bulgular başta sermaye piyasaları olmak üzere film hasılatları, seçim sonuçları, halk sağlığı ile ilgili araştırmalarda yaygın olarak kullanılmış olup, kullanılmaya devam etmektedir (Nguyen, vd., 2013:1). Sosyal medya ve internetin gelişimi ile toplumun ve tüketicilerin, pazarlama kampanyaları ve ürün tercihleri başta olmak üzere bir çok konu başlığında düşünce ve duygularını anlama imkanı araştırmacılara sunmaktadır. Sosyal medya ve sosyal network sitelerinde tüketicilerin ilettiği mesajların analizi ile tüketicinin belirli bir konudaki duygu ve düşüncesi konusunda detaylı bilgi edinebilme, elde olan önemli sayıda soruya cevap bulabilme imkanı yer almaktadır (Miranda ve Sassi, 2014:19-21). Sosyal medya, sosyal network ve mikro blog sitelerinde yer alan mesajların analizi gerçekleştiren çalışmalarda, kullanım alanlarındaki genişlemenin yanı sıra araştırmalardaki sayısal artışta yükselen bir ivme ile devam etmektedir.

Duygu Analizi ya da Fikir Madenciliği sayısal bir çalışma alanı olarak yer almaktadır. Çalışma alanı olarak işletmeler, diğer bireyler, sorunlar, olaylar, güncel başlıklar ve bunların özellikleri hakkında bireylerin fikirleri, değerlendirmeleri, tutumları, duyguları incelenmesi olarak yer almaktadır (Liu ve Zhang, 2012:415). Var olan verinin sayısal büyüklüğü, oluşturulma hızı göz önüne alındığında, ilgili verinin işlenmesinde, sınıflandırılması süreci hızlandıracak belirli yöntemlere ihtiyaç duyulmuştur. Sosyal medya izleme yöntemleri arasında en yaygın kullanılan yöntemlerden bir tanesi de Duygu Analizi olarak yer almaktadır. Yapılan çalışmanın farklı kullanım alanları yer almakta olup, bunlardan bir tanesi de duygu analiz sonuçları ile müşteri memnuniyetinin ölçülmesinde kullanılabilecek araçlardan bir tanesi de olabilmektedir (Miranda ve Sassi, 2014:18).

Bu araştırma, Twitter'dan elde edilen tweetlerin pozitif ve negatif olarak sınıflandırılması Duygu Analizi yöntemi ile gerçekleştirilmiştir. Telekomünikasyon sektöründe yer alan bir firmanın hisse

senedinin açılış ve kapanış değeri arasında fark alınıp, o gün için hisse senedi yükseliş yada düşüşte olduğu veri elde edilmiştir. Duygu analizi sonucunda elde edilen günlük polarite değeri ile hisse senedi günlük değer farkın kazanç yada kayıpta oluşu verisi arasında korelasyon analizi gerçekleştirilmiştir. Bu analiz sonucunda iki veri seti arasında orta düzeyde, negatif ve anlamlı bir ilişki olduğu ortaya çıkmıştır.

## **2.Literatür Özeti**

Mikro blog verileri ile sermaye piyasaları arasında ilişkiyi araştıran bir dizi çalışma gerçekleştirilmiştir. Bu konuda en fazla bilinenler arasında Bollen vd.,'nin 2011 ayında yaptığı çalışma gelmektedir. Çalışmada twitter üzerinde elde edilen tweetlerin duygu analizi ile pozitif ve negatif sınıflandırılmasının yanı sıra, toplumun genel ruh halini belirlemek amacıyla sakin, emin, kibar, mutlu, uyanık ve hayati olarak sınıflandırmıştır. İlk etapta, bu değerler ile Dow Jones Borsası Endüstri Endeksi arasındaki ilişki araştırılmıştır. Ek olarak, toplumun ruh hali verisi kullanılarak endeksin hareketinin yönünün öngörülüp öngörülmeceği konusunda çalışma gerçekleştirilmiştir. Öngörü aşamasında bulanık sinir ağları yöntemi kullanılmıştır. Toplumun genel ruh hali ile Dow Jones Borsası Endüstri Endeksi arasında yüksek düzeyde ilişki görülmüştür(Bollen vd., 2011:1).

Twitter üzerinde toplanan veriler üzerinde yapılan ve toplumun genel ruh halinin duygu analizi yöntemleri ile sınıflandırılmasını içeren çalışmada, toplumun genel ruh hali ile belli başlı finansal pazar göstergeler arasındaki ilişki araştırılmıştır. Bu veri setleri arasında Dow Jones, S&P 500, NASDAQ ve Chicago Board Options Exchange Volatility Index VIX yer almaktadır. Belirtilmesi gereken önemli bir nokta ise, Chicago Board Options Exchange Volatility Index VIX ile diğer indeksler arasında yüksek seviyede negatif yönde bir korelasyon yapısının var olmasıdır. Analiz kapsamında toplumun ruh hali kötümser ve iyimser olarak sınıflandırılmıştır. Çalışma kapsamında 30 Mart 2009- 7 Eylül 2009 tarihleri arasındaki altı aylık süreçte günlük 8.100 ile 43.040 arasında değişen tweetler üzerinden çalışma yapılmıştır. Çalışma kapsamında gerçekleştirilen aşamalar arasında, toplumun iyimser ve kötümser ruh halinde olması ile endeksler arasındaki korelasyon analiz edilmiştir. Toplumun ruh hali her iki durumda da Chicago Board Options Exchange Volatility Index VIX arasında pozitif yönde korelasyon göstermiştir. Dow Jones, S&P 500 ve NASDAQ endeksleri ile ise de negatif yönde bir korelasyon gösterdiği görülmüştür. Bu durum için varılan sonuç ekonomik belirsizlik durumlarında, endekslerin hareketinden bağımsız olarak, bireylerin umut, korku ve endişe halleri gibi duygusal durumları içeren mesajlar attıkları gözlenmiştir (Xue vd ,2010:58-60). Aynı dönemde yapılan başka bir çalışmada Twitter üzerinde yapılan duygu analiz sonuçları ile Dow Jones Borsası Endüstri Endeksi (DJIA), NASDAQ 100 Endeksi ve 13 teknoloji firmasının hisse senetleri fiyatlarının hareket yönü arasında korelasyon analizi gerçekleştirilmiştir. Duygu analizi kapsamında dört milyonun üzerinde tweet analize tabi tutulmuştur. Haziran 2010 ile Temmuz 2011 tarihleri arasında inceleme gerçekleştirilmiştir. Elde edilen sonuçlarda pozitif ve negatif olarak sınıflandırılmış toplumun ruh hali ile hisse senedi ve endekslerin hareket yönü arasında yüksek seviyede bir korelasyon ilişkisi gözlenmiştir (Rao ve Srivastava,2012:1-5).

2015 yılında yapılan bir başka araştırmada toplumun çeşitli firmalar hakkında attıkları tweetleri duygu analizi ile sınıflandırmış ve firmaların hisse senetlerindeki değer değişimi arasında ilişki araştırılmıştır (Dickinson ve Hu,2015:62). Normalize edilmiş duygu analiz sonuçları ile hisse senedi fiyatlarındaki değişim arasında ilişki olduğu tespit edilmiştir. Bu ilişki Walmart ve Microsoft gibi firmalar için yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu ortaya çıkarken, Goldman Sachs ve Cisco gibi firmalar için de yüksek düzeyde, negatif ve anlamlı bir ilişki olduğu ortaya çıkmıştır(Dickinson ve Hu,2015:67-70). 2016 yılında Microsoft firması için benzer bir çalışma yürütmüştür. Araştırma çerçevesinde Ağustos 2015 ile Ağustos 2016 tarihleri arasında Microsoft hakkında atılan 250.000 tweet toplanmış ve analize alınmıştır. İlgili metinler pozitif, negatif ve nötr olarak sınıflandırılmıştır. . Araştırmanın başlangıcında pozitif tweetlerin oranı ile hisse senetlerinin hareket yönü arasında ilişki olduğuna dair varsayımları elde ettikleri araştırma sonuçları tarafından desteklenir konumda olmaktadır. Araştırma sonucunda duygu analizi sonuçları ile hisse senetleri fiyatlarındaki değişim yönü arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu tespit

edilmiştir (Pagolu vd., 2016:2). Bu alanda ülkemizde yapılan çalışmalar arasında Eliaçık ve Erdoğan'ın 2016 yılında yaptıkları çalışma yer almaktadır. Bu çalışmada mikro bloklardaki finans topluluklarının duygu polaritesi ile Borsa İstanbul 100 endeksinin hareketleri arasında korelasyon analizi yapmışlar orta düzeyde, pozitif ve anlamlı bir ilişki olduğunu ortaya çıkarmışlardır (Eliaçık ve Erdoğan,2016:782).

### 3. Tasarım ve Yöntem

Araştırma kapsamında incelemeye alınacak zaman dilimi olarak 15.03.2017 -15.05.2017 tarihleri belirlenmiş ve Borsa İstanbul'un işlem gördüğü günlerde veri toplanmıştır. Twitter ortamında yer alan arama sayfasında içinde anahtar kelime olarak ilgili telekom firmasının adı ile arama gerçekleştirilmiştir. İlgili sayfa da dil olarak Türkçe seçilmiş, firmanın sektördeki rakiplerinin adlarını içeren tweetler seçim dışında tutulmuştur. Bu kapsamda ilgili tweetler python ve tweeter için yazılmış API'ları kullanarak veri tabanına aktarılmıştır. Bu yöntem ile toplam elde edilen tweet sayısı 9.584 adettir. Bu sayı içinde tekrarlanan tweetler silindikten sonra eldeki tekil tweet sayısı 8.475 olarak yer almaktadır. Bu aşamadan sonra veri temizleme aşamasına geçilmiştir. Eldeki metinler arasında yoğun olarak yabancı dilde kelimeler içeren tweetler, firmanın kendi hesabından gönderilen tweetler, sadece firmanın ismi geçen yer ve etkinliklerdeki yer bildirimleri içeren tweetler, 4 adedin üzerinde “#” veya “@” karakteri içeren tweetler değerlendirme dışında tutulmuştur. Bu kapsamda değerlendirmeye alınan tweet sayısı 6.009 adettir. Bu veriden 500 âdeti eğitim verisi olarak ayrılmıştır. Değerlendirme ve analize tabi olan tweet sayısı olarak 5.509 adet olarak yer almıştır.

Bu aşamadan sonra gereksiz karakter ve kelimelerin temizlenmesi aşamasına geçilmiştir. Bu kapsamda noktalama işaretleri ve basılamayan karakterler boşluk ile değiştirilmiştir. Tüm karakterler küçük harfe çevrilmiştir.

Eldeki veri seti içerisinde 500 âdeti eğitim seti olarak seçilmiştir. Veri setlerinde bulunan tweetlere el yordamı ile pozitif ve negatif sınıflandırılma işlemi yapılmıştır. Bu aşamadan sonra eldeki eğitim ve analiz veri setleri Weka(3.8) programında işlem yapılabilmesi için arff formatında kayıt edilmiştir.

Bir sonraki aşama ise, sınıflandırma algoritmasının seçimi olmuştur. Bu kapsamda değerlendirmeye alınacak sınıflandırma algoritmaları ZeroR, J48, K En Yakın Komşu Yöntemi ve Naive Bayes olarak yer almaktadır. Sınıflandırma aşamasında öznitelikler N-Gram model için 3-gram seçilmiş ve uygulanmıştır.

ZeroR Sınıflandırma algoritması en temel ve diğer algoritmalar için temel değer çıkarabilecek sonuçlar üretmektedir. Algoritma eğitim veri seti için frekans tablosu oluşturup, en yüksek frekansa sahip sonuç sınıf değerini tespit eder. Test, veri setinin tüm veri kalemleri bu değer olacağını ön görüp sonuç tablosu oluşturulmaktadır (Foster, vd.,2013:3).

J48 bazı kaynaklarda C4.5 olarak adlandırılan algoritma entropi ölçütüne dayalıdır. Karar ağacı yapısıdır. ID3 algoritmasının genişletilmiş halidir. Her bir düğümde nitelikler arasında en yüksek bilgi kazancına göre karar ağacının oluşumunu ve dallarını oluşturur (Akçetin ve Çelik,2014:49).

K En Yakın Komşu Yöntemi algoritmasında öğrenme kümesindeki verilerin benzerlikleri hesaplanmaktadır. Yakın olduğu düşünülen verinin ortalamasıyla verilmiş eşik değere göre sınıflandırma gerçekleştirilir (Çalışkan ve Soğukpınar, 2008,:122). Bu analizde, en yakın komşu sayısı k=3 alınmıştır.

Naive Bayes algoritması Bayes teorisi tabanına oturmuş sınıflandırıcıdır. Yapısı gereği olasılıksal sınıflandırıcıdır. Veride yer alan her bir sınıf için olasılık oranları hesaplanmaktadır. Sınıflandırma aşamasında ise olasılığı en yüksek sınıfı bulmayı hedeflemektedir (Sezgin ve Çelik,2013:3). Sınıflandırma gerçekleştirilirken Bayes teorisinin yapısı gereği tweetin ait olma olasılığı olan sınıfların koşullu olasılığı hesaplanır (Çalış vd,2013:3). Yaygın kullanımının nedenleri arasında yüksek başarı performansının yanında yüksek hesaplama hızı gelmektedir.

Bayes Teoremi Denklem 1 ifade edildiği şekilde formüle edilmektedir (Bahadır, 2008:50):

$$P(A/B) = (P(B/A) * P(A)) / P(B) \quad (1)$$

$P(A|B)$ : B olayı olduğunda A olayının gerçekleşme olasılığı

$P(B|A)$ : A olayı olduğunda B olayının gerçekleşme olasılığı

$P(A)$ : A olayının gerçekleşme olasılığı

$P(B)$ : B olayının gerçekleşme olasılığı

Algoritmaların Weka programı ile çalıştırılırken test yöntemi olarak "10-kat çapraz doğrulama" metodu kullanılmıştır. Bu yöntemde test verisi rassal olarak 10 parçaya ayrılır. Bunların 9 âdeti öğrenme bir tanesi de test kümesi olarak kullanılır. Öğrenme verisinden elde edilen sonuçlar çerçevesinde sınıflandırma test verisi üzerinde yapılır. Bu işlem 10 defa tekrarlanır. Performans sonuçlarının ortalaması alınarak işlem sonunda kullanıcıya iletilir.

Algoritmaların performans değerlendirilmesi aşamasında performans kriterleri olarak doğruluk, kesinlik, duyarlılık ve F-ölçütü olarak yer almaktadır. En temel performans ölçütü olarak doğru olarak sınıflandırılan veriler ile yanlış olarak sınıflandırılan verilerin sayısal dağılımı olarak yer almaktadır. Her test sonucunda başarı bilgileri karışıklık matrisi ile verilmektedir. Bu matriste satırlar test kümesindeki örneklere ait gerçek değerler, kolonlar ise modelin sınıflandırma sonuçlarını verir. Örnek karışıklık matrisi Tablo 1'de ve ilgili performans değerlendirme kriterlerine ait formüller aşağıda yer alan Formül 2,3,4 ve 5'de verilmektedir (Coşkun ve Baykal, 2011:53-54).

Tablo 1: Örnek Karışıklık Matrisi

		Model Sınıflandırması	
		Sınıf = 1	Sınıf = -1
Gerçek Değerler	Sınıf = 1	TP	FN
	Sınıf = -1	FP	TN

$$\text{Doğruluk} = \frac{[TP+TN]}{[TP+TN+FN+FP]} \quad (2)$$

$$\text{Keskinlik} = \frac{TP}{[TP+FP]} \quad (3)$$

$$\text{Keskinlik} = \frac{\text{Doğru Sınıflandırılmış Pozitif Örnek Sayısı}}{\text{Pozitif Sınıflandırılmış Örnek Sayısı}}$$

$$\text{Duyarlılık} = \frac{TP}{[TP+FN]} \quad (4)$$

$$\text{Duyarlılık} = \frac{\text{Doğru Sınıflandırılmış Pozitif Örnek Sayısı}}{\text{Pozitif Örnek Sayısı}}$$

$$F\text{-Ölçütü} = \frac{[2 * \text{Keskinlik} * \text{Duyarlılık}]}{[\text{Keskinlik} + \text{Duyarlılık}]} \quad (5)$$

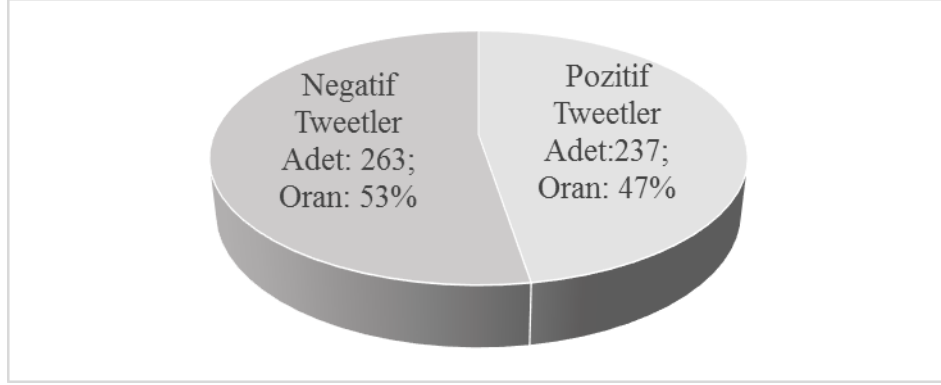
Analiz kapsamında seçilen firma, Borsa İstanbul' da 15 yılı aşan süreden beri işlem görmektedir. Telekomünikasyon sektöründe faaliyet gösteren firma sektöründe önemli firmalardan bir tanesi olarak BIST 30 Endeksinde yer almaktadır. Endekste ağırlık değeri üzerinden ilk beş sırada yer almaktadır. Firmanın hisselerinin %45'inden fazlası halka açık konumda olarak yer almaktadır. Firmanın, Borsa İstanbul'daki işlem geçmişi, sermaye büyüklüğü, işlem gören hisselerin büyüklüğü ve firmanın toplum tarafından bilinirliği göz önüne alındığında firmanın Borsa İstanbul'da işlem gören firmalar arasında önemli bir firma olduğu söylenebilmektedir. Bu firmanın hisselerinin açılış

ve kapanış değerleri msn.com web sitesinden elde edilmiştir. Araştırmada tarih aralığı olarak 15.03.2017 -15.05.2017 seçilmiştir. Bu tarih aralığında ilgili firma hissesi toplam 43 gün Borsa İstanbul'da hisse işlem görmüştür.

#### 4. Bulgular ve Tartışma

El yordamı ile sınıflandırılan eğitim seti 500 adet örnekten oluşmaktadır. Bunların pozitif ve negatif kayıtlarının dağılımı Şekil 1' de yer almaktadır.

Şekil 1: Eğitim Seti Veri Dağılımı



Eldeki veri setinin sınıflandırılması aşamasına geçmeden önce sınıflandırma algoritmasının seçilmesi gerekmektedir. ZeroR, J48, K En Yakın Komşu Yöntemi ve Naive Bayes değerlendirmeye alınan algoritmalar olarak yer almaktadır. Başarı ölçütleri olarak doğruluk, kesinlik, duyarlılık ve F-ölçütü seçilmiştir. Bu değerler aşağıda yer alan Tablo 2'de sunulmaktadır.

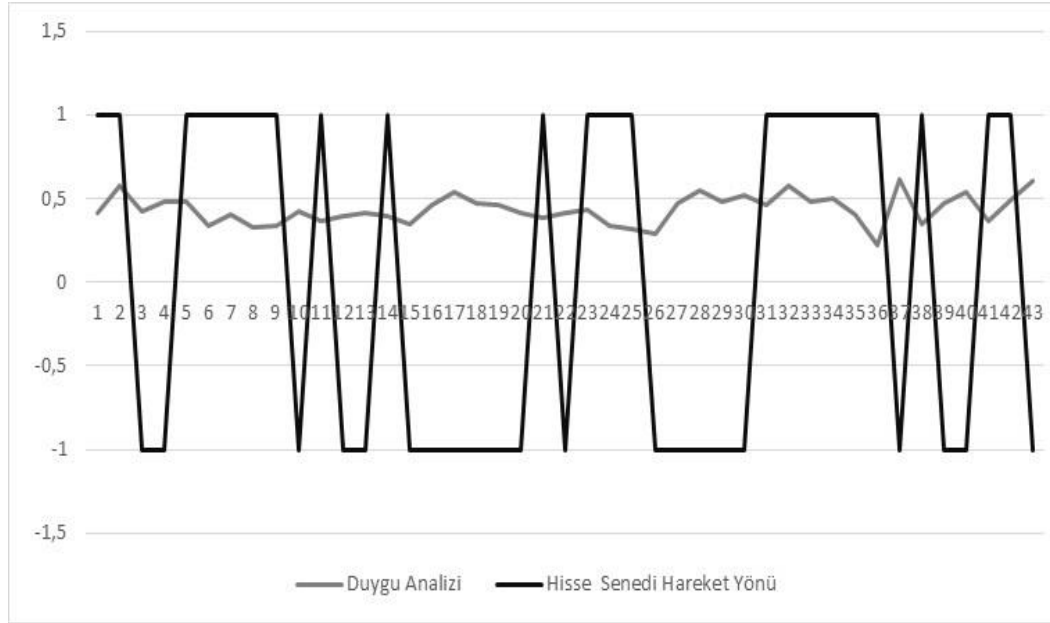
Tablo 2: Algoritma Başarı Ölçüt Değeri

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F-Ölçütü
ZeroR	53%	**	0%	**
J48	65%	71%	45%	55%
K En Yakın Komşu	47%	100%	64%	47%
Naive Bayes	70%	72%	60%	66%

Bu kapsamda en iyi performans gösteren algoritma Naive Bayes olarak yer almıştır. Bu aşamadan sonra eldeki analiz verisi olan 5.509 adet tweet eğitim verisi ile Naive Bayes algoritması ile sınıflandırma gerçekleştirilmiştir. Her bir gün için atılan pozitif ve negatif tweetlerin toplamı alınmıştır. Bir sonraki aşamada ise, her bir gün için o gün içinde atılan pozitif tweetleri sayısı o gün için atılan tweet sayısına bölünerek pozitif tweetlerin yüzdesel oranı elde edilmiştir.

Borsa verisi işleme sürecinde o gün için açılış değeri ile kapanış değerinin farkı alınmıştır. Kapanış değeri açılış değerinin üzerinde yer alması durumunda fark değeri pozitif bir değer almakta olup, hisse senedi değer yönü verisi olarak 1 verilmiştir. Tersisi durumunda ise bu değer -1 olarak ilgili gün için kayıt oluşturulmuştur. İlgili veriler Şekil 2 de verilmiştir.

Şekil 2: Duygu Analizi Sonuçları ve Hisse Senedi Fiyat Hareket Yönü



Tablo 3: Duygu Analizi Sonuçları ve Hisse Senedi Fiyat Hareket Yönüne İlişkin Shapiro-Wilks Testi Sonuçları

	İstatistik Değeri	Serbestlik Derecesi	p
Duygu Analizi Sonuçları	,988	43	,931
Hisse Senedi Fiyat Hareket Yönü	,637	43	,000

Tablo 4: Duygu Analizi Sonuçları ve Hisse Senedi Fiyat Hareket Yönüne Arasındaki İlişki Tablosu

		Duygu Analizi Sonuçları	Hisse Senedi Fiyat Hareket Yönü
Duygu Analizi Sonuçları	Spearman Korelasyonu	1,000	-,349*
	İstatistiksel Anlamlılık (2-yönlü)	-	,022
	N	43	
Hisse Senedi Fiyat Hareket Yönü	Spearman Korelasyonu	-,349*	1,000
	İstatistiksel Anlamlılık (2-yönlü)	,022	-
	N		43

\*Korelasyon değeri 0,05 düzeyinde anlamlıdır (2 yönlü)

Çalışma kapsamında olan değişkenlerin normallik sınaması için Shapiro-Wilks testi gerçekleştirilmiştir. Bu test sonucunda Hisse Senedi Fiyat Hareket Yönü değişkeninin anlamlılık düzeyleri ( $p < 0,00$ ) olduğu için normal dağılıma uygun olmadığı belirlenmiştir. İlgili analiz sonuçları Tablo 3'de sunulmuştur. Bu aşamadan sonra korelasyon analizi yapılırken Spearman'ın Sıra İlişkisi Testi tercih edilmiştir. Bu konu ile ilgili yapılan analiz sonucunda bireylerin firma hakkında gönderdiği pozitif tweetler ile hisselerin günlük fiyat hareket yönü arasında 0.05 anlamlılık düzeyi ve negatif yönde ( $r = -0,349$ ,  $p = 0,022$ ) ilişkisi olduğu tespit edilmiştir. İlgili analiz sonuçları Tablo 4'de sunulmuştur. Bu değerlere göre pozitif tweetlerin oranı artığında hisse senedi fiyatının yönü açılış değerine göre daha düşük değerler alabileceği söylenebilir. Determinasyon katsayısı ( $r^2 = 0,12$ ) dikkate alındığında hisse senedinin fiyat hareket yönündeki toplam varyansın %12'sinin bireylerin gönderdiği pozitif tweet yüzdesinden kaynaklandığı söylenebilir. Ancak teorik olarak açıklanan

varyans diğer değişken olan pozitif tweet yüzdesi içinde yorumlanabilir. Başka bir anlatımla , bu analiz bize neden-sonuç bağlamında yorumlama olanağı vermez, sadece değişkenlerin birlikte hangi düzeyde ve hangi yönde değiştikleri konusunda fikir verir (Büyüköztürk,2011:34).

### 5.Sonuç

Hisse senetlerinin fiyatlarının hareket yönü, işlem gördükleri platformun yapısı gereği değişik faktörlerden etkilenmektedir. Bu faktörlerin neler olduğunun belirlenmesi, hisse senetlerinin fiyat hareket yönünü etkileme güçleri konusunda bilgi elde edilmesi yatırım kararları alma aşamasında önem arz etmektedir. Borsa İstanbul'un da dahil olduğu gelişmekte olan pazarlardaki spekülasyon hareketlere gelişmiş ülke borsalarına göre daha fazla açık olduğu görülmektedir. Bu yapı, hisse senetleri fiyatlarının hareket yönünde hızlı yön değişimi göstermesine neden olmaktadır. Bu çerçeveden incelendiğinde, fiyatların hareket yönünü etkileyen faktörlerin neler olduğu ve etki güçleri konusunda yapılacak çalışma sonuçları Borsa İstanbul'da yatırım kararı alan kişiler açısından önem arz eden etmenlerden bir tanesidir. Sosyal medyada yer alan mesajlar ile Borsa İstanbul'da işlem gören hisse senetlerinin hareket yönü arasında ilişki, yoğun olarak araştırılan bir başlık olarak yer almamaktadır. Çalışmada, duygu analizi ile elde edilmiş polarizasyon değeri ile hisse senetlerinin fiyatlarının günlük hareket yönü arasında ilişki araştırılmıştır. Araştırmaya konu olan twittler üzerindeki duygu analizi sonucu elde edilen günlük duygu polaritesi değeri ile hisse senetlerinin fiyatlarının günlük hareket yönü arasındaki korelasyon analizinde ilgili hisse için üç temel sonuç çıkma ihtimali yer almaktadır. Literatürde başka sermaye piyasalarında aynı zaman aralığında farklı şirketler için yapılan korelasyon analizinde farklı sonuçların çıktığı görülmüştür. Yapılan çalışma sonucunda araştırmaya konu olan telekom firması için polarizasyon değeri ile günlük hisse hareket yönü arasında negatif yönde ilişki saptanmıştır. Bu sonuç elbette, Borsa İstanbul'da işlem gören diğer hisseler için genelleme yapılması imkânını sağlamaz. Sektör ve hisse senedi bazında daha geniş bir yapıyı inceleyen ileride yapılabilecek çalışmalar ile Borsa İstanbul'da işlem gören hisse senetlerinin, fiyat hareket yönü ile bireylerin duyguları arasındaki ilişkisi konusunda daha detaylı bilgiye sahip olunabilme imkânı doğacaktır.

Elektronik ağızdan ağıza iletişim konusunda yapılan araştırmaların, ürün ve hizmet değerlendirmesi, satın alma kararlarına etkisi üzerinde yoğunlaştığı görülmektedir. Firmaların finansal piyasalarda özellikle Borsa İstanbul'da işlem gören hisse senetlerinin hareketleri ile elektronik ağızdan ağıza iletişimin etkileşimi yoğun çalışılmamış bir başlık olarak yer almaktadır. Araştırma, elektronik ağızdan ağıza iletişim ile hisse senetlerinin hareket yönü arasında anlamlı ilişkinin varlığını göstermiştir. Araştırma verilerin alındığı zaman aralıkları değerlendirildiğinde ise, elektronik ağızdan ağıza iletişimin diğer başlıklarla nispeten kısa zaman diliminde etkileşime geçtiği ve sonuçların görüldüğü gözlemlenmektedir.

Pazarlama alanında çalışan profesyoneller açısından bakıldığında elektronik ağızdan ağıza iletişim konusunda gerçekleştirilecek analiz aşamaları mesajların elde edilmesi, ilintisiz mesajların temizlenmesi, doğal dil işleme süreçlerinin tamamlanması, makine öğrenme süreçlerini içeren sınıflandırmanın gerçekleştirilmesi başta olmak üzere dinamik yapıları içermektedir. Yapılan araştırma, ilgili süreçlerde kullanılacak bir dizi yöntem, model ve araç göstermesi açısından katkı sağlayacaktır.

Firmaların, bireylerin firma ile ilintili konulardaki duygu ve düşünceleri takip etmesinin gerekliliği her geçen gün artmaktadır. Firmalar, elektronik ağızdan ağıza iletişimde sadece pozitif mesajlara yoğunlaşmasının ötesinde, negatif mesajlar üzerinde durulmaması gerektiği kabul edilen bir değer olarak yer almaktadır. Polarizasyon değeri bütüncül yapısı gereği ölçülmesi ve izlenmesi gereken bir parametre olarak yer almaktadır. Yapılan araştırma, polarizasyon değerinin ölçülmesi ve kullanımını göstermesi açısından önem taşımaktadır.

Finansal piyasalarda işlem yapan yatırımcılar açısından bakıldığında, yatırımcıların sosyal medya özellikle twitter üzerinden elde edilecek bilgiye ilgisinin giderek arttığı görülmektedir. Borsada işlem yapan firmaların ve hedge fon yöneticilerin kısa vadeli işlem kararlarında, tweetler kullanılarak



gerçekleştirilen duygu ve fikir analiz sonuçlarından yararlanmaktadırlar (Rao ve Srivastava,2012:1-2). Borsa İstanbul'da işlem yapan yatırımcılar kısa vadeli işlemlerinde karar alma sürecinde belirli göstergeler çerçevesinde karar almaktadır. Elektronik ağızdan ağıza iletişimde izlenmesi gereken platformlardan bir tanesi olduğu gerçekleştirilen araştırmanın sonuçları arasında görülmektedir.

Yapılan çalışmadaki en önemli limit olarak görülebilecek nokta, incelenen firma sayısı olarak yer almaktadır. Özellikler takip edebilecek araştırmalar için sunulabilecek en önemli öneri firma sayısının artırılması olacaktır. Özellikle B2B ve B2C pazarlarda yer alan firmaların analize dahil edilmesi ile, analiz pazar bazında farklılaşan firmalar arasında olası benzerlikleri ve farklılıkları görebilme imkanına sahip olunabilecektir. Araştırma, tek bir borsa üzerinden yürütülmüştür. Farklı ülke borsalarında işlem gören aynı firmalara ait hisse senetleri üzerinde yapılabilecek çalışma ile sonuçlar arasında karşılaştırma imkânı verebilecektir. Araştırmanın diğer bir limiti ise, duygu analizi sonuçlarından sadece polarite değerinin kullanılması olmuştur. Pozitif ve negatif yorumların sayısı, bu sayıdaki değişim gibi diğer veri başlıklarında kullanılması araştırma konusunda daha detaylı bilgiye sahip olabilme imkânına kavuşabilecektir.

#### Kaynakça

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O ve Passonneau, R. (2011). Sentiment Analysis of Twitter Data. In Proceedings of The ACL 2011 Workshop On Languages in Social Media,30–38.
- Akçetin, E. ve Çelik, U. (2014). İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması. *Internet Uygulamaları ve Yönetimi*, 5(2),43-56
- Albeni, M. ve Demir Y. (2005). Makroekonomik Göstergelerin Mali Sektör Hisse Senedi Fiyatlarına Etkisi (İMKB Uygulamalı). *Muğla Üniversitesi SBE Dergisi*, 14, 1-18.
- Bahadır, İ. (2008). Bayes Teoremi ve Yapay Sinir Ağları Modelleriyle Borsa Gelecek Değer Tahmini Uygulaması,(Yayımlanmamış Yüksek Lisans Tezi). TOBB Ekonomi ve Teknoloji Üniversitesi, Ankara.
- Bollen, J., Mao, H. ve Zeng, X. (2011). Twitter Mood Predicts The Stock Market. *Journal of Computational Science*, 2(1), 1-8.
- Büyüköztürk, Ş. (2011).*Veri Analizi El Kitabı*. Ankara: Pegem.
- Çalışkan, S. K. ve Soğukpınar, İ. (2008). KxKNN: K-Means ve K En Yakın Komsu Yöntemleri ile Ağlarda Nüfuz Tespiti. *EMO Yayınları*,120–124.
- Çalış, K., Gazdağı, O. ve Yıldız, O.(2013). Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti. *Bilişim Teknolojileri Dergisi*, 6(1),1-7.
- Coşkun, C. ve Baykal, A. (2011). Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması. *Akademik Bilişim, Malatya*,51-58.
- Dickinson, B. ve Hu, W. (2015). Sentiment Analysis of Investor Opinions on Twitter. *Social Networking*, 62-71.
- Eliaçık, A. B. ve Erdoğan, N. (2015). Mikro Bloglardaki Finans Toplulukları için Kullanıcı Ağırlıklandırılmış Duygu Analizi Yöntemi. *UYMS*,782-793.
- Foster, M. E, A., Gaschler, A. ve Giuliani. M. (2013). How Can I Help You: Comparing Engagement Classification Strategies For A Robot Bartender. *Proceedings of ICMI'13, Sydney, Australia*, 1-7,
- Karaatlı, M., Güngör, İ., Demir, Y. ve Kalaycı, Ş. (2005). Hisse Senedi Fiyat Hareketlerinin Yapay Sinir Ağları Yöntemi ile Tahmin Edilmesi. *Balıkesir Üniversitesi İİBF Dergisi*, 2(1), 38-48.
- Liu, B. ve Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. Aggarwal, C. ve Zhai, C. (Ed.), *Mining Text Data*. Springer, 415–463.

- Miranda, M. ve Sassi, R. (2014). *Using Sentiment Analysis to Assess Customer Satisfaction in An Online Job Search Company. Business Information Systems Workshops*, Springer, Lecture Notes in Business Information Processing, Volume 183, 17 – 27.
- Moe, W. W. ve Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums, *Journal of Marketing Research*, 48 (3), 444-456.
- Nguyen, VD., Varghese, B. ve Barker, A. (2013). The Royal Birth Of 2013: Analysing and Visualising Public Sentiment in The UK Using Twitter. *IEEE International Conference on Big Data*, 1-9.
- Pagolu, V., Challa, K., Panda, G. ve Majhi, B. (2016). Sentiment Analysis of Twitter Data For Predicting Stock Market Movements. *SCOPES 2016*,1-5.
- Rao, T. ve Srivastava, S. (2012). Analyzing Stock Market Movements Using Twitter Sentiment Analysis. *Proceedings of the ASONAM 201*. IEEE Computer Society, 1-5.
- Rao, T. ve Srivastava, S. (2012). Twitter Sentiment Analysis: How To Hedge Your Bets in The Stock Markets. Working paper, Indian Institute of Technology, Delhi, 1-22.
- Sezgin, E. ve Çelik, Y. (2013). Veri Madenciliğinde Kayıp Veriler İçin Kullanılan Yöntemlerin Karşılaştırılması, *Akademik Bilişim Konferansı*, 1-4.
- Xue, Z., Fuehres, H. ve Gloor, P. A. (2010). Predicting Stock Market Indicators Through Twitter I Hope It Is Not As Bad As I fear. In *Collaborative Innovations Networks Conference (COINs)*, 55-62,
- MSN Finans,(2017). Erişim Adresi: <http://www.msn.com/tr-tr/finans/hissesenediayrintilari>
- Cross-Validation, (2017). Erişim Adresi: <http://www.cs.waikato.ac.nz/ml/weka/mooc/dataminingwithweka/transcripts/Transcript2-5.txt>

---

**INVESTIGATION OF RELATIONSHIP BETWEEN SOCIAL MEDIA AND STOCK PRICE  
DAILY MOVEMENT DIRECTION: SENTIMENT ANALYSIS IMPLEMENTATION**

---

**Extended Abstract**

**Aim:** The main aim of this research is briefly to investigate the correlation between daily stock price movements of a company and public sentiment towards to that company. The chosen company is Telco company and both operating on B2B and B2C markets. Its shares have been listed on the Borsa Istanbul over 15 years. The company is among the top five companies on BIST 30 Index in terms of its weight on the Index. Over 45 % of shares are traded on stock markets. The company has been chosen because of its long trade history on Borsa İstanbul, high trade volume large number shares, operating both on B2B and B2C markets and well known company among people. The stock price movement are being measured on binary scale. Up and downs are being classified according to differences between daily opening and closing stock prices of the company. In these research, in obtaining the public sentiment toward the company, the tweets which contains public opinion about the company have been used. Main purpose of sentiment analysis is to classify the writer's sentiment from given text. In this research, we only used tweets that contains sentiment towards the company. Supervised machine learning methods have used to classify the tweets in order to conduct the sentiment analysis In the research, a training set in telecom sector for sentiment analysis has been developed which was not only used in this research but could be used on the future sentimental analysis on telecom industry. Another aimed output of the research is to compose classification model that will yield satisfactory performance results. In this part, for different models each using different classification algorithms namely ZeroR, J48, K- nearest neighbors algorithm and Naive Bayes has been tested on the developed training set.

**Method(s):** In text mining phase, Natural Language Process (NLP) methods have allocated before commencing supervised machine learning methods. In these face, all relevant NLP actions, such as removal of unrelated characters and symbols, have been conducted. In supervised machine learning phase, 4 models that are using ZeroR, J48, K - nearest neighbors algorithm and Naive Bayes algorithms have been deployed. After retrieving the sentiment analysis and daily stock movement data, Shapiro-Wilks test have been used to test the normality of these two data sets. Hence the data sets foundnot to be normally distributed. Spearman's correlation test has been conducted on these data test to calculate Spearman's rank-order correlation coefficients. The Spearman's rank-order correlation coefficients measure statistically direction and strength of two ranked data sets' association.

**Findings:** The first finding is the model evaluation. We have found that best performing classification is the one using the Naive Bayes Algorithm. For the model, Accuracy value is 70 %, Precision value is 72 %,, Recall value is 60 % and finally F-measure value is 66 %. Naive Bayes Algorithm has the best performance values among the other tested models which are using ZeroR, J48 and K -nearest neighbors algorithms. In this study, correlation analysis between public sentiment analysis and stock price movement direction is conducted . In the final phase, correlation analysis between public sentiment analysis and stock price movement direction is conducted Hence the data sets are not nominally distributed. Spearman's rank correlation test is used. The obtained Spearman's rank correlation coefficient result showed , a negative and moderate correlation exists between daily chances in the company's stock prices with the public sentiments in tweets.

**Conclusion:** Social media is becoming one of main communication channels. Twitter is considered to be one of the major social media platform. Very large of amount content has been generated daily by public on Twitter each day. The people are expressing their opinions and sentiments on these tweets. Due to these reasons, twitter is considered to be one of the main data source to conduct sentiment analysis. In our research, we have created a model which is using supervised machine learning methods. The model is conducting sentiment analysis using the tweets which

contains sentiments on a Telco company. This model could be used to conduct sentiment analysis for any other Telco company. In the output of the sentiment analysis, daily polarity value has been obtained. Conducted Spearman's rank correlation test on obtained daily polarity values and Stock Price Daily Movement Direction values have shown that a a negative and moderate correlation exists between daily stock price movements and public sentiment. The one of the main contribution of this research is that it showed the existence of correlation between public sentiment and daily stock price changes for a company whose stocks are traded in Borsa Istanbul.