



# Düzce Üniversitesi Bilim ve Teknoloji Dergisi

Araştırma Makalesi

## Twitter’da Duygu Analizi Yöntemi Kullanılarak Bitcoin Değer Tahminlemesi

 Burak KÖKSAL <sup>a,\*</sup>,  Gözde ERDEM <sup>a</sup>,  Cansu TÜRKELİ <sup>a</sup>,  Zehra KAMIŞLI ÖZTÜRK <sup>a</sup>

<sup>a</sup> Endüstri Mühendisliği Bölümü, Mühendislik Fakültesi, Eskişehir Teknik Üniversitesi, Eskişehir, TÜRKİYE

\* Sorumlu yazarın e-posta adresi: burakkoksal.info@gmail.com

DOI: 10.29130/dubited.792909

### ÖZET

Bütün sektörler dahilinde finans sektöründe de müşterilere ait fikir ve düşüncelerinin belirlenmesi, firma ve kurumların ileriki dönemler için sunacağı hizmetleri etkilemektedir. Kripto para birimlerinin (Bitcoin, Ethereum, Ripple vb.) ekonomik ve sosyal etkileri hızla artmaya devam ettikçe, ilgili haber makalelerinin ve sosyal medya yayınlarının, özellikle de tweetlerin yaygınlığı da artmaktadır. Bu çalışmada, Twitter kullanıcılarının finans sektörü konularından biri olan Bitcoin ile ilgili yorumları derlenerek bir duygu analizi çalışması yapılmıştır. Kullanıcı yorumları, Twitter’ın sunmuş olduğu API hizmeti vasıtasıyla Python Programlama Dili kullanılarak alınmış; yorumlar olumlu, nötr ve olumsuz etiketler ile ayrıştırılmış, etiket bulutunda toplanmıştır. Naïve Bayes ve Lojistik Regresyon algoritmaları kullanılarak oluşturulan modellerde başarı oranları karşılaştırılmıştır. Naïve Bayes uygulamasının tweetlerin duygularını tahmin etmedeki başarı oranı %72,19 olurken, Lojistik Regresyon uygulamasında bu oran %75,53 olmuştur. Çalışmanın ikinci aşamasında ise, duygu analizinden sonra “Bitcoin” anahtar kelimesi içeren günlük pozitif tweet oranı ile Bitcoin günlük açılış değeri beraber kullanılarak Bitcoin kapanış değeri tahminlemesi yapılmıştır. Finans verileri Yahoo Finance web sitesi üzerinden alınmış; Doğrusal Regresyon ve Rastgele Orman Regresyon yöntemleri ile modeller oluşturulmuştur. Doğrusal Regresyon için  $r^2$  değeri %88,97 çıkarken, Rastgele Orman Regresyonu için ise %94,16 olmuştur.

**Anahtar Kelimeler:** Duygu analizi, Twitter, Bitcoin, Makine öğrenmesi, Veri madenciliği, Finans

## Bitcoin Price Prediction Using Sentiment Analysis on Twitter

### ABSTRACT

The identification of actual and potential customers' opinions before and after purchase shapes the services offered by companies in the financial sector as well as in every sector. Cryptocurrencies as their economic and social impact continues to increase rapidly, the prevalence of related news articles and social media posts, especially tweets, also increases. In this study, sentiment analysis was applied by collecting comments and thoughts about Bitcoin on the social media platform Twitter. User comments were received using the Python Programming Language via the API offered by Twitter. Compiled user comments were separated with positive, negative and neutral tags, and the results were analyzed using Naïve Bayes and Logistic Regression. Success rate of sentiment prediction with Naïve Bayes was found 72,19% and for Logistic Regression 75,33%. After sentiment analysis, this study attempts to predict Bitcoin daily closing value using percentage of positive tweets that include “Bitcoin” keyword and Bitcoin daily opening value. Yahoo Finance has been the source of the financial data used in this study. Models were created using Linear Regression and Random Forest Regression. The  $r^2$  value for Linear Regression was found 88,97%, for Random Forest Regression the  $r^2$  value was found 94,16%.

**Keywords:** Sentiment analysis, Twitter, Bitcoin, Machine learning, Data mining, Finance

# I. GİRİŞ

Günümüzün ticarileştirilmiş küresel manzarasında, daha erişilebilir ve şeffaf bir ortam için talep para birimi hızla büyümektedir. Özellikle 21. yüzyılın başından itibaren, finansal teknolojilerde çok büyük gelişme ve değişimler olmaktadır. İnternet erişiminin varlığı da yeni bir parasal sistemi tetiklemiş ve dijital para cinsi olarak bilinen “Kripto Para” (Cryptocurrency) kavramını hayatımıza sokmuştur. İnternetin bankasız ve devletsiz para birimi olan Bitcoin, İnternet üzerinde dijital varlıkların güvenli bir şekilde transfer edilmesini sağlamaktadır. Bu da Bitcoin, eski finansal araçlara olan ihtiyacı ortadan kaldırarak İnterneti daha demokratik ve efektif hale getirmektedir. Bitcoin ve ilgili teknolojiler tarım, lojistik ve sanayi sektörlerine kadar birçok farklı sektörde büyük çözümler sunmaktadır.

Sosyal medya madenciliği, veri madenciliği ilkelerinin sosyal medya üzerinde uygulanması ile ortaya çıkmıştır. Sosyal medya madenciliği çalışmalarında sosyal medya üzerinden veriler toplanır, düzenlenir ve analiz edilerek anlamlı sonuçlar çıkarılmaya çalışılır. Büyük verilere sahip en bilinen sosyal medya platformları Facebook, Google+ ve Twitter olarak verilebilir. Bu çalışma kapsamında ele alınan Twitter, kullanıcılarına 280 karakterle sınırlandırılmış “tweet” isimli mesajlar ile sosyal ağlarında bulunan takipçileri ile iletişim kurabilme olanağını sunmaktadır. Twitter üzerinden günde yaklaşık 500 milyon mesaj (tweet) paylaşılmaktadır. Atılan tweetlere ait bilgiler depolanırken içeriğin yanı sıra etiketler (hashtag), kullanıcı bilgisi, tweet saati ve tweet lokasyonu gibi bilgiler de kaydedilmektedir. Bu durumda tweet başına 1 KB ham veri boyutu oluşmaktadır. Yani dünyada sadece 1 günde atılan tweetlerin ham veri boyutu 500 GB civarındadır. Verilerin boyutları sosyal medya madenciliğinin bir büyük veri teknolojisi olduğunu göstermektedir.

Sosyal medya üzerinden toplanan bu verilerin analizinde, sosyal medya madenciliğinin en önemli çalışma alanlarından biri olan duygu analizi yöntemi kullanılmaktadır. Duygu analizi metinler içinde gizli bir şekilde yer alan düşünce ve duyguların belirlenmesi ve kullanılabilir bir biçime getirilmesidir. Duygu analizi bir yazarın veya bir yazının belirli bir konuya yönelik tutumunun olumlu (positive), olumsuz (negative), tarafsız (neutral) olarak değerlendirilmesi için bir metin parçasıyla belirtilen görüşleri/ifadeleri çeşitli algoritmalar aracılığı ile hesaplayıp tanımlama ve sınıflandırma süreçlerine verilen genel bir tanımdır. Buradan yola çıkarak, kripto para birimi fiyat dalgalanmasının birçok nedeni olsa da mevcut çevrimiçi sosyal medyadaki duygu analizinin bir Bitcoin fiyatının (yani, algılanan değer) yükselip yükselmeyeceği konusundaki tahminlerin bilgi verip veremeyeceğini araştırmak finans kullanıcıları için faydalı olacaktır. Duygu analizinde, genellikle kullanılan teknikler makine öğrenmesi ve anlamsal yönelim teknikleridir. Çalışma kapsamında, API’ler (Application Programing Interfaces-Uygulama Programlama Arayüzleri) vasıtasıyla şirketlere ve özel kişilere veriye kolay bir biçimde ulaşma imkanı tanınması sebebiyle Twitter sosyal ağı seçilmiştir. Veri madenciliği araçlarından Naïve Bayes ve Lojistik Regresyon algoritmaları ile de Twitter üzerinden veriler elde edilmiştir. Dolayısıyla, bu çalışmada, duygu analizi problemleri bir tür sınıflandırma problemi olarak tarif edilmiştir. Ayrıca, literatür çalışması kısmında da vurgulandığı üzere, Türkçe duygu analizi kullanılarak bu konuda yapılan çalışma sayısı sınırlıdır. Bu çalışma ile Bitcoin ile ilgili Türkçe duygu analizi literatürüne de bir katkıda bulunmak amaçlanmaktadır.

Bitcoin kullanımının yaygınlaşması ve İnternet kullanımının da anlık olması finanstaki değişimleri veya gelecekteki değişiklikleri tahmin etme ve yorumlama zorunluluğunu da beraberinde getirmektedir. Bitcoin fiyat dalgalanmalarının yönünü belirleyebilmek de günümüz piyasası için büyük bir önem kazanacaktır. Her alandaki yatırımcıya yol gösterici bir rehber niteliğinde olan bu çalışmada, duygu analizi aşamasından sonra Bitcoin kapanış değeri tahminlemesi yapılmıştır. Ancak, bu tahminleri yapmak tam anlamıyla da mümkün değildir. Bu değişimleri yorumlayabilmek amacıyla çeşitli modeller geliştirilmiştir. Bu modelleri oluştururken -kripto para piyasası açısından en güncel verilerin yer aldığı- finans verileri, Yahoo Finance web sitesi üzerinden anlık olarak çekilmiştir. Duygu analizinde kullanılan tweetlerin atıldığı tarih ve saat baz alınarak çekilen finans verileri oluşturulan modellerde kullanılmıştır. Yapılan bu modelde ise “Bitcoin” anahtar kelimesi içeren günlük pozitif tweet oranı ile Bitcoin günlük açılış değeri beraber kullanılarak Bitcoin kapanış değeri tahminlemesi yapılmıştır. Oluşturulan modellerde test ve eğitim setlerine dönüşüm sağlandıktan sonra

Doğrusal Regresyon ve Rastgele Orman Regresyon yöntemleri kullanılarak Bitcoin kapanış değeri tahminlemesi tamamlanmıştır.

Çalışmanın izleyen bölümlerinde sırasıyla; ikinci bölümünde Bitcoin ve duygu analizi ile ilgili literatüre değinilmiş, literatürdeki boşluğa vurgu yapılmıştır. Üçüncü bölümde duygu analizi için gerekli veri işlemleri ve kullanılan sınıflandırma algoritmaları; dördüncü ve beşinci bölümde uygulama ve elde edilen başarı oranları verilmiştir. Son bölümde ise gelinen nokta ve gelecek çalışmalara yer verilmiştir.

## **II. LİTERATÜR TARAMASI**

Çalışma kapsamında literatür taraması iki ana başlık altında ele alınmıştır. Öncelikle Bitcoin ile ilgili yapılmış çalışmalar incelenmiş, ardından, bu çalışmanın odak noktası olan duygu analizi ile ilgili çalışmalara yer verilmiştir.

Karasu vd. [1] tarafından yapılan çalışmada makine öğrenmesi algoritmalarından Doğrusal Regresyon (DR) ve Destek Vektör Makinesi (SVM) ile 2012-2018 yılları arasındaki günlük Bitcoin kapanış fiyatlarından oluşan zaman serisi kullanılarak Bitcoin tahmini yapılmıştır. Önerilen SVM modelinin DR modeline göre daha başarılı sonuç verdiği görülmüştür. Chen vd. [3]'nin yaptığı çalışmada ise, günlük Bitcoin fiyat tahminlemesi için istatistiksel yöntemlerden Lojistik Regresyon (LR) ve Lineer Diskriminant Analizi (LDA); makine öğrenmesi algoritmalarından da Rastgele Orman (RF), XGBoost (XGB), İkinci Dereceden Ayrımcı Analiz (QDA), Destek Vektör Makinesi (SVM) ve Uzun-Kısa Vadeli Hafıza Ağları (LSTM) kullanılmıştır. İstatistiksel yöntemler, günlük Bitcoin fiyat tahmini için %66'lık bir doğruluk elde ederken, makine öğrenme algoritmalarının en yüksek başarı oranı %65,3 olarak elde edilmiştir. Çalışmada yer alan çeşitli sınırlamalar nedeni ile elde edilen en yüksek başarı oranı bile diğer çalışmalara oranla yetersiz kalmıştır. Ayrıca çalışmanın performansı için diğer makine öğrenme algoritmalarından faydalanılmamıştır.

Rahman vd., makine öğrenmesi algoritmalarının uygulanarak yapıldığı duygu analizi ile Bitcoin'in fiyat değişikliği arasındaki ilişkiyi korelasyon kullanılarak incelemiştir. Bu çalışmada sınıflandırma için RFR ve Polinom Regresyon (PR) gibi beş farklı regresyon algoritması ve LR, K-En Yakın Komşu ve Naïve Bayes(NB) olmak üzere 11 farklı sınıflandırma algoritması kullanılmıştır. Regresyon yöntemleri ile elde edilen sonuçlar, makine öğrenmesi algoritmaları ile elde edilen sonuçlarına göre daha başarılı olmamıştır. Bunun en önemli nedeni ise kullanılan verilerin yetersiz olması ve tahminleme aralığının darlığıdır [4]. Lamon vd. de Bitcoin, Litecoin ve Ethereum kripto para birimlerindeki fiyat dalgalanmalarını Doğrusal SVM, Multinomial Naïve Bayes (MNB) ve Bernoulli Naïve Bayes (BNB) ile tahmin etmeyi denemiş ancak en iyi sonuçlar LR ile elde edilmiştir. Bu çalışmadaki en büyük eksiklik ise tahminleme zaman aralığında yer alan fiyat dalgalanmalarındaki büyük değişikliklerin tespit edilememesidir [7]. Valencia vd. tarafından yapılan çalışmada, Bitcoin, Ethereum, Ripple ve Litecoin kriptopara pazarının fiyat hareketini tahmin etmek için Twitter öğeleri kullanılarak Yapay Sinir Ağları (ANN), SVM ve RF algoritmaları kullanılmış ve piyasa verilerinin girdi özellikleri de kullanılarak tahminleme yapılmıştır. SVM algoritması Bitcoin, Ripple ve Litecoin pazarlarını tahmin etmede başarılı olurken, Ethereum tahmininde başarılı olamamıştır. RF algoritması da Bitcoin ve Litecoin tahmininde başarılı olmuştur [8]. ANN'in kullanıldığı bir diğer çalışma da Sakız vd. [5] tarafından yapılmıştır. Çalışma sonucunda tahminlemede büyük farklılık ortaya çıkmıştır. Veri yetersizliği, algoritmaların kullanımı Bitcoin gibi bir para biriminde çok doğru sonuçları vermemiştir.

Twitter üzerinden Bitcoin ile ilgili atılan tweet duygularının analiz edilerek yapılan Bitcoin fiyat tahmini çalışmalarına bakılacak olursa, ilk olarak Pant vd. tarafından yapılan çalışma örnek olarak verilebilir. Pant vd. [6], duygu analizinin ardından zaman serilerini kullanarak Bitcoin fiyatları tahmin etmeye çalışmıştır. Duygu analizinin, zaman serileri tahminlemesinden daha başarılı olduğu görülmüştür. Wimala Gunaratne vd. [9] ise kripto para fiyat tahmini için ANN, SVM, RF ve NB

algoritmalarını kullanmıştır. Twitter’da yapılan çalışmalarda verilerin doğruluğu çalışmanın başarı oranı için büyük öneme sahiptir. Colianni vd. [10] tarafından yapılan çalışmada, Twitter verilerinin kripto para birimleri ile ilgili olup olmadığını ispatlamak ve kripto para piyasası hareketliliğini tahminlemek adına çeşitli makine öğrenmesi algoritmaları kullanılmıştır. LR, Bernoulli Naïve Bayes ve SVM algoritmaları kullanılmış, Bernoulli Naïve Bayes en iyi sonucu vermiştir. Çalışmanın doğruluk oranını artırmak için, tekrarlanan tweet sayısını veri setinden kaldırmışlardır ancak bu da veri setini kabaca %50 oranında azaltmıştır. Bu sebeple çalışmada doğruluk oranı %59’larda kalmıştır. Türkiye borsalarında Bitcoin değerinin tahmini için de Aslan vd. [2] Eğri Uydurma ve Adaptif Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS) tekniklerini kullanmıştır.

Fikir madenciliği (opinion mining) ve duygu analizi dünyada hızla büyümekte ve çok farklı sektörlerde uygulanmaktadır [12], [15]. Twitter’da ifade edilen duygular tespit ve analiz edilip bu verilerin sonrasında kullanıcıların twitter platformu üzerindeki aktivitelerine göre genel veya kişiselleştirilmiş öneriler geliştirilmesi için kullanılabilir [23].

Rahman vd. Twitter’den toplanan gerçek verilerin duygu analizini gerçekleştirebilecek bir model sunmak için denetimli ve denetimsiz makine öğrenmesi algoritmalarının melez bir kullanımını önermiştir [14]. Agarwal ve Mittal [13] çalışmalarında, yeni bir özellik çıkarma yöntemi önermiştir. Duygu analizinde sözlük tabanlı yaklaşımlarda ise genelde İngilizce dili üzerinden çalışmalar yapılmakta, diğer dillerde ise çeviri yolu ile kullanım sağlanmaktadır. Loureiro ve Allo de 1.7 milyon tweet ile yaptıkları analizde EmoLex sözlüğünü kullanmışlardır. Sözlükte sekiz temel duygu (öfke, korku, beklenti, güven, sürpriz, üzüntü, neşe ve iğrenme) ve iki duygu (olumsuz ve olumlu) bulunmaktadır [22].

Bu çalışmanın odak noktası Türkçe tweetler üzerinden duygu analizi olduğundan, literatürdeki Türkçe duygu analizi çalışmaları da özellikle incelenmiştir. Akın ve Şimşek [19] bireylerin duygularını ve düşüncelerini ifade ettikleri metinlerin oluşturduğu veri setinin analizi sonucunda elde edilen çıktıları yorumlayabilecek bir model önermiştir. Çalışma, programlama alanında sınırlı çalışma yetkinliği olan fakat elde edilen çıktılarla değer yaratabilecek birçok araştırmacının duygu analizinden faydalanmasını mümkün kılan bir model sunmuştur. Ayata vd. [11] tarafından yapılan çalışmada, SVM ve RF algoritmaları Türkçe tweetlerin içeriklerine göre olumlu ve olumsuz olarak sınıflandırılmasını amaçlanmıştır. Türkçe tweetler içeren duygu problemlerinin, Türkçe ifadelerin kısa olması ve anlamsal açıdan farklı yorumlar içermesi nedeni ile zorlu olduklarını belirtmişlerdir. Metinlerin içeriklerine göre sınıflandırılmasına ek olarak bir başka yöntem ise emoji kullanılarak yapılmaktadır. Yurtöz [24], gerçekleştirdiği çalışmada tam olarak bu konunun üzerinde durmuştur. Türkçe metinler içerisindeki duyguları tanımlayabilme amacıyla emoji’ler kullanmıştır. Yaptığı duygu analizinde ise SVM, kategorik Naïve Bayes (NB), FastText, Evrişimli Sinir Ağı (ESA) sınıflandırıcıları ile, eğitim ve test kümeleri Twitter mesajlarından elde edilen veri kümesi üzerinde gerçekleştirmiştir. Çetin ve Eryiğit [21], bir metnin içinde yer alan farklı duyguların ilgili oldukları hedef varlıklar ile birlikte tespit edilmesini hedef tabanlı duygu analizi ile gerçekleştirmişlerdir. Bu çalışmada tanımlamaları üç temel alanla (hedef terim, hedef kategori ve duygu sınıfı) temsil etmişlerdir. Hedef kategori, hedef terim ve ikisinin aynı anda belirlenmesi görevleri için, kelime vektörleri ve doğal dil işleme çıktıları kullanan koşullu rastgele alanlara (CRF – conditional random fields) dayalı bir dizilim etiketleme algoritması tasarlamışlar ve her üç görevi de tek aşamada çözebilmişlerdir.

Mete vd. [20] tarafından yapılan çalışmada, duygu analizi ile yapılacak sınıflandırma için göz bebeği hareketleri ve boyutu girdi olarak kabul edilmiş; katılımcılara IAPS’ten seçilmiş 60 fotoğraf gösterilmiştir. Göz takip cihazı ile katılımcıların deney boyunca göz fiksasyon sayım ve sürelerini toplamış ve veri tabanına kaydedilmiştir. Sınıflandırma için k-NN, NB, SVM, Doğrusal Diskriminant Analizi, Karar Ağacı ve LR teknikleri uygulanmıştır.

Neethu ve Rajasree tarafından yapılan çalışmada, sosyal medyanın, duygu açısından zengin veri üretiyor olması sebebi ile önemi vurgulanmıştır. Atılan tweet’ler, durum güncellemeleri, blog gönderileri vb. duygu analizi açısından çok kullanışlı ve yararlıdır. Yapılan çalışmada duygu analizi için Matlab programı üzerinden SVM, Naive Bayes, Maximum Entropy algoritmaları kullanılmıştır.

Naive Bayes algoritması ile elde edilen doğruluk oranı %89.5 ile yüksek bir başarı göstermiştir [25]. Çağlar ve Yavuz tarafından yapılan çalışmada ise yapay zeka yöntemleri ile güncel finans haberlerinin, Bitcoin üzerindeki etkileri ortaya konmak istenmiştir. Bunun için 5 ulusal finansal dergisi baz alınarak tahminleme yapılmış, çalışma sonucunda %99'luk tahminleme başarısı gösterilmiştir. Ancak çalışmada Bitcoin içerikli gazete haberlerinin Bitcoin fiyat tahminlemesine etkisinin oldukça zayıf olduğu belirlenmiştir [26].

Haddi vd., yaptıkları çalışmada film yorumları içeren bir veri seti üzerinde ön işleme sürecinde çalışırken, gürültü olarak nitelendirilebilecek verileri, gereksiz gördüğü özellikleri çıkarmış ve bu işlemin çalışma üzerinde negatif bir etkisi olmadığını belirtmiştir. Duygu analizinde makine öğrenmesi tekniklerinin en iyi şekilde uygulanabilmesi için metin ön işlemenin önemini vurgulamış, iyi gerçekleştirilen bir ön işleme sürecinin kullanılan makine öğrenmesi yönteminin başarısına çok önemli katkılar sağlayabileceğini eklemiştir [27]. Renault, finans üzerine bir mikro blog olan StockTwits üzerinde gerçekleştirdiği duygu analizi çalışmasında 250,000 mesaj üzerinde çalışmış ve Naive Bayes, Maksimum Entropi Sınıflandırıcısı, Doğrusal Destek Vektör Sınıflandırıcısı, Rastgele Orman Sınıflandırıcısı ve Çok Katmanlı Perceptron sınıflandırıcısından faydalanmıştır. Renault, duygu analizinin, günlük frekansta büyük getirileri olacağına dair deneysel bir kanıtı rastlamadığını belirtmiştir [28]. Roy ve Ojha, Twitter'ın insanların anlık duygu ve düşüncelerini paylaştığı büyük bir altın madeni olduğunu ifade etmişler ve bundan yola çıkarak Twitter üzerinden duygu analizi çalışması gerçekleştirmişlerdir. Duygu analizi için üç derin öğrenme modeli oluşturulmuş ve karşılaştırma yapılmıştır. Google BERT, LSTM VE CNN algoritmaları kullanılmış ve BERT modelinin diğerlerinden daha iyi performans gösterdiği belirlenmiştir. Çalışmanın dilinin İngilizce olması doğruluk oranı için artırıcı bir etmen olmuştur [30].

Literatürde Türkçe tweetlerde duygu analizi uygulaması henüz yabancı dillerdeki kadar yaygın değildir. Bunun yanında kripto para kavramı da görece yeni bir çalışma alanı olup gelecek için büyük bir potansiyel taşıdığı düşünülmektedir. Bu iki kavramı bir araya getirerek yani duygu analizi tekniği gerçekleştirerek finanstaki değişimleri incelemek oldukça yeni bir çalışma alanıdır ve gelişmeye açıktır. Bu çalışma, bu alandaki eksiklik hedef alınarak gerçekleştirilmektedir.

Literatür incelendiğinde, duygu analizi yapılırken en çok sözlük tabanlı ve makine öğrenme temelli sınıflandırma yaklaşımlarının kullanıldığı görülmüştür. Sözlük tabanlı yaklaşımların kullanıldığı çalışmaların büyük bir çoğunluğu İngilizce dili üzerinden yapılmıştır. Bunun nedeni İngilizce diline ait daha önceden hazırlanmış geniş bir duygu sözlüğünün olmasıdır. Diğer dünya dillerinde ise İngilizce'den otomatik çeviri yapılarak otomatik duygu sözlükleri oluşturulmuş ve bu sözlük üzerinden duygu analizi çalışmaları yapılmıştır. Ancak otomatik çeviri yönteminde İngilizce ve çeviri yapılan dilin yapısı bire bir örtüşmediği bu yöntem en iyi sonuçları vermemektedir. Bu noktada diğer diller de için duygu sözlüklerinin oluşturulması gerekliliği ortaya çıkar. Diğer yöntem olan Makine öğrenme temelli sınıflandırma yaklaşımında ise en çok SVM, Bayes, karar ağacı ve ANN yöntemlerinin kullanıldığı tespit edilmiştir.

Tüm bu gelişmeler ışığında, Twitter verileri vasıtasıyla Bitcoin üzerine duygu analizi yapmak finans sektörü ve kripto para piyasası açısından yadsınamayacak bir kaynak oluşturacaktır. Fakat durum böyle olmasına rağmen özellikle Türkçe duygu analizi kullanılarak bu konuda yapılan çalışma sayısı sınırlıdır.

### **III. DUYGU ANALİZİ**

Çevrimiçi sosyal medyanın ve mobil iletişim teknolojilerinin ortaya çıkışı, kullanıcı tarafından oluşturulan çeşitli biçimdeki içeriğinin akışında hızlı bir artışa neden olmuştur. İnsanlar genellikle tepkilerini, beğenilerini, hayallerini ve şikayetlerini sosyal medyada uzun metin yazmaktan ziyade, metinsel parçalarla ifade eder. Örneğin, her gün birbirinden farklı birçok konuda milyonlarca tweet üretilmektedir. Ayrıca, radyo televizyon yayıncılığı yapan medyadan farklı olarak çevrimiçi sosyal

medyada üretilen içerik anlık olarak yazılır ve düzenlenmemiş haldedir. Bu nedenle, farklı alanlardaki araştırmacı, bu kalabalık bilgiyi analiz etmeye başlamıştır. Bu analiz süreci ise pratik metotları doğurmuştur. Bunlardan en çok yaygın olanlarından biri duygu analizidir.

Duygu analizini gerçekleştirebilecek birçok sosyal medya mümkündür. Bu sosyal medyalardan (Twitter, Facebook, Instagram vb.) hangisi kullanılarak bu çalışmanın gerçekleştirileceği büyük öneme sahiptir. Kullanılacak olan sosyal medyada uygulanacak duygu analizinde belirlenen finans alt başlığının sektöre hitap etmesi de bir o kadar önemlidir.

## **A. DUYGU ANALİZİNİN ÖNEMİ**

Karar verme süreçlerinde insanların ne düşündüğü her zaman ön planda olmuştur. Günümüzde, internetin son trendlerinin ortaya çıkması nedeniyle insanlar görüşlerini ve duygularını forumlarda, bloglarda, ağ ve içerik paylaşım servislerinde paylaşmaktadır.

Duygu analizi doğal dil işleme, istatistik ve programlama gibi farklı yöntemlerin kullanımı aracılığıyla internet ortamındaki kullanıcıların duygularını ve görüşlerini analiz etmekte kullanılan güncel bir araştırma alanıdır. Duygu analizinde temel motivasyon insanların durumlar karşısındaki düşüncelerini ortaya çıkarmaktır [29]. Görüş ve duygular karar verme sürecini belirleyen önemli etkenlerdir. Duygu analizi yöntemi, teknolojiye veya yeniliğe olumlu veya olumsuz referanslar aramaktan oluşur. Günümüzde bu başlıklar oldukça ön plandadır ve bu çalışmanın ana hedeflerini içermektedir.

## **B. TWİTTER'İN DUYGU ANALİZİNDEKİ ÖNEMİ**

Günümüz dünyasında rekabet halindeki firmalar, geleneksel pazarlama yöntemleri ile müşteriye ulaşmanın yeterli olmadığını görmüşler ve sahip olduğu potansiyel nedeninin etkisi ile gündemlerine sosyal medyayı almışlardır. Bu açıdan sosyal medya çok önemli bir veri kaynağıdır ve buradan elde edilen bilgi birçok sektöre yön vermektedir. Sosyal medyada yayılan fikirler ürün ve hizmetlerin ne kadar başarılı olduğunu yansıtmaktadır. Bu durum doğrudan satışları ve ekonomiyi etkilemektedir [16].

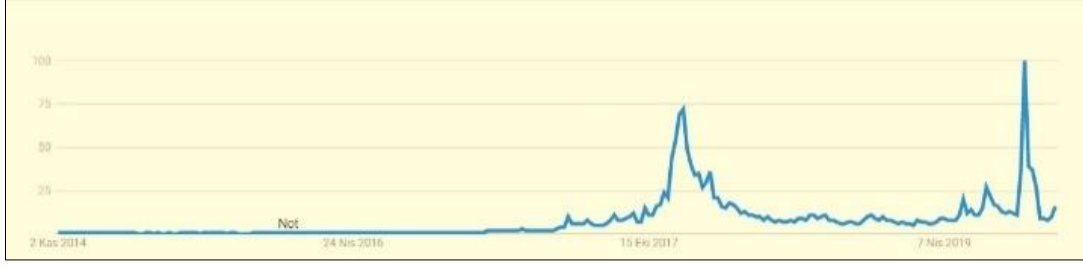
Twitter kullanıcıların, tweet olarak adlandırılan herhangi bir konu ile ilgili düşüncelerini paylaşabildikleri popüler sosyal ağlardan birisidir. Paylaşılan mesajlar, kullanıcıların farklı konular ile ilgili fikir ve duygularını içermektedir. Bu nedenle, bu çalışmada bir sosyal medya ortamı olan Twitter üzerinden elde edilen mesajlar duygu analizi sınıflandırılması probleminde kullanılmıştır [17].

Twitter tarafından yapılan bu sınırlandırmanın amacının, çok az kelime ile çok fazla şeyin paylaşılacağını göstermek ve kişilerin düşünmeden, ilk akıllarına gelen cümleler ile paylaşımında bulunmalarını sağlamak olduğu açıklanmıştır.

## **C. TWİTTER KULLANILARAK YAPILACAK OLAN ÇALIŞMADA FİNANSAL VARLIK TESPİTİ**

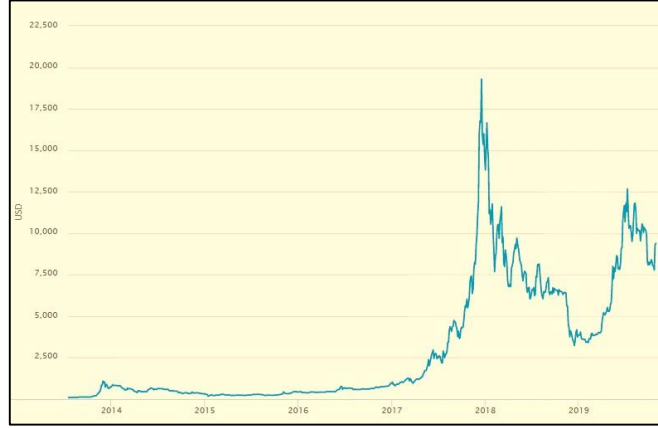
Günümüzde finans sektörü hayatın işleyişinde büyük öneme sahiptir. Teknolojinin zıplama çağı olan 90'lardan beri dijital bir para birimi oluşturmak için birçok girişimde bulunuldu. İlk kripto para olan Bitcoin dağıtık yapısı sayesinde çok kısa bir sürede günümüz finansal düzenine karşı yükselişe geçti. Düzenlemeye ya da kontrole tabi tutulmadığı için Bitcoin'in değeri sıfır noktasından, binlerce dolara kadar yükselmiştir.

Bu çalışmada finansal varlıklardan olan Bitcoin'in seçilmesinin en önemli nedeni son zamanlarda meydana gelen araştırılma sayısıdır. 2014-2016 yılları arasında bu değer oldukça aşağılardayken, son 2 yılda değişiklik göstermiştir (Şekil 1). Bu çalışma meydana gelen bu değişkenliği tahmin edebilmeyi mümkün kılacaktır. Bu sayede finansal konuda yapılacak eylemler, insanların bakış açıları tahmin edilebilecektir.



*Şekil 1. Son 5 yıla ait Bitcoin kavramını araştırma sayısı [31]*

Yapılan araştırmalar sonucu ile doğru orantılı olarak Bitcoin borsalarının da değiştiği görülmektedir (Şekil 2).



*Şekil 2. Son 5 yıla ait Bitcoin borsasındaki değişim [32]*

Dolayısıyla, bu çalışma ile Bitcoin'deki değişimlerin saptanabilmesi ve bireylerin veya kurumların bu değerler karşılığında eylemlere geçebilecek olması finansal açıdan büyük bir öneme sahip olacaktır.

#### **D. UYGUN YÖNTEMLERİN DEĞERLENDİRİLMESİ VE KULLANILACAK YÖNTEMLERİN TARİFİ**

Literatür taramasında bahsedilen çalışmalarda gözlemlendiği üzere en çok kullanılan makine öğrenmesi algoritmaları Destek Vektör Makinesi, Naïve Bayes, Rastgele Orman, Doğrusal Regresyon, Lojistik Regresyon ve Yapay Sinir Ağları'dır. Yapılan çalışmalarda farklı sonuçlar veren bu algoritmalar aynı zamanda en çok verim alınan algoritmalarlardır. Bu çalışmada Rastgele Orman ve Doğrusal Regresyon algoritmaları üzerinden duygu analizi için sınıflandırmalar yapılmıştır. Bu algoritmaların seçilmesinin en büyük nedeni kullanılan programlama diline kolay entegrasyonu ve bu çalışma için daha iyi verimlilik skorlarına sahip olmasıdır.

Makine öğrenmesi temelli duygu analizinde; metin ön işleme süreci, gerçek dünyadan alınan nesnelere doğrudan kullanılmadığı için yapılması gerekmektedir. Tokenlara ayırma, gereksiz sık kullanılan veya durak kelimelerin (stop-words) atılması ve kelime köklerini bulma (stemming) en yaygın kullanılan ön işleme yöntemlerindedir. Ayıklama işleminin ardından BoW (Bag of Words) ve N-Gram model kullanılarak öznelik çıkarılmaktadır. Bu işlemleri yapmamamızın sebebi hesapsal yük oranı düşürmek ve tahmin oranını (Accuracy değeri gibi) arttırmaktır.

Bayes teoremi esasına dayanan Naïve Bayes, kolaylıkla uygulanabilen en kullanışlı makine öğrenme algoritmalarındandır. Metin sınıflandırmada kullanılan bu algoritma, olasılık kurallarına göre belirlenmiş bir dizi hesaplama ile, sisteme sunulan verilerin sınıfını yani kategorisini tespit etmeyi amaçlar. Algoritmanın çalışma şekli bir eleman için her durumun olasılığını hesaplar ve olasılık değeri

en yüksek olana göre sınıflandırma yapar. Sınıflandırıcı, özniteliklerin birbirinden bağımsız olduğunu varsayar. Bu, bir özneliğinin var olmasının veya olmamasının, başka bir özneliğinin var olması, olmaması veya değerini etkilemeyeceği anlamına gelir. Bayes teoremi, koşullu olasılık üzerine kurulu bir teoremdir.

$$P(C_i|X) = \frac{P(X|C_i)*P(C_i)}{P(X)} \quad (1)$$

Bayes teoremi göz önüne alınarak, NB algoritması aşağıda verilmiştir:

1.  $D$ 'nin veri setini temsil ettiği ve  $D$ 'deki her  $X$ 'in sınıf etiketinin belli olduğu var sayılsın.  $X$ ,  $n$  tane öznelikten oluşan bir vektördür ve  $X = (x_1, x_2, \dots, x_n)$  olarak temsil edilmektedir.

2.  $C_1, C_2, \dots, C_m$  ile temsil edilen  $m$  tane sınıf olduğu var sayılsın. Naive Bayes sınıflandırıcısı bir  $X$  vektörünün  $C_i$  sınıfına ait olup olmadığını bulmak için, bütün sınıflar içinde en yüksek  $P(C_i|X)$  ardıl olasılığına sahip değeri bulmaya çalışır. Bu durum Bayes teoremi ile Eşitlik 1'de ifade edilmiştir.

3.  $P(X)$  değeri tüm sınıflar için aynı olduğundan, yalnızca  $P(X|C_i) * P(C_i)$  ifadesi maksimum yapılmalıdır.

4.  $P(C_i)$  ifadesi,  $C_i$  sınıfındaki eleman sayısının, tüm eleman sayısına oranıdır.

5.  $P(X|C_i)$  ifadesi ise,  $X$ 'in  $n$  tane değer içeren bir öznelik vektörü olduğu var sayıldığında Eşitlik 2 ile hesaplanır.

$$P(X|C_i) = \prod_{k=1}^n P(X_k | C_i) \quad (2)$$

6. Sonuçta, sınıflandırıcı en büyük  $P(X|C_i) P(C_i)$  ifadesine sahip olan  $C_i$  sınıfını,  $X$  vektörünün sınıfı olarak seçer [18].

Rastgele Orman algoritmasının çalışma prensibinde; algoritmayı başlatmak için kullanıcı tarafından iki parametre tanımlanır. Bu parametreler, en iyi bölünmeyi belirlemek için her bir bölümde kullanılan değişkenlerin sayısı ( $m$ ) ve geliştirilecek ağaçların sayısı  $N$ 'dir. Öncelikle eğitim veri kümesinin 2/3'ünden önyükleme örnekleri oluşturulur. Eğitim veri kümesinin 1/3'lük geri kalan kısmı hataları test etmek için yararlanılır. Daha sonra her önyüklemeli örnekten budama yapılmadan ağaç geliştirilir. Her bölümde  $m$  tane değişkeni tüm değişkenler içinden rassal olarak alınır ve bu değişkenler içinden en uygun dal belirlenir. Yeterli öngörü gücü ile yeterli seviyede düşük korelasyon sağlayan değişken miktarının seçilmesi çok önemlidir. Breiman'e göre toplam  $m$  tane değişkenin kare köküne eşit alınan  $m$  değişken miktarı genellikle optimum sonuca en yakın değeri verir. Rastgele Orman ağaç oluşturmak için CART (Classification and Regression Tree) algoritmasını kullanılır. Her bir bölümde dallar CART algoritmasının kriterine (örn. GINI indeksi) göre oluşturulur. GINI indeksi grup homojenliğini ölçer ve Eşitlik 3 ile ifade edilebilir.

$$\sum \sum_{j \neq i} \left( \frac{f(C_i, T)}{|T|} \right) \left( \frac{f(C_j, T)}{|T|} \right) \quad (3)$$

Burada  $T$  eğitim veri kümesi,  $C_i$  rassal belirlenen bir verinin ait olduğu sınıfı  $C_i$  ve  $\frac{f(C_i, T)}{|T|}$ , belirlenen örneğin  $C_i$  sınıfına ait olma olasılığını göstermektedir. GINI indeksi yükseldikçe grup heterojenliği artarken, GINI indeksi azaldıkça grup homojenliği artar [29]. Bu çalışmada rastgele orman algoritmasının seçilmesinin ana nedeni, yüksek boyutlu veri ile yüksek performans göstermesidir. Bu çalışmada bulunan veriler, Twitter platformu üzerinde insanların Bitcoin hakkında yazdığı yorumlardır. Her ne kadar Twitter'da karakter sınırı bulunsun da çok sayıda tweet elde edilince veri çok yüksek boyutlara ulaşabilmektedir. Bu nedenle bu duruma uygun bir algoritma olarak rastgele orman



algoritması tercih edilmiştir. Ek olarak, rastgele orman literatürde pek çok duygu analizi ile finans tahminlemesi gerçekleştirilen çalışmada tercih edilip olumlu sonuçlar vermiştir.

Lojistik regresyon (LR), bağımlı ve bağımsız değişkenler arasındaki ilişkiyi anlamamıza yardımcı olan istatistik temelli bir algoritmadır. LR, ikili 1 veya 0 olarak kodlanmış verileri içerir. Tahmin etmek istediğimiz değişken tipleri eğer kategorik (nitelik olarak belirli) ise bu durumda LR kullanmak faydalı olacaktır. LR algoritması ile veri setinde kullanılan örneklemeler üzerinden sınıflar belirlenir. Bir diğer özelliği ise lineer regresyonun aksine verileri doğrusal olarak değil, logaritmik olarak eğri üzerinde yakalamasıdır.

Doğrusal regresyon, değişkenler arası ilişkiyi açıklayabilmek ve en uygun düz çizgi ile hiper düzlemi bulabilmek için oluşturulan lineer bir model yaklaşımıdır. Basit doğrusal regresyonda 2 değişken bulunmaktadır. Bunlar, bağımlı ve bağımsız değişkenlerdir. En uygun düz çizgi ya da diğer adıyla regresyon çizgisi kullanılarak bağımlı değişken (Y) ile bağımsız değişken (X) arasında bir ilişki kurar. Doğrusal regresyondaki en önemli nokta, bağımsız değişkenden farklı olarak bağımlı değişkenin sürekli olmasıdır. Bu çalışmada doğrusal regresyonun tercih edilme sebebi ise uygulanabilirliği ve çalışmadaki bağımlı değişken ile bağımsız değişkenlerin doğrusal bir ilişkiye sahip olmasıdır. Bunun yanında, doğrusal regresyon yöntemi, geçmişte duygu analizi ile finansal tahmin yapılan çalışmalarda da tercih edilip başarılı sonuçlar vermiştir.

## **IV. TWITTER'DA BITCOİN ÜZERİNE DUYGU ANALİZİ**

### **A. DUYGU ANALİZİNDE KULLANILAN ARAŞTIRMA SEVİYELERİ**

Duygu analizinde araştırma seviyesi olarak üç farklı seviye bulunmaktadır.

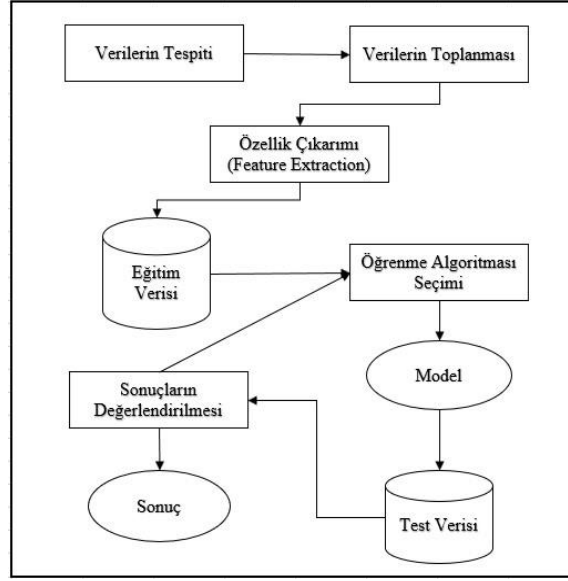
**Doküman Seviyesi:** Bu yöntemle bir dokümana göre tüm düşünceyi negatif ya da pozitif sonuç çıkaran yöntemdir. Bu analiz biçimi birden çok ürünü ya da durumu karşılaştıran dokümanlar için uygun değildir. Çünkü birden fazla durum ya da ürün karşılaştırması yapıldığında birden fazla sonuç çıkması beklenmektedir.

**Cümle Seviyesi:** Bu yöntemle cümle bazında pozitif, negatif ya da nötr sonucunu çıkarma işlemi yapılmaktadır. Nötr genelde duygu veya fikir belirtmeyen cümle olduğu anlamına gelmektedir. Birçok objektif cümle düşünce içerebilir fakat subjektif düşünce içermeyebilir bu durumu küçük bir örnekle açıklamak gerekmektedir. Örneğin: “Geçen ay araba aldık ve arabanın cam sileceği düştü.” cümlesi aslında olumsuzdur fakat cümle içinde olumsuzluk ifade eden bir kelime bulunmamaktadır. Bu nedenle de objektif bir çıkarım olarak görüntülenmektedir ve bir düşünce içermediği düşünülebilir ancak bu durum tam olarak böyle değildir.

**Varlık ve Görüş Seviyesi:** Varlık ve Görüş Seviyesi diğerlerine göre küçük taneli analiz yapmayı sağladığından daha doğru analizler yapmayı sağlamaktadır. Bu analiz yöntemiyle dil yapılarına bakmak yerine direkt duygu ile ilgilenir. Örneğin: “Servisi mükemmel olmamasına rağmen bu restoranı hala seviyorum.” cümlesinin pozitif bir tonu olsa da tam anlamıyla pozitif olduğu söylenemez. Burada restoranla ilgili pozitif düşünce olsa da negatif düşünce içermektedir.

### **B. DUYGU ANALİZİNİN AŞAMALARI**

Duygu analizinde belirli bir akış söz konusudur. Bu çalışmada da uygulanan akışın temel basamakları Şekil 3'te gösterilmiştir.



*Şekil 3. Duygu analizi metodolojisi*

Bu çalışmada Twitter kullanarak kripto para birimlerinden biri olan Bitcoin üzerine duygu analizi için yeni bir model önerilmiştir. Duygu analizinde, genellikle kullanılan teknikler makine öğrenmesi ve anlamsal yönelimdir. Makine öğrenmesi genellikle metin sınıflandırması problemlerinde kullanılır ve bu çalışmada da çeşitli makine öğrenmesi algoritmaları kullanılmıştır. Daha iyi tahmin performansı elde etmek için geleneksel sözlük yaklaşımı NB ve LR algoritması ile birleştirilmiştir. Bu algoritmaları Twitter duygusu ile Bitcoin'in finansal değerinin yükselişi ve düşüşü arasındaki etkileşimi araştırmak için kullanılmıştır. Bu sayede, bu ilişkide istatistiksel olarak anlamlı bilgi olup olmadığı analiz edilmiştir. Bu çalışmanın ana katkısı, geleneksel finans metotlarıyla karşılaştırıldığında bile, Twitter'ın finans sektöründe önemli bir veri kaynağı olarak kullanılmasını destekleyen ampirik bir kanıttır.

Ne var ki elde edilen verilerin (yorumların) işlenmemiş olması, büyüklüğü, hatalı yazım, kısaltma ve günlük konuşma dili gibi nedenlerle sosyal ağlar üzerinde çalışmak oldukça zordur. Binlerce hatta on binlerce verinin ham haliyle tek tek incelenmesi ve insan algısıyla anlaşılması güçtür [16]. Bu nedenle verilerin çeşitli programlar vasıtasıyla API sağlayıcılarından çekilmesi, doğal dil işleme yöntemleri ile işlenmesi ve analizinin doğru bir biçimde yapılması gerekir.

### C. VERİLERİN ÇEKİLMESİ

Twitter'da duygu analizi gerçekleştirmek için Python'da kullanılan kodun ilk kısmında <https://developer.twitter.com/> adresi üzerinden alınan kod ve tokenler devreye girmektedir. Kodun devamında anahtar kelimenin aranması ve anahtar kelimeyi içeren tweetlerin bir text dosyasına kaydedilmesi komutları bulunmaktadır. Kod çalıştırıldığında aranan anahtar kelimeleri twitter'da anlık olarak tarar ve bunları simultane olarak bir not defterine kaydeder. Tweet'in atıldığı tarih ve saat, tweet'i atan kişi, tweet'in retweet olup olmadığı, hangi cihaz ile atıldığı ve daha birçok bilgiye anlık olarak ulaşılır ve kaydedilir.

Kod yazılırken Twitter API'den yararlanılmıştır. API, Application Programming Interface yani Uygulama Programlama Arayüzü anlamına gelmektedir. Twitter API sayesinde Twitter'ın ana sayfasındaki yayın akışı farklı uygulamalar aracılığıyla takip edilebilmekte ve analizler yapılabilmektedir. Özellikle veri bilimi alanı için güncel ve popüler bir uygulama alanıdır. Yabancı literatürde çok sayıda uygulaması mevcuttur.

## D. VERİLERİN ÖNİŞLEME SÜRECİ

Öğrenme algoritmalarında kullanılacak olan eğitim ve test verilerini oluşturmak için, Tweepy kullanılmıştır. Tweepy, Twitter API'ye erişmek için kullanılan açık kaynaklı bir Python kütüphanesidir. Tweetler Python programlama dili aracılığıyla Twitter'da "bitcoin" anahtar kelimesi gerçek zamanlı olarak aranarak anlık olarak çekilmiştir. Çekilen tweetler anlık olarak bir not defteri dosyasına kaydedilmiştir. Bu not defterinde elde edilen tweetlerin tweetlerin hangi ülkeden atıldığı, ne zaman atıldığı, hangi cihazdan atıldığı gibi pek çok bilgi bulunmaktadır.

Python'da tweetler için bir dil filtresi oluşturulmuştur. Bu sayede "Bitcoin" anahtar kelimesini içeren herhangi bir dildeki tüm tweetleri çekmek yerine sadece Türkçe tweetler çekilebilmiştir. Sonrasında tweetler incelenip kötü sözcük içeren tweetler elenmiştir. Fazla boşluklar çıkarılmış, büyük küçük harfler düzenlenerek tweetler cümle formatına çevrilmiştir. Alfabetik olmayan tüm karakterler, yüz ifadeleri(smiley) ve emojiler tweetlerden çıkarılmıştır. Çalışmada çok sık tekrarlanan kısaltmalar olduğu gibi bırakılmış, tekrarlanmayan kısaltmalarda ise kelimenin açık hali yazılmıştır. Bunun amacı, aynı anlam kastedilmesine rağmen, tweetlerde birden çok defa açık hali verilmiş bir kelimenin yalnızca bir defa geçen bir kısaltmasına algoritmanın anlam veremeyecek olmasıdır. Toplanan tweetlerdeki hashtagler yalnızca hashtag sembolleri(#) çıkartılarak çalışmaya dahil edilmiştir. Bunun sebebi kelime havuzunun mümkün olduğu kadarıyla çok yüksek boyutlara ulaşmasına engel olmaktır. Tekrarlayan tweetlerin yalnızca bir tanesi kaydedilmiştir. Bunun amacı Twitter botlarının çalışma üzerindeki etkisi olmasını engellemektir. Ardından tweetler tweet ID (Atılan "tweet" in kimlik numarası), tweet'i atan kullanıcı adı (user name), tweet ve tweetin atıldığı tarihler ile birlikte bir Excel dosyasına aktarılmıştır. Önileme süreci boyunca tüm dillerde 20.000'in üzerinde tweet elde edilmiştir. Dil filtresi geliştirildikten sonra ise 4062 adet Türkçe tweet çekilmiştir. Önileme sürecinden sonra 3737 tweet ile model oluşturulmuştur. Tweetlerin önilemesi yaklaşık olarak 60 saat sürmüştür.

## E. VERİ SETİ

Bu çalışmada, Twitter'dan çekilen verilerin yanısıra Bitcoin'in borsadaki değerlerini içeren veriler de kullanılmıştır. Bu hedef ile farklı zamanlarda 3737 "Bitcoin" anahtar kelimesini içeren Türkçe tweet, Python'ın Tweepy kütüphanesi kullanılarak çekilmiştir. Veri seti pozitif, negatif ve nötr duyguları ayırt etmeye yoğunlaşmıştır. Bu çalışma bu duyguları ayırt etmeyi kelime bazlı duygu analizi yaparak gerçekleştirmeyi amaçlanmaktadır.

Bölünme sonucu eğitim seti eldeki tweetlerin %80'ini, test seti ise %20'sini oluşturmaktadır. Bu işlemin sonucunda eğitim setinde 2989 tweet bulunmakta olup bu tweetlerin %14,99'u negatif, %67,28'i nötr ve %17,73'ü pozitifdir. Test setinde 748 tweet bulunmaktadır. Bu tweetlerin %15,78'i negatif, %65,78'i nötr ve %18,45'i pozitifdir.

Bu işlemin ardından da tweetlerin duygularının üzerinde istatistiksel süreçlerin uygulanabilmesi için yazıların token matrislerine çevrimi sağlanmıştır. Bunu gerçekleştirebilmek için sklearn kütüphanesinden CountVectorizer fonksiyonu uygulanmıştır.

## F. VERİ GÖRSELLEŞTİRME

Bu çalışmada da çeşitli veri görselleştirme yöntemlerinden faydalanılmıştır. Bunun amacı çok sayıda tweet arasından ilk göze çarpabilecek verileri gözler önüne serilebilmektir. Bu doğrultuda kelime bulutu yöntemi uygulanan tekniklerden bir tanesidir. Bu yöntemin amacı belirli bir yazı dosyasında en çok tekrarlanan kelimeleri bir arada okuyucuya sunabilmektir. Pozitif ve negatif duygulu tweetler için elde edilen kelime bulutları sırasıyla Şekil 4 ve Şekil 5'te verilmiştir.



Şekil 4. Pozitif tweetler için uygulanan kelime bulutu çalışması



Şekil 5. Negatif tweetler için uygulanan kelime bulutu çalışması

## G. VERİLERİN SINIFLANDIRILMASI VE SONUÇLAR

Sınıflandırma, bir veri kümesini birbirinden farklı ve önceden belirlenmiş olan eğitim verisi kategorilerinden en uygun olanına atama işlemidir. Tweetlerin duygularına gibi sınıflandırılma işlemi, tweetlerdeki Bitcoin yorumlarına göre gerçekleştirilmiştir. Bu çalışmada amaç Bitcoin ile ilgili atılan tweetlerin pozitif, negatif ve nötr olmak üzere üç duygu ile sınıflandırılmasıdır. Buna göre Bitcoin hakkında olumlu bir yorum bulunan bir tweet pozitif olarak, olumsuz bir yorum bulunan bir tweet negatif olarak, herhangi bir duygu içermeyen tweetler ise nötr olarak sınıflandırılmıştır. Sınıflandırma işlemi makale yazarları ayrı ayrı gerçekleştirmiş, ardından fikir birliği baz alınarak tweetin duygusu belirlenmiştir. Sınıflandırma işlemi sonrasında ise bir finans uzmanından destek alınarak gerekli görülen düzenlemeler gerçekleştirilmiş, fikir birliği sağlanamayan tweetler için uygun bir sınıf bulunmuştur.

Tweetler eğitim ve test verisi olmak üzere ikiye ayrılmaktadır. Eğitim verisindeki tweetlerin pozitif, negatif ve nötr olacak şekilde sınıflandırılması yukarıda anlatıldığı gibi gerçekleşmiştir. Test kümesindeki verilerin sınıflandırma çalışmasında tweetlerde sıklıkla kullanılan ve bir duygu belirtmeyen edatlar, zamirler, bağlaçlar ve veriler toplanırken kullanılan sorgulama terimleri yasaklı kelimeler listesine dahil edilerek duyguların sınıflandırmasında yasaklı kelime listesi kullanılmamıştır. Test verilerine ait kelime sınıflandırma işlemi sonuçları Tablo 1’de verilmiştir.

Tablo 1. Test veri seti

Sınıf	Tweet Sayısı
Negatif	118
Nötr	492
Pozitif	138
Toplam	748

Çalışmada tweetler test ve eğitim setlerine ayrıştırıldıktan sonra sınıflandırma işlemini yapacak modeller için Naïve Bayes (NB) ve Lojistik Regresyon (LR) algoritmaları kullanılmıştır.

NB algoritması ile kurulan model, %72,19 başarı oranı ile gerçek tweetlerin duygularını tahmin etmeyi başarmıştır. LR algoritması ile kurulan modelde ise başarı oranı %75,53 olmuştur.

## V. DUYGU ANALİZİ İLE BITCOIN DEĞERİ TAHMİNLEMESİ

### A. FİNANSAL VERİ SETİ

Finans verileri ise Yahoo Finance web sitesinden elde edilen tweetlerin atıldığı tarih ve saatler baz alınarak çekilmiştir (Şekil 6), yani çekilen tweetlerin ve finans verilerinin tarih ve saatleri aynıdır. Bu durum elde edilen tweetlerin duyguları ile finans verilerini birbirleriyle karşılaştırabilmek amacıyla gerçekleştirilmiştir. Elde edilen veriler 2019 yılının Aralık, 2020 yılının Şubat ve Mart aylarına aittir.

	Tarih	Açılış	Yüksek	Düşük	Kapanış	Hacim
0	2019-10-13	8336.902344	8470.988281	8276.612305	8321.005859	13808286058
1	2019-10-14	8320.832031	8390.208984	8284.130859	8374.686523	15151387859
2	2019-10-15	8373.458008	8410.714844	8182.706543	8205.369141	15220412631
3	2019-10-16	8204.674805	8216.812500	7985.089844	8047.526855	16071646995
4	2019-10-17	8047.812500	8134.831543	8000.942871	8103.911133	14313052244

Şekil 6. Bitcoin verileri

Duygu analizi aşamasında günlük atılan tweetlerin duygusu belirlendikten sonra, literatürde sıkça geçen bir yöntem olan “Bitcoin” anahtar kelimesi içerip pozitif duygu içeren tweetlerin oranı belirlenmiştir. Bununla beraber Bitcoin günlük açılış ve değerleri kurulan modelde yer almıştır. 22 farklı gün için Şekil 7’de yer alan değerler modelde yer almıştır.

	Tarih	Pozitif Duygu(%)	Açılış Değeri	Kapanış Değeri
0	Ara 21	0.152	7220.593	7191.158
1	Ara 22	0.150	7191.188	7511.588
2	Ara 23	0.255	7508.902	7355.628
3	Ara 24	0.153	7354.393	7322.532
4	Ara 25	0.116	7325.755	7275.155

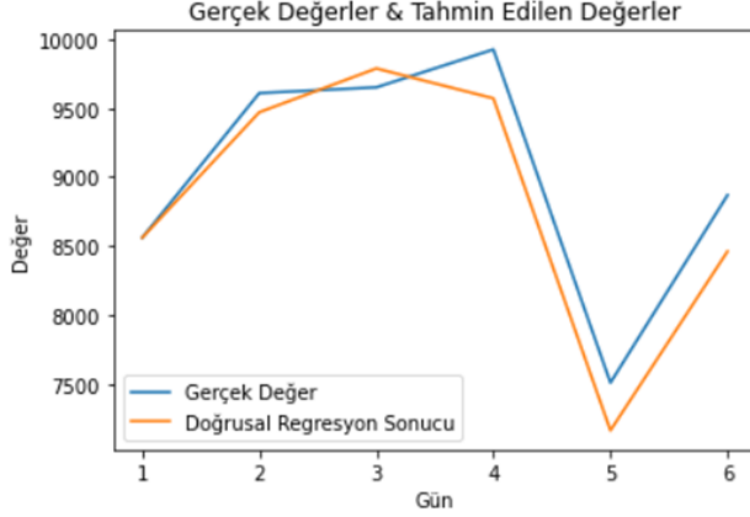
Şekil 7. Veri setindeki günlük pozitif tweet oranı, Bitcoin açılış ve kapanış değerlerinin bir kısmı

Eldeki veriler önce eğitim ve test veri seti olarak ikiye ayrılmış, ardından sklearn kütüphanesinden StandardScaler kullanılarak düzenlenmiştir.

### B. DOĞRUSAL REGRESYON UYGULAMASI

Bitcoin günlük açılış değeri ile günlük pozitif tweet oranı beraber kullanılarak Bitcoin günlük kapanış değeri tahminlemesinde bulunulmuştur.

Oluşturulan model test ve eğitim setlerine ayrıştırıldıktan sonra Bitcoin kapanış değeri tahmini uygulaması için kullanılan yöntemlerden birisi Doğrusal Regresyon(DR) olmuştur (Şekil 8). Sklearn kütüphanesinden linear\_model çağrılarak uygulanmıştır. Modelde oluşan eğriye denk gelmeyen her nokta artık (residual) olarak adlandırılmaktadır. Artıkların toplamının ortalamaların toplamına olan oranı ne kadar küçük ise  $r^2$  o kadar yüksek olacaktır. Bu nedenle  $r^2$ 'nin yüksek olması regresyon model uyumunun iyi olduğunu gösterir.

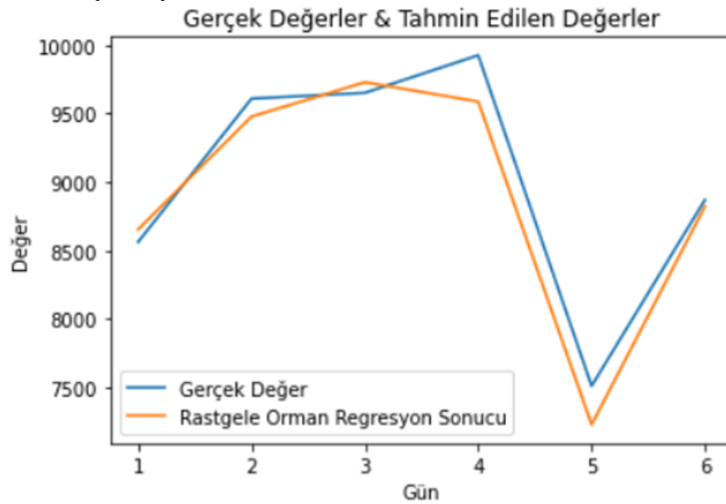


Şekil 8: Bitcoin günlük kapanış değeri ve doğrusal regresyon sonucu karşılaştırması

### C. RASTGELE ORMAN REGRESYON UYGULAMASI

Bitcoin günlük açılış değeri ile günlük pozitif tweet oranı beraber kullanılarak Bitcoin günlük kapanış değeri tahminlemesinde bulunulmuştur.

Oluşturulan model test ve eğitim setlerine ayrıştırıldıktan sonra Bitcoin kapanış değeri tahmini uygulaması için kullanılan yöntemlerden bir diğeri Rastgele Orman Regresyon(RFR) olmuştur (Şekil 9). Sklearn kütüphanesinden RandomForestRegressor çağrılarak uygulanmıştır. Buna göre elde edilen çalışma  $r^2$  değeri %94,16 çıkmıştır.



Şekil 9. Bitcoin günlük kapanış değeri ve rastgele orman regresyon sonucu karşılaştırması

## VI. SONUC

Duygu analizi aşamasında NB ve LR olmak üzere iki makine öğrenmesi metodu kullanılmıştır. NB uygulaması sonucunda tweetlerin duygularının tahmin etmedeki başarı oranı %72,19 olmuşken LR uygulaması sonucu bu oran %75,53 olmuştur.

Bitcoin tahminlemesi aşamasında “Bitcoin” anahtar kelimesi içeren günlük pozitif tweet oranı ile Bitcoin günlük açılış değeri beraber kullanılarak Bitcoin kapanış değeri tahminlemesi yapılmıştır. DR ve RFR yöntemleri kullanılmıştır. DR için  $r^2$  sonucu % 88,97 çıkmıştır. RFR için ise bu oran %94,16 olmuştur.

Yapılan çalışma finans sektörüne ait olduğundan, sektörde popüler bir kavram olan Bitcoin üzerinde durulmuştur. Teknolojinin gelişmesi başka düşüncelerin meydana gelmesine yol açmıştır. Bitcoin ise bu düşünceler arasında dikkat çeken bir kavram olarak yerini korumaktadır.

Uygulanan yöntemler sonucu atılan tweetlerin duygusu ile Bitcoin değeri arasında bir ilişki bulunduğunu söylemek kuşkusuz yerinde ve doğru olacaktır. Bitcoin’in değeri üzerinde birden çok faktörün etkili olduğu gözden kaçırılmaması gereken bir gerçek olsa da bu çalışma tweetlerin duygularının Bitcoin tahmininde önemli bir etken olabileceğini gözler önüne sunmaktadır. Bu çalışma yapılırken Dünya çapında birden çok olumsuz olayın gerçekleşmiş olmasının da piyasaları etkileyeceği düşünülürse daha huzurlu bir dönemde bu ilişki incelendiğinde başarı oranının artması oldukça gerçekçi ve mümkündür.

Gerçekleştirilmesi planlanan veya hali hazırda mevcut bulunan projelerin geliştirilmesi için bu çalışma kılavuz niteliği taşımaktadır. Seçilen metotların uygunluğunun tespit edilmesi, belirlenen yardımcı araçların tercih sebebinin neye bağlı olduğunun irdelenmiş olması çalışmayı daha açık kılmaktadır. Bu sayede yapılan çalışma bu alanla ilgili girişimde bulunacak herhangi bir çalışmacının işini oldukça kolaylaştıracaktır.

Çalışmanın konusu yalnızca Bitcoin ile sınırlı kalmamaktadır. Aynı şekilde başka yatırım araçları için de kullanılabilir. Yatırım sektörü dışında da birçok konu ile entegre edilebilir. Duygu analizi, finans sektörü dahil olmak üzere diğer bütün sektörler için kullanılabilir. Bu nedenle duygu analizi oldukça kapsamlı bir çalışma ortamına sahiptir. Yaşanılan dünyanın ana tüketimini insanlar oluşturmaktadır. İnsanların var olduğu bir dünyada ise onların düşüncelerini saptayabilmek değerli ve etkili çalışmaların yapıtaşını oluşturmaktadır. Bu nedenle bu çalışma yapılmış olan projelerin geliştirilmesinde ve gelecekte yapılacak olan projelerin etkili olmasında kullanılabilir. Böylece yapılan çalışmalardaki verimin artmasına ve pozitif bir kazanç sağlamasına olanak tanınmış olur.

Yapılan bu çalışma, kurumların sosyal medya analitiklerinden değer elde etmelerine yardımcı olarak entegre bir sosyal medya analitiği çerçevesi önermeyi ve böylece ticari büyümeyi ve sürdürülebilirliği arttırmayı amaçlamaktadır. Bu çok önemlidir, çünkü çoğu kurumun sosyal medya analitik hedeflerini yönlendirecek bütünlüklü çerçeveleri yoktur. Bu çalışma ile, kurumların yatırım veya ticaret yapış yapamayacaklarına karar vermelerine yardımcı olmak hedeflenmektedir. Bu şekilde bu çalışma yatırım danışmanlığı aracı olarak kullanılabilir. Böylece bunun bir girişimcilik olduğunu düşünülebilir. İkincisi, istatistiksel verilere dayanan yatırımın, bilimsel yaklaşımları kullanarak, şans faktörü karşıtı olduğunu kanıtlanır. Bu sayede borsalara yatırım yapmanın ekonomik açıdan tek yol olmadığını göstererek, parayı nakit olarak tutmanın bir başka yolu olduğunu gösterebiliriz.

Literatür araştırması sonucunda sosyal medya verilerinin kripto para hareketlerini tahmin etmede yardımcı olabileceği gösterilmiştir. Çalışmada atılan tweetler ve Bitcoin değeri arasında pozitif bir ilişki saptanırsa yöntem yatırımcıların başvurabileceği bir kaynak haline gelip yatırımcıların aldığı riski azaltıp daha güvenli yatırım yapmalarını sağlayacaktır. Ayrıca literatürde halen finans alanında Türkçe duygu analizi çalışmaları görece çok azdır. Bu çalışma ile literatüre katkı sunulmaktadır.

Çalışma, insanları madeni para ve banknotlardan (Euro, Dolar vb.) çok kripto paralara yatırım yapmaya teşvik edebilecek bir kaynak olma potansiyeli taşıdığından çevre anlamında da faydalı olması mümkündür. Bu durum örneğin gelecekte bankalar arası para taşıyan araçlara daha az ihtiyaç duyulmasını sağlayabilir. Bu sayede daha az karbonmonoksit, karbondioksit gibi egzoz gazlarının salınımı ve daha az yakıt tüketimini mümkün kılabilir. Bu da şüphesiz soluduğumuz havanın kalitesini arttıracaktır.

**TEŞEKKÜR:** Bu çalışma Eskişehir Teknik Üniversitesi Bilimsel Araştırma Projeleri Komitesi tarafından desteklenmiştir (ESTÜ ADP proje numarası: 19ADP048).  
Bu çalışma ayrıca TÜBİTAK 2242 Üniversite Öğrencileri Araştırma Proje Yarışmaları Asya Bölge Sergisi'ne katılmıştır.

## **VII. KAYNAKLAR**

- [1] S. Karasu, A. Altan, Z. Saraç, and R. Hacıoglu, "Prediction of bitcoin prices with machine learning methods using time series data," presented at the 26th Signal Processing and Communications Applications Conference (SIU), İKÇU, Izmir, 2018.
- [2] F. Aslan, I. Pençe, M. S. Çeşmeli, and A. Kalkan, "Bitcoin'in Türkiye piyasasındaki değerinin yapay zeka teknikleri ile tahmini," presented at the 5th International Management Information Systems Conference, Ankara, 2018.
- [3] C. Zheshi, L. Chunhong, S. Wenjun, Y. Efendiev and A. Keller, "Bitcoin price prediction using machine learning: an approach to sample dimension engineering," *Journal of Computational and Applied Mathematics*, Eds. Netherlands, vol. 36, 2020.
- [4] S. Rahman, J. N. Hemel, S. J. A. Anta, H. A. Muhee, and J. Uddin, "Sentiment analysis using r: an approach to correlate cryptocurrency price fluctuations with change in user sentiment using machine learning," presented at the 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, 2018.
- [5] B. Sakiz, and E. Kutlugün, "Bitcoin price forecast via blockchain technology and artificial intelligence algorithms," presented at the 26th Signal Processing and Communications Applications Conference (SIU), İKÇU, Izmir, 2018.
- [6] D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel, and B. K. Lama, "Recurrent neural network based bitcoin price prediction by twitter sentiment analysis," presented at the IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), Nepal, 2018.
- [7] C. Lamon, E. Nielsen, and E. Redondo, "Cryptocurrency price prediction using news and social media sentiment," in *SMU Data Sci. Rev.*, vol. 1, no. 3, pp. 1-22, 2017.
- [8] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," *Entropy*, vol. 21, no. 6, pp. 589, 2019.
- [9] M. Wimalagunaratne, and G. Poravi, "A predictive model for the global cryptocurrency market: a holistic approach to predicting cryptocurrency prices," presented at the 8th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Singapore, 2018.
- [10] S. Colianni, S. Rosales, and M. Signorotti, "Algorithmic trading of cryptocurrency based on twitter sentiment analysis," *CS229 Project*, California, 2015, pp. 1-5.



- [11] D. Ayata, M. Saraçlar, and A. Özgür, “Turkish Tweet sentiment analysis with word embedding and machine learning,” presented at the 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017.
- [12] B. Agarwal, and N. Mittal, *Prominent Feature Extraction For Sentiment Analysis*, vol. 28, Switzerland, 2016.
- [13] B. Agarwal and N. Mittal, “Semantic Feature Clustering For Sentiment Analysis of English Reviews,” *IETE Journal of Research*, vol. 60, pp. 414-422, 2014.
- [14] S. A. El Rahman, F. A. AlOtaibi, and W. A. AlShehri, “Sentiment Analysis of Twitter Data,” presented at 2019 International Conference on Computer and Information Sciences (ICCIS), Suudi Arabistan, 2019.
- [15] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera, “Opinion Mining And Sentiment Analysis On A Twitter Data Stream,” *International Conference on Advances in ICT for Emerging Regions (ICTer2012)*, Sri Lanka, 2012,
- [16] M. Meral and B. Diri, “Sentiment Analysis on Twitter”, in *22nd Signal Processing and Communications Applications Conference (SIU)*, Trabzon, 2014.
- [17] Ö. Çoban, B. Özyer, and G. T. Özyer, “Sentiment Analysis For Turkish Twitter Feeds,” in *23rd Signal Processing and Communications Applications Conference (SIU)*, Malatya, 2015.
- [18] B. İ. Sevindi, “Türkçe metinlerde denetimli ve sözlük tabanlı duygu analizi yaklaşımlarının karşılaştırılması,” Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Gazi Üniversitesi, 2013.
- [19] B. Karaöz and U. T. Gürsoy, “Adaptif Öğrenme Sözlüğü Temelli Duygu Analiz Algoritması Önerisi,” *Bilişim Teknolojileri Dergisi*, vol. 11, pp. 245-253, 2018.
- [20] S. Mete, O. Çakır, O. Bayat, D. Göksel Duru, and A. Duru, “Gözbebeği Hareketleri Temelli Duygu Durumu Sınıflandırılması,” *Bilişim Teknolojileri Dergisi*, vol. 13, pp. 137-144, 2020.
- [21] F. S. Çetin and G. Eryiğit, “Türkçe Hedef Tabanlı Duygu Analizi İçin Alt Görevlerin İncelenmesi – Hedef Terim, Hedef Kategori ve Duygu Sınıfı Belirleme,” *Bilişim Teknolojileri Dergisi*, vol. 11, pp. 43-56, 2018.
- [22] M. Loureiro and M. Alló, “Sensing Climate Change and Energy Issues: Sentiment and Emotion Analysis with Social Media in the U.K. and Spain,” *Energy Policy*, vol. 143, 2020.
- [23] K. Sailunaz and R. Alhajj, “Emotion and Sentiment Analysis From Twitter Text,” *Journal of Computational Science*, vol. 36, 2019.
- [24] Ç. Ü. Yurtöz, “Measuring The Sentiment Effects Using Emoticon Features For A General Turkish Corpus,” Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Galatasaray Üniversitesi, 2019.
- [25] M.S. Neethu and R. Rajasree, “Sentiment Analysis in Twitter using Machine Learning Techniques,” in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, India, 2013.
- [26] B. Çağlar and U. Yavuz, “Finansal Haberlerin Bitcoin Fiyatlarına Etkisinin Yapay Sinir Ağları İle Analizi,” *Bilişim Teknolojileri Dergisi*, vol. 14, pp. 65-78, 2021.
- [27] E. Haddi, X. Liu, and Y. Shi, “The role of text pre-processing in sentiment analysis,” *Procedia Computer Science*, vol. 17, pp. 26-32, 2013.

- [28] T. Renault, "Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages," *Digital Finance*, vol. 2, no. 1, pp. 1-13, 2020.
- [29] B. Pang, L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.
- [30] A. Roy and M. Ojha, "Twitter sentiment analysis using deep learning models", in *IEEE 17th India Council International Conference (INDICON)*, India, 2020.
- [31] (2019, November 1). Google Trends. [Online]. Available: <https://trends.google.com.tr/trends/explore?date=today%205y&geo=TR&q=%2Fm%2F05p0rrx>
- [32] (2019, November 1). Blockchain. [Online]. Available: <https://www.blockchain.com/tr/charts/market-price?timespan=all>