




## Metin Sınıflandırmada Yapay Sinir Ağları ile Bitcoin Fiyatları ve Sosyal Medyadaki Beklentilerin Analizi

Cihan ÇILGIN<sup>1\*</sup> , Ceyda ÜNAL<sup>2</sup> , Serkan ALICI<sup>3</sup> , Ekin AKKOL<sup>4</sup> ,  
Yılmaz GÖKŞEN<sup>5</sup> 

<sup>1</sup>Research Asst., Bolu Abant İzzet Baysal University, Department of Management Information Systems, Bolu, Turkey

<sup>2</sup>Research Asst., Dokuz Eylül University, Department of Management Information Systems, İzmir, Turkey

<sup>3</sup>Research Asst., Dokuz Eylül University, Department of Management Information Systems, İzmir, Turkey

<sup>4</sup>Research Asst., İzmir Bakırçay University, Department of Management Information Systems, İzmir, Turkey

<sup>5</sup>Prof. Dr., Dokuz Eylül University, Department of Management Information Systems, İzmir, Turkey

Geliş Tarihi/Received: 27.11.2019  
Kabul Tarihi/Accepted: 06.02.2020

Doi: doi.org/10.31200/makuubd.651904  
Araştırma Makalesi/Research Article

### ÖZET

Son yıllarda, bloglar, tweet'ler, forumlar, e-postalar gibi Web 2.0 hizmetleri iletişim kanalı olarak yaygın bir şekilde kullanılmaktadır. Ayrıca sosyal medya; gerek bilgi paylaşımı gerekse istek, şikayet ve dilekler gibi görüşleri belirtmenin en kolay ve en güncel yolu olarak düşünülmektedir. Sosyal medyanın, birçok alana olduğu gibi Bitcoin fiyatlarına olan etkisi de son yıllarda tartışılmaktadır.

Bitcoin yıllardır üzerinde durulan ve popülerliği her geçen gün artan bir yatırım aracıdır. Merkezi olmayan bir elektronik para birimi sistemi olan Bitcoin, çok sayıda kullanıcının ilgisini çeken, finansal sistemlerdeki köklü bir değişikliği ifade etmektedir. Bu çalışmada sosyal medyanın, özellikle Twitter kanalından elde edilen tweet'ler bazında, Bitcoin fiyatı ile etkileşimi ortaya konulmuştur. Bunun için 06.10.2018-19.05.2019 tarihleri arasında Twitter kullanıcıları tarafından atılan toplam 2.819.784 tweet üzerinden makine öğrenmesi yöntemlerinden sınıflandırma algoritmaları kullanılarak çeşitli analizler gerçekleştirilmiştir. Bulgular değerlendirildiğinde metin sınıflandırmada %90 ile en yüksek doğruluk oranına sahip olan Yapay Sinir Ağları kullanılmıştır. Ayrıca Bitcoin fiyatları ve sınıflandırılmış olumlu/olumsuz tweet oranları ile ikili korelasyon yapılmıştır. Elde edilen 0,681 korelasyon katsayısı ile pozitif yönde orta üstü kuvvetli ilişki tespit edilmiştir.

**Anahtar kelimeler:** Çok Katmanlı Yapay Sinir Ağları, Metin Sınıflandırma, Bitcoin, Twitter.

## **In Text Classification, Bitcoin Prices and Analysis of Expectations in Social Media with Artificial Neural Networks**

### **ABSTRACT**

In recent years, Web 2.0 services such as blogs, tweets, forums, emails have been widely used as communication channel. Also, social media; it is considered to be the easiest and most up-to-date way to both share information and express opinions such as requests, complaints, and wishes. As in many fields, the effect of social media on Bitcoin prices has been addressed in the last few years.

Bitcoin is an investment tool that has been underlined for years, and is increasing in popularity day by day. Bitcoin, an electronic currency system that is decentralized, states a radical change in financial systems that has attracted many users. In this study, interaction of social media with Bitcoin price was revealed, particularly based on tweets obtained from Twitter channel. For this purpose, various analyses were carried out by using classification algorithms in machine learning methods over a total of 2,819,784 tweets posted by Twitter users between 06.10.2018-19.05.2019. When the findings were evaluated, Artificial Neural Networks with the highest accuracy rate of 90% was used in text classification. In addition, bilateral correlations were made with Bitcoin prices and classified positive / negative tweet rates. The correlation coefficient of 0.681 was found to be positively correlated with higher than moderate strength.

**Keywords:** Multilayer Perceptrons, Text Classification, Bitcoin, Twitter.

### **1. GİRİŞ**

Merkezi olmayan dijital para birimi Bitcoin 2008 yılında Satoshi Nakamoto takma adını kullanan, kimliği bilinmeyen bir kişi tarafından yaratılmıştır (Nakamoto, 2008: 1). Bitcoin herhangi bir şekilde kontrol mekanizmasına tabi değildir ve bu nedenle aracısız ve kimliksiz bir biçimde işlemlerin gerçekleştirilmesine olanak tanımaktadır. Bitcoin, aynı zamanda borsa üzerinden satın alınabilmekte veya karmaşık/matematikselsel yapılar çözülerek madencilik (mining) ile elde edilebilmektedir (Shah ve Zhang, 2014: 411).

İnternet teknolojilerinin gelişmesi ve Web 2.0 kavramının hayatımıza girmesiyle birlikte, insanlara duygu ve düşüncelerini paylaşma imkânı sunan sosyal medyanın kullanımı

gün geçtikçe artmaktadır. Bu artış insan hayatında birtakım değişimleri de beraberinde getirmektedir. Sosyal medyadaki söylentiler kimi zaman dünya gündeminin değişmesine dahi yol açmaktadır.

İnsanların farklı konular üzerindeki görüşlerini analiz etmek araştırmacılar için önemli konulardan biridir. Bu nedenle sosyal medya platformları farklı konular hakkındaki fikirleri analiz etmek için yararlı bir kaynak olduğundan farklı sosyal medya platformlarındaki tartışmalar analiz için kullanılmaktadır (Khan vd., 2016: 855). Bununla birlikte, Twitter farklı alanlardaki kullanıcıların çok sayıda konuyu tartıştıkları ana platformlardan biridir (Kouloumpis vd., 2011: 538). Yapılan incelemeler sonucunda Twitter verileri kullanılarak çeşitli analizler yapılan birçok çalışma bulunmaktadır (Matta vd., 2015; Ceyhan vd., 2018; Alghobiri vd., 2019; Colianni vd., 2015; Kaminski, 2014). İnternetin yaygın olarak benimsenmesi, sosyal ağların popülaritesinin artması ve kurum içindeki bilgilerin dijitalleştirilmesi, geniş dijital belge koleksiyonlarının ortaya çıkmasına neden olmuştur. Bu durum göz önünde bulundurulduğunda tüm kategorilerde dijital bilgilerin işlenmesinde otomatik sınıflandırma ve etiketleme ihtiyacı ortaya çıkmaktadır (Ghiassi vd., 2012: 10967). Metin belgeleri için sınıflandırma, araştırmacılar ve uygulayıcılar tarafından sıklıkla kullanılan önemli bir tekniktir. Metin sınıflandırma araştırmaları 1960 yıllara kadar dayanmaktadır (Maron, 1961: 404). Metin kategorizasyonu olarak da bilinen metin sınıflandırma, içeriğine göre önceden tanımlanmış bir etiket veya kategori setinden bir belgeye bir veya daha fazla sınıf etiketi veya kategori atama işlemidir (Dumais, 1998: 148, Basu vd., 2003: 1). Günümüzde özellikle sosyal medya analizinde ve metin sınıflandırmada yapay sinir ağları modelleri kullanımı her geçen gün artmaktadır. Özellikle çalışma kapsamında uygulanan metin madenciliği yöntemleri ile elde edilen veriler sınıflandırılarak, gruplandırılarak veya metinler arasında ilişkiler ve istatistiksel sonuçlar oluşturularak modeller geliştirilmektedir. Geliştirilen bu model vasıtasıyla veri kümesinde yer almayan yeni bir metin hakkında sınıflandırma yapmak mümkündür (Coşkun ve Baykal, 2011: 1).

Özellikle Twitter gibi sürekli etkileşime açık mecralarda, çok sayıda takipçisi olan hesaplar spekülatif tweet'ler atarak kripto para borsasında ani dalgalanmalara sebep olmaktadır. Ancak bu konuda tweet atan bütün kullanıcıların genel etkisi anlamında herhangi bir yorumda bulunmak ilgili veriler analiz edilmeden mümkün değildir. Bu çalışmanın amacı Bitcoin etiketi (hashtag) baz alınarak elde edilen tweet'lere metin sınıflandırma uygulanarak yapay sinir ağları yöntemiyle Bitcoin fiyatları ile sosyal medya arasındaki etkileşimi ortaya koymaktır.

Çalışma, üç ana bölümden oluşmaktadır. İlk bölümde literatürdeki ilgili çalışmalardan söz edilmiştir. İkinci bölümde çalışma kapsamında kullanılan yöntemler detaylı bir biçimde açıklanmıştır. Üçüncü bölümde ise uygulama adımlarına yer verilmiştir. Sonrasında ise bulgular ve sonuçlar paylaşılmıştır.

## 2. İLGİLİ ÇALIŞMALAR

Literatürde bu konuyla ilgili yapılan bazı çalışmalar, Tablo 1’de karşılaştırılmalı olarak gösterilmiştir.

**Tablo 1.** İlgili çalışmaların karşılaştırılması

Yayın Yılı	Yazarlar	Makale İsmi	Kullanılan Metin	Kullanılan Algoritmalar/Modeller
2018	Kinderis, M., Bezbradica, M., & Crane, M	Bitcoin Currency Fluctuation	Haber Makaleleri, Tweet	k-En Yakın Komşu, Lojistik Regresyon, Karar Ağaçları, Destek Vektör Makineleri, Gaussian Naive Bayes, Doğrusal Ayrımçılık Analizi
2015	Matta, M., Lunesu, I., & Marchesi, M.	Bitcoin Spread Prediction Using Social and Web Search Media	Tweet, Web arama sonuçları	
2018	Ceyhan, K., Kurtulmaz, E., Sert, O. C., & Özyer, T.	Bitcoin movement prediction with text mining	Günlük haber başlıkları	Lojistik Regresyon, Destek Vektör Makineleri, Yapay Sinir Ağları, Adaboost ve Rastgele Orman
2019	Alghobiri, M.	Using data mining algorithm for sentiment analysis of users’ opinions about Bitcoin cryptocurrency	Tweet	Naive Bayes
2019	Shintate, T. & Pichl, L.	Trend prediction classification for high frequency bitcoin time series with deep learning	Bitcoin Fiyatları	Çok katmanlı algılayıcı, LSTM
2015	Colianni, S., Rosales, S. & Signorotti, M.	Algorithmic trading of cryptocurrency based on Twitter sentiment analysis	Tweet	Naive Bayes, Lojistik Regresyon ve Destek vektör makinesi
2017	Lamon, C., Nielsen, E., & Redondo, E.	Cryptocurrency price prediction using news and social media sentiment	Haber ve sosyal medya verileri	Lojistik Regresyon, Naive Bayes, Doğrusal destek vektör makinesi
2019	Burnie, A., & Yilmaz, E.	Social media and bitcoin metrics: which words matter	Reddit gönderileri	

**Kaynak:** Yazarlar

Kinderis vd. (2018), çalışmalarında haber makaleleri/tweet'lerden alınan veriler ve kripto para değerlerinin gidişatı arasındaki ilişkiyi veri madenciliği teknikleri ile ortaya koyan bir model önermişlerdir. Bunun için Uzun-Kısa Süreli Bellek Tekrarlayan Sinir Ağları (The Long Short-Term Memory Recurrent Neural Network- LSTM RNN) kullanılmıştır. Bu yöntem, tahmine dayalı makine öğrenmesinin yanı sıra, duygu analiz tekniklerinden oluşan karma bir modelin ortaya çıkarılmasına yardımcı olması amacıyla kullanılmıştır. Kestirimci modellemede, yatırım karar verme sürecinde ikili bir tahmin vermek için finansal ve duygusal verileri sınıflandırma amaçlı olarak çeşitli makine öğrenmesi algoritmaları kullanılmıştır. Quandl's API'sinden önemli miktarda finansal veri toplanmasının ardından, Bitcoin ile ilgili tweet'ler ve duygu analizi için gerekli makaleleri etkin bir şekilde harmanlamak için Selenium kütüphanesinden yararlanılmıştır. Belirli bir tarihte ortalama 6000 tweet toplanmasına rağmen, kaynaklar makinenin RAM kapasitesi ile sınırlandırılmıştır. 2013 yılından itibaren, bir ay içerisinde 8620 makale ve 7.000.000'in üzerinde tweet toplanmıştır. Bir duyguyu nitelendirmek için, TextBlob kütüphanesiyle Doğal Dil İşleme (Natural Language Processing), metinden bilgilendirici bir veri seti çıkarılarak gerçekleştirilmiş ve bu da Bitcoin hakkındaki düşünceye dair genel bir duygu göstergesi vermiştir. Farklı günlerin duygularını ve ertesi güne ilişkin finansal tahminleri açıklayan değerler belirlendikten sonra, Bitcoin'in fiyatının yönünü tahmin etmek için sınıflandırma algoritmalarından k-En Yakın Komşu (K-Nearest Neighbour), Lojistik Regresyon (Logistic Regression), Karar Ağaçları (Decision Trees), Destek Vektör Makineleri (Support Vector Machine), Gaussian Naive Bayes ve Doğrusal Ayrımcılık Analizi (Linear Discriminant Analysis) uygulanmıştır. Gaussian Naive Bayes ve Doğrusal Ayrımcılık Analizinin en doğru tahminleri verdiği gözlenmiştir (Sırasıyla %67.4 ve %67.6). Modelin başarısı, Bitcoin fiyat değişikliklerinin yönünü tahmin etme bağlamında değerlendirildiğinde, bulguların, piyasadaki diğer mevcut modellerle karşılaştırıldığında Bitcoin fiyatlarındaki dalgalanmaların daha doğru ve gerçek zamanlı tahminler verdiğini ortaya koymuştur.

Matta vd. (2015), Bitcoin fiyatlarının, tweet'lerin hacimleriyle veya Web Arama medya sonuçlarıyla ilgili olup olmadığını araştırmışlardır. Ayrıca, fiyat eğilimlerini Google Trends verileriyle, tweet'lerin hacmiyle ve özellikle de olumlu bir duyguyla ifade edenlerle karşılaştırmışlardır. Tweet'ler, Twitter Streaming API kullanılarak #Bitcoin veya @bitcoin hashtag'ini oluşturarak toplanmıştır. Twitter'da Ocak 2015-Mart 2015 (60 gün) arasında yayınlanan Bitcoin ile ilgili 1.924.891 tweet analiz edilmiştir. Bitcoin pazarındaki tweet'lerin zaman çizelgesi ve dalgalanmalarını karşılaştırarak daha iyi bir korelasyon değeri sağlayan belirli bir gün belirlenmiş ve daha sonra Twitter'dan çıkarılan yorumları değerlendirmek için

SentiStrenght uygulaması kullanılmıştır. Tüm tweet'lerin girdisi olarak sistem pozitif duygu için 1, negatif duygu için -1 ve nötr için 0 değeri verilmiştir. Özellikle Bitcoin fiyatı ile Google Trends verileri arasında anlamlı bir korelasyon değeri bulunmuştur (0.64). Zaman serileri arasındaki çapraz korelasyon analizinin sonuçlarından, pozitif tweet'lerin Bitcoin'in fiyatının birkaç gün içindeki hareketini tahmin etmeye katkıda bulunabileceği vurgulanmıştır.

Ceyhan vd. (2018), metin madenciliği ve yapay zekanın birleşiminden oluşan bir model ortaya koyarak Bitcoin hareket yönünü tahmin etmeyi amaçlamışlardır. Tahminleme amaçlı olarak Yapay Sinir Ağları, Destek Vektör Makineleri ve Lojistik Regresyondan yararlanılmıştır. Aynı zamanda günlük haber başlıkları metin madenciliği teknikleriyle analiz edilmiş ve tahminleme sonuçlarına olan etkileri gözlemlenmiştir. Çalışmada Bitcoin'in geçmiş verileri ve günlük haber başlıklarını içeren iki veri seti kullanılmıştır. Her iki veri seti için, Temmuz 2010 ve Haziran 2016 tarihleri arasında veri toplanmış ve 1466 adet örnek oluşturulmuştur. Çalışmada 250 ve 750 adet öznitelik üretilerek Bitcoin Geçmiş Verileri ile Günlük Haber Başlıkları veri setleri birleştirilmiş ve Lojistik Regresyon, Destek Vektör Makineleri, Yapay Sinir Ağları, Adaboost ve Rastgele Orman (Random Forest) modelleri uygulanmıştır. Modeller Bitcoin Geçmiş Verileri veri seti üzerinde %75 eğitim ve %25 test olarak gruplandırılmıştır. Model analizlerinde Lojistik Regresyon (LR), Destek Vektör Makineleri (SVM) ve Yapay Sinir Ağları (ANN) en yüksek performans gösteren modeller olarak gözlemlenmiş ve doğruluğu artırmak için optimizasyon uygulanmıştır. Ayrıca, model parametrelerinin son haline gelebilmesi için Grid Search yöntemi uygulanmış ve bunun sonucunda LR %65, SVM %66 ve ANN %67 doğruluk oranlarını vermiştir. Bitcoin Geçmiş Verileri ve Günlük Haber Başlıkları veri seti birleştirildiğinde ise, Bitcoin hareket yönü için tahmin başarısı Adaboost ile %91'e yükseltilmiştir.

Alghobiri (2019), çalışmasında, kullanıcı tweet'lerinin Bitcoin konusundaki duygu analizini yapmıştır. Bu amaçla, Bitcoin ile ilgili Twitter'dan elde edilen gerçek veri seti kullanılmıştır. Veri seti, daha iyi karşılaştırmalı analizler için tweet'ler, retweet'ler, bahsetmeler (mention), harici linkler içeren tweet'ler ve Bitcoin kripto para birimi ile ilgili tartışan kullanıcılar hakkında genel kapsamlı veri analizi de dahil olmak üzere beş ayrı bölüme ayrılmıştır. Sınıflandırma algoritması olarak Naive Bayes kullanılmıştır. Araştırmada kullanılan veri seti Kaggle'dan alınmıştır. Veri kümesi, farklı kullanıcıların 50.859 tweet'ini içermektedir. Veri seti, tweet'in oluşturulduğu zaman, tweet'i oluşturan kullanıcı, tweet'in metni ve tweet'in duygu (sentiment) skoru ile ilgili tüm ayrıntıları içermektedir. Verilerin sınıflandırılması aşamasında tweet'ler pozitif (4) ve negatif (0) olarak sınıflandırmak için tweet

veri setinde Naive Bayes modeli uygulanmıştır. Terim-doküman matrisi, her tweet'te bulunan belirteçler temelinde oluşturulmuş ve frekans, matrislerin her biri için hesaplanmıştır, daha sonra, frekansa dayalı olarak, Naive Bayes sınıflandırma modeli, test veri kümesi için duyguyu tahmin etme amaçlı eğitilmiştir. Sistemin performansı, standart performans değerlendirme teknikleri kullanılarak değerlendirilmiştir. Sonuçlar değerlendirildiğinde, Genel başarı %67 ile %77 arasında kalmıştır; bunların arasında kripto para birimli tweet'lerin sonuçları %67 iken, genel doğruluk sonuçları muhtemelen daha iyi bir eğitim ve test veri seti nedeniyle % 77 olarak ortaya çıkarılmıştır. Ayrıca, veri setinin genel duyguları olumlu olduğundan ve kullanıcıların Bitcoin hakkında çok az olumsuz görüşü olduğundan, duyarlılık sonuçları % 87 ile % 93 değerleri arasında kalmıştır.

Shintate ve Pichl (2019), durağan olmayan kripto para zaman serileri için rastgele örnekleme yöntemi (Random Sampling Method-RSM) olarak adlandırılan bir trend tahmini sınıflandırma yaklaşımı ortaya koymuşlardır. Bu çerçevede derin öğrenmeye dayanmaktadır. Girdi sınıflandırmak istenilen bir dizi (sequence) çifti ve son geçmişten örneklenen diziler (sırasıyla yukarı, aşağı ve statik olarak etiketlenmiş üç dizi) vardır. Daha sonra, her bir dizi bağımsız olarak kodlanmış ve gizli gösterimler elde edilmiştir ve sonuç olarak giriş dizisinin gizli gösterimini ve örneklenen dizilerin gizli gösterimleri karşılaştırılarak en benzer sınıf ortaya çıkarılmıştır. Bunun için, OkCoin Bitcoin market (Çin Yuanı ve Amerikan Doları için) zaman serisi verileri 13 Haziran 2013-18 Mart 2017 tarihleri arasında dakika bazlı olarak elde edilmiştir. Veri seti, Kaiko (Dijital Varlık Veri Sağlayıcısı) tarafından sağlanmıştır. Veri seti (BTCCNY); eğitim, doğrulama ve test bölümlerine ayrılmıştır. Temel modeller için, eğitim seti 920.484 dakika, 120.000 dakikalık doğrulama ve 120.000 dakikalık testten oluşmuştur. Mevcut yöntem için, rastgele örnekleme metodu, eğitim süresi 910,352 dakikaya kısaltılırken, doğrulama ve test setlerinin ikisi de 120,000 dakika olarak ortaya çıkmıştır. Sonuçlar, çok katmanlı algılayıcı (Multi Layer Perceptron) ve LSTM yöntemiyle karşılaştırılmış ve çalışma kapsamında ortaya konulan modelin en yüksek doğruluk elde ettiği gözlemlenmiştir (%57).

Colianni vd. (2015), Twitter'dan elde edilen verilerin denetimli yapay öğrenme yöntemleri kullanılarak kripto paraların gidiş yönünün tahminlenmesini amaçlamışlardır. Bu amaçla, öğrenme algoritmaları için bir eğitim ve test verisi oluşturulmuş ve bunun için açık kaynaklı bir Python kütüphanesi olan Tweepy kullanılmıştır. Anahtar kelime, Bitcoin, gerçek zamanlı olarak Twitter üzerinden aranmıştır. Twitter veri seti ile dijital para piyasası hareketini belirlemek için, metin sınıflandırma ve duygu analizi algoritmaları kullanılmıştır. Her bir algoritmanın amacı, Bitcoin fiyatının belirli bir zaman diliminde artacağını veya azaltacağını

tahmin etmektir. Metin sınıflandırma yaklaşımı için, Naive Bayes, Lojistik Regresyon ve SVM uygulanmıştır. Verilerin %70'i eğitim ve %30'u test için ayrılmıştır. Sınıflandırmanın ilk adımında, bir tweet'teki sözcükler, her sınıflandırıcı için özellik vektörünün öğeleri olarak ele alınmış ve Bernoulli Naive Bayes, günden güne %95.00 ve saatten saate %76.23'lük doğruluğu elde ederek tüm metin sınıflandırma algoritmalarından en iyi sonuca ulaşmıştır.

Lamon vd. (2017), haber ve sosyal medya verileri kullanarak, üç kripto para birimi için fiyat dalgalanmalarının tahmini üzerine bir çalışma gerçekleştirmişlerdir. Çalışmada Bitcoin, Litecoin ve Ethereum para birimleri kullanılmıştır. Günlük haberler ve sosyal medya verileri, olumlu ya da olumsuz duygulardan ziyade, her para birimi için gerçek fiyat değişikliklerine dayanarak etiketlenmiştir. Birçok farklı sınıflandırma algoritmasının kullanıldığı çalışmada, Bitcoin ve Litecoin fiyatları için en iyi sonucu Lojistik Regresyon (Logistic Regression), ethereum fiyatları için en iyi sonucu Bernoulli Naive Bayes vermiştir.

Burnie ve Yılmaz (2019), çalışmalarında Reddit gönderilerinin Google aramaları ile güçlü korelasyonuna dayanarak hangi kelimelerin Bitcoin fiyatlandırma dinamiğinde önemli olduğu tespit edilmiştir. Çalışma kapsamında Reddit seçilmesinin sebebi Google aramaları ile Reddit aramaları arasındaki uyumun yüksek oranda olmasıdır. Aynı zamanda kelimelerin fiyat dinamiğinde etkili olup olmadığını anlamak için DDPWI yöntemi geliştirilmiştir.

Mai vd. (2018), yaptıkları çalışmada metin madenciliği ve vektör hata düzeltme modellerini kullanarak sosyal medyanın Bitcoin değeri üzerindeki etkilerini araştırmışlardır. Çalışmada, forumlarda yazılanların Twitter gibi diğer mecralarda yazılanlara göre Bitcoin değeri üzerinde daha büyük etkisi olduğu sonucuna ulaşılmıştır. Ayrıca tüm sosyal medya mesajlarının eşit etkiye sahip olmadığı da ortaya konulmuştur.

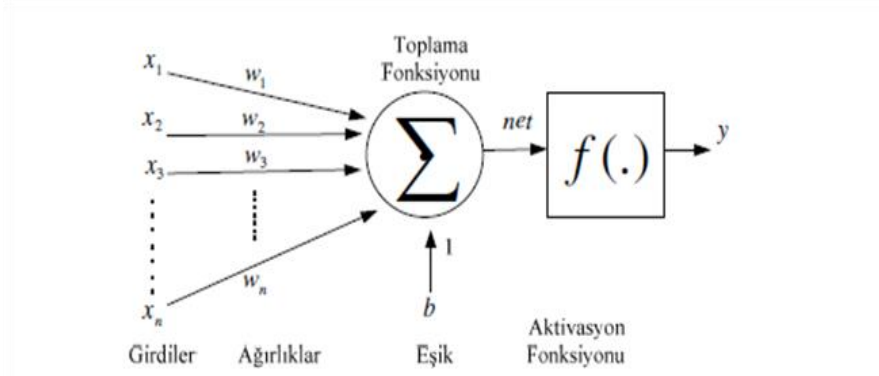
### **3. METODOLOJİ**

Belgeleri otomatik olarak sınıflandırmak için, son on yılda K-En Yakın Komşu yöntemi, Naive Bayes, çok değişkenli regresyon modelleri, karar ağaçları, Destek Vektör Makineleri, Yapay Sinir Ağları ve Genetik Algoritma tabanlı yöntemler gibi birçok istatistiksel ve makine öğrenme tekniği tasarlanmıştır (Deng vd., 2019: 3797). Bu çalışma kapsamında Yapay Sinir Ağları, Lojistik Regresyon, Destek Vektör Makineleri ve Naive Bayes sınıflandırma modelleri uygulanmıştır.

Yapay sinir ağları, insan beyninin özelliklerinden oluşan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım



almadan otomatik olarak gerçekleştirmek amacı ile geliştirilen bilgisayar sistemleridir. Kısaca, yapay sinir ağları insan beyninin çalışma mekanizmasından esinlenmektedir. Biyolojik sinir ağlarının sinir hücreleri olduğu gibi yapay sinir ağlarının da yapay sinir hücreleri bulunmaktadır. Yapay sinir hücrelerinin çalışmasına olanak sağlayan en küçük birimler, proses/işlem elemanı olarak adlandırılmaktadır. Her işlem elemanının girdiler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktılar olmak üzere beş temel elemanı vardır (Öztemel, 2016: 48). En basit anlamda bir yapay sinir hücresinin yapısı Şekil 1 de görülmektedir. Yapay Sinir Ağları ilk olarak metin sınıflandırmada Wiener vd. (1995) tarafından kullanılmıştır. Wiener vd. (1995), modellerinde, her kategoride giriş katmanından bir kategoriye doğrusal olmayan bir işlemeyi öğretmek için üç katmanlı bir sinir ağı yapısı kullanmışlardır.



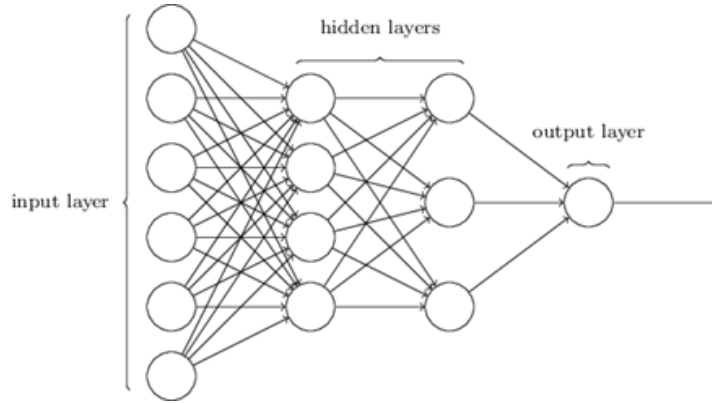
**Şekil 1.** Yapay sinir hücresinin yapısı

Girdiler, bir yapay sinir hücresine dış dünyadan gelen bilgileri kapsamaktadır. Ağırlıklar ise, hücreye gelen bilgilerin o hücre üzerindeki etkisini ve önemini göstermektedir. Toplama fonksiyonu, net girdinin hesaplanabilmesi için girdi değerleri ve ağırlıklar çarpılarak toplanmaktadır. Eşitlikte toplama fonksiyonunun matematiksel ifadesine yer verilmiştir.

$$net = \sum_{i=1}^n w_i x_i + b$$

Yapay sinir ağlarında nöronlar, katmanlar içerisinde bulunmaktadır. Bu katmanlar girdi katmanı, ara katmanlar ve çıktı katmanı olarak ifade edilmektedir. Tek katmanlı algılayıcılarda ara katman bulunmamaktadır. Ara katmanlar, girdi katmanından gelen verileri işleyerek bir sonraki katmana gönderme işlevini gerçekleştirmektedir. Yapay sinir ağlarının girdi ve çıktıları arasında doğrusal bir ilişki olmaması durumunda öğrenmenin gerçekleşebilmesi için Çok

Katmanlı Algılayıcılar (ÇKA-Multilayer Perceptron) kullanılmaktadır. Bu yüzden çalışma kapsamında bu metottan yararlanılmıştır (Şekil 2).



Şekil 2. Çalışmada kullanılan ÇKA modeli

Çalışmada kullanılan yapay sinir ağı topolojisi, bir giriş katmanı, gizli bir katman ve bir çıkış katmanından oluşan üç katmanlı, tamamen bağlantılı bir ileri besleme ağıdır. İlgili literatür çalışmaları incelendiğinde (Lam ve Lee, 1999: 195; Wiener vd., 1995: 332, Basu vd., 2003: 7) metin sınıflandırmada kullanılan yapay sinir ağları topolojilerinin genelde üç katmandan oluştuğu gözlemlenmiştir. Bu çalışma kapsamında kullanılan bu ağ yapısının alternatifleri üzerine yapılan deneyler sonucunda da ağın başarısının kayda değer bir oranda artmadığı gözlemlenmiş ve katman sayısındaki bir artışın değerlendirilecek veri miktarı göz önünde bulundurulduğunda ağın hem eğitim süresini hem de değerlendirme sürecini oldukça artırdığı gözlemlenmiştir. Giriş katmanında 512 giriş proses elemanı bulunmaktadır. Giriş katmanında, giriş birimi sayısı azaltılmış özellik alanının boyutuna eşittir. Başka bir ifade ile her metin (tweet) içinde yer alan kelimelerin sayı ile ifade edilen vektör matrisinin oluşturulmasında kelime maksimizasyonunun 512 seçilmesinden dolayı giriş katmanındaki giriş birim sayısı da 512 olarak belirlenmiştir. Şekil 2’de görüldüğü gibi uygulanan modelde bir gizli katman ve gizli katmanda 512 nöron bulunmaktadır ( $n=512$ ), gizli katmanlardan bir tanesi Dropout (Seyreltme) amaçlı kullanılmıştır. Katmanlarda belirlenen eşğin altındaki düğümler başarımın artması amacıyla seyreltilmiştir. Gizli katmanlarda 512 nöron bulunmaktadır. Gizli katmanda yer alan nöron sayısı yapılan çeşitli denemeler sonucunda ve bu alanda yapılan diğer çalışmaların incelenmesinin ardından belirlenmiştir. Bu çerçevede sürekli artan ve sürekli azalan bir iterasyon yapısı ile denenen gizli katman nöron sayılarındaki değişikliklerin sinir ağının performansı üzerinde olumlu anlamda bir etkisi olmadığı gözlemlenmiş ve bu sayıya karar verilmiştir. Aynı zamanda sinir ağında bir çıkış katmanı da bulunmaktadır. Gizli

katmanlarda aktivasyon fonksiyonu olarak ReLU (Rectified Linear Unit) ve çıkış katmanında “Softmax” fonksiyonu kullanılmıştır.

ReLU (Rectified Linear Unit) işlem karmaşıklığı çok az olan ve oldukça hızlı çalışan bir aktivasyon fonksiyonudur. Sonuçlara ulaşılma hızı fazla olduğundan derin öğrenme uygulamalarında sıklıkla tercih edilmektedir. ReLU fonksiyonu aşağıdaki gibi formülize edilmektedir.

$$f(x) = \max(0, x)$$

Özetle ReLU fonksiyonu gönderilen değer pozitif ise aynı değeri çıktı olarak verirken, negatif girdileri sıfır olarak değiştirmektedir. ReLU'nun çıktı aralığı  $[0, \infty)$  şeklinde ifade edilmektedir (Eriş, 2018: 38).

Softmax fonksiyonu, karşılıklı olarak  $k$  kadar sınıfı tahmin etmek için kullanılır ve  $\mathbf{1} \times \mathbf{k}$  boyutlu olasılık vektörünü çıkarmaktadır. Vektöre ait tüm olasılıkların toplamı bire eşit olmaktadır (Hinton vd., 2006: 1527). Çalışmada kategorizasyon işlemi uygulandığından Softmax fonksiyonu tercih edilmiştir. Softmax fonksiyonu eşitlikte gösterilmektedir.

$$f(x) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

Veri setinde yer alan yaklaşık 3 milyon verinin ağa katılabilmesi için hızlı olduğundan bu fonksiyon tercih edilmiştir. Optimizasyon fonksiyonu olarak “adam” fonksiyonundan yararlanılmıştır. Adam (Adaptive Moment Estimation), stokastik gradyan azaltma yöntemi yerine tercih edilebilecek daha verimli, adaptif bir optimizasyon algoritmasıdır (Gülcü ve Kuş, 2019: 508). Adam algoritması, her bir parametre için dinamik bir şekilde öğrenme oranını güncellemektedir. (Kingma vd., 2014: 1; Ruder, 2016: 1). Bunun yanında, hesaplama yükü olarak verimli olup, düşük bellek gereksinimleriyle etkili bir biçimde çalışabilmektedir. Adam optimizasyonu, en çok kullanılan yöntem olarak gözlemlenmektedir. Adam optimizasyon fonksiyonunda ağırlıkların güncellenmesi eşitlikteki gibi formülize edilmektedir.

$$\omega - \alpha \frac{v_{d\omega}}{\sqrt{\delta_{d\omega} + \epsilon}}$$

Naive Bayes sınıflandırıcı, istatistiksel bir sınıflandırıcıdır. Bayes teoremine dayanır. Belirli bir grubun belirli bir sınıfa ait olma olasılığını mevcut sınıflanmış durumdaki örnek verileri kullanarak yeni bir verinin mevcut sınıflardan herhangi birini hesaplayan bir yaklaşımdır. Naive Bayes sınıflandırıcıları, bir nitelik değerinin verilen bir sınıf üzerindeki

etkisinin, diğer niteliklerin değerlerinden bağımsız olduğunu varsaymaktadır (Han vd., 2011: 83).

$$L(X, Y) = \prod_{i=1}^m p(z^i, y^i)$$

Destek Vektör Makineleri, istatistiksel öğrenme teorisi alanında ortaya çıkmış bir öğrenme metodudur (Cortes ve Vapnik, 1995: 273). Destek Vektör Makineleri, Vapnik vd. (1998) tarafından, sınıflandırma görevleri için pozitif ve negatif veri örneklerinden oluşan bir dizi veri noktasını ayırmak için mümkün olan en geniş marjlı optimum bir hiper düzlem inşa etmek için yapısal risk minimizasyonu prensibine dayalı bir yöntem olarak tanıtılmıştır (Deng vd., 2019: 3797). Bu yöntem sınıflandırmayı doğrusal veya doğrusal olmayan bir fonksiyon kullanılmasıyla gerçekleştirir. Verileri ayırt etmek için en uygun fonksiyon tahmin edilmektedir (Özkan, 2016: 169).

$$R = \sum_{i=1}^l L_z + \Omega(l, h)$$

Lojistik Regresyon bir dizi tahmin unsuru değişkenin değerlerine dayanan niteliğin varlığı veya yokluğunu tahmin etmede kullanılan çok değişkenli bir analiz modelidir (Lee, 2004: 223). Lojistik regresyon analizi diğer analizlere göre sağladığı avantajlar sayesinde kategorik veri analizinde önemli bir yere sahiptir.

$$\sigma(x) = \frac{1}{1 + \exp(-W^T x)}$$

$$l(x) = \sum_{i=1}^m x^i \log \sigma(x^{(i)}) + (1 - y^{(i)}) \log(1 - \sigma(x^{(i)}))$$

Çalışma kapsamında ayrıca vektör oluşturmada öznitelik seçiminde BoW (Bag of Words) yöntemi kullanılmıştır. BoW yöntemi ile vektörü oluşturacak öznitelikler, veri kümesini oluşturan metinler üzerindeki kelimelerin köklerine ayrıştırılmasıyla veya en çok rastlanan kelimelerin atlanmasıyla belirlenmektedir (Blanzieri ve Bryl, 2008: 63).

### 3.1. Veri Seti

Bu çalışmada veri seti olarak kullanılan “Bitcoin” etiketi altında Twitter kullanıcıları tarafından atılan İngilizce tweet’ler Python Scrapy kütüphanesi kullanılarak elde edilmiştir.

Veri seti 06.10.2018-19.05.2019 tarihleri arasında Twitter kullanıcıları tarafından atılan toplam 2.819.784 İngilizce Twitter gönderisinden meydana gelmektedir.

JSON formatında elde edilen bu veri içerisinde tarih bilgisi, mesaj içeriği, kullanıcı bilgisi, URL, retweet sayısı, beğeni sayısı ve kullanıcı id gibi bilgiler yer almaktadır. Bu verilerden öncelikli olarak ilgisiz olan alanlar (URL, retweet sayısı, kullanıcı bilgisi gibi) çıkarılarak sınıflandırma yapmak için gerekli olacak metin bilgileri ve tarih bilgileri bir veri tablosu haline getirilmiştir. Sonrasında elde edilen bu tweet'ler Tablo 2'de gösterildiği üzere metin madenciliği ön işleme aşamalarından geçirilerek her birinden 500 adet olmak üzere pozitif (1), negatif (2) ve nötr (3) olarak etiketlenmiş ve eğitim verisi olarak belirlenip yapay öğrenme işlemi gerçekleştirilmiştir. Eğitim verisi için etiketlenen tweet'ler yaklaşık 200.000 tweet arasından seçilmiştir.

Özellikle tweet'lerin etiketlenmesi aşamasında olumlu yöndeki beklentiler, Bitcoin üzerine yapılan devlet düzenlemeleri veya şirketlerin Bitcoin kullanımına başlaması veya desteklemesi gibi tweet'ler 0 olarak (pozitif) etiketlenmiştir. Öte yandan Bitcoin kullanımı ve fiyatları üzerine yapılan olumsuz yorumlar ve beklentiler ise 1 olarak (negatif) etiketlenmiştir. Ayrıca çalışmada direkt kullanılmayacak olsa da elde edilen örneklemin yapısını bozmaması için tespit edilmek istenen nötr tweet'ler de bulunmaktadır. Bu tweet'ler genel itibari ile reklamlardan, anlık fiyat bildirimlerinden vb. Bitcoin üzerine direkt etkisi olmayan tweet'lerden oluşmaktadır. Bu nötr tweet'ler ise 2 (nötr) olarak etiketlenmiştir. Bulunan bu nötr tweet'ler öğrenmeyi olumsuz etkilediği tespit edilmiş olup ilerleyen aşamalardan öğretim veri setinden çıkarılmışlardır.

Çalışmada, kolay öğrenilebilen, kolay okunan ve birçok hazır veri bilimi kütüphanesi bulunduran Python (Python 3.6) programlama dilinden yararlanılmıştır. Veri bilimi çalışmaları için en temel ve en yaygın kullanılan; Numpy, Pandas, Scikit-Learn ve Keras gibi kütüphanelerini içerisinde barındırması sebebiyle Python yazılım dili tercih edilmiştir.

**Tablo 2.** Manuel olarak etiketlenen tweet'ler

Metin (Tweet)	Etiket
Arabia will be launching its own digital currency next year CryptoNewsReview <a href="http://dlvr.it/R50dnm">http://dlvr.it/R50dnm</a> Å btc bitcoin	0
There are 5 or 6 coin that will change the world economy. The rest are short term speculative trading until they fizzle out. cryptocurrency bitcoinSaudi	1
Are you are serious about EARNING FREE Bitcoin you sleep? Install CryptoTab browser and get real Bitcoin for using it. I have successfully done PAYOUTS Become financially free Crypto <a href="http://bit.ly/freebytes">http:// bit.ly/freebytes</a> Å fintech btc PassiveIncome eth cryptomining	2

#### **Kaynak:** Yazarlar

Metin sınıflandırma aşamasında ise Pozitif (1) 825.495, Negatif (2) 730.140 ve Nötr (3) 1.264.149 olmak üzere toplam 2.819.784 tweet sınıflandırılmıştır. Ayrıca sosyal medya (Twitter) ve Bitcoin fiyatı arasındaki ilişkiyi analiz etmek amacıyla tweet'lerin elde edildikleri tarih aralığındaki Bitcoin kapanış fiyatları ve sınıflandırılmış olumlu/olumsuz tweet oranları ile veri seti oluşturulmuştur.

#### **4. BULGULAR**

Çalışma kapsamında öncelikli olarak metin sınıflandırmada sıklıkla kullanılan Naive Bayes, Destek Vektör Makineleri, Lojistik Regresyon ve Yapay Sinir Ağları yöntemleri, belirlenen metrikler doğrultusunda en iyi başarı oranını elde etmek için karşılaştırılmıştır. Çalışma kapsamında kullanılan tüm sınıflandırma yöntemlerinde yukarıda da bahsedildiği üzere toplam 1500 tane etiketlenmiş verinin %70'i eğitim, %30 ise test veri seti olarak kullanılmıştır. Eğitim ve test veri setinin oluşturulması aşamasından doğru bir sınıflandırma yapılabilmesi açısından her tür sınıftan eşit sayıda verinin ayrılmasına dikkat edilmiştir. Sınıflandırma algoritmalarının etiketler bazında değerlendirilmesine ait başarı metrikleri aşağıda yer alan Tablo 3'te detaylı olarak gösterilmiştir.

**Tablo 3.** Sınıflandırma algoritmalarının etiketler bazında değerlendirilmesi

Algoritma	Etiketler (1=pozitif, 2=negatif, 3=nötr)	Kesinlik (Precision)	F- Ölçütü (F1-Score)
Naive Bayes	1	0.68	0.72
	2	0.70	0.73
	3	0.77	0.68
Destek Vektör Makineleri	1	0.74	0.74
	2	0.64	0.70
	3	0.63	0.66
Lojistik Regresyon	1	0.67	0.70
	2	0.68	0.70
	3	0.69	0.64

**Kaynak:** Yazarlar

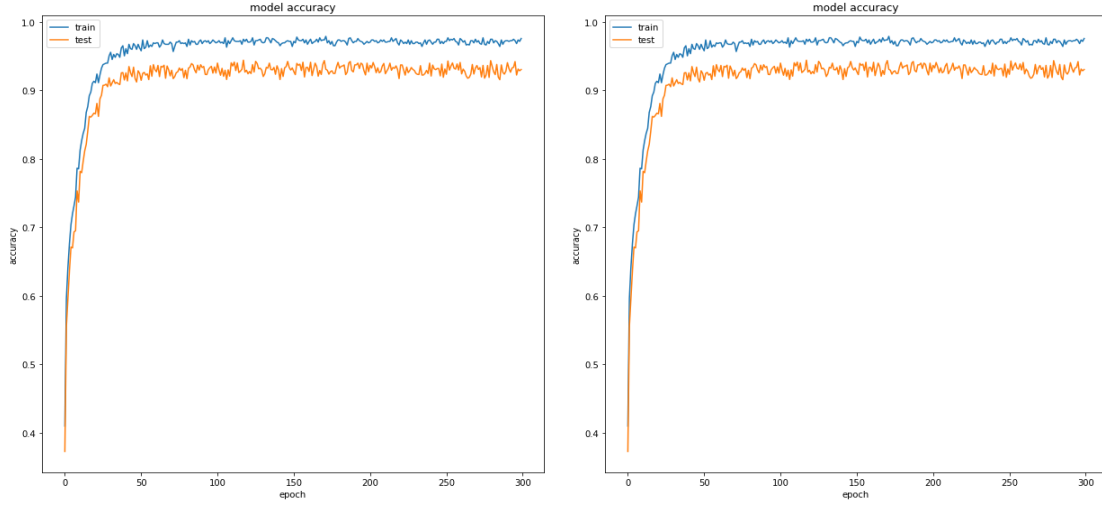
Doğruluk (accuracy) değerleri Tablo 4’de gösterilmiştir.

**Tablo 4.** Sınıflandırma algoritmalarının değerlendirilmesi

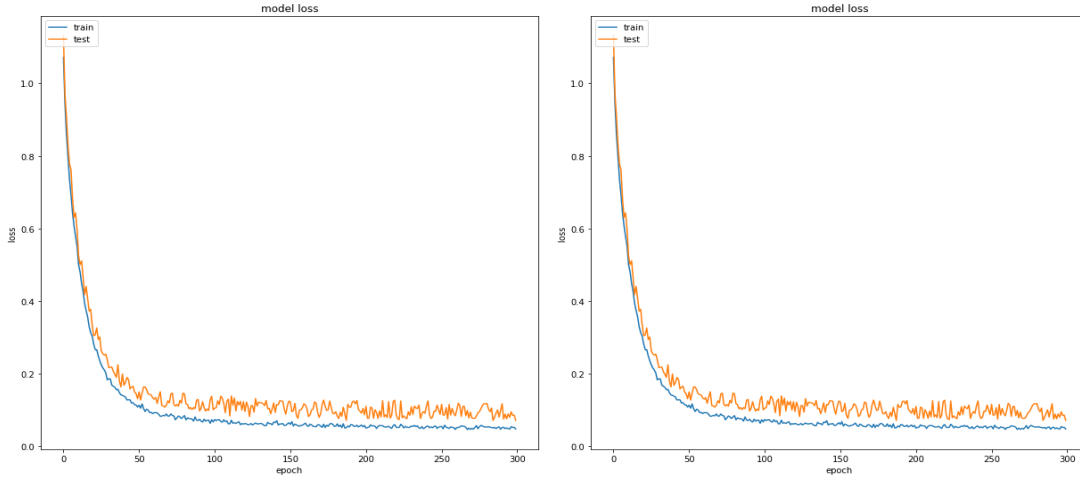
Algoritma	Doğruluk (Accuracy) Değeri
Naive Bayes	0.71
Destek Vektör Makineleri	0.70
Lojistik Regresyon	0.68
Yapay Sinir Ağları	0.9058

**Kaynak:** Yazarlar

Çalışma kapsamında kullanılan sınıflandırma algoritmalarının testi ve veri setinden elde edilen doğruluk oranları Tablo 5’te gösterildiği gibi olup 0.9058’lik başarı oranı ile yapay sinir ağları modelinin en başarılı olduğu görülmektedir. Bu nedenle metin sınıflandırma aşamasında yani sınıflandırılmamış tweet’lerin etiketlemesi aşamasında yapay sinir ağları modeli kullanılmıştır.



Şekil 3. YSA modeli eğitim ve test veri seti için doğruluk değerleri



Şekil 4. YSA modeli eğitim ve test veri seti için loss değerleri

Şekil 3 ve Şekil 4’te gösterildiği üzere oluşturulan sinir ağı topolojisi 300 epoch sonucunda yukarıda gösterilen doğruluk ve loss değerlerine ulaşmaktadır. Yukarıda yer alan Şekil 3 ve Şekil 4’ten de anlaşılacağı üzere 200 ile 300 epoch arasında hem başarı oranında hem de loss değerlerinde görüldüğü üzere, ağın öğrenmesi yavaşlamış, yapılan denemelerden sonra ise 300 epoch’tan sonra ağın öğrenme gerçekleştirmediği gözlemlenmiştir. Bu nedenle ağın eğitim veri seti ile eğitilmesinde daha fazla iterasyona gerek duyulmamıştır. Oluşturulan Yapay Sinir Ağının ağırlıklarının değiştirilme zamanı olarak bütün eğitim veri setinin ağ gösterilmesi yani epoch temelli eğitim kullanılmıştır. Bu yaklaşımda eğitim veri setindeki tüm örnekler ağa tek tek gösterilmekte ve daha sonra tüm eğitim veri setinin ağa gösterimi tamamlandıktan sonra eğitim setinde meydana gelen hatalar toplanmaktadır (Öztemel, 2016). Diğer bir ifade ile ağın ağırlık değerleri eğitim veri setindeki tüm örneklerin sinir ağına



gösterilmesinin ardından değiştirilmektedir. Aynı zamanda, kayıp fonksiyonu için kategorik çapraz entropi (categorical cross-entropy) fonksiyonu tercih edilmiştir. Bunun sebebi özellikle ikiden fazla etiket olmasıdır. Birden fazla çıktı sınıfını tahmin ederken kullanılan ortak kayıp fonksiyonu kategorik çapraz entropi fonksiyonu olarak adlandırılmakta ve aşağıdaki gibi formülize edilmektedir (Escontrela, 2018).

$$H(y, \hat{y}) = \sum_i y_i \log \frac{1}{\hat{y}_i} = - \sum_i y_i \log \hat{y}_i$$

Tasarlanan sinir ağının eğitim veri seti ile eğitim tamamlanmasının ardından etiketlenmek istenen metinler ağa tek tek sunulmuştur. Bu işlem ile sınıflandırma aşamasında ise Tablo 5'te gösterildiği üzere toplamda 2.819.784 adet tweet Pozitif, Negatif ve Nötr olarak etiketlenmiştir.

**Tablo 5.** Sınıflandırılmış (etiketlenilmiş) tweet sayıları

Sınıf (Etiket)	Sayı
Pozitif (0)	825.495
Negatif (1)	730.140
Nötr (2)	1.264.149

#### Kaynak: Yazarlar

Twitter üzerinden toplanan tüm tweet'ler etiklendikten sonra nötr olarak etiketlenen tweet'lerin analiz üzerine bir etkisi bulunmamasından dolayı çıkarılmıştır. Burada nötr tweet'lerin tespit edilme amacı pozitif veya negatif etkisi olmayan bu tweet'lerin Pozitif veya Negatif olarak etiketlenmesini engelleyerek daha doğru bir sınıflandırma elde etmektir. Bu aşamadan sonra elde edilen sınıflandırılmış metinler Tablo 6'da gösterildiği üzere tarih bazlı olarak toplanmış ve günlük bazda olumlu tweet oranları hesaplanmıştır.

**Tablo 6.** Tarih bazlı sınıflandırılmış (etiketlenilmiş) tweet sayıları

Tarih	Olumlu Tweet Sayısı	Olumsuz Tweet Sayısı	Olumlu Tweet Oranı
31.01.2019	3429	2849	0,5462
30.01.2019	3566	2946	0,54761
29.01.2019	3585	3116	0,5350
28.01.2019	3303	3209	0,5072
27.01.2019	2765	2459	0,4908
.	.	.	.
.	.	.	.
.	.	.	.

#### Kaynak: Yazarlar

Tablo 6’da gösterilen olumlu tweet oranı her gün için aşağıdaki şekilde hesaplanmıştır:

$$\text{Olumlu Tweet Oranı} = \frac{\text{Olumlu Tweet Sayısı}}{\text{Olumlu Tweet Sayısı} + \text{Olumsuz Tweet Sayısı}}$$

19.5.2019 ile 10.6.2018 tarihleri arasındaki tüm günler için hesaplanan olumlu tweet oranı ile günlük bazda Bitcoin fiyatlarının kapanış fiyatları alınarak Tablo 7’de gösterildiği gibi üç sütundan oluşan bir veri seti haline getirilmiştir.

**Tablo 7.** Tarih bazlı sınıflandırılmış (etiketlendirilmiş) tweet sayıları

Tarih	BTC Kapanış Fiyatı	Olumlu Tweet Oranı
31.1.2019	3.529,70	0,5462
30.1.2019	3.478,00	0,5476
29.1.2019	3.501,10	0,535
28.1.2019	3.597,10	0,5072
27.1.2019	3.630,70	0,4908
.	.	.
.	.	.
.	.	.

**Kaynak:** Yazarlar

Tablo 7’de gösterilmekte olan BTC Kapanış Fiyatı ve Olumlu Tweet Oranı üzerinde aşağıda gösterilmekte olan korelasyon formülü kullanılarak korelasyon analizi yapılmıştır.

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

BTC kapanış fiyatı ve sınıflandırılmış olan olumlu/olumsuz tweet oranları ile ikili korelasyon yapılmıştır. Elde edilen 0,681 korelasyon katsayısı ile bu iki değişken arasında pozitif yönde orta üstü kuvvetli ilişki bulunmuştur.

## 5. TARTIŞMA VE SONUÇ

Bitcoin fiyat dinamikleri, kripto para biriminin popülerliği her geçen gün arttığı ve daha geniş bir kitle tarafından tanındığından dolayı tartışmalı bir konu haline gelmiştir. Bu çalışmada, Twitter üzerindeki sosyal medya faaliyetlerinden elde edilen veriler ile Bitcoin fiyatlarının etkileşimi araştırılmıştır. Bitcoin kullanımı giderek yaygınlaştığından, fiyatlar ve Twitter verilerinin birbirlerine olan etkisi analiz edilmiştir. Atılan tweet’ler ile bitcoin fiyatlarının arasında bir etkileşimin var olup olmadığı, özellikle kripto para borsasında yatırım yapan yatırımcılara anlamlı ve yol gösterici nitelikte olacaktır. Kripto para yatırımcıları için çok hızlı değişim gösteren bitcoin fiyatını etkileyen etkenleri bilmek hayati önem taşımaktadır.

Özellikle sosyal medyadaki yorumlar yatırımcıların kararlarını büyük ölçüde etkilemektedir. Bu kapsamda Bitcoin etiketi (hashtag) altında 10.06.2018-19.05.2019 tarihleri arasında atılan 2.819.784 tweet kullanılarak yapay sinir ağları ile metin sınıflandırma yöntemi uygulanmıştır. Özellikle yukarıda bahsedilen diğer metin sınıflandırma yöntemlerine kıyasla metin sınıflandırmada yapay sinir ağları daha yüksek bir doğruluk oranı verdiği için yöntem olarak YSA modeli kullanılmıştır.

Çalışmanın kısıtları, çalışmada sosyal medya platformlarından sadece Twitter'ın kullanılması, Twitter'ın belirli bir kesime hitap etmesi ve baz alınan zaman olarak özetlenebilir.

Bitcoin fiyatları ve sınıflandırılmış olumlu/olumsuz tweet oranları ile ikili korelasyon yapılmıştır. Elde edilen 0,681 korelasyon katsayısı ile pozitif yönde orta üstü kuvvetli ilişki bulunmuştur. Genel olarak, kripto para birimleri özellikle Bitcoin hala gelişme döneminde olduğundan, sosyal medyadan elde edilen verilerin kripto para piyasasıyla olan etkileşimini tahmin etmek zorlaşmaktadır. Bu nedenle, çalışma kapsamında geliştirilen model, sadece Twitterdan elde edilen Bitcoin hashtagli verilerin Bitcoin fiyatlarına etkisini öngörmekle kalmayıp aynı zamanda bu iki durumun birbirleriyle nasıl bir etkileşim halinde olduklarını değerlendirme ve tahmin etme potansiyeline sahip olduğu için bu konuya ışık tutmaktadır. Gelecek çalışmalarda makine öğrenmesi algoritmalarının doğruluğunu artırmak için hata analizi alanında ek araştırmalar yapılabilir. Yapay sinir ağları ve yapay öğrenme yöntemleriyle zaman serisi tahminlerinde, metin madenciliği ile gerçekleştirilen tweet sınıfları tahmin gücünü artırmak üzere birleşik bir model olarak kullanılabilir. Aynı zamanda farklı kaynaklardan elde edilen verilerle niyet analizi çalışmalarına yön verilebilir. Ayrıca, farklı sosyal medya platformlarından elde edilecek veriler entegre edilerek birçok sosyal medya platformlarının Bitcoin fiyatına etkisi incelenebilir.

## REFERENCES / KAYNAKLAR

- Alghobiri, M. (2019). Using data mining algorithm for sentiment analysis of users' opinions about Bitcoin cryptocurrency. *Journal of Theoretical and Applied Information Technology*, 97(8), 2195-2205.
- Basu, A., Walters, C. & Shepherd, M. (2003). Support vector machines for text categorization. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 4(4), 1-8.
- Blanzieri, E. & Bryl, A. (2008). A survey of learning-based techniques of email spam filtering. *Artificial Intelligence Review*, 29(1), 63-92.
- Burnie, A. & Yilmaz, E. (2019). Social media and bitcoin metrics: which words matter. *Royal Society Open Science*, 6(10), 191068, 1-18.

- Ceyhan, K., Kurtulmaz, E., Sert, O. C. & Özyer, T. (2018). Bitcoin movement prediction with text mining. *26th Signal Processing and Communications Applications Conference*, 1-4.
- Cheque Cerda, G. & L Reutter, J. (2019). Bitcoin price prediction through opinion mining. *In Companion Proceedings of The 2019 World Wide Web Conference*, 755-762.
- Colianni, S., Rosales, S. & Signorotti, M. (2015). Algorithmic trading of cryptocurrency based on Twitter sentiment analysis (ss. 1-5). Erişim Tarihi: 25.07.2019, [http://cs229.stanford.edu/proj2015/029\\_report.pdf](http://cs229.stanford.edu/proj2015/029_report.pdf)
- Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Coşkun, C. & Baykal, A. (2011). Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması. *Akademik Bilişim*, 1-8.
- Deng, X., Li, Y., Weng, J. & Zhang, J. (2019). Feature selection for text classification: A review. *Multimedia Tools and Applications*, 78(3), 3797-3816.
- Dumais, S., Platt, J., Heckerman, D. & Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. *In: Proceedings of the seventh international conference on Information and knowledge management*, ACM, 148-155.
- Erdal, H. (2015). Makine öğrenmesi yöntemlerinin inşaat sektörüne katkısı: basınç dayanımı tahminlemesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 12(1), 109-114.
- Eriş, M. (2006). *Derin öğrenme yöntemleri kullanarak adli bilişim incelemelerinde delil çıkarımının gerçekleştirilmesi* (Basılmamış yüksek lisans tezi), Fırat Üniversitesi, Elazığ.
- Escontrela, A. (2018). *Convolutional neural networks from the ground up*. Erişim tarihi: 23.08.2019, <https://towardsdatascience.com/convolutional-neural-networks-from-the-ground-up-c67bb41454e1>.
- Ghiassi, M., Olschimke, M., Moon, B. & Arnaudo, P. (2012). Automated text classification using a dynamic artificial neural network model. *Expert Systems with Applications*, 39(12), 10967-10976.
- Gülcü, A. & Kuş, Z. (2019). Konvolüsyonel sinir ağlarında hiper-parametre optimizasyonu yöntemlerinin incelenmesi. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 7(2), 503-522.
- Han, J., Kamber, M. & Pei, J. (2011). *Data mining concepts and techniques (3rd Edition)*. ABD: The Morgan Kaufmann Series in Data Management Systems.
- Hinton, G. E., Osindero, S. & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527-1554.
- Kaminski, J. (2014). Nowcasting the Bitcoin market with twitter signals. *arXiv preprint arXiv:1406.7577*.
- Kinderis, M., Bezbradica, M. & Crane, M. (2018). Bitcoin currency fluctuation. *3rd International Conference on Complexity, Future Information Systems and Risk*, 31-41.
- Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Khan, R., Khan, H. U., Faisal, M. S., Iqbal, K. & Malik, M. S. I. (2016). An analysis of Twitter users of Pakistan. *International Journal of Computer Science and Information Security*, 14(8), 855-864.
- Kouloumpis, E., Wilson, T. & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! *Fifth International AAAI Conference on Weblogs and Social Media*, 538-541.
- Lam, S. L., & Lee, D. L. (1999). Feature reduction for neural network based text categorization. *Proceedings of 6th International Conference on Advanced Systems for Advanced Applications*, 195-202.
- Lamon, C., Nielsen, E. & Redondo, E. (2017). Cryptocurrency price prediction using news and social media sentiment. *SMU Data Sci. Rev*, 1(3), 1-22.

- Lee, S. (2004). Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. *Environmental Management*, 34(2), 223-232.
- Mai, F., Shan, Z., Bai, Q., Wang, X. & Chiang, R. H. (2018). How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems*, 35(1), 19-52.
- Maron, M. E. (1961). Automatic indexing: An experimental inquiry. *Journal of the ACM (JACM)*, 8(3), 404-417.
- Matta, M., Lunesu, I. & Marchesi, M. (2015). Bitcoin spread prediction using social and web search media. *Workshop Deep Content Analytics Techniques for Personalized & Intelligent Services*, 1-10.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
- Özkan, Y. (2016). *Veri madenciliği yöntemleri*. İstanbul: Papatya Bilim.
- Öztemel, E. (2016). *Yapay sinir ağları*. İstanbul: Papatya Yayıncılık.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.
- Shah, D. & Zhang, K. (2014). Bayesian regression and Bitcoin. 52nd annual Allerton conference on communication, control, and computing (Allerton). 409-414
- Shintate, T. & Pichl, L. (2019). Trend prediction classification for high frequency bitcoin time series with deep learning. *Journal of Risk and Financial Management*, 12(1), 1-15.
- Vapnik, V. N & Vapnik, V. (1998). *Statistical learning theory*. Wiley: New York
- Wiener, E., Pedersen, J. O. & Weigend, A. S. (1995). A neural network approach to topic spotting. *Proceedings of SDAIR-95, 4th Annual Symposium on Document Analysis and Information Retrieval*, 317-332.