POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

Investigation of fluctuations in cryptocurrency transactions with sentiment analysis

Duygu analizi ile kripto para işlemlerindeki dalgalanmalarin incelenmesi

***Yazar(lar) (Author(s))**: Uğur DEMİREL[1], Handan ÇAM[2]*

ORCID[1]: 0000-0002-4992-5632

ORCID[2]: 0000-0003-0982-2919

To cite to this article: Demirel U., Çam H., "Investigation of Fluctuations in Cryptocurrency Transactions with Sentiment Analysis", ***Journal of Polytechnic***, *(*): *, (*).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Demirel U., Çam H., "Duygu Analizi ile Kripto Para İşlemlerindeki Dalgalanmaların İncelenmesi", ***Politeknik Dergisi***, *(*): *, (*).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: 10.2339/politeknik.1518826

Investigation of Fluctuations in Cryptocurrency Transactions with Sentiment Analysis

Duygu Analizi ile Kripto Para İşlemlerindeki Dalgalanmaların İncelenmesi

Highlights

- *Obtaining a data set consisting of Turkish financial comments*
- *Performing pre-processing to prepare comments to be sent to sentiment analysis algorithms*
- *Application of machine learning algorithms for sentiment analysis*
- *Testing and comparing machine learning algorithms on the data set*

Graphical Abstract

This study collected a dataset from Telegram for sentiment analysis. We applied some preprocessing methods and machine learning classification algorithms to this dataset for sentiment analysis.

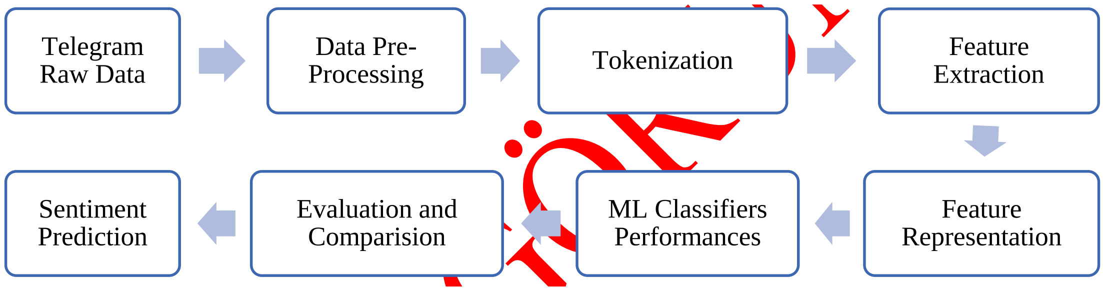

Figure. Flowchart of the Study

Aim

This study aims to evaluate whether sentiment analysis results are compatible with the trend change in the market by associating them with price fluctuations in the cryptocurrency market.

Design & Methodology

We build a dataset from Telegram via web scraping and apply some data preprocessing to this dataset. We train six classifiers on this dataset for the sentiment analysis.

Originality

Many sentiment analysis studies in the cryptocurrency domain are based on English texts. Our study addresses a significant gap by focusing on Turkish texts, thereby providing insights specific to the Turkish cryptocurrency market.

Findings

BTC (Bitcoin) is generally perceived negatively on platforms like Investing.com and Telegram. ETH (Ethereum) also exhibits more negative views. The results aid in understanding investor perceptions and market expectations towards cryptocurrencies.

Conclusion

This study provides a model for sentiment analysis of Turkish texts. Enhances the understanding of trend changes in the cryptocurrency market. Contributes to the development of new methods and approaches for future research.

Declaration of Ethical Standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Investigation of Fluctuations in Cryptocurrency Transactions with Sentiment Analysis

Research Article

Uğur DEMİREL[1*], Handan ÇAM[2]
[1]Irfan Can Kose Vocational School, Gumushane University, Türkiye
[2] Faculty of Economic and Administrative Science, Department of Management Information Systems, Gumushane University, Türkiye
(Geliş/Received : 19.07.2024 ; Kabul/Accepted : 12.09.2024 ; Erken Görünüm/Early View : 02.10.2024)

ABSTRACT

This study investigates public sentiment about popular cryptocurrencies listed on crypto exchanges in Turkey, using comments shared on social media platforms and online forums. The research seeks to enhance the existing body of knowledge by overcoming the shortcomings of sentiment analysis studies focused on Turkish texts. Data collected from social media and online forums were examined with sentiment analysis techniques. A total of 607,592 comments were analyzed, of which 89,986 were classified as negative, 72,655 as positive, and 444,951 as neutral. For binary classification, 89,986 negative and 72,655 positive examples were selected and machine-learning models were trained and tested on 162,641 examples. The study's methodology includes an in-depth examination of sentiment analysis results obtained using machine learning classifiers. The findings show how various cryptocurrencies are perceived on different social media platforms. For instance, BTC (Bitcoin) is generally perceived negatively on Investing.com and Telegram, while ETH (Ethereum) generally displays more negative views. These results help investors understand their perceptions and market expectations towards cryptocurrencies. This study deepens the role of social media sentiment analysis in cryptocurrency markets, contributing to the development of new methods and approaches for future research.

Keywords: Text mining, Natural language processing, Machine learning, Sentiment analysis, Cryptocurrency

Duygu Analizi ile Kripto Para İşlemlerindeki Dalgalanmaların İncelenmesi

ÖZ

Bu çalışma, sosyal medya platformlarında ve çevrimiçi forumlarda paylaşılan yorumları kullanarak, Türkiye'deki kripto borsalarında listelenen popüler kripto para birimleri hakkındaki kamuoyunun duyarlılığını araştırmaktadır. Araştırma, Türkçe metinlere odaklanan duygu analizi çalışmalarının eksikliklerini gidererek mevcut bilgi birikimini artırmayı amaçlamaktadır. Sosyal medya ve çevrimiçi forumlardan toplanan veriler duygu analizi teknikleriyle incelenmiştir. Toplam 607.592 yorum analiz edildi ve bunların 89.986'sı olumsuz, 72.655'i olumlu ve 444.951'i nötr olarak sınıflandırılmıştır. İkili sınıflandırma için 89.986 negatif ve 72.655 pozitif örnek seçildi ve makine öğrenimi modelleri eğitildi ve 162.641 örnek üzerinde test edildi. Çalışmanın metodolojisi, makine öğrenimi sınıflandırıcıları kullanılarak elde edilen duygu analizi sonuçlarının derinlemesine incelenmesini içermektedir. Bulgular, çeşitli kripto para birimlerinin farklı sosyal medya platformlarında nasıl algılandığını göstermektedir. Örneğin, BTC (Bitcoin) Investing.com ve Telegram'da genel olarak olumsuz algılanırken, ETH (Ethereum) genellikle daha olumsuz görüşler sergilemektedir. Bu sonuçlar, yatırımcıların kripto para birimlerine yönelik algılarını ve piyasa beklentilerini anlamalarına yardımcı olmaktadır. Bu çalışma, sosyal medya duyarlılık analizinin kripto para piyasalarındaki rolünü derinleştirerek gelecekteki araştırmalar için yeni yöntem ve yaklaşımların geliştirilmesine katkıda bulunmaktadır.

Anahtar Kelimeler: Metin madenciliği, Doğal dil işleme, Makine öğrenmesi, Duygu analizi, Kripto para

1. INTRODUCTION

As in the financial sector, determining customer ideas and opinions affects the services companies will offer. Cryptocurrency has attracted significant attention in recent years due to its increasing transaction volume and market value. Understanding the factors that cause volatility in cryptocurrency markets is critical for investors, analysts, and policymakers. As the economic and social impacts of cryptocurrencies grow, related news articles, social media posts, and forum posts are also becoming more common. The growing popularity of social media has simplified the process for individuals to share their viewpoints on social blogs. [1]. Comprehending the emotions expressed in these posts has become crucial for researchers, and the inability to manage the escalating volumes manually has necessitated the development of automated processing methods [2].

Just as stock market price predictions are important in developing economies, this is also important in cryptocurrency markets. Speculative movements in cryptocurrency markets increase price fluctuations, making it difficult for investors to make accurate predictions. Therefore, advanced analysis techniques and tools are required to increase the accuracy of price predictions and minimize the effects of speculative movements. In this regard, sentiment analysis was

*Corresponding Author
e-posta : udemirel@gumushane.edu.tr

conducted using natural language processing (NLP) and artificial intelligence approaches. One of the most common methods for computer analysis of the thoughts and feelings that people convey in text is sentiment analysis [3].

Although there are studies using the sentiment analysis method in Turkish texts [4–6], these studies generally focused on English texts and short sentences. In this study, comprehensive research was conducted using different platform types and Turkish texts for the data set. The study aims to investigate the public sentiment towards popular cryptocurrencies listed on crypto exchanges in Turkey. This research contributes to the literature by providing sentiment analysis of comments shared on social media and online forums. Investors will be able to make more informed investment decisions by predicting potential returns from various cryptocurencies under conditions of uncertainty and risk.

In this study, a sentiment analysis model for Turkish texts was developed by combining dictionary-based approaches with machine learning classifiers. The dataset was compiled from posts shared on the 'Binance Global Türkçe' Telegram channel between March 20, 2023, and January 4, 2024, specifically for use as input in these models. A total of 607,592 comments were reviewed, with 89,986 classified as negative, 72,655 as positive, and 444,951 as neutral. For binary classification, 89,986 negative and 72,655 positive samples were selected. The sentiments of the comments were categorized as negative, positive, and neutral.

To make predictions, sentiment analysis was performed using machine learning classifiers on untagged comments about cryptocurrencies shared on tr.investing.com, Twitter.com, and Telegram between April 7, 2024, and May 6, 2024. Data from the Investing.com and Twitter (X) platforms were collected using the Octoparse 8 web scraping tool [7], enabling efficient extraction of large datasets. Telegram data was obtained using its chat export tool, which allows users to easily save chats, photos, and other materials via the Telegram Desktop application. These datasets were specifically used for predicting cryptocurrency market behavior.

Finally, sentiment analysis results were evaluated in relation to cryptocurrency price fluctuations. This study contributes to understanding market trend changes by presenting a sentiment analysis model tailored for Turkish texts.

2. RELATED WORKS

The literature on sentiment analysis categorizes studies into three main approaches: machine learning, dictionary-based methods, and hybrid methods. This section provides an overview of relevant studies classified according to these methodologies.

Machine learning is a common approach used in sentiment analysis for cryptocurrency. In this method, text is preprocessed, labeled, and then fed into a machine learning model to classify the sentiment. For instance, [8] used machine learning algorithms to classify movie reviews with 82.9% accuracy using SVM. [1] trained three different models using NB, SVM, and ME on Twitter data and achieved an 83% success rate. [9–14] are other examples of studies using supervised machine learning algorithms on English texts.

Sentiment analysis of Turkish texts using machine learning has also been studied. [15] used NB, ME, SVM, and character n-gram models for sentiment analysis of Turkish political news with 77% accuracy. [16] tested machine learning methods such as NB, RF, SMO, DT(J48), and IB1 on Twitter messages to detect social media emotions. [4] used NB and SVM classifiers with feature selection methods on a Turkish movie review dataset. [17] performed sentiment analysis on Twitter data using NB, RF, and SVM with 90% F1 score. [6] gathered Turkish data from Twitter and utilized Bag of Words (BoW) and n-gram models for feature extraction, while employing Naive Bayes (NB), Multinomial Naive Bayes (MNB), Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN) algorithms for classification. [18] experimented with NB, centroid-based classifier, MLP, and SVM on Twitter data for sentiment analysis and achieved over 80% accuracy with MLP and SVM. [19] compared the performance of four feature selection methods using the EM classification algorithm on Twitter datasets. [5] combined NB and DVM algorithms with various ensemble methods and showed that combining classical classification methods improves sentiment analysis performance. [20] used classifiers such as NB, LR, SVM, DT, word bag, and fastText vector representations to predict social media sentiment on Turkish tweets using emojis. [21] tested NB multinomial, SVM, ME, and DT algorithms by measuring the effect of feature selection on different datasets and reported that NB multinomial outperforms other methods.

In recent years, studies have been conducted using deep learning algorithms along with machine learning algorithms for sentiment analysis. [22–25] are examples of studies using CNN and LSTM deep learning techniques for sentiment analysis. [26] performed sentiment analysis on Twitter datasets using LR, SVM, RF algorithms, and deep learning methods and observed that deep learning methods achieve higher accuracy than classical machine learning algorithms. [27] proposed fine-tuned multilingual and base models of BERT for Türkçe sentiment analysis on movie and hotel review datasets and stated that their models outperform existing methods. [28] developed a hybrid approach consisting of the hierarchical combination of RF and SVM classifiers, which have lower accuracy values than the hybrid approach for sentiment analysis.

In the dictionary-based approach, words expressing emotional states are searched through the dictionary and then scores expressing emotional states are obtained for each word. Various studies have been carried out on this method using resources such as SentiWordNet. [29]

created SentiWordNet based on English WordNet and assigned emotion scores. [30] developed the BlogMiner application, which extracts emotions and opinions from blog pages using SentiWordNet. Similar studies have been conducted for Turkish. [31] developed a Turkish word network within the scope of the Balkanet project. [32] created SentiTurkNet, which assigns sentiment scores to synsets in Turkish WordNet, and achieved varying levels of success. Similar approaches have been used in studies on Turkish social media data. [33] made a performance comparison by applying a dictionary-based approach and n-gram methods to Twitter data. [34] applied sentiment analysis to social media messages about Syrian refugees. [35] conducted dictionary-based sentiment analysis on different Turkish Twitter datasets.

Hybrid approaches in sentiment analysis have been created by combining machine learning algorithms and dictionary-based approaches. [36] conducted sensitivity analysis using SentiWordNet and SVM in their movie reviews. [37] achieved better performance by combining NB and genetic algorithms. [38] compared Lexicon-based and machine learning-based approaches. [2] presented a hybrid method by combining semantic rules and machine learning techniques. [39] achieved 73% success by using a combination of dictionary-based and machine-learning approaches in product reviews. [40] achieved the highest accuracy rate with Logistic Regression on Turkish datasets. [41] achieved a 7% accuracy increase by combining dictionary-based and machine learning methods with a hybrid approach.

Sentiment analysis studies on cryptocurrencies stand out as an important tool for understanding the complex relationship between social media platforms, digital data sources, and financial markets. [42] estimated the DIJA index of sentiments in daily tweets. They achieved 86% accuracy. [43] examined the relationship between Twitter posts and stock market indicators. They found a positive correlation with negative VIX. [44] used the Twitter sentiment for Apple stock prices. They showed success in two-day forecasts with Granger analysis. [45] examined the relationship between Bitcoin prices and Google Trends and Wikipedia searches. They found a two-way interaction. [46] analyzed stock performance with SeekingAlpha and Stocktwits data. They achieved 85% accuracy. [47] examined Bitcoin prices with Twitter data. They found that Twitter signals did not carry statistical significance. [48] examined the social interactions of Bitcoin users and reported having fewer social connections. [49] proposed to measure the socio-economic aspects of the cryptocurrency economy by analyzing digital footprints. [50] estimated the value of Bitcoin using Reddit and Google News data and showed that the model containing all parameters gave the best results. [51] suggested that Twitter data can be transformed into cryptocurrency trading strategies and achieved 76.23% accuracy. It shows that the use of emotional analysis in predicting price movements of cryptocurrencies is widespread and the impact of different data sources (Twitter, news sites, forums, etc.) is evaluated.

3. MATERIALS AND METHODS

The aim of this study is to contribute to investors' decision-making processes by examining the comments on popular cryptocurrencies listed on crypto exchanges in Turkey, shared on social media platforms and online forums, using sentiment analysis. In particular, it aims to help investors predict potential returns from various cryptocurrencies under uncertainty and risk. Previous studies have generally focused on Twitter data obtained from foreign sources.

In this study, we investigate public sentiment towards cryptocurrency stock market behavior by combining Turkish dictionary-based and machine learning (ML)-based approaches. The data set, created from the posts made on the "Binance Global Turkish" Telegram channel between 20.03.2023 and 04.01.2024, was used. Various cryptocurrencies have their own Telegram channels where investors can share information and communicate with others. Telegram offers a chat export tool that allows easy saving of chats along with photos and other materials. To obtain the data, the Telegram Desktop application is required on the computer.

The open-source Orange Program was used to analyze the data. The analysis process involves sentiment analysis steps performed after cleaning the data with text preprocessing. Text preprocessing is a crucial step in text

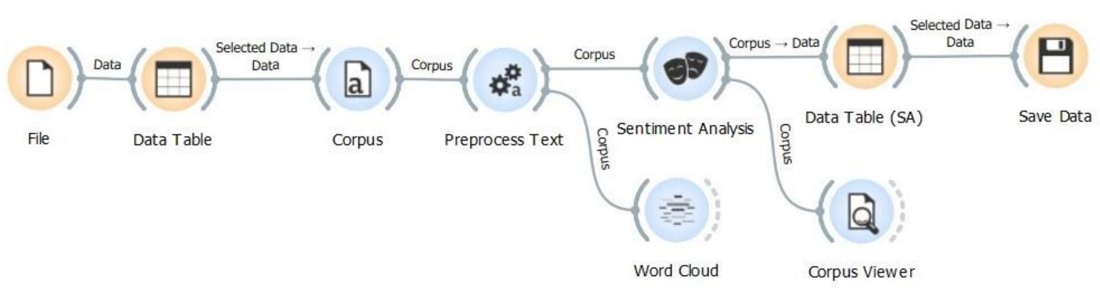

Figure 1. Lexicon-Based Method Analysis Design

mining and data mining applications, consisting of four processes. The transformation converts text to lowercase, removes diacritics, parses HTML, and removes URLs. Tokenization breaks the text into meaningful pieces. Normalization applies stemming and lemmatization to words. Filtering removes unnecessary words. After these steps are completed, word cloud can be created with the cleaned data and sentiment analysis can be performed [52]. Sentiment analysis was used to predict sentiment from Telegram data saved in Excel format. This method aims to identify positive, negative or neutral sentiments of comments using multilingual sentiment analysis methods for various languages, including Turkish [53].

Once the text preprocessing step is finished, the text data can be divided into individual texts and visualized as a word cloud, as illustrated in Figure 2. In this analysis, the most frequently used words in the comments included "çok, yok, btc, için, binance, iyi, değil, hocam, para, zaman, öyle, sonra, dolar, coin, güzel, arkadaşlar, bende, bnb, şimdi, hiç" ("very, none, btc, for, binance, good, not, sir, money, time, so, then, dollar, coin, nice, friends, me too, bnb, now, never" in English respectively). It is noteworthy that in the word cloud, in addition to words such as 'teacher' and 'friends' (addressing phrases in Turkish) originating from the chat platform, there are also important words related to stock market recommendations. This dataset offers appropriate data for developing a model using machine learning techniques for sentiment analysis.

The multilingual sentiment method introduced in the Orange was used to categorize comments according to negative, positive and neutral poles. A total of 607,592 comments were analyzed in this study. As a result of the analysis, 89,986 comments were classified as negative, 72,655 comments as positive and 444,951 comments as neutral. From this dataset, 89,986 negative and 72,655 positive examples as presented in Figure 3 were selected for training and testing machine learning models to perform binary classification.

The dataset underwent train-test splitting following the 80-20 rule. The training dataset comprises 80% of the total data, with the remaining 20% allocated for testing (10%) and validation (10%). Data from the test dataset was strictly excluded from both training and validation phases to mitigate the risk of potential model overfitting. The conceptual model outlining the procedures for sentiment forecasting on "Investing.com, Twitter (X), and Telegram" data is depicted in Figure 4.

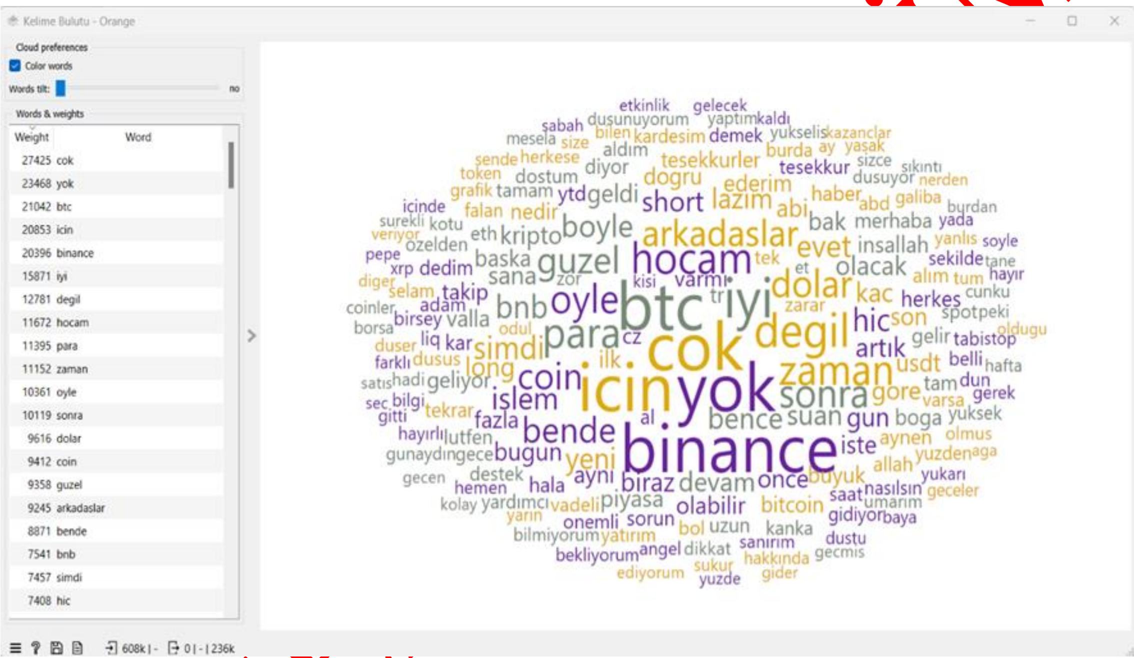

Figure 2. Word Cloud

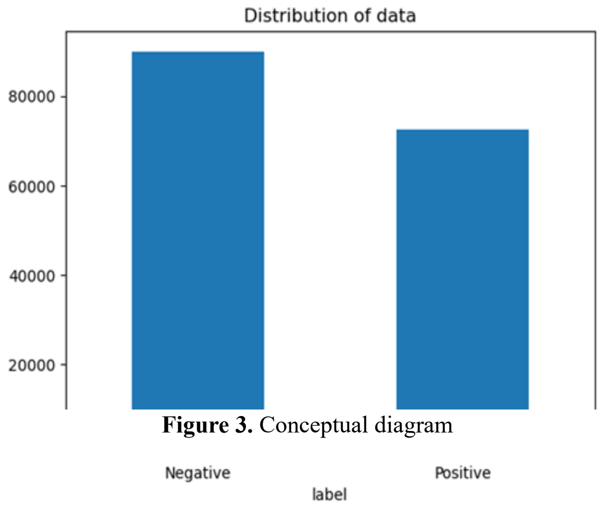

Figure 3. Conceptual diagram

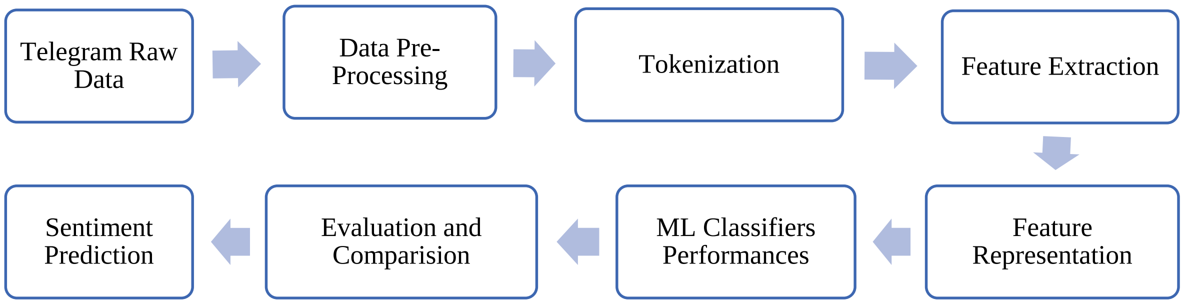

Figure 4. Distribution of dataset

The models were developed in Jupyter software, which is an open-source program that provides an interactive environment for various programming languages, using the programming language called Python. Jupyter Notebook is an open-source web application built for collaboration where users can write live code, equations, visualizations, and narrated text from one place, allowing programming in a web browser, then share Notebooks with others via email, GitHub, and Jupyter Notebook Viewer [54].

3.1. Machine Learning Classifiers

Supervised machine learning models were applied using lexicon-based features of texts to perform sentiment analysis on unlabeled Telegram data-related cryptocurrency from 20.03.2023 and 04.01.2024. Naive Bayes (NB), Logistic Regression (LR), Gradient Boosting (GB), K-Nearest Neighbor (KNN), Decision Trees (DTs), and Multilayer Perceptron (MLP) are machine learning classifiers used in the study.

Naive Bayes operates under the assumption that features are independent and equally significant, although this assumption may not always hold true in practice. Despite this limitation, Naive Bayes can be effectively applied to a wide range of machine learning problems. It is valued for its simplicity, relying on probabilistic principles compared to more intricate machine learning algorithms [5]. The feature independence assumption in Naïve Bayes is effective for text categorization when word features are taken into account. When there are a large number of features, the independence assumption allows the parameters of each feature to be learned separately, which significantly facilitates the learning process [37]. Naive Bayes uses the given word or data to estimate the probability of the class of a document or data. It also calculates the conditional probabilities of the class to which the data is likely to belong [55]. Bayes rule is formulated as in Equation 1 [1].

$$P(c|t) = \frac{P(c)P(t|c)}{p(t)} \tag{1}$$

where c represents a specific class (positive or negative) and t denotes the classified text. P(c) and P(t) are the prior probabilities of c (class) and t (text). P(t|c) signifies the probability of text appearing in a given class.

Logistic regression is a statistical procedure used for classification. The logistic regression algorithm creates a classification model using the examples in the training set and assigns the new examples to the class with the highest probability [56]. Logistic regression is a statistical method that calculates the probability of a particular outcome based on the values of only two variables. Linear regression is not suitable for binary values such as yes/no and true/false because it produces a logistic curve that ranges from 0 to 1. Logistic regression creates a model using the natural logarithm of the ratios of the target variable [57,58]. The logistic regression formula is shown in Equation 2.

$$\mathrm{Logit}(p) = \log\left(\frac{p}{1-p}\right) \tag{2}$$

The k-nn algorithm is one of the basic example-based learning algorithms. Example-based learning algorithms rely on the data in the training set to perform the learning process [59]. The K-nearest neighbor (KNN) method classifies new data by comparing it with previously categorized data, assuming that nearby data points belong to the same class. Testing examples are unclassified data, while learning examples are already categorized data. After calculating the Euclidean distance (Equation 3) between the test sample and each learning sample, KNN selects the k closest learning samples to the test sample. Then, it assigns the class (test sample class) based on the majority class among the selected k samples [57,60,61].

$$dis(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \tag{3}$$

Decision trees are a common classification method in data mining due to their cost effectiveness, interpretability, integration with database systems, and high reliability [62]. A decision tree classifier is a simple algorithm frequently used in data classification. The Decision Tree consists of internal nodes representing test conditions and leaf nodes representing class labels. This strategy involves asking questions about the characteristics of the test data set and correctly categorizing the data according to the answers to these questions [13,61,63,64].

The decision tree classifier creates a hierarchical structure in which data is divided according to attribute values. The condition or prediction is determined by the presence or absence of certain words. The data space is split iteratively until the leaf nodes reach a certain minimum number of records [65]. Figure 5 shows a standard decision tree. In the example, the probability that a customer at AllElectronics will purchase a computer is estimated. Rectangles represent internal nodes and ovals represent leaf nodes. While some algorithms only produce binary trees, others can also generate non-binary trees [61].

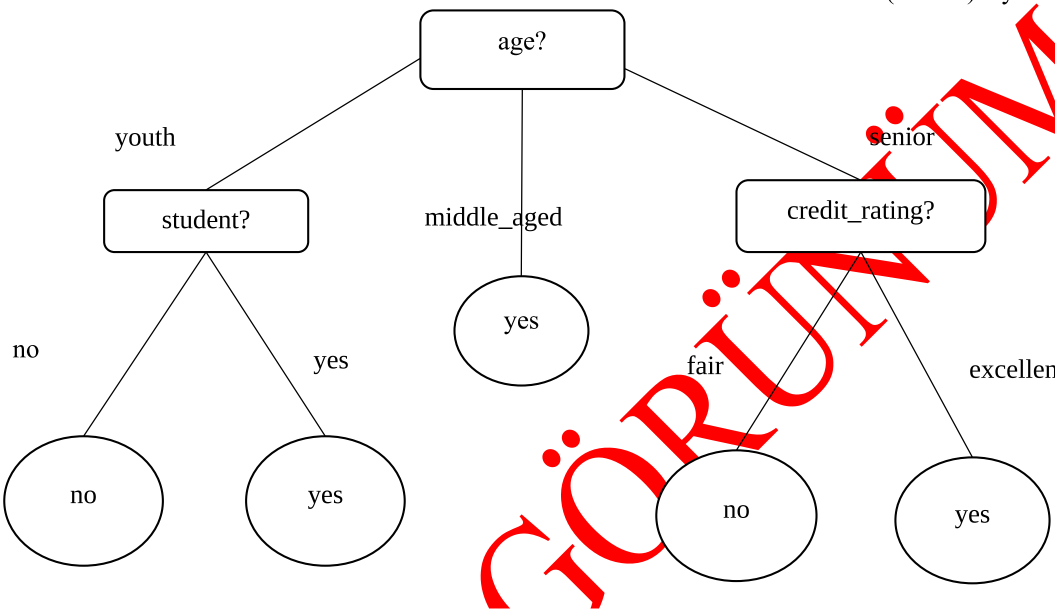

Figure 5. Decision tree concept

Decision trees are used in areas such as letter marketing (Direct Mail), credit decisions based on individuals' credit history (Credit Scoring), and determining recruitment processes by analyzing the characteristics of credit card holders. It is also widely used in various applications such as optimizing decision-making processes, identifying factors affecting sales, and finding variables that contribute to product defects through analysis of production data [62].

The gradient Boosting Algorithm is a powerful machine-learning technique that has gained significant popularity in recent years [66]. This method is a type of ensemble learning method that creates a strong prediction model by combining multiple weak learners, usually decision trees. The basic idea of gradient boosting is to iteratively build a sequence of weak learners; each learner tries to correct the previous learner's mistakes [67]. By taking the gradient (partial derivative) of the loss function, the algorithm determines the direction and magnitude of the corrections necessary to improve the performance of the model [66]. This process is repeated iteratively; each new weak learner focuses on the remaining errors and improves the model based on the improvements of previous learners [68]. The algorithm is known for its flexibility in handling complex relationships and its ability to process large, high-dimensional data sets. This powerful ensemble learning technique, widely used in machine learning, creates a robust prediction model by minimizing the overall prediction error through gradient descent optimization [69]. The solution is formulated as shown in Equation 4.

$$F(x) = \sum_{m=1}^{M} h_m(x) \tag{4}$$

Multilayer Perceptron (MLP) is an artificial neural network (ANN) that extends the capabilities of single-layer perceptrons. MLP has a feed-forward structure consisting of an input layer, an output layer, and usually one or more intermediate (hidden) layers. MLP performs learning using the backpropagation algorithm, so it is often referred to as backpropagation networks [57,70]. Single-layer perceptrons are limited to estimating only linear functions, but MLP overcomes these limitations thanks to its hidden layers and can also estimate non-linear functions [57,71]. Figure 6 details this difference.

In neural networks, particularly in classification tasks, the output layer is where the final determination of classes occurs. This layer may consist of either a single neuron or multiple neurons, the latter corresponding to

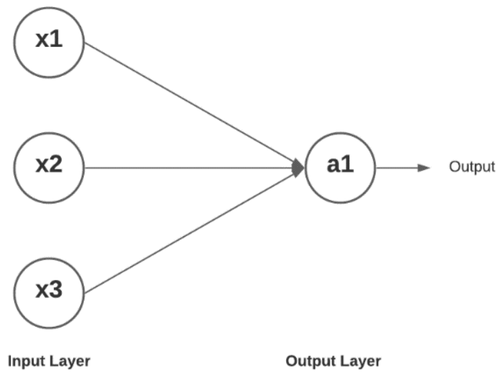

Figure 6. Single layer perceptron model [72]

the number of classes in the model. The hidden layer is the layer between the input and output layers where data is processed. The number of hidden layers and the number of neurons in each hidden layer are important factors that determine the training quality. In the MLP model, learning occurs in the form of feedforward. The training algorithm used in this model updates the weights to minimize the error. Feed-forward Artificial Neural Network (MLP) progresses from the previous layer to the following layer during the training process [18,57]. Equation 5 formulates the feedforward in the MLP model.

$$y_i = f\left(\sum_{j=1}^{n} x_j * w_{ji}\right) \tag{5}$$

4. EXPERIMENTS

The effectiveness of classifiers is generally evaluated based on two factors: the number of predictions and the F1 score. Prediction variation shows how changes made to the test data set affect the rate of correct classification of all samples [73]. Another performance measure, the area under the ROC curve (AUC), is the value that shows the prediction performance of the classification algorithm and takes a value between 0 and 1 [12]. After the pre-processing and modeling stages, the accuracy rate and F1 score of the models are examined and the results are compared. The decision matrix is used to determine precision, sensitivity (recall), accuracy (accuracy), and F1 score; True Negative (TN) and True Positive (TP) values are needed in this matrix. The precision value is calculated by dividing the number of TPs (True Positives) by the sum of TPs and False Positives (FP).

$$Precision = \frac{(TP)}{(TP+FP)} \tag{6}$$

$$Recall = \frac{(TP)}{(TP+FN)} \tag{7}$$

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \tag{8}$$

$$F1_{score} = 2x\left(\frac{Precision\ x\ Recall}{Precision+Recall}\right) \tag{9}$$

4.1. Experimental Results

Below are the performance measurements of all machine learning classifiers used in the study. Precision, sensitivity, and F1-score values for negative and positive classes are listed. Additionally, accuracy, area under the curve (AUC), and area under the precision-precision curve (AUPR) values are stated for all classes.

When the results of the applications are examined, it is seen that the Naive Bayes (NB) classifier exhibits a balanced performance in both negative and positive classes. While it gets slightly better results in the negative class, with an average of 90%, it performs slightly lower in the positive class, with an average of 88%. It is seen that the Logistic Regression (LR) classifier has a high accuracy of 94% for all classes and shows very high performance in both classes. K-Nearest Neighbor Classifier (KNN) performs close to the NB classifier in terms of accuracy for all classes. Although its sensitivity is high at 91% in the negative class, its sensitivity is 85% in the positive class. It is seen that the accuracy rate of the Decision Trees (Dts) classifier is 93% for all classes. It shows a balanced performance in both classes. It is seen that the Gradient Boosting (GB) classifier has a low performance of 84% for all classes. Finally, Multilayer Perceptron (MLP) shows a high performance of 94% for all classes. When a general evaluation is made, it is seen that the LR and MLP classifiers stand out with accuracy for all classes and a balanced performance for both classes, while the GB classifier has the lowest performance among all models.

Table 1. Comparison of Performance Measurements of Classifiers

Models	Classes	Precision	Recall	F1-score	Accuracy	AUC	AUPR
NB	Negative	90%	90%	90%	89%	0.8862	0.9028
	Positive	87%	88%	88%			
LR	Negative	93%	96%	95%	94%	0.9380	0.9509
	Positive	95%	92%	93%			
KNN	Negative	88%	91%	89%	88%	0.8780	0.9002
	Positive	89%	85%	87%			
DTs	Negative	93%	94%	93%	93%	0.9245	0.9380
	Positive	92%	91%	92%			
GB	Negative	78%	99%	87%	84%	0.8251	0.8955
	Positive	98%	66%	79%			
MLP	Negative	94%	95%	94%	94%	0.9364	0.9488
	Positive	94%	92%	93%			

Sentiment analysis was performed using six supervised machine learning models: Naive Bayes (NB), Logistic Regression (LR), Gradient Boosting (GB), K-Nearest Neighbors (KNN), Decision Trees (DTs), and Multilayer Perceptron (MLP) based on dictionary-based features for untagged comments on cryptocurrencies shared in tr.investing.com, Twitter.com, and Telegram between April 7, 2024, and May 6, 2024. Most cryptocurrencies experienced a decrease in their closing prices, indicating a potential bearish trend during this period. However, Tether (USDT) and Binance (BNB) saw slight increases. The sentiment analysis aimed to investigate potential correlations between sentiment predictions and price movements, providing insights into the relationship between public sentiment and market trends. Table 2 shows the opening and closing prices of various cryptocurrencies.

Table 2. Opening and closing prices of various cryptocurrencies between April 7, 2024, and May 6, 2024.

	Cryptocurrencies	Opening (07.04.2024)	Closing (06.05.2024)
1	Bitcoin (BTC)	69.001 $	63.172 $
2	Ethereum (ETH)	3.362,84 $	3.064,59 $
3	Ripple (XRP)	0,593399 $	0,540402 $
4	Tether (USDT)	0,998835 $	0,999814 $
5	Solana (SOL)	178,96 $	152,65 $
6	Litecoin (LTC)	101,24 $	80,74 $
7	Cardano (ADA)	0,583757 $	0,454169 $
8	Dogecoin (DOGE)	0,185693 $	0,156536 $
9	Binance (BNB)	585,39 $	588,00 $
10	Tron (TRX)	0,119514 $	0,118609 $

4.2. Results and Discussions

A sentiment analysis model for Turkish texts was developed by integrating dictionary-based methods and machine-learning classifiers. A corpus was created from posts shared on the "Binance Global Türkçe" channel on Telegram between March 20, 2023, and January 4, 2024, to be used as input for the models. Orange Program was used to analyze the data from Telegram. After connecting the corpus with the file icon, text preprocessing, including transformation, tokenization, normalization, and filtering, was performed.

Using the multilingual sentiment technique in the Orange, the sentiment analysis of comments classified the sentiment scores as positive, negative, or neutral. The data were then exported as an Excel file. A total of 607,592 comments were analyzed, with 89,986 classified as negative, 72,655 as positive, and 444,951 as neutral. For binary classification, 89,986 negative and 72,655 positive samples were selected, and machine-learning models were trained and tested on 162,641 samples. The 80-20 rule was applied for the training-test split, where 80% of the data was allocated for training, and the remaining 20% was divided equally between testing (10%) and validation (10%). The study's results evaluated different machine learning classifiers. Naive Bayes (NB) achieved an average accuracy of 90% in the negative class and 88% in the positive class, while Logistic Regression (LR) showed high accuracy across all classes and performed well in both classes. K-Nearest Neighbors (KNN) demonstrated accuracy close to NB but with slightly lower sensitivity in the positive class. Decision Trees (DTs) provided balanced performance with 93% accuracy, while Gradient Boosting (GB) showed lower performance with 84% accuracy. Multilayer Perceptron (MLP) exhibited a high accuracy of 94% across all classes. Overall, LR and MLP classifiers demonstrated higher accuracy and balanced performance with 94%, with AUC values of 0.9380 and 0.9364, and AUPR values of 0.9509 and 0.9488, respectively.

After completing machine learning classification algorithms, sentiment analysis for untagged comments on cryptocurrency shares from tr.investing.com, Twitter.com, and Telegram between April 7, 2024, and May 6, 2024, was conducted using six supervised machine learning models (NB, LR, GB, KNN, DTs, and MLP). Sentiment prediction results for each cryptocurrency were presented in tables, comparing the models' negative or positive sentiment predictions. Potential correlations between positive sentiment predictions and price increases, and negative sentiment predictions and price decreases, were examined. Model performance was evaluated to determine which model provided more consistent or accurate predictions.

The results of sentiment prediction for each cryptocurrency are presented in Table 3. Each cell in the tables represents the number of samples predicted as negative (Neg.) or positive (Pos.) sentiment by a specific model, reflecting the model's confidence in its prediction. Higher numbers generally indicate a stronger prediction.

By evaluating the performance of these models, it was aimed to determine which model provided more consistent or accurate sentiment predictions for cryptocurrencies. If models predict positive sentiment and the closing price of a cryptocurrency is higher than the opening price, it suggests a potential correlation between positive sentiment and price increase. Conversely, if models predict negative sentiment and the closing price is lower than the opening price, it indicates a potential correlation between negative sentiment and price decrease.

Table 3. Monthly Sentiment Predictions for Cryptocurrencies Using Machine Learning Approaches

Currencies	Data Sources	# of Comments	Sentiment Predictions Based on Binary Classification											
			NB		LR		KNN		DTs		GB		MLP	
			Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.
BTC	a	7688	4770	2918	5581	2107	5228	2460	4590	3098	6547	1141	4898	2790
	b	5569	2489	3080	3611	1958	3837	1732	3284	2285	4764	805	3173	2396

	c	7403	4193	3210	5627	1776	5247	2156	4847	2556	6524	879	4650	2753
ETH	a	811	499	312	544	267	509	302	483	328	692	119	480	331
	b	2086	983	1103	1169	917	1263	823	1135	951	1776	310	1170	916
	c	1117	679	438	758	359	662	455	674	443	987	130	671	446
XRP	a	5011	3135	1876	3512	1499	3126	1885	3255	1756	4391	620	2980	2031
	b	623	355	268	411	212	401	222	390	233	539	84	373	250
	c	347	257	90	261	86	209	138	222	125	311	36	181	166
USDT	a	219	115	104	141	78	132	87	138	81	186	33	122	97
	b	336	153	183	191	145	180	156	182	154	278	58	168	168
	c	855	614	241	649	206	442	413	452	403	772	83	493	362
SOL	a	844	465	379	568	276	523	321	523	321	736	108	489	355
	b	427	235	192	265	162	270	157	267	160	367	60	263	164
	c	585	419	166	436	149	354	231	382	203	514	71	398	187
LTC	a	188	107	81	118	70	117	71	118	70	151	37	115	73
	b	76	41	35	57	19	41	35	47	29	70	76	48	28
	c	591	277	314	414	177	399	192	396	195	520	71	398	193
ADA	a	290	162	128	183	107	181	109	167	123	240	50	160	130
	b	69	37	32	44	25	45	24	45	24	56	13	47	22
	c	61	53	8	53	8	51	10	53	8	55	6	48	13
DOGE	a	1945	1103	842	1217	728	1183	762	1177	768	1687	258	1071	874
	b	501	228	273	270	231	298	203	278	223	448	53	256	245
	c	348	117	231	162	186	203	145	176	172	306	42	154	194
BNB	a	22	11	11	12	10	13	9	12	10	18	4	11	11
	b	429	206	223	271	158	287	142	266	163	374	55	266	163
	c	972	545	427	716	256	677	295	681	291	874	98	592	380
TRX	a	70	42	28	38	32	41	29	43	27	62	8	36	34
	b	15	6	9	9	6	11	4	12	3	13	2	13	2
	c	132	68	64	85	47	80	52	95	37	120	12	72	60

a) Investing.com b) Twitter.com (X) c) Telegram

The general trends observed in sentiment prediction show that the Gradient Boosting (GB) model generally predicts the highest number of negative sentiments across all platforms and cryptocurrencies. In contrast, Naive Bayes (NB) and Logistic Regression (LR) models tend to predict higher positive emotions in some situations. When examined specifically on the platform, it is seen that Investing.com generally predicts more negative sentiment, Twitter shows a more balanced sentiment distribution, and some models predict higher positive sentiment. Telegram follows Investing.com's trend with more negative views. In terms of model performance, Gradient Boosting (GB) often appears to predict the highest number of negative emotions. GB has very high sensitivity (99%) and low precision (78%) in the negative class, indicating that GB classifies many samples belonging to the positive class as negative. On the other hand, Naive Bayes (NB) and Logistic Regression (LR) generally predict higher positive emotions compared to other models.

Sentiment analysis of cryptocurrencies from various platforms shows notable trends. BTC (Bitcoin) generally shows more negative sentiment on Investing.com and Telegram, while on Twitter the NB model shows more positive sentiment. ETH (Ethereum) exhibits higher negative sentiment on Investing.com and Telegram; While NB models make the most negative predictions, Twitter shows a positive trend in LR models. XRP (Ripple) shows dominant negative sentiment across all platforms. USDT (Tether) has more negative views on Investing.com and Telegram, while Twitter, NB, and KNN show more positive views on their models. SOL (Solana), LTC (Litecoin), and ADA (Cardano) exhibit dominant negative sentiment views. DOGE (Dogecoin) has strong negative sentiment across all platforms. BNB (Binance Coin) is showing mixed sentiment on Investing.com but a negative outlook on Twitter and Telegram. Similarly, TRX (Tron) offers mixed views on Investing.com but generally predicts more negative sentiment on Twitter and Telegram.

5. CONCLUSION

This study seeks to enrich existing literature by examining the sentiment expressed in discussions on social media platforms and online forums regarding popular cryptocurrencies listed on Turkish exchanges. As price predictions are crucial for securities markets in developing economies, they are even more critical in cryptocurrency markets. Speculative movements in cryptocurrency markets cause high volatility, making it difficult for investors to predict price movements, often leading to significant losses [74,75].

Investors can make better investment decisions if they can predict the returns of different cryptocurrencies under uncertain and risky conditions. Such foresight helps investors prepare for market fluctuations and minimize potential losses, enabling better portfolio management and asset allocation. Given the high volatility and uncertainty in cryptocurrency markets compared to traditional investment tools, market analysis supported by technical and fundamental analysis methods is essential. Therefore, sentiment analysis using natural language processing (NLP) and artificial intelligence approaches was conducted to detect emotions expressed in various sources. Sentiment analysis is a widely utilized method for computer-based analysis of opinions, perceptions, and emotions expressed in textual content. [3].

The study focuses on social media platforms and online forums, where individuals can quickly express their opinions and provide valuable insights into public reactions on various topics. Sentiment analysis was utilized to understand these opinions and emotions conveyed through texts and highlight their relationship with specific topics of interest. This analysis not only helps investors gain insights into market trends but also quantitatively evaluates public attention and interest through the volume of shared comments.

This study emphasizes the importance of expanding analysis by integrating data from diverse time periods and gathering insights from various sources. It was noted that the dominant user base on the social media platforms and online forums studied consisted of Turkish-speaking individuals residing in Turkey, highlighting the need for future analyses to consider geographic and linguistic factors in profiling the investors examined. To improve sentiment analysis accuracy, the study recommends optimizing parameters in a larger and more comprehensive dataset used for training machine learning models. Enhancing preprocessing steps for data obtained from platforms can lead to cleaner datasets, increasing the reliability and robustness of analysis results.

While there are studies using sentiment analysis methods on Turkish texts in the literature [4–6,15–17,19–21], sentiment analysis studies commonly focus on English texts and often utilize short-form sources like tweets due to character limitations [1,3,8–14]. These studies typically use models categorizing sentiments as positive/negative polarity. In contrast, our study stands out by using Turkish texts and integrating data from multiple platform types, addressing a gap in existing research focused primarily on English and Twitter data. Additionally, while existing studies often focus on major cryptocurrencies like Bitcoin or popular alternatives known as altcoins [76–78], our research aims to broaden this scope by examining cryptocurrencies listed on Turkish exchanges to investigate potential characteristic differences. Our study aims to bring new perspectives to the literature by analyzing sentiments expressed on social media platforms and online forums.

The study acknowledges various limitations, such as the intricate dynamics of the market and external factors, aiming to improve transparency. The primary data source focuses on sentiments expressed through posts related to financial matters on social media platforms and online forums. While the study provides insights into public opinion, it recognizes that it may not encompass all relevant information or opinions from other sources. It underscores the need for more accurate sentiment analysis methods and suggests further investigation into the impact of economic events and global trends on cryptocurrency prices. The study proposes exploring advanced natural language processing techniques and context-aware sentiment analysis models to improve sentiment analysis of financial posts. It also recommends incorporating multimodal data such as emojis, images, and links in posts to deepen sentiment analysis insights. Future research could explore how multimedia content affects public sentiment and subsequently market behavior across various social media platforms. By addressing these areas, the study aims to enhance understanding of the complex interplay between social media sentiment and market dynamics and pave the way for more sophisticated sentiment analysis methodologies in financial contexts.

DECLARATION OF ETHICAL STANDARDS

The authors of this article declare that the materials and methods used in their studies do not require ethical committee approval and/or legal-specific permission.

AUTHORS' CONTRIBUTIONS

Uğur DEMİREL: Wrote the manuscript. Performed the experiments and analysis the results.

Handan ÇAM: Performed the experiments and analysis the results.

CONFLICT OF INTEREST

There is no conflict of interest in this study.

REFERENCES

[1] Go A., Huang L., Bhayani R., "Twitter sentiment analysis", ***Entropy***, 17:252, (2009).

[2] Appel O., Chiclana F., Carter J., Fujita H., "A hybrid approach to sentiment analysis", ***2016 IEEE Congress on Evolutionary Computation (CEC), IEEE***, 4950–7, (2016).

[3] Pang B., Lee L., "Opinion mining and sentiment analysis", ***Foundations and Trends® in Information Retrieval***, 2:1–135, (2008).

[4] Akba F., Uçan A., Sezer E.A., Sever H., "Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews", ***8th European Conference on Data Mining***, 191:180–4, (2014).

[5] Catal C., Nangir M., "A sentiment classification model based on multiple classifiers", ***Applied Soft Computing***, 50:135–41, (2017).

[6] Çoban Ö., Özyer B., Özyer G.T., "Sentiment analysis for Turkish Twitter feeds", ***2015 23nd Signal Processing and Communications Applications Conference (SIU), IEEE***, 2388–91 (2015).

[7] ***www.octoparse.com***, "Octoparse", (2024).

[8] Pang B., Lee L., V"aithyanathan S. Thumbs up? Sentiment classification using machine learning techniques", ***arXiv Preprint Cs/0205070***, (2002).

[9] Al-Hadhrami S., Al-Fassam N., Benhidour H., "Sentiment analysis of english tweets: A comparative study of supervised and unsupervised approaches", ***2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), IEEE***, 1–5, (2019).

[10] Desai R.D., "Sentiment analysis of Twitter data", ***2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE***, p. 114–7, (2018).

[11] El Rahman S.A., AlOtaibi F.A., AlShehri W.A., "Sentiment analysis of twitter data", ***2019 international conference on computer and information sciences (ICCIS), IEEE***; 1–4, (2019).

[12] Kurniawati I., Pardede H.F., "Hybrid method of information gain and particle swarm optimization for selection of features of SVM-based sentiment analysis", ***2018 International Conference on Information Technology Systems and Innovation (ICITSI), IEEE***,1–5, (2018).

[13] Rane A., Kumar A., "Sentiment classification system of Twitter data for US airline service analysis", ***2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), IEEE***; 1:769–73, (2018).

[14] Ray P., Chakrabarti A., "A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis", ***Applied Computing and Informatics*** (2020).

[15] Kaya M., Fidan G., Toroslu I.H., "Sentiment analysis of Turkish political news", ***2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology***, ***IEEE***; 1:174–80, (2012).

[16] Nizam H., Akın S.S., "Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması", ***XIX Türkiye'de İnternet Konferansı***, (2014).

[17] Meral M., Diri B., "Sentiment analysis on Twitter", ***2014 22nd Signal Processing and Communications Applications Conference (SIU), IEEE***, 690–3, (2014).

[18] Kaynar O., Görmez Y., Yıldız M., Albayrak A., "Makine öğrenmesi yöntemleri ile Duygu Analizi", ***International Artificial Intelligence and Data Processing Symposium (IDAP'16)***, 234: 241 (2016).

[19] Parlar T., Saraç E., Özel S.A., "Comparison of feature selection methods for sentiment analysis on Turkish Twitter data", ***2017 25th Signal Processing and Communications Applications Conference (SIU), IEEE***, 1–4, (2017).

[20] Velioğlu R., Yıldız T., Yıldırım S., "Sentiment analysis using learning approaches over emojis for Turkish tweets", ***2018 3rd International Conference on Computer Science and Engineering (UBMK), IEEE***, 303–7, (2018).

[21] Parlar T., Özel S.A., Song F., "QER: a new feature selection method for sentiment analysis", ***Human-Centric Computing and Information Sciences***, 8:1–19, (2018).

[22] Ciftci B., Apaydin M.S., "A deep learning approach to sentiment analysis in Turkish", ***2018 International Conference on Artificial Intelligence and Data Processing (IDAP), IEEE***, 1–5, (2018).

[23] Cliche M., "BB_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs", ***arXiv Preprint arXiv:170406125***, (2017).

[24] Hassan A., Mahmood A., "Deep learning approach for sentiment analysis of short texts", ***2017 3rd international conference on control, automation and robotics (ICCAR), IEEE***, 705–10, (2017).

[25] Kamis S., Goularas D., "Evaluation of deep learning techniques in sentiment analysis from twitter data", ***2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), IEEE***; 12–7, (2019).

[26] Demirci G.M., Keskin Ş.R., Doğan G., "Sentiment analysis in Turkish with deep learning", ***2019 IEEE international conference on big data (big data), IEEE***, 2215–21, (2019).

[27] Acikalin U.U., Bardak B., Kutlu M., "Turkish sentiment analysis using bert", ***2020 28th Signal Processing and Communications Applications Conference (SIU), IEEE***, 1–4, (2020).

[28] Shehu H.A., Tokat S., "A hybrid approach for the sentiment analysis of Turkish Twitter data", ***The International Conference on Artificial Intelligence and Applied Mathematics in Engineering, Springer***, 182–90, (2020).

[29] Esuli A., Sebastiani F., "Sentiwordnet: A publicly available lexical resource for opinion mining", ***Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)***, (2006).

[30] Baloglu A., Aktas M.S., "An automated framework for mining reviews from blogosphere", ***International Journal of Advances in Internet Technology***, 3:234–44, (2010).

[31] Bilgin O., Çetinoğlu Ö., Oflazer K., "Building a wordnet for Turkish", ***Romanian Journal of Information Science and Technology***, 7:163–72, (2004).

[32] Dehkharghani R., Saygin Y., Yanikoglu B., Oflazer K., "SentiTurkNet: a Turkish polarity lexicon for sentiment analysis", ***Language Resources and Evaluation***, 50:667–85, (2016).

[33] Akgül E.S., Ertano C., Banu D., "Sentiment analysis with Twitter", ***Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi***, 22:106–10, (2016).

[34] Öztürk N., Ayvaz S., "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis", ***Telematics and Informatics***, 35:136–47, (2018).

[35] Yurtalan G., Koyuncu M., Turhan Ç., "A polarity calculation approach for lexicon-based Turkish sentiment analysis", ***Turkish Journal of Electrical Engineering & Computer Sciences***, 27:1325–39, (2019).

[36] Ohana B., Tierney B., "Sentiment classification of reviews using SentiWordNet", ***Proceedings of IT&T***, 8, (2009).

[37] Govindarajan M., "Sentiment analysis of movie reviews using hybrid method of naive bayes and genetic algorithm", ***International Journal of Advanced Computer Research***, 3:139, (2013).

[38] Türkmenoglu C., Tantug A.C., "Sentiment analysis in Turkish media", ***International Conference on Machine Learning (ICML)***, (2014).

[39] Rumelli M., Akkuş D., Kart Ö., Isik Z., "Sentiment analysis in Turkish text with machine learning algorithms", ***2019 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE***; 1–5, (2019).

[40] Uslu A., Tekin S., Aytekin T., "Sentiment analysis in Turkish film comments", ***2019 27th Signal Processing and Communications Applications Conference (SIU), IEEE***, 1–4, (2019).

[41] Erşahin B., Aktaş Ö., Kilinc D., Erşahin M., "A hybrid sentiment analysis method for Turkish", ***Turkish Journal of Electrical Engineering & Computer Sciences***, 27:1780–93, (2019).

[42] Bollen J., Mao H., Zeng X., "Twitter mood predicts the stock market", ***Journal of Computational Science***, 2:1–8, (2011).

[43] Zhang X., Fuehres H., Gloor P.A., "Predicting stock market indicators through twitter "I hope it is not as bad as I fear"", ***Procedia-Social and Behavioral Sciences***, 26:55–62, (2011).

[44] Smailović J., Grčar M., Žnidaršič M., Lavrač N., "Sentiment analysis on tweets in a financial domain", ***4th Jožef Stefan International Postgraduate School Students Conference***, 1:169–75, (2012).

[45] Kristoufek L., "BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era", ***Scientific Reports***, 3:1–7, (2013).

[46] Wang G., Wang T., Wang B., Sambasivan D., Zhang Z., Zheng H., et al. "Crowds on wall street: Extracting value from social investing platforms", ***arXiv Preprint arXiv:14061137***, (2014).

[47] Kaminski J., "Nowcasting the bitcoin market with twitter signals", ***arXiv Preprint arXiv:14067577***, (2014).

[48] Hernandez I., Bashir M., Jeon G., Bohr J., "Are Bitcoin Users Less Sociable? An analysis of users' language and social connections on twitter", ***International Conference on Human-Computer Interaction, Springer***, 26–31, (2014).

[49] Garcia D., Tessone C.J., Mavrodiev P., Perony N., "The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy", ***Journal of the Royal Society Interface***, 11, (2014).

[50] Garcia D., Schweitzer F., "Social signals and algorithmic trading of Bitcoin", ***Royal Society Open Science***, 2, (2015).

[51] Colianni S., Rosales S., Signorotti M., "Algorithmic trading of cryptocurrency based on Twitter sentiment analysis", ***CS229 Project***, 1, (2015).

[52] ***www.orangedatamining.com/widget-catalog/text-mining/preprocesstext/***, "Preprocess Text", (2022).

[53] ***www.orangedatamining.com/widget-catalog/text-mining/sentimentanalysis/***, "Sentiment Analysis", (2022).

[54] ***www.help.ku.edu.tr/ithelp/jupyter***, "Jupyter", (2024).

[55] Çalış K., Gazdağı O., Yıldız O., "Reklam içerikli epostaların metin madenciliği yöntemleri ile otomatik tespiti", ***Bilişim Teknolojileri Dergisi***, 6:1–7, (2013).

[56] Onan A., "Twitter mesajları üzerinde makine öğrenmesi yöntemlerine dayalı duygu analizi", ***Yönetim Bilişim Sistemleri Dergisi***, 3:1–14, (2017).

[57] Akpınar H., "Data: veri madenciliği veri analizi", ***2. basım. İstanbul: Papatya Yayıncılık Eğitim***, (2017).

[58] Osmanoğlu U.Ö, Atak O.N, Çağlar K., Kayhan H., Can T.C., "Sentiment analysis for distance education course materials: A machine learning approach", ***Journal of Educational Technology and Online Learning***, 3:31–48, (2020).

[59] Taşcı E., Onan A., "K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi", ***Akademik Bilişim***, 1:4–18, (2016).

[60] Çelik Ö., Aslan A.F., "Gender prediction from social media comments with artificial intelligence", ***Sakarya University Journal of Science***, 23:1256–64, (2019).

[61] Han J., Kamber M., Pei J., "Data mining: concepts and techniques", ***Morgan kaufmann***, (2011).

[62] Özekes S., "Veri madenciliği modelleri ve uygulama alanları", ***İstanbul Commerce University Journal of Science***, 3:3, (2003).

[63] Coşlu E., "Veri madenciliği", ***Akademik Bilişim***, 23–5, (2013).

[64] Soydaş S.S., Çam H., "Predicting Financial Failure in Companies by Employing Machine Learning Methods", ***International Journal of Social Science Research and Review***, 7:111–25, (2024).

[65] Medhat W., Hassan A., Korashy H., "Sentiment analysis algorithms and applications: A survey", ***Ain Shams Engineering Journal***, 5:1093–113, (2014).

[66] Si S., Zhang H., Keerthi S.S., Mahajan D., Dhillon I.S., Hsieh C-J., "Gradient boosted decision trees for high dimensional sparse output", ***International conference on machine learning, PMLR***, 3182–90, (2017).

[67] He Z., Lin D., Lau T., Wu M., "Gradient boosting machine: a survey", ***arXiv Preprint arXiv:190806951***, (2019).

[68] Hang H., Huang T., Cai Y., Yang H., Lin Z., "Gradient Boosted Binary Histogram Ensemble for Large-scale Regression", (2021).

[69] Friedman J.H., "Greedy function approximation: a gradient boosting machine", ***Annals of Statistics***, 1189–232, (2001).

[70] Hamzaçebi C., "Yapay sinir ağları: tahmin amaçlı kullanımı MATLAB ve Neurosolutions uygulamalı", ***Ekin Basım Yayın Dağıtım***, (2011).

[71] Alpaydin E., "Introduction to machine learning", ***Cambridge, Mass., MIT Press***, (2010).

[72] Öztemel E., "Yapay sinir ağlari", ***PapatyaYayincilik, Istanbul***, (2012).

[73] Sokolova M., Lapalme G., "A systematic analysis of performance measures for classification tasks", ***Information Processing & Management***, 45:427–37, (2009).

[74] Corbet S., Lucey B., Urquhart A., Yarovaya L., "Cryptocurrencies as a financial asset: A systematic analysis", ***International Review of Financial Analysis***, 62:182–99, (2019).

[75] Feng W., Wang Y., Zhang Z., "Informed trading in the Bitcoin market", ***Finance Research Letters***, 26:63–70, (2018).

[76] Erdinç U., Bursa N., "Covıd-19 Pandemi Sürecinde Twitter Yorumları İle Altcoın Kripto Para Piyasası Arasındaki Nedenselliğin Duygu Analizi İle İncelenmesi: Ripple Örneği", ***Yönetim ve Ekonomi Araştırmaları Dergisi***, 19:362–81, (2021).

[77] Kaplan C., Aslan C., Bulbul A., "Cryptocurrency Word-of-Mouth Analysis via Twitter", ***ResearchGate***, (2018).

[78] Köksal B., Erdem G., Türkeli C., Öztürk Z.K., "Twitter'da Duygu Analizi Yöntemi Kullanılarak Bitcoin Değer Tahminlemesi", ***Düzce Üniversitesi Bilim ve Teknoloji Dergisi***, 9:280–97, (2021).