Journal of Engineering Faculty, 2(2): 205-213, 2024

Journal of Engineering Faculty

| cumfad.cumhuriyet.edu.tr | Founded: 2023 Available online, ISSN Publisher: Sivas Cumhuriyet Üniversitesi

Predicting Bitcoin Price Direction Using Machine Learning Models

Tanju AÇI[1,a,*], Hakan Kekül[2,b]

[1] *Department of Computer Engineering, Institute of Science and Technology, Sivas Cumhuriyet University, Sivas, Turkey*
[2] *Department of Software Engineering, Faculty of Technology, Sivas Cumhuriyet University, Sivas, Turkey*
**Corresponding author*

Research Article

History

Received: 20/11/2024
Accepted: 10/12/2024

ABSTRACT

In the financial sector, as past economic and social events have shaken trust, this trust is being regained through the internet and computer technologies. Emerging in the 19th century, financial technology has led to a new economic understanding with digital money and especially bitcoin. The decentralized structure of bitcoin and the encryption systems used for security play an important role in preventing fraud and have become the center of attention of investors. As its value has increased, studies on price predictions have naturally increased. This study aims to predict the impact of data obtained from digital economy news sites on bitcoin price using natural language processing and machine learning techniques. In line with this goal, text vectorization was performed with the TF-IDF statistical method. Synthetic Minority Oversampling Technique (SMOTE) was applied to eliminate the imbalance in the vectorized data set. Classification models such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, K-Nearest Neighbor, Extra Trees, Bernoulli Naive Bayes and Multilayer Perceptron were applied to the obtained output.

According to the results of the performance of different machine learning models in predicting the direction of bitcoin price fluctuation, the Extra Trees Classifier model showed the highest performance with an Accuracy of 86.71%, recall of 86.71%, precision of 86.99% and F1 score of 86.59%.

Keywords: Artificial Intelligence, Bitcoin, Machine Learning, Natural Language Processing, Stock Market, Crypto

Makine Öğrenimi Modelleri Kullanarak Bitcoin Fiyat Yönünü Tahmin Etme

ÖZ

Finans sektöründe, geçmişte yaşanan ekonomik ve sosyal olaylar güveni sarsarken, internet ve bilgisayar teknolojileri sayesinde bu güven yeniden kazanılıyor. 19. yüzyılda ortaya çıkan finansal teknoloji, dijital para ve özellikle bitcoin ile yeni bir ekonomik anlayışa yol açmıştır. Bitcoin'in merkezi olmayan yapısı ve güvenlik için kullanılan şifreleme sistemleri dolandırıcılığın önlenmesinde önemli rol oynamış ve yatırımcıların ilgi odağı haline gelmiştir. Değeri arttıkça fiyat tahminleri üzerine yapılan çalışmalar da doğal olarak artmıştır. Bu çalışma, dijital ekonomi haber sitelerinden elde edilen verilerin bitcoin fiyatı üzerindeki etkisini doğal dil işleme ve makine öğrenmesi tekniklerini kullanarak tahmin etmeyi amaçlamaktadır. Bu hedef doğrultusunda TF-IDF istatistiksel yöntemi ile metin vektörleştirmesi yapılmıştır. Vektörleştirilen veri setindeki dengesizliği gidermek için Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) uygulandı. Elde edilen çıktılara Lojistik Regresyon, Karar Ağaçları, Rastgele Orman, Destek Vektör Makineleri, K-En Yakın Komşu, Ekstra Ağaçlar, Bernoulli Naive Bayes ve Çok Katmanlı Perceptron gibi sınıflandırma modelleri uygulanmıştır.

Farklı makine öğrenimi modellerinin bitcoin fiyat dalgalanmasının yönünü tahmin etmedeki performans sonuçlarına göre, Ekstra Ağaçlar Sınıflandırıcı modeli %86,71 Doğruluk, %86,71 hatırlama, %86,99 kesinlik ve %86,59 F1 puanı ile en yüksek performansı göstermiştir.

Anahtar Kelimeler: Yapay Zeka, Bitcoin, Makine Öğrenimi, Doğal Dil İşleme, Borsa, Kripto

[a] *20239257006@cumhuriyet.edu.tr* *ORCID: 0009-0008-5822-3393* | [b] *hakankekul@cumhuriyet.edu.tr* *ORCID: 0000-0001-6269-8713*

How to Cite: Açı T, Kekül H (2024) Predicting Bitcoin Price Direction Using Machine Learning Models, Journal of Engineering Faculty, 2(2): 205-213

Introduction

Economic events and phenomena in the past have from time to time damaged the economy and the financial system and caused a loss of trust. This damaged sense of trust is intended to be regained with the help of internet and computer technologies, which are among the most important developments of today. Originating in the 19th century, financial technology, which means technological innovation in the field of finance, is the world's newest economic understanding.While trying to regain the damaged sense of trust with this understanding, digital money bitcoin, which has a system that will deeply shake the foundations of the economy, has emerged [1].

The idea of digital money is relatively new, but it has not been successfully implemented until recently[2]. The pioneer of cryptocurrencies within digital currencies is bitcoin. It is a decentralized system. That is, it is a digital currency that is not managed by a bank or government. Transactions are tracked and verified through a security system called encryption. In this way, negative situations such as fraud and fraud are prevented [3].

While the value of bitcoin, which is taking strong steps forward in the economic sector, is increasing day by day, the number of studies on future price prediction has also increased. In addition to being a source of information for a society, it has been determined that newspaper news, from which economic perception is obtained, has an impact on price forecasts [1]. With the development of the internet and technology, news and newspapers have been moved to digital platforms, increasing the ease and speed of access.

Basically, in this study, the effect of the texts obtained from digital economy news sites on bitcoin price is tried to be predicted by using machine learning models with natural language processing techniques.

The study seeks to answer the following research questions.

R.Q. 1: Does news have an impact on predicting the direction of financial markets?

R.Q. 2: Can natural language processing and machine learning algorithms predict the direction of financial market instruments?

After this stage, the study is organized as follows. Section 2 presents a detailed analysis of the literature. The third section presents the proposed methodology. The fourth section presents the findings obtained. The study concludes with a discussion of the implications of the findings.

Literature Review

In recent years, when the demand for crypto assets has increased considerably, accurately predicting the cause of the asset's price movements and the direction of the price has become an important issue as it will realize the aim of reducing the risk of investors and providing more return. Although many studies have been conducted on this problem, its history is not old [4].

The effects of developments in computer and internet technologies on the economy have been previously analyzed in the literature. Over a period of 9 years, all news articles with positive or negative content published in 5 different newspapers were analyzed and associated with blockchain technology data. With the help of artificial neural network learning algorithm, the effects of news in national financial newspapers on bitcoin price are investigated. As a result, it is found that bitcoin news published in financial newspapers have a partial effect on price prediction. The study has become a reference for understanding the effects of Fintech innovations and artificial intelligence methods on the economy [1].

Some researchers believe that bitcoin and other cryptocurrencies could disrupt the way the traditional economy works. This is because blockchain technology completely eliminates the link between the financial authority and the banking sector. On the other hand, low-cost transactions, limited supply, preference against fluctuations in national currencies, and the fact that it can be easily transferred to very distant borders have made bitcoin an important stakeholder of the economy in recent years[5]. As of October 2024, the supply of bitcoin circulating in the market is 19,774,818. Its maximum supply is 21,000,000 and its approximate value on 29.10.2024 is 72,800 dollars. The market capitalization is $1,441,611,379,618 [6].

In a study on bitcoin price prediction, a correlation study was conducted on whether the price is correlated with the volume of tweets containing the bitcoin hashtag on twitter social media and web search media results. It was aimed to determine whether there is any relationship between the sentiment of tweets and the queries made by users on Google search engine and bitcoin price. In the first 3 months of 2015, the cross-correlation value between tweets, Google trends data on the same dates and bitcoin price was 0.64. This result confirms that tweet volumes and Google trends query volumes can predict bitcoin price fluctuations. As seen from the artificial neural network models and correlation results used in the studies, there was a positive relationship between the sentiment of tweets and bitcoin price[7], [8], [9], [10].

While econometric methods were frequently used in the past for stock market price prediction, machine learning methods, which have proven to be significantly more successful in recent years, are preferred. For this reason, the use of classical econometric methods has significantly decreased [11].

In studies measuring the reaction of markets to financial developments, it has been observed that financial markets react strongly to news about the legal status of cryptocurrencies. In addition to general bans on the use of bitcoin in financial transactions, markets were negatively affected by news that explicitly stated that cryptocurrencies would not be treated as currencies. On the other hand, news about restrictions on the interoperability of cryptocurrencies with the financial system in the fight against terrorism and money laundering also negatively affects bitcoin and the crypto market[12].

There have also been studies that propose a hybrid model that combines sentiment analysis and machine learning to predict Bitcoin's price movements. Researchers have analyzed data from news articles and social posts with deep learning techniques and machine learning algorithms and achieved successful results[13], [14], [15], [16], [17].

In addition to social media platforms such as twitter, reddit, forums, etc., there are also studies on predicting the bitcoin price direction. Successful results have been achieved with the advantage of high compatibility between Google searches and reddit searches[18]. It has also been concluded that what is written in forums has a greater impact on bitcoin[19].

The widespread use of Bitcoin and the increasing speed of internet use in instant communication brings with it the necessity to predict future changes in finance. Determining the direction of Bitcoin's price fluctuations will become very important. In this study, which can be used as a guide for investors, text classification was performed using natural language processing methods and machine learning models on a dataset of news and headlines from five different international economic news sites and the direction of bitcoin price fluctuations was tried to be predicted.

Methodology

In recent years, text classification methods have been used to predict the direction of bitcoin price fluctuations. In this study, classification models such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, Extra Tree Classifier, Multilayer Perceptron Classifier, Bernoulli Naive Bayes and K-Nearest Neighbor were applied.

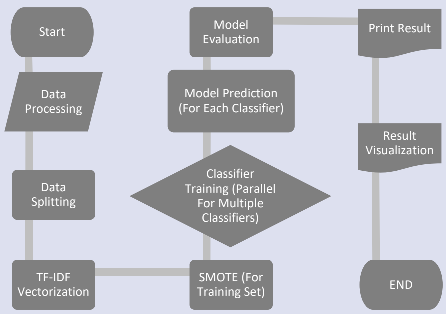

Figure 1. Model diagram

Logistic Regression

Although the Logistic Regression model resembles a model used for regression analysis, it is a supervised learning algorithm used for classification problems[20], [21].

Logistic regression calculates the probability of the dependent variable taking the expected value using independent variables. The Sigmoid Function in Figure 2 is used for classification. It compresses the output value between 0 and 1. It may not give good accuracy when the training data is small[22].

Mathematical representation of Sigmoid Function:

$$S(x) = \frac{1}{1+e^{-ax}} \quad (1)$$

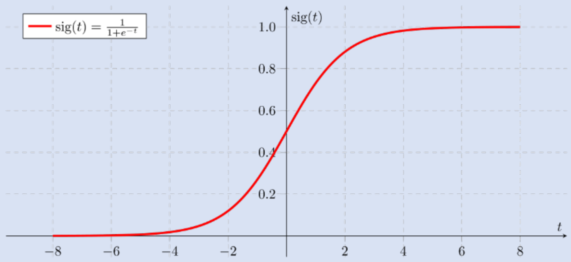

Figure 2. Sigmoid function graph

Decision Trees

Decision trees are a type of machine learning model that helps identify patterns in data. In this model, decision nodes and leaf nodes are created based on the training data. The algorithm divides the dataset into small subsets based on different criteria. By drawing optimal lines, it divides the data into clusters that will move the data to the maximum data point. It is suitable for complex data sets and can overlearn if the maximum depth is not set well. As the decision tree gets deeper, it becomes more complex to compute. It often requires longer training times. It is insufficient in regression applications [20].

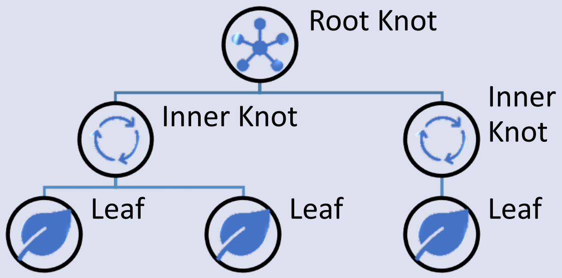

Figure 3. Decision Tree diagram

Random Forest

The Random Forest model is a machine learning method used in both classification and regression problems. It has a working logic similar to the structure of Decision Trees. The data set is divided into random parts and decision trees are created. In the prediction phase, the predictions of these decision trees are averaged. The reason for using the Random Forest model in this study is that it largely avoids the problem of overfitting the previous data used during training[23].

Support Vector Machines

Support Vector Machines originated in the field of statistical learning theory[24]. It produces a linear algorithm that generates a function to maximize the distance between classes. The class function is constructed using sample data on the edge of the class[25]. Creates a model that can predict class labels for a dataset with specified features. The data in the training set is expressed as a feature vector. Feature vectors are mapped to the feature space using the kernel function. A hyperplane that optimally separates the classes is created and tested [26].

K-Nearest Neighbor

K-Nearest Neighbor is a model that performs classification by using the closeness of a given feature to another feature that is closest to it[27].

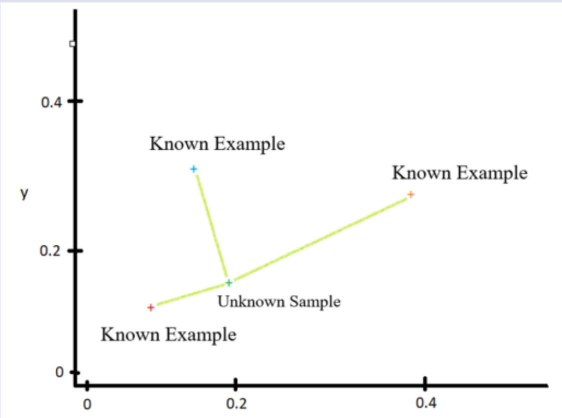

Figure 4. K-Nearest neighbor model

As shown in Figure 4, the distance of the unknown sample point (green point) to three known and close points is examined. By verifying the distance of the known red point to the green point, it is inferred that the green point is also red. Commonly used distance measures are Euclidean distance, Manhattan distance and Minkovski distance[27].

Extra Trees

It is a different version of the random forest classifier. It is a machine learning model that makes predictions using an ensemble of decision trees. Due to its similarity with random forest, it often performs similarly or better[28].

In this model, which uses a simpler algorithm, nodes branch randomly instead of using a decision criterion to make the best split. While this method reduces complexity and processing overhead, it does not perform well when analyzing large datasets with high noise. From a statistical point of view, it usually increases bias and decreases variance[29].

Multilayer Perceptron Classifier

It is one of the most widely used machine learning methods, especially in classification. The model uses a method called Delta Learning Rule. The purpose of this rule is to minimize the error between the desired output and the output produced in the network. It consists of 3 layers: input layer, hidden layer and output layer. Information from the input layer is introduced to the network, reaches the output layer through hidden layers and is transferred to the outside world. In the model using the learning-by-training method, both examples and outputs generated from these examples are given to the network. The network produces solutions in the problem space using these examples[30].

Bernoulli Naive Bayes

It implements naive Bayes classification algorithms designed for the case where each of the multi-class features is a binary (Bernoulli, boolean) variable. Therefore, the classifier requires feature vectors with binary values. Given other types of data, it can convert data to binary format depending on the binarize parameter. BernoulliNB can use word presence/absence vectors instead of word count vectors in text classification. BernoulliNB can perform better than the multinomial model, especially for data sets containing shorter documents[31].

Text Pre-Processing

For the bitcoin price fluctuation direction prediction process by applying machine learning models using the texts in the text field in the dataset, the class data labeled in the dictionary structure in the sentiment column was parsed as a separate column. Columns other than the text column and class columns were deleted from the dataset. Digitization was performed on the data in the class field. The neutral label was replaced with the numerical values 0, the positive label with 1 and the negative label with 2. For model training and predictions, the dataset was divided into 80% training group and 20% test group and performance metrics were obtained by training the models.

The raw text data in the Text field has been pre-processed to make it suitable for the model.

- ✓ All data has been converted to lower case.
- ✓ Numeric values have been removed, all special characters except letters A-Z have been replaced with spaces (instead of punctuation marks).
- ✓ English words that do not make sense on their own (the,a,or,and etc.) have been separated from the data set.

TF-IDF And Smote Application

Term Frequency - Inverse Document Frequency, or TF-IDF for short, is a numerical statistic that tells how important a word is in a document collection. It allows to focus on important words. "Important" here means frequently found in one document but rarely found in others. It balances term frequency (how often a word appears in a document) and inverse document frequency (how rare and common a word is in the whole corpus)[32]. Basically, TF-IDF is the product of the term frequency (TF) of terms and the inverse document frequency (IDF). In this way, the text data is digitized and normalized.

The TF-IDF calculation formula is as follows[33].

$$TF(t, d) = \log(1 + frekans(t, d)) \tag{2}$$

$$IDF(t,D) = \log(N\ say(d \in D : t \in d) \tag{3}$$

$$TF - IDF(t, d,D) = TF(t, d) * IDF(t,D) \tag{4}$$

Here, frequency(t,d) is the number of times word t occurs in the dth document; count(d∈D:t∈d) is the number of times word t is found in the documents that make up the corpus D[33].

For the text data in the Text domain, TF-IDF statistical text vectorization was applied as a term weighting method. As can be seen in Figure 5, when the data in the Sentiment class column is analyzed, it is seen that there is an unbalanced data set.

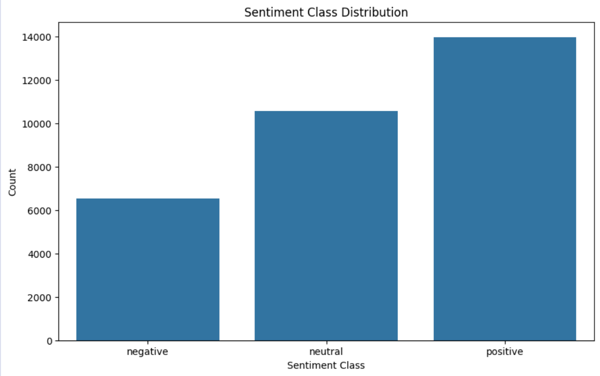

Figure 5. Sentiment class field data distribution

When there is a severe imbalance between classes, the performance of the model is adversely affected. SMOTE stands for Synthetic Minority Over-sampling Technique. It allows the training data to be made more balanced by artificially (synthetically) multiplying the samples especially on the minority class[34].

Data Set

The study used a dataset of 31037 news headlines and news texts obtained from the kaggle web application, taken from leading international cryptocurrency news sites such as cryptonews.com, cryptopotato.com and cointelegraph.com and published in English between 2021-10-12 / 2023-12-19 (Table 1 shows a cross-section of the dataset).

As can be seen in Table 2, the dataset consists of 7 columns: Date, Sentiment, Source, Subject, Text, Title and Url.

Table 1. An example from the data set

	Date	Sentiment	...	Text
1	*2023-12-19 06:40:41*	*{'class': 'negative', 'polarity': -0.1, 'subje...*	*...*	*Grayscale CEO Michael...*
2	*2023-12-19 06:03:24*	*{'class': 'neutral', 'polarity': 0.0, 'subject...*	*...*	*In an exclusive interview...*
3	*2023-12-19 05:55:14*	*{'class': 'positive', 'polarity': 0.05, 'subje...*	*...*	*According to the Federal...*
4	*2023-12-19 06:40:41*	*{'class': 'positive', 'polarity'...*	*...*	*Some suggest EVM...*

In the Source column, 3 values consisting of cryptonews, cryptopotato, cointelegraphy are observed. There are 13010 news records with cointelegraphy, 10459 with cryptonews and 7568 with cryptopotato.

In the subject column of the dataset, there are 6 different value classes: bitcoin, altcoin, blockchain, ethereum, nft, defi. There are 9968 records with bitcoin, 9278 with altcoin, 6947 with blockchain, 2274 with ethereum, 1533 with nft, and 1037 with defi.

In the text field with English news texts, the longest data consists of 513 characters and 87 words, and the shortest value consists of 26 characters and 5 words.

In the title field, which contains English news titles, the longest data consists of 254 characters, 42 words, and the shortest data consists of 1 character, 1 word.

While there are 30423 different data in the text field, all 31037 data in the url field are unique. This means that the values with the same news texts were published at different url addresses at different times.

In the sentiment column, in addition to the class section and the negative, neutral and positive impact of the bitcoin price, there are also polarity and subjectivity data. In the polarity field, the degree to which the text is positive or negative is expressed as a number between -1 and 1. -1 means completely negative, 0 means neutral and 1 means completely positive. For example, a value of -0.1 indicates that the text is slightly negative. In the Subjectivity domain, the degree of subjectivity of the text is expressed as a number between 0 and 1. 0 means completely objective (based on facts) and 1 means completely subjective (based on personal opinions). For example, a value of 0.8 indicates that the text is highly subjective. In other words, personal opinions, feelings or evaluations predominate in the text. In the study, the class data in the Sentiment field was used as target and the data in the text field was used as input.

Table 2. Data set column information

Column Name	Description
date	Date and time of the news
sentiment	Bitcoin's sensitivity to news (positive - negative - neutral)
source	Name of the news source website
subject	The subject of the content of the news item (in one word, the subject to which it relates)
text	News text
title	News headline

Results

According to the prediction results on 8 different machine learning models including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, K-Nearest Neighbor, Multilayer Perceptron Classifier, Extra Tree Classifier, Bernoulli Naive Bayes Classifier, it was observed that the Extra Tree Classifier model showed the best performance in the performance metrics obtained. The accuracy of the model was 0.8671, precision was 0.8699, F1 score was 0.8659 and recall was 0.8659. The K-Nearest Neighbor model showed the lowest success in predicting the direction of bitcoin price fluctuation. The accuracy of this model was 0.3412, precision was 0.5795, F1 score was 0.2727, and recall was 0.3412 (The classification metric values of all models are given in Table 3).

Table 3. Table of performance metrics of models

Model Name	Accuracy	Precision	F1 Score	Recall
LR	0,8241	0,8334	0,8238	0,8241
SVM	0,8194	0,8316	0,8163	0,8194
RF	0,8473	0,8535	0,8457	0,8473
DT	0,8096	0,8087	0,8091	0,8096
ETC	0,8671	0,8699	0,8659	0,8671
ML	0,7693	0,7915	0,7691	0,7693
BNB	0,7192	0,7163	0,7163	0,7192
K-NN	0,3412	0,5795	0,2727	0,3412

Precision and f1 score results are more important than other metrics as performance metrics in multiple classification models. The Extra Tree Classifier model shows that the accuracy of the predictions is high with a precision value of 86.99%. It has a balanced performance with an f1 score value of 86.59%.

The Confusion Matrix, which expresses how accurately the model predicts which classes or which classes have confusion in prediction, is presented in Figure 6. As can be seen from Figure 6, 1998 instances were correctly predicted as negative. A total of 79 examples were incorrectly predicted as negative when they should have been predicted as negative, 57 of which were neutral and 22 of which were positive. 2377 samples were correctly predicted as neutral, while 404 samples were incorrectly predicted as neutral. 998 samples were correctly predicted as positive and 352 were incorrectly predicted as positive. The most successful class in prediction is the negative class.

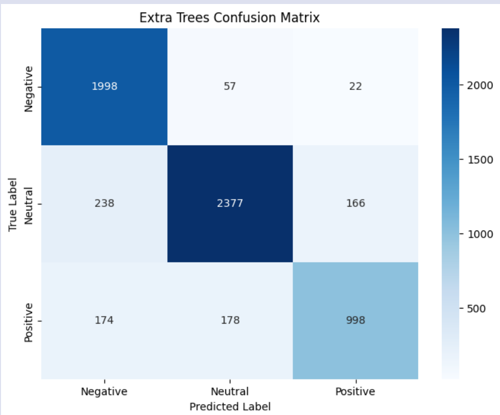

Figure 6. Extra tree classifier confusion matrix table

The ROC curve shows the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) of the model for different classes. Figure 7 also shows the AUC (Area Under Curve) value for each class. The AUC values for all classes are quite high (0.93, 0.94, and 0.95), indicating that the model is generally very successful in discriminating between the three classes. The slope of the ROC curves is quite steep and close to 1. This indicates that the model achieves high true positive rates even at low false positive rates.

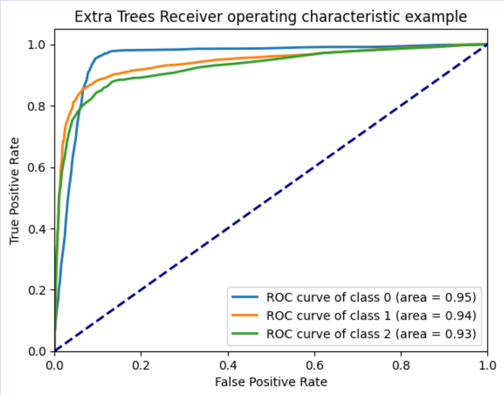

Figure 7. Extra tree classifier ROC curve plot

Conclusion and Discussion

This study aims to develop a text classification model for predicting the direction of bitcoin price fluctuation that can guide the decision-making process of investors in the crypto asset market. In the literature, the use of machine learning models in financial forecasting is increasing. This study aims to contribute to similar studies and evaluates machine learning models to predict bitcoin price frontier through texts obtained from news websites.

Extra Tree Classifier, which is the model used in the research, showed the highest classification success with performance metrics such as 86.71% accuracy rate, 86.99% precision, 86.59% F1 score and 86.71% sensitivity. According to the confusion matrix analysis, the model has a very high discriminative performance between the Negative and Neutral classes, but more misclassifications occur in the Positive class compared to the others. In the ROC curve analysis, AUC values of 0.93 and above were reached for all classes, and it was concluded that the model made a successful classification in general.

According to these results, ensemble methods such as Extra Tree Classifier can be used effectively in predicting bitcoin price direction. The findings confirm that financial markets can be influenced by text-based data, such as news content, with the help of the development of financial technology tools, and that such data can be useful in price prediction studies. By combining natural language processing techniques and machine learning models, the study can provide important layers for investors to reduce their risks and support their decision-making processes.

One of the limitations of the study is that the dataset used is limited to certain news sources. Model performance can be further improved by using a larger data set. Improvements can be made to increase the prediction accuracy of the positive class. In future studies, it is aimed to increase the predictive power of the model by using more advanced deep learning-based models other than machine learning models.

Acknowledgements

This study was produced as a part of the Master's thesis titled "Bitcoin Price Prediction Using Cryptocurrency News" prepared by Tanju Açı.

References

[1] B. ÇAĞLAR and U. YAVUZ, “Finansal Haberlerin Bitcoin Fiyatlarına Etkisinin Yapay Sinir Ağları İle Analizi,” Bilişim Teknol. Derg., vol. 14, no. 1, pp. 65–78, 2021, doi: 10.17671/gazibtd.703688.

[2] D. Vujičić, D. Jagodić, and S. Randić, “Blockchain technology, bitcoin, and Ethereum: A brief overview,” 2018 17th Int. Symp. INFOTEH-JAHORINA, INFOTEH 2018 - Proc., vol. 2018-Janua, no. March, pp. 1–6, 2018, doi: 10.1109/INFOTEH.2018.8345547.

[3] N. Gandal and H. Hałaburda, “Competition in the cryptocurrency market Working Paper / Document de travail Competition in the Cryptocurrency Market by Neil Gandal and Hanna Hałaburda,” Bank Canada Work. Pap., 2014.

[4] S. Nas and A. Ergin Ünal, “Bitcoin Fiyat Değişimlerinin Makine Öğrenmesi Yöntemi ile Tahmin Edilmesi,” İşletme Araştırmaları Derg., vol. 15, no. 4, pp. 2597–2608, 2023.

[5] Ş. Çetinkaya, “Kripto Paraların Gelişimi Ve Para Piyasalarındaki Yerinin Swot Analizi İle İncelenmesi,” Uluslararası Ekon. ve Siyaset Bilim. Akad. Araştırmalar Derg., vol. 2, no. 5, pp.

11–21, 2018, [Online]. Available: https://dergipark.org.tr/en/pub/joecopol/issue/42274/508677

[6] "CoinMarketCap." [Online]. Available: https://coinmarketcap.com/tr/

[7] C. Çilgin, "Metin Sınıflandırmada Yapay Sinir Ağları i le Bitcoin Fiyatları ve Sosyal Medyadaki Beklentilerin Analizi In Text Classification , Bitcoin Prices and Analysis of Expectations in Social Media with Artificial Neural Networks," vol. 4, no. 1, pp. 106–126, 2020.

[8] B. KÖKSAL, G. ERDEM, C. TÜRKELİ, and Z. KAMIŞLI ÖZTÜRK, "Twitter'da Duygu Analizi Yöntemi Kullanılarak Bitcoin Değer Tahminlemesi," Düzce Üniversitesi Bilim ve Teknol. Derg., vol. 9, no. 3, pp. 280–297, 2021, doi: 10.29130/dubited.792909.

[9] M. Matta, I. Lunesu, and M. Marchesi, "Bitcoin spread prediction using social and web search media," CEUR Workshop Proc., vol. 1388, 2015.

[10] S. Nitayaprapha, "How does social media affect cryptocurrency value? A case study of bitcoin in Thailand," Int. J. Sci. Res. Publ., vol. 11, no. 6, pp. 189–191, 2021, doi: 10.29322/ijsrp.11.06.2021.p11428.

[11] A. Fayeem, A. Kumar, R. Sagar, A. Aggarwal, and D. Jain, "Stock Price Prediction : Recurrent Neural Network in Financial Market," vol. 8, no. 01, pp. 259–264, 2022.

[12] R. Auer and S. Claessens, "Regulating cryptocurrencies: assessing market reactions," BIS Q. Rev., no. September, pp. 51–65, 2018.

[13] M. Alghobiri and M. Departmet, "Using Data Mining Algorithm for Sentiment Analysis of Users' Opinions About Bitcoin Cryptocurrency," J. Theor. Appl. Inf. Technol., vol. 97, no. 8, 2019, [Online]. Available: www.jatit.org

[14] K. Ceyhan, E. Kurtulmaz, O. C. Sert, and T. Ozyer, "Bitcoin movement prediction with text mining | Metin Madenciliği ile Bitcoin Hareket Yönü Tahmini," 26th IEEE Signal Process. Commun. Appl. Conf. SIU 2018, pp. 1–4, 2018.

[15] S. Colianni, S. Rosales, and M. Signorotti, "Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis," CS229 Proj., pp. 1–5, 2015, [Online]. Available: http://cs229.stanford.edu/proj2015/029_report.pdf

[16] M. Kinderis, M. Bezbradica, and M. Crane, "Bitcoin currency fluctuation," COMPLEXIS 2018 - Proc. 3rd Int. Conf. Complexity, Futur. Inf. Syst. Risk, vol. 2018-March, no. Complexis 2018, pp. 31–41, 2018, doi: 10.5220/0006794000310041.

[17] C. Lamon, E. Nielsen, and E. Redondo, "Cryptocurrency Price Prediction Using News and Social Media Sentiment," SMU Data Sci. Rev, vol. 1, no. 3, pp. 1–22, 2017.

[18] A. Burnie and E. Yilmaz, "Social media and bitcoin metrics: Which words matter," R. Soc. Open Sci., vol. 6, no. 10, 2019, doi: 10.1098/rsos.191068.

[19] F. Mai, Z. Shan, Q. Bai, X. (Shane) Wang, and R. H. L. Chiang, "How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis," J. Manag. Inf. Syst., vol. 35, no. 1, pp. 19–52, 2018, doi: 10.1080/07421222.2018.1440774.

[20] H. S. Taşan, "Machine Learning Dersleri - Sınıflandırma." [Online]. Available: https://www.kaggle.com/code/hasansezertaan/machine-learning-dersleri-s-n-fland-rma

[21] D. D. S. Uğuz, Makine Öğrenmesi Teorik Yönleri Ve Python Uygulamaları İle Bir Yapay Zeka Ekolü. Atlas Akademi Basım Yayın Dağıtım Tic. Ltd. Şti., 2023.

[22] İ. Aydoğan and G. Zirhlioğlu, "Öğrenci başarılarının yapay sinir ağları ile kestirilmesi," dergipark.org.trİ AYDOĞAN, G ZIRHLIOĞLUVan Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 2018•dergipark.org.tr, vol. 15, no. 1, pp. 577–610, 2018, doi: 10.23891/efdyyu.2018.80.

[23] S. ARSLANKAYA and Ş. TOPRAK, "Makine Öğrenmesi ve Derin Öğrenme Algoritmalarını Kullanarak Hisse Senedi Fiyat Tahmini," Uluslararası Muhendis. Arastirma ve Gelistirme Derg., vol. 13, no. 1, pp. 178–192, 2021, doi: 10.29137/umagd.771671.

[24] L. D. Avendaño-Valencia and S. D. Fassois, "Support Vector Networks," J. Phys. Conf. Ser., vol. 628, no. 1, pp. 273–297, 2015, doi: 10.1088/1742-6596/628/1/012073.

[25] P. Jain, K. R. Srinivas, and A. Vichare, "Depression and Suicide Analysis Using Machine Learning and NLP," J. Phys. Conf. Ser., vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012034.

[26] S. Erpolat and E. Öz, "Kanser Verilerinin Sınıflandırılmasında Yapay Sinir Ağları ile Destek Vektör Makinelerinin Karşılaştırılması," İstanbul Aydın Üniversitesi Fen Bilimleri Dergisi, vol. 2, no. 5. pp. 71–83, 2010. [Online]. Available: http://iaud.aydin.edu.tr/makaleler/yil2sayi5/IAUD_Yil_2_Sayi_5_Makale_4.pdf

[27] D. Kılınç, E. Borandağ, F. Yücalar, V. Tunalı, M. Şimşek, and A. Özçift, "KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi," Marmara Fen Bilim. Derg., vol. 28, no. 3, 2016, doi: 10.7240/mufbed.69674.

[28] J. Brownlee, "How to Develop an Extra Trees Ensemble with Python." [Online]. Available: https://machinelearningmastery.com/extra-trees-ensemble-with-python/

[29] S. EMANET, G. KARATAŞ AYDOĞMUŞ, and Ö. DEMİR, "Öznitelik Seçme Yöntemlerinin Makine Öğrenmesi Tabanlı Saldırı Tespit Sistemi Performansına Etkileri," DÜMF Mühendislik Derg., vol. 5, pp. 743–755, 2021, doi: 10.24012/dumf.1051340.

[30] H. Tanrıkulu, "Sald ı r ı Tespit Sistemlerinde Yapay Sinir A ğ lar ı n ı n Kullan ı lmas ı Saldırı Tespit Sistemlerinde Yapay Sinir Ağlarının Kullanılm a sı," no. January 2008, 2017.

[31] S.-L. Developers, "Naive Bayes." Accessed: Nov. 12, 2024. [Online]. Available: https://scikit-learn.org/dev/modules/naive_bayes.html#bernoulli-naive-bayes

[32] N. Ninja, "TF-IDF: Weighing Importtance In Text." [Online]. Available: https://letsdatascience-com.translate.goog/tf-idf/?_x_tr_sl=en&_x_tr_tl=tr&_x_tr_hl=tr&_x_tr_pto=tc

[33] A. GÜRAN and D. KINIK, "TF-IDF ve Doc2Vec Tabanlı Türkçe Metin Sınıflandırma Sisteminin Başarım Değerinin Ardışık Kelime Grubu Tespiti ile Arttırılması," Eur. J. Sci. Technol., no. 21, pp. 323–332, 2021, doi: 10.31590/ejosat.774144.

[34] N. Van Otten, "SMOTE Oversampling & Tutorial On How To Implement In Python And R." [Online]. Available: https://spotintelligence.com/2023/02/17/smote-oversampling-python-r/