Sosyoekonomi

2025, Vol. 33(63), 115-142

RESEARCH ARTICLE

ISSN: 1305-5577
DOI: 10.17233/sosyoekonomi.2025.01.06
Date Submitted: 30.03.2024
Date Revised: 02.09.2024
Date Accepted: 30.11.2024

The Effect of Cryptocurrency Ecosystem and Global Indicators on Bitcoin Price[1]

Ahmet AKUSTA *(https://orcid.org/0000-0002-5160-3210), Konya Technical University, Türkiye; ahmetakusta@hotmail.com*

Mehmet Nuri SALUR *(https://orcid.org/0000-0003-1089-1372), Necmettin Erbakan University, Türkiye; nsalur@erbakan.edu.tr*

Kripto Para Ekosistemi ve Küresel Göstergelerin Bitcoin Fiyatı Üzerindeki Etkisi[2]

Abstract

This research aims to forecast the price of Bitcoin by identifying the factors that influence its price movements. The study combines 396 variables, categorised into data concerning the cryptocurrency ecosystem and data about significant global indices. The analysis utilises a dataset spanning 90 days from October 2022 to December 2022. The dataset is divided into 85% for training and 15% for testing. Among the 18 machine learning methods, the model demonstrating the highest accuracy is selected. The findings show the solid overall performance of the model, as indicated by an R2 score of 0.909.

Keywords : Bitcoin, Cryptocurrency, Price Volatility, Blockchain, Machine Learning.

JEL Classification Codes : G15, G17, C52, C53.

Öz

Bu araştırma, Bitcoin'in fiyatını etkileyen faktörleri belirleyerek Bitcoin'in fiyatını tahmin etmeyi amaçlamaktadır. Çalışma, kripto para ekosistemiyle ilgili veriler ve önemli küresel endekslerle ilgili veriler olmak üzere toplamda 396 değişkeni bir araya getirmektedir. Analiz, Ekim 2022'den Aralık 2022'ye kadar olan 90 günlük bir veri setini kullanmaktadır. Veri seti, %85'i eğitim ve %15'i test için ayrılmıştır. 18 makine öğrenme yöntemi arasından en yüksek doğruluğa sahip olan model seçilmiştir. Bulgular, modelin, 0.909 R2 skoruyla iyi bir performans sergilediğini göstermektedir.

Anahtar Sözcükler : Bitcoin, Kripto Para, Fiyat Hareketliliği, Blokzincir, Makine Öğrenmesi.

---

[1] *This study is derived from the doctoral dissertation titled "Bitcoin Price Volatility Prediction with Machine Learning", completed in 2023 by Dr. Ahmet Akusta under the supervision of Assoc.Prof.Dr. Mehmet Nuri Salur in Necmettin Erbakan University, Institute of Social Sciences, Department of Business Administration.*

[2] *Bu makale, Necmettin Erbakan Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı'nda Doç.Dr. Mehmet Nuri Salur danışmanlığından Dr. Ahmet Akusta tarafından 2023 yılında tamamlanan "Bitcoin Fiyat Hareketliliğinin Makine Öğrenmesi ile Tahmin Edilmesi" başlıklı doktora tezinden türetilmiştir.*

1. Introduction

The most prominent cryptocurrency, Bitcoin, has experienced significant price fluctuations since its inception. Understanding the factors that drive these price movements is crucial for investors and market participants. This study aims to predict the price of Bitcoin by identifying the key factors that influence Bitcoin price movements.

Bitcoin volatility is a well-known problem in the cryptocurrency market. Despite its increasing popularity, Bitcoin's price exhibits high volatility, posing challenges for investors. Previous studies have explored the relationship between Bitcoin returns and volatility. For instance, (Bouri et al., 2017) found no evidence of an asymmetric return-volatility relationship in the Bitcoin market (Kayal & Balasubramanian, 2021). However, Bitcoin prices fluctuate and sometimes appear excessively volatile; they tend to stabilise over time. These findings highlight the complex nature of Bitcoin's price movements and the need for further investigation.

Economic crises also play a role in influencing Bitcoin's price. Some studies suggest that Bitcoin can hedge against local currency depreciation, but it is not considered a safe-haven asset during global crises (Zhao, 2022). Bitcoin's behaviour as a hedge or haven asset is contingent upon the level of uncertainty in global financial markets. According to (Zhou, 2019), Bitcoin may serve as a hedge in times of low uncertainty. However, with growing global economic anticipation, it will likely move in tandem with the markets, offering little protection against stock market crashes. Therefore, understanding the relationship between Bitcoin and economic crises is essential for predicting its price movements accurately.

To predict the price of Bitcoin accurately, a comprehensive analysis of various factors is necessary. Previous studies have highlighted the importance of demand shocks and liquidity in determining Bitcoin prices. Bitcoin demand shocks have been found to influence its price significantly (Gronwald, 2019). Liquidity, measured by trade volume, also plays a role, with exchanges that have larger trade volumes trading closer to Bitcoin's market price (Johnson, 2020). Furthermore, factors like supply quantity, global economic conditions, and user adoption have been identified as crucial determinants of Bitcoin's price (Liu & Zhang, 2023).

While significant strides have been made in understanding Bitcoin price movements, prior studies have often employed models with limited features, which may not fully capture the complexities of the cryptocurrency market (Dutta et al., 2020). There is a recognised need to include additional factors, such as national policies and social media activity, to enhance prediction accuracy (Huang et al., 2022). Moreover, there needs to be more in determining optimal preprocessing strategies for sentiment analysis of Bitcoin-related tweets, which could further improve machine learning models used for price prediction (Pano & Kashef, 2020).

A notable research gap exists in comparing machine learning algorithms for Bitcoin price prediction. While some studies have explored individual models, only some have comprehensively compared multiple algorithms. This study seeks to fill this gap by comparing 19 machine learning algorithms, providing insights into which models perform best under different conditions and data inputs.

This study also leverages a comprehensive dataset that includes 350 variables related to Bitcoin, categorised under network information, crypto exchange transactions, address data, price data, and social media data. Furthermore, it examines the relationship between Bitcoin and global indices using 46 important variables worldwide. By integrating these diverse data sources and comparing the performance of 19 different machine learning algorithms, this study aims to provide a more accurate and robust prediction model for Bitcoin prices, thereby contributing to the existing literature and offering valuable insights for investors and market participants.

In conclusion, this study aims to contribute to the existing literature by predicting the price of Bitcoin based on the key factors that influence its price movements. Understanding these factors will provide investors and market participants valuable insights, helping them make informed decisions and manage risks in the ever-evolving cryptocurrency market.

The following sections of this paper discuss relevant literature in the field, reviewing various studies that have explored Bitcoin price prediction and the factors influencing it. After thoroughly preprocessing the data, a comprehensive set of machine learning algorithms was applied to predict Bitcoin prices. The results of these predictions are then analysed and discussed, offering insights into the accuracy and performance of the models in capturing the dynamics of Bitcoin price movements.

2. Literature Review

Bitcoin price prediction has drawn significant academic attention due to its volatility and expanding role in global financial markets. Many studies have examined different variables and machine learning methods to enhance the understanding and forecasting of Bitcoin price movements. Factors that influence Bitcoin's price movements are multifaceted and encompass a variety of determinants. Research indicates that exchange rate variables significantly positively affect Bitcoin prices in the long run (Andrean, 2020).

One of the earliest contributions to Bitcoin price forecasting is from (Greaves & Au, 2015), who explored the predictive capacity of blockchain-based features. Their analysis focused on feature engineering and machine learning optimisation to classify Bitcoin's price movements. Despite the complexity of the task, their approach achieved an accuracy rate of around 55% in predicting price direction. This study highlights the importance of blockchain-specific data for predictive modelling, providing a foundational base for our research. We build on this by incorporating blockchain data, macroeconomic indicators, and social media sentiment in our predictive models to improve forecasting accuracy.

Wang et al. (2016) explored the influence of traditional macroeconomic indicators on Bitcoin prices, focusing on variables like the stock price index, oil prices, and Bitcoin's daily trading volume. They employed cointegration analysis and the Vector Error Correction (VEC) Model to investigate short-term dynamics and long-term equilibrium relationships. The study found that these factors interact dynamically with Bitcoin prices, emphasising the relevance of broader economic trends. In this context, Dyhrberg (2016) highlighted that the demand for Bitcoin as a medium of exchange plays a crucial role in influencing its returns, akin to a currency, rather than solely driven by temporary price shocks. Our study extends this approach by integrating a more comprehensive range of global indicators, such as bond and currency indices, to capture a more comprehensive picture of the economic forces influencing Bitcoin's volatility.

In a broader exploration of the cryptocurrency market, Virk (2017) used feature engineering and machine learning models, including support vector classifiers, random forests, and gradient boosting classifiers, to analyse the price relationships between ten cryptocurrencies. The study found strong correlations among the cryptocurrencies, suggesting market-wide factors influence their prices. Moreover, Bitcoin's price dynamics are influenced by a combination of factors, including market fundamentals such as supply and demand, attractiveness for investors, and global financial indicators (Ciaian et al., 2015). While Virk's research focuses on cross-cryptocurrency price dynamics, our study relies on Bitcoin, using diverse machine-learning models to predict its price. However, we incorporate some of Virk's insights by examining the impact of cryptocurrency market trends and liquidity on Bitcoin.

Wu et al. (2019) proposed a new prediction framework for daily Bitcoin price prediction, utilising two Long Short-Term Memory (LSTM) models: a traditional LSTM and an LSTM AR(2) model. The results demonstrated that the LSTM AR(2) model significantly improved prediction accuracy. Similarly, Roy et al. (2019) analysed Bitcoin data from 2013 to 2017, using time series approaches such as ARIMA models, and reported 90% accuracy in short-term volatility predictions. Both studies underscore the superiority of machine learning models, especially LSTM, over traditional time series models for Bitcoin price prediction. Our research builds on this by comparing the performance of 18 different machine learning models, including LSTM, across various factors influencing Bitcoin's price.

A comprehensive analysis by Kervanci & Akay (2020) categorised previous studies into machine learning methods, social media impact, time-frequency effects, and hyperparameter optimisation. Their study found that machine learning methods outperformed statistical models in Bitcoin price prediction, particularly when optimised hyperparameters. Similarly, Jana et al. (2021) compared six advanced forecasting models, including random forests, support vector machines, and multi-layer perceptron neural networks, finding that advanced regression frameworks provided higher prediction accuracy. These findings align with our research goals, where we compare multiple machine learning models and apply hyperparameter tuning to enhance prediction performance.

However, Cretarola et al. (2017) demonstrate that the high volatility of Bitcoin prices is influenced by factors such as sentiment and popularity within the market. Though not directly observable, these factors can be proxied by indicators such as the volume of Google searches and Wikipedia requests related to Bitcoin.

The impact of social media sentiment on Bitcoin prices has also been explored in recent studies. Aggarwal et al. (2019) examined the relationship between Twitter sentiment and Bitcoin prices, finding that sentiment significantly affects price movements, mainly when originating from influential users. Critien et al. (2022) reinforced this by demonstrating that Twitter sentiment could predict the direction and magnitude of price changes. These studies highlight the importance of incorporating sentiment analysis into price-prediction models. Some studies suggest that factors like usage in trade, money supply, popularity, and public interest influence Bitcoin prices. Fil & Krištoufek (2020) argue that Bitcoin's price behaviour cannot be fully explained by economic fundamentals but rather by the actions of buyers and sellers (Bouri et al., 2018). Our research includes social media data as one of the critical variables, integrating sentiment alongside technical and macroeconomic factors to improve prediction accuracy.

Munim et al. (2019) noted that long-term predictions present unique challenges for those who use ARIMA and Neural Network Autoregressive (NNAR) models for next-day Bitcoin price prediction. Their study showed that NNAR models outperformed ARIMA in the training sample but not in the test sample, indicating the complexity of capturing long-term trends. Similarly, Aghashahi & Bamdad (2022) concluded that the Fitnet network with 30 hidden neurons achieved superior accuracy in Bitcoin price forecasting over nine months. These studies underscore the challenges of long-term prediction due to Bitcoin's inherent volatility. On the other hand, Pele & Mazurencu-Marinescu-Pele (2019) identify a bidirectional causality between Bitcoin's price and network size, noting that expected price increases attract more investors, potentially resulting in super-exponential price growth. In response, our study focuses on short-term prediction, analysing Bitcoin's price movements over 90 days to better capture the dynamic nature of the market. Additionally, the fixed supply and predictable scarcity of Bitcoin create a strong link between public interest, user adoption, and price (García et al., 2014).

The literature review examines a wide range of studies on Bitcoin price prediction. The first group introduces critical factors such as exchange rates, demand, and macroeconomic conditions, providing a foundation for understanding price movements. The second group explores the methodologies used in Bitcoin prediction, including blockchain data, machine learning techniques, and the shift from traditional to more advanced models, emphasising the importance of broader data inputs like cryptocurrency trends and global indicators. The third group focuses on sentiment analysis and behavioural economics, particularly the influence of social media sentiment on Bitcoin's price. Finally, the last group discusses the challenges of long-term prediction due to Bitcoin's volatility and network effects, suggesting that short-term predictions may be more accurate.

These key thematic groups directly relate to the identified research gaps. This study addresses these gaps by incorporating a comprehensive dataset of 350 variables, including network information, exchange transactions, price data, and social media sentiment, which previous studies often overlooked. In addition to expanding the range of data inputs, the study also fills a gap in the literature by comparing the performance of 19 different machine learning algorithms, moving beyond the focus on individual models. The goal is to improve the accuracy of Bitcoin price prediction by developing more robust models that capture the complexities of the cryptocurrency market, providing valuable insights for investors and contributing to the existing research.

3. Data and Methodology

Initially, the intention was to analyse the entire duration of 2022. However, a whole year was deemed excessively protracted due to the inherent high volatility characterising instruments like Bitcoin. Consequently, the research scope was refined to a narrower timeframe, precisely 90 days from October 2022 to December 2022.

One potential limitation of using a short timeframe for Bitcoin price prediction is the risk of overfitting, mainly when working with small datasets. Small datasets can cause machine learning models to perform well on training data but fail to generalise to new, unseen data, leading to poor predictive performance (Charilaou & Battat, 2022). Overfitting is a common issue in machine learning, especially when more data is needed to capture the underlying patterns. To overcome this limitation, this study employs cross-validation techniques across all algorithms to ensure that the models are not just memorising the training data but are learning patterns that generalise to new data. This rigorous validation helps mitigate the overfitting problem and enhances the robustness of the analysis.

Another limitation of small datasets, particularly in time-series prediction, is the need for more information to train accurate models (Hayashi et al., 2020). This is especially problematic for long-term predictions, where the data size tends to be small and future trends are highly uncertain. However, this study focuses on short-term prediction, which naturally benefits from a smaller data horizon, as near-term predictions tend to have higher accuracy. The study leverages the available data more effectively by concentrating on short-term price movements. It avoids the additional complexities associated with long-term predictions, thereby improving the models' reliability within the dataset's constraints.

Subsequently, the dataset was divided into 85% for training and 15% for testing. The predictions generated from the training phase were subsequently visualised for the test data, providing valuable insights.

The test data was used to generate predictions using the selected machine learning method, which was subsequently compared with the actual results. This study's methodology consists of selecting data sources, data collection, processing and analysis, evaluating machine learning algorithms, and visualising the results.

3.1. Data Collection and Preprocessing

The research utilises data that can be divided into two primary categories. The initial category encompasses data associated with the cryptocurrency system, which is further organised into the following main sections: Network Information, Cryptocurrency Exchange Transactions, Address Data, Price Data, and Social Media Data. Obtained through Tradingview, this category reflects the overall performance of the cryptocurrency market and the price fluctuations of cryptocurrencies.

On the other hand, the second category comprises 46 variables linked to significant global indices. These variables are grouped into the following main sections: Bond Indices, Energy Indices, Currencies, Volatility Indexes, and Stock Indices. This category of data is acquired from Yahoo Finance.

3.2. Data Preprocessing

Techniques for preparing data are critical for extracting relevant information from it. Data is cleaned to remove noise and fix discrepancies. Data integration is the process of combining data from several sources into a usable database. Data transformation methods normalise or reduce data to prepare it for data mining. These strategies are used in data mining to get high-quality results while saving time (Oğuzlar, 2003).

Data cleaning and transformation activities help researchers prepare the data they have collected for the study. These activities are critical for improving data correctness and reliability. This section describes the cleaning and transformation techniques utilised on the study's data.

The data was first reviewed for missing or incorrect entries. Missing data were filled by copying the preceding entry since cryptocurrency exchanges operated on weekends while other exchanges were closed. To mitigate the potential for machine learning algorithms to overreact to the often minute price fluctuations of the highly volatile Bitcoin, the price was expressed in increments of thousands of dollars.

Subsequently, the data underwent normalisation utilising the z-score method, a crucial step in ensuring accurate and effective results, mainly when dealing with data on disparate scales. This normalisation process minimised potential misinterpretations and errors from dissimilarities across the dataset.

Furthermore, before analysis, the data underwent a meticulous cleaning process. In this regard, any excessive or missing data within the utilised dataset was systematically eliminated. Within Category 1, out of the initial pool of 350 variables, a careful selection led to the inclusion of only 73 variables for further analysis. Similarly, within Category 2, the selection process included 32 variables from the original set of 46. As a result, the dataset became more comprehensible and conducive to analysis.

These meticulous activities hold paramount importance in fostering researchers' accurate understanding of the data, thus significantly enhancing the reliability and credibility of the research outcomes. The research facilitates readers' comprehension and serves as a roadmap for future studies in this domain by providing a comprehensive exposition of the data cleaning and transformation techniques employed.

3.3. Feature Selection

3.3.1. Removal of Low Variance Variables

The quality of the characteristics employed directly relates to a model's success. As a result, variables with slight variance are often meaningless to the model and can harm its performance. The "Ignore Low Variance" approach was utilised in the investigation. This approach automatically discovers and removes variables with low variance during the model development. This enhances the model's performance and allows for more effective feature selection.

Choosing features is frequently one of the most critical tasks in the model construction process. This stage enables more accurate data analysis and generates more exact outputs from the model. The strategy improves feature selection by removing low-volatility variables, allowing the model to produce more accurate results.

In statistics, an outlier is a data point numerically far from the other data points. It is an observation that deviates significantly from the different sample members. While outliers are sometimes perceived as errors or noise, they can also contain essential information. Outliers frequently suggest a problem with model fitting or measurement inaccuracy; therefore, identifying them is critical for practical analysis (Oyeyemi et al., 2015).

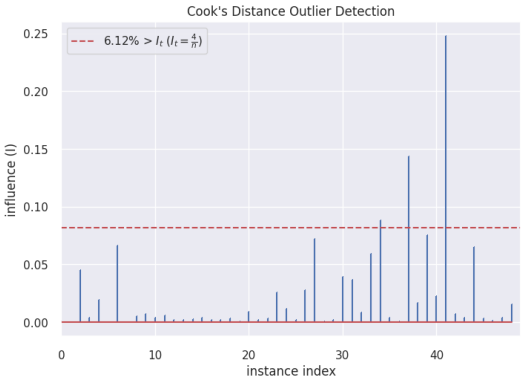
3.3.2. Removal of Outliers

Using a set threshold value to remove outliers is an effective way to improve a model's accuracy. This strategy identifies and removes Outliers in the dataset from the model. Outliers exceeding a specified threshold value were deleted from the model.

As shown in Figure 1, identifying outliers using a threshold value is crucial for improving model accuracy. The figure illustrates how data points above the threshold are considered outliers and removed from the model to enhance predictive performance.

The 0.07 threshold for identifying outliers was chosen because it led to better model performance during testing. Cook's Distance helps measure the influence of each data point, and by setting this threshold, the model removes only those data points that have too much influence and could distort predictions. This approach improves accuracy by reducing noise while keeping valuable data intact. The visual also supports this, showing how data points above the 0.07 threshold were effectively flagged as outliers.

Figure: 1
Identifying Outliers



3.3.3. Removal of Multicollinearity

Multicollinearity poses significant challenges in regression models, as it inflates the variance of the estimated coefficients, leading to unstable and unreliable model parameters. The statistical literature emphasises that the main problems associated with multicollinearity include unstable and biased standard errors, which result in very unstable p-values for assessing the statistical significance of predictors. This instability can lead to unrealistic and untenable interpretations of the model results (Vatcheva et al., 2016). This study removed features exhibiting perfect or near-perfect collinearity based on a correlation threshold 0.9. This threshold was chosen to address the redundancy between variables that could undermine model performance and increase the risk of overfitting.

While there is no universally accepted "hard" threshold for Variance Inflation Factors (VIFs) and tolerance, common heuristics provide a conservative approach to testing for multicollinearity ((Tappin et al., 2021). It is widely accepted that VIF values above certain thresholds (typically VIF > 5 or VIF > 10) indicate the presence of problematic multicollinearity, which requires corrective action (Guan et al., 2022). In this context, pairwise correlation thresholds such as 0.9 offer a straightforward method for preemptively addressing multicollinearity by identifying and removing highly correlated features before adversely impacting the model.

To further assess multicollinearity, the Variance Inflation Factor (VIF) metric was utilised (Wamuyu, 2022). However, in cases where pairwise correlations exceeded 0.9, removing the features directly was deemed more efficient. This strategy ensures that the model remains stable, interpretable, and less prone to overfitting, thereby improving the overall reliability and validity of the results.

3.3.4. Dimensionality Reduction

Principal Component Analysis (PCA) is a widely used statistical tool for dimensionality reduction in high-dimensional data analysis (Hung et al., 2012). As an essential tool for data exploration, PCA is based on a traditional approach favouring structures with significant variances. However, this approach can be sensitive to outliers and may obscure underlying patterns of interest (Akinduko & Gorban, 2014).

PCA employs sophisticated mathematical principles to transform a set of possibly correlated variables into a smaller number of uncorrelated variables, known as principal components. Generally, PCA reduces the dimensionality of large datasets by using a vector space transformation. Through mathematical projection, the original dataset, which includes many variables, can often be interpreted using only a few variables (the principal components). Consequently, examining a reduced-dimension dataset allows users to identify trends, patterns, and outliers in the data far more quickly than was possible before conducting principal component analysis (Richardson, 2009).

PCA decreases the dimensionality of a dataset by minimising variable dependencies while considering the relationships between variables in the dataset. This strategy achieves dimensionality reduction while retaining a considerable percentage of the variation of the variables, making the data more intelligible and reducing its size. In the current investigation, PCA was employed to reduce the dataset to three dimensions, a choice driven by its ability to preserve significant variance while simplifying the dataset. PCA operates under the assumption that components with higher variance contribute more to the predictive task, yet this may only sometimes hold in specific contexts (Zaman & Ahmed, 2019). However, in this study, reducing the data to three components demonstrated the highest accuracy, suggesting that PCA effectively retained the most critical information for this dataset (Hong et al., 2018). Furthermore, the three-dimensional output is visualisable and interpretable, balancing simplicity with accuracy and enhancing data comprehension.

According to the given information, the variance explained ratios of the data set reduced to three dimensions using the PCA algorithm are as follows:

The variance of the first component: 39.43692323

The variance of the second component: 18.72248772

The variance of the third component: 9.95660597

Figure: 2
Cumulative Variance Explanation Rate of Reduced Data

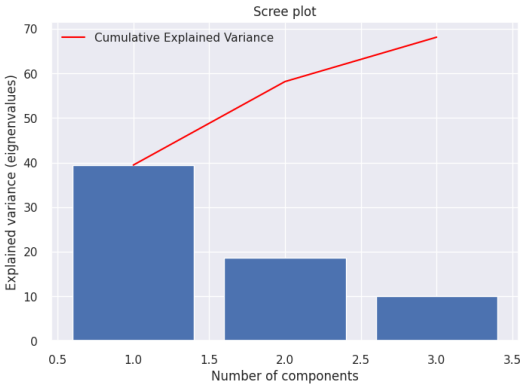


Figure 2 depicts the cumulative variance explained by the principal components. This figure highlights that the first three components account for 68.11% of the total variability, making them significant for subsequent analysis.

This information shows how much each component provides about the dataset's characteristics. For instance, since the variance of the first component is the highest, it represents the most critical features in the dataset. The variances of the other components provide information about the other features in the dataset.

Figure 3 visually represents the relationships between the reduced data components. This pair plot aids in understanding the distribution and correlation between the principal components derived from the PCA.

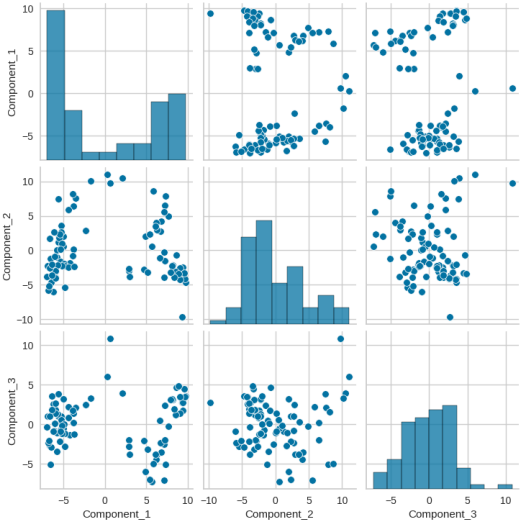
Making an overall interpretation that explains the relationship between the components in this graph is difficult because the components correspond to the original variables. Since it is unknown what those variables are, it is uncertain which relationships are meaningful. However, the spread of points in the subplots of the graph indicates whether there is a relationship between the variables. For example, in some subplots, we can observe that the points are distributed regularly, suggesting a relationship between those variables.

The diagonal graph shows the histogram of each variable. The distributions are roughly symmetric. However, the last variable (Dimension 3) is slightly skewed.

From the binary graph, we can observe that there is no strong linear relationship between the variables. However, we can discern clusters of dots in some scatterplots, particularly those containing Dimension 1, showing that Dimension 1 is distinct from the rest. We can say that Dimension 2 and Dimension 3 have a favourable relationship.

In general, the PCA technique keeps some of the underlying structure while decreasing the dimensionality of the data. As a result, it is a valuable tool in feature selection and can be used to understand the relationships between variables in a dataset.

Figure: 3
Pairplot Comparison of Reduced Data



3.3.5. Visualization of the Reduced Data Set

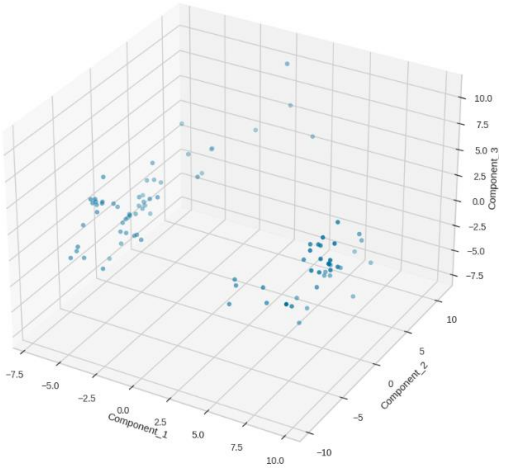
In the three-dimensional visualisation of the data reduced to three dimensions by the PCA algorithm, the X-axis represents Dimension 1, the Y-axis represents Dimension 2, and the Z-axis represents Dimension 3. Each point is generated from the original data merged in three dimensions and is displayed based on a colour-coded classification result. In this dataset, different colours represent different groups. It can be observed that the points are distributed along three different axes, with different groups being concentrated at various heights, indicating a dataset with three-dimensional dispersion.

In Figure 4, the three-dimensional visualisation of the reduced dataset provides insight into how different groups are distributed across the three principal components, facilitating better data interpretation.

In this study, 396 variables were examined, and as a result of the analysis conducted, they were effectively reduced to only three primary variables. However, the visual representation derived from this reduction in the form of a three-dimensional visualisation can pose significant challenges in terms of interpretation. Consequently, advanced visualisation techniques may be necessary to achieve a more comprehensive data analysis and facilitate a comprehensible interpretation. Therefore, employing appropriate methods

becomes crucial for enabling a more transparent and more understandable interpretation of the obtained data.

Figure: 4
Three-Dimensional Visualization of the Reduced Data



High-dimensional data can be seen in two or three dimensions using the statistical technique known as a t-distributed stochastic neighbour embedding (t-SNE). It is a variant of Stochastic Neighbor Embedding, first presented by Sam Roweis and Geoffrey Hinton and later proposed by Laurens van der Maaten. This nonlinear dimensionality reduction technique works by putting objects close together and different objects far apart. The Wikipedia page on T-distributed stochastic neighbour embedding offers more information about t-SNE.

Applying techniques such as t-SNE can help obtain more pertinent and comprehensible representations of the reduced data. This enables a deeper analysis of the interrelationships and patterns within the dataset.

The t-SNE algorithm operates in two distinct stages. Firstly, it generates a probability distribution wherein similar objects are assigned higher probabilities of being related. Subsequently, t-SNE reconstructs this probability distribution on a lower-dimensional map to minimise the divergence between the two distributions. Initially, Euclidean Distance is utilised as a similarity measure, but the algorithm can be adapted to accommodate different data types (Katubi et al., 2023).

A t-SNE manifold technique is employed to visualise the reduced data. This technique presents the data as if it were situated on a manifold-like surface, facilitating enhanced differentiation among data points.

Figure: 5
Visualisation of Reduced Data with the t-SNE Manifold Technique

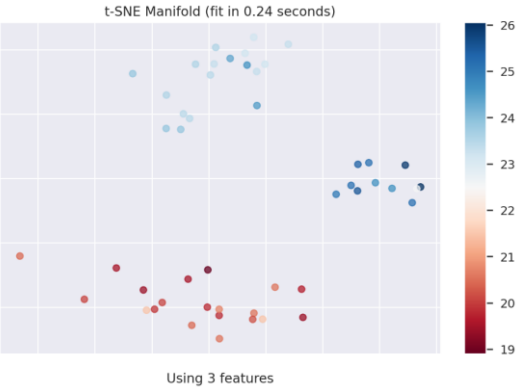


Figure 5 illustrates the t-SNE manifold technique, which effectively distinguishes between different data clusters, highlighting the nuanced relationships that were not evident in higher dimensions. One notable aspect of the image is that similar data points are grouped closely together. For instance, blue points are clustered closely to each other, while red points are more distant. After the PCA algorithm reduced the data to three dimensions, the binary comparison graph showed that Dimension 1 stands out distinctly from the other two dimensions. We can confidently say that the red points in the t-SNE plot represent Dimension 1, Dimension 2, and Dimension 3, both in blue and close to each other, indicating the similarity between the data points.

The visualisations in Figures 4 and 5 demonstrate the effectiveness of dimensionality reduction techniques in organising the data into distinct clusters, thereby enhancing model performance. In Figure 4, the PCA plot shows a clear separation of data points along three principal components. This indicates that the retained features differentiate groups meaningfully, improving the model's accuracy and generalizability. Figure 5, using t-SNE, captures local relationships and reveals distinct groupings through colour coding, further reinforcing the effectiveness of feature selection and dimensionality reduction. Both visualisations highlight critical features, and reducing data complexity leads to better model interpretation, more precise predictions, and overall improved performance.

3.4. Machine Learning Algorithms (Model Selection)

In this analysis, each model's default and fine-tuned parameters were systematically compared based on their respective predictive accuracies. Each model's performance evaluations were conducted using default and optimised configurations. The configuration that demonstrated superior accuracy was selected for further analysis. Subsequently, predictions were generated using the models with the better-performing parameters, and these results were utilized to identify the most effective model overall.

In regression analysis, several statistical methods are used to assess the quality of a model. One of these methods is R-squared (R2), which measures how well the independent variables' variations can predict the dependent variable's variations. R2 is a number between 0 and 1, where a value close to 1 indicates a higher accuracy rate and a better fit of the model. In contrast, a value close to 0 indicates fewer correct predictions and a poorer fit (American Institute of Certified Public Accountants., n.d.).

Table: 1
Comparison of Machine Learning Algorithms

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
K Neighbors Regressor	0,4244	0,3502	0,5231	0,8538	0,0271	0,0233	0,136
Extra Trees Regressor	0,5406	0,5679	0,6723	0,7648	0,0352	0,0301	0,306
Random Forest Regressor	0,5135	0,6473	0,6451	0,7468	0,0336	0,0286	0,573
Extreme Gradient Boosting	0,5341	0,7194	0,6926	0,7269	0,0368	0,0301	0,171
Gradient Boosting Regressor	0,5459	0,8454	0,7162	0,6767	0,0374	0,0304	0,197
AdaBoost Regressor	0,6159	0,9775	0,7996	0,6269	0,0417	0,0343	0,199
Decision Tree Regressor	0,5553	0,9041	0,7627	0,6246	0,04	0,0309	0,133
Bayesian Ridge	0,7444	0,7866	0,8618	0,624	0,0453	0,0411	0,134
Ridge Regression	0,7428	0,7886	0,8627	0,6212	0,0454	0,041	0,694
Least Angle Regression	0,7425	0,7897	0,8632	0,6203	0,0455	0,041	0,349
Linear Regression	0,7425	0,7897	0,8632	0,6203	0,0455	0,041	1,547
Huber Regressor	0,7383	0,819	0,878	0,6063	0,0464	0,0409	0,14
Light Gradient Boosting Machine	0,8265	1,1052	0,9916	0,5211	0,0515	0,0453	0,466
Orthogonal Matching Pursuit	0,9967	1,5383	1,155	0,2991	0,0585	0,0534	0,42
Elastic Net	0,9904	1,3366	1,1126	0,2413	0,0578	0,0539	0,383
Lasso Regression	1,0357	1,4508	1,1641	0,19	0,0605	0,0564	0,385
Lasso Least Angle Regression	1,0357	1,4508	1,1641	0,19	0,0605	0,0564	0,485
Passive Aggressive Regressor	1,089	1,9991	1,3091	-0,3522	0,0714	0,0596	0,133
Dummy Regressor	1,5527	2,9241	1,6832	-0,3867	0,0869	0,0856	0,134

Table 1 compares machine learning algorithms based on multiple performance metrics, including MAE, MSE, and R2. The table demonstrates that the K Neighbors Regressor and Extra Trees Regressor exhibit the highest R2 values, indicating superior performance. In the table, "TT (Sec)" represents each algorithm's time. When looking at the "TT (Sec)" values in the table, it can be observed that some algorithms are faster than others. For example, the K Neighbors Regressor is the fastest algorithm, while algorithms like Linear Regression, Lasso Regression, and Dummy Regressor are slower. However, more than relying on this criterion is needed, and evaluation should also be done based on other performance metrics.

The study compares the performance of 18 algorithms based on the R2 criterion. The study's main objective is to rank the algorithms in terms of performance using the R2 values as the performance metric. When examining the table, it can be seen that the K Neighbors Regressor and Extra Trees Regressor algorithms have the highest R2 values. This indicates that these algorithms perform better than the others. On the other hand, Lasso Regression, Lasso Least Angle Regression, Dummy Regressor, and Passive Aggressive Regressor algorithms show the lowest performance. Other performance metrics considered in the study also support this result, indicating that these algorithms are weaker in performance than others. However, it should be noted that the study considers the runtime to be insignificant. Being aware of this point is essential regarding the practical applicability of the algorithms in real-life scenarios.

3.4.1. Prediction Error for the Machine Learning Model

The graph represents the prediction error, another tool used to evaluate the performance of the selected machine-learning algorithm. The prediction error graph shows the differences between the actual and predicted values, indicating how accurate the model's predictions are.

Figure: 6
Prediction Error Curve

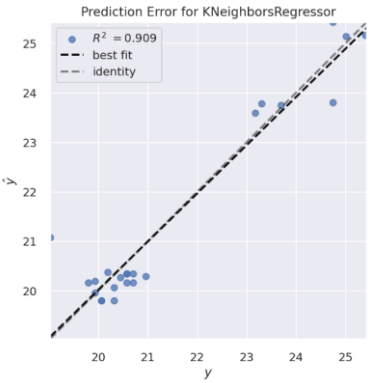


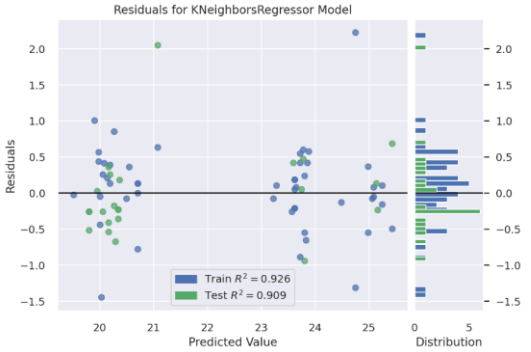
Figure 6 details the prediction error for the model. This curve demonstrates how the majority of points fall below the error line, indicating that the model's predictions are generally accurate. The R2 score is 0.909. This score indicates how well the model fits the data, with a value nearer 1 suggesting a better fit. A high R2 score indicates that the model fits the data well and that predictions are accurate.

3.4.2. Examination of Residuals for the Machine Learning Model

Residuals are the disparities between the machine learning algorithm's observed and predicted values of the target variable. These mistakes are plotted against the projected values in the residual plot. The points on residual plots, which are used to assess the success of the chosen machine learning algorithm, represent the disparities between the actual and anticipated values. If a model predicts perfectly, all of the points will be 0. If there is any deviation, the points on the graph will be spread around a line. The bigger the deviation, the further the points are from the line (Residuals vs. Fits Plot).

The residual plot in Figure 7 shows that the residuals are randomly scattered around the zero line, supporting the linear relationship between the predictor and target variables and confirming the model's appropriateness. This is an advantageous feature of a well-behaved residual plot. This implies that the predictor and target variables have a linear connection and that the assumption of a linear relationship is acceptable.

Figure: 7
Visualisation of Model Residuals



The graph also indicates that the selected machine learning algorithm has relatively tiny deviations between the actual values and the predictions. Most points are clustered very close to a line on the graph, indicating that the model performs well overall.

The graph also provides the training and test R2 scores. The R2 score measures how well a model fits the data, with a value closer to 1 indicating a better fit. While the train R2 score indicates how well the model fits the training data, the test R2 score predicts how well the model will perform on new data. A high test R2 score suggests a high likelihood of the model performing well on new data.

3.4.3. Model Evaluation and Learning Curves

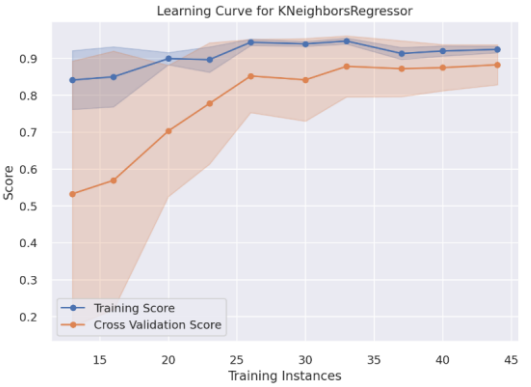
3.4.3.1. Learning Curve of the Machine Learning Model

A learning curve is a diagram of a machine learning model's learning behaviour. It demonstrates whether providing additional data to the model is advantageous for training. The curve contains both the cross-validation score and the training score.

The "K Neighbors Regressor" algorithm's learning curve is depicted in the figure. The score is plotted on the y-axis of this curve, and the number of training samples is plotted on the x-axis. The standard deviation is depicted in the plot by the darkened area. The training score is the training set's accuracy rating. The accuracy rating for the test set may be seen in the cross-validation score. The machine learning model was trained with poor-quality data if these two values, the training and cross-validation scores, overlap. (Katubi et al., 2023).

The data is used to train the machine-learning model because the training and cross-validation scores in the figure do not intersect. The model requires more training data since the training score is larger than the cross-validation score.

Chart: 1
Learning Curve of The Model



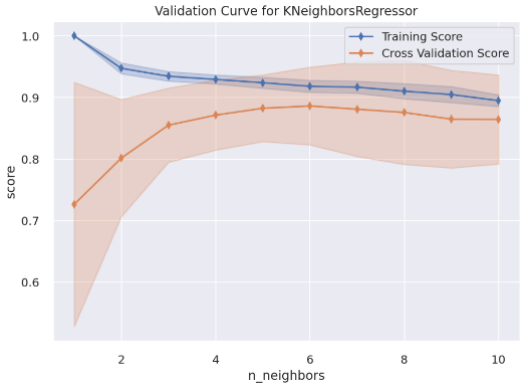
As depicted in Chart 1, the learning curve for the K Neighbors Regressor shows the relationship between the number of training examples and model accuracy. In the graph, as the number of training examples increases, the training score decreases while the validation score rises. Initially, both scores are low, and as more data examples are used, the training and validation scores increase. However, more than adding more data is required to improve the performance beyond a certain point, and the validation score becomes nearly constant. In our model, the cross-validation is terminated when the training and validation scores stabilise.

3.4.3.2. Validation Curve of the Machine Learning Model

A scoring function based on predictions is necessary to measure a model's accuracy. Optimising hyperparameters according to this scoring function becomes unreliable and biased. Test data is required for a correct generalisation prediction. However, visualising the effect of a single hyperparameter on the training score and the validation score can help determine whether the hyperparameter values are too appropriate or inappropriate.

Typically, achieving both a low training score and a high test score is impossible. When a predictor exhibits inadequate training and validation scores, it indicates underfitting. Conversely, overfitting becomes a concern when the training score is considerably high, but the validation score is disproportionately low. Consequently, striking an optimal balance in optimising hyperparameters becomes crucial for attaining accurate generalisation predictions, ensuring a proper equilibrium in the model's capacity to generalise effectively (<scikit-learn.org>, 2023).

Chart: 2
Validation Curve of The Model



The validation curve in Chart 2 illustrates the effect of varying hyperparameters on model accuracy. An optimal balance is achieved when the hyperparameter value is moderate, avoiding under and overfitting. This graph helps determine the optimal value of a hyperparameter by visualising its effect on the model's accuracy. By examining the graph, it can be understood that when the value of the hyperparameter is low, the model is insufficient initially. However, as the value of the hyperparameter increases, the model starts to fit the training data better, but the validation score decreases. This indicates that the model is overfitting the data.

Table: 2
10-Fold Cross-Validation Results of the Selected Algorithm

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0,1973	0,0802	0,2832	0,9095	0,015	0,0106
1	0,2096	0,0544	0,2332	0,9628	0,0126	0,0119
2	0,2309	0,0997	0,3157	0,8919	0,0162	0,0123
3	0,3561	0,1736	0,4166	0,9194	0,0209	0,019
4	0,283	0,0983	0,3135	0,9616	0,0151	0,0145
5	0,6791	0,677	0,8228	0,7792	0,0427	0,0357
6	0,3122	0,1165	0,3414	0,9648	0,0188	0,0181
7	0,7035	0,7007	0,8371	0,6307	0,0449	0,0391
8	0,7501	1,1441	1,0696	0,6307	0,0539	0,043
9	0,5223	0,3579	0,5982	0,8877	0,0306	0,0285
Mean	0,4244	0,3502	0,5231	0,8538	0,0271	0,0233
Std	0,2076	0,3511	0,2767	0,1229	0,0142	0,0116

The results of the 10-fold cross-validation are summarised in Table 2. The average MAE and R2 values across the folds indicate the model's robustness and generalisation capability on unseen data. A distinct test dataset is commonly designated to evaluate the efficacy of machine learning algorithms impartially. Nevertheless, specific scenarios necessitate addressing the concern of overfitting. To mitigate this inherent risk, the 10-fold cross-validation technique is employed to assess the performance of machine learning models. This technique divides the dataset into ten mutually exclusive sections, wherein

each segment is alternately used for training and testing the model across multiple iterations. Consequently, this approach yields a more objective appraisal of the model's ability to generalise effectively. Cross-validation thus emerges as a valuable technique for determining the performance and generalizability of a given model.

In this particular investigation, 10-fold cross-validation was used. The average MAE is calculated as 0.4244, suggesting a 0.4244 unit difference between the expected and actual values. Furthermore, the R2 value is 0.8538, indicating that the model has good explanatory power.

K-Nearest Neighbors (KNN) is a popular nonparametric model, especially effective in handling datasets with numerous features. Unlike parametric models, KNN does not require prior assumptions about the underlying data distribution, making it flexible for complex datasets. This flexibility allows KNN to achieve high accuracy in solving regression problems, particularly in cases with many features, as shown in this study with 105 features over 90 days. Previous studies have highlighted the utility of nonparametric models like KNN in capturing nonlinear relationships without predetermined functional forms, which is critical for modelling high-dimensional data accurately (Khatun & Siddiqui, 2023; Taylan, 2019).

However, KNN has notable computational efficiency limitations. The time complexity of KNN increases with the size of the dataset, making it computationally expensive for large-scale problems (Adhikary & Banerjee, 2022). This limitation can hinder the scalability of the model as the data grows. Various approaches, such as the KWKNN algorithm and kd-tree-based methods, have been proposed to reduce the computational burden while maintaining accuracy ((Aung et al., 2018; Rubio et al., 2009). These methods attempt to balance accuracy with speed, although the classic KNN algorithm still struggles with large datasets and higher values of kkk.

In this study, despite the inherent computational limitations of KNN, the model performed with high efficiency and accuracy due to the dataset's characteristics: short-term data and a manageable number of features. The relatively small dataset allowed KNN to execute quickly while maintaining high accuracy, mainly because KNN is known to perform well with heavy-tailed feature distributions (Zhao & Lai, 2019). This finding is consistent with prior research, which has shown that KNN's performance can be optimised in specific contexts, primarily when the selection of kkk is carefully managed to ensure a desirable tradeoff between bias and variance (Mladenova & Valova, 2023).

In the context of this study, KNN demonstrated both high efficiency and accuracy due to the dataset's specific characteristics, consisting of 105 features and 90 days of short-term data. The relatively small dataset size, in terms of period, allowed KNN to process the information quickly despite its usual computational challenges. The limited number of features and short duration reduced the complexity that typically hampers KNN's scalability. This scenario allowed KNN to maintain high accuracy, as the model is particularly effective

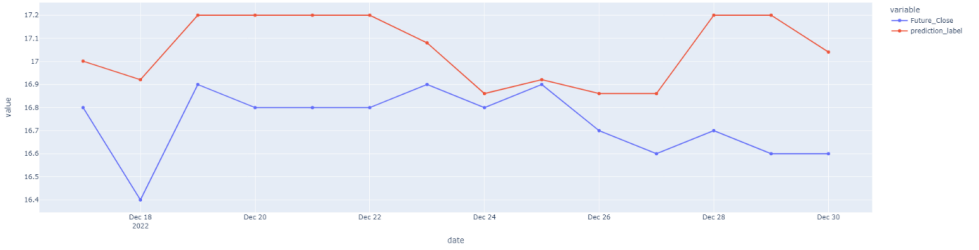
with datasets with heavy-tailed feature distributions, a condition that likely applied to this dataset (Zhao & Lai, 2019). Consequently, even though KNN is generally known for being computationally intensive with large datasets, the specific characteristics of the data in this study enabled it to achieve a desirable balance between computational efficiency and predictive accuracy (Mladenova & Valova, 2023).

3.5. Prediction Results

This research aims to examine the elements that influence Bitcoin price movements and determine their impact on prices. It focuses on predicting Bitcoin's Daily Closing in this scenario. Data were obtained from two types of data sources, and the analysis was initially intended to cover 2022. However, because of Bitcoin's significant volatility as an instrument, the time term was reduced to 90 days for the research from October 2022 to December 2022. The data was split into 85% for training and 15% for testing.

The data was processed, and the selected machine learning algorithm was trained based on the performance evaluation of the algorithms after choosing the sources for analysis and gathering the relevant data. The test data was then used to make predictions using machine learning algorithms and compared with the actual results.

Chart: 3
Model Forecast and Actual Daily Closing



The graph shows the 13-day predictions the model trained using the K Neighbors Regressor algorithm made with 90 days of Bitcoin data from October 2022 to December 2022 and the daily closing prices. The blue line represents the actual Bitcoin closing prices, while the red line represents the predicted prices using a model or algorithm.

As can be observed, Bitcoin closing prices fluctuate significantly and volatilely. The predicted closing prices by the model differ noticeably from the actual prices. The gap between the lines in this graph indicates the magnitude of the difference between the actual and predicted prices and can be used to understand the future trends of Bitcoin prices.

The difference between the model's predictions and daily closing prices alone must fully represent the model's success. However, a high correlation between the two lines is

evident upon examination. Among the displayed predictions, the algorithm made 13 predictions, and 10 of them were correct in terms of price direction. Careful observation also reveals that the model predicts sudden and significant price increases or decreases more accurately.

4. Discussion

This study examines the factors behind Bitcoin price volatility and their impact on the price. Machine learning models used for Bitcoin price prediction are systems that can perform statistical analysis on data to predict future events. However, financial markets, where fast movements and instant changes occur, require short-term data. Long-term data also present some issues. For instance, long-term trends can emerge in financial data, which can limit the information that machine learning algorithms need to learn from the dataset during training. This limitation can hinder the model's ability to make accurate predictions.

Data sources were collected and categorised into two different categories, and initially, the plan was to use all the data throughout 2022 for analysis. However, the time interval was narrowed due to Bitcoin's high volatility. Literature examples supporting this decision include the study by (Ye et al., 2022), where a proposed community deep-learning model achieved an improved accuracy of 88.74% in predicting Bitcoin's following 30-minute prices compared to daily prediction models. Similarly, (Azari, 2019) mentioned that models used for Bitcoin price prediction might lead to significant prediction errors during long-term predictions or sharp price changes as they fail to capture sharp fluctuations. Additionally, their research demonstrated a decrease in the mean squared error of prediction as the size of the testing window increased.

An effort has been made to include as much of the space influencing crypto price movements as possible in the analysis. The variables used in the study can be categorised into two main categories: crypto ecosystem variables and global variables. The crypto ecosystem is rapidly moving towards an alternative financial system. Fundamental crypto exchanges and sub-exchanges that provide hundreds of data streams about Bitcoin and other alternative coins have included variables believed to impact the direction of Bitcoin price in the analysis. This idea is supported by the study (Sovbetov, 2018), where the authors analysed the factors influencing the prices of five popular cryptocurrencies, including Bitcoin, Ethereum, Dash, Litecoin, and Monero. They found that total market prices, trading volume, and volatility significantly affected these five cryptocurrencies in both the long and short terms.

Georgoula et al. (2015), Aggarwal et al. (2019), and Critien et al. (2022) have investigated the relationship between fundamental economic variables, sentiment analysis from Twitter posts, and Bitcoin prices. All the data belonging to the cryptosystem in the first category are related to Bitcoin and consist of 350 different variables. These variables can be categorised under network information, crypto exchange transactions, address data, price data, and social media data.

The increasing volume of Bitcoin within the global financial system leads to its growing dependence on the world economy. If a prediction of Bitcoin's price is to be made, it is necessary to include critical global indices in the analysis. The study by Bouoiyour & Selmi (2017) attempts to identify the relationship between Bitcoin and factors across the fundamental, macroeconomic, financial, speculative and technical determinants. Wang et al. (2016) demonstrated that short-term analysis, oil price, and Bitcoin volume had a limited impact on Bitcoin price, while the stock price index had a more substantial effect. Chen (2023) argues that oil price, ETH price, and U.S. stock market indices (NASDAQ, DJI, and S&P500) all contribute to the impact of the Bitcoin price bubble.

The findings from this study have significant implications for investors and market participants, particularly in understanding the drivers of Bitcoin's price volatility. First and foremost, the need to rely on short-term data for accurate Bitcoin price predictions highlights the importance of timely and agile trading strategies. Investors involved in Bitcoin trading need to be aware that models based on long-term data might miss critical, rapid changes in the market, potentially leading to losses if not accounted for. Therefore, short-term analysis and high-frequency trading could offer better predictive power and risk management.

Identifying two main categories of influencing factors -crypto ecosystem and global variables- suggests that investors must adopt a more holistic approach to market analysis. Investors should monitor on-chain data (such as transaction volumes, network activity, and social media sentiment) and off-chain data (including global economic indicators and financial indices) to make informed decisions. This dual focus could help mitigate risks and capitalise on opportunities arising from micro-level cryptocurrency market dynamics and macro-level global economic conditions.

Additionally, the study's findings that trading volumes, volatility, and market sentiment significantly affect cryptocurrency prices reinforce the importance of behavioural finance in understanding market movements. Investors should consider sentiment analysis as part of their decision-making process, recognising that market sentiment, mainly reflected in social media platforms like Twitter, can substantially and immediately impact Bitcoin's price.

For institutional investors and market makers, the growing interdependence between Bitcoin and traditional financial markets underscores the need to integrate Bitcoin analysis into broader portfolio management strategies. As Bitcoin continues to be influenced by global indices such as stock markets and commodity prices, its role as an alternative asset class becomes more complex, with traditional market factors increasingly playing a role in its price dynamics.

Furthermore, the implications for regulatory bodies and policymakers are equally important. The study indicates that regulatory actions and global economic policies can significantly impact Bitcoin volatility. Understanding the regulatory landscape and anticipating potential policy changes could provide a strategic advantage for market

participants. Investors should stay informed about developments in cryptocurrency regulation across different jurisdictions, as these can lead to abrupt market movements.

Lastly, the study's emphasis on the limitations of long-term predictions due to Bitcoin's unique volatility suggests that investors should exercise caution when relying on traditional investment frameworks. The high volatility and rapid evolution of the cryptocurrency market mean that strategies successful in traditional markets may translate poorly to Bitcoin trading. As such, investors may need to adopt more dynamic and adaptable trading strategies, possibly incorporating advanced machine learning techniques to stay ahead in a fast-moving market.

5. Conclusions

These references indicate the importance of considering various factors from the crypto ecosystem and global variables to understand and predict Bitcoin price movements effectively. This study examines the relationship between Bitcoin and global indices, which is a limited aspect. Forty-six different essential variables worldwide have been used for the analysis. These variables can be categorised under main headings such as bond and stock indices, energy indices, exchange rates, and volatility indices. Unlike similar studies, this research utilises a comprehensive set of variables for a more extensive analysis. This aims to predict the relationship between global indices and Bitcoin prices more accurately.

Including as many sections of the crypto space as possible in the analysis will enable more accurate predictions. Therefore, the chosen artificial intelligence model should be trained with a wide-ranging dataset. This idea raises the question of selecting a suitable model and anticipates that machine learning algorithms can make more precise predictions. Multiple studies have demonstrated that machine learning-based models outperform traditional statistical models in predicting Bitcoin prices. For example, (Mudassir et al., 2020) showcased machine learning-based classification and regression models to predict short-term and medium-term Bitcoin price movements and prices. (Ji et al., 2019) found that deep neural network (DNN) models yielded the best results in predicting price increases and decreases (classification), while long short-term memory (LSTM) models, which incorporate memory, outperformed other prediction models for Bitcoin price forecasting. Dhande et al. (2022) identified deep learning for Bitcoin price prediction and used methods like gradient descent, random forest, and linear regression. A study by Kervanci et al. (2020) emphasises that machine learning methods perform better in Bitcoin price prediction.

With machine learning, data analysis uses a predetermined model, allowing predictions of future events. However, the accuracy of the algorithm used in the data analysis process is crucial. Therefore, further advanced studies have been conducted to select the machine learning algorithm that will provide the most accurate prediction. The first step in the study is to split the data, with 85% training and 15% test data. Then, predictions are made on the test data using 18 machine learning algorithms. The K Nearest Neighbors Regression algorithm with the highest accuracy rate is selected among these algorithms.

The results obtained using machine learning algorithms are a highly effective option among data analysis methods used in various fields. However, appropriate methods must be employed to increase the algorithm's accuracy. The results are pretty successful, with a correct prediction rate of around 90% for the test data. These results indicate that the analysed variables were chosen and processed correctly, and the algorithm parameters were optimised, as shown in the literature.

Aside from the effectiveness of Bitcoin, other digital currencies exist in the market. While not attaining the same level of prevalence as Bitcoin, examining the price movements of alternative cryptocurrencies and their relationships with Bitcoin and the changes over the years in terms of direction and degree can help us understand Bitcoin's volatility. Investigating the relationship between Bitcoin's volatility and macroeconomic factors is very important. Our study drew attention to macroeconomic factors addressed under global indices that affect Bitcoin prices. These factors can be further expanded. For example, examining the effects of economic crises, interest rates, and inflation on Bitcoin volatility can contribute to our understanding of market dynamics.

Investigating the impact of regulatory actions and rules on Bitcoin volatility is also critical. Comparing regulatory approaches and frameworks in different countries and their implications on Bitcoin volatility can help us understand how regulation affects market stability. Understanding the significance of investor behaviour and psychological variables in determining Bitcoin volatility is a topic that needs to be researched further. Future research might look into risk appetite, market sensitivity, herd behaviour, and the impact of significant market participants on Bitcoin volatility. Additionally, while cryptocurrencies such as Bitcoin and Ethereum are not typically considered safe havens for currencies affected by dollar exchange rate volatility - such as the Turkish lira - they have emerged as alternative financial tools during times of geopolitical tension. For instance, during the Russia-Ukraine conflict, Russia leveraged Ethereum and Bitcoin as alternatives to the ruble in natural gas transactions with EU countries, using them as strategic assets.

References

Adhikary, S. & S. Banerjee (2022), "Introduction to Distributed Nearest Hash: On Further Optimizing Cloud Based Distributed kNN Variant", *Procedia Computer Science*, 218, 1571-1580.

Aggarwal, A. et al. (2019), "Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction", *2019 12th International Conference on Contemporary Computing, IC3 2019*.

Aghashahi, M. & S. Bamdad (2022), "Analysis of different artificial neural networks for Bitcoin price prediction", *International Journal of Management Science and Engineering Management*, 18(2), 126-133.

Akinduko, A.A. & A.N. Gorban (2014), "Multiscale principal component analysis", *Journal of Physics: Conference Series*, 490, 012081.

American Institute of Certified Public Accountants (n.d.), *Audit guide : analytical procedures*, <https://books.google.com/books/about/Audit_Guide.html?hl=tr&id=0vlotgEACAAJ>, 08.03.2023.

Andrean, G. (2020), "Determinant of the Bitcoin Prices as Alternative Investment in Indonesia", *Indicators - Journal of Economic and Business*, 1(1), 22-29.

Aung, S.S. et al. (2018), "A high-performance classifier from k-dimensional tree-based Dual-kNN", *IEIE Transactions on Smart Processing and Computing*, 7(3), 184-194.

Azari, A. (2019), *Bitcoin Price Prediction: An ARIMA Approach*, <https://arxiv.org/abs/1904.05315v1>, 08.03.2023.

Bouoiyour, J. & R. Selmi (2017), *The Bitcoin price formation: Beyond the fundamental sources*, <https://arxiv.org/abs/1707.01284v1>, 08.03.2024.

Bouri, E. et al. (2017), "On the Return-Volatility Relationship in the Bitcoin Market Around the Price Crash of 2013", *Economics the Open-Access Open-Assessment E-Journal*, 11(1), 2.

Bouri, E. et al. (2018), "Testing for Asymmetric Nonlinear Short- And Long-Run Relationships Between Bitcoin, Aggregate Commodity and Gold Prices", *Resources Policy*, 57, 224-235.

Charilaou, P. & R. Battat (2022), "Machine learning models and over-fitting considerations", *World Journal of Gastroenterology*, 28(5), 605-607.

Chen, J. (2023), "Analysis of Bitcoin Price Prediction Using Machine Learning", *Journal of Risk and Financial Management*, 16(1), 51.

Ciaian, P. et al. (2015), "The Economics of BitCoin Price Formation", *Applied Economics*, 48(19), 1799-1815.

Cretarola, A. et al. (2017), "A Sentiment-Based Model for the Bitcoin: Theory, Estimation and Option Pricing", *SSRN Electronic Journal*, <https://doi.org/10.2139/ssrn.3042029>.

Critien, J.V. et al. (2022), "Bitcoin price change and trend prediction through twitter sentiment and data volume", *Financial Innovation*, 8(1), 1-20.

Dhande, A. et al. (2022), "Cryptocurrency Price Prediction Using Linear Regression and Long Short-Term Memory (LSTM)", *International Journal for Research in Applied Science and Engineering Technology*, 10(12), 1591-1598.

Dutta, A. et al. (2020), "A Gated Recurrent Unit Approach to Bitcoin Price Prediction", *Journal of Risk and Financial Management*, 13(2), 23.

Dyhrberg, A.H. (2016), "Bitcoin, Gold and the Dollar - A GARCH Volatility Analysis", *Finance Research Letters*, 16, 85-92.

Fil, M. & L. Krištoufek (2020), "Pairs Trading in Cryptocurrency Markets", *Ieee Access*, 8, 172644-172651.

García, D. et al. (2014), "The Digital Traces of Bubbles: Feedback Cycles Between Socio-Economic Signals in the Bitcoin Economy", *Journal of the Royal Society Interface*, 11(99), 20140623.

Georgoula, I. et al. (2015), "Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices", *SSRN Electronic Journal*, <https://doi.org/10.2139/SSRN.2607167>.

Greaves, A. & B. Au (2015), *Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin*.

Gronwald, M. (2019), "Is Bitcoin a Commodity? On Price Jumps, Demand Shocks, and Certainty of Supply", *Journal of International Money and Finance*, 97, 86-92.

Hayashi, S. et al. (2020), "Long-term prediction of small time-series data using generalized distillation", *Transactions of the Japanese Society for Artificial Intelligence*, 35(5), 1-9.

Hong, D. et al. (2018), "Asymptotic Performance of PCA for High-Dimensional Heteroscedastic Data", *Journal of Multivariate Analysis*, 167, 435-452.

Huang, W. et al. (2022), "Time Series Analysis and Prediction on Bitcoin", *BCP Business & Management*, 34, 1223-1234.

Hung, H. et al. (2012), "On Multilinear Principal Component Analysis of Order-Two Tensors", *Biometrika*, 99(3), 56-583.

Jana, R.K. et al. (2021), "A differential evolution-based regression framework for forecasting Bitcoin price", *Annals of Operations Research*, 306(1-2), 295-320.

Ji, S. et al. (2019), "A Comparative Study of Bitcoin Price Prediction Using Deep Learning", *Mathematics*, 7(10), 898.

Johnson, J. (2020), "Bitcoin, Corruption and Economic Freedom", *Journal of Financial Crime*, 27(1), 58-66.

Katubi, K.M. et al. (2023), "Machine learning assisted designing of organic semiconductors for organic solar cells: High-throughput screening and reorganization energy prediction", *Inorganic Chemistry Communications*, 151, 110610.

Kayal, P. & G. Balasubramanian (2021), "Excess Volatility in Bitcoin: Extreme Value Volatility Estimation", *Iim Kozhikode Society & Management Review*, 10(2), 222-231.

Kervancı, I.S. & F. Akay (2020), "Review on Bitcoin Price Prediction Using Machine Learning and Statistical Methods", *Sakarya University Journal of Computer and Information Sciences*, 3(3), 272-282.

Khatun, M. & S. Siddiqui (2023), "Estimating Conditional Event Probabilities with Mixed Regressors: a Weighted Nearest Neighbour Approach", *Statistika*, 103(2), 226-234.

Liu, X. & X. Zhang (2023), "The Analysis of the Influencing Factors of Virtual Currency Price Based on Multiple Regression Method", *Frontiers in Business Economics and Management*, 7(1), 156-159.

Mladenova, T. & I. Valova (2023), "Classification with K-Nearest Neighbors Algorithm: Comparative Analysis between the Manual and Automatic Methods for K-Selection", *International Journal of Advanced Computer Science and Applications*, 14(4), 396-404.

Mudassir, M. et al. (2020), "Time-Series Forecasting of Bitcoin Prices Using High-Dimensional Features: A Machine Learning Approach", *Neural Computing and Applications*, https://doi.org/10.1007/s00521-020-05129-6.

Munim, Z.H. et al. (2019), "Next-Day Bitcoin Price Forecast", *Journal of Risk and Financial Management*, 12(2), 103.

Oğuzlar, A. (2003), "Veri Ön İşleme", *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21, 67-76.

Oyeyemi, G. et al. (2015), "Comparison of Outlier Detection Procedures in Multiple Linear Regressions". *American Journal of Mathematics and Statistics*, 5(1), 37-41.

Pano, T. & R. Kashef (2020), "A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets During the Era of COVID-19", *Big Data and Cognitive Computing*, 4(4), 33.

Pele, D.T. & M. Mazurencu-Marinescu-Pele (2019), "Metcalfe's Law and Log-Period Power Laws in the Cryptocurrencies Market", *Economics the Open-Access Open-Assessment E-Journal*, 13(29), 1-26.

Richardson, M. (2009), *Principal component analysis*, <http://aurora.troja.mff.cuni.cz/nemec/idl/09bonus/pca.pdf>, 02.03.2024.

Roy, S. et al. (2019), "Bitcoin Price Forecasting Using Time Series Analysis", *21st International Conference of Computer and Information Technology, ICCIT 2018*.

Rubio, G. et al. (2009), "Parallelization of the nearest-neighbour search and the cross-validation error evaluation for the kernel weighted k-nn algorithm applied to large data dets in matlab", *Proceedings of the 2009 International Conference on High Performance Computing & Simulation*, Leipzig, Germany, 145-152.

Sovbetov, Y. (2018), "Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litcoin, and Monero", *Journal of Economics and Financial Analysis*, 2(2), 1-27.

Tappin, B.M. et al. (2021), "Rethinking the Link Between Cognitive Sophistication and Politically Motivated Reasoning", *Journal of Experimental Psychology General*, 150(6), 1095-1114.

Taylan, P. (2019), "On foundations of estimation for nonparametric regression with continuous optimization", in: F.P. Garcia Marquez (ed.), *Handbook of Research on Big Data Clustering and Machine Learning* (177-203), IGI Global Scientific Publishing.

Vatcheva, K. et al. (2016), "Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies", *Epidemiology Open Access*, 6(2), 227.

Virk, D.S. (2017), "Prediction of Bitcoin Price using Data Mining", *Doctoral Dissertation*, National College of Ireland, Dublin.

Wamuyu, P.K. (2022), "Use of Cloud Computing Services in Micro and Small Enterprises: A Fit Perspective", *International Journal of Information Systems and Project Management*, 5(2), 59-81.

Wang, J. et al. (2016), "An Analysis of Bitcoin Price Based on VEC Model", in: *Proceedings of the 2016 International Conference on Economics and Management Innovations* (180-186), Atlantis Press.

Wu, C.H. et al. (2019), "A new forecasting framework for bitcoin price with LSTM", *IEEE International Conference on Data Mining Workshops, ICDMW* (168-175), 2018-November.

Ye, Z. et al. (2022), "A Stacking Ensemble Deep Learning Model for Bitcoin Price Prediction Using Twitter Comments on Bitcoin", *Mathematics*, 10(8), 1307.

Zaman, S. & B. Ahmed (2019), "Hybrid Subspace Detection Based on Spectral and Spatial Information for Effective Hyperspectral Image Classification", *International Journal of Computer Applications*, 178(41), 37-43.

Zhao, J. (2022), "Do Economic Crises Cause Trading in Bitcoin?", *Review of Behavioral Finance*, 14(4), 465-490.

Zhao, P. & L. Lai (2019), "Minimax Regression via Adaptive Nearest Neighbor", *IEEE International Symposium on Information Theory - Proceedings* (1447-1451), 2019-July.

Zhou, S. (2019), "Exploring the Driving Forces of the Bitcoin Currency Exchange Rate Dynamics: An EGARCH Approach", *Empirical Economics*, 60, (557-606).