

Cryptocurrency Price Prediction Using Social Media Sentiment Data

Bashir Sani Sabo^{1*} and Bashir Shehu Galadanci²

¹Department of Information and Communications Technology, Jigawa State College of Education, Gumel

²Department Of Software Engineering, Faculty of Computer Science and Information Technology, Bayero University, Kano

bsanisabo@gmail.com^{1*}, bsgaladanci.se@buk.edu.ng²

*Corresponding Author

Abstract

Due to the boom and bust of Bitcoin in 2017, there has been increased interest in cryptocurrency by organizations, academia, businesses, industries, and individuals with many considering cryptocurrencies as investment assets rather than just digital coins. However, due to the sudden drop in Bitcoin price, academic researchers' attention has been drawn to developing a prediction model on which to base investment. Although there is existing research work done on predicting cryptocurrency price, very few consider the two major data sources (Social media sentiment data and Historical price data) for comparison purposes, and to the best of our knowledge, none has considered merging the two data sources for price prediction. This paper proposed a novel cryptocurrency price prediction model based on tree-based machine learning algorithms namely: Decision Tree, Gradient Boosting Regressor, CATBOOST Regressor, and Random Forest using social media sentiment data (tweets) and cryptocurrency historical price data for the three most popular cryptocurrencies: Bitcoin, Ethereum, and XRP. We scrap data from Twitter with a sample size of 100 tweets daily for 2 years. We then pre-process the data and run sentiment analysis on the 100 tweets to arrive at a single sentiment for the 100 tweets. The historical price data obtained from Yahoo Finance is then merged with the tweet sentiment for prediction using our machine learning algorithms. The proposed model performs best with Gradient Boosting Regressor with an accuracy of 98.94%, Mean Absolute Error (MAE) of 0.00400, Means Squared Error (MSE) of 0.00028, and R2 score of 0.93 for XRP. The gradient Boosting regressor also performs best for the BTC model with an accuracy of 97.65%, MAE: 482.2100, MSE: 508309.8100, RMSE: 712.9600, and an R2 score of 0.96. The same is also applied to the Ethereum model with an accuracy of 99%, MAE: 12.7133, MSE: 295.1983 RMSE: 17.1813, and R2 score of 0.99.

Keywords: Cryptocurrency, Decision Tree, Gradient Boosting Regressor, CATBOOST Regressor, Random Forest.

1. Introduction

Blockchain as the building blocks of cryptocurrency is based on Cryptography which is simply referred to as the science of hiding information against third-party intrusion. Cryptography is mostly used in blockchain to protect user privacy and secure transaction information.

Cryptocurrencies are digital assets that are designed to work as a medium of exchange that uses cryptography to secure their transactions, control the creation of additional units, and verify the transfer of assets (Aytan et al., 2019). It is rising in popularity and being widely adopted across the globe where millions of people are estimated to be actively trading in cryptocurrencies today, and many are already using cryptocurrency to pay for items such as hotels, games, and even their rent (Chang et al., 2023). Bitcoin is the most accepted cryptocurrency in the world, which makes it attractive for investors and traders. Bitcoin stands as the first decentralized cryptocurrency, primarily based on blockchain technology (Hachicha et al., 2023). It has helped remarkably in facilitating electronic payments among individuals without the need to go through the intermediary of a third party. Since its emergence in 2009 as

the first digital cryptocurrency, more than 5,600 cryptocurrencies have surfaced (Tanwar et al., 2021). Bitcoin was the first devised digital currency and remains the market-leading cryptocurrency in terms of stock market capitalization (Fakhfekha & Ahmed, 2019). Some of these cryptocurrencies are designed for specific purposes such as in-app purchases, game rewards, and so on while others are targeted for general transactions most of which are aimed to serve like traditional currencies (Higbee, 2018).

The Bitcoin price is selected among the available digital currencies as it is considered the single largest cryptocurrency asset with more than 410 billion worth of market value (Zhong et al., 2023; Basher & Sadorsky, 2022; Tanwar et al., 2021). Although there are other variants of cryptocurrency, the choice of Bitcoin among the alternatives is mainly informed by its prominence as the most traded cryptocurrency coupled with the fact that it has sufficiently large datasets when compared to other cryptocurrencies. (Afeez et al., 2018). Investors make use of many online discussion channels when deciding to make investments in stock markets. Such information is presented within Financial Discussion Boards (FDBs), news corporations (e.g., Financial Times), broker agency websites, and social media platforms. Recently, Twitter has become a popular platform for investors to disseminate stock market information and discussion (Brown, 2012). Many large organizations are also using Twitter as a platform to obtain and share information relating to their products and services (Evans et al., 2019).

Cryptocurrencies have gained significant investment attention in recent times, attracting individuals, large tech companies, and financial institutions. Facebook, for instance, introduced its cryptocurrency called Libra. It is, however, crucial to note that all cryptocurrencies are highly volatile, meaning their prices can experience dramatic fluctuations within short timeframes. For instance, in 2017, the value of Bitcoin fluctuated between \$900 and \$20,000. This volatility makes cryptocurrency investment inherently risky, as the price may change within a very short period (Abraham et al., 2018; Tanwar et al., 2021; Chang et al., 2023). Consequently, there is a need for a model that can forecast cryptocurrency prices. While the literature offers a range of approaches to address this challenge (Abraham et al., 2018; Alghamdi et al., 2022; Basher & Sadorsky, 2022; Chang et al., 2023; Chen, et al., 2020; Hosure et al., 2022; Khedr, et al., 2021; Salman & Ibrahim, 2020; Tanwar, et al., 2021; Wardak & Rasheed, 2022; Zhong et al., 2023), many existing models have not adequately considered the unique characteristics of the data used for price prediction. As a result, these models often perform poorly during extended market downturns and struggle to adapt to predicting the prices of new coins.

This study, therefore, aims to address these limitations by developing a model that predicts cryptocurrency prices and market trends. The model combines historical price data with social media sentiment data (specifically, Tweets) to train decision trees (DT), gradient boosting regressors (GBR), random forests (RF), and CATBOOST machine learning models. By incorporating sentiment analysis from social media alongside historical price data, this research seeks to enhance the accuracy and robustness of cryptocurrency price prediction models.

2. Related Works

In this section, we discuss the contribution of related works based on the type of machine learning technique utilized and their limitations. The section is majorly divided into two sub-sections: cryptocurrency price prediction using conventional machine learning models and cryptocurrency price prediction using deep learning.

2.1 Cryptocurrency Price Prediction Using Conventional Machine Learning Models

A literature survey conducted by Khedr et al. (2021) deduced that most of the research conducted on predicting cryptocurrency prices was based on machine learning models rather than traditional statistical techniques. The work further asserted that 45.97% of the sampled papers are based on statistical techniques while 54.03% belong to machine learning techniques. Also, among the machine learning-based papers, 29.78% are based on Deep Learning/Reinforced Learning categories. The research recommended the use of more sophisticated prediction algorithms, feature engineering techniques, and advanced ensemble methods. It also recommended the use of social media and long short-term memory (LSTM) models for more price prediction.

Chen et al. (2020) proposed a machine learning model including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Quadratic Discriminant Analysis (QDA), and Support vector machine (SVM) for bitcoin 5-minute interval price prediction and it was found that the models are superior to statistical methods with accuracy reaching 67.2%. The research suggested limitations in its data sources and that further studies should focus on more useful elements of the sentiment using text mining and analysis of social networks. Jaquart et al. (2021) proposed a model based on different machine learning algorithms to evaluate different prediction models using four categories of features (Technical, Asset-based, Blockchain-based, sentiment-/interest-based). It was concluded from the research results that the Recurrent Neural Network (RNN) and gradient Boosting Classifier (GBC) accuracy in predicting the bitcoin market outperformed the accuracy of other models. It further asserts that the models were able to predict binary market movement with accuracies ranging from 50.9% to 56.0%. It is concluded that the research focus is now on models using machine learning for cryptocurrency price prediction.

Accordingly, Hosure et al. (2022) proposed a hybrid machine learning model including Support vector regression (SVR) and Autoregressive Integrated moving average (ARIMA) for one month's Cryptocurrency price prediction to be superior to other machine learning with accuracy reaching 90%. The proposed model used the merge method to combine both data frames considering the test values of both algorithms that are the same. However, the research cannot be used to predict new coin that has been launched recently and has less historical data because it only focuses on cryptocurrency that has a larger dataset.

While there is existing literature on forecasting bitcoin prices, there are several important gaps in the literature that lead to research questions such as *How important are macroeconomic variables like interest rates, inflation, and market volatility for forecasting bitcoin prices?* In Basher and Sadorsky (2022), the use of tree-based ensemble machine learning classifiers like bagging and random forest was proposed to predict bitcoin and gold price direction. This was because of the research findings that tree-based ensemble machine learning methods have high accuracy when predicting stock price direction (Basher & Sadorsky, 2022). It was then concluded from the result that technical indicators are important predictors for both bitcoin and gold price and their importance tend to be higher than that of the macroeconomic variables considered. However, this research is Bitcoin-centric and only predicts the price direction and not the actual price of the cryptocurrency.

2.1.1 Prediction Using Conventional Machine Learning Models with Sentiment Analysis

Lamon et al. (2020) assert that there is a relationship between media sentiment and the price of cryptocurrency coins. The work was able to answer the research question *Can sentiment analysis on news article headlines and/or social media posts produce accurate predictions prices of Bitcoin, Litecoin, and Ethereum?* The proposed model was able to predict a 43.9% price increase correctly and 61.9% price decrease correctly for Bitcoin and a 75.8% price increase for Ethereum.

Additionally, Abraham et al. (2018) proposed a model using tweet volumes and tweet sentiment analysis to predict changes in Bitcoin and Ethereum prices. It was concluded that tweet volume is more reliable than tweet sentiment as a predictor of cryptocurrency price direction for Bitcoin and Ethereum. However, due to the model's limitation in terms of the machine learning algorithm used, it was recommended that more complex models and not just linear ones should be considered

2.2 Cryptocurrency Price Prediction Using Deep Learning Models

Hegazy and Mumford (2016) were able to predict with 57% accuracy the actual prices of bitcoin using historical prices and supervised learning methods. Jiang and Liang (2017) took a different approach using a deep reinforcement learning algorithm to make trading decisions on a bitcoin portfolio that resulted in 10 times increase in the portfolio's value. It should however be noted that these works were bitcoin centric and used purely historical data. Moving away from historical price data; in price prediction, some researchers have used tweets and social media data to predict market directions. Most notably Laskowski and Kim (2016) as well as Dehua et al. (2018) were able to find a direct correlation between the number of mentions on Twitter and the prices of Bitcoin and Dogecoin.

Colianni et al. (2015) were able to predict the market directions of a handful of cryptocurrencies including bitcoin using supervised learning algorithms for sentiment analysis of tweets. Stenqvist and Lönnö (2017) were able to achieve 79% accuracy in predicting bitcoin prices. Due to the limitation of some of the proposed machine learning models to predict the price of cryptocurrencies in real-time, Tanwar et al. (2021) proposed a deep-learning-based hybrid model using Gated Recurrent Units (GRU) and LSTM to predict the price of two cryptocurrencies namely, Litecoin and Zcash. The result obtained from the model shows that it outperformed the accuracy of the existing models. However, the research recommends the addition of sentimental factors such as Twitter and Facebook posts and messages for feature cryptocurrency price prediction. Earlier on, Salman and Ibrahim (2020) proposed a model using machine learning-based Neural Networks. The work further analyzed bitcoin prices using machine learning and sentiment analysis. The proposed model was able to predict the prices of bitcoin using a machine learning-based neural network with an accuracy of 94.89% under all circumstances of technical trade indication. However, the research was limited to developed countries with access to technology.

Wardak and Rasheed (2022) proposed a deep learning-based model using LSTM Recurrent Neural Network to predict the price of Bitcoin using a variety of factors that influence its value. The research utilizes the historical dataset of bitcoin over the last five years and was able to achieve an accuracy of 95.7%, Validation loss of 0.00065, and Root Mean Square of 0.05. The result was compared with other existing methods such as RNN, Multilayer Perception (MLP), ARIMA, and other LSTM models, and was found that LSTM with hyperparameter tweaking is one of the most effective methods for predicting bitcoin prices. Do-Hyung et al. (2019) proposed a model for the classification of cryptocurrency price time series using LSTM. The paper proposed the use of historical cryptocurrency price time series data for prediction. The data is encoded into the three-dimensional price tensor representing the past price changes of cryptocurrencies and was able to achieve a performance improvement of about 7% compared to using the Gradient Boosting model. However, the researchers recommend the use of additional features to build a better LSTM model for performance improvement.

Furthermore, Alghamdi et al. (2022) proposed and developed two models to predict cryptocurrency prices. One model was to forecast cryptocurrency prices using LSTM deep learning technique and the other model was to analyze Twitter user’s sentiments using SVM and Naïve Bayes. The research aimed to determine the relationship between investors' sentiment and the volatility of cryptocurrency prices. It was found that based on the sentiment analysis model, the SVM model perform better than Naïve Bayes with an accuracy of 95.59% and 93.95% respectively. Also, based on the forecasting model, the model was able to achieve a Mean Absolute error (MAE) of 0.2528 and a Root Mean Squared Error (RMSE) of 0.5028. However, the research was only able to utilize Twitter data for five months. Also, the research performed predictions on only two cryptocurrencies, Bitcoin (BTC) and Ethereum (ETH). Table 1 presents a summary of related works in terms of the objective, models proposed, accuracy achieved, and limitations.

Table 1: Summary of Related Works

References	Objectives	Models	Accuracy	Limitations
Chen et al. (2020)	To propose an ML model for Bitcoin 5-minute interval price prediction	RF, XGBoost, QDA, and SVM	67.2%	The research suggested limitations in its data sources and only bitcoin was considered.
Jaquart et al. (2021)	To propose a model based on different machine learning algorithms to evaluate different prediction models	RNN, GBC, LSTM, RF, GRU, feedforward neural network, logistics regression (LR)	50.9% to 56.0%	The research only considered the prediction of the Bitcoin price

Hosure et al. (2022)	To propose a Hybrid ML model for one month's cryptocurrency price prediction	SVR and ARIMA	90%	Less historical data was used and the proposed model cannot be used to predict new recently launched coin
Hegazy and Mumford (2016)	Prediction of the actual prices of bitcoin using historical prices and supervised learning methods	LR, recurrent reinforcement learning (RRL), gaussian discriminant analysis (GDA), weighted linear regression, and boosted classifiers	57%	The research only considered the prediction of the Bitcoin price
Colianni et al. (2015)	To predict the market directions of Bitcoin using supervised learning algorithms for sentiment analysis of tweets	LR, Naïve Bayes, and SVM	76.23%	The research only considered the prediction of bitcoin.
Stenqvist and Lönnö (2017)	To analyze Bitcoin-related tweets for sentiment fluctuations to predict a price change	1h-3, 30min-4, 45min-3, 4h-1, 2h-2, 5min-2, 15min-4	79%	Only bitcoin was considered in the work.
Salman and Ibrahim (2020)	To propose a ML-based Neural Networks	-	94.89%	Only bitcoin was considered in the work.
Wardak and Rasheed (2022)	To propose a deep learning-based model to predict the price of Bitcoin	LSTM Recurrent Neural Network	95.7%	Only bitcoin was considered in the work.
Alghamdi et al. (2022)	To propose a deep learning model to predict the prices of Bitcoin and Ethereum	SVM and Naïve Bayes	95.59% and 93.95% respectively	Only five months Twitter dataset were utilized

From the review conducted, it can be deduced that there are no general-purpose models that can be easily extended for the use of predicting new cryptocurrency prices as various models have proven suitability for different coins. Furthermore, the models in the literature perform badly over/after long periods of price decrement, and this can be attributed to the overtraining based on one predictive set of market features while turning a blind eye to other market predictive features alongside the conservative nature of said autoregressive models. Additionally, from Table 1, it can be seen that most of the existing research work focused on the prediction of Bitcoin currency. Therefore, there is a need for a price prediction model that will handle these limitations.

3. Methodology

In this paper, the three most popular cryptocurrencies, Bitcoin, Ethereum, and XRP, were considered and four different models were developed using four different supervised machine learning techniques which include Decision Tree (DT), Random Forest (RF), CATBOOST, and Gradient Boosting Regressor (GBR). Two types of

datasets were collected namely, Historical Price data from Yahoo Finance and Sentiments data from tweets. The first dataset which is Historical price data has the following metadata.

1. Daytime – date and time of the day
2. Open price – The opening price of the coin for that day
3. High Price – The highest price reached by the coin on that day.
4. Low Price – The lowest price of the coin
5. Closing Price – The last price for that day
6. Volume of Coin – The total number of coins traded on that day.

The sentiment dataset was obtained by running a custom library called sncrape. Sncrape is an advanced Twitter scraping tool written in Python that allows for scraping tweets from Twitter without using Twitter API. The justification for the use of this library is as follows

1. It allows users to scrape tweets from specific users.
2. Scrape Tweets relating to certain topics.
3. Allows easy access to Hashtags and trends.
4. Allow for the Sorting of Sensitive Information from tweets.

In addition to these, the sncrape library can fetch almost all tweets, can be used anonymously, and has no rate limitations.

3.1 Proposed Model

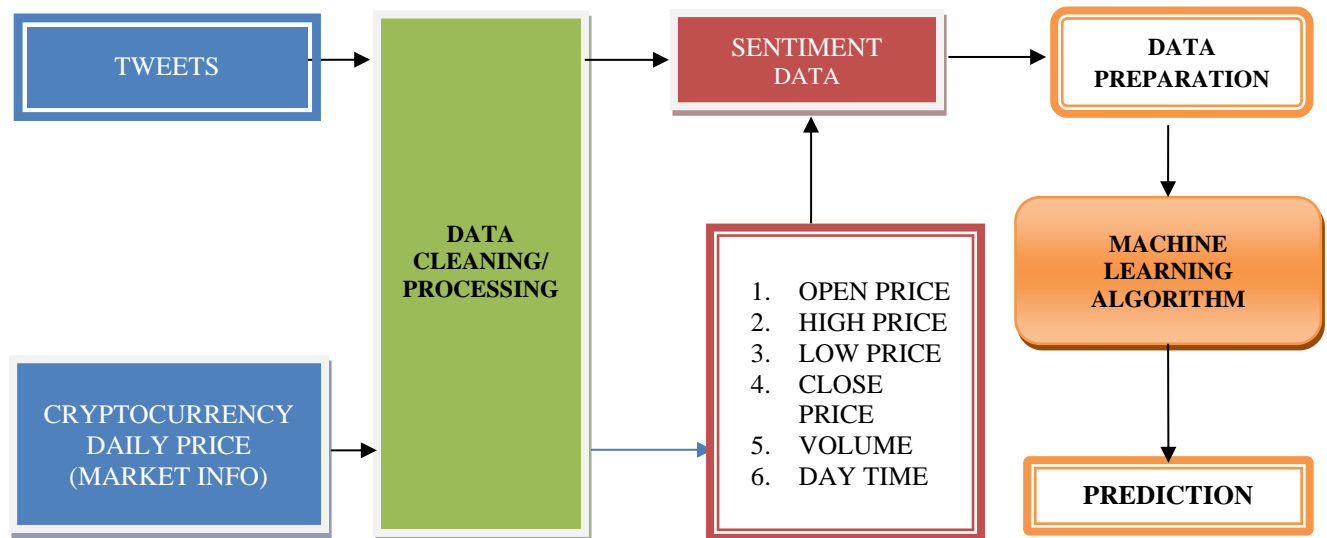


Figure 1: Proposed Model

Based on the proposed model presented in Figure 1, the Python library, sncrape, was used to scrap data from Twitter by specifying the start and stop date which were 01/01/2020 and 31/12/2022 respectively. A sample size of 100 tweets per day was utilized that is, pulling 100 tweets per day for a total of 365 days each year. At the end of our data scraping, the result was multiple CSV files where only the CSV file with the stop date on it was needed. Preprocessing of data obtained, such as data arrangement, cleaning, and wrangling, was conducted to properly represent the data in a format that can be useful for machine learning algorithms. After this data preparation was done such as data merging and further formatting. Finally, model training and testing were conducted where 80% of the prepared dataset was used to train the four different machine learning algorithms while the remaining

20% was utilized for testing. The result of these models was then compared for better performance in terms of the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R2 score, and Mean Absolute Percentage Error (MAPE).

4. Results and Discussion

In this section, we discuss the result of our developed models using the three most widely used cryptocurrencies. The section is divided into three subsections: the BTC model, the Ethereum model, and the XRP model, with each subsection discussing the result of model performance on the cryptocurrency.

4.1 BTC Model

This section provides the results and evaluation of the four tree-based machine-learning models implemented using the BTC dataset. Table 2 gives the summary of the performance of the developed BTC models in terms of the selected metrics.

Table 2: Evaluation Metrics for the BTC model

Models	DT	GBR	RF	CATBOOST
MAE	598.98	482.21	751.83	693.89
MSE	561724.89	508309.81	938368.13	730152.50
RMSE	872.76	712.96	968.69	854.48
R2 SCORE	0.94	0.96	0.92	0.94
MAPE	2.85	2.35	3.62	3.39
ACCURACY	97.15%	97.65%	96.38%	96.61%

Close observation from Table 2 shows that the Decision Tree Machine learning technique performs very well with an Accuracy of 97.15% and a Mean Absolute error (MAE) of approximately 598.98. Figure 2 presents the comparison between the actual and predicted price of the BTC model using the Decision Trees model.

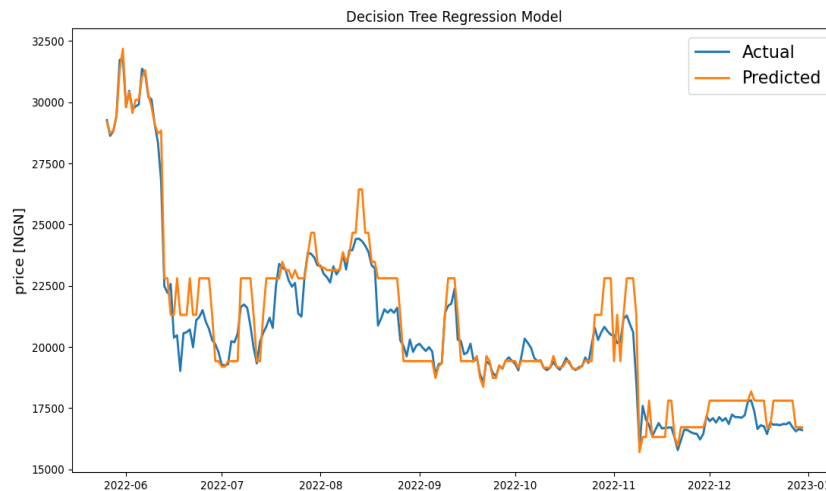


Figure 2: BTC Model Plot Using Decision Tree

Compared to the BTC model using Decision Trees, the Gradient Boosting Regressor Machine Learning algorithm performed excellently with an Accuracy of 97.65% and, a mean absolute error of approximately 482.21. Figure 3 presents the comparison between the actual and predicted prices.

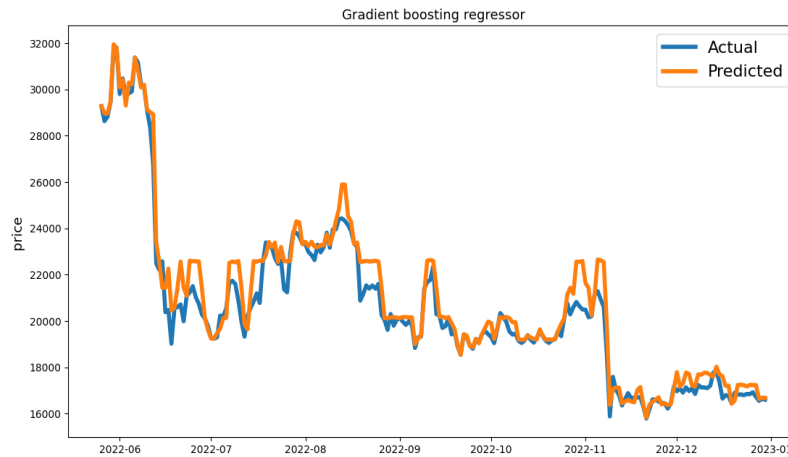


Figure 3: BTC model Plot using Gradient Boosting Regressor

For the BTC model using the Catboost Ensemble ML algorithm, an accuracy of 96.61% and a Mean Absolute Error of 693.89 were obtained as seen in Table 2. Figure 4 presents the comparison of the actual and predicted prices.

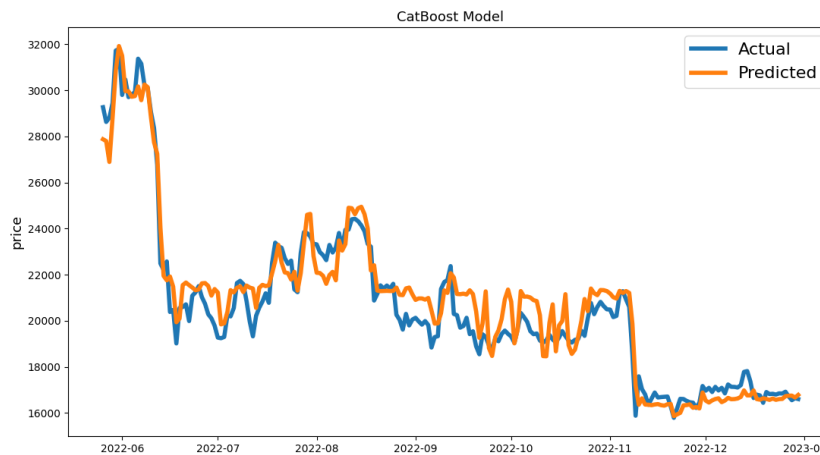


Figure 4: BTC model Plot using Catboost

Random Forest was also utilized for the BTC model, and from Table 2, it has an Accuracy of 96.38% and the highest MAE value of approximately 751.83. Figure 5 presents the comparison of the actual and predicted prices.

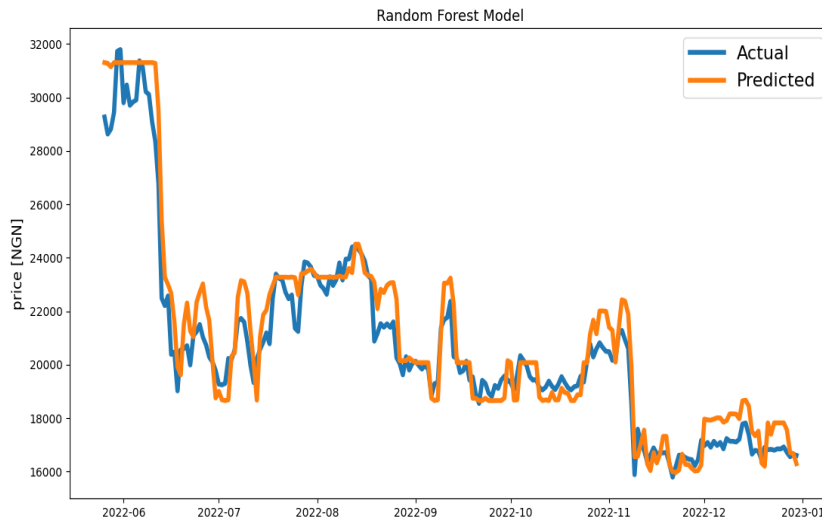


Figure 5: BTC model Plot using Random Forest

4.2 Ethereum Market Model

This section provides the results and evaluation of the four tree-based machine learning models implemented using the Ethereum dataset. Table 3 provides a summary of the performance of the developed Ethereum models.

Table 3: Summary of Evaluation metrics for the Ethereum model

Models	DT	GBR	RF	CATBOOST
MAE	18.23	12.71	29.58	20.16
MSE	678.87	295.20	1463.25	769.05
RMSE	26.06	17.18	38.25	27.73
MAPE	1.77	1.00	2.14	1.52
R2 SCORE	0.98	0.99	0.97	0.98
ACCURACY	98.23%	99%	97.86%	98.48%

Like the BTC model, it can be seen from Table 3 that the decision trees model performs well with an accuracy of 98.23%, a Mean Absolute error of approximately 18.23, and an R2 score of approximately 0.98. Figure 6 presents the comparison of the actual and predicted prices.

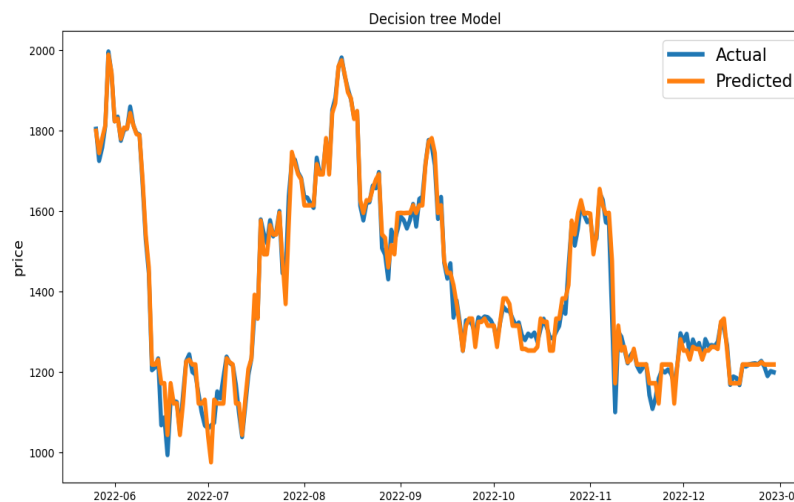


Figure 6: Ethereum model Plot using Decision Tree

The Catboost Regressor model also performed well with an accuracy of 98.48%, however, compared to the decision trees, it has a higher mean absolute error value of 20.16 but the same R2 score of 0.98. A comparison of the actual and predicted prices using this model is presented in Figure 7.

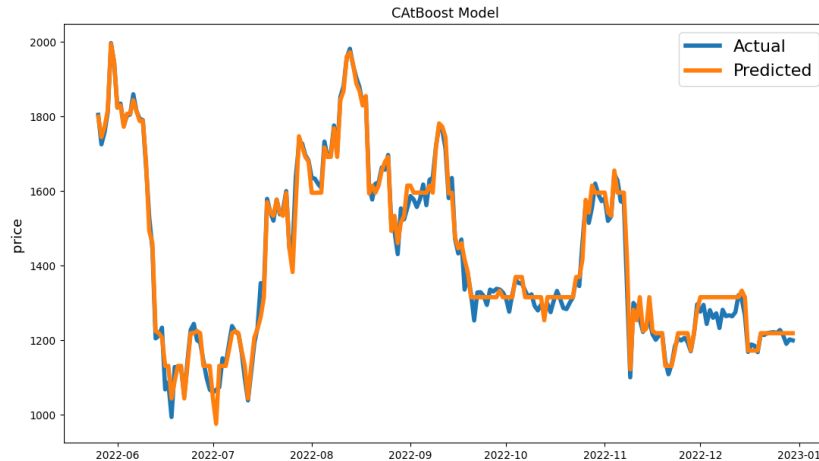


Figure 7: Ethereum model Plot using Catboost Regressor

Compared to the DT and Catboost, the Gradient Boosting Regressor performs excellently with an accuracy of 99% and has the lowest mean absolute error value of 12.71. Additionally, GBR has a higher R2 score of 0.99. Figure 8 presents the comparison between the actual and predicted price values.

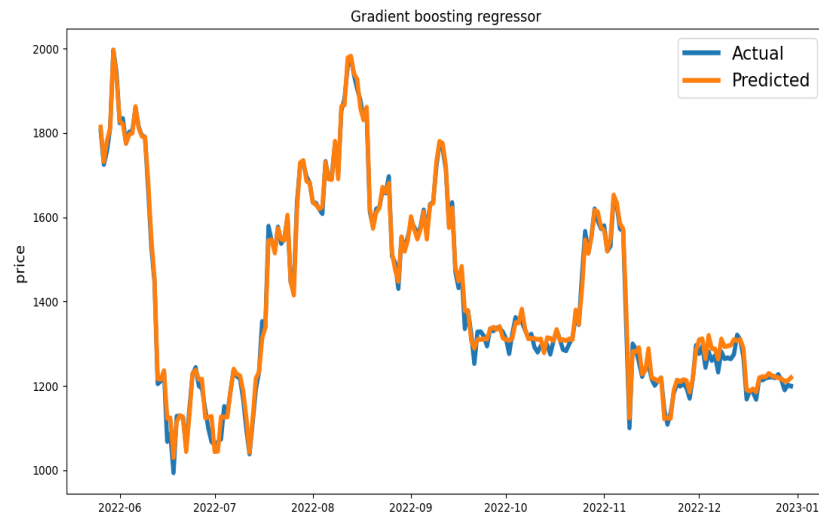


Figure 8: Ethereum model Plot using Gradient Boosting Regressor

Finally, for Ethereum model is the Random Forest model, which has an accuracy of 97.86%. It also has the highest MAE value compared to the other models but with the lowest R2 score of 0.97. Figure 9 presents the comparison of actual and predicted price values for different dates and times.

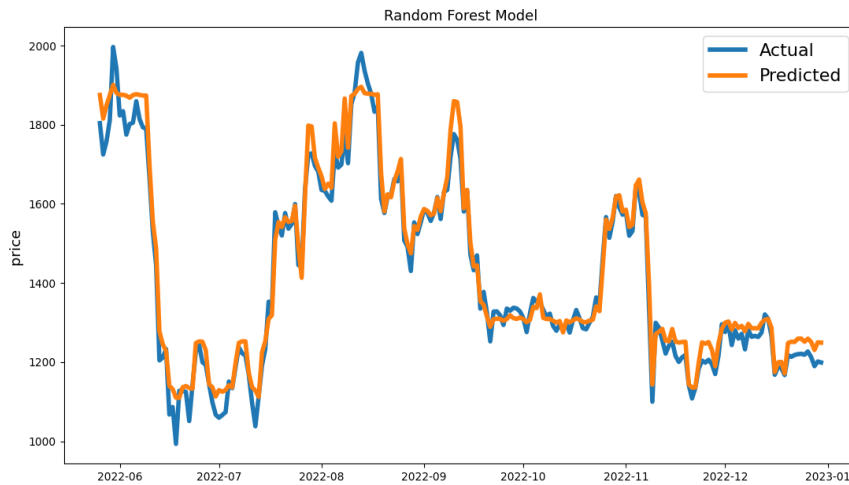


Figure 9: Ethereum Model Plot using Random Forest Regressor

4.3 XRP Model

This section provides the results and evaluation of the four tree-based machine learning models implemented using the XRP dataset. Table 4 provides a summary of the performance of the developed XRP models.

Table 4: Summary of Evaluation Metrics for XRP Models

Models	DT	GBR	RF	CATBOOST
MAE	0.0085	0.0041	0.0047	0.0130
MSE	0.0001	0.000017	0.000022	0.0002
RMSE	0.0141	0.0055	0.0070	0.0162
MAPE	2.26	1.06	1.21	3.06
R2 SCORE	0.93	0.99	0.98	0.91
ACCURACY	97.74%	98.94%	98.79%	96.94%

From Table 4, the XRP model using decision trees was seen to perform well with an Accuracy of 97.74%, a Mean Absolute error of approximately 0.0085, and an R2 score of approximately 0.93. Figure 10 presents the comparison of the actual and predicted price values for different dates and times.

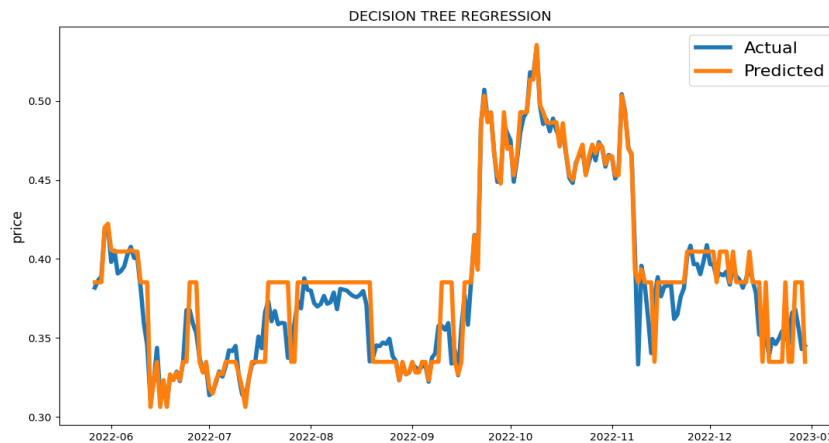


Figure 10: XRP Model Plot using Decision Tree

Catboost model, compared to Decision Tree, performed better in terms of the mean absolute error with a value of 0.0130 however, has a higher R2 score of 0.91 and an accuracy of 96.94% as seen in Table 4. Figure 11 presents the comparison of the actual and predicted price values for different dates and times.

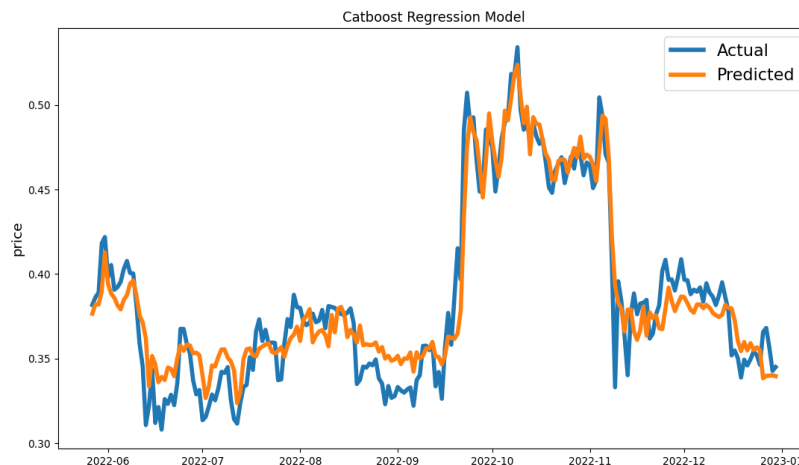


Figure 11: XRP Model Plot using Catboost

For the Gradient Boosting Regressor, the model performs excellently with an accuracy of 98.94%, a Mean Absolute error of approximately 0.0041, and an R2 score of 0.99. Figure 12 presents the comparison of the actual and predicted price values for different dates and times.

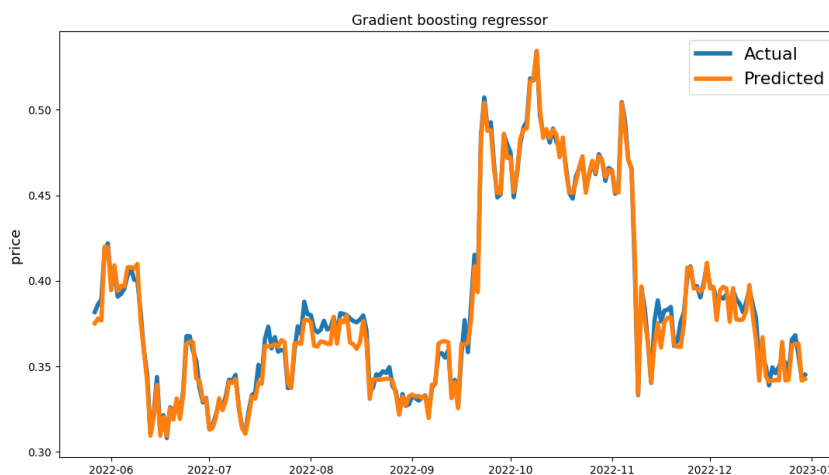


Figure 12: XRP Model Plot using Gradient Boosting Regressor

Finally, for Random Forest, the model performs excellently with an accuracy of 98.79% Mean Absolute error of approximately 0.0047, and an R2 score of 0.98. Figure 13 presents the comparison of the actual and predicted price values for different dates and times.

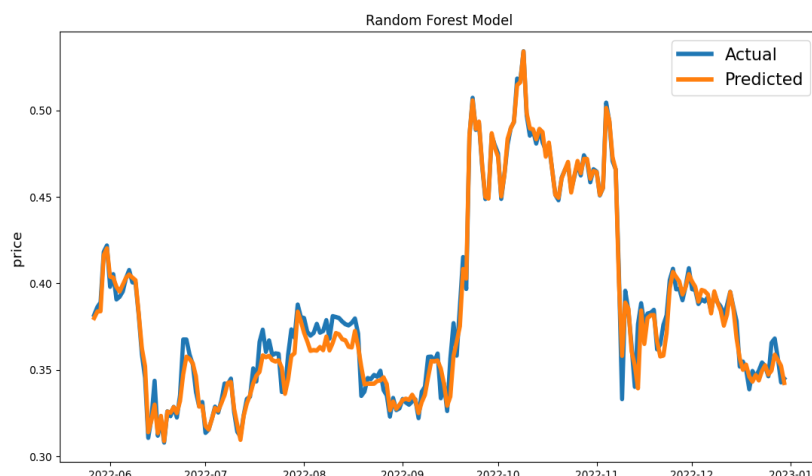


Figure 13: XRP model plot Using Random Forest

In summary, our Ethereum model using **GBR** performs best with an accuracy of 99%, Mean absolute error of approximately 12.71, Mean squared error of approximately 295.20, Root Mean Squared Error of 17.18, Mean Absolute Percentage Error of 1.00 and R2 score of 0.99 and. To the best of our knowledge, very few of the research works were able to get close to the accuracy obtained using our developed model.

5. Conclusion

Due to its volatile nature, cryptocurrency has attracted lots of attention from Individuals, Businesses, and Institutions for investment purposes. As a result of this, there is a need for an accurate cryptocurrency price prediction model for asset allocation. In this research, we developed four models using four different tree-based machine-learning algorithms (DT, RF, CATBOOST, and GBR). The models were trained and tested using two types of datasets. The first dataset was scraped from Twitter while the second dataset was obtained from the Yahoo finance database. The Performance of the four models was evaluated and results showed that the Ethereum Model using the Gradient Boosting Regressor machine learning technique outperformed other models with an accuracy of 99%, mean absolute error of 12.71, mean squared error of 295.20, Root Mean Squared error of 17.18 and R2 score of 0.99. Therefore, based on our results the Gradient Boosting Regressor performs best in all the models developed. Further research work should leverage more sources of data such as Facebook and News headlines. Also, the use of unsupervised machine learning techniques and deep learning should be considered for large amounts of datasets.

References

- Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. *SMU Data Science Review*.
- Afees A., S., Kazeem, I., & Akannid, L. O. (2018). Improving the predictability of stock returns with Bitcoin prices. *North American Journal of Economics and Finance*.
- Alghamdi, S., Alqethami, S., Alsubait, T., & Alhakami, H. (2022). Cryptocurrency Price Prediction Using Forecasting and Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, 891-900.
- Aytan, A., Karasu, S., & Stelios, B. (2019). Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. *Chaos, Solitons and Fractals*, 325-336.
- Basher, A. S., & Sadorsky, P. (2022). Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility. *Machine Learning with Applications*.
- Brown, E. (2012). Will Twitter Make You a Better Investor? A Look at Sentiment, User Reputation and Their Effect on the Stock Market. *SAIS 2012 Proceedings*. <http://aisel.aisnet.org/sais2012/7/>

- Chang, T.-J., Lee, T.-S., Yang, C.-T., & Lu, C.-T. (2023). A ternary-frequency cryptocurrency price prediction scheme by ensemble of clustering and reconstructing intrinsic mode functions based on CEEMDAN. *Expert Systems with Applications*.
- Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computation and applied Mathematics*.
- Colianni, S., Rosales, S., & Signorotti, M. (2015). Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis. *CS229 Project, 1(5)*, 1–5. http://cs229.stanford.edu/proj2015/029_report.pdf
- Dehua, S., Andrew, U., & Pengfei, W. (2018). Does twitter predict Bitcoin? *Economics Letters*.
- Do-Hyung, K., Ju-Bong, K., Ju-Sung, H., Kim, I. C.-M., & Youn-Hee, H. (2019). Time Series Classification of Cryptocurrency Price Trend Based on a Recurrent LSTM Neural Network. *Journal of Information Processing Systems*, 694-706.
- Evans, L., Owda, M., Crockett, K., & Vilas, A. F. (2019). A methodology for the resolution of cashtag collisions on Twitter –A natural language processing & data fusion approach. *Expert Systems With Applications*, 353-369.
- Fakhfekha, M., & Ahmed, J. (2019). Volatility dynamics of crypto-currencies' returns: Evidence from asymmetric and long memory GARCH models. *Research in International Business and Finance*.
- Hachicha, F., Masmoudi, A., Abid, I., & Hassan, O. (2023). Herding behavior in Exploring the predictability of price clustering in cryptocurrency market. *Finance Research Letters*.
- Hegazy, K., & Mumford, S. (2016). Comparitive automated bitcoin trading strategies. *CS229 Project, 27*, 1–6. <http://cs229.stanford.edu/proj2016/report/MumfordHegazy-ComparitiveAutomatedBitcoinTradingStrategies-report.pdf>
- Higbee, A. (2018). The role of crypto-currency in cybercrime. *Computer Fraud and Security*.
- Hosure, N. S., Gaikwad, J. V., Shrivani, M. R., Kulloli, N., & Madhusudhan, H. S. (2022). Cryptocurrency Price Prediction Using Machine Learning. *International Journal of Advanced Research In Computer and Communication Engineering*.
- Jaquart, P., Dann, D., & Weinhardt, C. (2021). Short-term bitcoin market prediction via machine learning. *The journal of Finance and Data Science*, 45-66.
- Jiang, Z., & Liang, J. (2017). Cryptocurrency portfolio management with deep reinforcement learning. *2017 Intelligent Systems Conference (IntelliSys), 2018-Janua*, 905–913. <https://doi.org/10.1109/IntelliSys.2017.8324237>
- Khedr, A. M., Arif, I., Pravija, R. P., El-bannany, M., Alhashmi, S. M., & Sreedharan, M. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. *Wiley*.
- Lamon, C., Nielson, E., & Redondo, E. (2020). Cryptocurrency Price Prediction Using News and Social Media Sentiment. *The journal of Finance and data science*.
- Laskowski, M., & Kim, H. M. (2016). Rapid Prototyping of a Text Mining Application for Cryptocurrency Market Intelligence. *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, 448–453. <https://doi.org/10.1109/IRI.2016.66>
- Salman, M. K., & Ibrahim, A. A. (2020). Price Prediction of Different Cryptocurrencies Using Technical Trade Indicators and Machine Learning. *IOP Conference Series: Materials Science and Engineering*. Istanbul: IOP Publishing Ltd.
- Stenqvist, E., & Lönnö, J. (2017). Predicting Bitcoin price fluctuation with Twitter sentiment analysis [KTH ROYAL INSTITUTE OF TECHNOLOGY]. In *Diva*. <https://www.diva-portal.org/smash/get/diva2:1110776/FULLTEXT01.pdf>
- Tanwar, S., Patel, N. p., Patel, S. N., Patel, J. R., Sharma, G., & Davidson, I. (2021). Deep Learning-based Cryptocurrency Price Prediction Scheme with Inter-dependent Relations. *IEEE Access*.
- Wardak, A. B., & Rasheed, J. (2022). Bitcoin Cryptocurrency Price Prediction Using Long Short-Term Memory Recurrent Neural Network. *European Journal of Science and Technology*, 47-53.
- Zhong, C., Du, W., Xu, W., Huang, Q., Zhao, Y., & Wang, M. (2023). LSTM-ReGAT: A network-centric approach for cryptocurrency price trend Prediction. *Decison Support Systems*.