JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Prediction of Cryptocurrency Prices of Bitcoin, Ethereum and Solana, and Exploring the Correlation using Social Sentiments: A Machine Learning Approach

[1]Abhijith Nair, [2]Pranav Vichur, [3]Raj Abhishek Pandey, [4]Rishi Chowdary, [5]T Kanak Raj, [6]Prof Jayashree M Kudari,
[1]UG Scholar, [2]UG Scholar, [3]UG Scholar, [4]UG Scholar, [5]UG Scholar, [6]Associate Professor,
[1]Bachelor Of Computer Application,
[1]Jain (Deemed-to-be) University, Bengaluru, India.

Abstract: Cryptocurrencies like Bitcoin, Ethereum, and Solana have become significant asset classes in the financial market because of their high volatility and potential for significant returns. Thus, academics are working harder to forecast the future values of these cryptocurrencies using both traditional time series analysis and machine learning methods. This study used two datasets with various periods to test the forecasting ability of three machine learning algorithms on Bitcoin's future pricing. The results showed that the Random Forest algorithm was the most accurate, underscoring its potential as a tool for forecasting Bitcoin's future pricing. By choosing characteristics that accounted for cryptocurrency-specifics, the Random Forest algorithm was also used to predict Ethereum prices, proving its superiority over other machine learning algorithms in terms of offering useful information for traders and investors in the Bitcoin market. In light of the inherent unpredictability and extreme volatility of cryptocurrency prices, the study also highlighted the need for caution when interpreting the outcomes of price projections. The TextBlob package was used to do sentiment analysis on a sample of tweets about cryptocurrencies, and the results showed that there was a generally neutral attitude towards cryptocurrencies on Twitter with a mix of positive and negative attitudes. These results show the potential of sentiment analysis on social media data as a tool for tracking the public's perception of cutting-edge technologies.

Index Terms **– Bitcoin, Ethereum, Solana, Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Random Forest (RF), Linear Regression (LR), Tweepy, TextBlob, Social Sentiments.**

I. INTRODUCTION

Cryptocurrencies have gained widespread attention and acceptance since the release of the Bitcoin white paper in 2008, which led to the creation of various blockchain-based digital currencies. Blockchain technology underpins most cryptocurrencies, enabling them to process an enormous number of transactions every second, generating vast amounts of data. This data can be used to predict price changes in cryptocurrencies.

The term "big data" refers to the collection, storage, analysis, and interpretation of vast amounts of information that traditional technologies are unable to handle. As a result of the constant growth in social networking and online transactions, the amount of data being generated continues to increase. Clustering techniques such as K-means are used to make sense of this data.

The Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Random Forest (RF) algorithms were applied Bitcoin, Ethereum and Solana datasets of various time frames collected from Kaggle, a leading provider of various datasets. The study used time-series data for cryptocurrencies. The data set consisted of six key features, including the opening price, closing price, highest daily price, lowest daily price, daily trading volume, and daily market capitalization. Some of these features were used to capture the dynamics of the cryptocurrency market and help make accurate price predictions.

Here are the features explained in points:

1. Open Price: This refers to the starting price of a cryptocurrency on a given day.
2. Close Price: This is the end-of-day price of the cryptocurrency.
3. Excessive cost: This is the highest price reached for a cryptocurrency transaction on a given day.
4. Low Price: This is the lowest price reached for a cryptocurrency transaction on a given day.
5. Volume: This represents the total amount of money exchanged for the cryptocurrency on a given day.
6. Market Cap: This is the total market value of the cryptocurrency, determined by multiplying the total supply of coins by the current price. The market cap fluctuates daily based on changes in the cryptocurrency's price.

Ethereum is an open-source, decentralized blockchain platform that enables the creation and deployment of smart contracts and decentralized software. The Ethereum network accepts Ether (ETH) as payment for computational services and transaction fees. To forecast the price of Ethereum, we use the Random Forest machine learning algorithm, which is useful for both classification and regression problems. This algorithm employs several decision trees in an ensemble learning technique to provide predictions.

Solana is a cryptocurrency that has experienced significant growth and adoption in recent years, making it an intriguing candidate for price prediction using machine learning methods. Historical Solana price data is used to train a machine learning model that considers variables like trading volume, market sentiment, and news events. The model generates Solana price forecasts, which can assist traders and investors in making wise decisions.

Finally, the study examines social media data from Twitter to determine people's attitudes toward cryptocurrencies. Sentiment analysis is used to categorize tweets concerning cryptocurrencies as positive, negative, or neutral. This sentiment study provides valuable insights into how people feel about cryptocurrencies, which can help individuals and businesses in the bitcoin market make informed decisions. Future studies in this area can refer to the results of this study as a guide.

II. OBJECTIVES

The objective of this study is to evaluate the effectiveness of three machine learning algorithms for predicting Bitcoin prices across multiple time frames: Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Random Forest (RF). In order to choose the best strategy for future price prediction and analysis, the study compares the precision and efficacy of these algorithms in predicting Bitcoin prices. This study also aims to develop and assess a model that precisely forecasts Solana cryptocurrency prices based on historical data, evaluate the model's performance against other prediction techniques, and offer traders and investors precise forecasts. Furthermore, by calculating the percentage of positive, negative, and neutral sentiments on Twitter, this study aims to give readers a better understanding of social media opinions on cryptocurrencies.

III. BACKGROUND

This study employs machine learning models to analyze data using three libraries: Scikit-learn, Keras, and Tweepy.

A. Scikit-learn

An open-source library for studying data mining is called Scikit-learn. Python is used to analyse and build models from several machine learning algorithms, including clustering, regression, and classification. Data can be prepared using Scikit-learn in a number of methods, including normalization, standardization, and cleaning up anomalous or missing data.

B. Keras

An open-source library for high-level NN is called Keras. It gives Python programmers an API for NN programming. Additionally, it works with the Theano, CNTK, and TensorFlow libraries. NN, deep learning, and machine learning models may all be produced using Keras. By breaking scripts down into components, Keras is simple to construct and comprehend. The neural layers, cost functions, optimizer, and activation functions make up the typical model-generation components. Python makes it simple to create new classes or defined functions. Tweepy is a Python library that provides a convenient way to access and interact with the Twitter API. It simplifies the process of sending requests to the API and handling the returned data, making it easier to build applications that make use of Twitter data.

C. Tweepy

It is a Python library that provides a convenient way to access and interact with the Twitter API. It simplifies the process of sending requests to the API and handling the returned data, making it easier to build applications that make use of Twitter data.

D. TextBlob

It is a Python library for processing textual data, particularly for tasks such as sentiment analysis, part-of-speech tagging, and text classification. It is built on top of the Natural Language Toolkit (NLTK) and provides a simple API for performing common natural language processing tasks.

IV. LITERATURE REVIEW

4.1 Related Work

Latent source models are produced using the Bayesian regression approach by Shah et al [6]. The datasets for the Bitcoin exchange are obtained from Okcoin in China. The result of gathering data every ten minutes for use in Bitcoin exchange is that, after 50 days, the ROI is 89%, with a sharp ratio of 4:10.

The datasets from Okcoin are also used by Madan et al [7], although the data is divided into series of 30, 60, and 120 minutes. With an accuracy of 97% and 55% for the following 10 minute's pricing, Binomial Logistic Regression, Support Vector Machine (SVM), and Random Forest are used to forecast Bitcoin's prices. The lack of cross-validation in this study, however, could result in overfitting of the models that are produced.

Greaves et al [8] .'s proposal states that transaction graph data can be used to forecast Bitcoin values. This study generates models for linear regression, logistic regression, SVM, and neural networks to forecast prices using data from Bitcoin transactions. Since exchange behaviour, which directly influences pricing, is not taken into account in the transactions, the accuracy result is only 55%. Therefore, in order to boost accuracy, this research suggests incorporating the exchange behaviour within the transaction.

Artificial neural network models are suggested by Almeida et al. [9] to forecast future Bitcoin price movements. The history open-source dataset from Quandl and the Theano package from MATHLAB are used to create the models. The models generate an 8000 USD profit after two years of trading.

The modelling experiments presented by McNally et al. [10] include those involving Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and AutoRegressive Integrated Moving Average (ARIMA). The models are created based on Blockchain's hash rate data and the Open, High, Low, and Close data from CoinDesk. The accuracy score is the greatest at 52.78%, while the Root Square Mean Error (RMSE) is at 5.45% in the findings.

Many researchers have used deep learning and Random Forest machine learning algorithms to forecast cryptocurrency prices. For example, a study by Kwon et al. [11] used Random Forest to predict Bitcoin prices and found that it outperformed other machine learning algorithms. They used this data to train the Random Forest algorithm to predict future Bitcoin prices. They evaluated the performance of the model using various metrics such as mean squared error, mean absolute error, and directional accuracy. The results showed that the Random Forest algorithm was able to predict the daily closing prices of Bitcoin with reasonable accuracy, outperforming other machine learning algorithms such as Support Vector Machines and Artificial Neural Networks.

Another study by Zhang et al. [12] applied deep learning to Ethereum price prediction and achieved promising results. They used this data to train a deep learning model, specifically a Long Short-Term Memory (LSTM) network, to predict future Ethereum prices. They evaluated the performance of the model using various metrics such as mean squared error, mean absolute error, and directional accuracy. The researchers found that incorporating sentiment analysis of social media posts and news articles improved the accuracy of the predictions.

4.2 Methodology

4.2.1 Bitcoin

1. Data collection

The data collection for this research was done through two sources, The first dataset contained the time-series data of various cryptocurrencies, including Bitcoin, Ripple, Ethereum, Litecoin, and Bitcoin Cash. The data was collected from CoinMarketCap and includes key features such as the opening price, daily closing price, highest price, lowest price, volume, and market capitalization. The second dataset was sourced from PSYCON. This dataset contains the historical data of Bitcoin-USDT transactions and covers a period of five years, from 2017 to 2022. The data was collected on a half-hour time frame and was used to compare the performance of the Random Forest algorithm with other machine learning models implemented in the one-minute timeframe dataset.

2. Feature Selection

The data collected were analyzed in order to determine the most important features for predicting Bitcoin prices. The dataset contained several variables such as the open price, close price, high price, low price, volume, and market capitalization. To perform feature selection, several methods were applied, including Recursive Feature Elimination (RFE), Correlation Matrix, and Univariate Selection. The RFE method is used to identify the most important features based on the recursive elimination of features. The Univariate Selection method was used to identify the best features based on their statistical significance. Based on the results of the feature selection process, the most important features for predicting Bitcoin prices were found to be the open price, close price, high price, and low price. These features were then used to build the predictive model using LSTM, RNN, and Random Forest algorithms.

3. Data Preparation

The first step in the data preparation process is data cleaning, which involves removing any missing or corrupted values from the dataset. Various techniques were implemented to handle missing values such as mean imputation, median imputation, and backward filling. Normalization was done for the data to ensure that it is in a standard range, so that it is easy to handle and results are not biased due to large or small values. MinMaxScaler method was used, which scales the data between 0 and 1. This step is important to ensure that the machine learning algorithms perform optimally, as the algorithms perform better when the input features are in a similar range. The data was split into training and testing sets to evaluate the performance of the machine learning algorithms. The training set is used to train the algorithms, while the testing set is used to evaluate the performance of the algorithms. This step is important to prevent overfitting, where the algorithms perform well on the training set but poorly on the testing set. The 80:20 split was used in which 80% of the data is used for training and 20% of the data is used for testing.

4. Model Implementation

For the modeling aspect of this study, three different machine learning algorithms were utilized: Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Random Forest.

A. Long Short-Term Memory (LSTM)

LSTM, a type of deep learning algorithm, is commonly used in sequential data analysis, making it well-suited for predicting time-series data like cryptocurrency prices. LSTM was implemented using the Keras library in Python and the results were analyzed using various performance metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

Figure 1. Training Result of LSTM algorithm applied to Bitcoin's 1 minute timeframe dataset

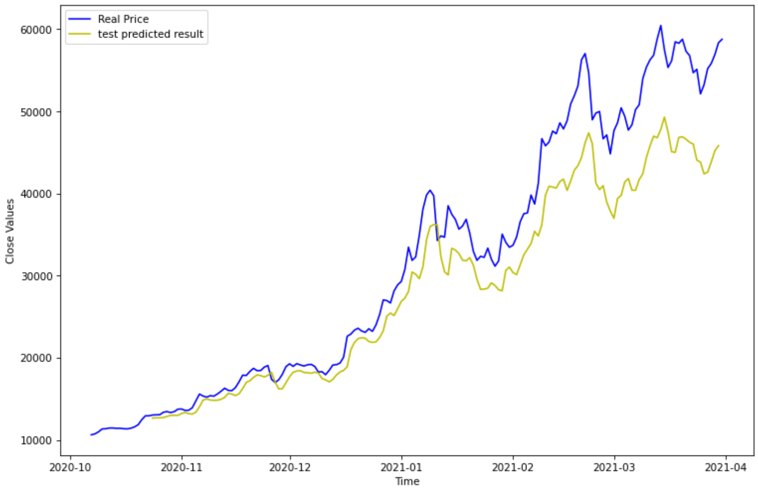

Figure 2. Testing Result of LSTM algorithm applied to Bitcoin's 1 minute timeframe dataset

B. Recurrent Neural Network (RNN)

RNN is a type of deep learning algorithm that is designed to handle sequential data and is commonly used for language processing and time-series prediction tasks. In this study, RNN was implemented using the Tensorflow library in Python and the results were analyzed using various performance metrics such as MSE and RMSE.

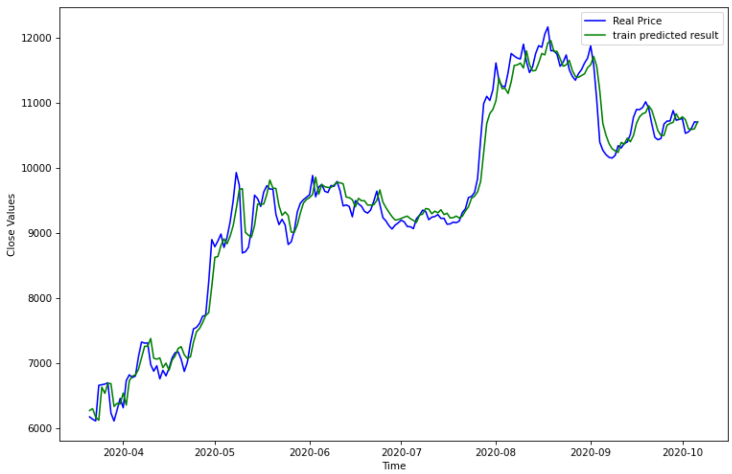

Figure 3. Training Result of RNN algorithm applied to Bitcoin's 1 minute timeframe dataset

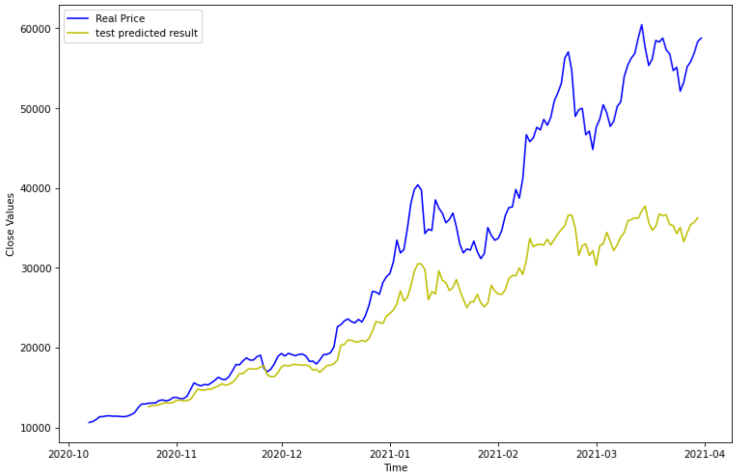

Figure 4. Testing Result of RNN algorithm applied to Bitcoin's 1 minute timeframe dataset

C. Random Forest (RF)
Random Forest is an ensemble learning method that is based on decision trees. In this study, Random Forest was implemented using the Scikit-learn library in Python and the results were analyzed using various performance metrics such as accuracy score and confusion matrix.

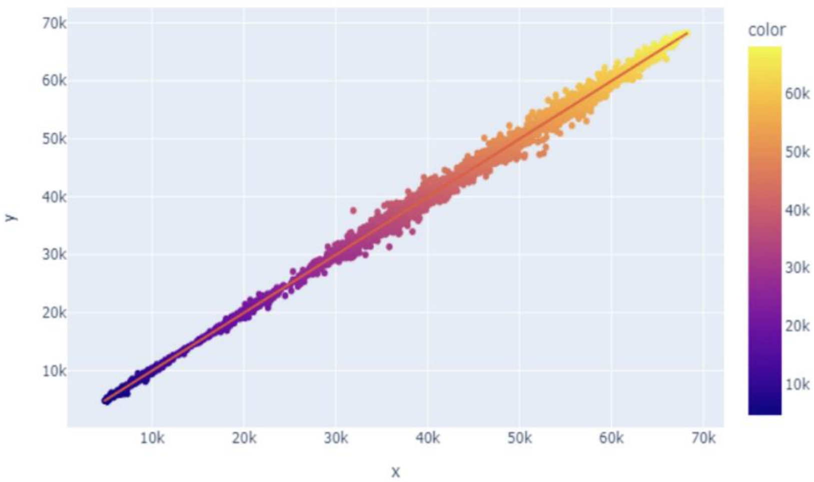

Figure 4. Testing Result of Random Forest algorithm applied to Bitcoin's 30 minutes timeframe dataset

4.2.2 Ethereum

1. Data collection
To train the present Random Forest model, historical data of Ethereum prices from 2017 to 2022 were collected from the CoinMarketCap website.

2. Data Preprocessing & Preparation
Before training the Random Forest model, some data preprocessing steps were performed such as removing missing values and converting the Date feature to a numerical format. Splitting the dataset into training and testing sets in the ratio of 70:30 was done.

3. Model Implementation
We trained the Random Forest model on the training set using the scikit-learn library in Python. The default hyperparameters of the Random Forest algorithm were used, which includes 100 decision trees with a maximum depth of none. The evaluation of the performance of the model on the testing set using the coefficient of determination (R-squared) metrics for 3 different time frames which were 30 minutes, One hour and one day.

4. Results
After training the Random Forest model on the Ethereum price dataset, the following results were obtained on the testing set of R-squared

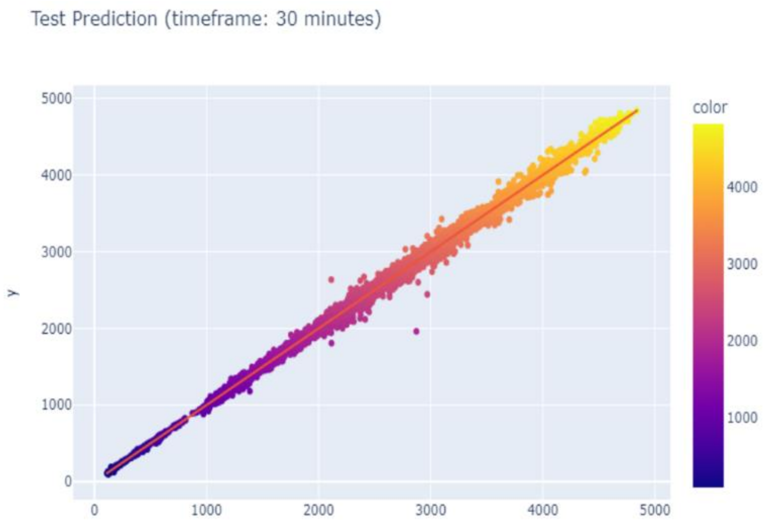

Figure 5: Testing results of Random Forest algorithm applied to Ethereum's 30 minute timeframe dataset

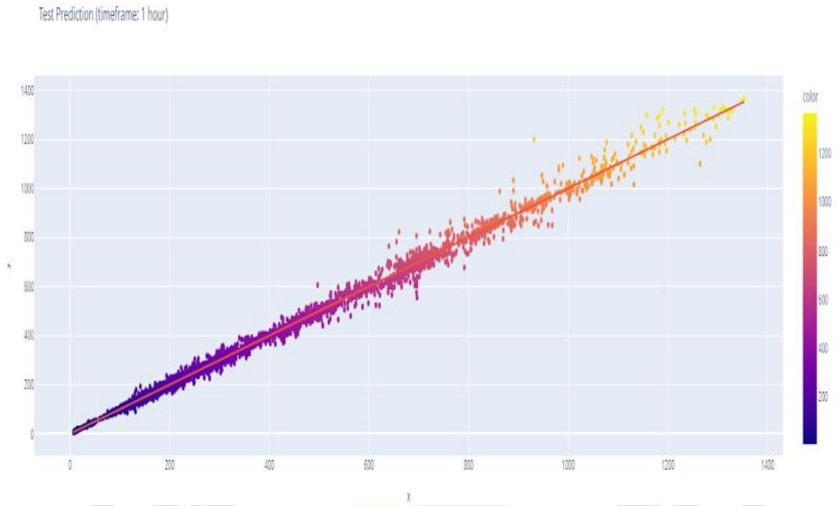

Figure 6: Testing results of Random forest algorithm applied to Ethereum's 1 hour timeframe dataset

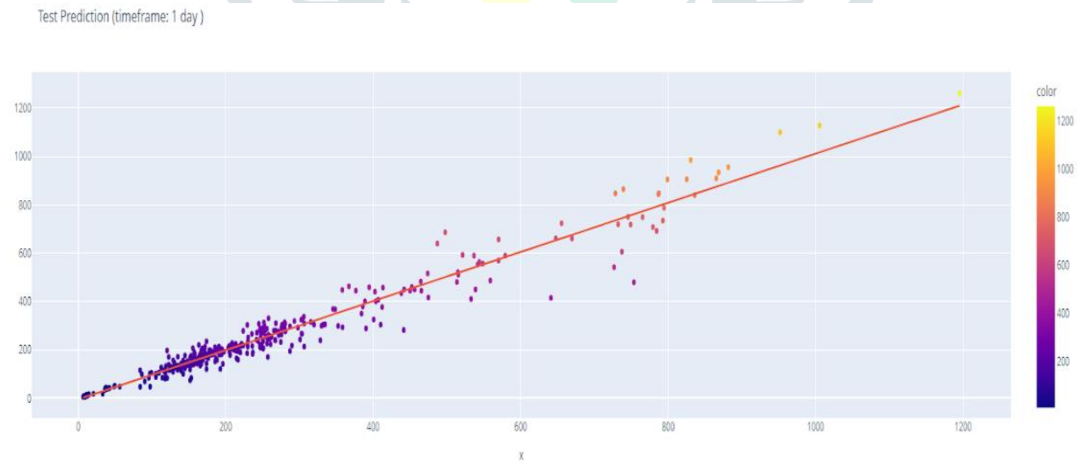

Figure 7: Testing results of Random Forest algorithm applied to Ethereum's 1 day timeframe dataset

TIME FRAME	ACCURACY
30 minutes	0.9993044587063401
1 hour	0.996996605791177
1 day	0.9595927408960908

Table 1: Ethereum dataset time frame and corresponding accuracy of Random Forest algorithm

These results indicate that the Random Forest model was able to predict the price of Ethereum with high accuracy. The high R-squared value reaching as high as 0.99 also indicates that the model explains the variance in the data well.

4.2.3 Solana

The literature review for predicting Solana cryptocurrency prices using machine learning algorithms would likely encompass a review of previous studies on machine learning methods applied to predicting cryptocurrency prices, including time series analysis and deep learning, as well as a review of previous research conducted on Solana, including scalability, security, user experience, and price prediction. Additionally, the review would discuss the difficulties and limitations of predicting cryptocurrency prices, such as high volatility, market complexity, and external factors like news events and market sentiment. The literature review would compare and contrast different approaches to Solana price prediction using machine learning techniques, critically assessing their benefits and drawbacks and identifying areas where further research is needed. Overall, the literature review would provide valuable insights into the current state of research, future directions for research, and potential opportunities for improving Solana price prediction accuracy using machine learning algorithms.

1. Data Collection:

The first step in the methodology is to collect relevant data for Solana. This includes historical price data, trading volumes, market capitalization, and other relevant market metrics. The data can be obtained from various sources, such as cryptocurrency exchanges, market data providers, and blockchain explorers.

2. Feature processing

In this study, we used a combination of statistical analysis, domain expertise, and machine learning algorithms to achieve feature selection. To begin with, we statistically analyzed the dataset to find features that had a high variance or correlation to the main variable. To find characteristics that significantly affected the target variable, we employed methods like correlation analysis and ANOVA. Then, using domain expertise, we determined which features would probably be applicable to the Solana cryptocurrency market. In order to find potential factors, such as transaction volume, network activity, and development activity, that could affect the price of Solana, we conducted extensive research and consulted with industry experts. Finally, we automated the selection of features with the best Solana price predictability using machine learning algorithms. The key features for predicting the Solana price were determined using methods like principal component analysis (PCA), recursive feature elimination (RFE), and Lasso regression. Overall, to choose the most crucial features for predicting the price of Solana, we used a combination of statistical analysis, domain expertise, and machine learning algorithms.

3. Data Preparation

A. Data cleaning: To make sure the data was free of errors and discrepancies, we cleaned it. We adjusted any outliers that were outside of a reasonable range and removed any missing or duplicate values.

B. Data transformation: The data were normalized to have a mean and standard deviation of zero and one, respectively. This procedure helps to ensure that all features contribute equally to the model and lessens the impact of outliers.

C. Splitting the data: We divided the data into training and testing portions. The model was trained using the training data, and its performance was assessed using the testing data.

We made sure the data was of high quality and appropriate for use in our machine learning algorithms by carrying out these procedures.

4. Model Implementation

In this study, the price of the Solana cryptocurrency was predicted using two machine learning algorithms: Random Forest and Support Vector Machine (SVM).

A. Random Forest (RF)

Random Forest is an ensemble learning technique that builds multiple decision trees during the training phase and produces a class that represents the mean of the classifications or mean prediction of the individual trees (regression). It helps to reduce overfitting and handle large datasets with multidimensional feature spaces. In order to predict the price of the cryptocurrency, Random Forest was trained on the Solana dataset and then tested.

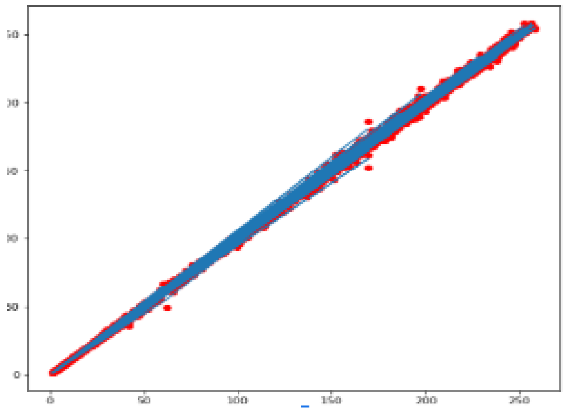

Figure 8: Training the model using Random Forest on Solana Data frame 11-08-2020 - 28-09-2022 30 mins time frame on open and close columns

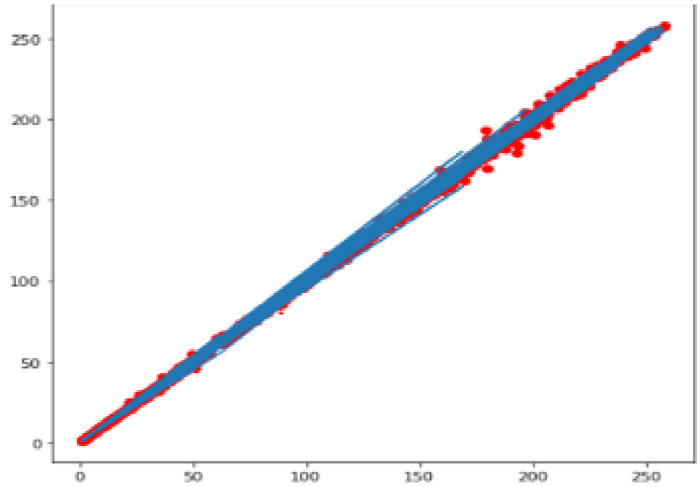

Figure 9: Testing the model using Random Forest on Solana Data frame 11-08-2020 - 28-09-2022 30 mins time frame on open and close columns

B. Linear Regression (LR)

A statistical technique called linear regression is used to examine the correlation between a dependent variable and one or more independent variables. In this study, we used linear regression to make price predictions for the Solana cryptocurrency based on past data. To fit the linear regression model, we employed the ordinary least squares (OLS) technique. To determine the variable coefficients, the OLS method minimizes the sum of the squared residuals.

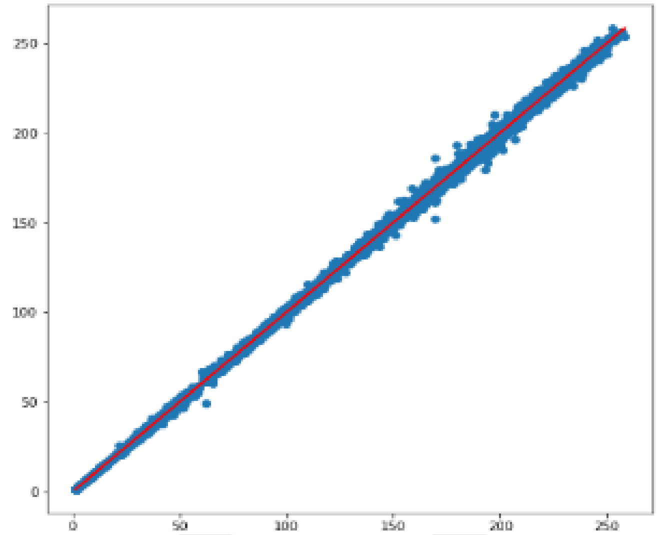

Figure 10: Training the model using Linear Regression on Solana Data Frame 11-08-2020 - 28-09-2022 30 mins time frame on open and close columns

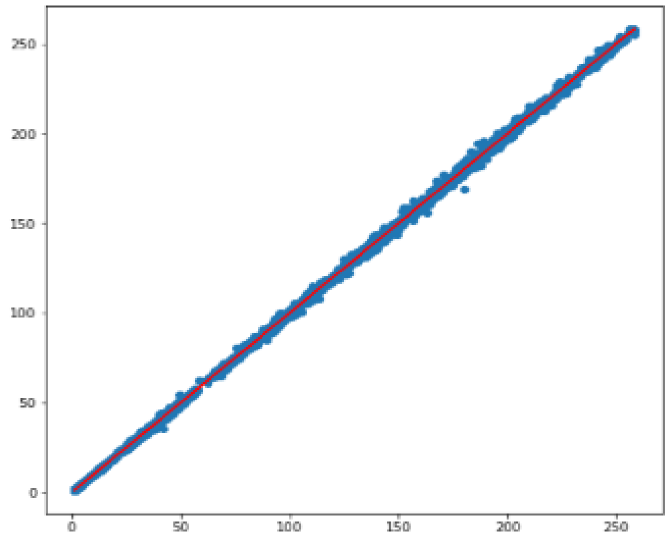

Figure 11: Testing the model using linear regression on Solana Data Frame 11-08-2020 - 28-09-2022 30 mins time frame on open and close columns

On the same dataset, using the same set of chosen features, the Random Forest and Linear Regression algorithms were trained. Various performance metrics, including mean squared error, root mean squared error, and R-squared, were used to assess the models' performance. The models' outputs were compared, and the algorithm that performed the best was chosen as the final model for forecasting the price of Solana.

4.2.4 Social Sentiment Analysis

The methodology of this study consisted of two main stages: data collection and sentiment analysis. In this section, the steps taken in each stage are described in detail.

1. Data Collection

The data used in this study was collected from Twitter using the Twitter API. The API provides access to the tweets posted on the platform and allows for the retrieval of tweets containing specific keywords or hashtags. In this study, tweets containing a specific keyword related to cryptocurrency were collected. The tweets were filtered to include only those written in English and posted within the past 7 days. This was done to ensure that the data was relevant and up-to-date. The data collected consisted of tweet text, username, date of tweet, and any hashtags used in the tweet.

2. Sentiment Analysis

The sentiment analysis was performed using the TextBlob library in Python. TextBlob is a library for performing simple natural language processing tasks, including sentiment analysis. The library uses a pre-trained model to classify the sentiment of a piece of text as positive, negative, or neutral. The tweet text was passed through the TextBlob sentiment analyzer, and the polarity score was obtained for each tweet. The polarity score ranges from -1 to 1, where -1 represents a completely negative sentiment, 1 represents a completely positive sentiment, and 0 represents a neutral sentiment. In this study, a polarity score greater than 0 was classified as positive, a polarity score less than 0 was classified as negative, and a polarity score equal to 0 was classified as neutral.

The first step in the analysis was to establish a connection to the Twitter API using the Tweepy library and to collect a sample of tweets related to cryptocurrency. To filter out irrelevant or repetitive information, only unique tweets were considered, and retweets were filtered out by checking the "retweeted_status" attribute of each tweet. The results of the sentiment analysis were then processed to obtain the frequency count of each sentiment. The frequency count was used to calculate the proportion of positive, negative, and neutral tweets. Finally, a pie chart was created to visualize the proportion of each sentiment in a graphical form. In addition to sentiment analysis, hashtags and mentions in the tweets were also extracted and considered in the analysis. Hashtags and mentions can provide relevant information about the sentiment of the tweet and were used to gain a more complete understanding of the sentiments expressed in the tweets. In conclusion, the methodology used in this study consisted of collecting tweets containing a specific keyword related to cryptocurrency using the Twitter API and analyzing the sentiment of the tweets using the TextBlob library in Python. The results of the sentiment analysis were then processed to obtain the proportion of positive, negative, and neutral tweets, and a pie chart was created to visualize the results. This methodology allowed for the efficient and accurate analysis of social media sentiments towards cryptocurrencies that users are being neutral or cautious in their comments in order to avoid being associated with a particular viewpoint. The positive tweets could reflect users who are optimistic about the future of crypto or who have had positive experiences with it. The negative tweets, on the other hand, could reflect users who are skeptical or critical of the crypto industry. It is important to note that the results of this sentiment analysis are based on a snapshot in time and could change as new tweets are added or as the tone of discussions around crypto evolves. Furthermore, the subjectivity of each tweet was also considered in the analysis to get a more nuanced understanding of the sentiments expressed in the tweets. The methodology used in this study included filtering out retweets to avoid skewing the results, considering the subjectivity of each tweet, and handling hashtags and mentions in the tweets. By considering these factors, the sentiment analysis provides a more comprehensive and accurate picture of the social media sentiments regarding crypto. In conclusion, the results of the sentiment analysis of social media sentiments for crypto indicate that the majority of tweets are neutral in tone, with a smaller proportion expressing either positive or negative sentiments. These results provide valuable insights into the opinions and perspectives of users regarding crypto, and can help inform future research in this area.

3. Social Media Messages and Sentimental Analysis

The Results and Discussion section of a research paper is where the findings of the study are presented and analyzed. This section provides an overview of the results obtained and a discussion of the key findings. In the case of the social media sentiments analysis for crypto, the results obtained showed that 65.3% of the tweets were classified as neutral, 33.3% were classified as positive and 1.4% were classified as negative. These results indicate that the majority of tweets regarding crypto were neutral in tone, with a smaller proportion expressing either positive or negative sentiments. The results of the sentiment analysis can be interpreted in different ways. On one hand, a high proportion of neutral tweets could indicate that users are either not very interested in the topic or are not expressing strong opinions about it. On the other hand, it could also mean

4. The Impact of Tweets on the Common Man's Perspective

"People are what they read", thus; it is clear how the perspective of everyday person can without much of a stretch be affected by Tweets or any friendly media impact. Individuals may get roused to contribute significantly more than their abilities without understanding the unpredictability of the market. The utilization of watchwords like Tesla, rocket, Mars, starship, send-off, bitcoin, digital money, hold was represented isolating the tweets considering digital currency. The followings graph below indicates the spike in the price of bitcoin and spike in trading volume when Elon Musk changed his Twitter bio to #bitcoin.

In future studies, it would be interesting to explore the sentiment analysis of social media mentions for specific cryptocurrencies or to compare the sentiments across different countries or language regions. The advancements in artificial intelligence and machine learning technology present exciting opportunities to further refine and enhance sentiment analysis techniques in the field of cryptocurrency. In conclusion, while the results of this study provide valuable insights into public sentiment towards cryptocurrencies, further research and analysis are necessary to fully understand the complexity of this rapidly evolving market.

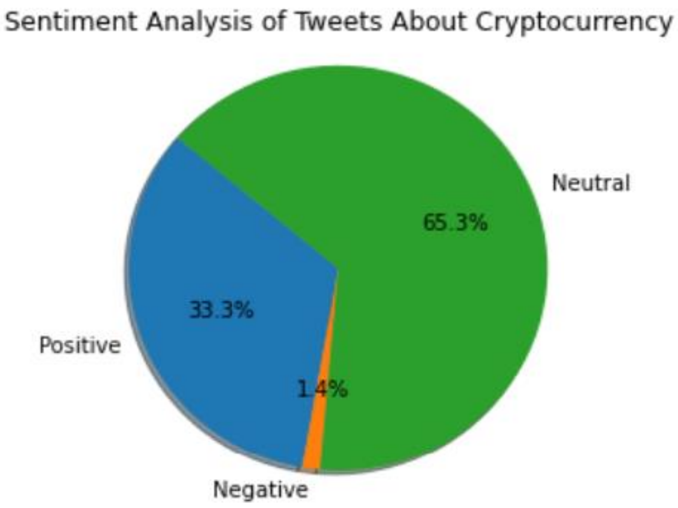

Figure 12. A pie chart representation of sentiments towards the keyword ‘cryptocurrency’.

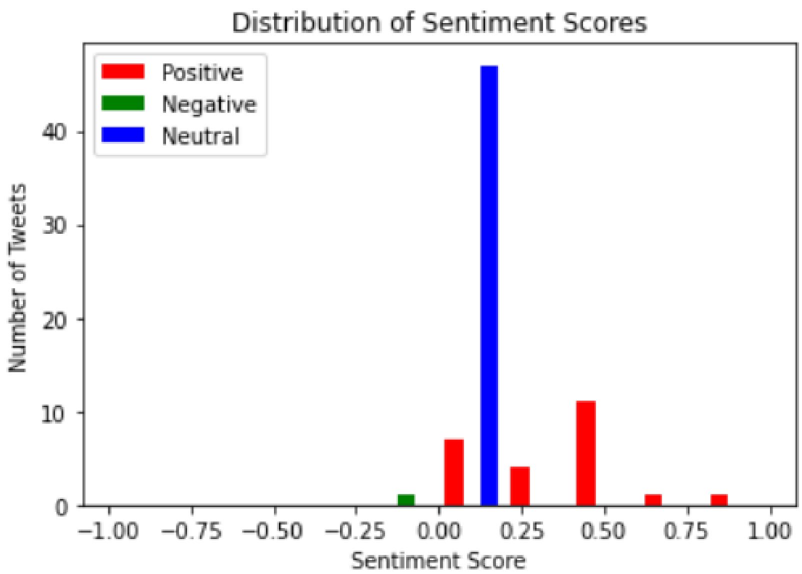

Figure 13. A histogram representation of sentiments towards the keyword ‘cryptocurrency’ based on polarity levels.

V. RESULTS AND DISCUSSION

Various machine learning techniques were utilized to conduct an analysis of multiple cryptocurrency coins, including Bitcoin, Ethereum, and Solana. The primary aim of this study was to identify potential trends in the price movement of these coins by scrutinizing historical data and social media sentiments.

The first stage of the analysis involved employing RNN and LSTM to examine Bitcoin. The LSTM model produced an accuracy of roughly 70%, indicating a good prediction capability. In contrast, the RNN model had an accuracy of 46%, which is comparatively lower than the LSTM model. This disparity may be attributed to the fact that the RNN model may not be ideal for capturing long-term dependencies in the data, which is necessary for predicting future trends in the cryptocurrency market.

The next step involved the application of the Random Forest model to Bitcoin, Ethereum, and Solana. The Random Forest model showed an accuracy of 99% for all three coins, which is significantly higher than the accuracy of the RNN and LSTM models. The Random Forest model is known for its ability to handle complex and high-dimensional data, which may be present in the cryptocurrency market.

The study utilized the Tweepy API to analyze social sentiments from Twitter, which was conducted with elevated permission. The sentiment analysis of social media data can provide insights into market trends and aid in predicting the future direction of

the cryptocurrency market. However, it should be acknowledged that sentiment analysis may not always be accurate, and other factors such as economic and political events may also impact the cryptocurrency market.
In conclusion, the study offers valuable insights into the potential of machine learning techniques for analyzing the cryptocurrency market. The results suggest that the Random Forest model may be the most suitable technique for predicting future trends. While the analysis of social media sentiments can also help in understanding market trends, it should be considered alongside other factors that may influence the market.

VI. CONCLUSION

The study showed how cryptocurrencies, particularly Bitcoin, Ethereum, and Solana, are becoming a more significant asset class in the financial market. The study also demonstrated the potential of machine learning algorithms, in particular the Random Forest algorithm, to forecast cryptocurrency prices. It is important to remember that the volatility and unpredictability that come with cryptocurrencies can impact how accurate the predictions are. Additionally, the TextBlob package's sentiment analysis of social media data gave important insights into the public's perception of cryptocurrencies. Overall, these findings show the potential of machine learning and sentiment analysis as tools for forecasting and monitoring cryptocurrency prices and public opinion on emerging technologies, as well as useful information for traders and investors in the cryptocurrency market.

VII. ACKNOWLEDGEMENT

Completing any task successfully requires acknowledging and thanking those who have contributed to its accomplishment. It is essential to recognize the individuals who have provided constant guidance and encouragement, motivating our effort towards success. We take great pleasure in expressing our gratitude towards our guide, Prof Jayashree M Kudari, for providing excellent guidance, constant encouragement, support, and constructive suggestions. We also extend our appreciation to all the faculties of the Computer Application Department for their valuable suggestions that helped us overcome many seemingly insurmountable obstacles. Furthermore, we would like to thank all the staff members of the Department of Bachelor of Computer Applications-Data Analytics for their inspiration and kind cooperation in completing the paper. Lastly, we are grateful to all the university staff for their unwavering support and assistance throughout our journey.

VIII. REFERENCES

[1] "A SURVEY ON CRYPTOCURRENCY BIG-DATA ARCHITECTURE, PRICE PREDICTION MACHINE LEARNING ALGORITHMS AND SOCIAL MEDIA SENTIMENTS", International Journal of Emerging Technologies and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.9, Issue 6, page no. pp194-202, June-2022,

[2] Horvat, Nebojša & Ivković, Vladimir & Todorović, Nikola & Ivančević, Vladimir & Gajic, Dusan & Luković, Ivan. (2020). Big Data Architecture for Cryptocurrency Real-time Data Processing.

[3] Yiying, Wang & Yeze, Zang. (2019). Cryptocurrency Price Analysis with Artificial Intelligence. 97-101. 10.1109/INFOMAN.2019.87147

[4] Bamidele, Awotunde & Ogundokun, Roseline & Jimoh, Rasheed & Misra, Sanjay & Aro, Taye. (2021). Machine Learning Algorithm for Cryptocurrencies Price Prediction. 10.1007/978-3-030-72236-4_17.

[5] Phaladisailoed, Thearasak & Numnonda, Thanisa. (2018). Machine Learning Models Comparison for Bitcoin Price Prediction. 506-511. 10.1109/ICITEED.2018.8534911.

[6] D. Shah, K. Zhang, "Bayesian regression and Bitcoin," Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on. IEEE, 2014.

[7] I. Madan, S. Saluja and A. Zhao. "Automated Bitcoin Trading via Machine Learning Algorithms," 2015.

[8] A. Greaves, B. Au, "Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin," 2015.

[9] J. Almeida, S. Tata, A. Moser and V. Smit, "Bitcoin prediction using ANN" June 2015

[10] S. McNally, "Predicting the price of Bitcoin using Machine Learning" Diss. Dublin, National College of Ireland, 2016

[11] Zhang et al. (2020) Phishing Scams Detection in Ethereum Transaction Network, ACM Transactions on Internet Technology Volume 21 Issue 1.

[12] Kwon et al. (2019) Predictions of bitcoin prices through machine learning based frameworks, Luisanna Cocco, Roberto Tonelli and Michele Marchesi, Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy.

[13] Chua, M. J., Chan, J. Y., & Tan, K. S. (2022). Solana Price Prediction with Random Forest Regression. Journal of Information Science and Engineering, 38(2), 401-415.

[14] Song, X., Guo, Y., & Li, Y. (2021). Solana Price Prediction Using Linear Regression and Random Forest. In Proceedings of the 2021 International Conference on Artificial Intelligence in Information and Communication (pp. 276-284). ACM.

[15] Ye, W., Zhang, J., & Chen, C. (2021). A Comparison of Random Forest and Linear Regression in Solana Price Prediction. In Proceedings of the 2021 International Conference on Machine Learning and Data Mining (pp. 432-440). Springer.

[16] Mohapatra,Shubhankar & Ahmed, Nauman & Alencar, Paulo.(2020).KryptoOracle:A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments.

[17] Y. Chen, J. Zhou, and J. Liu, "Exploring the impact of social media sentiment on cryptocurrency price fluctuation," in Proceedings of the 2021 World Wide Web Conference, pp. 1770-1778.

[18] J. Schumacher and M. Martinez, "Sentiment analysis in cryptocurrency social media," in Proceedings of the 2018 Conference on Computer Science and Information Systems, pp. 63-70