

Prediction Analysis: Bitcoin Price using Machine Learning

E.Sujatha¹ and Jaswanth N.S²

^{1,2}Department of Computer Science and Engineering, Saveetha Engineering College (Autonomous), Chennai, 602 105
Email: sujathae@saveetha.ac.in, nsjaswanth@gmail.com

Abstract—‘Bitcoin Price Prediction’, the ruler of cryptocurrency plays important role in blockchain technology. - In this project, we proposed to predict the Bitcoin price accurately taking into consideration various parameters that affect the Bitcoin value. we aim to understand and find daily trends in the Bitcoin market while gaining insight into optimal features surrounding Bitcoin price. Our data set consists of various features relating to the Bitcoin price and payment network over the course of every year, recorded daily. Features such as the opening price, highest price, lowest price, closing price, volume of Bitcoin, volume of currencies, and weighted price were taken into consideration so as to predict the closing price of the next day. Random forest model designed and implemented on scikit learn frameworks to build predictive analysis and evaluated them by computing various measures such as the RMSE (root mean square error) and r (Pearson's correlation coefficient) on test data. Flask framework was used to make prediction in webpages and Beautiful Soup is used to scrap the data from 'URL': 'https://bitinfocharts. The future prediction of bitcoin is predicted as a result from today real time data. Support Vector Regressor, Auto-Regressive Integrated Moving Average (ARIMA), Facebook Prophet, Unidirectional LSTM, Bidirectional LSTM, Stacked LSTM. Moreover, the research experiments are repeated several times to achieve the best results by employing hyperparameter tuning of each algorithm. This involves selecting an appropriate kernel and suitable data normalization technique for SVR.

Index Terms— Machine learning; Neural networks; Bitcoin price prediction; Deep learning; Regression; Classification

I. INTRODUCTION

Bitcoin is a cryptographic money which is utilized worldwide for advanced instalments or basically for speculation purposes. Bitcoin is decentralized for example it isn't possessed by anybody. Bitcoins are put away in an advanced wallet which is essentially similar to a virtual financial balance. The record of the considerable number of exchanges, the timestamp information is put away in a spot called Block chain. Each record in a block chain is known as a square. Each square contains a pointer to a past square of information. The information on block chain is scrambled. During exchanges the client's name isn't uncovered, however just their wallet ID is made open. The Bitcoin's worth fluctuates simply like a bitcoin though in an unexpected way. There are various calculations utilized on financial exchange information for value forecast. Notwithstanding, the parameters influencing Bitcoin are extraordinary. In this manner it is important to anticipate the estimation of Bitcoin so right venture choices can be made. The cost of Bitcoin doesn't rely upon the business occasions or mediating government not at all like securities exchange. Hence, to anticipate the worth we feel it is important to use AI innovation to foresee the cost

of Bitcoin. We will predict the sign of the daily price change with highest possible accuracy. Since the barter system, trade and investment is an integral part of economic development of a country.

Over a time period significant changes have been made in trade which leads to the growth of national economy. The changes like conversion from barter system to money-based trades, introduction of digital currency like bitcoin (BTC) are significant. The initial value of a BTC in the year of 2009 was \$0.0008 and over the years the price of BTC has drastically increased to \$46,434.40. This indicates the popularity of BTC among the current society. This cryptocurrency is used extensively to add data into blockchain. Blockchain technology is used by many countries in various fields like health care, banking, business etc, mainly because of its high-level security and scalability.

II. RELATED WORKS

Researchers have used a wide range of machine learning algorithms to study and predict personality. Almost all of the research in this area went through several stages, including data collection, pre-processing, feature extraction, and classification to assess the model's accuracy [1][2]

Machine learning (ML) and neural networks (NNs) have been widely used in bitcoin price prediction in recent years [3]. ML algorithms can learn from historical bitcoin data to identify patterns and relationships that can be used to predict future bitcoin prices. NNs are a type of ML algorithm that are particularly well-suited for bitcoin price prediction because they can learn complex non-linear relationships in the data [4].

A subset of artificial intelligence (AI), machine learning (ML) enables systems to learn and develop on their own from data without explicit programming [5]. We find supervised learning, unsupervised learning, and reinforcement learning among the various machine learning algorithms. Notably, the most widely used techniques are supervised learning and unsupervised learning [6].

Input data, also known as training data, and a corresponding target or outcome variable define supervised learning. A number of features in the input data help the model predict the desired output. Linear regression, logistic regression, and decision trees are a few examples of supervised learning algorithms. When the output variable is divided into distinct classes, classification is a machine learning technique used to predict the outcome of a given sample. Naive Bayes Classifier and Support Vector Machine (SVM) are notable classification algorithms [7][8][9]. A bitcoin price prediction system using machine learning (ML) is a system that uses ML algorithms to predict the future price of a bitcoin [10]. These systems work by analysing historical bitcoin data, such as price, volume, and other financial indicators, to identify patterns and trends. Once the system has identified these patterns, it can use them to make predictions about future bitcoin prices [12].

Machine learning (ML) is a type of artificial intelligence (AI) that allows systems to learn and improve from data without being explicitly programmed [13] [14]. ML algorithms can be used to predict bitcoin prices by analysing historical data, such as future bitcoin prices.

One common ML approach to bitcoin price prediction is to use supervised learning. Supervised learning algorithms are trained on a set of labelled data, where each data point has an input and an output. In the case of bitcoin price prediction, the input data could be historical bitcoin prices, trading volume, and news sentiment, and the output data could be the future bitcoin price. Once the algorithm is trained, it can be used to predict the future bitcoin price for a given set of input data [15][16].

Another common ML approach to bitcoin price prediction is to use unsupervised learning. Unsupervised learning algorithms are trained on a set of unlabelled data, where each data point does not have an output. Unsupervised learning algorithms can be used to identify patterns and trends in the data, which can then be used to make predictions about the future bitcoin price [17][18].

A popular supervised learning algorithm that excels at binary classification tasks is logistic regression [19]. Logistic Regression focuses on modelling the relationship between input features and the likelihood of an instance belonging to a particular class, as opposed to Random Forest, which uses a collection of decision trees [20].

The simplicity and interpretability of Logistic Regression are its main advantages. It uses a logistic (sigmoid) function to analyse a linear combination of input features to determine the likelihood that a given instance belongs to one of two classes. The model's coefficients, which are assigned to each feature, shed light on the importance and trajectory of their influence on the result [21].

III. METHODOLOGY

Workflow Analysis

Bitcoin values are very valuable but very difficult to predict correctly for any individual human being. This project tries to solve the problem of bitcoin price prediction by using deep learning models, LSTM (Long-Short Term

Memory) neural network algorithm to predict future bitcoin values. In this project, historical data of bitcoin values that were publicly listed on NSEPY website were used. This can be thought of as time series analysis, a specialized branch of statistics widely used in fields such as econometrics and operations research.

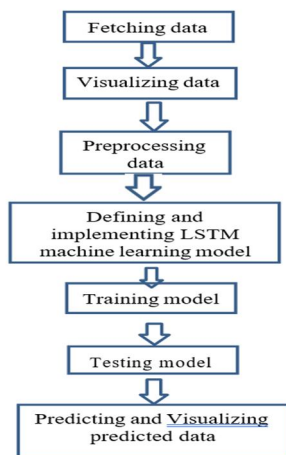


Figure 1. Work Flow Diagram



Figure 2. Overall Architecture for Cryptocurrency Price Prediction

A System architecture or systems architecture is the conceptual model that defines the structure, behaviour, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviours of the system

The above figure 2 explains the prediction of Bitcoin prices. First collected data is preprocessed to remove missing values, normalize the data, and convert it into a suitable format for machine learning algorithms. The preprocessed data is split into training and testing datasets, typically with a ratio of 70:30 or 80:20. Machine learning algorithms such as random forest regression, is trained on the training dataset to learn the patterns and relationships between the features and Bitcoin prices. Hyperparameter tuning can be performed to optimize the performance of the machine learning algorithms. The best-performing model can be selected based on metrics such as mean squared error, mean absolute error, and R-squared. The trained model can be tested on the testing dataset to evaluate its performance and generalization ability. Metrics such as mean squared error, mean absolute error, and R-squared can be used to measure the accuracy of the model. The model can be further evaluated by comparing its performance with other models or with the baseline performance of a simple model such as linear regression. Web scraping techniques can be used to extract real-time data on Bitcoin prices and other relevant features from cryptocurrency exchanges, news websites, and social media platforms. The predicted Bitcoin prices can be displayed to the user in a user interface such as a mobile app or web application.

IV. MACHINE LEARNING MODELS

Performing machine learning involves creating a model that is trained on some training data and then can process additional data for predictions. Different types of models have been used and investigated for machine learning systems. Some of models are ANNs, SVMs, Decision trees and Regression models.

1. Artificial Neural Networks

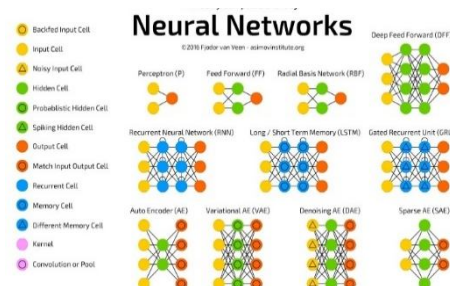


Figure 3. Neural Networks Model

Artificial neural networks (ANNs), or connectionist systems, are computational systems vaguely inspired by the biological neural networks that make up animal brains. Such systems "learn" to perform tasks by considering the examples, usually without being programmed with any task-specific rules.

2. Decision Trees

Decision tree learning uses a decision tree as a predictive model to go from observations about an item (represented in branches) to inferences about a target value of an item (represented in leaves). It is one of the most predictive modelling approaches used in statistics, data mining and machine learning.

3. Support-Vector Machines

Support vector machines (SVMs), also known as support vector networks, are a set of related supervised learning methods used for classification and regression. Based on a set of training examples, each labelled as belonging to one of two categories, the Support Vector Machine training algorithm builds a model that predicts whether a new example falls into one category. The Support Vector Machine training algorithm is a probabilistic, binary, linear classifier, although there are methods such as Platt scaling for using SVM in a probabilistic classification setting. In addition to performing linear classification, SVMs can efficiently perform nonlinear classification using what is called a kernel trick, implicitly mapping their inputs to high-dimensional feature spaces.

4. Regression analysis

Regression analysis includes a large number of statistical methods to estimate the relationship between the input variables and their associated features. Its most common form is linear regression, where a single line is drawn that best fits the given data according to a mathematical criterion, such as ordinary least squares. The latter is often augmented with regularization methods to mitigate offset and bias, as in ridge regression. When dealing with non-linear problems, the go-to models include polynomial regression (used for example to fit a trend line in Microsoft Excel), logistic regression (often used in statistical classification), or even kernel regression, which introduces non-linearity by using a kernel trick to implicitly map input variables into multidimensional space.

5. Training Model

Predicting bitcoin prices is a complex task and involves various factors. Machine learning can be a useful tool for this purpose, but it's important to note that predicting bitcoin prices with high accuracy is challenging due to the inherent uncertainty and randomness in financial markets. However, you can build models that capture patterns and trends based on historical data. Here's a general outline of how you can approach training a model for bitcoin price prediction using machine learning.

6. Data Fetching

The historical bitcoin data required for the machine learning model training is fetched online directly from the NSE website using nsepy module available in python libraries. The bitcoin data fetched from nsepy module will be in form of time-series. So, the data can be treated as Data frame data structure in pandas python module.

Data fetching is a crucial step in bitcoin price prediction using machine learning. It involves gathering historical and real-time bitcoin market data to train and evaluate the predictive models. The quality and relevance of the data significantly impact the accuracy of the predictions.

Financial Data APIs: APIs like Alpha Vantage, Quandl, and IEX Cloud offer access to historical and real-time bitcoin prices, financial ratios, and other relevant market data.

Financial Data Platforms: Platforms like Yahoo Finance, Google Finance, and Investing.com provide free access to historical and real-time bitcoin prices, along with company financials and market news.

Web Scraping: For specific data requirements, web scraping can be employed to extract data directly from financial websites or news sources.

Datasets Analysis

Bitfinex Open price: the open represents the first price traded during the candlestick.

High price: the high is the highest price traded during the candlestick.

Low: the low shows the lowest price traded during the candlestick.

Close: the close is the last price traded during the candlestick.

Volume (btc): volume, in btc traded in the stock market during a given measurement interval.

Volume (currency): volume, in used, traded on stock market during a given measurement interval.

Weighted price: measure of the average price.

Data Preprocessing

Data preprocessing is a crucial step in preparing data for machine learning models. It involves cleaning and transforming the data to make it ready for modeling. The main goal is to remove any inconsistencies, redundancies, or errors in the data and convert it into a format that the machine learning algorithm can use. Remove unwanted columns. Check nan values.

Test/Train

RANDOM FOREST/LINEAR REGRESSION – GRID SEARCH METHOD

Now we will fit the random forest algorithm to the training set. To fit it, we will import the random forest classifier class from the sklearn.ensemble library. Grid search is a process that searches exhaustively through a manually specified subset of the hyperparameter space of the targeted algorithm. Random search, on the other hand, selects a value for each hyperparameter independently using a probability distribution.

EVALUATION- TESTING

Rmse (root mean square error) - root mean square error (rmse) is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed.

R (pearson's correlation coefficient) the pearson's correlation coefficient depicts the linear association between two variables. It helps to figure if two sets of data move in the same direction.

WEB SCRAPING

Send an http request to the url of the webpage you want to access. The server responds to the request by returning the html content of the webpage. For this task, we will use a third-party http library for python-requests. Once we have accessed the html content, we are left with the task of parsing the data. Since most of the html data is nested, we cannot extract data simply through string processing. One needs a parser which can create a nested/tree structure of the html data. There are many html parser libraries available but the most advanced one is html5lib. Now, all we need to do is navigating and searching the parse tree that we created, i.e. Tree traversal. For this task, we will be using another third-party python library, beautiful soup. It is a python library for pulling data out of html and xml files.

V. RESULTS AND DISCUSSION

The results obtained by comparing the efficiency of the proposed system with the existing system. The purpose of this system will increase accuracy, fastest prediction, and consistency of results. It is useful to classify the cryptocurrency outcome from trained data set for accurate detection. This makes the future convenient for the user. Most of these ideas pertaining to the cryptocurrency domain revolve around trustworthiness of the results and transparency of the model.

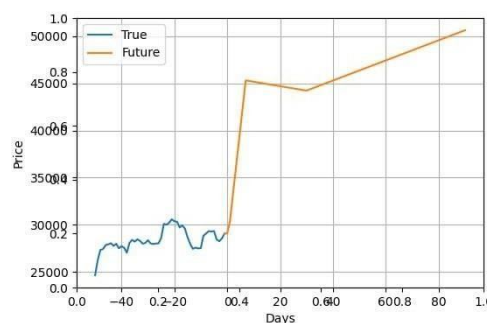


Figure 4. Accuracy Price Prediction

In this graph, the blue line represents the actual Bitcoin price over time, while the orange line represents the predicted Bitcoin price generated by the machine learning algorithm. The orange line is generated by the machine learning algorithm using a combination of historical data and mathematical models that identify patterns and relationships in the data. The algorithm learns from the historical data to identify trends and patterns that can be used to predict future prices.

The accuracy of the predicted prices depends on the accuracy of the historical data and the ability of the algorithm to identify meaningful patterns and relationships.

The above screenshot shows the current Bitcoin price is 29186.99, which means that as of now, this is the market price of Bitcoin as traded on various cryptocurrency exchanges.



Figure 5. User interface for cryptocurrency price prediction – Day 1



Figure 6. User interface for cryptocurrency price prediction – Day 7

On the other hand, the predicted first day price is 30319.96, which is the predicted price of Bitcoin for the first day in the future, based on the machine learning algorithm used for Bitcoin price prediction.

The current Bitcoin price is 29126.55, which represents the current market price of Bitcoin as traded on various cryptocurrency exchanges.

The predicted 7th day price of 45315.97, on the other hand, is the estimated price of Bitcoin for 7 days in the future, as predicted by the machine learning algorithm used for Bitcoin price prediction

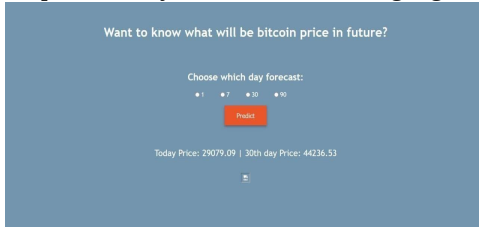


Figure 7. User interface for cryptocurrency price prediction – Day 30

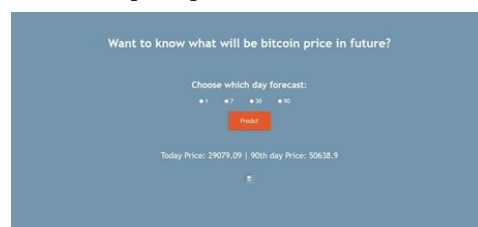


Figure 8. User interface for cryptocurrency price prediction – Day 90

The current Bitcoin price is 29079.09, which represents the current market price of Bitcoin as traded on various cryptocurrency exchanges.

The predicted 30 day price of 44236.53, on the other hand, is the estimated price of Bitcoin for 30 days in the future, as predicted by the machine learning algorithm used for Bitcoin price prediction.

The current Bitcoin price is 29079.09, which represents the current market price of Bitcoin as traded on various cryptocurrency exchanges.

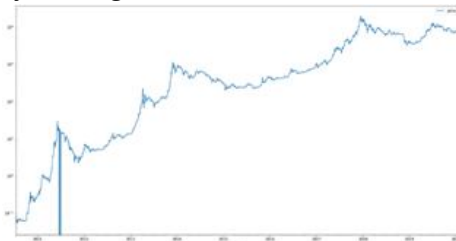


Figure 9. Microsoft Bitcoin Price History



Figure 10. Microsoft Bitcoin Price History



Figure 11. Starbucks Bitcoin Price History

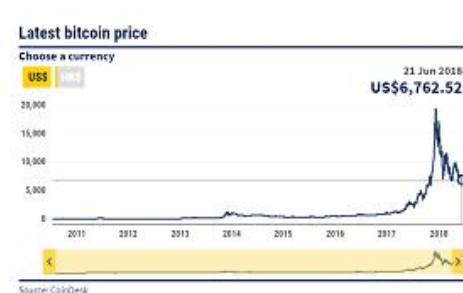


Figure 12. Twitch Bitcoin Price History

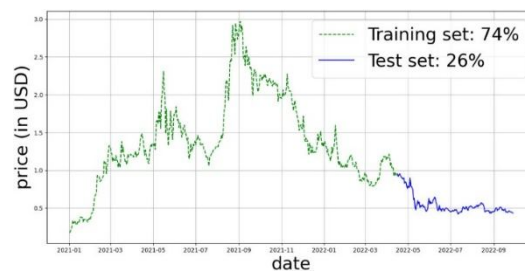


Figure 13. Tesla Bitcoin Price

VI. CONCLUSION AND FUTURE WORK

Conclusion

Based on the project's findings, it is evident that Bitcoin's price is influenced by various factors. The proposed predictive model, which incorporates multiple parameters, was able to accurately forecast Bitcoin's closing price for the following day. The model can be helpful for investors and traders looking to make informed decisions about buying or selling Bitcoin. Furthermore, the project highlights the importance of data analysis and machine learning in understanding and predicting trends in the cryptocurrency market. As blockchain technology continues to evolve, it is likely that Bitcoin will remain a crucial player, and its value will continue to fluctuate. By monitoring and analyzing relevant data points, we can gain insights into the future of Bitcoin and wide range of the bitcoin supermarket and has a large scope in cryptocurrency

Future Works

Improved data analysis: As more data becomes available, machine learning models can become more accurate at predicting the value of cryptocurrencies. This can include incorporating sentiment analysis of news articles and social media posts related to cryptocurrencies, as well as using data from cryptocurrency exchanges to better understand market trends.

Integration of blockchain technology: Blockchain technology can be used to improve the accuracy and transparency of cryptocurrency prediction models. For example, smart contracts can be used to automate trades based on predicted price movements.

Collaboration between industry experts: Collaboration between experts in cryptocurrency, blockchain technology, and data science can lead to more accurate predictions. By pooling their knowledge and resources, they can create more sophisticated prediction models.

Adoption of decentralized prediction markets: Decentralized prediction markets allow individuals to make predictions on the value of cryptocurrencies, and the wisdom of the crowd can be used to generate more accurate predictions. **Incorporation of macroeconomic factors:** Cryptocurrency prediction models can be improved by taking into account macroeconomic factors such as inflation rates, interest rates, and geopolitical events. These factors can have a significant impact on the value of cryptocurrencies, and incorporating them into prediction models can lead to more accurate predictions.

REFERENCES

- [1] Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach Mohammed Mudassir¹ Shada Bennbaia¹ Devrim Unal Mohammad Hammoudeh 16 June 2020 Springer-Verlag London Ltd., part of Springer Nature 2020.
- [2] Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis S M Rajua Ali Mohammad Tarif.
- [3] Crypto-Currency price prediction using Decision Tree and Regression techniques Karunya Rathan¹, Somarouthu Venkat Sai², Tubati Sai Manikanta IEEE Xplore Part Number: CFP19J32-ART 2019.
- [4] Prediction of Bitcoin Prices with Machine Learning Methods using Time Series Data Zonguldak, Türkiye 2018 IEEE
- [5] Aggarwal, Apoorva, Isha Gupta, Novesh Garg, and Anurag Goel. 2019. Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction. Paper presented at 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, August 8–10.
- [6] Akyildirim, Erdinc, Oguzhan Cepni, Shaen Corbet, and Gazi Salah Uddin. 2021. Forecasting mid-price movement of Bitcoin futures using machine learning. *Annals of Operations Research*, 1–32.
- [7] Awoke, Temesgen, Minakhi Rout, Lipika Mohanty, and Suresh Chandra Satapathy. 2021. Bitcoin Price Prediction and Analysis Using Deep Learning Models. In *Communication Software and Networks*. Singapore: Springer, pp. 631–40.

- [8] Basak, Suryoday, Saibal Kar, Snehanishu Saha, Luckyson Khaidem, and Sudeepa Roy Dey. 2019. Predicting the direction of bitcoin market prices using tree-based classifiers. *The North American Journal of Economics and Finance* 47: 552–67
- [9] Baur, Dirk G., and Lai Hoang. 2021. The Bitcoin gold correlation puzzle. *Journal of Behavioral and Experimental Finance* 32: 100561. Baur,
- [10] Dirk G., and Thomas Dimpfl. 2021. The volatility of Bitcoin and its role as a medium of exchange and a store of value. *Empirical Economics* 61: 2663–83.
- [11] Blake, R. 2019. An Econometric Analysis of the Relationship between Bitcoin & Gold. Available online: https://medium.com/@blake_richardson/an-econometric-analysis-of-the-relationship-between-bitcoin-gold-2018-584b4c63a17 (accessed on 21 September 2022).
- [12] Carbó, José Manuel, and Sergio Gorjón. 2022. Application of Machine Learning Models and Interpretability Techniques to Identify the Determinants of the Price of Bitcoin. Banco de Espana Working Paper No. 2215. Available online: <https://ssrn.com/abstract=4087481> (accessed on 1 October 2022).
- [13] Chen, Wei, Huilin Xu, Lifan Jia, and Ying Gao. 2020a. Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting* 37: 28–43
- [14] García-Medina, Andrés, and Toan Luu Duc Huynh. 2021. What Drives Bitcoin?
- [15] “Emotional and Sentimental Prediction from Textual Comments Using Machine Learning Techniques” Published in *NeuroQuantology*, August 2022, Volume 20, Issue 10, Page 1735-1749, DOI: 10.14704/nq.2022.20.10.NQ55157. SNIP – 0.38.
- [16] Safe Trade - A Stock Recommender using Machine Learning Algorithms, Julian, A., Ganeshreddy, E., Reddy, V.R.D., Somasekhar, D.V.V., 2023 International Conference on Computer Communication and Informatics, ICCCI 2023, 2023
- [17] “Sensor Based Automatic Neonate Respiration Monitoring System”, published in *International Journal of Innovative Technology and Exploring Engineering*, Vol.9, Issue 2, December 2019, pg.1296-1299.