PRICE PREDICTION MODEL OF BITCOIN USING DECISION TREE CLASSIFICATION

[1]KANAKA TULASI PRAGADAREDDY, [2]NAKKA MAMATHA VIOLET, [3]KALAGARA SURENDRANATH, [4]NAMUDURI VYAGHRESWARA KIRAN DATHA

[1,2,3]Assistant Professor, Department of Computer Science and Engineering, BVC College of Engineering, Andhra Pradesh, India

[4]B. Tech Student, Department of Computer Science and Engineering, BVC College of Engineering, Andhra Pradesh, India

ABSTRACT: Over the past few years, Bitcoin has been a topic of interest of many, from academic researchers to trade investors. Bitcoin is the first as well as the most popular cryptocurrency till date. It made Bitcoin a very profitable market for investment but Bitcoin saw many ups and down as well. Trading of Bitcoin has proved to be very profitable to many people but the risk in trading is huge as the market of Bitcoin is very volatile. However, many previous works on Bitcoin price prediction forecast short-term Bitcoin price, have low accuracy and have not been cross-validated. This paper presents Price Prediction model of Bitcoin using Decision Tree Classification. The dataset till current date is taken with open, high, low and close price details of Bitcoin value. Exploiting the dataset machine learning module is introduced for prediction of price values. The aim of this work is to derive the accuracy of Bitcoin price prediction using different machine learning algorithm and compare their accuracy. Decision Tree outperforms other classifiers with accuracy of 97 percent, precision 96.7 percent, and recall 96 percent.

KEYWORDS: Bitcoin price prediction, using Decision Tree Classification, Accuracy, Machine Learning.

I. INTRODUCTION

Bitcoin is a crypto currency which is used worldwide for digital payment or simply for investment purposes. Bitcoin is decentralized i.e. it is not owned by anyone. Transactions made by Bitcoins are easy as they are not tied to any country. Investment can be done through various marketplaces known as "bitcoin exchanges". These allow people to sell/buy Bitcoins using different currencies. The largest Bitcoin exchange is Mt Gox [1].

Bitcoins are stored in a digital wallet which is basically like a virtual bank account. The record of all the transactions, the timestamp data is stored in a place called Blockchain. Each record in a blockchain is called a block. Each block contains a pointer to a previous block of data. The data on blockchain is encrypted [2].

Bitcoin is one of the most valuable cryptocurrency in the world and the world's first decentralised cryptocurrency, in other terms it is a form of electronic cash. A cryptocurrency in essence is a digital asset meaning it exists in a binary format and comes with the right to use and the data that do not possess that right are not considered assets, and it is designed to work as a method of exchange that uses robust cryptography to ensure reliable financial transactions, and substantiate the transfer of assets. Bitcoin is the new kind of money and an innovative payment network. It is an opensource and its design is public and nobody owns or administers it and everyone can take part to make assets. After the release of Bitcoin in 2009, over 4000 alternative variants of Bitcoin which is referred to as altcoins.

The Bitcoin's simply and similar like a stock but in different way. There are various algorithm using machine learning are utilized on price prediction on stock value. The features influencing Bitcoin are unique. For investors it is mandatory to predict Bitcoin prices. Bitcoin price do not affect by business announcements or government

announcements and it is not at all like securities exchange. Thus, we exploit machine learning techniques to foresee the cost of Bitcoin [3].

Volatility as a proportion of value fluctuations, it significantly affects exchange methodologies and investment choices just as on alternative estimating and proportions of fundamental risk. Bitcoin, as a pioneer in the blockchain money, it assumes a predominant job in an entire cryptographic money showcase capitalization. Subsequently, it is of extraordinary interest are growing these days on data mining and machine learning network to most likely predict Bitcoin value variances.

While senders of conventional electronic installments are generally distinguished, investors of bitcoin work in anonymous. No central authority is available for bitcoin, investors don't have to recognize themselves when sending bitcoin to another client. Investors are identified by the address of wallet. It is mandatory for Exchanges to check identity of the clients handled by them, they are not permitted to make buy or sell trade without checking their identity [4]. Bitcoin is more secure for investors. The objective of the proposed study is for price prediction of bitcoin by feature selection of different machine learning techniques. Intuitively, idea is to first transform order book data into features over time, referred as feature series and then to develop prediction models to consume volatility and feature series simultaneously.

The following is a breakdown of the paper's structure: Section II summarizes the literature review; Section III explains the technique and how the model work; and Section IV displays the Results and Analysis. The paper's conclusion is in Section V.

II. LITERATURE SURVEY

H. Jang and J. Lee, et. al. [5] analyzing what has been done to predict the U.S. stock market. the conclusion of his work is the mean square error of the prediction network was as large as the standard deviation of the excess return. However, the author is providing evidence that several basic financial and economic factors have predictive power for the market excess return. Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, Chang Hun Kim, et. al. [6] studied the price fluctuation model in crypto-currency using comments given by users. Apart from bitcoin, they considered Litecoin and Ripple, which are the next two major currencies. User crawled data is taken for their study and they segregated sentiment opinions of five types such as very positive, positive, neutral, negative and very negative.

A, Vishal & Sonawane, Sheetal, et. al. [7] presents a method for finding the sentiment using multiple machine learning techniques. The author collected twitter dataset and performed the analysis, they handle various stages of pre-processing techniques such as removing URL, spelling corrections, and emoticons are replaced with their corresponding polarity values. They used classification techniques such as Naive Bayes algorithm and Support Vector Machine. Using NLTK package, they find the polarity of tweets, from which they arrived positive and negative sentiment.

P. Mondal, L. Shit, and S. Goswami. et. al. [8] conducted a study on effectiveness of time series analysis basically ARIMA model for predicting stock market price in India. They have worked with past twentythree months data. Their models can predict the stock market movement with an accuracy of above 85%. M. Daniela and A. BUTOI, et. al. [9], a comparison between Multi-Layer Perceptron (MLP) and Non-linear

autoregressive exogenous (NARX) model is made. They conclude that MLP can also be used for stock market prediction even though it does not outperform NARX model in price prediction. The authors made use of MATLAB's neural network toolbox to build and evaluate the performance of the network.

F. A. d. Oliveira, L. E. Zarate, M. d. A. Reis and C. N. Nobre, et. al. [10] instead of directly forecasting the future price of the stock the authors predicts the trend of the stock. The trend can be considered as a pattern. They perform both short-term predictions day or week predictions and also long-term predictions months they found that the latter produced better results with 79% accuracy. Another interesting approach the paper reflects is the performance evaluation criteria of the network. Based on the predicted output the performance evaluation algorithm decides to either buy and sell or hold the stock.

III. PRICE PREDICTION MODEL OF BITCOIN

The architecture of Price Prediction model of Bitcoin using Decision Tree Classification is represented in below Fig. 1. As Bitcoin is a kind of stock traded in stock market, dataset will be available in plenty with all time intervals. Live data from 2011 to till date is collected from quandl.com, which provided us the most comprehensive bitcoin price in date wise data. Dataset is extracted to CSV file. Though there are many authorize websites are available for collecting bitcoin dataset for study, CoinMarketCap is one of the other authorized websites, which provides the transactions that bitcoin traded for the 24 hours of a day. These data are fed from various exchanges handling crypto currency. The dataset collected from 2011 to till data is plotted on the below chart, Fig. 2, which

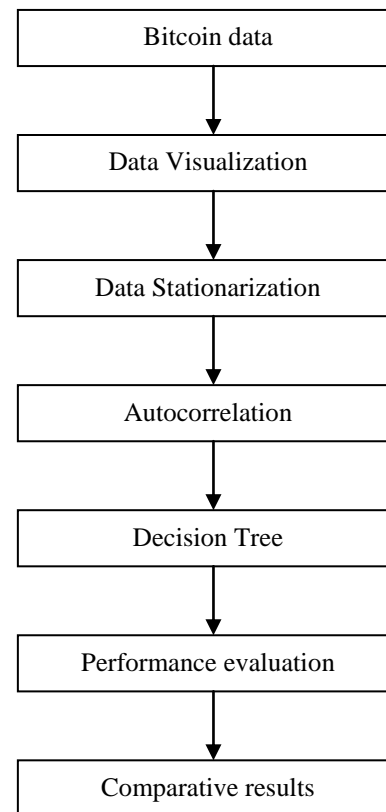clearly depicts that bitcoin is a positive market.

Bitcoin data
↓
Data Visualization
↓
Data Stationarization
↓
Autocorrelation
↓
Decision Tree
↓
Performance evaluation
↓
Comparative results

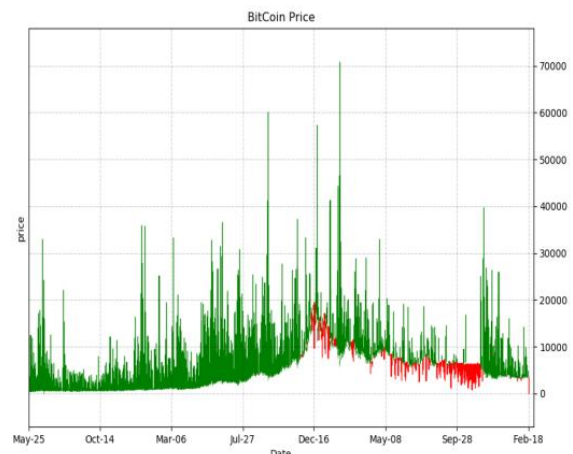**Fig. 1: ARCHITECTURE OF PRICE PREDICTION MODEL OF BITCOIN**

Fig. 2: DATA VISUALIZATION OF BITCOIN PRICE

To check the stationarity, Augmented Dickey-fuller test has been used based on the nature of our dataset. The Augmented Dickey Fuller test is a sort of factual test

called a unit root test and is good for larger and complicated time series data. The instinct behind a unit root test is that it has decided how unequivocally a time series was characterized by a pattern. There are number of unit root tests and ADF (Augmented Dickey-Fuller Test) is a standout amongst the most generally utilized. Dickey-Fuller test includes an intercept and time trend.

Autocorrelation is a measurement of the inter connection inside a time series. It is a method for estimating and clarifying interior relationship between perceptions in a time series analysis. According to the concept of auto-correlation, if the first element is closely related to the second, and the second to the third, then the first element must also be somewhat related to the third one. Autocorrelation function (ACF) helps to determine the order of moving average (MA) model in the dataset.

Decision tree is one of the learning models that are generally utilized in classifications. In this strategy, we split dataset into at least two sets. Internal nodes in Decision tree indicate a test on the features, branch portrays the result and leafs are decisions made after subsequent processing. Decision Tree works as follows:

- Place the best feature of the dataset as root of tree.
- Split the dataset into train and test set. Subsets ought to be made so that every subset contains information with a similar feature attribute.
- Above steps repeated on every subset until we get leaf in the tree.

The result is compared with the score value to identify the performance parameters value.

IV. RESULT ANALYSIS

We downloaded dataset from quandl.com with necessary authentication keys. The data downloaded contains up-to date data. The dataset is 80% considered as train set and 20% considered as test set. Decision tree Machine learning algorithm is applied. Five days forecast price prediction is done using decision tree. The values are compared. Accuracy, Precision and Recall are used parameters in this study for performance analysis.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \dots (1)$$

$$Recall = \frac{TP}{TP + FN} \dots \dots (2)$$

$$Precision = \frac{TP}{TP + FP} \dots \dots (3)$$

Where, TN represents true negative, TP represents true positive, FP represents false positive and FN represents false negative. The comparative analysis among different machine learning classifiers as Convolution neural networks (CNN), Artificial neural networks (ANN) and Decision Tree (DT) based Price Prediction model of Bitcoin model is represented in below Table. 1.

Table 1: PERFORMANCE OF DIFFERENT CLASSIFIERS

Classifiers	Accuracy	Recall	Precision
CNN	87	87.5	88
ANN	85	85.2	84.3
DT	97	96	96.7

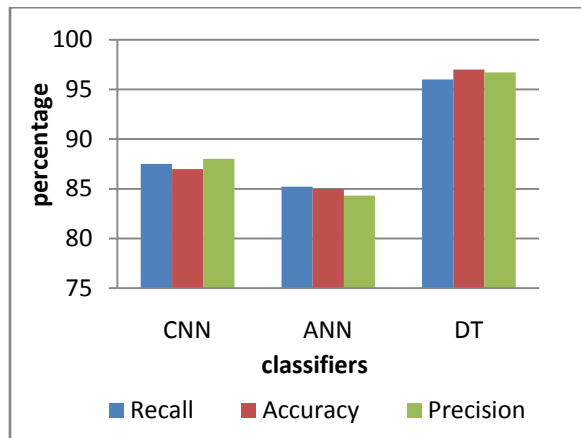The graphical representation of Recall, Accuracy and Precision is shown in below Fig. 3.

Fig. 3: COMPARATIVE PERFORMANCE ANALYSIS

We can see that the best result was obtained by using the Decision Tree (DT) based Price Prediction model of Bitcoin model. Decision Tree outperforms other classifiers with accuracy of 97 percent, precision 96.7 percent, and recall 96 percent.

V. CONCLUSION

In this paper, Price Prediction model of Bitcoin using Decision Tree Classification is described. Bitcoin is a booming crypto-currency market, and various researches have been studied in fields of economics and price prediction. The aim of this work is to derive the accuracy of Bitcoin price prediction using different machine learning algorithm. The dataset till current date is taken with open, high, low and close price details of Bitcoin value. The dataset is 80% considered as train set and 20% considered as test set. Decision tree is one of the learning models that are generally utilized in classifications. Accuracy, Precision and Recall are used parameters in this study for performance analysis. We can see that the best result was obtained by using the Decision Tree (DT) based Price Prediction model of Bitcoin model.

VI. REFERENCES

[1] S.Venkata Lakshmi, N.G. Swadthi, V.Suba Shree, M. Swethamura, "Investigation on Bitcoin Prognosis", 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Year: 2021

[2] E. Mahendra, H. Madan, S. Gupta, S. V. Singh, "Bitcoin Price Prediction Using Deep Learning and Real Time Deployment", 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Year: 2020

[3] Thearasak Phaladisailoed, Thanisa Numnonda, "Machine Learning Models Comparison for Bitcoin Price Prediction", 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), Year: 2018

[4] Huisu Jang, Jaewook Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information", IEEE Access, Volume: 6, Year: 2018

[5] H. Jang and J. Lee, "An Empirical Study on Modelling and Prediction of Bitcoin Prices with Bayesian Neural Networks based on Blockchain Information," in IEEE Early Access Articles, 2017.

[6] Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, Chang Hun Kim, "Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies" Interdisciplinary Program in Visual Information Processing, Korea University, Seoul, Korea Aug 2016

[7] A, Vishal & Sonawane, Sheetal, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications, 2016, 139. 5-15. 10.5120/ijca2016908625

[8] P. Mondal, L. Shit, and S. Goswami. "Study of effectiveness of time series modeling (arima) in forecasting stock prices." International Journal of Computer Science, Engineering and Applications (IJCSEA), 29, April 2014.

[9] M. Daniela and A. BUTOI, "Data mining on Romanian stock market using neural networks for price prediction". informatica Economica, 17,2013

[10]F. A. d. Oliveira, L. E. Zarate, M. d. A. Reis and C. N. Nobre, "The use of artificial neural networks in the analysis and prediction of stock prices," in IEEE International Conference on Systems, Man, and Cybernetics, 2011.