Price Forecasting and Analysis of Bitcoin

Marriammal R[1], Reni Hena Helen R[2], Rubika M[3], Sowbhagya T[4]

Assistant Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4]
Dhanalakshmi College of Engineering, Chennai, India

Abstract: *Bitcoin, the king of cryptocurrencies, is central to blockchain technology. A fixed amount of bitcoins is required for each transaction stored in the blockchain. The price of bitcoins fluctuates wildly and is unaffected by any company or marketing techniques, creating both curiosity and terror in the minds of traders. It is possible for consumers to study and invest in bitcoin by anticipating the bitcoin price, which promotes the use of digital money. As a result, a high-prediction-rate prediction model is required. The goal of this project is to employ a variety of machine learning models to predict the price of bitcoin. The best model for predicting bitcoin value is given based on the error percentage of these machine learning algorithms.*

Keywords: Bitcoin, Blockchain Technology, Grid Search, Cryptocurrency

I. INTRODUCTION

Since the barter system, trade associated investment is an integral part of economic development of a country. Over time, major changes are created in trade which leads to the expansion of the national economy. Changes such as the conversion from the barter system to primarily cash-based exchanges, the introduction of digital currency such as bitcoin (BTC) are significant

The initial value of a BTC in the year 2009 was $ 0.0008 and over the years the value of BTC has drastically increased to $ 46,434.40. This indicates the recognition of BTC in this company. This cryptocurrency is widely used to enter information into the blockchain. Blockchain technology is used by different countries in numerous fields such as healthcare, banking, business etc mainly due to its high level security and scalability. Hence the BTC forecast is crucial to persuade traders to take a position in BTC which can lead to the expansion of the world economy.

II. LITERATURE REVIEW

A. Bitcoin value prediction there's already abundant relevant analysis within the literature on the subject of Bitcoin price prediction. Here during this section, we have a tendency to in short introduce some recent papers on this topic.[3] used Google News knowledge and Reddit messages data for sentiment analysis victimization Valence Aware wordbook and sentiment thinker (VADERical price of Bitcoin, Litecoin, and Ethereum are combined for coaching the machine learning models. (LSTM, GRU, 1D-CNN). All information was collected from Gregorian calendar month 1, 2018, to Nov 20, 2019, with a measure of 1 hour, except Google News, that was collected on a per day basis. Once mistreatment of all the features, the LSTM model with the sum of all the sentiment analysis values offers an all-time low check RMSE of 434.87. However, a stimulating finding is that, while not adding sentiment analysis data, the LSTM model provides a test RMSE of 116.36, which is the overall low. Model, multilayer perceptron neural network, statistical procedure Linear regression, AdaBoost, Bayesian crest regression, call tree, ElasticNet and their hybrid, i.e. average of all models. completely different cryptocurrencies are considered, including Bitcoin (BTC), Ethereum (ETH), Electronium (ETN), Ripple (XRP), ZEC money (ZEC) and Monero (XMR).

Twitter sentiments associated with Google Trends knowledge were used as input features.The authors conducted an empirical experiment by commercialism $100 on the BitBay cryptocurrency exchange over one month and the account balance stood at $114.82. In contrast, once they used CryptoBot, a widely known tool for cryptocurrency trading, they managed to convert $100 into $102.45 at intervals identical period.[5] collected data for 5 currencies in OTC, currency, and contract market, particularly Bitcoin (BTC), Ethereum (ETH),Tether USD (USDT), EOS, Ripple (XRP), litecoin (LTC).

As a result, it also scanned day-to-day people's criticisms on online forums from the 0.33 zone in 2018 to the main zone

in 2019. The specific statistical objects consist of the content material and the amount of title, comments, responses and time of publication. . They concluded that the LSTM version performs very well and adding comment sentiment can greatly improve the accuracy of the prediction. Additionally, customers consult buying and selling records and comment records 1-7 days earlier than when making buying and selling decisions. [6] examined Bitcoin's debit predictability, use of BTC-USD intraday yields, and investor sentiment and investor interest as measured by StockTwits records. They also don't forget an ability opposite to causation - returns that impact investor sentiment - by running Granger causation tests. The end result indicates that the sentiment element is statistically vast once they don't forget the delay of up to 15 minutes, and investor interest, measured by the full amount of messages over a given time frame, doesn't always impact. significantly on the performance of Bitcoin.

B. Impact of COVID-19 on the financial market Some research has targeted the impact of COVID-19 on the inventory market. [7] tested the effect of COVID-19 on rising inventory markets.It determined that the negative effect decreased steadily and everything started to decrease with the help of the mid-April usage. It also determined that the best effect occurs in Asian markets and the least in European bull markets. [8] tested the March 2020 inventory market crash caused by the use of COVID-19. Several sectors show extraordinary influences. Shares in natural gas, food, health care and software programs achieve excessive effective returns. Shares of oil, real estate, entertainment and hospitality are falling dramatically.

[9] also tested the predictability of inventory return during the COVID-19 crisis. A new and robust Cauchy error lasso technique was followed for predictive regressions. The aftermath confirmed that corporate bonds, all levels of funding and excessive yields, had tremendous predictive power. The cryptocurrency market is similarly studied during COVID-19.

The exploration of the cryptocurrencies market potency before and once the COVID-19 pandemic through a multifractal analysis was conducted in [10]. COVID-19 was unconcealed to have an effect on the efficiency of all the 5 cryptocurrencies.

In [11], thirty six applied math tests are performed to envision for variations between periods of your time (pre- versus throughout COVID-19 pandemic samples) on the one hand, furthermore as check for differences between markets (cryptocurrencies versus stocks), on the opposite hand. Crypto currency showed more instability and more irregularity during the COVID-19 pandemic compared to international stock markets
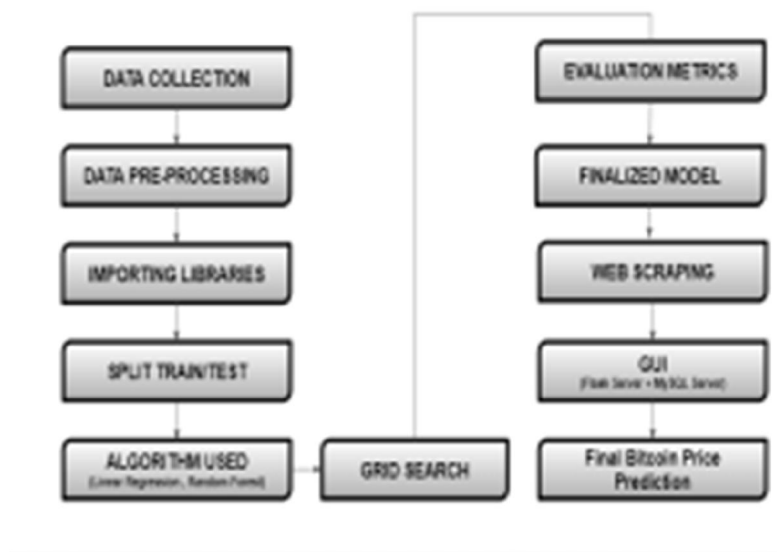
III. FLOW DIAGRAM

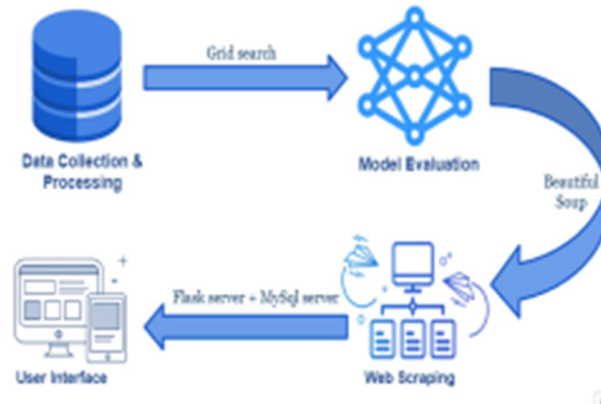Figure 1-Flow Diagram

IV. BLOCK DIAGRAM

Figure 2-Block Diagram

V. DATASET

The Bitcoin dataset accumulated and saved as a bitcoin.csv document is considered as a dataset, which is divided into a learning set and a control set. We have taken into account 80% of the information as the school enters so that our device knows all the teaching rules of the version. The last 20% of the information is taken into account as a control for the prediction of the final result. These technical signals are calculated using exceptional intervals which include the end of the day, 7, 30 and ninety days. End-of-day final expenses are treated as gross values.

VI. DATA PRE-PROCESSING

Data preprocessing is a system for preparing raw records. A true international record usually contains noises, missing values, and possibly in an unusable layout that cannot be immediately used for the device to know the patterns. Data preprocessing is required to clean up records and make it appropriate for a device to know the version, which will also increase the accuracy and performance of a device to know the version. During preprocessing, the dataset was mixed and split into two sets: training set and validation set. 20% of the data was saved for validation and 80% of the data was used for training.

VII. IMPORTING LIBRARIES

1. Pandas: Pandas is an open-source Python package most ordinarily used for knowledge science/data analysis and machine learning tasks.
2. Matplotlib: Matplotlib may be a cross-platform, knowledge visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open supply different to MATLAB
3.Sklearn: Scikit-learn could be a free machine learning library for Python. It offers varied algorithms like support vector machine, random forests, and k-neighbors, and it additionally supports Python numerical and scientific libraries like NumPy and SciPy .

VIII. SPLIT TRAIN/TEST DATA

The data set is divided into two parts. The first subset is known as the training data and is part of the actual data set that is fed to the machine learning model to discover and learn patterns. This is how our model is trained. The other is known as test data. The results of this correlation are used to refine the model. This iterative process is called model fitting. The correctness of the training data set or validation data set is critical to the correctness of the model. A machine language training model is a feed process involving knowledgeable nursing algorithm rules to determine and learn meaningful values for all relevant functions. In this module, we will investigate the use of a machine learning model trained with a test dataset. Quality assurance is essential to ensuring that your code works according to your requirements. All

parameters used exclusively for program testing must be clearly specified in the specification document.

IX. ALGORITHM USED

A. LINEAR REGRESSION

Linear regression is the way to find a line that best fits the available data points on plot, so we can use it to predict the output values for the inputs present in the dataset that we don't have with the belief that these results will come into A simple linear regression real life example might mean that you are tracing the relationship between revenue and temperature, with a sample size for revenue as the dependent variable. In the case of multiple variable regression, you can find the relationship between temperature, pricing, and the number of workers to revenue.
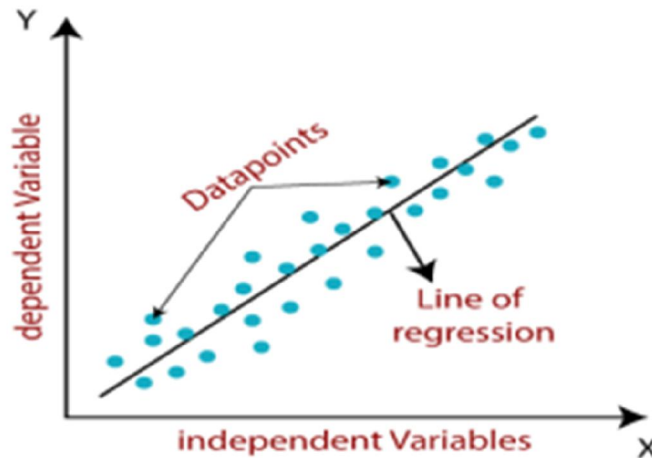

Figure 3-Linear Regression Diagram

B. RANDOM FOREST

Random Forest is a machine learning algorithm that is supervised and used for classification and regression problems. It builds a decision tree on different samples and takes their majority vote for classification and average in case of regression. The random forest model is trained to predict the closing price on the next day. At the highest level, random forest is a set of decision trees used for classification and regression activities. Once grown, each decision tree classifies an unlabeled point by voting, and the random forest carries the label or value with the most votes.

XI. EVALUATION METRICS

The mean absolute error (MAE) is the loss function, like the MSE, it only uses the distance between the predicted and predicted values but does not consider the nth direction for this difference. The following formula is used to calculate
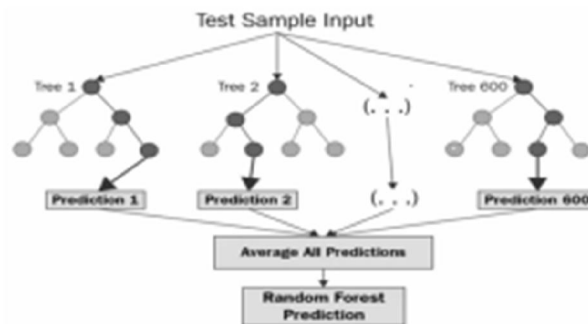
$$MAE = \frac{1}{n} \Sigma_{(i=1)}^{n} |e_i|$$

Figure 4-Random Forest Diagram

732

GRID SEARCH

Grid Search (GS) is an ingenious approach that attempts to explore the data space in depth using a manually defined subset of hyperparameters from the search space for the selected algorithm while random search selects values for all hyperparameters independently based on their probability distribution. Fine Tuning means finding the best parameters for machine learning algorithms to improve results. Optimized planning is an effective practical step that can lead to a significant increase in the production of the random forest method. The optimal parameters can be found automatically in the grid search.

Figure 5-Grid Search

The mean square error between the predicted and expected values. The following formula used to calculate

$$MSE = \frac{1}{n} \Sigma_{(i=1)}^{n} (e_i^2)$$

Root Mean Squared Error (RMSE) : If the effect of the MSE is derived, another loss function is constructed as the "square root of error" which is briefly shown with RMSE.

$$RMSE = \sqrt{\frac{1}{n} \Sigma_{(i=1)}^{n} (e_i^2)}$$

XII. FINALIZED MODEL

The finalized model shows that bitcoin price prediction is efficient using random forest regression algorithm comparing with linear regression algorithm. our proposed outperforms existing work accuracy.

In Mean Square Errors (MSE), the values are positive. It also has a more significant impact on large errors. The loss function computes
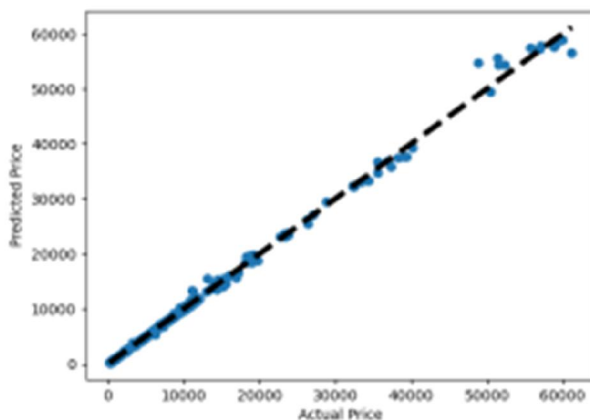

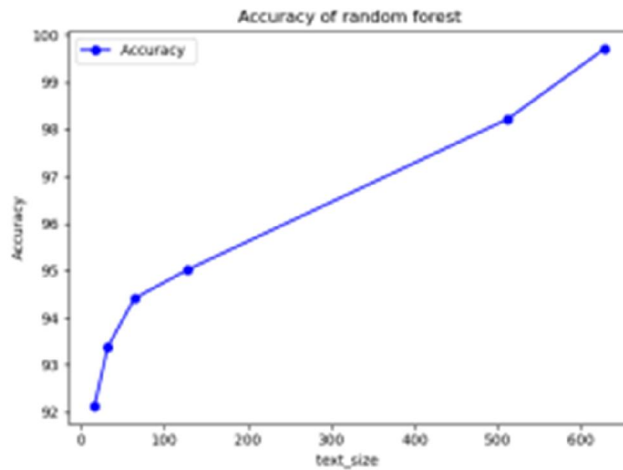Figure 6-Accuracy in Linear regression algorithm

Figure 7-Accuracy in Random forest algorithm

XIII. WEB SCRAPING

In this model, **Beautiful Soup** is used for scraping and parsing the data .

Step1: We are going to scrape data from the following url https://bitinfocharts.com//bitcoin/

Step2: Code for collecting data from url is implemented in Python using BeautifulSoup.

Step3: Collected data is used as per the requirement.

Figure 7-Web scraping

XIV. USER INTERFACE

- Flask is a Python-based web framework. It is classified as a microframework as it does not require unique tools or libraries. FrontEnd technology is very important as it is the user interface that a user will actually see on the client side, to simplify we have used core technologies on the frontend such as HTML, CSS and Javascript so that the user can view and interact with them data.

- The goal of frontend technology is to provide easy access to the tool without any hassle. Html is the backbone of websites, serves as the skeleton of any network application trendy version of html is html5 pos Css controls web page style which makes website components clean and trouble free at hand for the person without any problem on any instrument.

- MySQL is an open source Relational Database Model System (RDBMS) that follows Structured query language (SQL) as syntax and is licensed under GNU, is publicly available to everyone. MySQL is used by around 80% of existing business applications, including large companies like Facebook, NetFlix, Google, Amazon as a database system to store most of their data

XV. FINAL PRICE PREDICTION

- The result shows that bitcoin price prediction is efficient using the random forest regression algorithm. Random forest regression achieves approximately 97.5% accuracy in price prediction, while linear regression achieves 95.8% accuracy. Our proposed work exceeds the accuracy of existing work.

- User input:
- The user will open the website and log in (enter username and password) if they are already registered or register first and then log in.
- The user will be given an option in the heading Start Prediction. The user has to click on it.

XVI. CONCLUSION

Bitcoin price fluctuation is very unstable, most of the traders, middle class people are afraid to invest in bitcoin. Our model is built with the aim of predicting the price of bitcoin in the easiest way and with the highest accuracy. Our models allow us to predict the price of bitcoin after 7 , 30 and 90 days which will be displayed through the user interface.

REFERENCES

[1]. Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach Mohammed Mudassir1 Shada Bennbaia1 Devrim Una lMohammad 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