



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Bitcoin Prediction using Machine Learning

Prof. Karpagam V¹, Sankarapandi S², Santhoshkumar M³, Seeni Barith Ali S⁴, Vasanthraj P M⁵

Professor and Faculty Mentor, Dept. of Computer Science and Engineering, K L N College of Engineering, Madurai,
Tamil Nadu, India¹

B. E Student, Dept. of Computer Science and Engineering, K L N College of Engineering, Madurai,
Tamil Nadu, India²⁻⁵

ABSTRACT: This project aims to precisely predict prices of Bitcoin using the analysis of various parameters with significant impact on its market price. To meet this objective, we have sourced and examined information from different studies and incorporated valuable findings for instant application. Because Bitcoin is an extremely volatile and dynamic digital asset, various studies have put forth distinct methodologies to predict its prices. We use two well-known machine learning algorithms—K-Nearest Neighbor (KNN) and Support Vector Machine (SVM)—in this project to build predictive models. We chose these algorithms based on their efficiency in classification and regression problems with financial data. Input features were historical price trends, trading volume, and time-based factors. The performance of SVM and KNN models was analyzed in terms of traditional metrics like accuracy, mean squared error, and F1-score. A comparison was subsequently made to analyze which algorithm provides more accurate predictions. The result of this project adds to the realm of financial analytics by providing insights into algorithmic efficiency to forecast cryptocurrency prices..

I. INTRODUCTION

Bitcoin, the best-known and most traded cryptocurrency, has been incredibly price volatile from its early years, depending on market demand, investor attitudes, world economic events, and the actions of policymakers. Herein, Bitcoin price forecasting via machine learning—particularly through algorithms like Support Vector Machine (SVM) and K-Nearest Neighbor (KNN)—symbolizes a remarkable combination of finance analytics and computer intelligence. These models have the ability to reveal intricate patterns and trends in large data sets, providing valuable insights for traders and analysts. As computational power increases and real-time, high-quality data become more accessible, the precision and speed of machine learning-based forecasts are likely to increase, guiding the future of digital financial forecasting and facilitating more informed decision-making in the cryptocurrency market.

II. LITERATURE SURVEY

In [1] Sean McNally, Jason Roche, and Simon Caton (2024) in their work Predicting the Price of Bitcoin Using Machine Learning, IEEE PDP, applied Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models to predict Bitcoin prices. These models are especially good at learning sequential patterns and temporal relationships in time-series data. The work showed good performance in learning from past price trends. Nevertheless, one significant drawback of their method is the substantial computational expense and longer training time of deep models like LSTM..

In [2] In Anticipating Cryptocurrency Prices Using Machine Learning, Khadija Mushtag, Muhammad Rizwan, and Abdul Wahab (2023) implemented a few supervised learning classifiers like Random Forest, Support Vector Machine (SVM), and Gradient Boosting for forecasting cryptocurrency prices. The research underscored the need for preprocessing, feature selection, and hyperparameter optimization to increase model accuracy. Low error rates were achieved by SVM and Gradient Boosting models. The paper is strong in its comparison across models; however, there is a need to ensure the careful management of overfitting and scalability when the models are applied to larger data.

In [3] Mudassir et al. (2022), in their paper Time-Series Forecasting of Bitcoin Prices Using High-Dimensional Features published in Neural Computing and Applications, used Support Vector Regression (SVR), Decision Tree, and Random Forest models with heavy feature engineering. With the use of high-dimensional input data, the model attained

better accuracy in prediction. The study also identified the threat of the dimensionality curse, which required efficient feature selection methods. This paper emphasizes the advantage of feature diversity towards improving model performance but at the cost of increased computational burden.

III. EXISTING SYSTEM

Current systems of Bitcoin price forecasting typically integrate a combination of technical indicators, machine learning models, and sentiment analysis to enhance the accuracy of predictions. But their performance is usually inconsistent under different market scenarios. Models based mainly on historical data and technical indicators generally perform poorly during times of extreme market volatility because they cannot promptly respond to sudden changes or external shocks. One of the most important factors affecting the accuracy of these models is the quality of the training data. Missing values, noise, and outliers in historical price data can severely impair performance. In addition, most models are susceptible to overfitting, where they learn patterns that are specific to the training data—including the noise—instead of general trends, leading to poor prediction performance on new, unseen data. Another significant shortcoming is that the majority of systems disregard the influence of outside occurrences, including regulatory action, geopolitical situations, or sentiment in the market, which tend to have an immense impact on pricing Bitcoin.

IV. PROPOSED SYSTEM

By combining various machine learning algorithms like SVM, KNN, and LSTM through an ensemble method, the system takes advantage of the bias of each model towards improving its performance. The system also includes real-time data processing, enabling it to update predictions constantly with the most current market data. This is particularly important in the cryptocurrency market since prices can change quickly. Real-time data updates guarantee the system gives timely analysis, allowing traders to make decisions promptly. In addition, the suggested system considers not just historical price data but also technical indicators, sentiment analysis of the market, and macroeconomic factors from the outside world. Considering this diversity of factors by a holistic approach allows better precision in predictions by assuming a wider set of variables that affect the price of Bitcoin. By combining sentiment from news and social media, along with worldwide economic conditions, the system provides a more complete picture of possible market movement.

V. METHODOLOGY

5.1 Data Preprocessing

Data preprocessing is a step of getting the raw data ready and making it appropriate for a machine learning model. Data preprocessing is necessary steps for cleaning the data and making it appropriate for a machine learning model which also enhances the accuracy and efficiency of a machine learning model. We split the dataset up into training and test set, and standardize its features. Standardization is good practice as it reduces over fitting in cases where variance for some features may be higher than others. The LSTM model requires us to organize the data in blocks. Our data is grouped at one-minute intervals and we'll use blocks of 50 minutes to predict the next block. Bitcoin is currently being traded in a great deal of online exchanges with the largest being BitFinex, Bitthumb, GDAX, Poloniex, and Bitstamp. I gathered data from Bitstamp as it's one of the older exchanges and have had no hacking or market manipulation issues. We have gathered the data from 2013 to 2018 for Bitstamp available through their API. This data includes opening price, closing price, low, high, volume per record.

5.2 Weighted Price

Stock technical analysis is in a way very similar to feature engineering. Technical indicators and technical analysis are employed by stock market professionals and traders to forecast the future price. Each trader has his or her own technical indicator with each indicator performing differently in a different environment. Instead of using solely the closing price, it is also better to know the volume weighted price. A price shift because of a movement of a significant amount of volume tends to impact the market more than a price shift because of a lesser volume movement. Therefore a Volume Weighted price is an effective technical indicator..

5.3 Data Preparation and Price Prediction

Instead of transforming this into a precise price forecast, It has transformed it into a classification problem because that is what the traders are looking for. As the price of Bitcoin has also shifted from a single digit to now five digits, it is also challenging for the neural network to learn this pattern if the price is not standardized or normalized. We defined classes to categorize the future value: Within 0.4% of the closing price. Between 0.4% to 0.8% above the closing price. Between 0.4% to 0.8% lower the closing price. More than 0.8% higher than the closing price.

5.4 Simulation Result

The outputs were never able to bring the accuracy to a single point, the higher the neural network learned the better the output achieved. It can be executed to obtain the following range of classification accuracy: Train dataset accuracy 75%-95% Test dataset accuracy 70%-90% I think that for trading, these accuracies are excellent predictions. Considering the volatility of Bitcoin, it is reasonable that the prediction accuracy varies with time. Machine (SVM). Each model is trained on the training set of the dataset and tested on the held-out test data. The system employs major evaluation criteria such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) to evaluate the precision and effectiveness of every algorithm. Of these, Random Forest tends to perform most optimally because it can manage non-linear relationships and prevent overfitting. The module of evaluation ensures that the model deployed is most accurate and dependable among all the tested models.

VI. SYSTEM ARCHITECTURE

A software architecture diagram graphically depicts a system's overall outline, indicating components, their interactions, dependencies, data flows, and sometimes deployment. It assists developers and stakeholders in comprehending the system's design, facilitating collaboration and decision-making during the software lifecycle.

The bitcoin prediction architecture diagram is as shown in Fig 1.

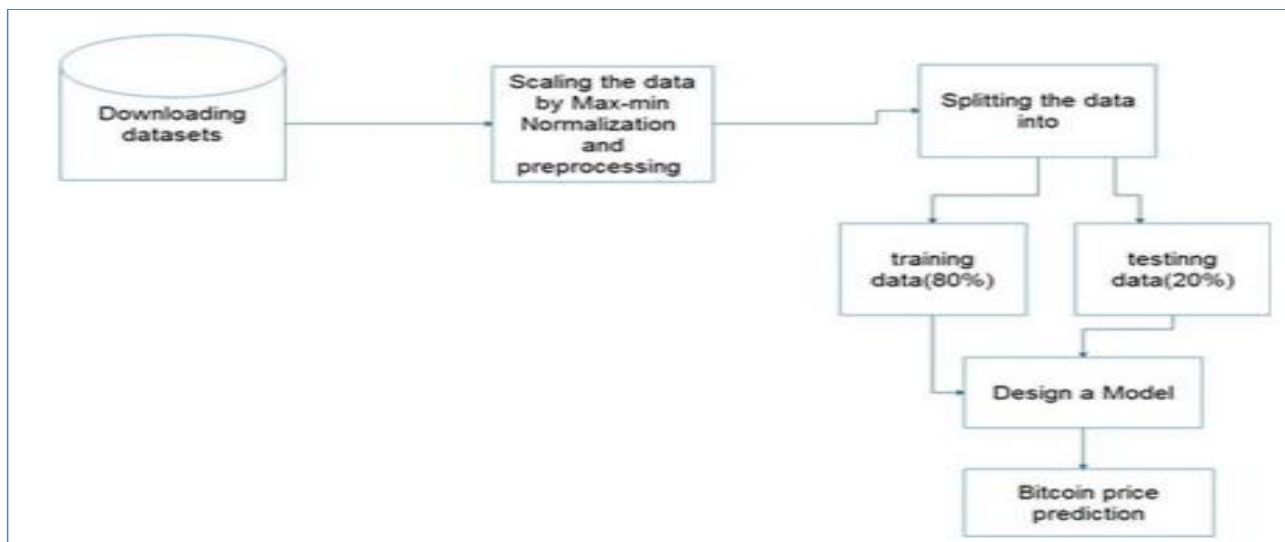


FIG1.ARCHITECTURE OF THE SYSTEM

The System shows the process flow of house bitcoin prediction with machine learning models. It begins with the data acquisition, which is followed by data preprocessing to deal with missing values, categorical data encoding, and feature normalization. The pre-processed data is divided into a training set and a testing set. The training set is utilized to build Random Forest and Gradient Boosting models, which are robust ensemble methods. These models are ultimately combined in the final ensemble model to increase precision and reliability. The trained model is ultimately used for predicting home prices, with the test set being utilized in order to monitor model performance. A software architecture diagram is a visual representation of the overall structure and interaction of a software system. It gives a

top-level view of the parts of the system, interactions, and how they work together to provide the system's functionality. These diagrams assist in showing how different parts of software components interact with each other. They are components of the system modules, layers, or services, and specify how components communicate with one another through methods, protocols, or APIs. The diagram also indicates the dependency between components, i.e., which ones depend on others for functionality, and where data moves from one component to another or from one subsystem to another and how data is processed. In certain situations, it might also illustrate the logical or physical arrangement of components in hardware or software environments. Software architecture diagrams provide a roadmap architects.

VII. EXPERIMENTAL RESULTS

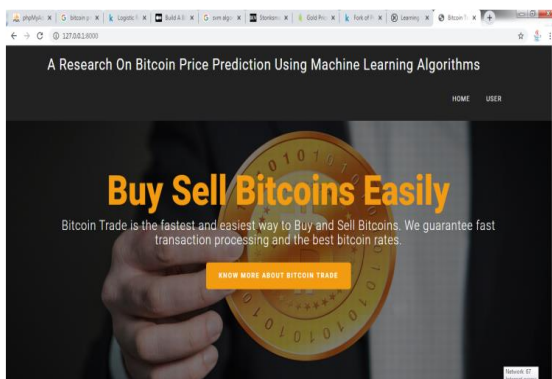


FIG2.HOMEPAGE

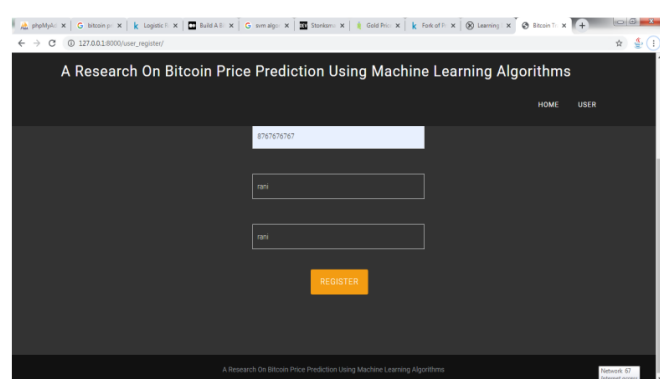


FIG3.USERREGISTRATION

In Fig2 Show Bitcoin Prediction System login page offers safe access to users of the system. It makes sure that only registered users can enter the system and read or interact with the prediction information. The login form contains fields for the password and username, with minimal validation.

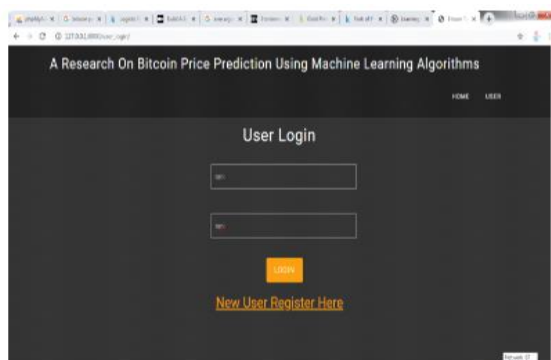
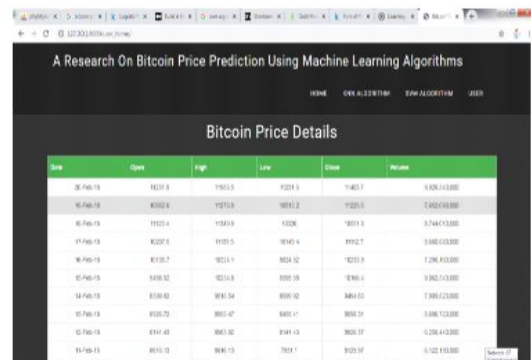


FIG4.USERLOGIN



Date	Open	High	Low	Close	Volume
30-Mar-19	8237.9	7953.5	8229.9	7483.7	9.528.143.000
01-Apr-19	8302.9	7975.9	8051.2	7720.5	7.080.030.000
02-Apr-19	7702.4	7580.9	7328	7071.6	5.744173.000
03-Apr-19	8207.5	7701.5	8140.4	7702.7	9.980.030.000
04-Apr-19	8138.7	8251.1	8034.32	7823.9	7.296.830.000
05-Apr-19	8436.32	8253.9	8395.39	7890.4	9.980.030.000
06-Apr-19	8139.82	8191.34	8091.92	8404.03	7.980.030.000
07-Apr-19	8108.72	8051.47	8405.11	8090.39	5.096.720.000
08-Apr-19	8341.48	8061.92	8341.13	8638.37	5.276.830.000
09-Apr-19	8610.18	8696.19	7929.7	8329.97	5.122.830.000

FIG5.BITCOINPRICEDetails

In Fig 4 shows the user login page enables registered users to access the Bitcoin prediction system securely. It has fields for inputting a username and password. Once logged in successfully, users can see Bitcoin price forecasts, monitor trends, and view personalized information. This page ensures user information is secure and only authorized users can access the system.

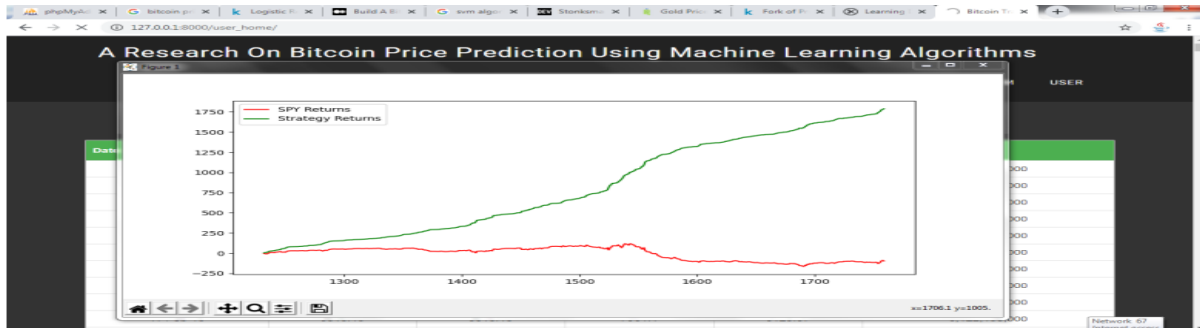


FIG 6. PERFORMANCE ANALYSIS

In Fig 6 Shows the output the analysis of the Bitcoin Prediction System's performance is directed towards assessing the model's accuracy and efficiency in making predictions.

VIII. CONCLUSION

In summary, the application of machine learning algorithmic approaches such as Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) for predicting the price and data-reliant system for analyzing and predicting cryptocurrency activity. Due to the extremely high volatility and constant changes in the Bitcoin market, conventional forecasting procedures tend to remain MSE Yet, with the incorporation of machine learning models, the system proposed here can learn intricate patterns in past data, recognize market trends, and make more precise predictions. The use of a hybrid model that integrates several algorithms, coupled with stringent testing approaches—unit testing, functional testing, and integration testing—guarantees system reliability, precision, and user satisfaction

REFERENCES

- [1] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015) suggested stock and index movement prediction through a method of using Trend Deterministic Data Preparation alongside machine learning techniques. Their process improved the input data quality prior to the usage of ML models, resulting in improved prediction accuracy.
- [2] McNally, S., Roche, J., & Caton, S. (2018) employed Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models to forecast Bitcoin prices. These models are efficient in capturing temporal and sequential dependencies in financial time series data.
- [3] Lahmiri, S., & Bekiros, S. (2019) presented deep learning chaotic neural networks in predicting cryptocurrency prices. Their system utilized chaotic characteristics in market behavior to enhance forecast accuracy in strongly volatile markets.
- [4]. Mudassir, M., Bennbaia, S., Unal, D., & Hammoudeh, M. (2020) applied high-dimensional input features using Support Vector Regression and Random Forest models in the forecast of Bitcoin price. Improved accuracy through dimension reduction and feature engineering was obtained with their technique.
- [5]. Abraham, J., Higdon, D., & Nelson, J. (2018) emphasized sentiment analysis through the application of Tweet volumes and emotions derived from Twitter data to forecast cryptocurrency prices. They illustrated how public sentiment has a considerable impact on price movements.
- [6]. Gopichand Vemulapalli, Sreedhar Yalamati, Naga Ramesh Palakurti, Naved Alam, Srinivas Samayamantri, Pawan Whig, "Predicting Obesity Trends Using Machine Learning from Big Data Analytics Approach," pp. 1-5, 2024.
- [7]. Alessandretti, L., ElBahrawy, A., Aiello, L. M., & Baronchelli, A. (2018) compared various machine learning models such as SVM, Gradient Boosting, and Random Forest for predicting different cryptocurrency prices. They concluded that there is no model that performs optimally for all currencies and currency-specific models are required.
- [8]. Kim, Y., & Kim, H. Y. (2019) utilized LSTM-based models to forecast cryptocurrency prices based on historical data like price and trading volume. Their deep model accurately captured time-series trends and demonstrated better performance than conventional models.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details