Statistical and Computational Methods for Bitcoin Price Forecasting

Sowmya B[1], Sneha T J[2], Geetha M M[3]
[1,2] *Lecturer, Department of Science, Government Polytechnic, Nagamangala, Karnataka, India*
[3]*Lecturer, Department of Science, Government Polytechnic, Mirle, K R Nagara, Karnataka, India*

Abstract*—Cryptocurrency, particularly Bitcoin, has gained immense popularity among investors due to its decentralized nature and blockchain-based technology. This research aims to predict Bitcoin prices by analyzing various factors that influence its value. The investigation is divided into two phases. The first phase focuses on identifying daily market trends and extracting optimal features related to Bitcoin's price movements. The dataset includes daily records of Bitcoin's payment network metrics, trading volumes, and external factors like stock market indices and sentiment analysis. Understanding these trends will provide a comprehensive insight into the factors driving Bitcoin's value. In the second phase, machine learning algorithms, particularly Long Short-Term Memory (LSTM) networks, are employed to predict the direction of daily price changes with high accuracy. LSTM's capability to analyse time-series data makes it well-suited for capturing the temporal dependencies in cryptocurrency markets. The inclusion of diverse data sources such as blockchain analytics, market sentiment, and coin tracking platforms enhances the predictive model's robustness. Bitcoin's value is unique as it does not depend on traditional economic events or government policies, making it highly volatile and unpredictable. Machine learning methods simplify this complexity by modeling nonlinear relationships within the data. By integrating advanced computational techniques and statistical analysis, this study offers a framework for reliable price forecasting, addressing the challenges faced by investors in navigating the dynamic cryptocurrency market. This research not only contributes to the financial literature but also supports decision-making for both novice and seasoned investors.

***Index Terms*—Patrolling, IOT, Raspberry pi module, Noob software**

I. INTRODUCTION

Bitcoin has emerged as a revolutionary force in the global financial system, representing the forefront of blockchain technology. It is a decentralized cryptocurrency that operates without any centralized authority, enabling seamless digital payments and serving as an investment vehicle. Transactions made through Bitcoin bypass traditional banking systems, offering advantages such as faster processing, lower transaction fees, and enhanced privacy. This independence from traditional financial institutions is made possible through blockchain, a distributed ledger that maintains a secure, transparent, and immutable record of transactions. Each "block" in the blockchain contains data encrypted to protect user identities, ensuring that only wallet IDs are visible during transactions.

Investors access Bitcoin through specialized marketplaces called Bitcoin exchanges, such as Mt. Gox, where they can buy or sell using various fiat currencies. These exchanges facilitate Bitcoin's integration into global financial ecosystems, contributing to its widespread adoption. Despite its innovative features, Bitcoin is characterized by high volatility, making it a focal point for financial analysts, traders, and machine learning researchers. Fluctuations in Bitcoin's price influence trade strategies, systemic risk evaluations, and investment decisions. Understanding and predicting this volatility is thus crucial for stakeholders across the cryptocurrency and data science domains.

The primary goal of Bitcoin price prediction research is to determine the accuracy of various machine learning algorithms in forecasting its price trends. This involves using historical price data, trading volumes, and other market indicators to develop robust predictive models. The study's methodology begins by transforming order book data into a structured "feature series" format, enabling machine learning models to simultaneously analyse price volatility and key features. These models aim to extract meaningful insights from complex and non-linear relationships inherent in Bitcoin market data, addressing one of the most significant challenges in financial analytics.

Machine learning algorithms such as Random Forests, Support Vector Machines (SVMs), and neural networks, particularly Long Short-Term Memory (LSTM) models, have shown promise in this domain.

LSTMs are especially suited for analyzing time-series data, as they can capture long-term dependencies and trends. By incorporating volatility as a central feature, these models align closely with investment strategies and risk assessment frameworks. The study also highlights the importance of feature selection to optimize model performance, leveraging statistical and computational techniques to refine the input data for predictive accuracy.

Bitcoin’s role as the pioneer of blockchain technology underscores its significance beyond being just a cryptocurrency. Its introduction has catalysed a broader financial revolution, giving rise to numerous alternative cryptocurrencies and reshaping the global financial landscape. As a result, Bitcoin accounts for a dominant share of the cryptocurrency market capitalization. For researchers and practitioners in machine learning and data mining, Bitcoin price prediction offers a compelling problem that combines the challenges of financial forecasting with the technical intricacies of blockchain technology.

The proposed study addresses these challenges by developing predictive models that not only aim for accuracy but also contribute to a deeper understanding of market dynamics. By utilizing diverse data sources, including stock market indices, market sentiment analysis, and blockchain metrics, the research provides a comprehensive framework for analyzing Bitcoin’s price behavior. The insights gained from this work are expected to benefit investors, financial analysts, and policymakers, enabling them to navigate the complexities of the cryptocurrency market with greater confidence.

II. RELATED WORK

Ciaian, Rajcaniova, and Kancs (2016) examine Bitcoin price formation by integrating traditional economic factors like supply and demand with cryptocurrency-specific influences, such as Bitcoin's appeal to investors and users. Their analysis of five years of daily data (2009–2015) using time-series methods reveals that market forces and Bitcoin's attractiveness significantly impact its price. However, their findings challenge previous research by showing that macroeconomic factors do not drive Bitcoin’s price in the long term. This study provides valuable insights into the unique dynamics of cryptocurrency markets, emphasizing internal factors over broader financial trends in determining Bitcoin's value.

In the paper "Predicting the Price of Bitcoin Using Machine Learning" by McNally (2016), the author aims to predict the direction of Bitcoin price fluctuations in USD using machine learning techniques. The data is sourced from the Bitcoin Price Index, and the task is tackled using Bayesian-optimized recurrent neural networks (RNN) and Long Short-Term Memory (LSTM) networks. Among these, the LSTM model performs the best, achieving 52% classification accuracy and a Root Mean Square Error (RMSE) of 8%. In comparison, the ARIMA model, a traditional time-series forecasting method, performs poorly. The study also benchmarks the deep learning models on both GPU and CPU systems, with the GPU implementation outperforming the CPU by 67.7% in training time. This research highlights the superior performance of non-linear deep learning methods over traditional models like ARIMA in predicting Bitcoin price trends.

In the paper "Automated Bitcoin Trading via Machine Learning Algorithms" by Madan, Saluja, and Zhao (2015), the authors explore the use of machine learning techniques to predict Bitcoin price movements. The project is divided into two phases. In the first phase, the goal was to understand daily trends in the Bitcoin market and identify key features influencing Bitcoin price. The dataset, spanning five years and consisting of over 25 features, included data related to Bitcoin's price and payment network. Using this data, the authors achieved an impressive 98.7% accuracy in predicting the sign of daily price changes.

In the second phase, the focus shifted to Bitcoin price data alone, with predictions made at higher granularities (10-minute and 10-second intervals). This allowed the authors to evaluate predictions at varying levels of time-series noise and granularity. The price prediction problem was framed as a binomial classification task, where the authors utilized a custom algorithm combining random forests and generalized linear models. At these shorter intervals, the prediction accuracy dropped to 50-55%, reflecting the increased difficulty and noise in short-term forecasting.

This study demonstrates the potential of machine learning in Bitcoin price prediction, with different results based on data granularity and the complexity of time-series modeling.

In the study "Volatility Estimation for Bitcoin: A Comparison of GARCH Models" by Katsiampa (2017), the author investigates the optimal conditional heteroskedasticity model for Bitcoin price data. The analysis focuses on identifying the best-fitting

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. Among the models tested, the AR-CGARCH model emerges as the most effective, as it incorporates both short-run and long-run components of conditional variance. This dual-component approach captures the complex volatility structure of Bitcoin prices more accurately than simpler GARCH models. The findings underscore the importance of accounting for different volatility horizons in analyzing Bitcoin's highly dynamic market behavior.

In the paper "Hedging Capabilities of Bitcoin: Is It the Virtual Gold?" by Dyhrberg (2016), the author examines Bitcoin's potential as a hedging tool using the asymmetric GARCH methodology commonly applied to analyze gold. The findings reveal that Bitcoin shares some hedging properties with gold. Specifically, Bitcoin effectively hedges against stock market risks, as demonstrated in its relationship with the Financial Times Stock Exchange Index, and provides short-term hedging capabilities against the US dollar. These results suggest that Bitcoin can be considered a complementary hedging instrument, expanding the toolkit available to market analysts for managing market-specific risks.

In the paper "A Comparative Study of Bitcoin Price Prediction Using Deep Learning" (2019), the authors compared state-of-the-art deep learning methods, including LSTM, DNN, CNN, and deep residual networks, for Bitcoin price prediction. They found LSTM-based models performed slightly better in regression tasks, while DNN-based models excelled in classification tasks for predicting price movements. Classification models were also shown to be more effective for algorithmic trading than regression models. This study highlights the practical utility of combining deep learning methods for varying prediction objectives in cryptocurrency trading.

In "Forecasting the Movements of Bitcoin Prices: An Application of Machine Learning Algorithms" (2019), four machine learning algorithms—Support Vector Machines (SVM), Artificial Neural Networks (ANN), Naive Bayes (NB), and Random Forest (RF)—were evaluated alongside logistic regression as a benchmark. Results showed RF had the highest forecasting accuracy on continuous datasets, while ANN excelled with discrete datasets. The study also highlighted the improved performance achieved by discrete datasets in all models tested, demonstrating the potential of tailored machine learning approaches for cryptocurrency price forecasting

Dimitriadou and Gregoriou (2020) studied Bitcoin price predictions using machine learning with 24 explanatory variables, including cryptocurrency data, exchange rates, and macroeconomic factors. They found that logistic regression outperformed support vector machines and random forests, achieving 66% accuracy. This research provided insights into inefficiencies in the Bitcoin market and demonstrated the potential of combining various financial variables for predictive modeling

Chen et al. (2020) employed ensemble learning models for Bitcoin price prediction, combining random forests and gradient boosting techniques. Their study revealed improved accuracy compared to standalone models, emphasizing ensemble approaches' adaptability in handling Bitcoin's volatility.

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are widely used to analyze Bitcoin's price volatility over time. Studies examining Bitcoin's statistical and economic properties often highlight its potential as a financial asset, exploring aspects such as inefficiencies under the efficient market hypothesis, hedging capabilities, and speculative bubbles. Additionally, researchers investigate the correlation between Bitcoin prices and search trends on platforms like Google and Wikipedia. Advanced techniques, including wavelet analysis, are also applied to reveal temporal and frequency domain dynamics, further enriching the understanding of Bitcoin's price behavior.

Bitcoin price prediction done by P. Ciaian, M. Rajcaniova, and D. Kancs evaluates Bitcoin price formation based on a linear model by considering related information that is categorized into several factors of market forces, attractiveness for investors, and global macro-financial factors. They assume that the first and second factors mentioned above significantly influence Bitcoin prices but with variation over time.

S. McNally et.al predicts the Bitcoin pricing process using machine learning techniques, such as recurrent neural networks (RNNs) and long short-term memory (LSTM), and compare results with those obtained using autoregressive integrated moving average (ARIMA) models. A machine trained only with Bitcoin price index and transformed prices exhibits poor predictive performance.

Existing techniques compares the accuracy of predicting Bitcoin price through binomial logistic regression, support vector machine, and random forest.

A. System Analysis

Computer Aided Diagnosis is a rapidly growing dynamic area of research in financial industry. The recent researchers in machine learning machine learning promise the improved accuracy of price predictions. Here the computers are enabled to think by developing intelligence by learning. There are many types of Machine Learning Techniques and which are used to classify the data sets.

B. Requirement Analysis

Software Requirement Specification (SRS) is the starting point of the software developing activity. As system grew more complex it became evident that the goal of the entire system cannot be easily comprehended. Hence the need for the requirement phase arose. The software project is initiated by the client needs. The SRS is the means of translating the ideas of the minds of clients (the input) into a formal document (the output of the requirement phase.)

Under requirement specification, the focus is on specifying what has been found giving analysis such as representation, specification languages and tools, and checking the specifications are addressed during this activity.

The Requirement phase terminates with the production of the validate SRS document. Producing the SRS document is the basic goal of this phase.

The purpose of the Software Requirement Specification is to reduce the communication gap between the clients and the developers. Software Requirement Specification is the medium though which the client and user needs are accurately specified. It forms the basis of software development. A good SRS should satisfy all the parties involved in the system.

C. Hardware Interfaces

Ethernet on the AS/400 supports TCP/IP, Advanced Peer-to-Peer Networking (APPN) and advanced program-to-program communications (APPC).

To connect AS/400 to an Integrated Services Digital Network (ISDN) for faster, more accurate data transmission. An ISDN is a public or private digital communications network that can support data, fax, image, and other services over the same physical interface. can use other protocols on ISDN, such as IDLC and X.25.

D. Operational Requirements

- The developed product is economic as it is not required any hardware interface etc.
- Statements of fact and assumptions that define the expectations of the system in terms of mission objectives, environment, constraints, and measures of effectiveness and suitability (MOE/MOS). The customers are those that perform the eight primary functions of systems engineering, with special emphasis on the operator as the key customer.
- The software may be safety-critical. If so, there are issues associated with its integrity level. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions. If a system must be of a high integrity level and if the software is shown to be of that integrity level, then the hardware must be at least of the same integrity level. There is little point in producing 'perfect' code in some language if hardware and system software (in widest sense) are not reliable. If a computer system is to run software of a high integrity level then that system should not at the same time accommodate software of a lower integrity level. Systems with different requirements for safety levels must be separated. Otherwise, the highest level of integrity required must be applied to all systems in the same environment.

E. System Requirements

H/W Requirements
Processor : Any Processor above 500 MHz.
Ram : 4 GB
Hard Disk : 4 GB
Input device : Standard Keyboard and Mouse.
Output device : VGA and High-Resolution Monitor.

S/W Requirements
Operating System: Windows 7 or higher
Programming : Python 3.6 and related libraries

III.PROPOSED SYSTEM METHODOLOGY

We have collected the dataset for the document with following details from quandl.com and we applied machine learning algorithm such as decision tree and regression for prediction and price forecast.

IV. DATASET DETAILS

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.

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.

A	B	C	D	E	F	G	H
Date	Open	High	Low	Close	Volume (BTC)	Volume (Currency)	Weighted Price
9/13/2011	5.8	6	5.65	5.97	58.37138238	346.0973894	5.929230648
9/14/2011	5.58	5.72	5.52	5.53	61.14598362	341.8548132	5.590797514
9/15/2011	5.12	5.24	5	5.13	80.1407952	408.2590022	5.094271914
9/16/2011	4.82	4.87	4.8	4.85	39.9140068	193.7631466	4.854515047
9/17/2011	4.87	4.87	4.87	4.87	0.3	1.461	4.87
9/18/2011	4.87	4.92	4.81	4.92	119.8128	579.8431027	4.839575594
9/19/2011	4.9	4.9	4.9	4.9	20	98	4.9
9/20/2011	4.92	5.66	4.92	5.66	89.28071068	481.0492629	5.388053693
9/21/2011	5.7	5.79	5.66	5.66	17.62932238	100.5942336	5.706074879

Figure 1: Dataset used for study

Table 1: Dataset used for study

Variable name	Short description
Date	Trading Date
Open	Bitcoin Open price for particular time
High	Bitcoin High price achieved for particular time
Low	Bitcoin Low price achieved for particular time
Close	Bitcoin Close price for particular time
Volume (BTC)	Coin volume traded
Volume (Currency)	Coin value traded
Weighted price	Price per coin traded

The dataset collected from 2011 to till data is plotted on the below chart, Figure 2, which clearly depicts that bitcoin is a positive market.

A. Data Visualization Of Botcoin From 2011 To Till Data

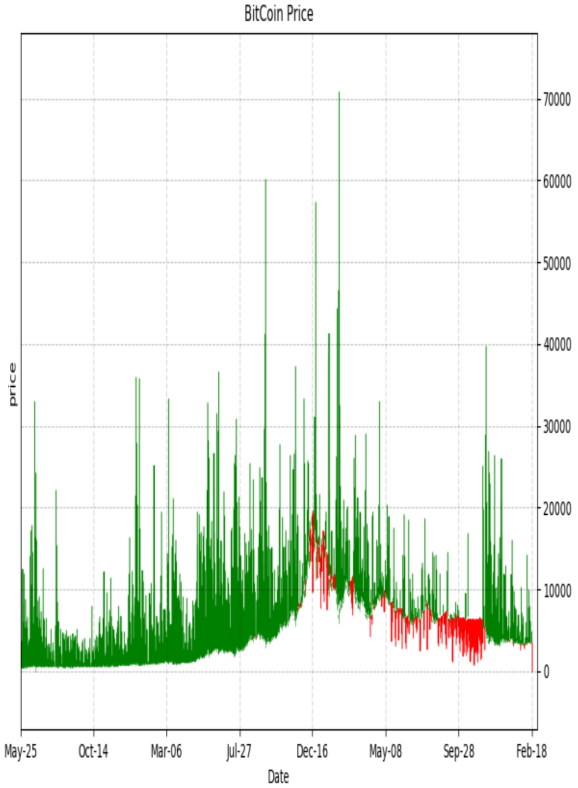


Figure 2: Data Visualization of Bitcoin price

B. MODULES DESCRIPTION

i. PRICE PREDICTION:

Price prediction on the considered dataset is done using two different machine learning algorithms such as Decision Tree and linear regression. The predicted value is compared for the predicted accuracy and error values.

ii. Decision Tree

Decision tree is a type of supervised learning algorithm that is mostly used in classification problems. It works for both categorical and continuous input and output variables. In this technique, we split sample into two or more homogeneous sets (or sub-populations) based on most significant splitter / differentiator in input variables. In decision tree internal node represents a test on the attribute, branch depicts the outcome and leaf represents decision made after computing attribute.

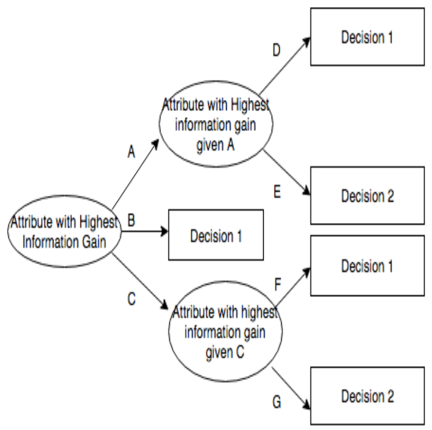


Figure 3: Flow chart Decision Tree algorithm

Decision Tree works in following manner

- Place the best attribute of the dataset at the root of the tree.
- Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute.
- Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

In decision trees, for predicting a class label for a record we start from the root of the tree. Then compare the values of the root attribute with record's attribute. On the basis of comparison, follow the branch corresponding to that value and jump to the next node. Decision Tree Classifier is a class capable of performing multi-class classification on a dataset.As with other classifiers, Decision Tree Classifier takes as input two arrays: an array X, sparse or dense, of size [n_samples, n_features] holding the training samples, and an array Y of integer values, size [n_samples], holding the class labels for the training samples.

iii. Linear regression

In simple, linear regression, predict scores on one variable from the scores on a second variable. The variable that predicted is called the criterion variable and is referred to as Y. The variable base for predictions on is called the predictor variable and is referred to as X. When there is only one predictor variable, the prediction method is called simple regression. In simple linear regression, the topic of this section, the predictions of Y when plotted as a function of X form a straight line.

i. Algorithm

Decision Tree Algorithm Pseudocode

Step 1: Compute the entropy for data-set

Step 2: For every attribute/feature:

a.calculate entropy for all categorical values

b.take average information entropy for the current attribute

c.calculate gain for the current attribute

Step 3: Pick the highest gain attribute.

Step 4: Repeat until we get the tree we desired

In decision trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

ii. Flow chart of Logistic Regression algorithm:

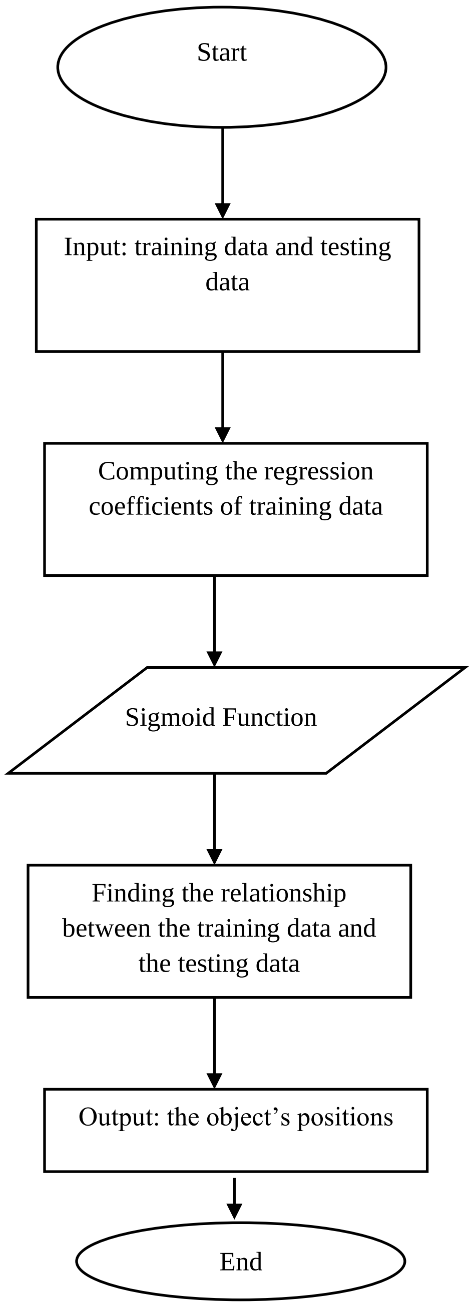


Figure 4: Flow chart Decision Tree algorithm

V. RESULTS AND DISCUSSION AND PERFORMANCE ANALYSIS

The proposed work is implemented in Python 3.6.4 with libraries scikit-learn, pandas, matplotlib and other mandatory libraries. 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. Machine learning algorithm is applied such as decision tree and regression. Five days forecast price prediction is done using decision tree and regression. The values are compared.

The result shows that bitcoin price prediction is efficient using liner regression algorithm. Linear regression achieves around 97.5% accuracy in price prediction, whereas decision tree achieves 95.8% accuracy.

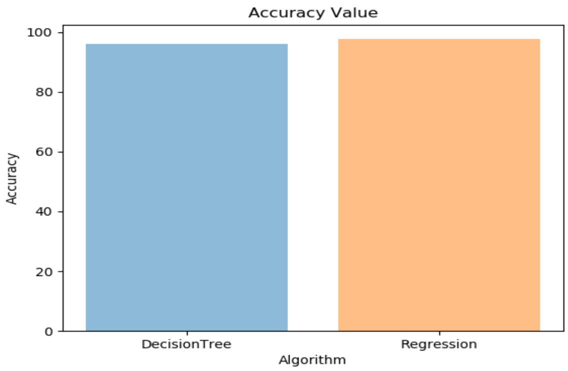
The following table shows the accuracy arrived in our experimental study.

Table 2: Results for Decision tree and Regression

Algorithm	Accuracy
Decision Tree	95.88013
Regression	97.59812

The below figure shows the accuracy comparison of our proposed work, in which regression method outperforms decision tree.

Price forecast is done for 5 days using machine learning techniques such as Decision tree and regression. The result is compared with the score value to identify the accuracy value and plotted. Predicted price forecast using our proposed method is shown in the below figure



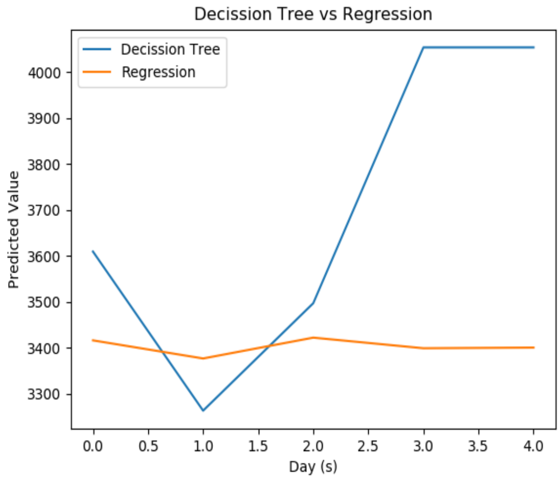


Figure: Price Forecast for five days

V. CONCLUSION

Bitcoin is a successful cryptocurrency, and it has been extensively studied in fields of economics and computer science. In this study, we analyse the time series of Bitcoin price with a Decision Tree and Linear regression models. Also the price forecast for five days is done using Lasso and Linear regression models. After establishing the learning framework and completing the normalization, we intend to use the two methods mentioned above and choose the best method to solve the crypto currency prediction problem. The experimental results show that linear regression outperforms the other by high accuracy on price prediction.

ACKNOWLEDGMENT

We extend our heartfelt gratitude to our mentors and colleagues for their invaluable support and guidance throughout this work. We also appreciate the resources provided by the research community that facilitated our work. Finally, we thank our families and friends for their unwavering encouragement during this journey.

REFERENCES

[1] Shah and K. Zhang, “Bayesian regression and Bitcoin,” in 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2015, pp. 409-415.

[2] Huisu Jang and Jaewook 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, vol. 99, pp. 1-1.

[3] Andrade de Oliveira, L. Enrique ZÃ¡rate and M. de Azevedo Reis; C. Neri 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, pp. 2151-2155.

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

[5] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, 2008.

[6] G.S. Atsalakis, I.G. Atsalaki, F. Pasiouras and C. Zopounidis, "Bitcoin price forecasting with neuro-fuzzy techniques", European Journal of Operational Research, vol. 276, no. 2, pp. 770-780, 2019.

[7] S. Roy, S. Nanjiba and A. Chakrabarty, "Bitcoin Price Forecasting Using Time Series Analysis", 2018 21st International Conference of Computer and Information Technology (ICCIT), pp. 1-5, 2018.

[8] D.U. Sutiksno, A.S. Ahmar, N. Kurniasih, E. Susanto and A. Leiwakabessy, "Forecasting Historical Data of Bitcoin using ARIMA and a-Sutte Indicator", Journal of Physics: Conference Series, vol. 1028, no. 1, pp. 012194, 2018.

[9] N. Hitam and A. Ismail, "Comparative performance of machine learning algorithms for cryptocurrency forecasting", Ind. J. Electr. Eng. Comput. Sci, vol. 11, no. 3, pp. 1121-1128, 2018.