

DATA ANALYTICS SYSTEM FOR DIGITAL CURRENCY PRICE PREDICTION USING REGRESSION ALGORITHM

Parthiban.V¹, Dhiwakar N. R², Venkatesh.KSB³, Mr R. Rajesh Kanna⁴

^{1,2,3}B. E, Department of Computer Science and Engineering, Agni College of Technology, India

⁴Assistant Professor, Dept. of Computer Science and Engineering, Agni College of Technology, India

ABSTRACT

Similar to coins, euros, and paper money, cryptocurrency is a contemporary type of money. But rather of having the support of a reputable bank or government, it functions through internet transactions protected by robust encryption. A web-based tool helps with the intricate calculations used to establish the value of cryptocurrency. This platform makes use of the LSTM (Long Short-Term Memory) algorithm, a specialised algorithm, to make highly accurate predictions about the value of cryptocurrencies. It can be difficult to quantitatively forecast the value of Bitcoin. However, the suggested framework, which employs the LSTM algorithm, shows encouraging outcomes in precisely predicting cryptocurrency prices.

Key Words: Bayesian regression, Bitcoin, Bitcoin prediction, Blockchain, cryptocurrency, generalized linear model (GLM), machine learning.

1. INTRODUCTION

In-depth information on cryptocurrency prices as well as pertinent variables like trade volume and market sentiment are gathered and analysed to produce price predictions using machine learning. The machine learning model is then trained using this data, allowing it to predict future prices based on patterns and trends found in the acquired data. It should be emphasised, nonetheless, that a number of variables, such as unforeseen occurrences and world market conditions, can affect a cryptocurrency's value. Therefore, it's possible that machine learning algorithms won't be able to accurately predict future prices in a straightforward manner.

1.1 BITCOIN

Bitcoin is a well-known cryptocurrency that can be used as payment as well as for speculative purposes. Bitcoin is not held by anyone or any organisation because it is decentralised. The simplicity of transactions, which are not limited to any one nation, is one of its distinguishing aspects. Users can buy and sell bitcoin using specified currency denominations on a useful site called "Bitcoin Trading." Mt. Gox was historically the biggest Bitcoin exchange. Like a virtual bank account, a digital wallet is often where bitcoin is kept. Through a process called mining, transactions are logged and timestamped, and the data is saved in blocks referred to as blockchains. A chain of information is formed by the fact that each block in the blockchain has a reference to the block before it. During transactions, the ownership information is concealed, but the wallet address connected to the transaction is displayed.

1.2 Prediction

The perception of Bitcoin's value is gradually evolving, leading to the exploration of various methods for determining its worth. Unlike traditional stocks, Bitcoin's value is not influenced by trading events or intermediaries, making it crucial to forecast its value accurately. Multiple calculations are employed, taking into account factors such as inventory demand, to estimate the value of Bitcoin. Understanding the fluctuations in Bitcoin's value is essential in predicting potential gains or losses.

2. LITERATURE REVIEW

[1] This paper investigates the use of three different machine learning algorithms to forecast the values of BTC, ETH, and LTC. The effectiveness of the prediction models was evaluated by an executive analysis, and a comparison between the actual and anticipated expenses was established. The results show that the GRU algorithm performs better than other algorithms, with MAPEs for ETH, BTC, and LTC, respectively, of 0.8267, 0.2454, and 0.2116. The GRU model's RMSE is 26.59174.129, which indicates that ETH, BTC, and LTC performed satisfactorily. According to these results, GRU appears to be a reliable and practical method for estimating the value of centralised cryptocurrencies. Bi-LSTMs, on the other hand, perform inconsistently when forecasting actual and anticipated fees for BTC and ETH because to their reduced sensitivity when compared to GRUs and LSTMs. This calls into doubt the suitability of Counterfeit Insights algorithms for precise and verified bitcoin forecasts. All algorithms demonstrate promising predicting skills, while the GRU algorithm beats the LSTM and bi-LSTM models.

[2] The main goal is to produce a line that closely resembles the observed data, demonstrating a thorough and aesthetically attractive fit, even with small variations. The term "residual" is used to describe any departure from the regression line.

Understanding LSTMs and RNNs is crucial for producing predictive insights since they can be used to forecast future values and discover patterns and trends in the underlying data. A sizable number of datasets that have been carefully vetted are needed for data preparation. When the data is steady, methods like loss functions, activation functions, and optimizers are used to improve the training process. This approach involves training the data iteratively, and the predicted values and actual values are then compared using statistical significance tests to validate the model.

[3] While machine learning has been widely utilised to increase the accuracy of Bitcoin price prediction, this study paper notes that only a small number of studies have concentrated on using a variety of data structures and crucial aspects. By categorising Bitcoin value based on different factors including daily price and high-frequency pricing, the authors of the research sought to anticipate Bitcoin prices using ML techniques. For the purpose of predicting the daily price of bitcoin, a variety of feature dimensions were taken into account, including network and transaction characteristics, market demand and trade indications, gold spot price, and sentiment. Additionally, short-term intraday price prediction within 5-minute intervals was performed using initial trading features collected from a cryptocurrency exchange.

[4] This study investigates the use of various methods, including Random Forest (RF) and Artificial Neural Networks (ANN), to forecast the closing stock values of five businesses from various industries the next day. In order to provide more input variables for the prediction models, the study includes financial data that is specific to stocks. Standard assessment metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used to assess the performance of the models. Lower values of these indicators show that the models are more accurate at predicting closing stock prices.

[5] Deep neural networks have been used to predict bitcoin prices using RNNs (Recurrent Neural Networks), MLPs (Multilayer Perceptrons), and LSTMs. Particularly LSTM has proven to be successful at correctly predicting the price movement of cryptocurrencies.

3. PROPOSED METHODOLOGY

The offered example demonstrates the use of machine learning methods to develop a system for predicting the value of Bitcoin using publicly available datasets from online databases. To evaluate Bitcoin, the suggested system specifically uses LSTM, a kind of RNN. The agent for the system is the Python module “boa constrictor”. Detailed descriptions of the techniques used in this system are provided. The first stage entails gathering information using a Rest-API to get the highest Bitcoin price from a database online. In order to promote targeted research and produce results that are appropriate for the needs of the system, the collected data is subsequently organised into a structured manner that is in line with the established problem statement. For exploration and validation reasons in creating the demonstration, outdated dataset columns are eliminated, and pertinent data is extracted and added to the CSV file. To handle attribute values, pre-processing techniques are used, which typically entail noise reduction and data homogeneity. In order to estimate Bitcoin’s daily value, LSTM (RNN) models are trained in the demonstration using the prepared dataset. By adding different RNN model layers and evaluating their effectiveness, the predictions may be examined.

4. PRELIMINARIES

The purpose of this portion is to prepare the audience for the subsequent segment. It starts off by giving a succloriThrough a quick chart that highlights recent performance or pertinent trends, it first gives a succinct summary of the Bitcoin market. Following that, the focus of the conversation switches to highlighting the significance of the DC-net (Decentralised Network) contract and exploring the difficulties involved in its implementation.

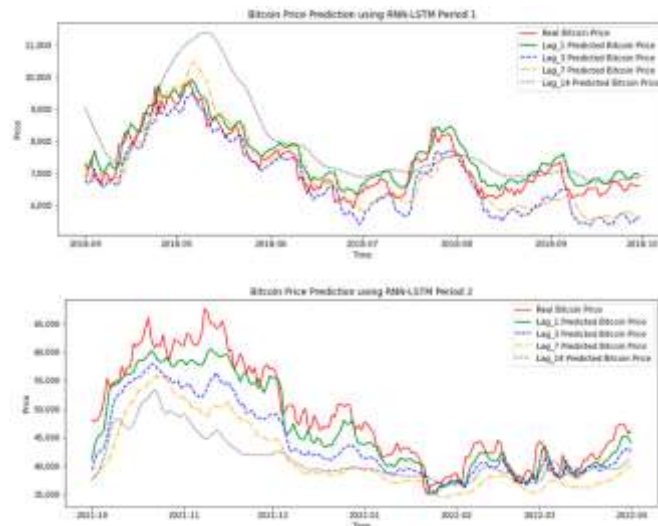
4.1 Overview Of Bitcoin

The purpose of this portion is to prepare the audience for the subsequent segment. It starts off by giving a succloriA well-known and well-recognized digital money and payment mechanism is bitcoin [27]. Direct exchanges of bitcoins between users, often known as “addicts,” take place within the Bitcoin network. The use of intermediaries, such as central banks, is no longer necessary with this peer-to-peer system. Decentralised verification is carried out by miners to guarantee the validity and integrity of transactions. As they verify freshly produced transactions, put them in new blocks, and add these blocks to the blockchain, which acts as a post-settlement record, miners play a critical role in the Bitcoin network. Miners are compensated with bitcoins for their computational labour. Notably, Bitcoin mining operations can use a lot of energy, with some Chinese data centres requiring up to 135 megawatts [31]. A cryptographic hash of the preceding block is included in every block of the Bitcoin blockchain, creating a safe chain of blocks and establishing a sequential relationship. Over time, this blockchain maintains its consistency and immutability. Participants in Bitcoin transactions are identifiable by their distinct Bitcoin addresses, also referred to as aliases. Input addresses and transaction addresses, respectively, are the addresses of the payer and the payee. Based on the user’s public key, these Bitcoin addresses are created using robust cryptographic hash techniques.

5. FLOW OF PAPER

The paper's major objective is to get data from databases. In particular, Quandl Coin and ask Cap databases were used to get our data on bitcoin values. We need to standardise and smooth this time-series data, which is collected five times every day at varying intervals. We have used a variety of normalisation techniques to achieve this, including log transformation, z-score normalisation, box Cox normalisation, and others. The data is then smoothed throughout the entire period after that. The next step is to choose the parameters that will help in sequence prediction. Some of the many features that are accessible are listed below

Comparison of the related works on cryptocurrency price prediction.



Using sophisticated machine learning techniques, such as deep learning, can be essential for solving difficult, nonlinear issues and making accurate predictions from enormous volumes of data. Due to their inherent volatility and wide range of values, cryptocurrencies are notoriously difficult to predict with any degree of accuracy, but deep learning techniques may hold the key. Ji et al. (7) evaluated Long Short-Term Memory (LSTM) and Deep Neural Networks (DNNs) and discovered that LSTM performed better in predicting bitcoin values than other regression models, proving its efficacy. Their strategy centred on using regression analysis to analyse deep learning models, which was beneficial for bitcoin trading. Shintateetal. (8) demonstrated a method for analysing and predicting trends in non-stationary bitcoin time series data using deep learning. According to the findings, LSTM fared exceptionally well when analysing profit possibilities employing the purchase and hold strategy. The results showed that LSTM produced accurate price predictions for cryptocurrencies. In a different study, (9) suggested combining Support Vector Machine (SVM) and Generalised Autoregressive Conditional Heteroskedasticity (GARCH) for predicting bitcoin prices. The purpose of this method was to deal with the heteroskedasticity and volatility of bitcoin data. The proposed model, which combined GARCH and SVM, demonstrated promising outcomes in predicting bitcoin values. These studies demonstrate how several methods, including LSTM, deep learning, GARCH, and SVM, can be used to predict bitcoin prices. Each strategy has its own benefits and solves particular difficulties related to cryptocurrency data. Researchers work to improve the accuracy and efficiency of machine learning by using sophisticated machine learning algorithms.

6. METHODS

Different models have been looked at in the challenge of predicting the directionality of Bitcoin price changes. There has been evaluation of bracket models like Support Vector Machine and Linear Regression. The Autoregressive Integrated Moving Average (ARIMA) and other regression-based models have also been taken into consideration. Recurrent neural network (RNN)-based models have also been put into practise and tested. The effectiveness of each of these models has been carefully evaluated, and the findings have been analysed. To investigate how each model's fundamental assumptions might affect its performance, different models were explored. The theories and ideas that underpin these models have also been succinctly summarised. Researchers wanted to learn more about the variables affecting the performance of the models, so they looked at several models and their associated theories. The numerous methods and the underlying presumptions provide for a thorough knowledge of the various approaches used to forecast Bitcoin price directionality.

A. Logistic Regression

A statistical framework for analysing a dataset of independent factors that affect an outcome is described in the following paragraph. A binary variable is used to measure the result, which has just two possible outcomes (e.g., 1/0, yes/no, true/false). A set of independent variables are used to predict binary outcomes using this paradigm. It is a predictive

regression model with a categorical dependent variable. Maximum Likelihood Estimation is used in the framework to calculate the odds of a given outcome occurring in the regression model.

$$\left(1 - \sum_{k=1}^p \alpha_k L^k\right)(1 - L)^d X_t = \left(1 - \sum_{k=1}^q \beta_k L^k\right) \epsilon_t$$

B. Support Vector Machine

With fewer dataset-related assumptions than logistic regression, the support vector machine approach also generates a binary classification model.

The classifier is obtained by optimizing:

$$\begin{aligned} \min_{\gamma, w, b} \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \dots, m \end{aligned}$$

C. Auto Regressive Integrated Moving Average(Arima)

The Autoregressive Integrated Moving Average (ARIMA) model is a popular one for forecasting and time series analysis. Data from time series that can be converted into a stationary time series are used to test this model. Internal regressions that take into account elements like temporal differences and moving averages provide the foundation for the predictions. The Statsmodels package's version of ARIMA is used in this situation (Seabold and Perktold, 2010). Data is differenced in ARIMA, which translates price values into differences between successive prices. Let's write $\hat{}$ to represent the different data. P, D, and Q are hyperactive parameters that are optimised in the equation below. A model is trained using previous price data for each time point (t) in order to predict the price at that particular time (t), and the prediction is based on the sign of the price change.

D. Recurrent Neural Networks (RNN)

Elman, a researcher, created the recurrent neural network (RNN) in its basic form. Its structure is comparable to that of the Multilayer Perceptron (MLP), but it differs in that it permits iterative forward and backward signal flow. A new subcategory called the Context Layer has been added to allow for both the forward and backward flow. The output of each subcategory is fed into the context subcategory, which is used in the following subcategory with the next input, in addition to passing input between layers. At each time step in this scenario, the state is updated. In contrast to giving the same weights to all inputs, this has the advantage of allowing the network to give different weights to events that occur in a sequence.

BITCOIN PRICE CHANGE PREDICTOR

ACCURACIES

Method	Accuracy
Logistic Regression	45%
SVM	49%
ARIMA	55%
RNN	50%

7. CONCLUSIONS

In our study, we looked at historical Bitcoin sales data and used factors such as price and time to forecast future Bitcoin prices. Logistic regression, Support Vector Machine (SVM), Recurrent Neural Network (RNN), and ARIMA were the four models we used to forecast prices. Table 2 displays the degree of forecast accuracy for these four models. ARIMA fared well for short-term forecasts among the four models, but badly for long-term predictions. It was shown that using the prices from the previous few days can predict prices for the following 5-7 days with accuracy. RNN, however, consistently outperformed other methods for up to 6 days. When a divisible hyperactivity plane is present, the logistic regression-based model held true to its basic assumptions and demonstrated its suitability for direct classification. Overall, our analysis demonstrates how different models perform differently in predicting Bitcoin values, with ARIMA excelling in the near term, RNN being reliable up to a certain timescale, and logistic regression offering classification capabilities under certain circumstances.

8. REFERENCES

- [1] CharlesBovaird.2017. Why the crypto markets the appreciated more than 1,200% this year. Forbes Magazine [Online;posted[17-November-2017]. <https://www.forbes.com/sites/cbovaird/2017/11/17/whythe-crypto-market-has-appreciated-more-than1200-this-year>.
- [2] MarshallVanAlstyne.2014. Why bitcoin has value .Communications of the ACM 57(5):30–32.<https://doi.org/10.1145/2594288>.
- [3] KareemHegazy and Samuel Mum ford.2016.Comparitive automated bitcoin trading strategies.<http://cs229.stanford.edu/proj2016/repot/MumfordHegazyComparitiveAutomatedBitcoinTradingStrategiesreport.pdf>.
- [4] LipoWang and EdwinSin School of Physical and Mathematical Sciences Nanyang Technological University Singapore presented–Bitcoin Price Prediction Using Ensembles of Neural Networks|
- [5] Nehamangla,||UnstructuredDataAnalysisAndProcessingUsingBigDataTool-HiveAndMachineLearningAlgorithm-Linear Regression||,International Journal of Computer Engineering & Technology (IJCET)
- [6] NehaMangla, ||Machine Learning Approach for UnstructuredData Using Hive, ||International Journal of Engineering Research||, –