BITCOIN PRICE PREDICTION USING LSTM

A PROJECT REPORT

Submitted by

K. ANANDARAJ	**712219104004**
M. HARISHKUMAR	**712219104011**
K. JAYAPRAKASH	**712219104012**
V. SIVABALAN	**712219104025**

In partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

PARK COLLEGE OF ENGINEEING AND

TECHNOLOGY

KANIYUR, COIMBATORE-641659

ANNA UNIVERSITY:: CHENNAI 600 025

MARCH 2021

BONAFIDE CERTIFICATE

Certified that this project report **"BITCOIN PRICE PREDICTION USING LSTM"** is the Bonafede work **"K. ANANDARAJ , M. HARISHKUMAR , K. JAYAPRAKASH , V. SIVABALAN**" who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
Dr.P. Vivekanandan, M.Tech., Ph.D.	Mr. G. Prabanchan, M.E.
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Professor	Assistant Professor
Department of Computer Science	Department of Computer Science
and Engineering,	and Engineering,
Park college of Engineering	Park college of Engineering
And Technology,	And Technology,
Kaniyur, Coimbatore-641 659	Kaniyur, Coimbatore-641 659

Submitted for the University Project VIVA -VOCE examination held on_____

Internal Examiner External Examiner

ACKNOWLEDGEMENT

First and fore most we praise and thank god and our parents, from the bottom of our heart for bestowing his benediction on us throughout the course of this project.

We wish to express our sincere thanks to our beloved Chairman **Dr.P.V.RAVI Ph.D.,MISTE**, for providing an opportunity to work on this project.

We would like to express our deep sense of gratitude to the Chief Executive Officer **Dr.ANUSHA .R B.E.,M.S.(USA), Ph.D,** for her valuable support and encouragement.

We extend sincere thanks to our Principal **Dr. G.MOHAN KUMAR B.E., M.E., Ph.D, MBA, PGDMRM, MIE, MISTE,** for

supporting and encouraging to carry out our project.

We are highly indebted and graceful to our beloved Head of the Department **Dr.P.VIVEKANANDAN M.Tech, Ph.D**, for his valuable and his constant suggestions and persistent encouragement.

We express our gracious gratitude to our beloved guide **Miss.A.VANISHREE M.E,** for her guidance and support throughout our project.

We also thank all the faculty members, teaching and non-teaching staff and lab in-charge of our department who helped us to complete our project successfully.

ABSTRACT

Bitcoin is the first digital decentralized cryptocurrency that has shown a significant increase in market capitalization in recent years. The objective of this paper is to determine the predictable price direction of Bitcoin in USD by machine learning technique. We explored several algorithms of machine learning using supervised learning to develop a prediction model and provide informative analysis of future market prices. Due to the difficulty of evaluating the exact nature of a Time Series model, it is often very difficult to produce appropriate forecasts. Then we continue to implement long short-term memory cells (LSTM) algorithm. Thus, we analyzed the time series model prediction of bitcoin prices with greater efficiency using long short-term memory (LSTM) techniques. The proposed LSTM model which used the individual features(Open, High, Low, Close) as input value. The models are evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The Results show 0.189 (MAE) and 0.039 (RMSE) for Open price , 0.143 (MAE) and 0.027 (RMSE) for Low price , 0.233 (MAE) and 0.057 (RMSE) for High price and 0.222 (MAE) and 0.052 (RMSE) for Close price. The low error value of these two metrics show that the models are efficient in predicting predicting price.

CONTENTS

CHAPTER NO	TITLE	PAGE NO
	FRONT PAGE	**i**
	BONAFIDE CERTIFICATE	**iii**
	ACKNOWLEDEMENT	**iv**
	ABSTRACT	**v**
	CONTENTS	**vi**
	LIST OF FIGURES	**viii**
	LIST OF TABLES	**ix**
	LIST OF SYMBOLS AND ABBREVIATION	**x**
1	**INTRODUCTION**	**1-2**
1.1	BITCOIN PRICE PREDICTION	1
1.2	MACHINE LEARNING	1
1.3	TIME SERIES ANALYSIS	1
1.4	LONG SHORT TERM MEMORY	2
2	**LITERATURE SURVEY**	**3-6**
3	**SYSTEM ANALYSIS**	**7**
3.1	EXISTING SYSTEM	7
3.2	PROBLEM STATEMENT	7
3.3	PROPOSED SYSTEM	7
4	**SYSTEM DESIGN**	**8-10**
4.1	SYSTEM MODEL	8
	4.4.1 DATA COLLECTION	9
	4.4.2 PRE-PROCESSING	9
	4.4.3 LSTM MODEL	9
5	**IMPLEMENTATION**	**12**
5.1	OPEN PRICE	12
5.2	HIGH PRICE	12
5.3	LOW PRICE	13

5.4 CLOSE PRICE 13

6 RESULTS AND DISUSSIONS 14-15

6.1 RESULT ANALYSIS FOR BITCOIN PRICE DATASET 14

7 CONCLUSION 16

REFERENCES 17

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO
FIGURE 4.1	SYSTEM WORKFLOW	8
FIGURE 4.2	INTERNAL STRUCTURE OF LSTM	10
FIGURE 4.3	LSTM MODEL SUMMARY	11
FIGURE 5.1	OPEN PRICE	12
FIGURE 5.2	HIGH PRICE	12
FIGURE 5.3	LOW PRICE	13
FIGURE 5.4	CLOSE PRICE	13
FIGURE 6.1	PERFORMANCE OF BITCOIN PRICE DATASET	14

LIST OF TABLES

TABLE NO	TITLE	PAGE NO
TABLE 6.1	METRICS FOR BITCOIN PRICE DATASET	14

LIST OF SYMBOLS AND ABBREVIATIONS

ACRONYM	ABBREVIATION
LSTM	LONG SHORT TERM MEMORY
USD	UNITED STATE DOLLER
MAE	MEAN ABSOLUTE ERROR
RMSE	ROOT MEAN SQUARE ERROR
ML	MACHINE LEARNING
GRU	GATED RECUIREMENT UNITS
BI.LSTM	BIDIRECTIONAL LONG SHORT TERM MEMORY
BTC	BITCOIN
LTC	LITECOIN
ETH	ETHEREUM
MAPE	MEAN ABSOLUTE PERCENTAGE ERROR
MLP	MULTI LAYER PERCEPTRON
ARIMA	AUTOREGRESSIVE INTEGRATED MOVING AVERAGE
RNN	RECURRENT NEURAL NETWORK
SVR	SUPPORT VECTOR REGRESSION
LR	LOGISTIC REGRESSION
BSE	BOMBAY STOCK EXCHANGE
ANN	ARTIFICIAL NEURAL NETWORK
RF	RANDOM FOREST

CHAPTER 1

INTRODUCTION

1.1 Bitcoin Price Prediction:

Bitcoin (₿) is a decentralized digital currency that can be transferred on the peer-to-peer bitcoin network. Bitcoin transactions are verified by network nodes through cryptography and recorded in a public distributed ledger called a blockchain. The cryptocurrency was invented in 2008 by an unknown person or group of people using the name Satoshi Nakamoto. The currency began use in 2009 when its implementation was released as open-source software.

Cryptocurrency markets are highly volatile and your investments are at risk. According to experts' Bitcoin price prediction, the price of BTC will reach $100,000 by the end of 2022, rising to $256,117 by the end of 2025 and $1,010,923 in 2030.

1.2 Machine learning:

Machine learning (ML) is a type of artificial intelligence that can predict the future based on past data. ML-based models have various advantages over other forecasting models as prior research has shown that it not only delivers a result that is nearly or exactly the same as the actual result, but it also improves the accuracy of the result.

1.3 Time Series Analysis:

Time series is a sequence of observations of numeric variables indexed by a date or timestamp. A clear example of time series data is the time series of a stock price.

The most common application of time series analysis is forecasting future values of a numeric value using the temporal structure of the data. This means, the available observations are used to predict values from the future.

In statistics, a moving average is a calculation used to analyze data points by creating a series of averages of different subsets of the full dataset. In finance, a moving average is a stock indicator that is commonly used in technical/time series analysis.

The reason for calculating the moving average of a stock is to help smooth out the price data over a specified period of time by creating a constantly updated average price.

1.4 Long Short Term Memory (LSTM):

LSTM, which stands for Long Short Term Memory, is a type of neural network which is particularly useful in the case of time series forecasting. LSTM network is the most effective solution to time series analysis and thus Bitcoin price Prediction.

LSTM is a kind of Recurrent Neural Networks. LSTM is capable of learning “long term dependencies”. LSTMs are very powerful in sequence prediction problems because they are able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price.

CHAPTER 2

LITERATURE SURVEY

1. **Improving the Cryptocurrency Price Prediction Performance Based on Reinforcement Learning [2]**

   This paper proposed a machine learning-based approach to price prediction for a financial institution. The proposed system contains the blockchain framework for secure transaction environment and Reinforcement Learning algorithm for analysis and prediction of price. The main focus of this system is on Litecoin and Monero cryptocurrencies. The results show the presented system accurate the performance of price prediction higher than another state-of-art algorithm.

2. **A novel cryptocurrency price prediction model using GRU, LSTM and BI-LSTM machine learning algorithms [5]**

   The main aim of this paper is to predict the next-day cryptocurrency price and proposes three types of recurrent neural network (RNN) algorithms used to predict the prices of three types of cryptocurrencies, namely Bitcoin (BTC), Litecoin (LTC), and Ethereum (ETH). The models show excellent predictions depending on the mean absolute percentage error (MAPE). Results obtained from these models show that the gated recurrent unit (GRU) performed better in prediction for all types of cryptocurrency than the long short-term memory (LSTM) and bidirectional LSTM (bi-LSTM) models. Therefore, it can be considered the best algorithm. GRU presents the most accurate prediction for LTC with MAPE percentages of 0.2454%, 0.8267%, and 0.2116% for BTC, ETH, and LTC, respectively. The bi-LSTM algorithm presents the lowest prediction result compared with the other two algorithms as the MAPE percentages are: 5.990%, 6.85%, and 2.332% for BTC, ETH, and LTC, respectively. Overall, the prediction models in this paper represent accurate results close to the actual prices of cryptocurrencies. The importance of having these models is that they can have significant economic ramifications by helping investors and traders to pinpoint cryptocurrency sales and purchasing.

3. **Stochastic neural networks for cryptocurrency price prediction [4]**

   This paper proposed a stochastic neural network model for Cryptocurrency price prediction. The proposed approach is based on the random walk theory, which is widely used in financial markets for modelling stock prices. The proposed model induces layer-wise randomness into the observed feature activations of neural networks to simulate market volatility. Moreover, a technique to learn the pattern of the reaction of the market is also included in the prediction model. We trained the Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) models for Bitcoin, Ethereum, and Litecoin. The results show that the proposed model is superior in comparison to the deterministic models.

4. **Bitcoin price prediction using machine learning: An approach to sample dimension engineering [1]**

   In this paper, predict Bitcoin price at different frequencies using machine learning techniques, it first classifies Bitcoin price by daily price and high-frequency price. A set of high-dimension features including property and network, trading and market, attention and gold spot price are used for Bitcoin daily price prediction, while the basic trading features acquired from a cryptocurrency exchange are used for 5-minute interval price prediction. Statistical methods including Logistic Regression and Linear Discriminant Analysis for Bitcoin daily price prediction with high-dimensional features achieve an accuracy of 66%, outperforming more complicated machine learning algorithms. Compared with benchmark results for daily price prediction, these works achieve a better performance, with the highest accuracies of the statistical methods and machine learning algorithms of 66% and 65.3%, respectively. Machine learning models including Random Forest, XGBoost, Quadratic Discriminant Analysis, Support Vector Machine and Long Short-term Memory for Bitcoin 5-minute interval price prediction are superior to statistical methods, with accuracy reaching 67.2%. Their investigation of Bitcoin price prediction can be considered a pilot study of the importance of the sample dimension in machine learning techniques.

5. **Real-time prediction of bitcoin price using machine learning techniques and public sentiment analysis [3]**

   The objective of this paper is to determine the predictable price direction of Bitcoin in USD by machine learning techniques and sentiment analysis. Twitter and Reddit have attracted a great deal of attention from researchers to study public sentiment. They applied sentiment analysis and supervised machine learning principles to the extracted tweets from Twitter and Reddit posts, and analyze the correlation between bitcoin price movements and sentiments in tweets.Explored several algorithms of machine learning using supervised learning to develop a prediction model and provide informative analysis of future market prices. Due to the difficulty of evaluating the exact nature of a Time Series(ARIMA) model, it is often very difficult to produce appropriate forecasts. Then they continue to implement Recurrent Neural Networks (RNN) with long short-term memory cells (LSTM). Thus, we analyzed the time series model prediction of bitcoin prices with greater efficiency using long short-term memory (LSTM) techniques and compared the predictability of bitcoin price and sentiment analysis of bitcoin tweets to the standard method (ARIMA). The RMSE (Root-mean-square error) of LSTM are 198.448 (single feature) and 197.515 (multi-feature) whereas the ARIMA model RMSE is 209.263 which shows that LSTM with multi feature shows the more accurate result.

6. **Effects of activation functions and optimizers on stock price prediction using LSTM recurrent networks [7]**

   The research was carried out on historical stock price dataset from Spanish Stock Company to analyse future market/stock values. In Linear Regression (LR), Support Vector Regression (SVR) and Long Short Term Memory (LSTM) algorithms were used to predict stock market prediction. And also compared the different activation functions with different optimizers and concluded that the tanh activation with the Adam algorithm performs best with an accuracy of 98.49%. The results reveals that the performance of LSTM model was better than the other models.

7. **Stock Price Prediction using LSTM on Indian Share Market [6]**

   The Experiment an LSTM model to predict the stock price of selected companies were collected from Bombay Stock Exchange (BSE) website. The

selected features were used to Date and Closing Price. The output of the results was analysed the different pattern of the share price of different companies and prediction of growth rate of different companies in different time spans. The results showed that the Error value was below 0.944.

8. **Stock Closing Price Prediction using Machine Learning Techniques [8]** The main aim of this paper is to predict the next day closing price of stock using Artificial Neural Network (ANN) and Random Forest (RF) algorithms. The research was carried out on historical stock price dataset of top 5 companies from Yahoo Finance. The models were evaluated using standard strategic indicators: Mean Absolute percentage Error (MAPE) and Root Mean Squared Error (RMSE). The results showed that the best values obtained by ANN model gives Root Mean Square Error (0.42) and Mean Absolute percentage Error (0.77).

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Statistical methods including Logistic Regression for Bitcoin daily price prediction with an accuracy of 66%. In this paper, compared with benchmark results for daily price prediction, machine learning models including Random Forest, Support Vector Machine for predicted the bitcoin price and its accuracy of 66% and 65.3%, respectively.

3.2 PROBLEM STATEMENT

Bitcoin is the most complex cryptocurrency which value change in every second. Investing money for bitcoin is more risk and less profit.

3.3 PROPOSED SYSTEM

In our project, we used a time series model (LSTM) to leverage machine learning technology to predict the real-time price of Bitcoin. However, machine learning literature is lacking verification of whether or not the stock evaluation strategies are legitimate for the cryptocurrencies, and if so, how they may be modified. That is what features want to be eliminated or introduced as a foundation for price prediction, whether current machine learning algorithms work for cryptocurrencies, and which technique yields the excellent outcomes.

CHAPTER 4

SYSTEM DESIGN

4.1 SYSTEM MODEL

The proposed system is composed by four steps: Data Collection, Pre-processing, Train the LSTM model and performance analysis.

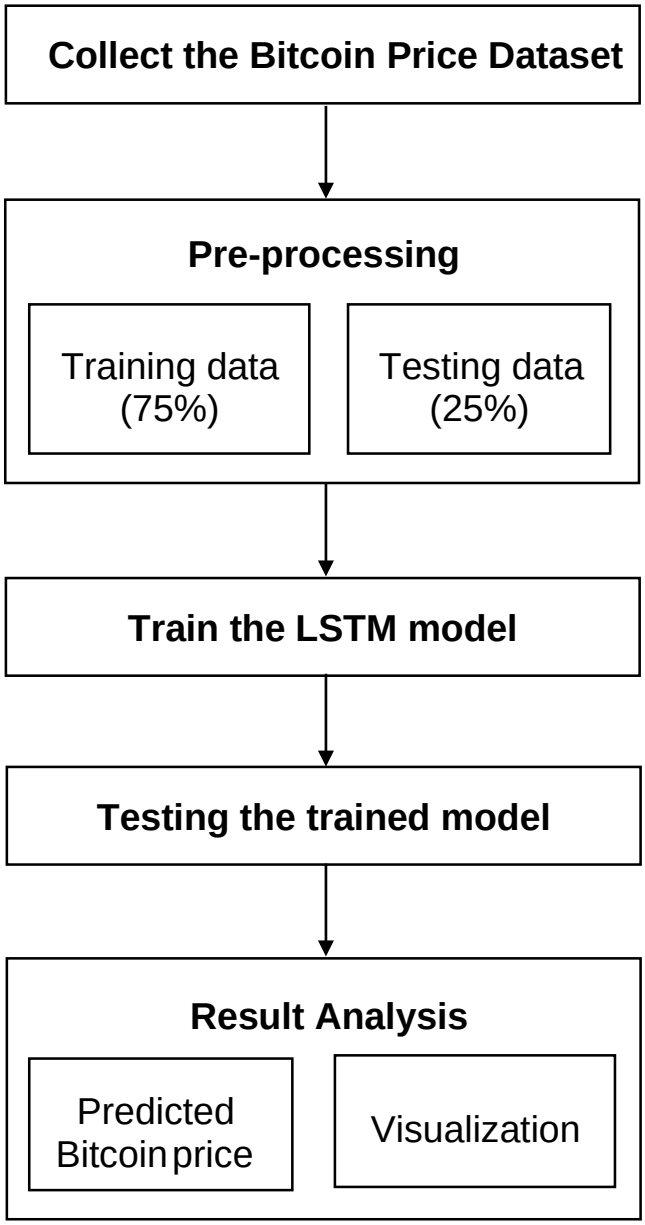

Figure 4.1 System Workflow

4.4.1 Data Collection

In our project, we are collecting dataset of 8 years 2012-01-01 to 2020-09-14 from Kaggle site.

The data contains information about the bitcoin such as,

- **Timestamp** **:** Specifies trading date and time
- **High** **:** Maximum price during the day
- **Low** **:** Minimum price during the day
- **Open** **:** Open price of the day
- **Close** **:** Close price of the day
- **Volume (BTC)** **:** Volume of BTC transacted during the day
- **Volume (Currency):** Volume of corresponding currency transacted during the day
- **Weighted_Price** **:** Volume Weighted Average Price

4.4.2 Pre-processing

i. **Data cleaning** **:** The unwanted attributes and empty rows were filled by some filling methods.

ii. **Data split** **:** Then data set is split into training and testing data with respective percentages of 75% and 25%

iii. **Feature Scaling:** To scale the bitcoin prices between (0, 1) to avoid intensive computation. LSTM model to take normalization, particularly when working on RNN with a sigmoid function in the output layer.

4.4.3 LSTM Model

Long Short Term Memory (LSTM) is a many types of recurrent neural network capable of learning data from past stages and use it for future predictions. In general, a neural a neural networks consists of three layers: input layers, hidden layers and output layer.

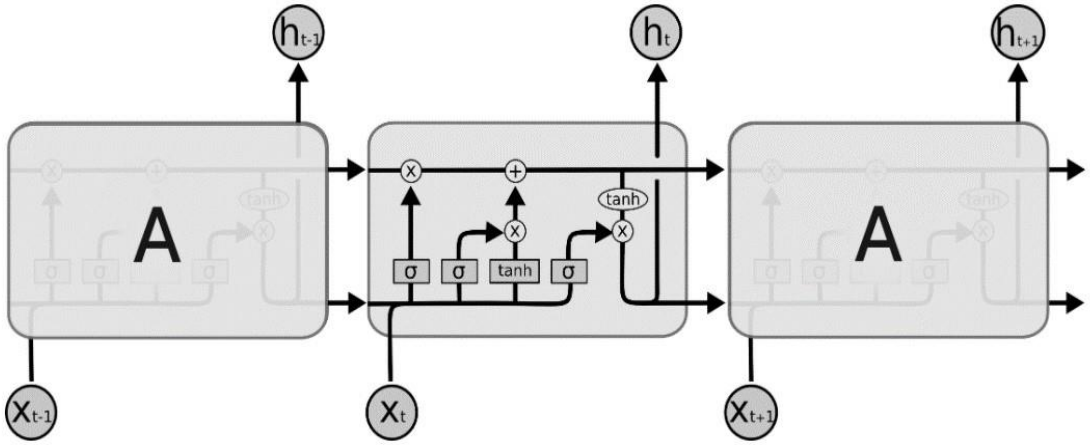

Figure 4.2 Internal Structure of LSTM

There are three gates to protect and control the cell states: Forget gate, Input gate and Output gate. Input Gate chooses which new data will be stored in the cell. First, a sigmoid layer “input door layer” chooses which values will be changed. Next, a tanh layer makes a vector of new candidate values that could be added to the state. Forget Gate outputs a number between 0 and 1, where 1 illustration “completely keep this”; whereas, 0 indicates “completely ignore this”. Output Gate decides what will be the output of each cell. The output value will be based on the cell state along with the filtered and freshest added data.

Train the LSTM Model

The neural network is trained by feeding the training dataset. The model is initiated using random weights and biases. LSTM model consists of a sequential input layer followed by 3 LSTM layers, later add a few Dropout layers to prevent over-fitting and then a dense layer with activation. Add the LSTM layer with the following arguments: 50 units which is the dimensionality of the output space. Return sequences = true, which determines whether to return the last output in the output sequence, or the full sequence. Input shape as the shape of our training set.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 100, 50)           10400

 dropout (Dropout)           (None, 100, 50)           0

 lstm_1 (LSTM)               (None, 100, 50)           20200

 dropout_1 (Dropout)         (None, 100, 50)           0

 lstm_2 (LSTM)               (None, 100, 50)           20200

 dropout_2 (Dropout)         (None, 100, 50)           0

 lstm_3 (LSTM)               (None, 50)                20200

 dropout_3 (Dropout)         (None, 50)                0

 dense (Dense)               (None, 1)                 51

=================================================================
Total params: 71,051
Trainable params: 71,051
Non-trainable params: 0
_________________________________________________________________
```

Figure 4.3 LSTM Model Summery

After that, compiled the model using the ‘adam’ optimizer and set the loss as the ‘mean squared error’. Next, compiling and fitting the training set to run on 50 epochs with 32 batch sizes.

CHAPTER 5

IMPLEMENTATION 5.1

5.1 Open price:

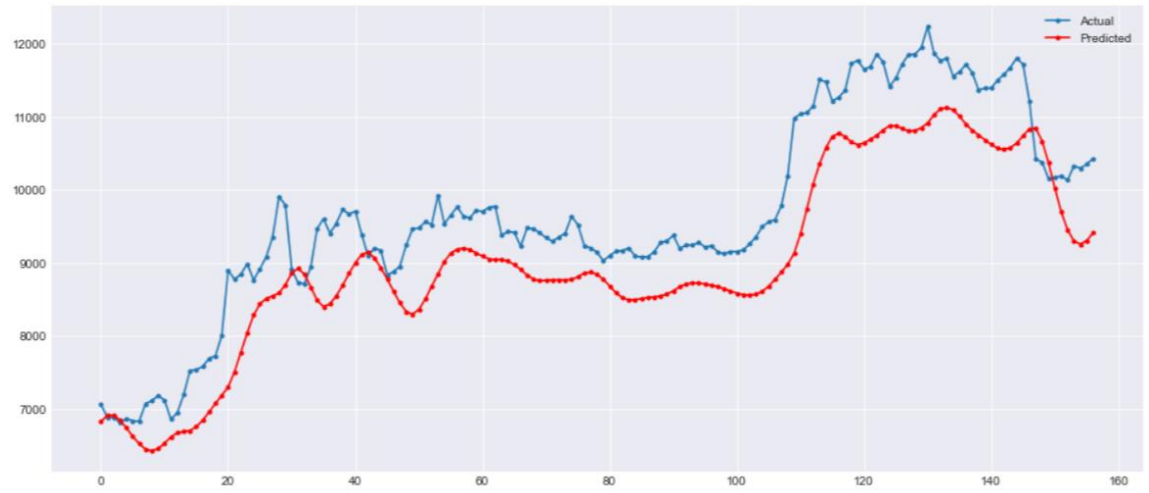

Figure 5.1 Bitcoin price prediction of Open price

5.2 High price:

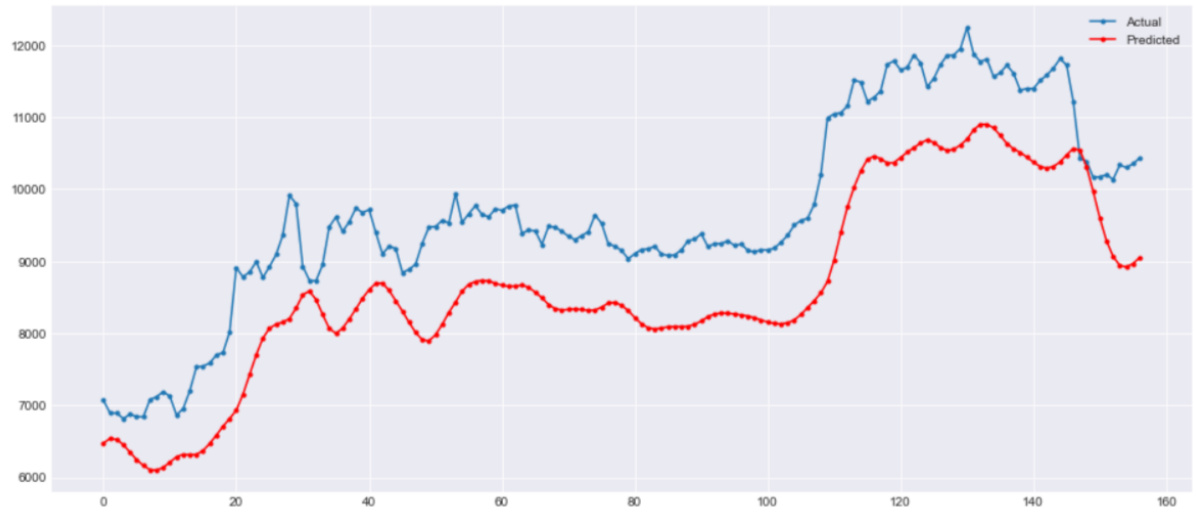

Figure 5.2 Bitcoin price prediction of High price

5.3 Low price:

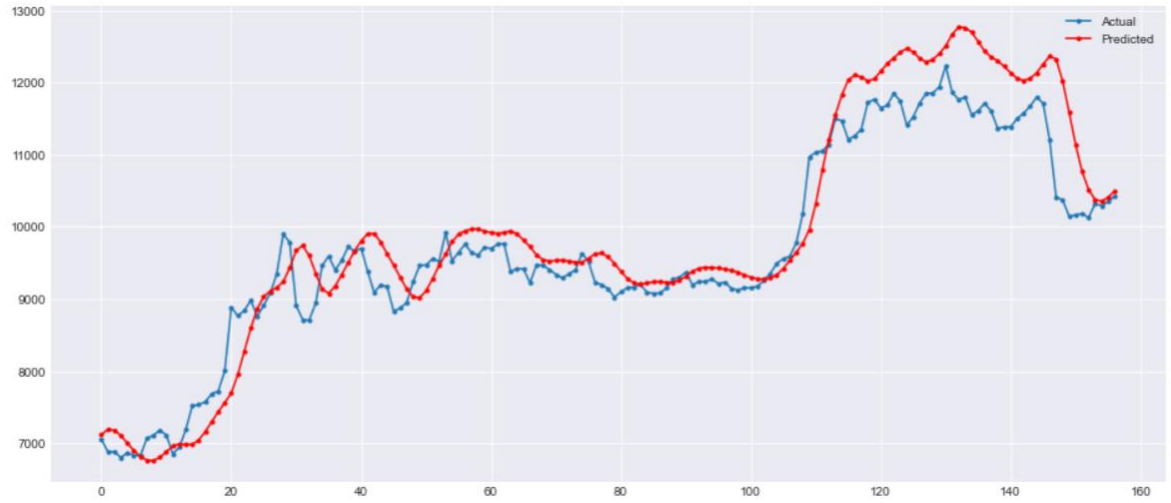

Figure 5.3 Bitcoin price prediction of Low price

5.4 Close price:

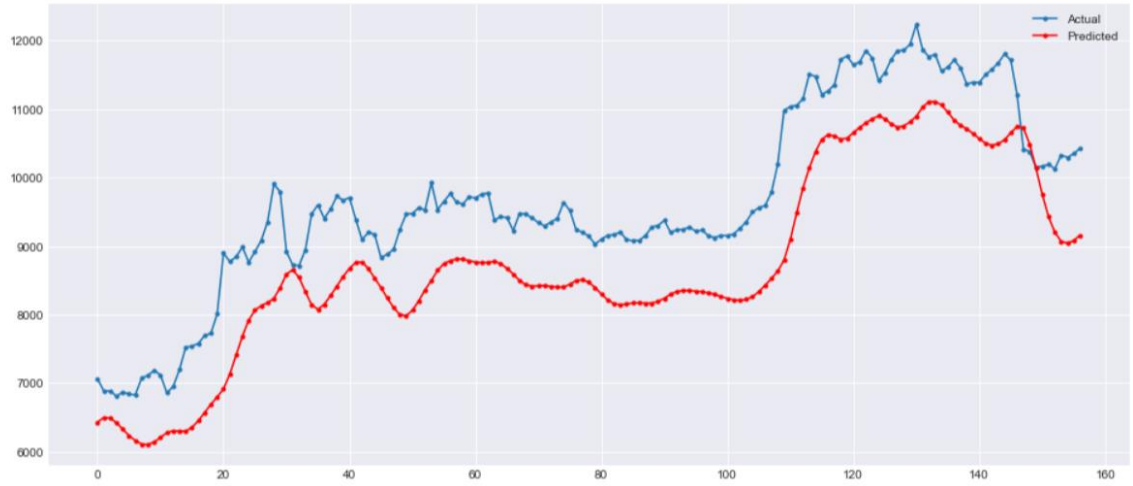

Figure 5.4 Bitcoin price prediction of Close price

CHAPTER 6

RESULTS AND DISCUSSIONS

6.1 Result Analysis for Bitcoin Price dataset

The analysis was carried out on Bitcoin Price dataset. In LSTM model, the Mean Absolute Error shows 0.189, Root Mean Squared Error shows 0.039 for open price, Mean Absolute Error shows 0.143, Root Mean Squared Error shows 0.027 for low price, Mean Absolute Error shows 0.233, Root Mean Squared Error shows 0.057 for high price, Mean Absolute Error shows 0.222, Root Mean Squared Error shows 0.052 for close price. So, the overall performance shows better results.

Table 6.1 Metrics for Bitcoin Price dataset

Features/Metrics	MAE	RMSE
OPEN	0.189	0.039
LOW	0.143	0.027
HIGH	0.233	0.057
CLOSE	0.222	0.052

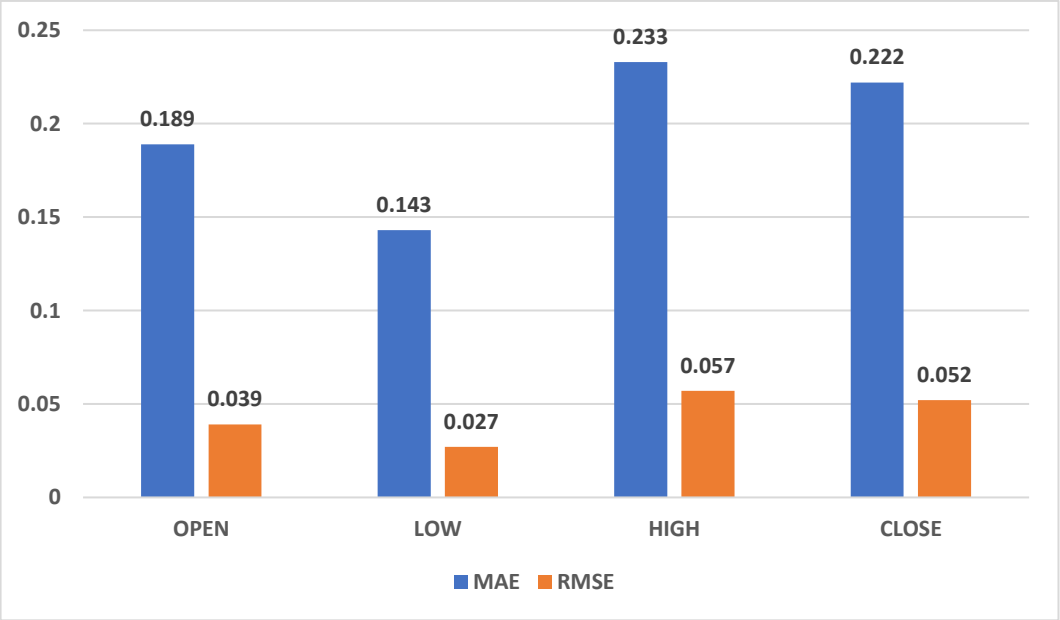

Figure 6.1 Performance of Bitcoin Price dataset

To evaluate the effectiveness of the model, compared and analysed the performance of 50 epochs with the 32 batch size on Bitcoin Price Dataset using LSTM model. Predicted Bitcoin prices were evaluated to Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for finding the final minimalized errors in the predicted price.

Mean Absolute Error (MAE):

MAE is obtained by obtaining the absolute difference between the model predictions and the true (actual values).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - x|$$

where 'x_i' refers to the actual price, 'x' refers to the predicted price and 'n' refers to the total window size.

Root Mean Squared Error (RMSE):

RMSE represents the standard deviation of the error i.e difference between the model predictions and the true values (actual value).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - F_i)^2}{n}}$$

where 'O_i' refers to the actual price, 'F_i' refers to the predicted price and 'n' refers to the total window size.

CHAPTER 7

CONCLUSION

Predicting cryptocurrency market is a challenging task due to consistently changing price values. The analysis was carried out on dataset for predicting bitcoin price values using Long Short Term Memory (LSTM) model. And also compared and analysed the performance of 50 epochs for training data with the 32 batch sizes. The Mean Absolute Error shows 0.189, Root Mean Squared Error shows 0.039 for open price, Mean Absolute Error shows 0.143, Root Mean Squared Error shows 0.027 for low price, Mean Absolute Error shows 0.233, Root Mean Squared Error shows 0.057 for high price, Mean Absolute Error shows 0.222, Root Mean Squared Error shows 0.052 for close price. So, the overall performance shows better results.

REFERENCES

1. Chen, Z., Li, C. and Sun, W., 2020. Bitcoin price prediction using machine learning: An approach to sample dimension engineering. Journal of Computational and Applied Mathematics, 365, p.112395.
2. Shahbazi, Z. and Byun, Y.C., 2021. Improving the Cryptocurrency Price Prediction Performance Based on Reinforcement Learning. IEEE Access, 9, pp.162651-162659.
3. Springer, Cham. Raju, S.M. and Tarif, A.M., 2020. Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis. arXiv preprint arXiv:2006.14473.
4. Jay, P., Kalariya, V., Parmar, P., Tanwar, S., Kumar, N. and Alazab, M., 2020. Stochastic neural networks for cryptocurrency price prediction. Ieee access, 8, pp.82804-82818.
5. Hamayel, M.J. and Owda, A.Y., 2021. A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. AI, 2(4), pp.477-496.
6. Ghosh, A., Bose, S., Maji, G., Debnath, N. and Sen, S., 2019, September. Stock price prediction using LSTM on Indian share market. In Proceedings of 32nd international conference on (Vol. 63, pp. 101-110).
7. Rana, M., Uddin, M.M. and Hoque, M.M., 2019, December. Effects of activation functions and optimizers on stock price prediction using LSTM recurrent networks. In Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence (pp. 354-358)
8. Ghosh, A., Bose, S., Maji, G., Debnath, N. and Sen, S., 2019, September. Stock price prediction using LSTM on Indian share market. In Proceedings of 32nd international conference on (Vol. 63, pp. 101-110).