International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 01 | Jan 2023 www.irjet.net p-ISSN: 2395-0072

# Bitcoin Price Prediction and Recommendation System using Deep learning techniques and twitter sentiment analysis

**Gargi Singh [1], B.Indra Thanaya [2]**

[1] *Masters Student, Department of Computer Science and Engineering ,Indira Gandhi Delhi Technical University For Women Delhi*
[2] *Assistant Professor, Department of Computer Science and Engineering, Indira Gandhi Delhi Technical University For Women Delhi*

---

**Abstract –** *Prediction of bitcoin market trends is one of the important tasks that need dedicated attention as predicting bitcoin prices successfully leads to attractive profits by making the proper decision. Bitcoin market is challenging due to its non-stationary, blaring and chaotic data, and thus the prediction becomes challenging for investors to invest their money for making profits. The bitcoin price prediction based on historical data or textual information have shown to be unsatisfactory ,That's why we have used both for predicting the bitcoin price and recommendation system to buy or sell the bitcoin as per the fluctuation in the market. As the bitcoin price prediction is a Time series problem we have used different machine learning and deep learning techniques such as LSTM, ARIMA, and Linear Regression. Out of these ARIMA has performed really well with an accuracy of 80%.Existing Studies in the sentiment analysis have shown that there is a correlation between the fluctuation of bitcoin prices and twitter tweets. We have performed sentiment analysis on the latest 100 tweets about the bitcoin which will provide a recommendation to buy or sell the bitcoin.*

***Key Words*: LSTM, ARIMA, Machine Learning, Sentiment Analysis, Bitcoin Trade Open, Bitcoin Trade Close.**

## 1. INTRODUCTION

Bitcoin is a crypto currency that is being used globally for digital payment ,investment or for trading. Bitcoin prices are difficult to predict due to their extreme volatility, which is influenced by a variety of political and economic issues, as well as changes in leadership, investor attitude, and a variety of other factors. A model which is considering one component may not be reliable. As a combining both the tweets and the historical price data might improve the accuracy . There are primarily two approaches for predicting markets trends. Technical analysis and the fundamental analysis are two types of analysis. Fundamental analysis uses previous price and volume to forecast future trends, but technical analysis does not. Fundamental analysis of a current bitcoin prices, on the other hand, entails evaluating financial data to get insights. The efficient market theory, which holds that bitcoin market prices are basically unpredictable, casts doubt on the usefulness of both technical and fundamental analysis. The goal of this research work is to build a model which predicts the bitcoin market trends. Three models are used as a part of this research work. The model are ARIMA,LSTM, and Linear Regression. Sentiment Analysis is performed on the latest tweets about bitcoin.

LSTM model was first introduced by Hoc Hochreiter & Schmidhuber [1] which was capable of learning long term dependencies. Later on, many researched improves this work in[2][3][4].

The rest of the paper is organized as follows. Section 2 includes the research state of the bitcoin price prediction. Section 3 includes the Data collection & Preprocessing. Section 4 consists of methodologies used. Section 5 includes the Experimental Results. Section 6 concludes the paper.

## 2. LITRATURE SURVEY

The Siliverstovs of Manh Ha Duong Boris [5], investigated the relationship between equity prices and combined finances in key European countries such as the United Kingdom and Germany. Acceleration in European country investments is likely to result in a stronger link between European nation equity prices. If innovations in bitcoin markets effect actual financial instruments like investment and consumption. Lui Li[6],Examines the technical indicators and procedures of trading.

Kunal Gaur[7], the stock market time series prediction and recommendation system is developed using the parallel combination of LSTM-ARIMA-Linear Regression with Twitter Sentiment analysis which has show and improved accuracy of 83%.

R Batra[8] experimented the Sentiment analysis for better Prediction of Stock Price Movement use the twitter sentiment analysis for developing the prediction tool on the basis of tweets polarity.

A Mittal[9] proposed a cross validation method for financial data and obtained 75.6% accuracy in price prediction using Self Organization Fuzzy Neural Networks on the Twitter feeds.

A Raheman[10] had analyzed that the "out-of-the-box" Aigents model which had a correlation of ~0.33, and after

fine-tuning,"aigents+" has a correlation of ~0.57. "ensemble(all)" corresponds to average metrics across all models, and '"ensemble(top 3)" corresponds to the average of the best three models (aigents+, aigents and finBERT).

Pour [11] proposed that the cryptocurrency market prediction has been done with Deep Learning tools such as LSTM and Bayesian Optimization has been tested on stock market indices .The minimum batch size of training Epochs in deep learning algorithms is set equal to 32.

From all these researches done we were not able to find any work which have used more than 1 models and based on Polarity of tweets a recommendation of buy/sell is given to the user.

## 3. DATA COLLECTION AND PREPROCESSING

### 3.1 *Data Collection*

The data has been collected from three sources .

- The historic data has been downloaded from the year 2014 to 2022 dataset for Bitcoin Prices by using the Yahoo Finance API.
- To validate the data on real time data I have used the cryptoCMD API.This library allows to collect the latest data from CoinMarketCap I have used the predefined scraper object to get the data from Coin cap Market API.
- To perform the sentiment analysis of the tweets related to Bitcoin prices I have used the Twitter API which is accessible from the developers account.

### 3.2 *Data Preprocessing*

To make the data from the mode of entry appropriate for trustworthy analysis, it has to be pre-processed.

- We pre-processed the historical data in the following manner :-
- Data Pre-processing For ARIMA Model :-

| S.No. | Techniques |
|---|---|
| 1. | Filling the null values with backward fill method. |
| 2. | Taken 80% of the data for training and 20% data for testing. |

- Data Pre-processing for LSTM Model :-

| S.No. | Techniques |
|---|---|
| 1. | Scaled the values using Min- Max Scaler |
| 2. | Storing trends of a particular company from 7 days before current day to predict 1 next output and storing them to training part |
| 3. | Converting training list into numpy arrays |
| 4. | Adding 3rd Dimension to training part. |

- Data Pre-processing for Linear Reg Algorithm :-

| S.No. | Techniques |
|---|---|
| 1. | Declaring number of days (n) to be forecasted in future. |
| 2. | Declaring new dataframe with relevant data |

- Data Pre-processing for Tweets :-

| S.No. | Techniques |
|---|---|
| 1. | Cleaning up the tweets. |
| 2. | Passing the tweets to TextBlob for calculating the Polarity. |

## 4. METHODOLOGIES

### 4.1 ARIMA (Auto Regressive Integrating Moving Average)

ARIMA stands for auto regressive integrated moving average. It is a statistical analysis model that uses time series data to better understand the data set or anticipate future trends. If a statistical model predicts future values based on previous values, it is called autoregressive. For example, an ARIMA model may try to anticipate a company's earnings based on prior periods or predict a stock's future pricing based on historical performance. The model's final goal is to forecast future time series movement by looking at disparities between values in the series rather than actual values. When there is evidence of non-stationarity in the data, ARIMA models are used. Nonstationary data are always turned into stationary data in timeseries analysis.

We can break down the model into smaller components based on the name:

- The AR which stands for Autoregressive Model, shows a random process. The output of the model is linearly dependent on its prio value , such as the

number of lagged data points or previous observations.

- Integrative [I] :denotes the separating of raw observations so that the time series can become stationary [i.e. data values are replaced by the difference between the data values and the previous values ].
- Moving Average : It takes into account the relationship between an observation and a residual error from a lagged moving average model.

### 4.2 LSTM (long Shorth Term Memory Network)

It's a unique type of recurrent neural network that can learn long-term data relationships. This is possible because the model's recurring module is made up of four layers that interact with one another. An LSTM module has a cell state and three gates, giving it the ability to learn, unlearn, or retain information from each of the units selectively. By permitting only a few linear interactions, the cell state in LSTM allows information to travel across the units without being altered. Each unit contains an input, output, and a forget gate that adds or removes data from the cell state. The forget gate utilizes a sigmoid function to determine which information from the previous cell state should be ignored. The input gate uses a point-wise multiplication operation of 'sigmoid' and 'tanh' to control the information flow to the current cell state. Finally, the output gate determines which data should be transmitted on the next hidden state.

LSTM can be used in many applications such as for weather forecasting, NLP, speech recognition, hand writing recognition, time-series prediction, etc .The cell state is represented by the horizontal line that runs across the top of the figure. The condition of the cell is similar to a conveyor belt. This flows straight down the chain with just minimal linear interactions. The ability of LSTM to add or delete information from the cell state is controlled by gates. Gates are used to allow information to pass through if desired.

A sigmoid neural net layer plus a point wise multiplication operation make up gates. The sigmoid layer produces values ranging from 0 to 1, indicating how much of each component should be allowed to pass. Let nothing through with a value of 0, and everything through with a value of 1! To safeguard and govern the cell state, an LSTM contains three of these gates.

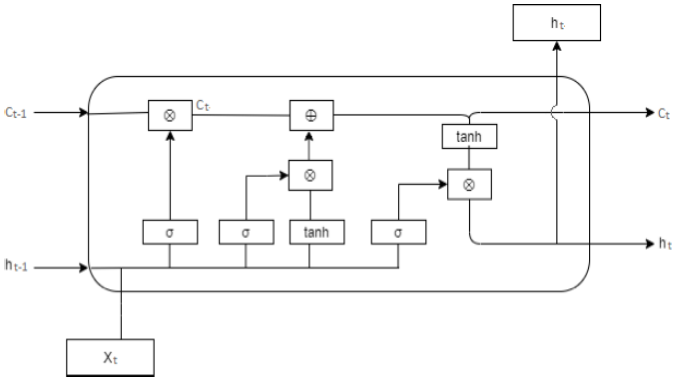


Fig 4.1: LSTM Architecture

The prior hidden state(ht-1), previous cell state(Ct-1) and present input are the inputs to the current cell state (Ct), as illustrated in Figure 3.4.1 (Xt). The forget gate , input gate and output gate are the three gates that make up the cell.

### 4.3 Linear Regression

Linear regression is a supervised machine learning algorithm. It carries out a regression task. Based on independent variables, regression models a goal prediction value . It carries out a regression task. Based on independent variables, regression models a goal prediction value. It is mostly utilised in forecasting and determining the link between variables. Different regression models differ in terms of the type of relationship they evaluate between dependent and independent variables, as well as the amount of independent variables they employ. Linear regression is used to predict the value of a dependent variable (y) given an independent variable (x). As a result of this regression technique , a linear relationship between x (input) and y output is discovered (output). Linear Regression gets its name from the equation

Y=m*x+c

### 4.4 Research Methodology

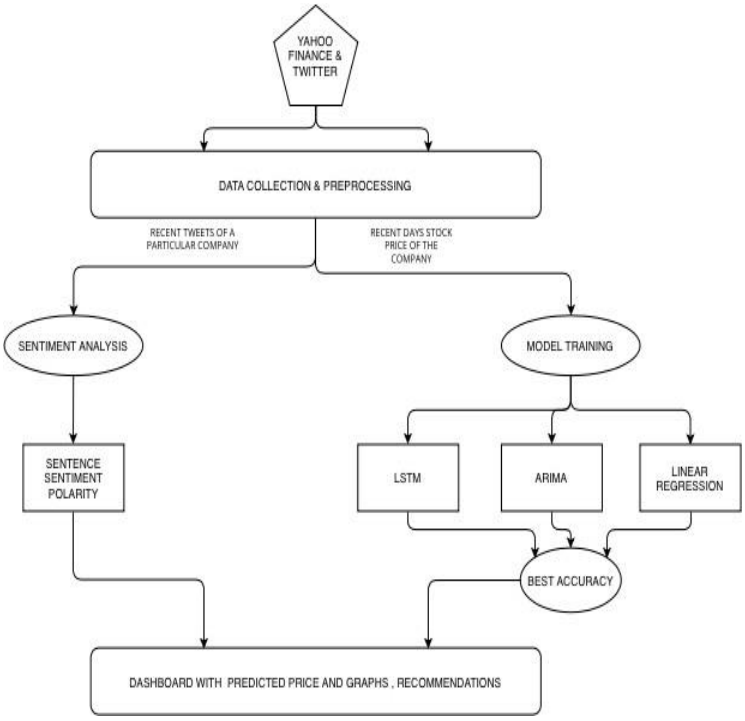


Figure IV.2 : System Architecture

Real-time data is retrieved through the CryptoCMD API, historic data is retrieved from the Yahoo Finance API, and sentiment analysis data is retrieved from the Twitter API. ARIMA,LSTM, and Linear Regression models were developed using the functions ARIMA ALGO(),LSTM ALGO(), and LIN REG ALGO(). These models accept the dataset variable as a parameter, and each model returns projected values and root mean square error. These models are trained for predicting its future price up to 7 days.

The Twitter Sentiment analysis model was developed using the function retrieving_tweets_polarity .

This model is accepting a string variable i.e. "btc" a short form for Bitcoin Currency. The tweets were fetched from twitter API by authenticating it.

The TextBlob library for Natural Language Processing has been to analyse the polarity and subjectivity of the sentence. The polarity of texblob function lies between [-1,1].I have counted the positive polarity and negative polarity of the latest tweets.

This function is returning the polarity ,tweets list , tweets pool generating the overall polarity of the sentences, count of positive polarity , count of negative polarity and the count of neutral polarity.

These function are called individually and their value has been stored in Global variable .

These global values are passed into the recommending function which is accepting historic-dataset , global_polarity , real-time-dataset and the mean value which is the prediction value of the most accurate model i.e. ARIMA.

This function is returning the idea and decision , where the idea describes the next observed pattern in market i.e. Rise or Fall and decision is giving the user suggestion to buy or sell the bitcoin as per the fluctuation of market.

## 5. EXPERIMENTAL RESULTS

The models ran for recent bitcoin prices. Some illustrations are given below. We'll be seeing the RMSE(Root Mean Squared Error) for these bitcoin price.

- ARIMA:-ARIMA model was on the latest bitcoin prices. We have achieved an RMSE value of 19.17.

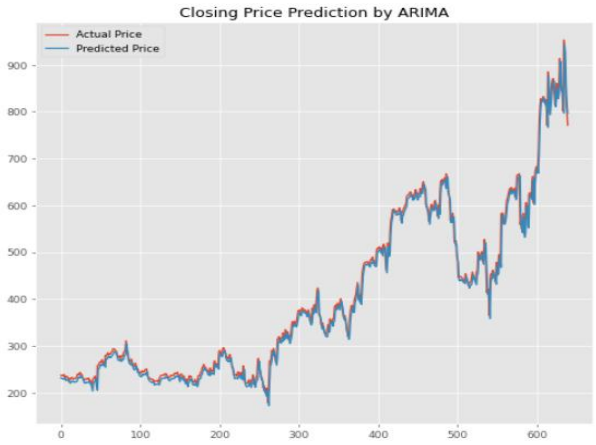


Fig 5.1: ARIMA Model Predictions

- LSTM :-LSTM model was on the latest bitcoin prices for various epochs as depicted in the table below. After 30 epochs we got a good RMSE value.

| Epochs | RMSE |
|---|---|
| 5 | 150.150 |
| 10 | 302.187 |
| 15 | 390.529 |
| 30 | 78.27 |

Table 5.1 (LSTM On Various Epochs)

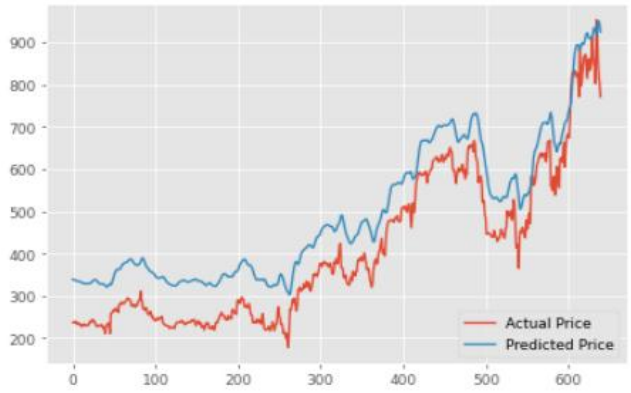


Figure 5.2. :- LSTM Model Accuracy

- Linear Regression :-Linear Regression model was ran and got an RMSE of 89.53.Following is the graph plotted for Linear Regression Model.


Figure 5.3. :- Linear Regression Model Accuracy

- Sentiment Analysis :-
- Taken latest 100 tweets about the bitcoin .
- Calculated Polarity by using TextBlob and got overall polarity as Positive.

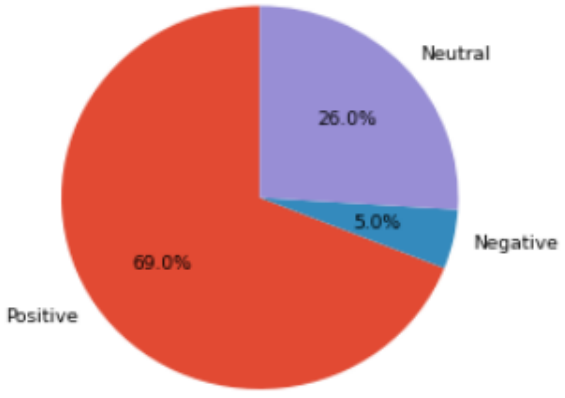


Figure 5.4 :- Polarity From Tweets

## 6. CONCLUSION

In recent years, most people have been investing in the bitcoin market in order to make quick money. At the same time, an investor stands a good probability of losing all of his or her money. To comprehend future market trends, the user will need an effective model.

Many prediction models exist that can anticipate whether the market is going up or down, but they are inaccurate. A model for predicting the bitcoin market movement for the next day has been attempted. A model has been constructed and evaluated using diverse bitcoin market data accessible open source, taking into account numerous patterns such as continuous up/down, volume traded each day, and also includes corporate sentiment.

On the considered dataset, LSTM and ARIMA model are performing best.

We have also performed sentiment analysis on twitter data to detect polarity of that particular tweet. Recommendation System is running well with the help of polarity of each tweet.

In Future we can make this research broader by predicting other crypto currencies prices and including different advanced model for sentiment analysis.

## ACKNOWLEDEGEMNT

On this auspicious occasion of accomplishment of our internship project on " Bitcoin Price Prediction and Recommendation System using Deep learning techniques and twitter sentiment analysis ", I would like to thank Prof. Indra Thanaya, Professor, CSE Department, IGDTUW, who has supported us through all the ups and downs in the completion of this project. He not only guided us throughout the project but also taught life values such as never giving up attitude and eagerness to learn and explore more in the field of ML.

We would be thankful to our honourable vice chancellor Dr.Amita Dev, Indira Gandhi Delhi Technical University for including the internship program and providing an opportunity to gain practical experience in research work.

We are grateful to our honorable HOD, Prof. Seeja. K.R., Computer Science and Engineering Department, IGDTUW for showing her trust and choosing us for this internship project work and encouraging us further.

Finally, the support and coordination received from all the team members is great and their contribution was vital for the completion of this internship project. I hope we will achieve more in our future endeavors.

## REFERENCES

[1] Hochreiter, Sepp, and Jürgen Schmidhuber. "LSTM can solve hard long time lag problems." *Advances in neural information processing systems* 9 (1996).

[2] Bengio, Yoshua, Patrice Simard, and Paolo Frasconi. "Learning long-term dependencies with gradient descent is difficult." *IEEE transactions on neural networks* 5.2 (1994): 157-166.

[3] Hochreiter, Sepp, and Jürgen Schmidhuber. "LSTM can solve hard long time lag problems." *Advances in neural information processing systems* 9 (1996).

[4] Hochreiter, Sepp. "The vanishing gradient problem during learning recurrent neural nets and problem solutions." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6.02 (1998): 107-116.

[5] Duong, Manh Ha, and Boriss Siliverstovs. "The stock market and investment." *unpublished Manuscript, Available at http://www. finprop. de/Paper5_The_Stock_Market. pdf (accessed 2 June 2010)* (2006).

[6] Liu, Li. "Are Bitcon returns predictable?: Evidence from technical indicators." *Physica A: Statistical Mechanics and its Applications* 533 (2019): 121950.

[7] Kunal Gaur. "*STOCK PRICE PREDICTION AND RECOMMENDATION USINGMACHINE LEARNING TECHNIQUES AND TWITTERSENTIMENT ANALYSIS*" IRJET 2020

[8] R. Batra and S. M. Daudpota, "*Integrating StockTwits with sentiment analysis for better prediction of stock price movement,*" 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2018, pp. 1-5, doi: 10.1109/ICOMET.2018.8346382.

[9] Mittal, Anshul, and Arpit Goel. "Stock prediction using twitter sentiment analysis." *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)* 15 (2012): 2352.

[10] Raheman, Ali, et al. "Social Media Sentiment Analysis for Cryptocurrency Market Prediction." *arXiv preprint arXiv:2204.10185* (2022).

[11] Pour, Ehsan Sadeghi, et al. "*Cryptocurrency Price Prediction with Neural Networks of LSTM and Bayesian Optimization." European Journal of Business and Management Research* 7.2 (2022): 20-27.

[12] Abraham, Jethin, et al. "Cryptocurrency price prediction using tweet volumes and sentiment analysis." *SMU Data Science Review* 1.3 (2018): 1.

[13] Rather, Akhter Mohiuddin. "*LSTM-based Deep Learning Model for Stock Prediction and Predictive Optimization Model.*" EURO Journal on Decision Processes 9 (2021): 100001.

[14] Li, Xinyi, et al. "DP-LSTM: Differential privacy-inspired LSTM for stock prediction using financial news." arXiv preprint arXiv:1912.10806 (2019).

[15] Gandhmal, Dattatray P., and K. Kumar. "Wrapper-Enabled feature selection and CPLM-based NARX model for stock market prediction." The Computer Journal 64.2 (2021): 169-184.

[16] Kumar, K., and Dattatray P. Gandhmal. "An intelligent Indian stock market forecasting system using LSTM deep learning." Indones J Electr Eng Comput Sci 21.2 (2021): 1082-1089.

## BIOGRAPHIES

Name: Gargi Singh
Masters Student, Department of Computer Science and Engineering ,Indira Gandhi Delhi Technical University For Women Delhi

Name: B.Indra Thanaya ,Assistant Professor, Department of Computer Science and Engineering, Indira Gandhi Delhi Technical University For Women Delhi