Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction

Apoorva Aggarwal[1], Isha Gupta[1], Novesh Garg[1] and Anurag Goel[2]
Department of Computer Science
Jaypee Institute of Information Technology
Noida, India
Email: [1]{apoorvaaggarwal11, isha.gupta998, noveshgarg}@gmail.com and [2]anurag.goel@jiit.ac.in

***Abstract* – Investment in cryptocurrency has been in trend from last many years. Bitcoin is one of the most popular and valuable cryptocurrency. Many studies have been done on bitcoin price prediction using various parameters which includes bitcoin factors, social media etc. In this paper, a comparative study of the various parameters affecting bitcoin price prediction is done based on Root Mean Square Error (RMSE) using various deep learning models like Convolutional Neural Network (CNN), Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). We have studied the effect of Gold price on the price of bitcoin.**

Keywords - Bitcoin Price Prediction, Twitter Sentiment Analysis, Deep Learning, Cryptocurrency

I. INTRODUCTION

With the enhancement of technologies in various sectors, digitalization has taken over many industries which benefits customers as well as the organizations. Cryptocurrencies has become popular as one of the digitalization in finance sector over a period of time. Bitcoin is the first decentralized digital currency since it is not governed by any central bank or authority. It was first originated in 2009 but comes in trend in 2017.Bitcoin is used worldwide for digital payments as well as for investment purposes. It is recommended for the investors to do proper investment risk analysis before investing in bitcoin.

In the recent years, many studies have been done on bitcoin price prediction. There are various factors which affect the price value of bitcoin. Initially, only bitcoin factors like opening and closing price values of bitcoin, highest and lowest price values of bitcoin reached in a day were taken into consideration to predict the bitcoin price[1]. Some recent studies also focused on social platforms like twitter as now-a-days social networking sites are used as central communicating channel for cryptocurrencies. The various features of twitter like tweets sentiment score, number of followers of the user, number of retweets, number of likes etc. were used as parameter to increase the accuracy of the prediction results of bitcoin price [8].

978-1-7281-3591-5/19/$31.00 ©2019 IEEE

II. LITERATURE SURVEY

The previous work uses bitcoin parameters and social media content to predict the bitcoin price using various neural network models.

A. Bitcoin Parameters

In [1], Artificial Neural Network (ANN) is used for predicting the next day price of bitcoin. They used four ANN algorithms namely Neuro Evolution of Augmenting Topologies (NEAT), Genetic Algorithm Neural Network (GANN), Genetic Algorithm Backpropagation Neural Network (GABPNN) and Backpropagation Neural Network (BPNN). These methods are compared based on their algorithmic complexity and accuracy of the results. In [2], bitcoin price prediction is modelled as a binomial classification problem using machine learning algorithms like generalized linear models and random forests. [3] focuses on optimal features affecting bitcoin prices and proposed price prediction approach using different machine learning algorithms like Bayesian Regression and Random forest. [4] uses Bitcoin Price Index as data source and compares the ARIMA model with various deep learning models like Bayesian optimised recurrent neural network (RNN) and a Long Short Term Memory (LSTM) network on the basis of accuracy and Root Mean Square Error (RMSE) for bitcoin price prediction. [5] compares linear regression model with polynomial regression models like Support Vector regression and KNN regression and on the basis of RMSE, it concludes that Support Vector models have been outperformed by KNN.

B. Social Parameters

[6] analyses twitter data and uses supervised learning models like Naive Bayes classifier, logistic regression and support vector machines for day-to-day and hour-to-hour price prediction of bitcoin. [7] uses hidden Markov Model with different cryptocurrencies to predict bubbles in the digital currency. [8] uses multi-linear regression approach on two cryptocurrencies - Litecoin and Bitcoin. These cryptocurrencies have a large number of users and huge market size. [9] analyses

fluctuations in three different cryptocurrencies' prices and is based on user comments in online cryptocurrencies communities.

III. METHODOLOGY

In this section, we have shown the methodology which has been used to predict the bitcoin price. We have included both social and financial factors to predict the bitcoin price to ease out predictions associated with bitcoin over years. We have also used a new parameter Gold in predicting the price of bitcoin. Gold is one of the most important financial aspects which has a great influence in determining a country's economy. Therefore, the trends of gold price in the international market can also be seen as a parameter to give a better approach for bitcoin price prediction.

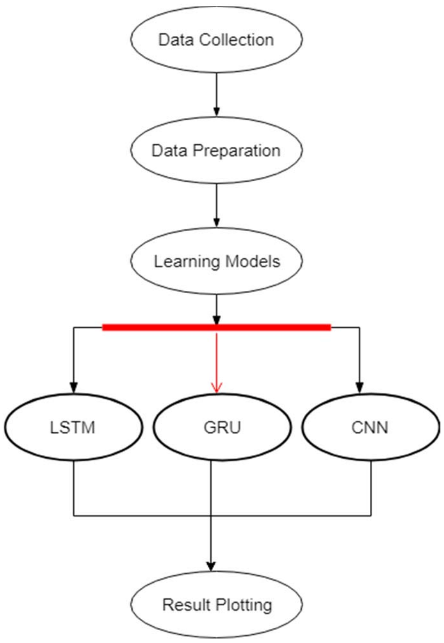

Fig. 1. Methodology Used

We have applied three deep learning algorithms to estimate the price of bitcoin namely Convolutional Neural Network (CNN), Long Short Term Memory (LSTM) and Gated Recurrent Units (GRU) and analysed their performance to identify which algorithm giving least RMS error.

A. Convolutional Neural Network (CNN)

A one-dimensional Convolutional Neural Network is a class of deep learning which is used for image recognition, image classification, object detection, face recognition etc. Fig. 2 shows the internal wiring of CNN model.

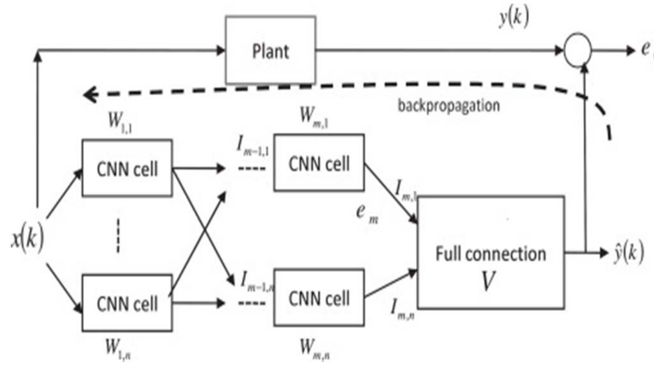

Fig. 2. CNN Internal Wiring

B. Long Short Term Memory (LSTM)

LSTM is a variation of Recurrent Neural Network (RNN). Vanilla RNN has a vanishing gradient problem which was overcome by LSTM. It uses three gates -input, output and forget gate. Therefore, it manages when to remember and when to forget. Hence, they are capable of learning long-term dependencies. Fig. 3 shows the internal wiring of LSTM model.

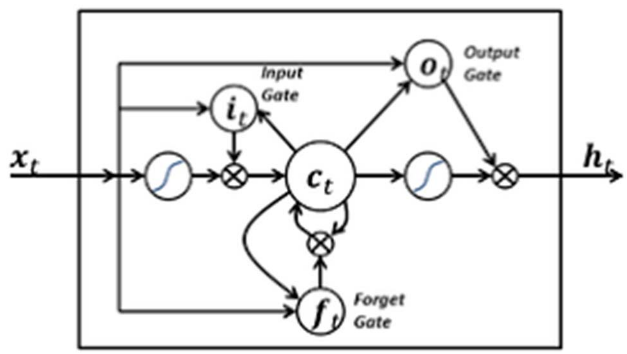

Fig. 3. LSTM Internal Wiring

C. Gated Recurrent Units (GRU)

GRU can be considered as a variant of LSTM. GRU solves the vanishing gradient problem by update gate and reset gate. Fig. 4 shows the internal wiring of GRU model.

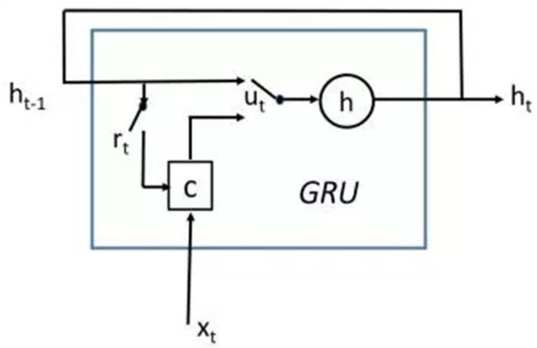

Fig. 4. GRU Internal Wiring

IV. IMPLEMENTATION

A. Datasets Used

We have used dataset from Poloniex which records the bitcoin price on an interval of 5 minutes. The dataset for gold price is taken from datahub.io which records the gold price on daily basis. The corpus of tweets for analysing social factors affecting bitcoin price is collected using Twitter API. Dataset for positive and negative words was downloaded in order to analyse the tweet sentiment score. [10]

B. Data Preprocessing

Since the available dataset of bitcoin was of five minutes interval basis, it is first mapped to daily basis so as to maintain the consistency in both the bitcoin price and gold price datasets. The data is then parsed and feed to the various deep learning models for bitcoin price prediction. Further, tweets are analysed by VADER (Valence Aware Dictionary for Sentiment Reasoning) algorithm for sentiment analysis. A score is generated based on sentiment, number of followers, likes and the number of tweets in a time range.

IV. RESULT ANALYSIS

Fig. 5, 6, 7 shows the results of bitcoin price prediction of CNN model, LSTM model and GRU model respectively using the bitcoin parameters. In the graph, the blue line shows the Actual Bitcoin price while the red dotted portion shows the Predicted Bitcoin price for the different months of year 2018.

The accuracies of these models are tabulated in Table 1. The Root Mean Square Error (RMSE) of LSTM model is least. Thus, LSTM proves to be more capable for predicting long-term dependencies as compared to CNN and GRU as shown in Table 1. This is due to the presence of dependency on past prices which are captured by LSTM model but not captured by the GRU and the CNN models.

Fig. 5. CNN Results Using Bitcoin Price Parameters

Fig. 6. LSTM Results Using Bitcoin Price Parameters

Fig. 7. GRU Results Using Bitcoin Price Parameters

Models	Number of Layers	RMSE
CNN	2	61.23
LSTM	4	47.91
GRU	1	55.98

Table 1. Accuracy Results of Deep Learning Models Using Bitcoin Price Parameters

The economy of a country determines the change in prices of gold which progressively brings a change in prices of various financial instruments. Fig. 8, 9, 10 shows the results of bitcoin price prediction of CNN model, LSTM model and GRU model respectively using the gold price as the parameter. In the graph, the blue line shows the actual price and the predicted price of bitcoin respectively for the different months of year 2017.

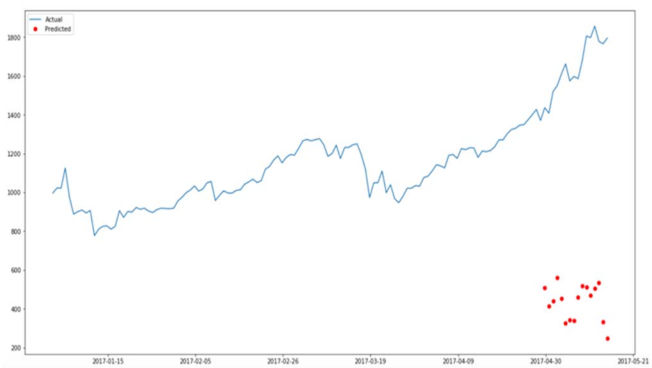

Fig. 8. CNN Results Using Gold Price Parameters

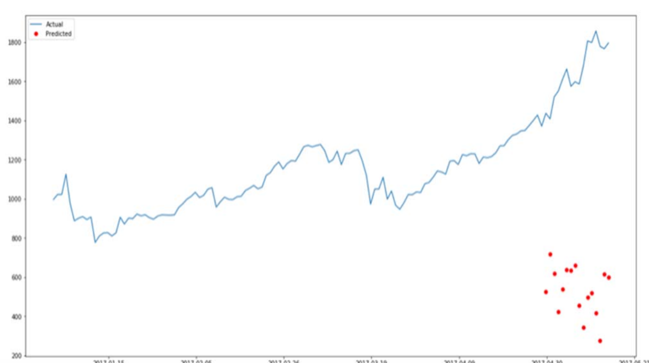

Fig. 9. LSTM Results Using Gold Price Parameters

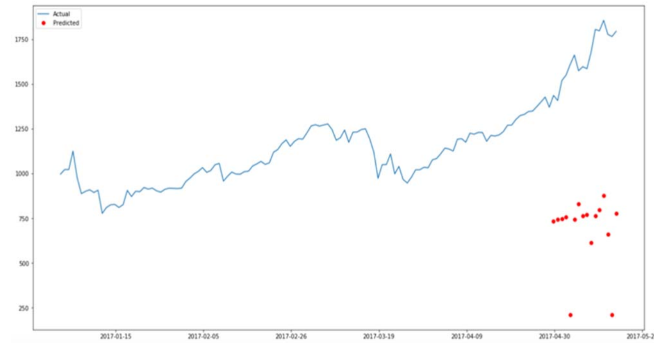

Fig. 10. GRU Results Using Gold Price Parameters

The accuracies of these models are tabulated in Table 2. The Root Mean Square Error (RMSE) values of all the three models namely LSTM, CNN and GRU are quite high which clearly states that there does not exist a positive correlation between gold price and bitcoin price.

Models	Number of Layers	RMSE
CNN	2	201.34
LSTM	4	151.67
GRU	1	179.23

Table 2. *Accuracy Results of Deep Learning Models Using Gold Price Parameters*

Fig. 11. LSTM Results Using Twitter Sentiment Analysis

Fig. 11 shows the relative movement of predicted price of bitcoin with changes in sentiment score calculated based on the tweets collected from twitter during the same time frame.

Negativity in tweets correlates with the drop in bitcoin price value while positivity in tweets brings a hike in bitcoin prices.

Model	Number of Layers	RMSE
LSTM	4	32.98

Table 3. *Accuracy Results of Deep Learning Model Using Twitter Sentiment Analysis*

V. CONCLUSION AND FUTURE WORK

In this paper, a comparative analysis of various deep learning models on various parameters affecting bitcoin price prediction has been done on the basis of RMSE values. The results show that the various deep learning models are evidently effective for the prediction of bitcoin price. LSTM results into the least RMSE value when bitcoin parameters are considered for price prediction. On the other hand, when gold price is used as the parameter for bitcoin price prediction, LSTM does not show a positive correlation between the gold price and bitcoin price. As a result, the graph of actual versus predicted values shows a huge deviation. The results of bitcoin price prediction using twitter sentimental analysis show a positive correlation and is shown when a positive tweet is posted in context of bitcoin, the prices are expected to hike and when a negative tweet is passed related to the bitcoin, the prices do get affected and are expected to fall depending on the popularity of the twitterer which can be measured using the number of followers of twitterer on twitter. In future, further research can be done to quantify the effect of twitterer influence on the bitcoin price. To improve the accuracy of results further, live dataset input streams of various parameters affecting bitcoin price can be used for price predictions of bitcoin.

REFERENCES

[1] Radityo, Arief, Qorib Munajat, and Indra Budi. "Prediction of Bitcoin exchange rate to American dollar using artificial neural network methods." In *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 433-438. IEEE, 2017.

[2] Madan, Isaac, Shaurya Saluja, and Aojia Zhao. "Automated bitcoin trading via machine learning algorithms." *URL: http://cs229. stanford. edu/proj2014/Isaac 20Madan 20 (2015).*

[3] Velankar, Siddhi, Sakshi Valecha, and Shreya Maji. "Bitcoin price prediction using machine learning." In *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pp. 144-147. IEEE, 2018.

[4] McNally, Sean, Jason Roche, and Simon Caton. "Predicting the price of Bitcoin using Machine Learning." In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, pp. 339-343. IEEE, 2018.

[5] Singh, Hrishikesh, and Parul Agarwal. "Empirical Analysis of Bitcoin Market Volatility Using Supervised Learning Approach." In *2018 Eleventh International Conference on Contemporary Computing (IC3)*, pp. 1-5. IEEE, 2018.

[6] Colianni, Stuart, Stephanie Rosales, and Michael Signorotti. "Algorithmic trading of cryptocurrency based on Twitter sentiment analysis." *CS229 Project* (2015): 1-5.

[7] Phillips, Ross C., and Denise Gorse. "Predicting cryptocurrency price bubbles using social media data and epidemic modelling." In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-7. IEEE, 2017.

[8] Jain, Arti, Shashank Tripathi, Harsh DharDwivedi, and Pranav Saxena. "Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis." In *2018 Eleventh International Conference on Contemporary Computing (IC3)*, pp. 1-7. IEEE, 2018.

[9] Kim, Young Bin & Gi Kim, Jun & Kim, Wook & Im, Jaeho & Kim, Tae Hyeong & Jin Kang, Shin & Kim, Chang-Hun. (2016). Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. PLOS ONE. e0161197. 10.1371/journal.pone.0161197.

[10] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168-177. ACM, 2004.