

Bitcoin Price Prediction

Mareena Fernandes

Student,

Fr. Conceicao Rodrigues College of

Engineering

(Affil. to University of Mumbai)

Mumbai, India

themareenafernandes@gmail.com

Saloni Khanna

Student,

Fr. Conceicao Rodrigues College of

Engineering

(Affil. to University of Mumbai)

Mumbai, India

salonikhanna9@gmail.com

Leandra Monteiro

Student,

Fr. Conceicao Rodrigues College of

Engineering

(Affil. to University of Mumbai)

Mumbai, India

monteiroleandra17@gmail.com

Anu Thomas

Student,

Fr. Conceicao Rodrigues College of Engineering

(Affil. to University of Mumbai)

Mumbai, India

anu@puthenpurackal.in

Garima Tripathi

Assistant Professor,

Fr. Conceicao Rodrigues College of Engineering

(Affil. to University of Mumbai)

Mumbai, India

garima@fragnel.edu.in

Abstract - Advancement in technological developments introduced virtual currency exchange methods viz Bitcoin, Litecoin, Ethereum and so on which are evolving rapidly. Cryptocurrencies were introduced to eliminate financial intermediaries leading to direct peer-to-peer transactions. With the spread of the global Coronavirus pandemic, the relationship between Bitcoin and the equity market has expanded. Cryptocurrencies are highly volatile but can also prove to be good investments. Cryptocurrency, being a novel technique for transaction systems, has led to a lot of confusion among investors and any rumours or news on social media has been claimed to significantly affect the prices of cryptocurrencies. The huge percentage increase/decrease in Bitcoin's price over a short period of time is an intriguing phenomenon that cannot be foreseen. For a long time, bitcoin price prediction has been a hot topic of study.

In this paper, we discuss the implementation and results of the Deep Learning Bitcoin Price Prediction Model and prepare a strategy to maximize gains for investors. The paper covers to framework with a set of deep learning models, analysis methods with a fixed set of factors to predict daily Bitcoin prices and design-integration of price prediction of different cryptocurrencies using RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory) and GRU (Gated recurrent units). The idea of incorporating Public Sentiment in the prediction of the hikes and falls of the Bitcoin market from Social Media platforms like Reddit and Twitter leading to meaningful predicted results. This prediction can bring confidence to the common man to invest with lesser risk and more profit. Also, this can enable the digital new-age currency to become a primary method of transaction.

Keywords – RNN, LSTM, GRU, sentiment analysis, lexicon model, bitcoin, cryptocurrency, MSE, MAE

I. INTRODUCTION

In 2008 an academic white paper titled 'Bitcoin: A Peer-to-Peer Electronic Cash System' was published in theory and the design was a system for a digital currency free of control from any organisation or government^[1]. In the year 2009, Bitcoin

was created as a way for people to send money over the internet. These digital currencies were intended to provide an alternative payment system that would operate free of central control but otherwise be used just like traditional currencies. Bitcoin among other cryptocurrencies is fast gaining popularity thanks in part to increasing endorsement by companies and pop culture icons alike. But all this fanfare has left the common man out of the picture and led some to blindly put their money expecting bumper returns. The objective is to make the working man confident enough to invest in this new-age currency.

The price of a cryptocurrency, in this case Bitcoin is directly related to market supply and demand.

These industry forces can change for several reasons, including public opinion, the press, and social media. Along with the technical and fundamental analysis, investigating the Bitcoin market sentiment can be a valuable addition to a trader's toolkit. Market sentiment is the collective pool of attitude of traders and investors towards a financial asset or market. Social-media sentiment does have the power to influence market cycles.

II. METHODOLOGY

A. System Architecture

To train the prediction model, we extracted historical Bitcoin APIs from Coinbase, CoinMarketCap and CoinDesk. For sentiment analysis, the social media platforms considered were Twitter and Reddit, and their datasets were extracted using Tweepy for Twitter and public library pushshift.io for Reddit. The predictions would be based on predefined time epochs paired with metrics based on key features like date and closing price, including only those which would yield the most optimum outcomes. Before implementing any model data needs to be prepared by normalizing and reshaping. For obtaining the polarities of the sentiments extracted a lexicon model was developed. Lexicon Model, works on the concept of 'bag of words', wherein a pre-defined polarity is assigned to a word and saved as a .csv file which is used to build the model. The sentiment data is then grouped by date and average polarity and merged^[7] with the historical bitcoin dataset, which is further divided into training^[4] and testing datasets^[8]. The

system is trained for three Deep Learning models viz RNN, LSTM and GRU.

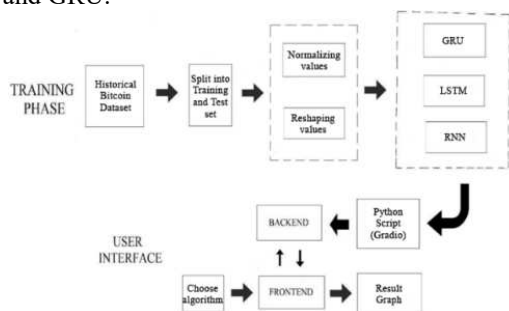


Fig. 2.1: System Architecture

B. Implemented Algorithms

We have studied and implemented 3 deep learning model for Real-time Bitcoin Price Prediction:

RNN - Recurrent Neural Networks are a class of Artificial Neural Networks that can process a sequence of inputs in deep learning and retain its state while processing the next sequence of inputs. Traditional neural networks will process input and move onto the next one disregarding its sequence.

LSTM - Long Short-Term Memory networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a skill that is required in a variety of complicated issue domains, including machine translation, speech recognition, and others. LSTMs are a complex area of deep learning^[5].

GRU - Gated Recurrent Unit aims to solve the vanishing gradient problem which comes with a standard recurrent neural network. GRU can also be considered as a variation on the LSTM because both are designed similarly and, in some cases, produce equally excellent results.

III. IMPLEMENTATION AND RESULTS

A. Sentiment Analysis

For extracting the polarities of the sentiments, a lexicon model was developed that decided sentiment value for data mapping. The prediction model training required the previous day's sentiment to predict the next day's prices. Training of the model was implemented using the 12L dataset from Kaggle, which was first passed through the lexicon model and then grouped by date and average sentiment score.

The grouping of the sentiments, the output resulted with just 79 rows which was not sufficient to train the model. The lexicon model is unable to detect sarcasm, which would result in invalid outputs. Hence, the sentiment factor in regulating bitcoin price was dropped.

B. Deep Learning Models

The following plotted graphs are for the test dataset.

RNN

Using 1 tanh layer, 1 sigmoid layer, 1 tanh layer, Dropout of 0.28 after layer, Units 20, Epoch 75

The UI is a website, developed using HTML, CSS, Javascript, AJAX, Python, and Flask. For prediction, Gradio is used which is an open-source library to create customizable UI components around TensorFlow or PyTorch models or even arbitrary Python functions. Components can be mixed and matched to support any combination of inputs and outputs.

For the prediction system embedded Python notebooks were used and then presented on a webpage. It automatically generates a public link we can share with colleagues that let users interact with the model from their computer remotely from their own devices. User can upload a dataset of cryptocurrency and generate a prediction graph by selecting which model needs to be implemented.

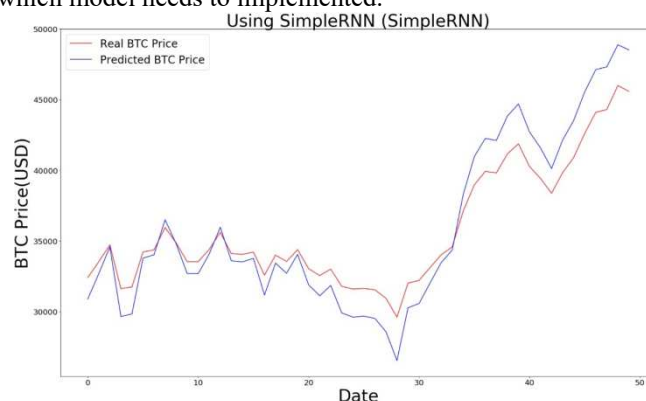


Fig.2.2: Bitcoin Price Prediction using RNN

	Predicted Price	Real Price
0	34447.550781	32404.33036
1	36327.257812	33532.25785
2	38243.718750	34718.47493
3	33122.703125	31622.08932
4	33321.933594	31739.29021
5	37436.117188	34213.01274
6	37680.757812	34365.18578

Fig. 2.3: Compared Bitcoin Price with RNN results

GRU

Using 1 tanh layer, 1 sigmoid layer, 1 tanh layer, Dropout of 0.28 after layer, Units 20, Epoch 75

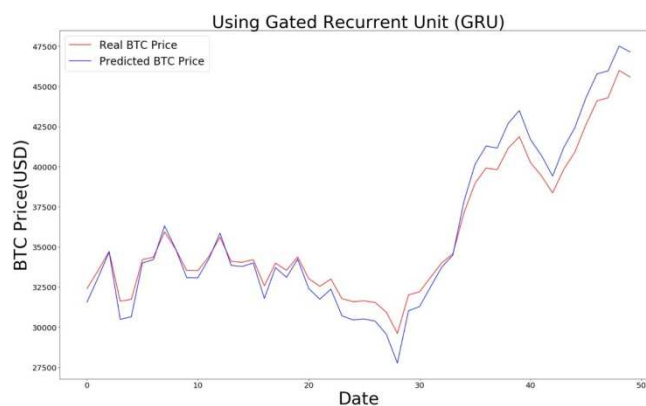


Fig. 2.4: Bitcoin Price Prediction using GRU

	Predicted Price	Real Price
0	31557.460938	32404.33036
1	33089.617188	33532.25785
2	34687.781250	34718.47493
3	30492.607422	31622.08932
4	30652.119141	31739.29021
5	34009.097656	34213.01274
6	34213.828125	34365.18578

Fig. 2.5: Compared Bitcoin Price with GRU results

LSTM

Using 1 tanh layer, 1 sigmoid layer, 1 tanh layer, Dropout of 0.28 after layer, Units 20, Epoch 75

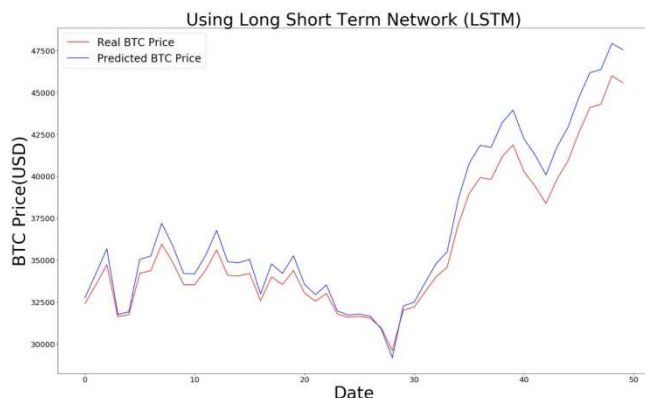


Fig. 2.6: Bitcoin Price Prediction using LSTM

	Predicted Price	Real Price
0	32755.900391	32404.33036
1	34188.058594	33532.25785
2	35677.218750	34718.47493
3	31756.667969	31622.08932
4	31906.591797	31739.29021
5	35045.230469	34213.01274
6	35235.925781	34365.18578

Fig. 2.7: Compared Bitcoin Price with LSTM results

After due comparison the 3 deep learning models and constant parameters LSTM was found to be the most efficient and reliable algorithm.

Deep Learning Model	MSE	MAE
RNN	52691476.890904665	1528.8687720375
GRU	44180507.654957004	861.0022885125001
LSTM	42147659.553459354	1016.5750596874999

Table 1: Compared MSE and MAE values

IV. CONCLUSION

Bitcoin is seen as the most commonly used digital currency, attracting investor's attention making its growth exponential. Although, the prediction of Bitcoin price is not an easy task because of high fluctuations in the Market.

Due to the data mapping issue faced while developing the sentiment analysis model, it had to be dropped and use only historical Bitcoin transactions data for building the model. The system can produce near accurate results leading to positive investments. We successfully implemented LSTM, GRU, RNN taking various test cases into consideration, for example increasing the number of activation function layers, and increasing the number of units. We also utilized a new library named Gradio with which we designed our own user interactive UI, which lets the user input the .csv file that could be implemented in any of the 3 models according to the users' preference. Did web scraping from popular Indian news sites and picked up news only related to the live cryptocurrency market.

Also, for prediction models, LSTM has the greatest accuracy as compared to other algorithms concerning Sentimental Analysis. Bitcoin prices have 6 times higher volatility than fiat currencies so the prediction requires a model suitable for handling long-term time series dependencies which again can be satisfied in LSTM. Time-series observation will be better when taken in a 30 minutes interval in the prediction modelling to fully decipher the workings of the cryptocurrency market.

V. FUTURE SCOPE

Artificial Intelligence and Cryptocurrency especially Bitcoin have come a long way in the last decade. But there is still a lot of work required to develop secure Bitcoin markets for the common man to trade. Bitcoin is open-source and implements a decentralized, peer-to-peer digital cash payment system; it does not require any trusted intermediaries to operate which can be extremely profitable if invested correctly.

Unlike the BSE, NYSE, SGX, or any stock market for that matter, there is no major body governing it. High volatility and major hike and fall in the market prices over a short period is an interesting phenomenon yet makes it difficult to predict. Inclusion of public sentiments from multiple social media platforms like Facebook, LinkedIn related to cryptocurrency and Bitcoin mainly can be included in a comprehensive collection of public opinion to understand the Bitcoin Price regulation. Normalizing Digital Currency for transactions, that is trade without a moderator will be revolutionizing.

These prediction models can be applied to other cryptocurrencies like Litecoin (LTC), Ethereum (ETH), Zcash (ZEC), and many more to further improve our understanding of cryptocurrencies. Implementing using Functional API to better the efficiency of the prediction process. Further implementation using combined configurations of two or more Deep Learning Models can be tested to obtain better results. Completely different models, such as neuro-fuzzy ANFIS predictive system have been also used in the time series field in the past, can be tested as well. Inclusion of more variables to estimate the price, such as the volume of trading operations done in a day and some statistics about them, such as the price balance (positive or negative), the total quantity of bitcoins exchanged and the number of trading events.

REFERENCES

- [1] D.L.K. Chuen, Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data, Academic Press, 2015.

- [2] Zheshi Chen, Chunhong Li, Wenjun Sun, Bitcoin price prediction using machine learning: An approach to sample dimension engineering, *Journal of Computational and Applied Mathematics*, Volume 365, 2020, 112395, ISSN 0377-0427, <https://doi.org/10.1016/j.cam.2019.112395>.
- [3] S. McNally, J. Roche and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning," 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), 2018, pp. 339-343, DOI: 10.1109/PDP2018.2018.00060.
- [4] Phaladisailoed, Thearasak&Numnonda, Thanisa. (2018). Machine Learning Models Comparison for Bitcoin Price Prediction. 506-511. 10.1109/ICITEED.2018.8534911.
- [5] C. Wu, C. Lu, Y. Ma, and R. Lu, "A New Forecasting Framework for Bitcoin Price with LSTM," 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 2018, pp. 168-175, DOI: 10.1109/ICDMW.2018.00032.
- [6] Patrick Jaquart, David Dann, Christof Weinhardt, Short-term bitcoin market prediction via machine learning, *The Journal of Finance and Data Science*, Volume 7, 2021, Pages 45-66, ISSN 2405-9188, <https://doi.org/10.1016/j.jfds.2021.03.001>.
- [7] Zaiqiao Meng, Richard McCreadie, Craig Macdonald, IadhOunis, "Exploring Data Splitting Strategies for the Evaluation of Recommendation Models", arXiv:2007.13237, 2020
- [8] Zahraa Said Abdallah, Lan Du, and Geoffrey I Webb, "Data Preparations", *Encyclopedia of Machine Learning and Data Mining*, Springer, 2017
- [9] Gökhan Aksu, CemOktayGüzeller and Mehmet Taha Eser, "The Effect of the Normalization Method Used in Different Sample Sizes on the Success of Artificial Neural Network Model", *International Journal of Assessment Tools in Education*, 2019, Vol. 6, No. 2, 170-192, DOI: 10.21449/ijate.479404
- [10] Yun Xu and Royston Goodacre, "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning", *Journal of Analysis and Testing* (2018), DOI: 10.1007/s41664-018-0068-2
- [11] D. E. Birba, "A Comparative study of data splitting algorithms for machine learning model selection", Dissertation, 2020.
- [12] ZulfadzliDrus, Haliyana Khalid, Sentiment Analysis in social media and Its Application: Systematic Literature Review, *Procedia Computer Science*, Volume 161, 2019, Pages 707-714, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2019.11.174>.
- [13] Stephan A. Curiskis, Barry Drake, Thomas R. Osborn, Paul J. Kennedy, An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit, *Information Processing & Management*, Volume 57, Issue 2, 2020, 102034, ISSN 0306-4573, <https://doi.org/10.1016/j.ipm.2019.04.002>.
- [14] Shrestha, Prasha and Sathanur, Arun and Maharjan, Suraj and Saldanha, Emily and Arendt, Dustin and Volkova, Svitlanam, Multiple social platforms reveal actionable signals for software vulnerability awareness: A study of GitHub, Twitter, and Reddit, *PLOS ONE*, 2020, 15, 3, e0230250, 2021/05/28, pub.1125865271, <https://doi.org/10.1371/journal.pone.0230250>.
- [15] Jing-Zhi Huang, William Huang, Jun Ni, "Predicting bitcoin returns using high-dimensional technical indicators", *The Journal of Finance and Data Science*, Volume 5, Issue 3, 2019, Pages 140-155, ISSN 2405-9188, <https://doi.org/10.1016/j.jfds.2018.10.001>.
- [16] H. B. Hosseinabad, E. F. Ersi, and A. Vahedian, "Detection of early signs of self-harm on Reddit using a multi-level machine," *CLEF*, 2020. <https://www.semanticscholar.org/paper/Detection-of-Early-Sign-of-Self-harm-on-Reddit-Hosseinabad-Ersi/97be314f3f84429318f48c80cb166c196bb9da56>
- [17] Singh, M. (2020, October 8). Automate Sentiment Analysis Process for Reddit Post: TextBlob and VADER. Medium. <https://towardsdatascience.com/automate-sentiment-analysis-process-for-reddit-post-textblob-and-vader-8a79c269522f>
- [18] Ashutosh Shankhdhar et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1131 012009 - Bitcoin Price Alert and Prediction System using various Models <https://iopscience.iop.org/article/10.1088/1757-899X/1131/1/012009>
- [19] M.FatihAkay, I.SibelKervancı, VOL. 3, NO. 3, DECEMBER 2020 DOI: 10.35377/saucis.03.03.774276 - Review on Bitcoin Price Prediction Using Machine Learning and Statistical Methods <http://saucis.sakarya.edu.tr/en/download/article-file/1217451>
- [20] Aygün, B., GünayKabakçı, E. (2021). Comparison of Statistical and Machine Learning Algorithms for Forecasting Daily Bitcoin Returns. *European Journal of Science and Technology*, (21), 444-454. <https://dergipark.org.tr/en/download/article-file/1383140>
- [21] Mrs. JayaBharathiChintalapati, P.P.S.Pratyusha, V.Uma Keerthi, K.Shiva Jatin4 Assoc. Professor, UG Scholar2, Vol 11, Issue 5, May/2020 ISSN NO:0377-9254 Bitcoin Price Prediction Using Time Series Analysis Observations Marked with a Timestamp
- [22] Hua, Yiqing.E3S Web of Conferences; Les Ulis, Vol. 218, (2020). DOI:10.1051/e3sconf/202021801050 Bitcoin price prediction using ARIMA and LSTM
- [23] A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," 2017 7th International Annual Engineering Seminar (InAES), 2017, pp. 1-4, DOI: 10.1109/INAES.2017.8068556.
- [24] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis," 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), 2019, pp. 205-208, DOI: 10.1109/BigDataService.2019.00035.
- [25] Huppmann, Sebastian. (2019). Bitcoin Price Prediction using Sentiment Analysis of Twitter Data. 10.13140/RG.2.2.32123.13605.
- [26] Yogesh Candra, Antoreep Jana, "Sentiment Analysis using Machine Learning and Deep Learning", 7th International Conference on Computing For Sustainable Global Development (INDIACom)
- [27] Wasiat Khan, Usman Malik, Mustansar Ali Ghazanfar, Muhammad Awais Azam, Khaled H. Alyoubi, Ahmed S. Alfakeeh "Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis", Springer-Verlag GmbH Germany, part of Springer Nature 2019, <https://doi.org/10.1007/s00500-019-04347-y>