

Prediction of dogecoin price using deep learning and social media trends

Basant Agarwal¹, Priyanka Harjule^{2,*}, Lakshit Chouhan¹, Upkar Saraswat¹, Himanshu Airan¹, Parth Agarwal¹

¹Department of Computer Science and Engineering, Indian Institute of Information Technology Kota (IIIT Kota)

²Department of Mathematics, Malaviya National Institute of Technology Jaipur (MNIT Jaipur)

Email: basant.cse@iiitkota.ac.in , priyanka.maths@mnit.ac.in , 2018kucp1041@iiitkota.ac.in , 2018kucp1049@iiitkota.ac.in , 2018kucp1141@iiitkota.ac.in , 2018kucp1050@iiitkota.ac.in

Abstract

INTRODUCTION: Cryptocurrency is a digital, decentralized form of money based on blockchain technology, which makes it the most secure method of making a transaction. There has been a huge increase in the number of cryptocurrencies in the past few years. Cryptocurrencies such as Bitcoin and Ethereum have become an interesting subject of study in fields such as finance. In 2021, over 4,000 cryptocurrencies are already listed. There are many past studies that focus on predicting the price of cryptocurrencies using machine learning, but the majority of them only focused on Bitcoin. Moreover, the majority of the models implemented for price prediction only used the historical market prices, and do not utilize social signals related to the cryptocurrency.

OBJECTIVES: In this paper, we propose a deep learning model for predicting the prices of dogecoin cryptocurrency. The proposed model is based on historical market price data as well as social trends of Dogecoin cryptocurrency.

METHODS: The market data of Dogecoin is collected from Kaggle on the granularity of a day and for the same duration the verified tweets have also been collected with hashtags “Dogecoin” and “Doge”. Experimental results show that the proposed model yields a promising prediction of future price of Dogecoin, a cryptocurrency that has recently become the talk of the town of the crypto market.

RESULTS: Minimum achieved RMSE in predicted price of Dogecoin was 0.02 where the feature vector consisted of OCVP (Open, Close, Volume, Polarity) values from combined dataset.

RESULTS: Experimental results show that the proposed approach performs efficiently.

Keywords: Dogecoin, Cryptocurrency, Deep Learning, Sentiment Analysis.

Received on 10 July 2021, accepted on 13 September 2021, published on 29 September 2021

Copyright © 2021 Basant Agarwal et al., licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.29-9-2021.171188

*Corresponding author. Email: priyanka.maths@mnit.ac.in

1. Introduction

Cryptocurrencies were introduced back in 2009 with the announcement of Bitcoin by Satoshi Nakamoto as the first peer-to-peer electronic medium of exchange. All cryptocurrencies are a major application of the blockchain which is merely a more secure method of creating applications, and are now highly accepted by virtue of its secure nature of transactions and decentralized form of operation [35]. While it was not so popular in 2009, the prices started skyrocketing from 2013 and in April, 2021 bitcoin reached an all-time high of almost \$65000. Inspired by bitcoin, other currencies such as Ethereum and Ripple have also gained popularity in the crypto market. At present, the currency has gained immense popularity due to its relevance with the meme material dog face, the Shiba, which is also the face of Dogecoin.

Since more and more people have started investing in crypto currencies, now it has become an interesting area of research to be able to predict future prices of cryptocurrencies with high accuracy. For over a decade, machine learning algorithms have gained popularity in predicting the stock market [18,20]. But with cryptocurrencies, the solvability of the problem reduces due to multiple factors such as economic and political conditions, as well as human behavior. Also, according to academic research, crypto currencies do not behave in a random manner but at the same time not in a linear, rather a dynamic way. Unlike the stock market, the prices of cryptocurrencies are much more volatile and difficult to predict [17-20].

Bitcoin has become the most used cryptocurrency, as asserted by the authors of [4,38]. At present, there are over 4000 listed cryptocurrencies. Bitcoin is not the only cryptocurrency that can offer a profitable trade. The second quarter of 2021 has proved to be a major growth period for Dogecoin [34]. Investors have gained huge sums of money and this along with events such as famous celebrities showing their enormous support for Dogecoin, have resulted in a huge increase in the current price of the crypto currency and the volume that is being traded each day.

In the case of Dogecoin though, very few or no significant studies are available in the literature yet [31,32]. This, along with the fact that Dogecoin has witnessed huge growth in terms of volume and value, has motivated the authors to focus their work on Dogecoin. In this paper a working algorithm that can predict the price of Dogecoin developed. The development of such a model can help in making a profitable trade in Dogecoin. Following are the significant motivating factors for this study:

1. Knowledge of price movements by a fraction of a second can lead to high profit investors can make, which makes this study a major motivation for the researcher.
2. It helps determine the intrinsic value of a company including tangible and intangible factors.
3. The results can be used for comparison with currencies' market value and finding out whether the currency is undervalued on the crypto market or not.

Neural networks, Support Vector Machines and Random Forests are few of the most widely used techniques used for stock market prediction [21,26]. On the other hand, deep learning approaches derived from Neural Networks have been used to predict prices of different crypto currencies like Bitcoin and Ethereum. A study conducted in [27, 39,40] shows that Random Forests gave the least error among other models when the model was trained with ten technical parameters.

Apart from the publicly available market data, social media has been noticed to have a strong influence on cryptocurrency prices. Social media platforms such as Twitter have played a vital role in this domain [3,22-25]. In this regard, Dogecoin prices were also affected due to recent tweets from events that took place on Twitter [36,41]. Thereby, it is a novel approach to utilize social signals and sentiment analysis for the prediction of prices of cryptocurrencies [1,2,29,33].

The number of machine learning and deep learning models implemented to predict prices of the stock market still outnumber those of the crypto market [28]. With the increasing growth in popularity of cryptocurrencies, it becomes absolutely essential to have a model that can predict the crypto market with an accuracy that is at par with that of the stock market.

Majority of the existing algorithms have been developed for predicting the prices of cryptocurrencies Bitcoin, Ethereum and Ripple. These outnumber the studies and research that has been made towards Dogecoin. On the contrary, while the algorithms made for the aforementioned cryptocurrencies have attained good accuracy, in case of the latter, algorithms that exist have considerable error and henceforth, there arises a need for the improvement of these models or the development of a new one.

While these models use different approaches and neural networks such as RNN [4], Bayesian neural networks (BNN) [30], LSTM and Gated Recurrent Units (GRU) [9], models such as [4] and [6], predict the future price of bitcoin using only attributes from historical market data. In [4], Open, High, Low, Close (OHLC) values were used through the implementation of a Bayesian optimized RNN and LSTM network. A highest classification accuracy of 52% and RMSE of 8% was achieved. Majority of these papers work using some or the other attributes available in the historical market data of a cryptocurrency. A few studies [37] use social media trends from Twitter but the area of study is Bitcoin. Henceforth, it is worth mentioning that at the time of writing this paper, no model exists that combines

attributes from market data and social media trends for Dogecoin and uses them for price prediction.

This paper aims at predicting real time prices of Dogecoin using market and twitter sentiment data and analyzing it for further comparative studies. Further, this paper aims to design a model that can consistently predict the price of Dogecoin.

The proposed model is based on feeding the LSTM network with a feature vector that not only comprises a set of values from the historical market data, but also contains useful information from sentiment analysis of social media data, most of which originates from twitter. This useful information from sentiment analysis can be the polarity or subjectivity of the tweets.

In the proposed model, initially, the historical market data is collected from Kaggle and tweets regarding Dogecoin are collected using Twint. After cleaning of market data, sentiment analysis is performed on tweets using Textblob. The two datasets are then combined over date and null values in the dataset are removed. The feature vector using the information from previous market history and sentiment analysis of tweets related to dogecoin is prepared, and finally this feature vector is fed to the LSTM network for future price prediction.

Predicting the exact price can be very difficult, thereby the key focus of this paper is being able to achieve the least error and the greatest accuracy. The objective is to develop a model that leads to Dogecoin price prediction accuracy by incorporating public sentiments from social media.

This paper is divided into six sections. Following the introduction section, the second section presents a literature review of the various related works that have already been carried out in this domain. The third section introduces the proposed methodology along with the problem statement. The fourth section provides a detailed explanation of the materials, methods and the experimental settings used, followed by the results obtained in the fifth section. Further, the conclusion and future scope is presented followed by the research papers and articles used as references for the study purpose of this paper.

2. Related Work

As discussed in the previous section, several studies [7-16] have been conducted on the price prediction of cryptocurrencies like Bitcoin and Ethereum. In [5], a feature vector was constructed by taking ten values, five each from market data and twitter sentiment data. This

feature vector was then fed into three different models, namely Multi-Layer Perceptron (MLP), Support Vector Machines (SVM) and Random Forests. The results obtained were compared and a conclusion was made that SVM was the best performing model. In another study [6], data was retrieved from Kaggle in USD. After preprocessing and splitting of the data into training set and testing set, an RNN model was used for the computation of price prediction. Another study [30] employed Bayesian neural networks (BNN). Blockchain data and macroeconomic variables were used in this for predicting the Bitcoin prices to get a rmse of 0.23.

Most of the research done earlier did not include all the parameters affecting the price of Dogecoin. Some researchers [6] considered only market data while predicting the prices which resulted in wrong predictions when a change in sentiment caused a sudden change in the price of Dogecoin. On the other hand, certain researches focused solely on the sentiments of people [5] and made it the base of their research in predicting prices. Few studies [9] considered only opening and closing prices of each day to predict the prices. This research lacked the day-to-day changes occurring in the markets which the sentiments were unable to depict. They also lacked the coordination between all these factors to obtain a model that combined all these factors to obtain the result. To overcome these loopholes this study combined the market price and sentiment component instead of focusing on the two aspects separately in a single feature vector to predict the price of dogecoin. A significant decrease in the rmse value in this model can be observed as compared to the earlier models. Table 1 summarizes the various deep learning models employed in predicting the prices of various cryptocurrencies.

Dogecoin is an emerging cryptocurrency, it is important to put an emphasis on the fact that the existing crypto-trading and cryptocurrency price prediction algorithms only work for cryptocurrencies like Bitcoin and Ethereum. Due to the increasing trading volume of Dogecoin in 2021, there arises a need for development of an algorithm for efficient and accurate prediction of its prices. Several magazines such as The Express UK [42] have mentioned and discussed the importance and scope of dogecoin[43-45]. The methodologies discussed in table 1 summarize several different models for Bitcoin. As such no scientific algorithm exists in the literature for predicting the price of dogecoin. Several events across the globe play an important role in deciding the uprise or downfall of a cryptocurrency. Examples of such events include banning of a cryptocurrency by a nation or on the contrary, the adoption of a cryptocurrency by a nation as its primary currency or medium of exchange. This gives a hint towards working on including these events in the model somehow. In this paper we use a hybrid model of existing methodologies to solve the problem of price prediction of Dogecoin.

Table 1. Comparison of existing work

Author	Model Used	Cryptocurrency	Dataset	Error of the paper	Remarks
[4]	Long short-term memory network (LSTM)	Bitcoin	Bitstamp, Bitnex, Coinbase, OkCoin and itBit	6.87	Attributes used in LSTM are open,high,low,close from market data but the error increased because of the increase in the number of activation function
[4]	Recurrent Neural Network (RNN)	Bitcoin	Bitstamp, Bitnex, Coinbase, OkCoin and itBit	5.45	Attributes used in RNN are same(open,high,low,close) but less error achieved due to less activation function than LSTM
[6]	Use of RNN which use LSTM algorithms	Bitcoin	Kaggle https://www.kaggle.com/	3.38	Only features from historical market data were used(total volume , time)
[9]	Gated Recurrent Unit (GRU)	Bitcoin	Kaggle https://www.kaggle.com/	0.069 (RMSE)	Forecasting here is based on the market data of bitcoin (open,high,low,close)
[9]	Long short-term memory network (LSTM)	Bitcoin	Kaggle https://www.kaggle.com/	0.0728 (RMSE)	Forecasting is based on open, high, low, close parameters but LSTM takes greater compilation time than GRU
[30]	Bayesian neural networks (BNN)	Bitcoin	Bitcoincharts https://bitcoincharts.com/markets/	0.23	Blockchain data and macroeconomic variables are used in predicting the Bitcoin prices

3. Proposed Approach

The proposed model is developed for predicting the dogecoin cryptocurrency prices based on the historical market data and public opinion towards dogecoin over the social media. Figure 1 depicts the proposed approach. The proposed model works as follows. Initially, we collect the historical and social trend data of the dogecoin cryptocurrency for the same time duration in a granularity of a day. Further, important key attributes such as retweet count, polarity and subjectivity from Twitter using sentiment analysis are extracted, and in parallel, the historical market price values such as market close, open, high and low prices are considered. Further,

the information from both kinds of data is extracted and a feature vector is formed which has the important information for predicting the dogecoin market price. The feature vector is further used to develop different deep neural networks-based models such as long short-term memory (LSTM), and Gated Recurrent Units (GRU). We explain each step of the proposed approach in detail in the subsequent sections.

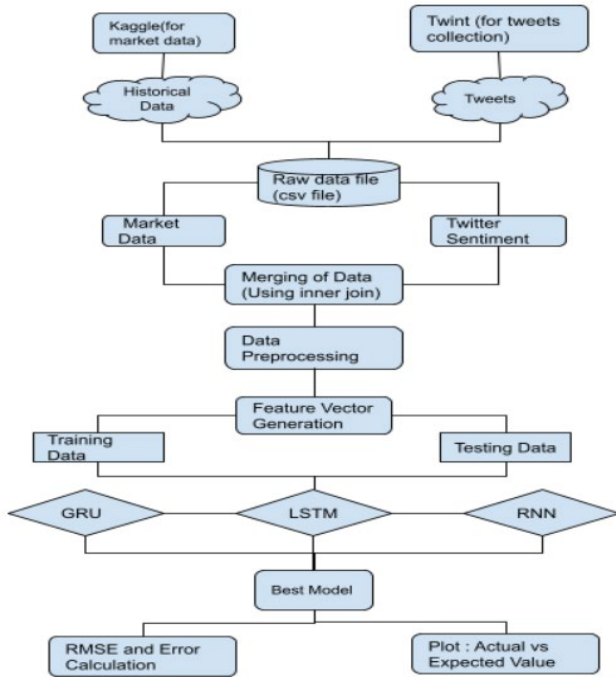


Fig. 1. Overview of the proposed approach for predicting the dogecoin market price

3.1 Preprocessing

Firstly, the collected data is preprocessed in order to clean it for further processing. The tweets dataset has inconsistencies, for some dates it has multiple entries/tweets while some dates are missing (no tweet on that day). These inconsistencies are removed. The previous day’s sentiment value is considered for that day and for the dates with multiple entries, the mean of the polarity scores was taken for each day.

3.2 Feature extraction

The price of dogecoin cryptocurrency depends on a number of factors. Different factors affect the price of dogecoin differently in magnitude. The following factors are key factors that may affect the change in price of dogecoin. The parameters included from market data covers the change in price due to statistical changes in the market while tweets polarity covers the change in price due to the sentimental changes. The information related to the social media trends are extracted using sentiment analysis carried out on the twitter dataset using textblob, whereas, the historical data related features on the corresponding days are extracted from the data collected from the Kaggle.

The feature vector has the following attributes:

Open price: The opening price of the Dogecoin on the day of observation. Open price means the first price for which a cryptocurrency is traded during a day.

Close price: The closing price of the Dogecoin on the day of observation. Close price means the official last price for which the last trade of cryptocurrency was made.

Low price: The lowest price of Dogecoin on the day of observation.

High price: The highest price of Dogecoin on the day of observation.

Volume: Total volume of cryptocurrency traded on the day of observation.

Polarity: After performing the sentiment analysis on tweets, polarity is obtained. The polarity obtained is assigned values based on the sentiment (<0 for negative, 0 for neutral and >0 for positive).

Before dividing the data into training and testing sets, these values are reshaped into the value range 0 to 1.

Table 2. Feature set for the deep learning model

close	closing price in the time period
high	highest price in the time period
low	lowest price in the time period
open	opening price in the time period
vol	trading volume for the time period
pol	mean polarity of each day

3.3 Deep learning model

After the feature vector is formed, it is fed into a deep neural network for learning the pattern. The LSTM is used to train the model. The Long Short-Term Memory (LSTM) network is a variant of recurrent neural networks, which makes it easier to remember past data in memory [6]. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. LSTM units include a 'memory cell' that can maintain information in memory for long periods of time. A set of gates is used to control when information enters the memory, when it's output, and when it's forgotten. Long Short-Term Memory (LSTM) networks are capable of learning order dependence in sequence prediction problems hence can learn the order dependence between items in a sequence. LSTMs have the promise of being able to learn the context required to make predictions in time series forecasting problems, rather than having this context pre-specified and fixed.

In this paper, we focus on the Dogecoin closing price for the development of the predictive model. The increase or decrease in the Dogecoin price with the higher volatility makes it harder to predict, but the LSTM model is well suited to predict Dogecoin prices. The proposed models are evaluated on the basis of root mean square error.

4. Dataset and Experimental Setting

4.1. Dataset collection

The market data is collected from Kaggle, Dogecoin market data is available at the granularity of a day. In parallel to the market data, verified tweets containing the hashtags “dogecoin”, “doge” are collected. Both datasets are collected over a period of Jan 2016 to May 2021.

Historical market data was obtained from the top performing cryptocurrency exchanges. The data was fetched from the kaggle dataset from source [2]. The data obtained had a daily granularity and contained the closing

price, opening price, lowest price, highest price and the transactional volume traded for the day. Time duration of historical market data used: January 1, 2016 to May 15, 2021.

Further, all tweets featuring the hashtags "Dogecoin", "Doge", "Cryptocurrency" were collected using the Twint library. The selected tweets had the following attributes:

1. --verified: The verified attribute was used to eliminate dump and meaningless tweets.
2. --en: The tweets selected were in the English language.
3. --since "2016-01-01 00:00:00": To fetch the tweets from January 1, 2016 to May 1, 2021.

Total number of samples for the historical market data were 1958, and the number of twitter dataset were 22147 for the same duration. The multiple tweets of the same day are combined as described below. Further, the data for training is divided into 70 percent and 30 percent for testing, so the training dataset consists of 1372 samples and the test dataset contains 588 samples. Few sample tweets are shown in Table 3 in order to give an idea about how the opinion expressed over twitter social media is about dogecoin.

Table 3. Sample Tweets related to the Dogecoin cryptocurrency

Timestamp	Tweet	Polarity
2021-05-10 12:09:45	Elon Musk's SpaceX to Launch Dogecoin-Funded Mission To The Moon #DOGE1 @elonmusk @SpaceX	~0 (Since the tweet does not explicitly term that can yield positive sentiment, the polarity averages out to close to 0)
2021-02-14 12:30:06	The cryptocurrency dogecoin was created for laughs, serves no purpose, and is now worth more than Western Union, Under Armour and Xerox Holdings	0 (Since the tweet suggests a strong negative meaning in the first half and a strong positive meaning in the second, the overall polarity the tweet averages out to 0)
2021-05-09 22:27:46	Elon Musk Calls Dogecoin 'The Future of Currency' on 'SNL' Appearance & ' predicts it will 'take over the world #doge #ToTheMoon	1 (The terminology used here explicitly proposes a positive meaning and thereby yields a polarity score of 1)
2017-04-10 19:31:05	CoinDesk examines the darker side of dogecoin, chronicling the scams that almost killed the cute, iconic 'doge'	-1 (Terms such as darker, scams and killed indicate a clear negative impression and thereby the polarity of the tweet averages out to -1)
2021-04-03 08:02:15	@jaymichaels13 @ItsDogeCoin @DogecoinRise I agree with you. ❤️ \$doge people are great people to be around.	0.8 (Agree indicates agreement, great proposes positive meaning along with heart emoji, hence the polarity is positive)
2021-04-02 13:31:15	@SusanLiTV @DaytonaNissan @SlimJim Is that the devil (I mean \$doge) on my shoulder? 😊 I like it! He whispers all kinds of things in my ear, especially about taking over the world! 🌍😍👍	-0.195 (Terms such as devil and whisper indicate a downside and positive emojis counteract against it, yet the score comes out to be negative)

4.2 Data Visualization

In order to better understand the dataset, we visualize different attributes such as open, close and volume of dogecoin prices with respect to the time.

- In Fig 2, the graph depicts the changes in closing price of dogecoin versus time over a duration of January 1, 2016 to May 15, 2021. The graph makes it clearly evident that till the later part of 2020, the close price did not show a significant rise. On the other hand, another observation can be made that the close price has increased by over 800% in 2021 alone.

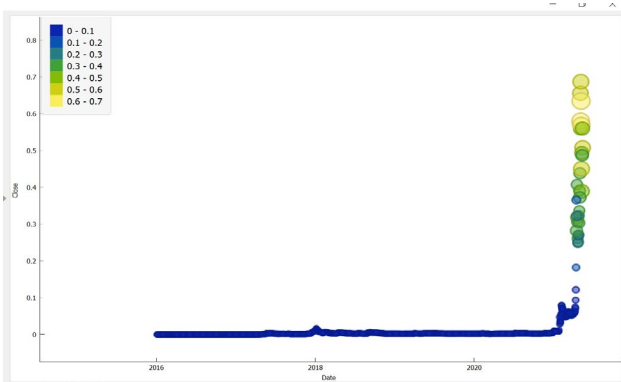


Fig. 2. Data Visualization of closing price of Dogecoin

- In Fig 3, the graph depicts the changes in closing price of dogecoin versus time. The graph makes it clearly evident that the high price of dogecoin shows a similar plot as the close price. One reason for this is that it is the first time that dogecoin has reached an all-time high.

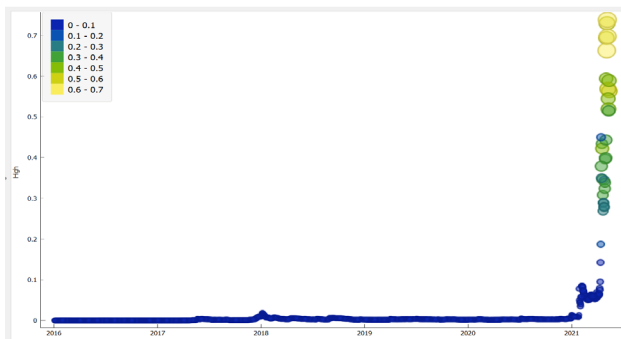


Fig. 3. Data Visualization of high price of Dogecoin

- Fig 4 depicts the volume of dogecoin that was being traded over a period of time. It was steady till the early part of 2021 and then we saw a rapid increase in the volume traded.

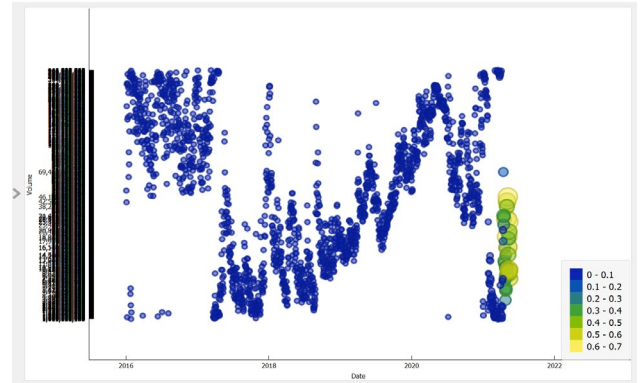


Fig. 4. Data Visualization of Volume of trade

4.3 Evaluation measure

Time series usually focus on predicting real values, which are called regression problems. Hence, the performance measurements in this paper focus on real-value prediction evaluation methods. The most commonly used method to measure error in time series models is root-mean squared error (RMSE) [11]. The root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample values and predicted values of a model).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

4.4. Hyperparameter tuning

We have performed hyperparameter tuning to optimize the parameters for the proposed models. Firstly, eight different optimizers such as Adadelta, Adagrad, Adam, Adamax, Ftrl, Nadam, Rmsprop, and Sgd were compared against one another whilst keeping the values of other hyperparameters constant. Fig 5 demonstrates the comparison of RMSE values obtained using different optimizers. The best performing optimizers were reported to be adam and nadam, whilst the worst performing were sgd and adadelta. Adam was chosen as the appropriate optimizer due to its lower degree of reported errors when compared to nadam, adadelta, ftrl and adagrad. Adam was chosen as the best optimizer; a similar method was adopted to obtain the other hyperparameters such as the epochs value and loss function.

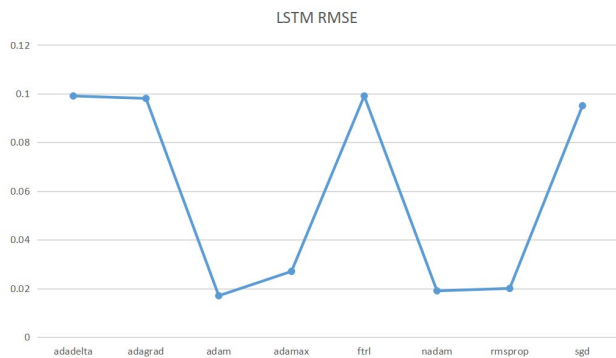


Fig. 5. Comparison of RMSE values whilst using different optimizers

Fig 6 demonstrates the comparison between RMSE values obtained by changing epochs values while keeping other hyperparameters constant and using adam optimizer.

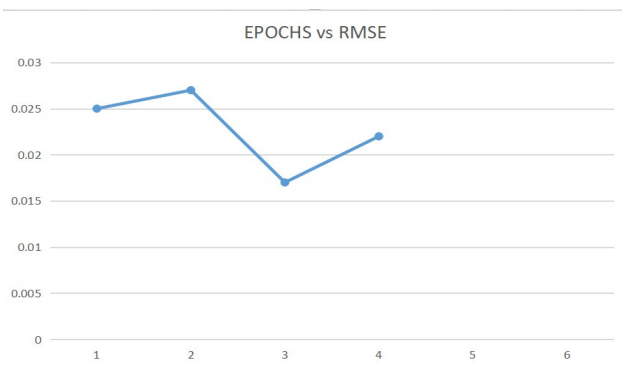


Fig. 6. Comparison of RMSE values whilst using best performing optimizer (adam) and different epochs values

Fig 5, Fig 6 and Fig 7 demonstrate the comparison between the different possible hyperparameter values optimizers, number of epochs and loss function. Further, the loss functions namely Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Squared Logarithmic Error (MSLE) were compared against each other. Upon comparison, it was observed that mean absolute error proved to be the ideal loss

function as shown in Figure 7.

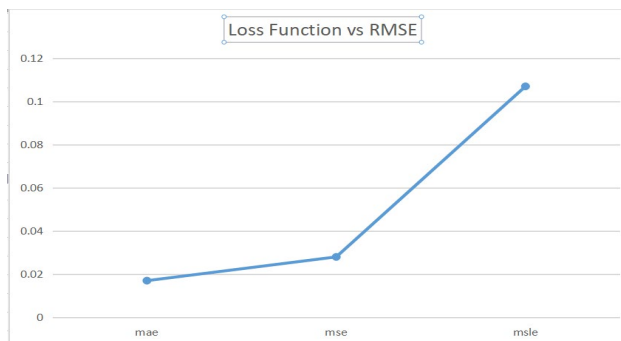


Fig. 7. Comparison of RMSE values whilst using different loss functions, best performing optimizer and epochs value.

Table 4. Hyperparameters with their corresponding chosen values

Hyperparameter	Name/Value of Hyperparameter chosen
Optimizer	Adam
Learning rate	0.01
Epochs	300
Batch size	100
Loss Function	MAE
Look back	1

5. Results and Discussion

The results obtained with different combinations of the proposed models varying different deep neural network models are presented and described in the following cases.

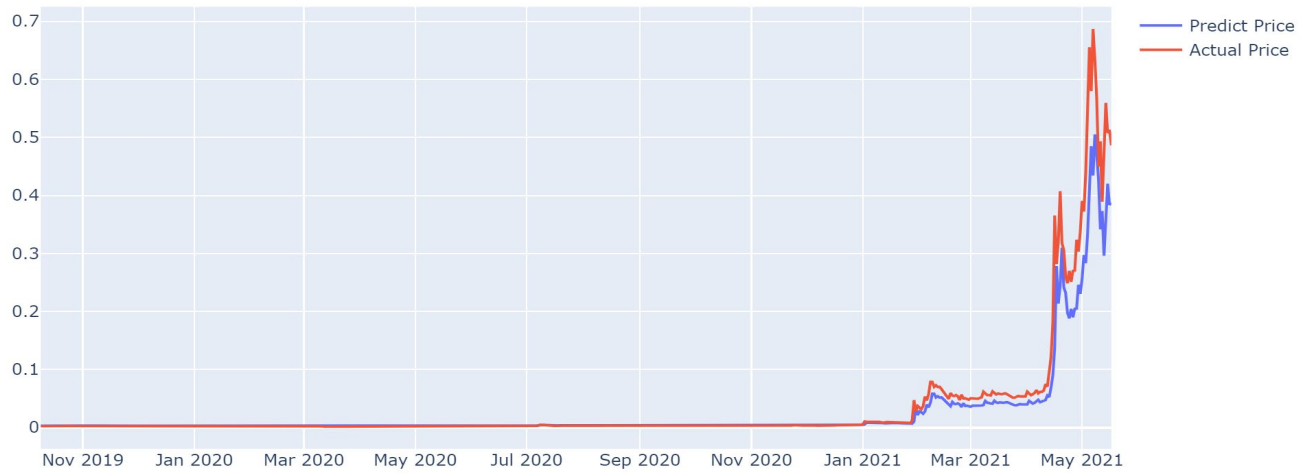


Fig. 8. Price prediction graph when only historical market data is used

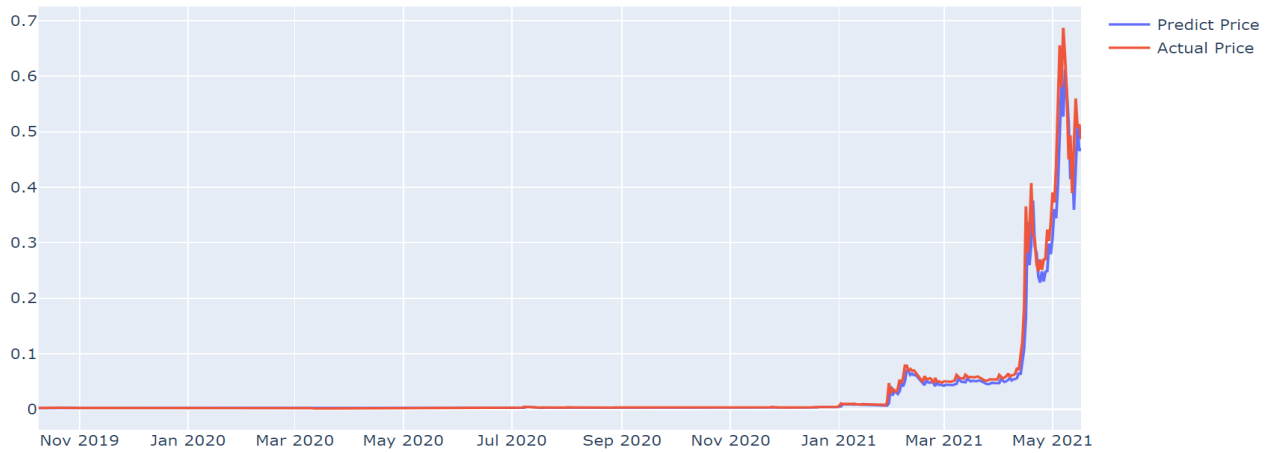


Fig. 9. Price prediction graph when complete feature vector (as discussed above) is used

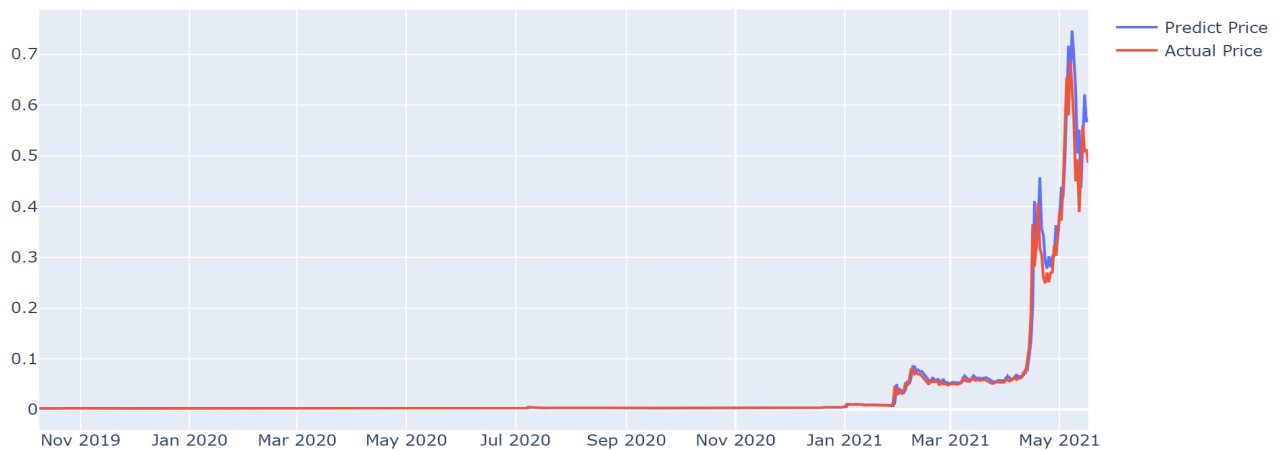


Fig. 10. Price prediction graph when two features (High & Low) are removed from feature vector

Case 1. Analysis using historical Market Data HIST_LSTM model

In this case, only the historical data related features are considered, and we discard the sentiment analysis related features. Figure 8 depicts the price prediction graph when the traditional market data was fed to the LSTM without adding any twitter sentiment data. RMSE value of 0.03 was achieved and the approximate percentage error calculated was 31.3 .

Case 2. Analysis using historical market data (with all features) and twitter sentiment SENTI_HISTO_LSTM MODEL

Fig 9. depicts the price prediction graph when the feature vector composed of all values from market data (open, close, high, low prices and volume traded) were used along with polarity from processed twitter data. RMSE value of 0.027 was achieved and the approximate percentage error

calculated was 20.8%. There is a slight improvement over the conventional “only market data” feature vector.

Case 3. Analysis using historical market data (without High-Low price values) and twitter sentiment

Fig. 10 depicts the price prediction graph when the feature vector consists of open, close prices and volume traded from market data and polarity from processed twitter data. This feature vector yielded an RMSE value of 0.020 and calculated percentage error was approximately 15%. Upon detailed analysis, the conclusion drawn about the reason for the improvement of this model over that in Case 2 was that the high and low prices of a day show heavy deviation over closing price in case of Dogecoin in the recent time.

Table 5 shows the results of various combinations of the proposed models with varying features in order to understand the main reasons which contribute to the prediction of the dogecoin prices.

Table 5. Models and their symbolic representation

Model Used	RMSE
LSTM using historical market data (histo_lstm)	0.024
GRU using historical market data (histo_gru)	0.020
LSTM using filtered historical market data (open, close and volume)- (filt_histo_lstm)	0.017
GRU using filtered historical market data (open, close and volume) (filt_histo_gru)	0.022
LSTM using all features from historical market data and polarity from twitter sentiment (lstm_complete)	0.027
GRU using all features from historical market data and polarity from twitter sentiment (gru_complete)	0.091
LSTM using historical market data (without high and low prices) and polarity from twitter sentiment (filt_lstm_complete)	0.020
GRU using historical market data (without high and low prices) and polarity from twitter sentiment (Filt_gru_complete)	0.159

A detailed comparison with other papers with similar proposed methodologies on other cryptocurrencies like bitcoin and ethereum cannot be ideally done because of a distinction between sources of data. Another reason for the same is that the witnessed uprise in Dogecoin is relatively recent, thereby showing a lot of fluctuations in prices as

compared to Bitcoin and Ethereum.

6. Discussion

Papers such as [4], [6] and [9] discussed in section II have shown extensive levels of performance for Bitcoin. Since the focus of this

paper is on price prediction of Dogecoin, the results discussed previously in section II are compared with the results yielded by the proposed model in this paper. Proposed model was tested over three cases:

1. Feature vector with only values from market data.
2. Feature vector with all values as discussed in section
3. Feature vector with open price, close price, volume traded and polarity from sentiment analysis data.

The proposed model gave the least error in case 3. The minimum achieved RMSE over all test runs was 0.020. It can be very well observed that in the very low-price range in which Dogecoin is traded as of May 2021, a reduction in the RMSE by 0.01 from case 1 to case 3 is remarkable. This clearly indicates that using social signals from twitter have increased the prediction accuracy of the proposed model. The result obtained in case 3 has been compared with several similar models in Table 5 . Fig 11 also compares the error of the proposed model with other similar models in graphical format.



Fig. 11. Comparison between best performing LSTM and GRU models.

7. Conclusion and Future Work

The aim of this paper is to develop a deep neural network-based model for predicting dogecoin prices by combining the market data of Dogecoin and sentiments extracted from the tweets. The developed model using LSTM and sentiment analysis is more accurate than the already existing traditional models. LSTM has been used because of its capability to recognize long-term dependencies. Two deep learning models namely LSTM and GRU were used in different combinations. Experimental results show that the error is minimum when historical market data (without high and low prices) and polarity from twitter sentiment are used in the feature vector employed in GRU model. Therefore, it can be concluded that this approach of combining the historical market data of dogecoin prices with that of the

sentiment polarity is more efficient in predicting the dogecoin prices. The authors wish to extend the scope of this work in future by expanding the feature vector by including a new feature vector with news sentiment analysis that can prove to be promising in terms of higher accuracy and lower error.

References

- [1] Agarwal B. , Mittal N. (2016) Prominent feature extraction for review analysis: an empirical study. *Journal of Experimental & Theoretical Artificial Intelligence* **28**(3), pp: 485-498. 2016.
- [2] Agarwal B. , Mittal N. , Bansal P. , Garg S. (2015) "Sentiment analysis using common-sense and context information", *Computational Intelligence and Neuroscience*, Article ID 715730, 9 pages.
- [3] Pant D. , Neupane P. , Poudel A., Pokhrel A. Kumar, Lama B. Kumar (2018), "Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis". In Proceedings of 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), Kathmandu, Nepal, 25-27 October, 2018 DOI : 10.1109/CCCS.2018.8586824
- [4] McNally S., Roche J. , Simon C. (2018) "Predicting the Price of Bitcoin Using Machine Learning" In Proceedings of 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP) Cambridge, UK, 21-23 March, 2018 339343. 10.1109/PDP2018.2018.00060.
- [5] Valencia F, Gómez-Espinosa A, Valdés-Aguirre B. (2019) "Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning". *Entropy*; 21(6):589. <https://doi.org/10.3390/e21060589>
- [6] Samiksha Marne, Delisa Correia, Shweta Churi, Joanne Gomes (2020), "Predicting Price of Cryptocurrency - A Deep Learning Approach", *NTASU - 9*(3)
- [7] Cocco, Luisanna & Tonelli, Roberto & Marchesi, Michele. (2021). Predictions of bitcoin prices through machine learning based frameworks. *PeerJ Computer Science*. 7(3), e413. 10.7717/peerj-cs.413.
- [8] Patrick J., David D., Christof W. (2021), Short-term bitcoin market prediction via machine learning, *The Journal of Finance and Data Science*, 7, Pages 45-66, ISSN 2405-9188.
- [9] Muniye, Temesgen & Rout, Minakhi & Mohanty, Lipika & Satapathy, Suresh. (2020). Bitcoin Price Prediction and Analysis Using Deep Learning Models. 10.1007/978-981-15-5397-4_63.
- [10] Reddy, Lekkala S. and Dr.P. Srirama (2020), "A Research on Bitcoin Price Prediction Using Machine Learning Algorithms." *International Journal of Scientific & Technology* , 9 : 1600-1604.
- [11] Pang, Yan & Sundararaj, Ganeshkumar & Ren, Jiewen. (2019). Cryptocurrency Price Prediction using Time Series and Social Sentiment Data. 35-41. 10.1145/3365109.3368785.
- [12] Onuodu E. ,Ufio N. (2020), "An Enhanced Cryptocurrency Prediction Model Using Sentiment Analysis and Machine Learning", *International Journal of Engineering Science Invention (IJESI)*, 9(4) Series III , pp 14-24
- [13] Mohammed S. and Abdullahi I. (2020), "Price Prediction of different cryptocurrencies using technical trade indicators and machine learning" *IOP Conf. Ser.: Mater. Sci. Eng.* 928 032007.
- [14] Vrbka, Jaromir & Rowland, Zuzana. (2017). Stock price development forecasting using neural networks. *SHS Web of Conferences*. 39(8). 01032. 10.1051/shsconf/20173901032.

- [15] Hochreiter, Jürgen S. (1997). "Long Short-term Memory. Neural computation", 9(8), 1735-80. 10.1162/neco.1997.9.8.1735.
- [16] F. A. Gers, J. Schmidhuber and F. Cummins (1999), "Learning to forget: continual prediction with LSTM," 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), pp. 850-855, 2, doi: 10.1049/cp:19991218.
- [17] Hiransha, M.; Gopalakrishnan, E.A.; Menon, V.K.; Soman (2018), K.P. NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Comput. Sci.* 132, 1351–1362.
- [18] Fang Z., George K. (2017) Application of machine learning: An analysis of Asian options pricing using neural networks. In *Proceedings of the 2017 IEEE 14th International Conference on e-Business Engineering*, Shanghai, China, 4–6 November ; pp. 142–149.
- [19] Fama E.F. (1965) "The Behavior of Stock-Market Prices". *J. Bus.* 38(1), pp. 34-105.
- [20] Kaizhu H., Haiqin Y., I. King and M. R. Lyu(2006), "Local Support Vector Regression for Financial Time Series Prediction," *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006, pp. 1622-1627, doi: 10.1109/IJCNN.2006.246628.
- [21] Bollen J., Mao H., Zeng X. (2010) Twitter mood predicts the stock market. *J. Comput. Sci.*, 2, 1–8.
- [22] Li Q., Wang T., Li P., Gong Q., Chen Y. (2014) "The effects of news and public mood on stock movements". *Inf. Sci.* 2014, 278, 826–840.
- [23] Bordino I., Battiston S., Caldarelli G., Cristelli M., Ukkonen A., Weber I. (2012) "Web search queries can predict stock market volumes". *PLoS ONE* 7(7), e40014.
- [24] Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. (2015) "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques". *Expert Syst. Appl.* 42(1), 259–268
- [25] Mai, F.; Shan, Z.; Bai, Q.; Wang, X.; Chiang, R. (2018) How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. *Manag. Inf. Syst.* 35(1), 19–52.
- [26] Kohara, K.; Ishikawa, T.; Fukuhara, Y.; Nakamura, Y. (1997) Stock price prediction using prior knowledge and neural networks. *Int. Syst. Account. Finance. Manag.* 6(1), 11–22.
- [27] Chen, R., Lazer, M.: (2011) Sentiment analysis of twitter feeds for the prediction of stock market movement. *Stanford Computer Science* 229
- [28] Pimprikar, R., Ramachadran, S., Senthilkumar, K.: (2017) Use of machine learning algorithms and twitter sentiment analysis for stock market prediction. *International Journal of Pure and Applied Mathematics* pp. 521-526
- [29] Bollen, J., Mao, H., Zeng, X.: (2011) Twitter mood predicts the stock market. *Journal of Computational Science* 2(1), 1-8.
- [30] H. Jang and J. Lee, (2018) An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information, in *IEEE Access*, 6, pp. 5427-5437
- [31] Kolla, Bhanu. (2020). Predicting Crypto Currency Prices Using Machine Learning and Deep Learning Techniques. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(4). 10.30534/ijatcse/2020/351942020.
- [32] Mehta P., Sasikala E. (2020) Prediction of Bitcoin using Recurrent Neural Network, *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, 8(6)
- [33] Agarwal B., Mittal N. (2016) Prominent Feature Extraction for Sentiment Analysis, In *Springer Book Series: Socio-Affective Computing series*, Springer International Publishing, ISBN: 978-3-319-25343-5, DOI: 10.1007/978-3-319-25343-5 .
- [34] <https://www.statista.com/statistics/1200235/dogecoin-price-index> Accessed on: May 15, 2021
- [35] URL: <https://www.investopedia.com/terms/c/cryptocurrency.asp>, Accessed on: May 11, 2021
- [36] Harjule P., Gurjar A., Seth H., Thakur P (2020). Text Classification on Twitter Data, 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, 2020, pp. 160-164, doi: 10.1109/ICETCE48199.2020.9091774.
- [37] Stenqvist, Evita and Jacob Lönnö. (2017) Predicting Bitcoin price fluctuation with Twitter sentiment analysis. p. 31, E53, Osquars backe 14, KTH Campus, Stockholm
- [38] Suhwan Ji, Jongmin Kim and Hyunseung Im (2019) A Comparative Study of Bitcoin Price Prediction Using Deep Learning, *Mathematics*, 7(10), 898; doi:10.3390/math7100898
- [39] Mohammed M., Shada B., Devrim U. and Mohammad H. (2020) Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach. *Neural Computing and Applications* doi:10.1007/s00521-020-05129-6 .
- [40] Sharmila Begum, G. Jayashree, Z. Mahaboob Asfia, R. Subhasri and L. Vishnu Priya (2021), Predicting The Prices of Bitcoin Using Data Analytics. *Turkish Journal of Computer and Mathematics Education*, 12(10), 1487-1501.
- [41] Harjule P., Sharma A., Chauhan S. and Joshi S (2020) Reliability of News, 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, pp. 165-170.
- [42] <https://www.express.co.uk/finance/city/1425009/how-high-will-dogecoin-go-prediction-expert-comments-EVG>, Accessed on: August 15, 2021.
- [43] Bhardwaj J. , Negi R. , Nagrath P. , & Mittal M. (2022). Applications of Blockchain in Various Domains . In *Smart IoT for Research and Industry* (pp. 1-30). Springer, Cham.
- [44] Aggarwal A., Mittal M. , & Battineni G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 100004.
- [45] Bhardwaj J., Jain P., Nagrath P., Mittal M. (2021). "File Authentication Ownership Using Blockchain". In *Book Integrated Intelligence Enable Networks and Computing* (pp. 43-52). Springer, Singapore.
- [46] Malhotra J., Sethi J. K., Mittal M. (2021). Analysis of big data using two mapper files in hadoop, *International Journal of Security and Privacy in Pervasive Computing* , 13(1), 69-77.