Received : 28th October Revised : 30th November 2021 Accepted: 15th December 2021

BITCOIN PRICE PREDICTION USING LSTM

Mrs.D. RagaVamsi[1], Parise Sri Tirumala[2], MandapatiYaswanth Reddy[3], PetetiHemanthVenkatesh[4], Motupalli Sarah Chaitanya[5]

1 M. Tech Assistant Professor, CSE Department & SR GEC, Gudlavalleru, Krishna
2Student, CSE Department &SR GEC, Gudlavalleru, Krishna
3Student, CSE Department &SR GEC, Gudlavalleru, Krishna
4Student, CSE Department &SR GEC, Gudlavalleru, Krishna
5Student, CSE Department &SR GEC, Gudlavalleru, Krishna

Abstract:

Bitcoin is a very volatile cryptocurrency that is becoming increasingly popular. It is a cryptographic protocol-based digital currency. It is an unregulated kind of electronic cash that can be used for online transactions or as a medium of exchange between any two parties and is purely electronic cash in a peer-to-peer format [1].Bitcoin is a particularly volatile currency since it has no central regulatory authority and is governed by the whole population. Its price is influenced by socially generated ideas. Hence Bitcoin's total market value is affected in some way, either directly or indirectly by Twitter attitude about Bitcoin. Cryptocurrency investors frequently consider which cryptocurrency to invest in.Cryptocurrency price prediction is critical for this purpose. The LSTM model can estimate the value of pricing data because it is time series data. For the best results, socially generated opinions (tweets data) are combined with bitcoin price data.

Introduction:

Cryptocurrency (CC)[2] can be defined as any medium of exchange, apart from real world money, that can be used in many financial transactions whether they are virtual or real transactions. Cryptocurrency is a digital representation of valuable and intangible assets that may be used in a wide range of applications and networks, including as social networking websites, social media games, virtual worlds, and peer-to-peer networks.There are 16,598 cryptocurrencies in circulation as of January 2022. The total trading volume of all cryptocurrencies is currently $75 billion per 24 hours. With the "dead" cryptos removed, there are roughly 9,631 active cryptocurrencies, which is a significant quantity. With so many cryptocurrencies accessible, one may wonder which one to invest in. Having an overview and understanding of bitcoin future values is really beneficial in this case. Bitcoin's price is forecasted using a variety of time series forecasting models.

Time series forecasting methods can be divided into two categories named parametric methods like AR[9],MA,ARMA,ARIMA models and non-parametric models such as nonparametric regression , neural network prediction, support vector machine (SVM) and the combination of these algorithms[3]. Parametric methods won't achieve good performance when time series data shows irregular variations[3].

Both bitcoin price data and tweet data about bitcoin come under time series data. Using the mood of users' tweets about bitcoin to predict the price can greatly improve the model's accuracy. The LSTM model is being used to forecast the bitcoin price in this case.

LSTM (Long Short Term Memory) model:LSTM network is a special kind of RNN. By treating the hidden layer as a memory unit,The LSTM network can deal with short and long-term correlation within time series, making it capable of learning order dependence in sequence prediction tasks. In LSTMs, information is transmitted using a mechanism known as cell states. In this way, LSTMs can consciously recall or forget things. There are three types ofreliance on information at a given cell state.

These dependencies can be used to solve any problem in the following way:

1. The cell's former condition (i.e. the information that was present in the memory after the previous time step)
2. The previously unknown state (i.e. this is the same as the output of the previous cell)
3. The input for the current time step (i.e. the new information that is being fed in at that moment)

Sentiment analysis:

Sentiment analysis is the appraisal of a speaker's, writer's, or other subject's opinion on a given topic. Twitter is being utilised to collect a big and diversified data set that represents current popular sentiment on bitcoin. To assess public attitudes, the gathered tweets are analysed using a lexicon-based technique. We calculate the polarity and subjectivity metrics for the gathered tweets in this study, which aid in understanding user opinion on bitcoin price. For example,FarhaNausheen and SayyadaHajera Begum [3] employed sentiment analysis to forecast what would happen in US presidential election results 2016, in which Donald Trump, Hillary Clinton and Bernie Sanders were among the top election candidates. The opinion of the public for a candidate will impact the potential leader of the country.

Literature Survey (related work):

A univariate time series is defined as a collection of measurements of the same variable through time. The measurements are usually taken at regular intervals. One of the distinguishing features of a time series is that it is a collection of observations in which the order is important. Because there is reliance, ordering is crucial, and changing the order could change the meaning of the data.

Various models, such as AR, MA, ARIMA, and Neural Network-based models, can be used to analyse time series data.

1.AR (Auto-Regressive) model:

An autoregressive (AR) model predicts future behaviour based on historical data. It is useful for anticipating whether there is an association between the values in a series data and the values that come before and after them. The procedure entails a linear regression of the current series' data against one or more previous values from the same series.

Thevalue of the outcome variable (Y) at time t is closely attributed to the predictor variable in an AR model just as it is in "ordinary" linear regression (X). The main difference between simple linear regression and AR models is that Y is dependent on X and prior Y values. It has the following representation:

$$Yt = \beta_1 * y_{-1} + \beta_2 * y_{\square-2} + \beta_3 * y_{\square-3} + ... + \beta_{\square} * y_{\square-\square}$$

Where Yt = forecasting variable at time period t

β_1 , β_2 ,.. are bias terms

2. MA(Moving Averages) model:

A moving average model(MA)[11] employs past forecast mistakes in a regression-like model, unlike the AR model, which uses past values of the predicted variable in a regression. The current value is linearly dependent on the current and previous error terms, according to this formula.

$$r_t = c + \theta_1 \epsilon_{t-1} + \epsilon_t$$

where r_t = forecasting variable at time t

c = constant factor

θ_1 =numeric coefficient for the value associated with the 1st lag.

ϵ_t = residual at time t

ARIMA(autoregressive integrated moving average) model:

ARIMA is a type of regression analysis that shows how strong a dependent variable is in relation to other variables that change. It's the result of combining AR and MA models.

The data must be stationary to apply all of the following models (AR,MA). The qualities of a

stationary time series are independent of the time at which it is viewed. Time series containing patterns or seasonality, on the other hand, are not stationary. When there is evidence of non-stationarity in the data, ARIMA models are used. In time series analysis, non-stationary data is always processed into stationary data by differencing the data.

The ARIMA model is abbreviated as ARIMA (p, d, q), and the parameters p, d, and q are as follows:

- p: the lag order of the autoregressive model AR, or the number of temporal lags (p)
- d: the degree of differencing, or the number of times the data has been subtracted from a previous value.
- q: what is the order of the MA model of moving averages? (q)

Neural network based RNN model:

The most up-to-date sequential data algorithm is recurrent neural networks (RNN). It is the first algorithm with an internal memory that remembers its input, making it ideal for machine learning problems involving sequential data. This is why they're the algorithm of choice for series data, speech, text, financial records, audio, video, climate, and a variety of other sequential data types. The formula for the current state in RNN is

$$h_t = f(h_{t-1},x_t)$$

$$h_t= tanh(W_{hh}\, h_{t-1} +W_{xh}\, X_{t-1})$$

where $h_t$is the output at time 't'

$W_{hh}$is the weightmatrix(kernel)associated with h_{t-1}

X_{t-1} is the input given at time t-1

Exploding gradients and vanishing gradients are two important challenges that RNNs have had to overcome. Long-term dependencies are impossible for RNNs to handle.

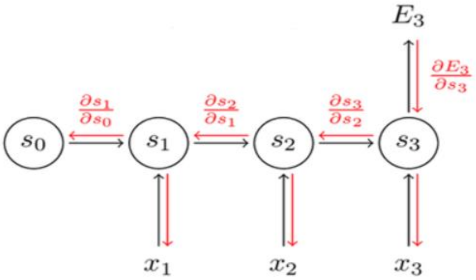

E_3
$\frac{\partial E_3}{\partial s_3}$
$\frac{\partial s_1}{\partial s_0}$
$\frac{\partial s_2}{\partial s_1}$
$\frac{\partial s_3}{\partial s_2}$
s_0
s_1
s_2
s_3
x_1
x_2
x_3

Figure 1. Backpropagation in RNN

Backpropagation in RNN is determined using the chain rule, which involves multiplying the derivatives of all prior state activation functions. As a result, if anyintermediate derivative is close to zero, the updated weight is zero, resulting in a vanishing gradient problem.

Dataset:

Time series data:

Data of a set of variables collected over time is called time series data. Then What constitutes a time series?

A simple additive dissected version will look like this, without digging into too much notation:

$$x_t = m_t + s_t + e_t$$

where:

m_t is the trend

s_t is the seasonality

$e_t$is the error or random white noise

To put it simply, the components of a time series model are defined as[12]:

The climbing or dropping data in a series are referred to as a trend.

Seasonality is defined as the series' repeated short-term cycles.

arbitrary white noise - a series of random variations

Python modules are used to retrieve historical bitcoin price data through the Coinbase Pro API. The data is downloaded as csv files and the date column is preprocessed. On the other hand, data from tweets on the bitcoin topic (keyword) is retrieved using Python's scraping packages, such as snscrape. It appears to be the columns below.

Date (Object)	Low (float)	High (float)	Open (float)	Close (float)	Vol ume (int)

Sentiment analysis:

All of the tweets that have been retrieved have been preprocessed in order to retrieve the polarity and subjectivity. Each tweet (string data) is processed by the textblob function, which determines the polarity and subjectivity of each tweet. Create a new 'Sentiment' column with the resultingpolarity and subjectivity in appropriate columns. On a given day, the polarity of all tweets is summed up, and a new column named polarity is formed and added to bitcoin price history data, resulting in the final dataset.

The probabilities of positive and negative statements are illustrated in table below.

- text1=TextBlob("Bitcoin price is increasing")
- print(format(text1.sentiment))
- Sentiment(polarity=0.5, subjectivity=0.6

The results are divided into two groups: polarity and subjectivity. Polarity can have a value of -1 to +1. In this case, 0 denotes a neutral statement, -1 denotes a severely negative statement, and +1 denotes a highly positive one. Subjectivity, on the other hand, has a value ranging from 0 to 1. Here, 0 denotes a statement that is extremely objective, while +1 denotes a statement that is highly subjective.

Sentiment polarity is 0.5, and subjectivity is 0.6 in the above-mentioned result.

Because the polarity is 0.5, the statement is definitely positive, and the subjective score is 0.6, the statement is subjective.

Date (Object)	Total polarity per day (int)

Proposed method:

Because parametric models for forecasting time series data, such as AR, MA, and ARIMA, require input data to be stationary, we prefer neural networks and use LSTM (Long Short Term Memory), an advanced RNN that can manage long-term dependencies. Unlike RNNs, which have vanishing and exploding gradients, LSTMs keep a cell state that acts as long-term memory and feature three gates that allow the network to pick which information to keep and which to discard.

We are using our own dataset, which was created by combining bitcoin price data with tweet sentiment scores over a one-year period. The Long Short Term Memory (LSTM) model is trained using this data.

Normalizing the data:

Because unscaled input variables could produce a lethargic or inaccurate learning process, all of the data is normalized to ensure that the quantities are within the range [0,1].

Dataset Generation:

The time series data must be divided into samples, each with its own input and output elements. To create the dataset, a lookback is chosen first.

For instance, take the following sequence:

11, 12, 13, 14, 15, 16, 17, 18, 19, 20

Then the samples for training the model will look like:

Input	Output
11, 12, 13, 14, 15	16
12, 13, 14, 15, 16	17
13, 14, 15, 16, 17	18

To create the dataset, we used 7 as the lookback period.

Training with LSTM model:

Recurrent backpropagation requires a long time to learn to store information over long time periods, owing to insufficient, fading error backflow [4]. This was addressed by introducing long short-term memory, a revolutionary, efficient gradient-based technique (LSTM). Error signals travelling backwards in time with traditional "Back-Propagation Through Time" (BPTT) or "Real-Time Recurrent Learning" (RTRL) tend to either (1) blow up or (2) vanish.

LSTMs deal with both Long Term Memory (LTM) and Short Term Memory (STM), and the concept of gates is used to make the calculations simple and effective.

Forget Gate: When LTM enters the forget gate, it discards information that is no longer helpful.

Learn Gate: The event (current input) and the STM are merged such that the relevant information from the STM may be applied to the current input.

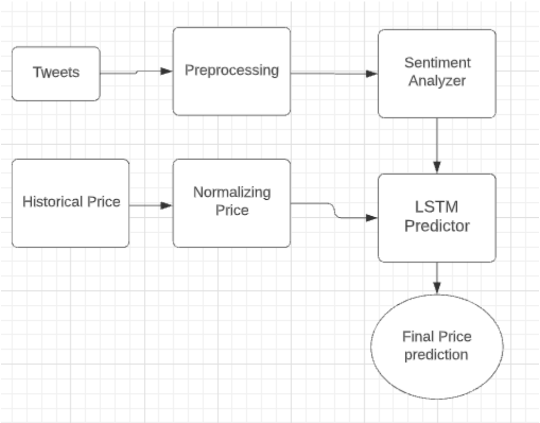
Remember Gate: LTM information that we don't want to forget, as well as STM and Event information, are integrated in Remember Gate, which functions as an updated LTM.

Use Gate: This gate predicts the output of the current event using LTM, STM, and Event, and acts as an updated STM.

The complete dataset is separated into training and testing datasets after it isgenerated with a lookback time. The input layer neurons in the LSTM model are of the train X shape.

We used Relu as the activation function as it’s derivative doesn’t cause vanishing or exploding gradient problem.

Block Diagram:



Experimental Results:

The below figure shows the training loss and validation loss. When the loss becomes constant model training is stopped with early stopping.

Validation loss is 0.0019

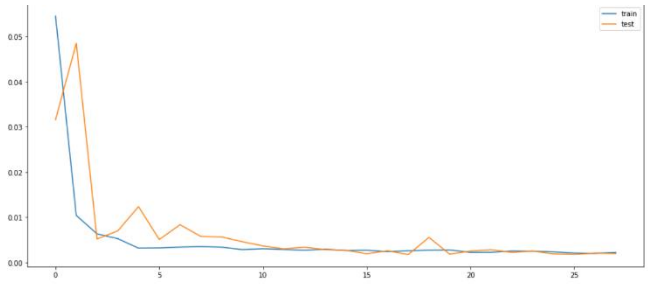


Figure 2.validation loss vs training loss

The testing data prediction plot is included below

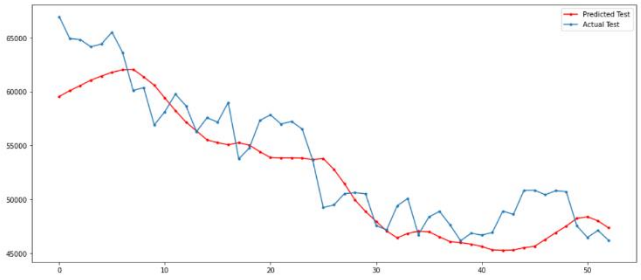


Figure 3. predicted test data vs actual test data

Below is the prediction over Training data

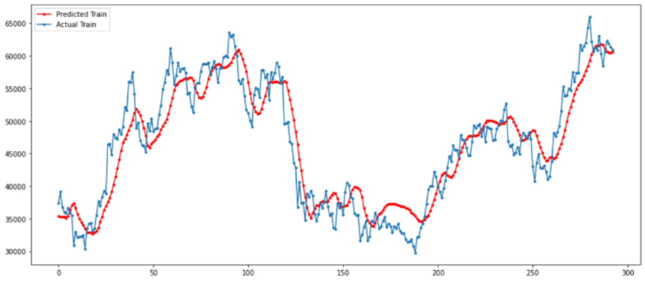


Figure 4. predicted train data vs actual train data

RMSE :

The root of the square mean of the square of all errors (RMSE) [6] is the root mean squared error. The use of RMSE value is fairly prevalent, and it is regarded as a good all-around error statistic.

The error function utilised here is the root mean square error (RMSE)[6]. The RMSE over test data is 2855.296, down from 8788.126 when the lookback period is adjusted. The RMSE on train data is now 3489.402, down from 6079.801 previously.

Conclusion:

Using LSTM model over the historical data of 1 year from 2021 January to 2021 December, we got best results compared to RNN. The error values have decreased from 8788.126 to 2855.296 . we conclude that

for data with random depth, LSTM is suitable fit.

Further considering the public opinion as a feature by including polarity column decreased the error value.

The concept could also be improved by sending users notifications when the bitcoin price reaches a certain level. This threshold (limit price) is initially set by the user.

References:

[1] Satoshi Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”, 2008

[2] ShailakJani, “The Growth of Cryptocurrency in India: Its Challenges & Potential Impacts on Legislation”, 2018.

[3] Zheng Zhao, Weihai Chen, Xingming Wu, Peter C. Y. Chen, Jingmeng Liu, “LSTM network: a deep learning approach for short-term traffic forecast” 2017

[4] FarhaNausheen, SayyadaHajera Begum “Sentiment analysis to predict election results using Python” ,2018

[5] Sepp Hochreiter “Long Short-term Memory”,2016

[6] Simon P. Neill, M. Reza Hashemi, “Root-Mean-Squared Error:Ocean Modelling for Resource Characterization” 2018

[7] Praveen Gujjar J ,Prasanna Kumar H R “Sentiment Analysis:Textblob For Decision Making”, 2021

[8] Poongodi M, Vijayakumar V, Naveen Chilamkurti , “Bitcoin Price Prediction using ARIMA Mode”, 2018

[9] Xi Chen, Hongzhi, Yanjie Wei, Jianzhong Li, Hong Gao, “Autoregressive-Model-Based Methods for Online Time Series Prediction with Missing Values: an Experimental Evaluation”, 2019

[10] Zoran Ivanovski, Ace Milenkovski, “Time Series Forecasting Using a Moving Average Model for Extrapolation of Number of Tourist”, 2018

[11] Sepp Hochreiter, JurgenSchmidhuber “Long Short-Term Memory”, 1997

[12] George Edward Pelham Box, Gwilym M. Jenkins, “Time Series Analysis: Forecasting and Control”, 1994