Predicting Price of Cryptocurrency - A Deep Learning Approach

Samiksha Marne
Department of Information Technology
St. Francis Institute of technology of Mumbai University
Mumbai, India

Shweta Churi
Department of Information Technology
St. Francis Institute of technology of Mumbai University
Mumbai, India

Delisa Correia
Department of Information Technology
St. Francis Institute of technology of Mumbai University
Mumbai, India

Joanne Gomes
Department of Information Technology
St. Francis Institute of technology of Mumbai University
Mumbai, India

Abstract—Bitcoin, a type of cryptocurrency is currently a thriving open-source community and payment network, which is currently used by millions of people. As the value of Bitcoin varies everyday, it would be very interesting for investors to forecast the Bitcoin value but at the same time making it difficult to predict. Bitcoin is a cryptocurrency technology that has attracted investors because of its big price increases. This has led to researchers applying various methods to predict Bitcoin prices such as Support Vector Machines, Multilayer Perceptron, RNN etc. To obtain accuracy and efficiency as compared to these algorithms this research paper tends to exhibit the use of RNN using LSTM model to predict the price of cryptocurrency. The results were computed by extrapolating graphs along with the Root Mean Square Error of the model which was found to be 3.38.

Keywords—Recurrent-Neural-Network (RNN), Long-Short- Term-Memory (LSTM), Deep Learning, Bitcoins.

I. INTRODUCTION

The Bitcoin is a highly cryptic and virtual currency used by many investors throughout the world. Satoshi Nakamoto was the inventor of Bitcoins in 2009[1]. Thus, Bitcoin is a blockchain based currency that encompasses a public records of all the transactions performed under monitoring. Many researchers have worked in this field to predict and analyze the trends and patterns of the Bitcoin prices. Initially, with very less data and limited scope in algorithms and tools the accurate representation and factual prediction of values was difficult but with the advancements in technology and higher scopes in domains like machine learning and deep learning researchers have been keen on developing models that can provide an insight to the estimation of monetary values. A literature survey that consisted of some prominent work in the respective domain provided quite notable results. There is high volatility in the market, and this provides an opportunity in terms of prediction as mentioned in [1]. The writer of [2] proposes a solution to double spending problem using peer to peer distributed server. The authors of [3] assert

that Bitcoin is the world's most valuable cryptocurrency and is traded on over 40 exchanges worldwide accepting over 30 different currencies The study in the paper [4] reveals that the authors have executed the result of Bayesian Neural Networks (BNNs) by analyzing the time series of Bitcoin process. The paper [5] proposes that the model for prediction of time series data based on the concept of sliding window using Artificial Neural Network (ANN) technique which is Radial Basis Function Network (RBFN). It depicts certain limitations such introduction of hybrid or ensemble techniques with new features. The paper [6] attempts to identify and understand daily trends in Bitcoin market by gathering optimal features surrounding Bitcoin prices and plot a graph using normalization. The authors of the work [7], use rolling window Long Short- Term Memory (LSTM) model to predict Bitcoin price by selecting the input features such as macroeconomics, global currency ratios, and block chain information. In the work [8], the authors explore Neural Network ensemble approach called Genetic Algorithm based Selective Neural Network Ensemble by using back tracking strategy where the author suggested that some input information might be missing as more processing of the data was required. The authors of [9] establish a study of Binomial Classification algorithms such as Generalized Linear Model (GLM), SVM and Random Forest. The authors suggest that K-means is a bulbous scope. The work done in paper [10] aims on calculating quantitative gradation and predict values using hand designed features or technical pointers. The research work as proposed in [11] makes use of multivariate linear regression to predict highest and lowest price of cryptocurrencies by using features like open, low and close. The research work presented in [12] attempts to predict the Bitcoin price precisely taking into consideration various constraints that affect the Bitcoin value. For principal phase of the analysis, it aims to know and identify day-to-day fashions within the Bitcoin marketplace while gaining perception into best features

surrounding Bitcoin price. It m v akes use of Recurrent neural networks and LSTM comparison ARIMA model.

This proposed work tends to exhibit the use of Recurrent

Neural Network (RNN) model using Long Short-Term Memory (LSTM) regression algorithm on the acquired Cryptocurrency dataset for predicting the prices of cryptocurrency (Bitcoin) by analyzing the dataset and applying deep learning algorithms. Thus, for this research the dataset used consists of various parameters of Bitcoins data values . The goal of this research is to design a model that will consistently be able to predict the price of Bitcoin. Predicting the exact price is very hard. Therefore, we simplify the problem; we only try to predict whether the price will increase, decrease or stay the same within certain thresholds. The prediction analysis would be carried out based on the resultant values from the given algorithms. The objectives of proposed model is to create model that leads to the Bitcoin price prediction accuracy by incorporating RNN elements.

The brief description of various sections provides an insight that integrates the flow of this work. The second section represents various related works under this domain. The third section provides a generalized methodology used in this work. The fourth section gives an understanding of the proposed methodology which we have undertaken to accomplish our objectives. The results and observations constitute of the fifth section. The sixth section conforms the conclusions and future scope. The seventh and the final section enlists all the references used to assimilate our theories.

II. RELATED WORK

Various data scientists and researchers have worked to find out the prediction of the price of cryptocurrency by the means of different algorithms and approaches. The work proposed in [4] makes use of Bayesian Neural Networks (BNNs) by analyzing the time series of Bitcoin process which describes fluctuation in a timeseries format. The authors suggest the use of different machine learning algorithms to improve variability which the authors fails to sustain. The work presented in [7] aims to analyses a timeseries data of bitcoin prices by using various variables. The authors suggest the scope that the time series data can be modeled by predicating price of the bitcoins using LSTM algorithm. It gives an insight about the backend processing of bitcoins followed by the use of a rolling window LSTM and empirical study for price prediction. The research work exhibited in [11] makes use of Multivariate Linear Regression to predict highest and lowest price of cryptocurrencies by using features like open, low and close. According to the authors of [11], this work fails to provide enough information for long term analysis. Therefore, the authors of this work propose a scope of using LSTM to analyze various cryptocurrencies

by creating data frames for training set. Hence, taking this into consideration LSTM is implemented in our thesis.

III. PROPOSED METHODOLOGY

This proposed method as shown in the Figure.1. depicts the use of Recurrent Neural Network which uses the Long Short-Term Memory algorithms. This proposed method begins with A. visualizing and analyzing the bitcoin dataset followed by B. implementation of RNN model using LSTM algorithm in this model.
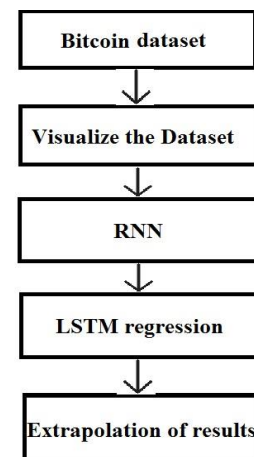

Figure .1. Work flow of the proposed model

A. *Data Visualization of Bitcoin dataset*
The data used here is obtained from Kaggle because it presents
Bitcoin exchanges from the time period of January 2014 to January 2019 . It provides minuscule updates of the bitcoin exchange considering attributes like Open, High, Low, close, Volume, currency, and weighted Bitcoin price. Unix timestamps are available for the same.
Data visualization is done using the Orange Tool. It helps to understand and analyze the data set and different patterns and trends that can help to incorporate various algorithms to predict and perform various operations. The initial working can be shown in Figure .2. where the visualization toolset is available at the left corner with different options to visualize the data. First the dataset file is selected which has to be visualized. Then by selecting scatter plot and distribution function from the toolset it is connected to the datafile.

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
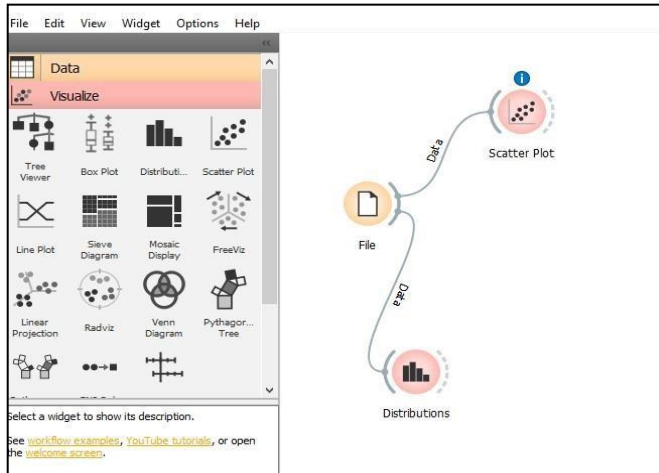**ISSN: 2278-0181**
NTASU - 2020 Conference Proceedings

Figure 2. Data Visualization of the Bitcoin dataset

The graph in the Figure .3. depicts the clusters made by the total volume of Bitcoin traded on the Y-Axis over the average data collected in the span of 24 hours in dollars on the X-Axis. It is evidently visible that between the range 200 to 600 the clusters appear to be dense.
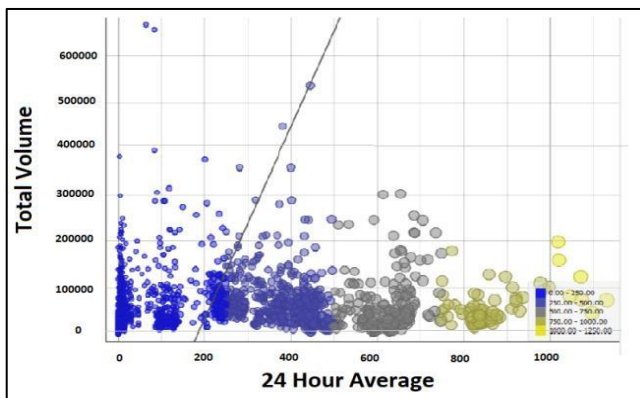

Figure 3. Total volume with respect to average data collected over 24 hours

The graph in the Figure .4. depicts the clusters made by the total volume of Bitcoin traded on the Y-Axis over the data been bided in dollars on the X-Axis. It can be observed that between the range 200 to 650 the clusters appear to have the highest bids. We can see that for a bid of 250 to 300 dollars the Total Volume was the highest.

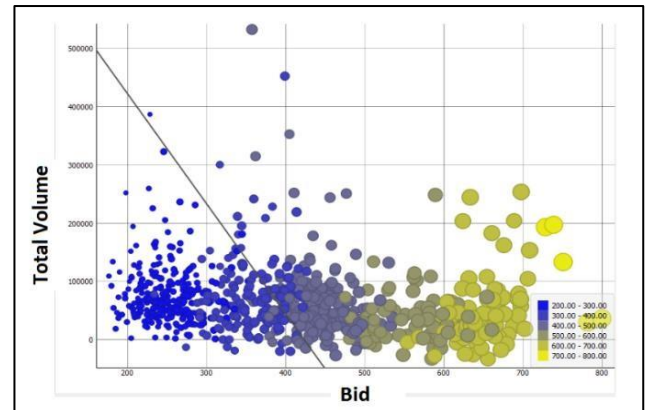Predicting price of cryptocurrency- A deep learning approach

Figure .4. Plot of total volume with respect to the value been bided of the Bitcoins data

The linear plot as seen in the Figure.5. shows that the last amount at which the Bitcoin was bided on the Y-Axis is almost similar to the average data collected over the 24 Hour that was referenced on the X-Axis. This shows the similarity between the two quantities.
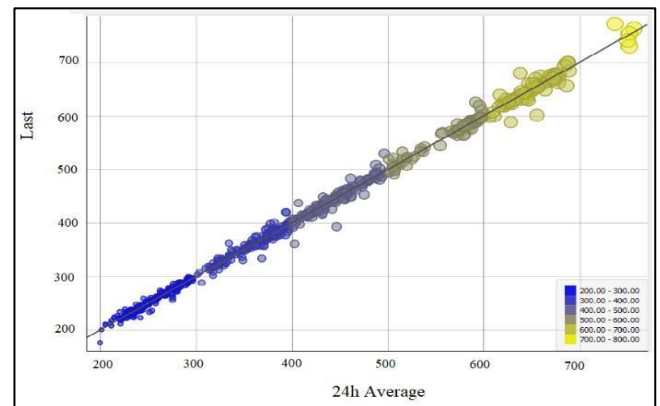

Figure .5. Last price of Bitcoin bided with respect to   the average of data collected over 24 hours

The Figure .6. shows that frequency on the Y-Axis of all the last value of Bitcoins bided on the X-Axis was computed. Along with this, the standard deviation of 388 and variance of 135.57. The frequency of the last value is observed to be highest at 240 dollars with a frequency greater than 140
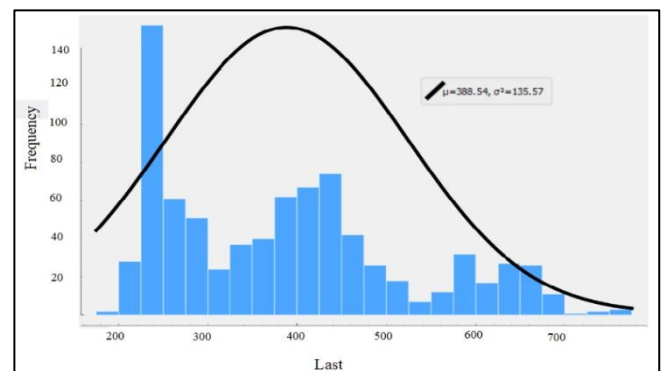

Figure .6. Frequency of last value bided values of

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NTASU - 2020 Conference Proceedings

Bitcoins

The frequency of the total volume bided was calculated to find the greatest and the least volume for the given data set as shown in Figure .7 The standard deviation and variance were calculated for the same. The X-axis represents the total volume whereas the Y- Axis represents the frequency for each value of the volume. The highest frequency was greater than 500 with a mean of 57501 and standard deviation of 55035
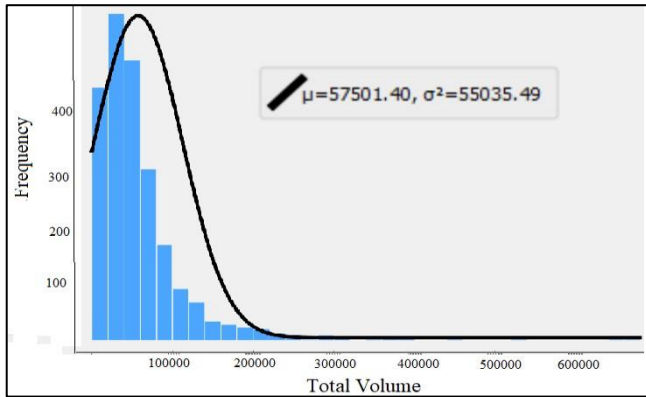

Figure 7. Frequency of total volume bided

B. Implementation of RNN algorithm using LSTM

The flow of the working process is described in Figure 8. Initially, the bitcoin data is retrieved from data sources available online such as Kaggle. This data consists of various bitcoin attributes such as high, low, open, volume, timestamp etc. Out of these attributes the total volume exchanges and timestamp are used for the prediction process in this model.   Next, the deep learning environment is set up. . After the data has been retrieved, some modification needs to be done before deep learning can be applied such as converting into proper format and type.
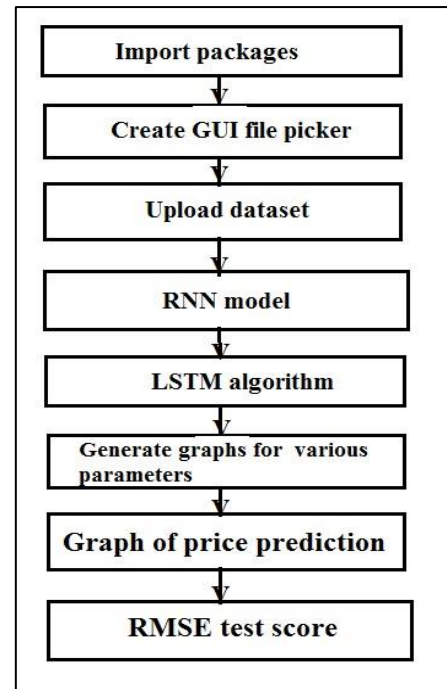

Figure .8. Process Flow Chart

First, the dataset is obtained for USD out of all other country's currencies. Secondly, the data has to be matched and parsed by timestamp date. Thirdly, the data type was changed to remove inconsistencies by converting the data to proper format. After that, the parts of data set with null values is removed. Finally, the data will be split in a train and test set and the data will be ready for machine learning to train a model with. The data is split in a train set, validation set, and   test set based on various parameters. The distinction considered for testing part was the last 30 rows of the dataset whereas for training the rest of the dataset was used so that better training can take place.

The train set is the first part of the data, the validation set is the second part of the data and the test set is the last part of the data. Further, RNN model is used for the complete computation of the price prediction.

Recurrent Neural Networks (RNNs) are a collection of deep learning methods, which has become a widely used method for extracting patterns from temporal sequences [7], making it possibly effective for predicting time series like the Bitcoin price trend. The Figure .9. referenced from the work [3] shows the block flow of how an RNN functions like.
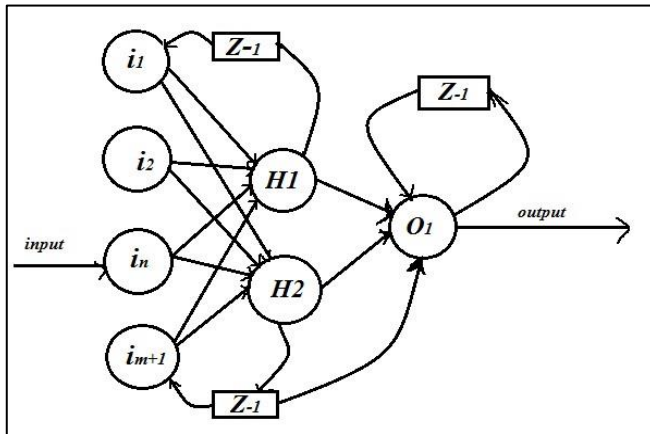
Figure 9. Recurrent Neural Network

An RNN is actually an ANN prepared with historical but temporal memory, as it takes a sequence as input. For deep learning models, parameters are chosen with the help of some options available such as heuristic search model like genetic method and grid search, data pre-process stage is carried out to train the data and reshape into three dimensional arrays. Lastly, after reshaping the data it is finally fed to the LSTM regression model. This model consists of 2 hidden layers that

are used for better computation and performance. Long Short-Term Memory networks can learn long-term dependencies. Thus, these networks can remember information for long periods of time by default. An architectural flow of an LSTM model is shown in the Figure .10. which is referenced from the work proposed by author of [9]
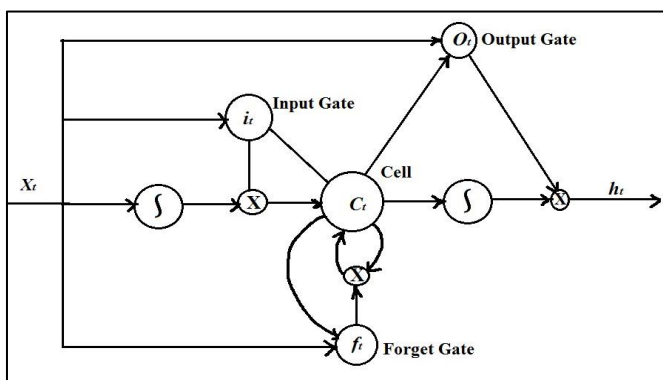

Figure .10. Long Short-Term Memory

Recurrent Neural Networks have chains of repetitive iterating modules of neural network. In standard RNNs, these modules are a simple connection of network layers. LSTM's have the cell state, and different gates. The cell is used to transfer information throughout the sequence also known as memory of the network This information is remembered by the memory in long term dependencies. So

even information from the earlier can be used later thus eliminating the use of short-term memory. The resultant was used to predict the training and testing score and root mean square error. Thus, graph for the same was plotted. For ease of user manifestation, a GUI file picker was created.

IV. RESULTS AND OBSERVATIONS

A JavaScript code for a GUI file picker was used for the ease of data selection by users as shown in Figure 11. This also allows users to use different data files in .csv format thus increasing the scope of the project by not limiting to a particular dataset.
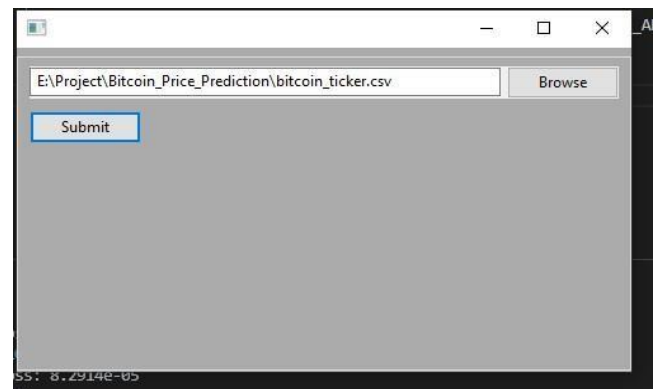

Figure 11. GUI For dataset selection

Figure .12. shows that the X-Axis represents the dates from 2015 to 2019 and Y-Axis represents the mean USD Dollars, and the exchange between these quantities shows that January 2018 had the highest exchange of Bitcoins
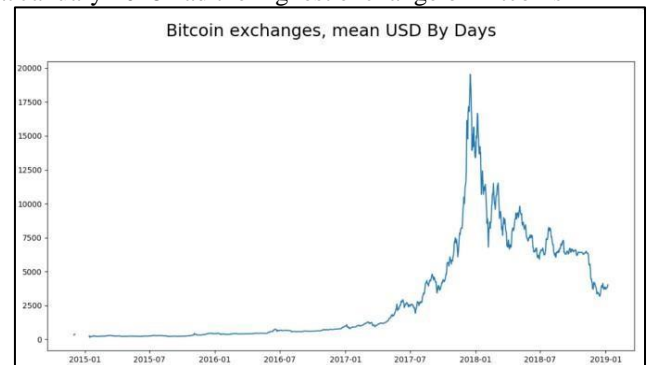

Fig.12. Bitcoin exchanges mean USD by days

The Bitcoin exchange over the month with respect to mean USD Dollars is represented as given in the graph shown in Figure .13. The highest mean of USD exchange was observed in January

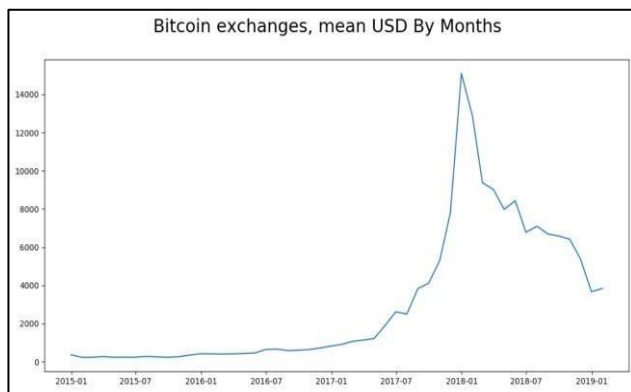Predicting price of cryptocurrency- A deep learning approach

Fig.13. Bitcoin exchanges mean USD by months

The Bitcoin exchange over the mean USD by Quarters is shown in the graph in Figure .14. The quarters between January 2018 to January 2019 showed the largest exchange of Bitcoins
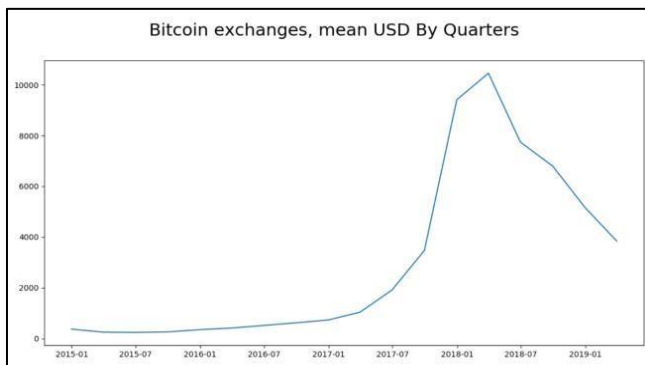

Fig.14. Bitcoin exchanges Mean USD By Quarters

The Bitcoin exchange graph from mean USD as per different years is shown in the Figure .15. Out of all the years considered the span between 2018 and 2019 has the highest exchange of Bitcoins over the past four years.
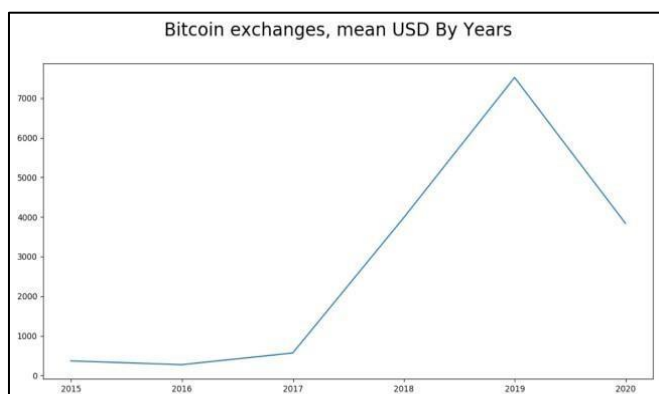

Fig. 15. Bitcoin exchange as per mean USD by Years

In the Figure .16. the X-Axis of the graph represents the price is USD The Y-Axis on the graph represents the timestamp dates of the data generated as per the dataset

csv file. It gives the predicted line graph with respect to the actual line graph of the values
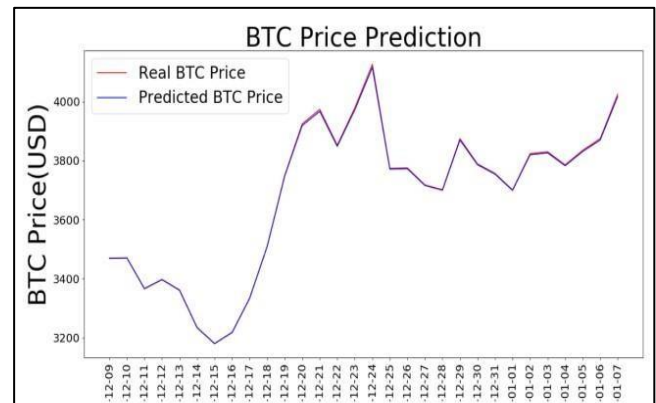

Figure 16. Graph Output for Price Prediction

The red line represents the actual value of the Bitcoin price whereas the blue line of the Bitcoin data represents the predicted value. It is very evidently observed that the difference between the actual and predicted vale is very minute. With every epoch and different ratios of datasets different variations of the graph can be extrapolated. The Root Mean Square Error (RMSE) calculated was 3.3% of the Testing data set.

V.   CONCLUSION AND FUTURE WORK

The optimized implementation of RNN using LSTM for real time datasets and models was studied. The use of deep learning and its usage to real time problem of crypto currency price prediction was performed. The implementations of the data preprocessing and filtering to give precise, sound and consistent data was executed successfully by creating a GUI File Picker for ease of users and greater scope. The usage of RNN using LSTM algorithm was done effectively. The accurate representation of the system design along with the precise threshold outputs at the display unit was done. The results were noted, and outputs were recorded by plotting a graph. The RMSE Test Score value was computed as 3.38. The proposed model further would have advancements in terms of design and functionality. A sentimental analysis using twitter dataset is a prominent scope. Along with this, more complex and advanced algorithms supporting high level neural networks can be used. On comparison with other reviewed papers in the literature survey the difference between this calculated RMSE value of the test score was computed as shown in TABLE I.

TABLE I. COMPARISON OF RMSE OF TRAINING AND TESTING PHASE

Reference Number	ROOT MEAN SQUARE ERROR	
	Error of reference papers	*Error of our paper*
[3]	Test: 8.07	Test:3.38
[4]	Test: 0.23	Test:3.38
[5]	Test:9.1	Test:3.38

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NTASU - 2020 Conference Proceedings

VI. REFERENCES

[1] M. Briaere, K. Oosterlinck, and A. Szafarz,Virtual currency, tangible return: Port- folio diversification with Bitcoins, Tangible Return: Portfolio Diversification with Bitcoins , 2013.

[2] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System", Available at: https://Bitcoin.org/Bitcoin.Accessed on 2008.

[3] McNally, Sean & Roche, Jason & Caton, Simon. (2018). Predicting the Price of Bitcoin Using Machine Learning. 339343. 10.1109/PDP2018.2018.00060.

[4] H. Jang and J. Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information," in *IEEE Access*, vol. 6, pp. 5427-5437, 2018

[5] Hota HS, Handa R & Shrivas AK, "Time Series Data Prediction Using Sliding Window Based RBF Neural Network", International Journal of Computational Intelligence Research, Vol.13, No.5, (2017), pp.1145-1156

[6] Siddhi Velankar,Sakshi,Valecha,Shreya Maji, aBitcoin Price Prediction using Machine Learninga,20thInternational Conference on Advanced Communication Technology(ICACT)on,vol.5,pp.855-890,2018

[7] Jang Huisu,Jaewook Lee,Hyungjin Ko,Woojin Le, aPredicting Bitcoin price using Rolling Window LSTM modela ,DSF, ACM ISBN 123-4567-24- ˆ567/08/06,vol.4,pp.550-580,2018.

[8] Sin, Edwin & Wang, Lipo. (2017). Bitcoin price prediction using ensembles of neural networks. 666-671. 10.1109/FSKD.2017.8393351.

[9] Isaac Madan, Shaurya Saluja, Aojia Zhao, Automated Bitcoin Trading via Machine Learning Algorithms, Stanford:Department of Computer Science, Stanford University, 2015.

[10] John Mern1; Spenser Anderson1 ; John Poothokaran1 ,a Using Bitcoin Ledger Network Data to Predict the Price of Bitcoin

[11] Ruchi Mittal;Shefali Arora;M.P.S Bhatia; "Automated cryptocurrencies price prediction using machine learning" a,ICTACT JOURNAL ON SOFT COMPUTING, VOLUME: 08, ISSUE: ˆ04 JULY, 2018

[12] Amin Azari,a Bitcoin Price Prediction: An ARIMA Approach ˆ a,Available at: https://www.researchgate.net/publication/328288986, 2018

[13] http://wp.firrm.de/index.php/2018/04/13/building-a-lstm-network-completely-from-scratch-no-libraries/