



A comparative study of bitcoin price prediction using SVR and LSTM

¹J. Arumugam, ²S.Sabarichvarane, ³Dr. V. Prasanna Venkatesan

¹Ph.D. Scholar, ²B.Tech Student, ³Professor

¹Department of Computer Science, ²Department of Computer Science & Engineering, ³Department of Banking Technology
^{1,3}Pondicherry University, Puducherry, India, ²Sri Manakula Vinayagar Engineering College, Puducherry, India,

Abstract: The most significant objective of this paper is to determine an outstanding model for forecasting the price of a bitcoin for the future days. All along the way accepting various factors and methods in finding a price can be utilized with different techniques like Long-Short Term Memory (LSTM) and Support Vector Regression (SVR). This paper deals with analyzing and reviewing the great practicable methods used to accurately predict bitcoin prices. The most significant aspect that should be taken into account is that the bitcoin prices from earlier years should be taken from the database to perform prediction. For analysis, the dataset should be pre-processed and adjusted for the actual review process. The most significant feature of this paper is to pre-process the dataset which is found to be uncooked. Besides, pre-processing the dataset, on the other side, the utilization of various models like Support Vector Regression, Linear Regression, and Long-Short Term Memory, and the generation of accurate results will also be explained in this paper. Additionally, the proposed work will show the applications of predicting frameworks in real life, and issues related to accuracy are also added. The paper further presents Machine Learning models with the best accuracy, the impact of training and testing ratios, and the removal of outliers to predict the price of bitcoin. The efficiency of bitcoin price prediction will be the greatest solution to the financial markets.

Index Terms - Support Vector Regression, LSTM, Financial market, Bitcoin, Machine Learning.

I. INTRODUCTION

The term Bitcoin refers to a peer-to-peer system of digital payment which was introduced in the year 2009. It is the first open deliver crypto overseas currency, an open-source software program algorithm that controls the application of bitcoin transactions, and also with the help of the internet a bitcoin is created which can be further used to report and verify its transactions. Funding needs to be feasible through terrific business centers known as "bitcoin exchanges". Being the digital currency, Bitcoin uses the standards of cryptography to manipulate the advent and alternative of Bitcoin. The arrangement of more bitcoins in the digital wallet essentially portrays a digital report. The transaction history of all the records and the statistics of timestamps are stored in a place known as Blockchain. The blockchain consists of a single file and each file is referred to as a block. Every block consists of a pointer to a preceding block containing the data. Blockchain is responsible for encrypting the information. Within the path of transactions, the client's name isn't located, however, most effectively their wallet identity is made public.

The digital currency is essentially a collection of diverse customers and companies of bitcoins. A bitcoin is responsible for taking care of various considerations that are claimed by a certain person or a group of people associated with an organization. The cryptographic money forecast is defined as the process of selecting and evaluating the future bitcoin exchange. The conjecture is always focused on maintaining strongness, caution, and success. The system needs to fill in as verified with the aid of real situations and ought to be suitable for actual settings. The structure seems to be in an equal manner which should be taken into account for each one of the components that result in the impact of quality and execution of bitcoin. There are numerous techniques and strategies for finishing the belief models which include essential technical analysis, essential analysis, market Mimicry, Time association in a perspective organization, and Machine Learning. The forecast has been circulating in the progressive regions due to the high-level development of time forecasting. The highly promising and significant methods for the prediction process include the utilization of Recurrent Neural Networks and Artificial Neural Network which is essentially the usage of Artificial Intelligence. AI consists of man-made reasoning that enhances the model to take part and develop from previous experience instead of redoing again and again which is highly time-consuming. Preferred strategies for forecasting in AI use a calculation like Backward Propagation, in any case, were given back to spread errors. Overdue, numerous experts are using a greater distinguished degree of collecting learning techniques. It would use minimum time and effort postponements to foresee future highs even as some other frameworks will use loosened increases to expect a rise in the future. These conjectures had been utilized in shaping the bitcoin's prices.

With the given market capitalization and pervasiveness of bitcoin, it is exceptionally a crucial stage to cope with the precariousness associated with greater circumstances within it and verify the eccentrics for taking risks and adventures results. In recent times, despite different individual monetary benefactors, various institutional monetary supporters have recalled Bitcoin for their portfolios. The worth advancement for Bitcoin isn't equivalent to those of other money-related assets considering its eccentricities [1]. It has been proved that Bitcoin is more critical with high precariousness when compared to stocks and it is found to be an irrelevant and poor association with some assets [2,3,4]. This prompts the opportunity of making supporting frameworks and forming a lot of separate portfolios by using Bitcoin [5]. Regardless of the component of Bitcoin being unsteady, it fabricates the threat for adventure execution, whether or not a more critical eccentrics much of the time prompts a more raised degree of benefits. Subsequently, predicting the insecurity even more correctly accomplishes pivotal risings in the pay of monetary sponsors by diminishing the peril.

The term bitcoin refers to a cryptocurrency, that deals with digital payment and uncomplicated investment requirements throughout the world. Bitcoin cannot be owned by anyone as it is considered to be decentralized. Bitcoin has made transactions easier and more reliable as they are not associated with any country. Various investments can be carried out by different marketplaces called "Bitcoin exchanges". These make the people engage themselves in buying or selling bitcoins by using various currencies. Mt Gox is the largest bitcoin exchange in the world. The digital wallet is quite similar to a digital bank account and plays an important role in storing bitcoins. A blockchain is a place where the timestamp data and records of every transaction are stored. A single record, which is present in the blockchain is termed a block. Each block consists of a pointer to the block containing the previous set of data. The data which is present on the blockchain is highly encrypted. The name of the user will not be revealed during every transaction, but the wallet's identity is visible to the public.

The machine learning process is divided into three main categories, Supervised Learning is the most significant when compared to other types. In the supervised learning technique, it is highly essential to know about the concept of two sets namely, the test set and the train set, which is found to be useful in making better predictions about unnoticed observations.

The largest question which is required to answer by everyone is: "how does the machine learning algorithm work?" To respond to this question, it is essential to hold a group of data, known as the test set, which is never used for learning and making hyperparameters or architectural decisions. The most significant feature of the bitcoin price prediction model focuses on predicting the object variable i.e., the price of a given day by using various information from the dataset which includes opening price, closing price, and various other factors. The technique which is used in the prediction of bitcoin prices incorporates the prediction of time series forecasting with different technical analyses and predicting the accurate price of a bitcoin, and illustration of machine learning. The fundamental aim of this analysis is to bring the model into a correct structure. With the help of collected data, a machine learning model is used to calculate the price patterns for the future according to different turns of events. The bitcoin price prediction model lacks inefficiency when it is under classification, and it has high performance when a regression model is used. The Support Vector Machine (SVM) method plots the segment of every piece of information in n-dimensional space (where n refers to the number of highlights which are required to access the dataset) by calculating the highlights for a specific range. Different predictive models like LSTM are used for similar purposes. This Random Forest algorithm concentrates on taking subsamples from the dataset; this reduces the over-fitting of the model and makes the predictive analysis more effective.

II. RELATED WORK

Shah has executed the Bayesian regression with a divided time duration of three segments of 180,360,720 are used for pattern finding, parameter learning, and evaluating algorithms [6], and Madan has applied the tree-based regression techniques and deep learning, their considered training set contains the first 70% of the dataset, while test set is the remainder [7]. Arun Upadhyay et. al. (2012) created the Multinomial Logistic Regression (MLR) model to anticipate the outflanking stock. The author utilized financial ratios as usable choice standards for deciding execution in the financial exchange dependent on the stock return contrasting and market return. In a comparative rise or decline arrangement issue, in [8], a highly corrected rate of 56% across the tests using PCA the dimensionality reduction process was accomplished by taking up a diverse arrangement of specialized pointers. To appraise the course of bitcoin, Recurrent Neural Networks (RNNs), Convolutional Neural Organizations (CNN), and long - short term memory networks (LSTM) are utilized as deep learning architectures, experiments are done on the English Twitter dataset [9].

Derbentsev et. al [10] employed the Stochastic Gradient Boosting Machine (SGBM) and Random Forest (RF) models to anticipate the three widely used cryptocurrencies such as Ripple, Ethereum, and Bitcoin (XRP, BTC, ETC) were between the range of 0.92-2.61%. The dataset was partitioned into test and training subsets in a proportion of 20% and 80% to train the model, tuning and fitting their parameters. Ramesh Medar et. al. tried different things with a model prepared to utilize Artificial Neural Network (ANN) is utilized for climate forecasting. The input dataset contains 365 samples out of which 20% of data were utilized as test data, 20% of samples were used in the validation of data and the remaining 60% of samples were utilized as preparing information. The Mean Square Error is utilized to complete the assessment of performance Justin Cheng et al [11] According to the path used in social media like Facebook, and Twitter, the functionality of reposting or resharing allows the content of others with all the followers/friends. The photo sample reshapes the cascade on each of the social media and finds performance in the predicting to grow of the future time Pradeep Singh et al

[12] In this digital world, social media is a major attraction for everyone at this time. The information which is to be published is to be given properly so that it does not get misunderstood by anyone. The main result is to make them filter from each other and take ahead of the other methods and make us get the output needed. Dancho Panovski et al [13] This article introduces a singular system mastering technique committed to predicting the bus arrival instances within the bus stations over a provided period. They encompass conventional system mastering techniques, together with linear regression and support vector machines (SVM), but additionally enhanced, the tremendously non-linear neural community which is primarily based on the total processes. They additionally display those substantial enhancements that may be done over the country of the artwork techniques

Ayush Varma et al [14] Real estate is considered to be the least transparent industry in our environment. The price of homes keeps changing in our day-to-day life and is sometimes inflated unlike, relying on valuations. The results might not be the only identification of a technique and various regression techniques were used in this path, to give more accurate results, a weighted means of different techniques is applied. Han Seung Jang et al [15] Due to the great increase in the huge number of large-scale PV farms (PV), the penetration of solar radiation into the main grid has gradually developed in recent years. On account of the wide variation in meteorological conditions, there will be some fluctuations in the power output of the PV farms. So, compensation should be made in advance to overcome this effect. This paper proposes a solar prediction model from different satellite images by using the Supporting Vector Machine (SVM) algorithm.

Gabriel et al [16] A software fault called prognostication is occurred due to variation of data. Software engineers find it difficult to handle the software fault i.e., imbalanced data for the previous prediction of software faults. Pravas Ranjan Bal et al [1] The railway infrastructure and rolling stock have found some defects in the wheels on railway wagons and it is identified to be a continuous source of damage. This results in the production of vibration emissions and noises which requires a high cost to repair. This paper provides two machine learning models which are used in the automatic detection of defects in the wheels. A sensor system is installed on the railway network to monitor the virtual force of the wheel. Pradeep Singh et al [12] In this digital world, social media is the major attraction for everyone at this time. The information which is to be published is to be given properly so that it does not get misunderstood by anyone. The ML Approach, this one is also successfully done many features are used to train the models of machine learning for the desired task. The main result is to make them filter from each other and take ahead of the other methods and make us get the output needed.

Poonam L. Patil et al [17] Different statistical methodologies are broadly divided into social media applications like Twitter and Facebook. The research deals with the information of a person and what they communicate with the help of their updated statuses which will be highly useful in determining the human personality. Based on some machine learning models, the personality traits of Facebook users have been predicted and it is a part of the Big Five model. This paper uses a dataset of various Facebook users from their traits. This paper uses the concept of having data to extract data from the existing feature, after that choosing the feature will be done. XGBoosting Ranking Form is used for character prediction. Berina Alic et al [18] This paper gives an outline of various machine learning techniques involved in distinguishing cardiovascular diseases (CVD) and diabetes by using Bayesian Networks (BN) and Artificial Neural Networks (ANN). The papers which are published from 2008 to 2017 are selected for the comparative study. Levenberg-Marquardt's learning algorithm along with a multilayered neural network which is a part of ANN is found to be used more on selected papers. On the contrary, Naïve Bayesian is the highly used algorithm of the Bayesian Network and it provides efficient results in classifying CVD and diabetes with an accurate value of 99.51% and 97.2% respectively. Jue Wang et al [19] The most familiar deep learning models are used for recognizing actions in the videos, and various independent predictions are made for the short clips which are further combined to produce an action label, which is assigned for the full video segment. Underlying actions are not characterized by all frames, but some multiple actions are most common, and various combined schemes are responsible for providing equal importance to every frame that is unfavored. To overcome this problem, the discriminative pooling analysis is proposed, based on the idea that deep learning features are generated to the short clips for characterizing the actions. In identifying the usual features, it is believed to be a negative bag consisting of various features that are found to be trivial. The results are reported from the experiments conducted on datasets of eight computer vision benchmarks, providing different kinds of tasks related to video and displaying the performance of art related to these tasks.

In the next section, we describe the machine learning techniques which can be inserted into the bitcoin price dataset. The results of the application for these methods have been presented to the dataset. In conclusion, various steps for separating training and testing sets from a given sample of data are discussed.

III. PROPOSED STUDY

An enormous amount of data is analyzed, which results in successful outcomes and assessments that are considered to be the most important benefits of machine learning. This analysis however completely depends on the correctness and accuracy of using the data. The evaluation of the performance of data splitting used in machine learning models is considered to be highly efficient, which provides a way for selecting reliable data splitting for the ML model. Hence, this study focuses on two widely used machine learning models such as Support Vector Regression (SVR) and Long-Short Term Memory (LSTM) for modeling.

3.1 SVR

The concept of regression in Support Vector Machine (SVM) was introduced in the year 1996 and it was named Support Vector Regression (SVR). The SVR focused mainly on a group of training data because the training points which lie behind the boundary are not taken into account by the cost function while building up the model. SVRs are more accurate in performing the calculation, which is portrayed to be utilized into pieces, minima from surroundings, small conditions during the arrangement, and limiting the control from numbered support vectors, topmost of the edge, etc. For understanding the prediction accuracy of machine learning models, the dataset is divided for training/test in different ratios. The impact of training and testing ratios is recorded based on their performances. Outliers cause the model to degrade.

3.2 LSTM

Long short-term memory (LSTM) plays an important role in deep learning and it is a part of an artificial recurrent neural network (RNN) which is widely utilized in research activities carried out in deep learning, consists of a feedback connection, not the same as the standard feedforward Artificial Neural Network and hence results in the formation of the general-purpose computer. The outcomes of the predicted prices and real prices are illustrated in Figure 4, where different sizes of the window are 7 days and 3 days.

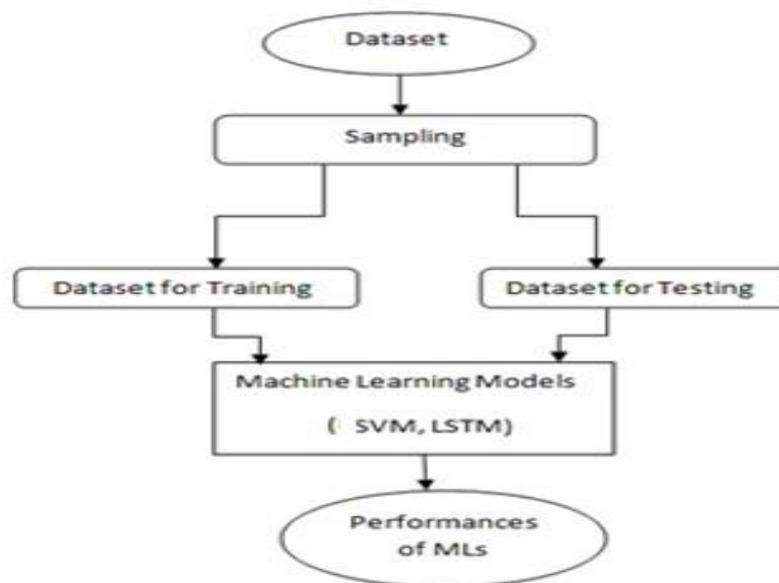
IV. PROPOSED METHODOLOGY

Various methods that affect the training and testing of machine learning models will be discussed in the section. The fundamental goal of this approach is to detect the suitable model which can be employed in the training and testing of data. The major instance of this model is to illustrate choosing the appropriate model by defining from the primary step to the final step. This example provides explains the selection of methods at each stage. The flow diagram in figure 1 shows the step-by-step procedure involved in this study.

4.1 Data Collection

The most significant and primary step toward the analysis is the Data Collection. It deals with the collection of accurate datasets for analysis. The dataset which is collected should be utilized for market prediction and must be differentiated based on various perspectives. Our dataset mostly deals with the bitcoin prices of previous years. Initiating with the first step, the dataset should be analyzed with the correctness of its values, which will be used in the model for accurate predictions of future prices.

Fig. 1. Flow diagram



4.2 Data Pre-Processing

The term data preprocessing plays a major role in the mining of data, which is responsible for transforming the raw data to a clearer configuration. There is an enormous number of errors in the raw data, therefore it is considered to be consistent. The data pre-processing contains various processes like detecting the qualities which have been missed, finding the uncategorized qualities, and dividing the dataset into test and preparation sets. Finally, to limit the scope of factors that depends on the surroundings, an element scaling should be performed.

4.3 Training the Machine

The methodology which is being used in the development of the model focuses on detecting a few initial qualities and further strengthening the parameters which are required for the model. A cross-validation is done during the development of the model, to gain a high range of evaluation when the model gets executed with the help of prepared data. The process involved preparing and training the machine, which is similar to the supervision of data with great care for the algorithm to complete the test data. Since the model is not considered enough to make decisions about the unnoticed data, the test data is kept untouched.

4.4 Data Scoring

The strategy of utilizing the prediction model in various numbers of data is termed scoring of information. The process- which is carried out in processing the dataset is called SVR Algorithm. The fluorescence of learning models aims in accomplishing the possible outcomes. Based on specific parameters, the rise or decline of prices is predicted with the help of findings which will be discussed in the final module. In parallel, it displays the fragility of certain bitcoin or some other materials. The elements which are beneficial in getting the outcomes are ensured and approved by the client confirmation System Control.

V. EXPERIMENTAL STUDY

We have collected datasets from the following link <https://www.coindesk.com/price/bitcoin>. The timestamp of the dataset contains USD 1-day data from July 2010 to March 2021. The dataset consists of 3874 rows and 6 columns. The columns denote different parameters of the collected data while rows represent values from each day of the corresponding days. The columns are discussed as follows: Date represents the price of a particular day recorded. The open and close represents the price value at the beginning and end of the day. The high and low columns represent the highest and lowest values of prices that occurred on a particular day. The volume denotes the volume of a bitcoin price traded on a day, whereas the change% is the difference between the highest and lowest price of a bitcoin.

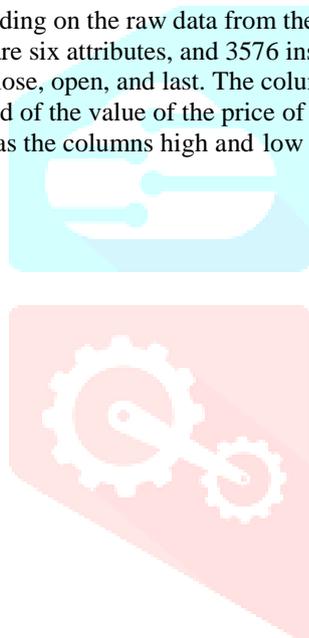
This data set contains day-wise data on bitcoin value Content Period: July 2010 – February 2021 (10 years 7 months) Granularity: Day Wise. The word embedding methods at each training set percentage and classification accuracies of deep learning models [7].

	Date	Price	Open	High	Low	Vol.	Change %
0	18-07-2010	0.1	0.0	0.1	0.1	0.08K	0.00
1	19-07-2010	0.1	0.1	0.1	0.1	0.57K	0.00
2	20-07-2010	0.1	0.1	0.1	0.1	0.26K	0.00
3	21-07-2010	0.1	0.1	0.1	0.1	0.58K	0.00
4	22-07-2010	0.1	0.1	0.1	0.1	2.16K	0.00
...
3869	19-02-2021	55906.6	51590.1	56238.5	50816.8	139.43K	0.08
3870	20-02-2021	55923.7	55922.0	57523.8	54124.1	127.85K	0.00
3871	21-02-2021	57433.8	55907.6	58335.1	55502.7	87.26K	0.03
3872	22-02-2021	54111.8	57437.6	57480.8	48353.8	244.95K	-0.06
3873	23-02-2021	49773.5	54114.1	54115.0	48750.4	268.61K	-0.08

3874 rows × 7 columns

Table 1. Bitcoin dataset

Depending on the raw data from the xlxs record, the results can be published easily. To find the rise or decline in the bitcoin price, there are six attributes, and 3576 instances are used for this prediction process. The attributes are listed as follows the date, high, low, close, open, and last. The column data represents the price value of a day. The columns open and close denote the starting and end of the value of the price of bitcoin for a certain day. The column last denotes the last recorded price of a particular day whereas the columns high and low denote the highest and lowest price of a day.



	Price
0	0.1
1	0.1
2	0.1
3	0.1
4	0.1
...	...
3869	55906.6
3870	55923.7
3871	57433.8
3872	54111.8
3873	49773.5

3874 rows × 1 columns

Table 2. Bitcoin price day-wise for the training set

5.1 Test and Training set ratios

How long the test should be? The value $O(1/N)$ represents the estimation error of the test performance, Where N denotes the size of the test set, it is simply based on how the Monte Carlo sampling method is used for reducing the standard error. For the prediction process, the default value of choosing the train and test set will be in percentages of 80% and 20%, but the percentage of the test set might be less out of the overall data if the dataset is larger and there will be a huge confusion while choosing the test. A common situation is found to occur while trying to fix the model, the goal for predicting the bitcoin price moves faster by varying the underlying distribution of data. For instance, when identifying the Netflix prize, a dataset was gathered and released by Netflix that contains huge data from various users throughout the world. When comparing the other competitors, the test has first taken for the prediction purpose and then the training set was collected. From this observation from Netflix, they focused on forwarding the predictions faster. If an algorithm is chosen randomly for the dataset for all the years, there will be great difficulty in predicting future prices. Hence, it will be ideal to split the datasets into two parts test and training sets as illustrated in Figure 3, for predicting an accurate value.



Fig.2 Random training/test split



Fig.3 Temporal training/test split

To predict the price of a bitcoin by using Machine Learning algorithms, the Root Mean Square Error (RMSE) value should be utilized in the evaluation of performance.

RMSE is denoted as:

$$RMSE = \sqrt{(q - b)^2} \quad (1)$$

Where q denotes the value to be predicted and b represents the real value.

This paper deals with machine learning techniques used for regression problems in particular bitcoin price prediction. We have taken different training and test dataset size to get an optimum result from various machine learning techniques. The ratio of training and testing dataset are defined below:

<u>Training(%)</u>	<u>Test(%)</u>	Training size	Test size	SVR Normalized MSE	LSTM Normalized MSE
40	60	1549	2325	6.53971E-05	0.003671773
50	50	1937	1937	7.84791E-05	0.006783988
60	40	2324	1550	9.67046E-05	0.011204557
70	30	2711	1163	0.000125415	0.002729209
80	20	3099	775	0.000146273	0.001398904
90	10	3486	388	0.000257713	0.003696527

Table 3. training/test dataset ratios

5.1 Outliers Removal

Outlier removal is considered to be similar in the meaning of pruning which refers to implementation, with some significant differentiation. However, the majority of the implemented code corresponds to the definition of pruning, which can be utilized again in the removal of outliers.

VI. RESULTS AND DISCUSSIONS

The bitcoin dataset has been applied to the Support Vector Regression Algorithm (SVR), and Long Short-Term Memory (LSTM) algorithms. This algorithm divides the dataset into two forms testing and training. The Root Mean Square Error values of all the two algorithms differ from each other with a slight variation of values. A time series plot of bitcoin price displays the predicted bitcoin price for the future using the testing and training values. Figure 3. shows the trend of bitcoin price prediction and figure 4. exhibits different MSE values of machine learning techniques of SVR and LSTM. Table.3 shows that training/testing dataset ratios of LSTM and SVR are 80:20 and 40:60 respectively for optimum models. Applying machine learning algorithms on a dataset in different sizes for training and test purpose is not fixed for better results, 80:20 in particular. It is varied based on the application to use the model. In general, an 80:20 split of training/test ratio has been taken for model learning. We use LSTM on the bitcoin dataset, it adopted this ratio only but not in the case of SVR.

We have taken training and test split ratio not only considering machine learning algorithm perspective, depending upon train/test split ratio performance improved on 40:60 of SVR algorithm and also 80:20 of LSTM which gives optimum value.

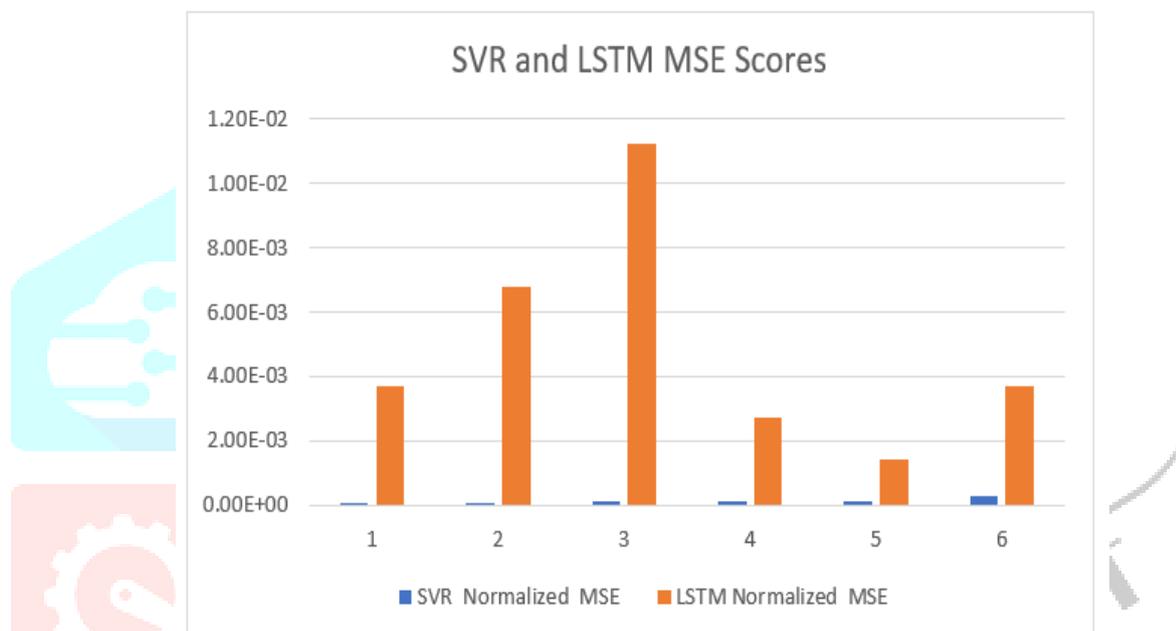


Fig. 4. SVR and LSTM MSE Scores

6.1 Limitations

The study on Support Vector Machine, linear regression, and Long Short-Term Memory has been done on the bitcoin dataset and these algorithms help in predicting the prices of bitcoin for future days. From the above three algorithms, the Long Short-Time Memory (LSTM) provides the best-predicted result. We consider the price attribute of the bitcoin dataset, not taking any other features impacting the bitcoin market like gold price, oil price, and other parameters of the world economy.

VII. CONCLUSIONS

In today's world Bitcoin has become highly essential, which can affect the entire society and the world in various aspects, especially in terms of safety and security. Thus, predicting the changes or trends in Bitcoin price will be significant. But it is highly challenged when compared to traditional stock markets. The introduction of machine learning in artificial intelligence brings new ideas. The proposed idea in predicting the price of a bitcoin is to determine and also to design the patterns associated with price and volume movement over a particular interval of time. Two machine learning models have been proposed – based on LSTM and SVR – that deal with analyzing and predicting how the changes in price or volume of a bitcoin interact over a period which leads the price either to increase or decrease in the following next days. The experimental results display that the proposed models are considered to be high sufficient for capturing long-range dependencies and stating the art of performances.

In this paper, various machine learning models like SVR, LSTM, and LR have been utilized in the prediction of bitcoin prices. To compare these algorithms, the Root Mean Squared Error Value (RMSE) is applied in measuring the accuracy. Thus, the Results display the LSTM contributed the best performance among all the three algorithms. Here, three machine learning algorithms have been utilized in the prediction process. Various other Machine Learning algorithms can also be found in predicting the Bitcoin price. When coming to this paper, only one dataset has been used for training and testing purposes from the dataset. Multiple datasets may also be used in getting more efficient and accurate results.

The investigation intends to decide the processes influencing the training and testing phases in machine learning and to pick the proper techniques for this interaction. In the article, a definite clarification is made for machine learning training and testing. To estimate the RMSE value for different algorithms, the most practical algorithm for expecting bitcoin on various information centers

from the chronicled information is LSTM. The LSTM is considered to be an incredible asset for delegates and economic experts for placing assets in the financial market like cryptocurrency since it is placed on an enormous assortment of recorded information and has been selected after being tried on example information. These undertaken machine learning models assist to foresee the bitcoin price with high accurateness when contrasted and as of late executed machine learning models. The training/testing dataset ratios of LSTM and SVR are 80:20 and 60:40 respectively for optimum models.

VIII. FUTURE ENHANCEMENT

In the future, the prediction of a bitcoin price can further be enhanced by adding a dataset containing a huge amount of data compared to the one which is currently being used. The accuracy of bitcoin price prediction can be improved greatly. Moreover, various other machine learning models can also be observed and used for testing the rate of correctness resulting from the dataset. The future development of this work will be dealing with the addition of a greater number of parameters and factors that includes financial quantities, an enormous number of instances, and some other entities. If a large number of parameters are taken into consideration, the results of the accuracy might be greater. To analyze the medium of open feedback along with the lines making the decision pattern can be carried out by the algorithm. Regular using of data mining and traditional algorithm strategies can similarly assist in predicting the overall execution structure of an organization.

IX. ACKNOWLEDGMENT

The work was supported by the University Grants Commission (UGC), Ministry of Education, India. No.3674(NET- DEC2018)

REFERENCES

- [1] Pavel Ciaian, Miroslava Rajcaniova d'Artis Kancs "The economics of BitCoin price formation" Pages 1799-1815, 13 Nov 2015
- [2] D. Yermack Is Bitcoin a real currency? An economic appraisal. In Handbook of digital currency 2015 Academic Press 31 43 10.1016/B978-0-12-802117-0.00002-3.
- [3] Briere, M., Oosterlinck, K., Szafarz, A. (2015). Virtual currency, tangible return: Portfolio diversification with bitcoin. *Journal of Asset Management*, 16(6), 365–373. <https://doi.org/10.1057/jam.2015.5>
- [4] Begušić, S., Kostanjčar, Z., Stanley, H. E., Podobnik, B. (2018). Scaling properties of extreme price fluctuations in Bitcoin markets. *Physica A: Statistical Mechanics and its Applications*, 510, 400–406. <https://doi.org/10.1016/j.physa.2018.06.131>
- [5] Chuen, D. L. K., Guo, L., Wang, Y. (2017). Cryptocurrency: A new investment opportunity? *The Journal of Alternative Investments*, 20(3), 16–40. <https://doi.org/10.3905/jai.2018>.
- [6] Shah, Devavrat, and Kang Zhang. "Bayesian regression and Bitcoin." arXiv preprint arXiv:1410.1231 (2014).
- [7] Madan, Isaac, Shaurya Saluja, and Aojia Zhao. "Automated bitcoin trading via machine learning algorithms." URL: <http://cs229.stanford.edu/proj2014/Isaac>
- [8] S. Selvin, Vijayakumar, R., Dr. E. A. Gopalakrishnan, Menon, V. K., and Dr. Soman K. P., "Stock Price Prediction using LSTM, RNN, and CNN-sliding Window Model", in 2017 International Conference on Advances in Computing, Communications, and Informatics (ICACCI), 2017.
- [9] Manish Agrawal, Asif Ullah Khan, Piyush Kumar Shukla. Stock Price Prediction using Technical Indicators: A Predictive Model using Optimal Deep Learning 2019.
- [10] Qasem A. Al-radaideh, Adel Abu Assaf, Eman Alnagi. Predicting Stock Prices Using Data Mining Techniques
- [11] Justin Cheng, Lada A. Adamic, Jure Leskovec, Jon Kleinberg, "Prediction of the Cascades", Vol. 7, Nov 2014.
- [12] Pradeep Singh, Satish Chand, "Prediction of the popularity of rumors", p.12, Dec 2018
- [13] Dancho Panovski, Titus Zaharia "Bus Arrival Time Prediction with Traffic Density Matrix" p.16, Vol. 7, Mar 2015
- [14] Ayush Varma, Abhijit Sarma, Sagar Doshi, Rohini Nair, "House Price Prediction", Vol. 3, Nov 2019
- [15] Han Seung Jang, Kuk Yeol Bae, Hong-Shik Park, Dan Keun Sung, "Solar Power Prediction on Images and Support Vector Machine", Vol. 1, Nov 2017
- [16] Gabriel, Cheng Soon Ong "Weighted Regularization Extreme Machine for Imbalance Learning in Software Fault Prediction", p.11, Vol.4, Jan 2019
- [17] Poonam L. Patil and S. R. Jadhao "User Personality Prediction on Social Media", p.5, Vol. 3, Sep 2014
- [18] Berina Alić, Lejla Gurbeta, Almir Badnjević Classification of Diabetes and cardiovascular diseases, p.7, Vol. 1, Mar 2019
- [19] Jue Wang, Anoop Cherian, "Discriminative video representation", Vol. 2, Nov 2020
- [20] www.cs.princeton.edu/sites/default/files/uploads/Saahil_magde.pdf
- [21] Hakob GRIGORYAN, "A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)", DSJ 2016
- [22] Raut Sushrut Deepak, Shinde Isha Uday, Dr. D. Malathi, "Machine Learning Approach In Stock Market Prediction", IJPAM 2017.
- [23] Pei-Yuan Zhou, Keith C.C. Chan, Member, IEEE, and Carol Xiaojuan Ou, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining", IEEE 2018.
- [24] K. Hiba Sadia, Aditya Sharma, Adarsh Paul, Sarmistha Padhi, Saurav Sanyal. Stock Market Prediction Using Machine Learning Algorithms 2019
- [25] Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan. Stock Market Prediction Using Machine Learning 2019
- [26] Aparna Nayak, M. M. Manohara Pai and Radhika M. Pai. Prediction Models for Indian Stock Market 2016