OPEN ACCESS
Indo-JC
Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 7, Issue. 1, April 2022. pp. 51-60
doi:10.34818/indojc.2022.7.1.612

Stock Market Price Forecasting Using Recurrent Neural Network

Pragya Bhardwaj [#1], Jayant Kwatra [*2]
Programmer Analyst Trainee, Cognizant Technology Solutions, Pune, India
**Welingkar Institue of Management, Mumbai, India*
[1]pragyabhardwaj.2@gmail.com
[2]jayantkwatra.2@gmail.com

ABSTRACT

A stock refers to the ownership of the organisation and its investors. A market where these stocks are sold or purchased is known as stock market. The prices of the stock is listed over National Stock Exchange or Bombay Stock Exchange for all Indian Companies. In this work, a machine learning approach is used to predict and forecast the prices of a company listed in NSE and BSE for 30 days using recurrent neural network known as stacked long-short term memory model. The results show that the model worked highly effective in performing the task. The model in the evaluation phase gave a root mean square error of 3.00 on the training data, 0.03 on testing data. R2 score for training data was 0.99 and 0.97 for the testing data. The prices when compared by the client organisation showed that they matched the predicted values to upto 90%. Thus, stacked LSTM models are one of the best models to make predictions of stock related data.

Keywords –LSTM, Machine Learning, Recurrent Neural Network, Stock Market,

I. INTRODUCTION

A place where people perform selling or buying of shares of a company that is listed publicly is known as Stock Market. In general terms, an aggregation of sellers and buyers of shares or stocks is known as a stock market. This is also known as representative of claims on ownership of organisations dealing with businesses which can further include public or private securities[1]. Any investment that intends to be performed in a stock market can be done using brokers or platforms that are established online, for eg: Upstocks, Trading Giant etc. A stock market works in a manner where there is a network of exchanges, examples could be Nasdaq, New York Stock Exchange. An initial public offering or IPO is being opened by the company for the investors. Any organisation opens an IPO so as to increase its business and maximise the profit. After an IPO is opened, the investors can either buy or sell the shares based on the supply and demand, any buyer or seller initially offers a bid for the stock and for facilitating the trade, a buyer has to increase the price and seller decreases the price[10]. This task is too complicated and complex, performing it on your own is not easy. Thus, complex computer algorithms are used so that the people who are engaged in this market can have high profit and incurred low losses. In this work, a stacked LSTM model is used to forecast the prices of closing value for a NSE and BSE listed company for 30 days.

II. LITERATURE REVIEW

Multiple studies are used to predict the prices of financial markets. Machine learning plays a vital role in producing different predictions models. Linear models include SVM, gradient boosting techniques, decision trees and artificial neural networks. However, they are less preferred when it comes to prediction of stock prices as they cannot handle the multicollinearity of the data. In this scenario comes the non-linear models - deep neural nets that comprise CNN, RNN. The main objective of this paper is to analyse the performance of LSTM for forecasting the closing prices of an asset portfolio. A LSTM model is capable of taking information from the past and forecast the future based on that information. A typical ANN comprises three layers - input, hidden and the output layer In these models the task of memorisation is done using the gates present through the memory lines that are being present. Therefore, the capability of an LSTM model to remember the patterns of the data makes them a very special kind of RNN model. Each node present in this model comprises cells that are responsible for the storage of streams of data that are passed [1]. The study develops a RNN model - LSTM on the share price of a company

Received on Januari, 2022. Accepted on March, 2022

for making predictions on a particular period of time. The data is taken from NSE India. NSE is an Indian stock exchange market that provides investors a facility for their investments. The features include closing prices, opening price, high day prices, low day prices, trading date, Total Quantity traded, prices of previous day and the turnover amount. As per the study each LSTM model is made up of an information door, an entrance and a view door. The data used in this work comprises 3000 records. The loss incurred is 0.0024. The prediction made by the model on the 300th day was Rs. 166 and the actual price was Rs. 172 [2]. Making predictions about the future stock prices is one of the key areas of research. However, when it comes to such a volatile market the trust of the investors is quite less. But with the advent of deep learning models based on the performance on the past data, near close to actual values. This study used the data from NSE for 5.5 years. This work collates the results obtained by the researchers from different models which include - multivariate regression analysis, Decision Tree Regression Result, Boosting regression, Bagging regression, Random Forest regression, ANN and SVM. This study concluded that the results generated by the LSTM models work fine in the dataset provided as compared to other models [3]. Domains related to finance, demand, market analysis comprises time series data which is used to find the pattern in the past data and predict the future accordingly. A time series analysis method can be grouped into two broad categories - statistical methods and machine learning model methods. Though, as of now any so-called "reliable" method is not developed for forecasting these data, RNN's have proved to be much more accurate than other methods. RNN's fall into the category of machine learning models and are extensively used in predictions, classification and forecasting. There are also hybrid models that are working highly effectively on the problem of modern day. The study discusses different models used for forecasting the time series data. It describes all the mathematical information related to recurrent neural networks [4]. This research performs a forecasting of stock market indexes using a novel approach by developing a hybrid model. The authors have used CNN along with the LSTM structure. CNN are very powerful networks in extracting the features. Each layer performs extraction of features, the hybrid models made by the authors show different results on the data. The models proposed by the authors have shown the highest accuracy on both one week and one month data. This shows that the proposed model is effective in making predictions of the stock market [5]. The evolution of technology has facilitated the analysts and data scientists across the countries to make better return off investments pertaining to the stock market. This study uses the data of the past for predicting the future and then making a comparative study of the model presented with the past approaches. The researchers have used a random forest and stacked LSTM for projecting the output. They have used the hardware setup of the i7 processor with a RAM of 8GB. The data used is from a time period of one year i.e July-2018 - July 2019 using python programming language and environment setup in jupyter. The model has been developed in three stages - preprocessing data, model building and model evaluation. Data preprocessing involved standardisation of the closing price of stock value in a range 0-1 using the formula:

$$\mathbf{x' = (x1 - mean(x1))/std(x1)} \quad \mathbf{(1)}$$

The model requires a timestamp value that will determine the number of t past data input to be used to predict t+1 output value. This study uses t value as 60 and prediction is made for the 31st Day. The output layer comprises an activation function which is Relu as the nature of the data is non-linear. Adam optimizer is used in this model. Inorder to remove the problem of overfitting, dropout layer is used where some amount of neurons are removed from the model. Evaluation is done using RMSE metric. Based on the computations and results, LSTM has shown much better results as compared to MLP[6]. This study is a review of the recent advancements made for predicting the stock market prices using different machine learning and deep learning models. Most important aspect is finding the depth of the data that is needed to be analysed, short t(erm data aren't effective in analysing the stock prices. Thus, quality and historical data is needed to find the insights. Preprocessing of such data involves methods - missing data corrections, removal of noise and extraction of features. The model for building used in these papers are - convolution neural network and recurrent neural network. Further improvement of these deep learning models, restricted boltzmann machine and sequence to sequence machine. Probable machine learning models - linear regression, autoregressive integrated moving average, generalised autoregressive conditional heteroskedasticity, logistic regression, support vector machine, k-nn. The model evaluation is done using multiple metrics, some can be used in prediction models are: classification reports, regression evaluation metrics like root mean square, mean square error, normalise mean square error, r2 score, profit analysis, significance analysis, accuracy, precision, recall and sensitivity analysis [7]. In this study, authors have done a comparative study of different regressors models like boosting, bagging, blending and superstacking. The data is taken from Ghana stock exchange, johannesburg stock exchange, bombay stock exchange and new york stock exchange from the period of 6 years that include January 2012 to December 2018. The stock market is one the biggest source of data, it generates structured and unstructured data in huge volume everyday. The baseline of this paper is the use of traditional machine learning algorithms like decision trees and support vector machines. The ensemble techniques used in this paper are weighted average, max voting and averaging. The advanced ensemble techniques are: bagging, boosting, stacking and blending. The steps in this study are as follows: Preprocessing Data, Building heterogeneous and homogeneous ensemble classifiers and regressor model and Evaluation of model. Data cleaning or preprocessing is done using the wavelet transform. The formula for which is as follows:

$$X_{\omega}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\varphi\left(\frac{t-b}{a}\right)dt \tag{2}$$

The transformation of the data is done using the following formula:

$$b' = \frac{b - b_{\min}}{b_{\max} - b_{\min}} \tag{3}$$

The models developed are twenty five in total and results showcase that homogeneous models have proven to be better in predictions as compared to the heterogeneous models [8]. This paper presents a time series prediction using LSTM stacked auto encoders, 2-Dimensional CNN, and wavelet transforms. These models are combined together to perform the task proposed by the authors. The preprocessing or denoising of the data is done by using wavelet decomposition. Analysis of financial markets comprises analysing data from the stock prices, defining loan rates, exchange rates or predicting stock market indices. There are numerous models proposed and worked for analysing this data that includes both linear and non-linear models. Linear models comprise ARIMA - auto regressive integrated moving average, SVM's, statistical analysis etc. With the advancements in technology and rapid developments in the area's of deep learning - a subset of AI, non linear models like neural net, LSTM, convolution neural networks are also being widely used. These models have proven to be more effective in varied domains like driving assistant systems, traffic management system, anomaly detections and the time series data forecasting. The proposed model outperformed in terms of accuracy and other metrics. The paper concluded by showcasing that testing the proposed model on these datasets have proven that this model can be highly effective in predictions of time series data using the approach [9]. This study gives a comprehensive understanding of the long short term memory deep learning model. The authors have given a comprehensive understanding about how, where and when to use these networks. LSTM models developed in the late 90's are one of the most powerful recurrent neural networks. These models solve the issue of persistence of information but rather stores information for longer periods of time. They consist of input, output and hidden layers and in the hidden layer they have memory gates or the memory unit. This unit also comprises three gates within itself known as the input, output and hidden gate. In other words, these networks are recurrently connected sub-network sets that are known as memory blocks. The main work of the memory block is to maintain a particular state at a time and then regulate the information flow [10]. This paper uses the LSTM model for performing stock market analysis. The weights are trained and adjustments are made for individual data points using the stochastic gradient descent approach. The paper also shares the importance of online and batch learning algorithms. As per the study in an online learning process, the optimization step can be stopped during the middle of learning, which is indeed helpful for datasets that consist of large records. Whereas, a batch learning involves keeping the system weights at a constant value and finding the error with each data point. As per the authors, LSTM is an ideal model to investigate variation in one price of stock affecting other prices. They also help in making adequate decisions about the time duration by which the prices predicted are reliable to in understanding the trends in future. This paper uses the LSTM model for performing stock market analysis. The weights are trained and adjustments are made for individual data points using the stochastic gradient descent approach. The paper also shares the importance of online and batch learning algorithms. As per the study in an online learning process, the optimization step can be stopped during the middle of learning, which is indeed helpful for datasets that consist of large records. Whereas, a batch learning involves keeping the system weights at a constant value and finding the error with each data point. As per the authors, LSTM is an ideal model to investigate variation in one price of stock affecting other prices. They also help in making adequate decisions about the time duration by which the prices predicted are reliable to in understanding the trends in future. These results show that when a prediction model is created using LSTM and ANN, better results are obtained in the LSTM model as per the paper. The paper concludes that with more addition of the data, an in depth analysis of the pattern could be obtained with adjustment in the weight layers [11].

III. PROPOSED WORK

The objective of this work is to develop a stacked-LSTM model that can forecast the stock closing value from the data provided for 30 days based on the past trends. The following figure shows the flow of work of the proposed model.

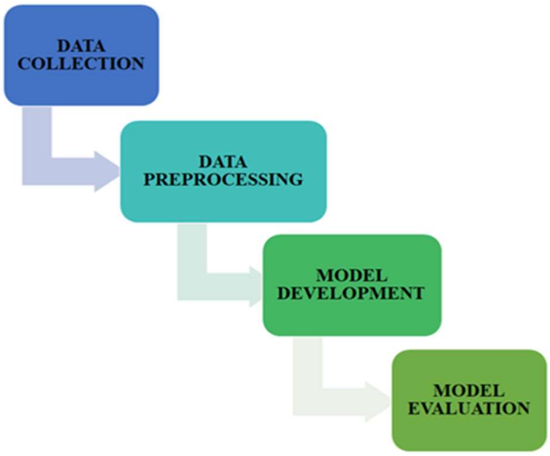

Fig 1: Design flow of proposed model

The proposed model is divided into following steps.

- **Data Collection**

Data has been provided by the client and comprises the stock price values from the year 2015-2020.

- **Data Pre-processing**

Data has been pre-processed which included removal of null values (if any) and then making the value into a standardised form using the Min - Max scalar function. This function transforms the target value in the range of 0-1[13]. This step is important and integral as it helps in stable and fast execution when the model is created. The formula for min - max scaling is as follows:

$$\mathbf{X\ (scaled) = x\text{-}min(x)/x\text{-}max(x)}$$

- **Model Establishment**

A stacked LSTM model is used based upon the client requirement. These networks were introduced by Schmidhuber Hochreiter and are those special categories of RNN's that are able to learn dependencies that are long. They are widely used in many areas due to their feature of remembering the information taken from a specific data for a long period of time. They form a structure of a chained neural network and comprises a repetitive module which has four layers that interact with each other in a totally different way. Fig below represents the stacked LSTM model.

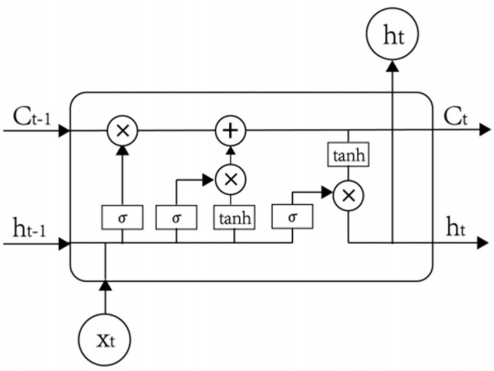

Fig 2: Long Short Term Model Structure

- **Model Evaluation**

The model is further evaluated using Mean-Squared Error and Root Mean Square Evaluation metrics. The Mean square error is referred to as an average of the errors present in the model. It helps us compute the difference between estimated value and the actual value. The formula for MSE is as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

The root mean square error or RMSE is the square root of MSE. The formula for RMSE is as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}$$

IV. DESIGN

A. Proposed Model Design

The proposed model intends to develop a stacked-LSTM model that can predict the stock closing value from the data. Below fig represents the design flow of the proposed model:

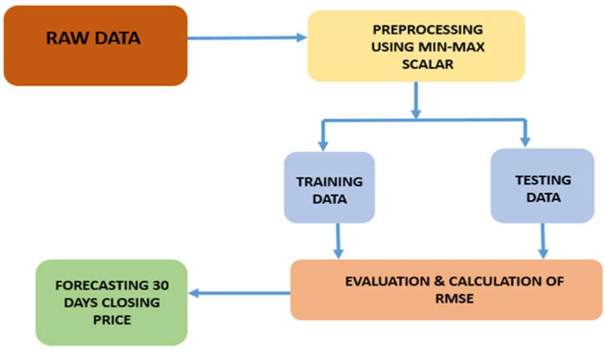

Fig 3: Design flow of proposed model

The dataset has been provided by the company and comprises 1257 records collected from the year May 2015 - May 2020. The various attributes present in the data are as follows:
- Date
- Close
- Open
- High
- Low

B. IMPLEMENTATION

1. Importing data using pandas in google collab.
2. EDA is performed on the data and relevant checks for null values and other vital information like column datatypes and number of records is seen. It is clear from here that the data has no null values and the total records are 1258 wherein each attribute's data type is shown above.
3. The work involves taking the closing price value for the company stock, thus the close index is taken from the data and plotted.

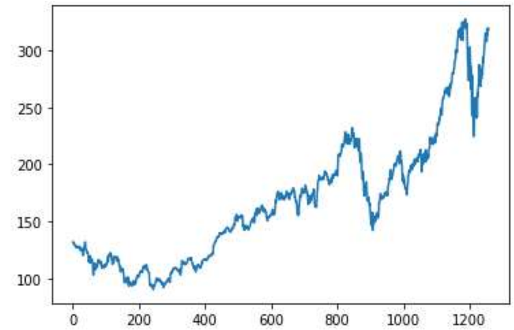

Fig 4: Visualisation of closing price of the historical data

4. As the closing prices have different scales and require preprocessing. For standardising and normalising the values to a scale 0-1 using the min-max scalar.
5. Before, building the model data is splited into training and testing using the scikit learn in the library. 65% of the data is taken for training and remaining data is taken for testing.
6. A LSTM model requires a timestamp value, here we have taken the value for timestamp as 100, this means that 100 days data would be covered for training the model.
7. After developing the dataset, based on the timestamp values, the total data in the training and testing is as follows: X_train: 716,100 , Y_train: 716, X_test: 340,100, Ytest: 340
8. The stacked LSTM model is now being developed using tensorflow and keras. Model is sequential and three layers of LSTM are created. Loss value is calculated by taking the RMSE and the optimizer used is Adam.
9. The data is trained with 100 epochs and a batch size of 64. An epoch is the time the algorithm will work on the training datasets. A batch refers to the internal number of samples to be covered before making any update to internal parameters.
10. After the training is completed, the relevant predictions are made on the testing data and training data.
11. At the initial stages, transformation was done for standardisation values, now an inverse transform is being done to bring the values to normal state and generating final results.
12. The model is evaluated using the root mean square error metric and r2 score. The values are shown in the results section.
13. Now the forecasting of the price value for the next 30 days is being done. The values predicted from the above code are plotted using the matplolib library and is shown in the figure below.

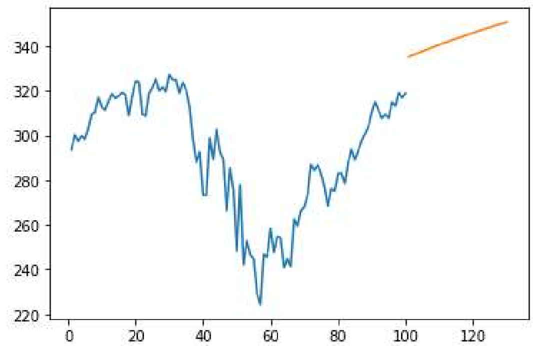

Fig 5: Visualisation of forecasted values

V. RESULT

The model is evaluated on the basis of RMSE values and R2 score. The root mean square error and r2 score found after the building of the model is as follows:

Table 1: Results on model evaluation

METRIC	TRAINING DATA	TESTING DATA
RMSE	3.00	0.03
R2 Score	0.99	0.97

This shows that the model has outperformed on the given data.

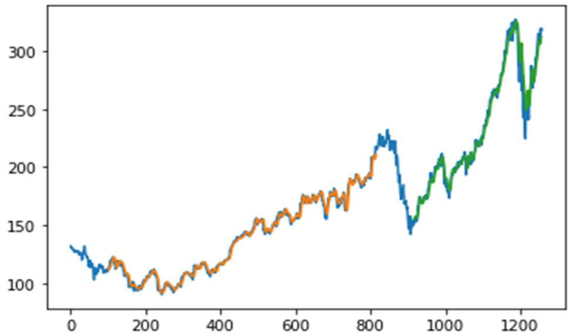

Fig 6: Past values, predicted values on training and testing data

The blue graph represents the visualisation of the stock price value as predicted by the model, the blue line indicates the past values or the complete dataset, the green line represents the predictions of the test data and the orange line indicates the train data predictions. As it is seen that the model is well fitted on the actual values and has performed extremely well in predicting the prices. The below graph represents the forecasted values for the client company and thus we can see that the closing price is now increasing and the predictions made by the model are in coherence with the actual data.

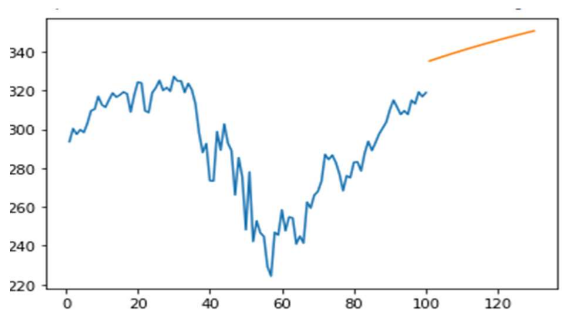

Fig 7: Final Forecasted Stock Values

VI. CONCLUSION

Financial markets are considered to be one of the most volatile markets as they have sensitive information with multiple features. In olden days due to less advancement and technological constraints, analysts used to perform predictions on their own. However, with the advent of machine learning and deep learning models, pricing predictions and forecasting have become a lot easier. Multiple machine learning models are used to predict the prices of stocks and with the removal of old technological constraints deep learning is proven to be one of the most effective technologies. Some of the deep learning models like ANN's, CNN's and RNN's are used highly in this domain. With the emergence of RNN, the forecasting of time series data has become a lot easier. In this project, a stacked LSTM model was used to forecast the stock prices for the client company. The results show that the model worked highly effective in performing the task. The model in the evaluation phase gave a root mean square error of 3.00 on the training data, 0.03 on testing data. R2 score for training data was 0.99 and 0.97 for the testing data. The prices when compared by the client organisation showed that they matched the predicted values to upto 90%. Thus, stacked LSTM models are one of the best models to make predictions of stock related data.

ACKNOWLEDGEMENT

The authors would like to thank their organizations for providing them the opportunity to work on this problem statement and coming up with a solution and also present their thanks for the same.

REFERENCES

[1]. Budiharto, W. (2021). Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM). *Journal of big data*, *8*(1), 1-9. DOI - https://doi.org/10.1186/s40537-021-00430-0.

[2]. Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, *170*, 1168-1173.DOI - https://doi.org/10.1016/j.procs.2020.03.049

[3]. Pramod, B. S., & PM, M. S. (2020). Stock Price Prediction Using LSTM. Available at - https://www.researchgate.net/profile/MallikarjunaPm/publication/348390803_Stock_Price_Prediction_Using_LSTM/links/5ffc6a23a6fdccdcb84a20f8/Stock-Price-Prediction-Using-LSTM.pdf

[4]. Mehtab, S., & Sen, J. (2020, November). Stock price prediction using CNN and LSTM-based deep learning models. In *2020 International Conference on Decision Aid Sciences and Application (DASA)* (pp. 447-453). IEEE DOI - 10.1109/DASA51403.2020.9317207.

[5]. Elsworth, S., & Güttel, S. (2020). Time series forecasting using LSTM networks:A symbolic approach. a rXiv preprint arXiv:2003.05672. Available at - https://arxiv.org/pdf/2003.05672.pdf.

[6] Hao, Y., & Gao, Q. (2020). Predicting the trend of stock market index using the hybrid neural network based on multiple time scale feature learning. Applied Sciences, 10(11), 396.
DOI- https://doi.org/10.3390/app10113961.

[7] Indian Stock-Market Prediction using Stacked LSTM AND Multi-Layered Perceptron Siddharth Banyal, Pushkar Goel, Deepank Grover – 2020. DOI - 10.35940/ijitee.C8026.019320

[8]. Jiang, W. (2020). Applications of deep learning in stock market prediction: recent progress. a rXiv preprint arXiv:2003.01859.DOI - https://doi.org/10.1016/j.eswa.2021.115537

[9]. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. J ournal of Big Data, 7 (1), 1-40. DOI- https://doi.org/10.1186/s40537-020-00299-5

[10]. Essien, A., & Giannetti, C. (2019, July). A deep learning framework for univariate time series prediction using convolutional LSTM stacked autoencoders. In 2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA) (pp. 1-6). IEEE. DOI - 10.1109/INISTA.2019.8778417.

[11]. Van Houdt, G., Mosquera, C., & Napoles, G. (2020). A review on the long short-term memory model. Artificial Intelligence Review, 5 3, 5929-5955. DOI- 10.1007/s10462-020-09838-1

[12]. Ojo, S. O., Owolawi, P. A., Mphahlele, M., & Adisa, J. A. (2019, November). Stock market behaviour prediction using stacked LSTM networks. In 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (pp. 1-5). IEEE. DOI - 10.1109/IMITEC45504.2019.9015840.

[13] Nandakumar, R., Uttamraj, K. R., Vishal, R., & Lokeswari, Y. V. (2018). Stock price prediction using long short term memory. International Research Journal of Engineering and Technology, 5 (03). Available at - https://www.irjet.net/archives/V5/i3/IRJET-V5I3788.pdf

[14]. Gao, S. E., Lin, B. S., & Wang, C. M. (2018, December). Share price trend prediction using CRNN with LSTM structure. In 2018 International Symposium on Computer, Consumer and Control (IS3C) (pp. 10-13). IEEE. DOI - https://doi.org/10.1080/23080477.2019.1605474

[15]. Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. Procedia computer science, 132, 1351-1362 DOI -https://doi.org/10.1016/j.procs.2018.05.050.

[16]. Althelaya, K. A., El-Alfy, E. S. M., & Mohammed, S. (2018, April). Evaluation of bidirectional LSTM for short-and long-term stock market prediction. In 2018 9th international conference on information and communication systems (ICICS) (pp.151-156). IEEE. DOI - 10.1109/IACS.2018.8355458.

[17]. Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: deep portfolios. Applied Stochastic Models in Business and Industry, 3 3(1), 3-12. DOI - https://doi.org/10.1002/asmb.2209

[18]. Budhani, N., Jha, C. K., & Budhani, S. K. (2014, August). Prediction of stock market using artificial neural network. In 2014 International Conference of Soft Computing Techniques for Engineering and Technology (ICSCTET) (pp. 1-8). IEEE. DOI - 10.1109/ICSCTET.2015.7371196.

[19]. Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. Expert systems with applications, 112,353-371. DOI - https://doi.org/10.1016/j.eswa.2018.06.032

[20]. Gharehchopogh, F. S., Bonab, T. H., & Khaze, S. R. (2013). A linear regression approach to prediction of stock market trading volume: a case study. International Journal of Managing Value and Supply Chains, 4(3), 25. DOI - 10.5121/ijmvsc.2013.4303

[21] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. DOI - https://doi.org/10.1007/978-3-030-82014-5_28

[22]. https://colah.github.io/posts/2015-08-Understanding-LSTMs/