Scientific Journal of Impact Factor (SJIF): 5.71
E-ISSN (O): 2348-4470
P-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 10, Issue 02, February 2023

A Review Paper on Forecasting of Stock Prices using Machine Intelligence

Hitesh Momaya[1], Venus Patel[2], Vansh Momaya[3]

[1]PG student, U. V. Patel College of Engineering, Ganpat University, [1]hitesh.momaya@ganpatuniversity.ac.in
[2]Asst. Professor, U. V. Patel College of Engineering, Ganpat University, [2]vrp02@ganpatuniversity.ac.in
[3]Student, B. S. Patel Polytechnic, Ganpat University, [3]vanshmomaya21@gnu.ac.in

Abstact – *The emerging trends in the fundamental perceptions of information technology have changed the route of stockholders, where they can buy and sell stocks through electronic media. Nowadays, most stockholder connects with online trading platforms. In this context, forecasting the new price of selected stocks has an excellent achievement for the systems analyst. The paper's objective focuses on forecasting prices for individual stocks, sectors, and index values of the Indian and global stock markets. In the recent era, the modern approach to forecasting stock values has a deep learning-based approach and sentiment analysis. The generic process of stock price prediction has three strategies: historical data, data pre-processing, and deep learning model training. For the calculation of the loss function, the actual stock price collects from the historical data set, and the predicted stock price is by training the desired model. Furthermore, the comparison performs with several deep learning techniques which results in better accuracy for stock as well as Index values.*

Keywords-*Stock Forecasting, Deep Learning, Support Vector Regression, Convolution Neural Network, LSTM*

Paper Received:*05/02/2023* ***Paper Accepted:****17/02/2023* ***Paper Published:*** *04/03/2023*

I. INTRODUCTION

Investors or traders can perform their investment strategies at the stock exchange by purchasing or selling shares or contracts at a mutually settled price. However, it has a risk of volatility, inflation, and global market outcomes that depend on many factors, such as demand and supply, government strategy, and fiscal indicators. The emerging trends in the fundamental perceptions of information technology have changed the route of stockholders, where they can buy and sell stocks through electronic media. The shareholders can earn good returns by receiving regular dividends or selling their shares at an increased rate than they purchased. Nowadays, most stockholders linked with online trading platforms have become a center of attention for software developers, those who are concerned with technological developments. Thus, stakeholders search for algorithm strategies, auto-trading software, and methods that improve the risk-reward ratios.

Forecasting the trend and next price level is an exciting task for effective methodology. The easy and active estimate model helps investors and experts to predict the stock trend. Thus, earning maximum gains with stock trading is an optimized feature. However, the non-linear characteristics and liability to rapid change in the price level, therefore, forecasting the new price of selected stocks has an excellent achievement for the systems analyst. With technological advances, software developers can prepare a model for the investor to anticipate the development of the industry. Based on historical price data combined with current data, the sentiments on television media data can survey to forecast the momentum and trend to earn greater returns.

Thus, system analysts and software developers work in area of neural networks [11]. Therefore, the performance of computer algorithm systems relies mainly on quality. Researchers have used specific strategies to enhance the stock-explicit features, and more attention is needed to feature extraction and selection mechanisms. This paper includes a historical dataset, duration for time series analysis, proposed techniques, and calculation of loss functions. It has five categories: s*ection I* suggest an introduction to the stock market and the need for a software developer for an online trading platform; section *II* includes a literature survey based on the past dataset and proposed techniques; section *III* identifies methodology based on various regression and deep learning models, *Section IV* gives an overall comparison with statical data, and *Section V* represents the conclusion of each author and suggested future scope in preparation of forecasting model for the stock or indices.

1.1 Classical Approach

In formulating the prediction, the classical approach [6] to forecast the next price of the industry is fundamental and technical analysis. Moreover, the fundamental analysis of the stock is about analyzing the business model and balance sheet to understand its current valuation and the proposed growth potential of the industry. However, technical analysis plays a pivotal role in predicting the next price of stock performance with

the help of historical data, real-time data, volume, trend, price momentum, charts, and recent news on media channels.

1.2 Modern Approach

However, the stock market has variable nature; hence it is not easy to prepare a forecasting model. In the era of digitization, machine learning-based algorithms discover data patterns to generate accurate results from many structured and unstructured, heterogeneous data sets. The modern approach [11] for predicting stock values has a machine learning-based approach and sentiment analysis.

II. LITERATURE REVIEW

The literature survey focuses on predicting prices for individual stocks, sectors, and index values of the Indian and global stock markets. The research was performed for independent stocks by [1], [3], [8], [10]. The authors of [1] suggested an easy and effectual approach for prediction on Microsoft Corporation using back-propagation with feed-forward architecture. They performed on one year of data in two strategies; firstly, by updating weights and bias, and secondly, in batch mode with 113 epochs. The authors concluded that ANN could not predict exact values but could be used to find the stock trend.

A paper [3] by Bruno Miranda Henrique et al., 2018 proposed to predict the stock price of global indices using a moving training window strategy based on average return and volatility. They performed on a fixed training set on a daily price of two years of data; they also suggested that SVR predicted price using the random walk-based model results in better accuracy. In a recent paper by [10], Madhumita, G. et al., 2022 proposed to predict the stock price of Google using Regression Techniques and k-Nearest Neighbour algorithms. They performed on two algorithms, SVR and k-NN, on a daily price of one year of data, also authors suggested that SVR using three kernels predicts price with better accuracy.

The authors of [8] Dilip Singh et al., 2021 proposed changes in the closing price of Tata Power as compared to its previous daily closing price with variable window size. They performed on a closing price of ten years of data. They built the input layer using LSTM-RNN, whereas the output layer used Dense Network. RMSE as a loss function with different sizes of rolling windows. They suggested that the rolling window concept with different window sizes improves accuracy. Next, the model was performed for other sectors by [4]; Achyut Ghosh et al., 2019 analyzed stocks of different sectors with the best period for predicting future prices using the LSTM-based model. They performed on the closing price of three years of data for various sectors like IT, Pharma, FMCG, Aviation, and Bank on long-time historical data. The calculation best results of average error value for the banking sector. They suggested applying LSTM based model on long-time historical data.

Next, the research was performed for NIFTY 50 index values by [6], [9] and the global index by [2]. The authors of [6] Sidra Mehtab et al., 2020 proposed an optimized approach for predicting the NIFTY 50 index values using five DL-based regression models as a hybrid approach. They took more than five years of data, but 60% of the records obtained the best performance of the model for training and 40% of records for testing a model with two weeks of data. They also suggested that univariate encoder-decoder convolutional LSTM was the most accurate model. In a paper [9], Mehtab S. et al., 2020 proposed a robust and reliable framework to predict the price of the NIFTY 50 index using four approaches in LSTM network-based models for medium-term investors. They collected four years of data. They performed with the hybrid approach for ML and DL regression models. They also suggested that the performance of the LSTM-based deep learning regression models was superior.

Hakob Grigoryan et al., 2017 [2] proposed a hybrid approach to detecting trend and directional changes in DJIA Index. He performed with a dataset of 400 days with price, volume, and seven technical indicators. He examines the correctness of the integrated model using a peeling algorithm and a SVM. Adil Moghar et al. (2020) [7] proposed RNN based LSTM model for GOOGLE and NKE with large historical data sets of about 15 years. They performed by increasing the number of epochs to reduce loss function.

III. THE GENERIC PROCESS

The generic process of stock price prediction has three strategies, type of data, data pre-processing, and machine learning model. Here, the data types may be historical price or text data. In machine learning algorithms, to train a dataset, 80% of records are to train a model, and the rest of 20% of records collected to test a model. Further, regression techniques prefer individual stock with long historic data for designing an accurate model, LSTM models prefer different sectors, and a hybrid approach of CNN with LSTM results in better accuracy for Index values.

For the calculation of the loss function [5], the actual stock price collects from the historical data set, and the predicted stock price is by training the desired model. Here, we have to calculate the relationship between the actual and predicted prices in every statistical analysis to estimate the model's accuracy.

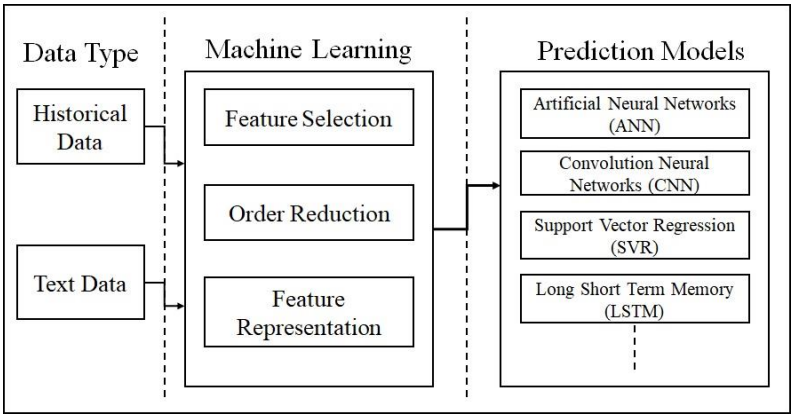

Fig.1 The Generic Process

IV. LITERATURE COMPARISON

Reference	Historical Dataset	Dataset Duration	Techniques Proposed	Result	Conclusion
[1]	Microsoft Corporation	1 Year	ANN	Accuracy 98% MSE = 0.043	ANN could not predict exact values but could be used to find the stock trend
[2]	NIFTY 50 index	2 Weeks	CNN, LSTM	RMSE = 0.035	Univariate encoder-decoder convolutional LSTM was the most accurate model.
[3]	Brazilian, American, and Chinese stocks	2 Years	SVR	RMSE = 0.411	Support Vector Regression (SVR) predicted price using the random walk-based model results in better accuracy.
[4]	Google Stock	1 Year	SVR, KNN	RMSE = 12.4582 Accuracy= 99.37%	Support Vector Regression (SVR) using three kernels to predict price with better accuracy.
[6]	NIFTY 50 index	6 Years	LSTM	RMSE = 0.1711	Theysuggested model of regression was good.
[7]	Google, NKE	15 Years 10 Years	RNN, LSTM	RMSE = 4.97E-04 RMSE = 8.74E-04	By increasing the number of epochs,maximize the predictive accuracy
[8]	Sectoral Index	3 Years	LSTN	Accuracy 98% MSE = 0.043	By applying LSTM based model on long-time historic data
[9]	Tata Power	10 Years	RNN, LSTN	RMSE = 0.035	The rolling window concept with different window sizes results in better accuracy.
[10]	Dow Jones Index	400 days	SVM	RMSE = 0.411	Correctness of the integrated model using a peeling algorithm and an SVM

V. METHODOLOGY

Initially, the data collection process justifies the duration of the historical dataset for estimation. For long-term investors, gather historical prices of more than ten years, approximately 2000 data records. For medium-term investors, gather historical prices of about four years, which has about 800 data records. For short-term investors, the historical price of nearly two weeks to two years of data may vary from 10 to 400 data records. Traders may prefer real-time data, but the investor has to collect a few years of a dataset. We can use daily, weekly, or monthly stock prices available on NSE or BSE servers. Here, values of stock, sector index, and

NIFTY 50 are present. Yahoo Finance is one of the most recommended websites for downloading stock prices to train the desired algorithms.

The format of the dataset is CSV file format, which easily supports libraries. It includes open price, high and low price, and close and adjusted close price with the record of volume of the stock and indices. Such data records cannot be referred directly, as there are possibilities of additional data, improper values, and noisy or unwanted text like a bonus, merger, stock split, or company announcements. Hence, to solve such a problem, data preparation is to be performed. The most vital step in learning algorithms is data pre-processing. It is an essential parameter to build the most accurate model. In the learning process of a model, it is necessary to spend about 80% of the time duration on training or data pre-processing and about 20% on the actual analytic performance of the model. The data from stock exchanges or different sources might be in a raw format like missing data, noisy data, and inconsistent data, which is not feasible for the model to train. Therefore, it is necessary to convert them by cleaning, where the initial raw data converts into an actual dataset.

VI. CONCLUSION AND FUTURE SCOPE

The authors of [1] concluded that ANN could not predict exact values but could be used to find the stock trend. In the future, identify best suit methods with comparative training and learning functions. The authors of [2], [3], [9], [10] concluded that the Support Vector Regression (SVR) is a robust algorithm for better accuracy in global stocks and indices. The predicted price using the random walk-based model of regression results in better accuracy [3], [10]. The future scope of the model includes independent variables, time series analysis, and various fundamental indicators for improvement in the learning algorithm. The author also suggested the correctness of the integrated model using a hybrid approach of a peeling algorithm and an SVM [2]. The authors of [4], [6]-[8] compared CNN and LSTM models with different strategies. They concluded that applying LSTM based model on long-time historical data with the rolling window concept results in better accuracy. Also, to improve accuracy in future scope, they suggested a predictive model using generative adversarial networks (GANs) [6], [9]. The authors of [7] select the long dataset with an increasing number of epochs to train a model. Hence, sentiment analysis based on the government policy framework, geopolitical conditions, global news, and elections may consider for a better accurate model.

REFERENCES

[1] K. Abhishek, A. Khairwa, T. Pratap, and S. Prakash, “A stock market prediction model using artificial neural network,” in 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT’12), 2012, pp. 1–5. doi: 10.1109/ICCCNT.2012.6396089.
[2] H. Grigoryan, “Stock market trend prediction using support vector machines and variable selection methods,” in Proceedings of the 2017 International Conference on Applied Mathematics, Modelling and Statistics Application (AMMSA 2017), Atlantis Press, 2017/05, pp. 210–213, isbn: 978-94-6252-355-5. doi: https://doi.org/10.2991/ammsa-17.2017.45.
[3] B. M. Henrique, V. A. Sobreiro, and H. Kimura, “Stock price prediction using support vector regression on daily and up to the minute prices,” The Journal of Finance and Data Science, vol. 4, no. 3, pp. 183–201, 2018, issn: 2405-9188. doi: https://doi.org/10.1016/j.jfds.2018.04.003.
[4] A. Ghosh, S. Bose, G. Maji, N. Debnath, and S. Sen, “Stock price prediction using lstm on indian share market,” in Proceedings of 32nd International Conference on Computer Applications in Industry and Engineering, Q. Yuan, Y. Shi, L. Miller, G. Lee, G. Hu, and T. Goto, Eds., ser. EPiC Series in Computing, vol. 63, EasyChair, 2019, pp. 101–110. doi:10.29007/qgcz.
[5] K. H. Kim T, “Forecasting stock prices with a feature fusion lstm-cnn model using different representations of the same data,” PLoS ONE, vol. 14(2), 2019. doi: https://doi.org/10.1371/journal.pone.0212320.
[6] S. Mehtab and J. Sen, “Stock price prediction using cnn and lstm-based deep learning models,” in 2020 International Conference on Decision Aid Sciences and Application (DASA), 2020, pp. 447–453. doi: 10.1109/DASA51403.2020.9317207.
[7] A. Moghar and M. Hamiche, “Stock market prediction using lstm recurrent neural network,” Procedia Computer Science, vol. 170, pp. 1168–1173, 2020, The 11th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 3rd International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops, issn: 1877-0509. doi: https://doi.org/10.1016/j.procs.2020.03.049.
[8] D. S. et al., “Closing price prediction of nifty stock using lstm with dense network,” International Conference on Advances in Distributed Computing and Machine Learning (ICADCML-2021), Bhubaneswar, 2021.
[9] S. Mehtab, J. Sen, and A. Dutta, “Stock price prediction using machine learning and lstm-based deep learning models,” in Machine Learning and Metaheuristics Algorithms, and Applications, S. M. Thampi, S.

Piramuthu, K.-C. Li, S. Berretti, M. Wozniak, and D. Singh, Eds., Singapore: Springer Singapore, 2021, pp. 88–106, isbn: 978-981-16-0419-5. [Online]. Available: https://doi.org/10.1007/978-981-16-0419-5_8.
[10] M. Ghosh and R. Gor, "Stock price prediction using support vector regression and k-nearest neighbors: A comparison," International Journal of Engineering Science Technologies, vol. 6, no. 4, 1–9, 2022. doi: 10.29121/ijoest.v6.i4.2022.354.5
[11] P. Soni, Y. Tewari, and D. Krishnan, "Machine learning approaches in stock price prediction: A systematic review," Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012 065, 2022. doi: 10.1088/1742-6596/2161/1/012065.

