A SURVEY ON STOCK PRICE PREDICTION USING MACHINE LEARNING

Manavi Mishra, Manjushree Patil, Geetanjali Raut
Student, Department of EXTC
St. John College of Engineering, Palghar
Maharashtra, India

Tushar Chaudhari
Project Guide, Asst. Professor
Department of EXTC
St. John College of Engineering, Palghar
Maharashtra, India

Abstract— **Stock returns are very fluctuating in nature. They rely upon various factors like previous stock prices, current market trends, financial news, etc. To feature their annual income, people have now started watching stock investments as a remunerative option. There are many tools available to investors using technical analysis to form decisions. With expert guidance and intelligent planning, we will almost double our annual income through stock returns. These days, social media has become a mirror. It reflects people's thoughts and opinions on any particular event or news. Sentiments of the general public associated with an organization can have an upshot on its stock prices. This paper surveys various machine learning techniques and algorithms employed to boost the accuracy of stock price prediction.**

Keywords— **Stock prices, Technical analysis, Stock Returns, Social Media, Sentiments**

I. INTRODUCTION

Stock price prediction is one of the foremost important problems faced by many investors. This increases the danger of investing our earned money with the fear of losing it. Hence, many are scared to speculate within the stock exchange. Investing within the Stock markets generates significant revenue for the country. If a successful model for stock prediction is developed, we will gain insight into the market movement over time, spotting the trends that might otherwise not have been observed. Together with the increasing computational power of computers, machine learning is going to be a coherent method to resolve this problem. Our main motivation for this project is to develop a much better stock price prediction system that the investments increases and therefore the investors can maximize their profits.

II. LITERAUTRE SURVEY

1. Stock Market Prediction using Machine Learning Algorithms: A Classification Study

The objective of this paper is to classify different machine learning algorithms. This paper introduced the concept of economic derivatives just like the "no arbitrage" principle and people of the predictive model, like stochastic process theory and efficient market hypothesis (EMH). News articles were analyzed and also the prediction was made on their basis. Steps involved in prediction were data preparation, analysis, aggregation, and visualization. The results were neutral, positive, and negative. Data were collected daily and predictions were made on Logistic Regression. The accuracy obtained was 70%. A generalized linear model (LM) (binomial family) is often used as a logistic regression model. [1]

2. Stock Market Prediction using Machine Learning Techniques

The paper predicts the market performance of the Karachi stock market (KSE) using different machine learning techniques. Various methodologies were discussed like Single Layer Perceptron, Multi-Layer Perceptron, Radial Basis Function, and Support Vector Machines. The technique proposed is to use various factors impacting the market as input attributes for the machine learning model. Many different factors were found to own a major impact on the market performance. [2]

3. Survey of Stock Market Prediction Using Machine Learning Approach

This paper has proposed the survey of a widely known efficient regression approach to predict the stock exchange price from the stock exchange database. Various regression techniques like polynomial regression, rectilinear regression, RBF regression, and sigmoid regression were described. Polynomial regression is that the variety of statistical regression which is that the best fit exchange prediction because it is largely for the non-linear relationship between one dependent and another experimental variable. [3]

4. Stock Market Prediction using Data Mining Techniques

In this paper, the prediction of the exchange trends was done using machine learning models just like the Random Forest model and Support Vector Machine. The historical data, further as sentimental data, were analyzed as they players an important role in market performance. Researchers applied sentiment analysis on Twitter feeds to find the interrelationship among "public sentiment" and therefore the "market sentiment". Data retrieved from Twitter was wont to predict the general public mood. A Self Organizing Fuzzy Neural Network was used on predicted mood from the Twitter feeds and Dow-Jones Industrial Average values from the previous day to predict the movement of the securities market for the longer term. [4]

5. Stock Closing Price Prediction Using Machine Learning

This paper predicted the price of the stocks within the stock market of Thailand (SET). Various methodologies wont to predict the damage of the stock were the Multi-Layer Perceptron model, Support Vector Machine model, and Partial Least Square Classifier. The experimental result proves that Partial Least Square is that the best algorithm of the three algorithms to predict the stock terms. [5]

6. Empirical Study on Stock Market Prediction Using Machine Learning

This paper aims to review the securities market prediction using multiple Traditional, Machine learning, and Deep learning algorithms. together with these algorithms, the survey has also focused on various datasets used for securities market prediction, the features of those datasets were selected as input parameters, and also the evaluation metrics used for comparing the results of predictions. All of these were taken into consideration to create an efficient prediction. [6]

7. DP-LSTM: Differential Privacy-inspired LSTM for Stock Prediction Using Financial News

This paper proposes a novel profound deep neural network DP-LSTM for stock value prediction. It coordinates news stories as restricted data and joins distinctive news sources through the differential security system. In view of the autoregressive moving average model (ARIMA), an estimation ARIMA is assembled by contemplating the data of monetary news stories in the model. At that point, an LSTM-based profound neural organization is outlined. It has three parts: LSTM, VADER model, and differential protection (DP) component. The planned DP-LSTM plan can altogether decrease forecast blemishes and expands vigor. The paper incorporates the profound neural organization with the well-known NLP models (VADER) to distinguish and remove suppositions inside a given content, joining the stock change close cost and compound score to lessen the speculation hazard. In the event that we accept the information separated from the news for expectation completely, we may expand inclination because of some non-target reports. Accordingly, the DP-LSTM improves the power of the forecast model. [7]

8. Predicting NASDAQ and NSE Stocks using Machine Learning Algorithms: ARIMA, LSTM & Linear Regression

This paper predicts NASDAQ (American) and NSE (Indian) stock costs. The Machine Learning calculations utilized were ARIMA, Linear regression, and Long Short Term Memory. These calculations were applied to the verifiable stock information of the previous 2 years of ongoing stock costs. The stock information was gotten from Yahoo Finance API and Alpha Vantage API for NASDAQ and NSE separately. They derived that ARIMA and LSTM models were more compatible for determining NASDAQ (American Company) stocks. For NSE (Indian Company) stocks, LSTM and Linear Regression demonstrated more productivity.

9. Prediction of Stock Market using Machine Learning Algorithms

This paper introduced an examination of AI-supported calculations to assess the stock costs later on to break down market conduct. R factual programming language has been utilized to make the examination. It gathered that Logistic-regression gives exactness in examining market development course and market expectations contrasted with different models like Random Forest, ARIMA, and K-NN. Clarification of these models with their test concentrate on the datasets for various stocks was arranged understanding rate exactness, merits, and demerits. [9]

10. Social Media Monitoring using Sentiment Analysis of Twitter Data

The paper examined the expectation of future stock costs utilizing the assumption esteems for each stock. It focuses on two different techniques, Word2vec and N-gram, for deriving insights from sentiments in tweets. Use of sentiment analysis in addition to machine learning algorithms to tweets extracted from Twitter and analyzing the correlation between the stock market movement of company and sentiments in tweets. It proposes a hybrid approach that combines unsupervised learning to cluster the tweets and then performing supervised learning methods for Classification. The creator applied diverse AI methods: Naive Bayes (NB), Maximum entropy, support vector machine (SVM), and so forth, and presumed that NB, SVM with 89.4% exactness beats different strategies in opinion characterization. It searched for a correlation between Twitter sentiments with stock costs and figured out which words in tweets associate to change in stock cost by doing a post-investigation of value change and tweets. [10]

11. Stock Price Prediction Using ARIMA Model

The creators utilized the ARIMA model to anticipate the stock cost on the information got from New York Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE). They have utilized data sets comprising four features: open, low, close, and high cost. They have accepted the end cost as the objective viewpoint to be anticipated. The explanation for this was that the closing price is the most relevant cost by the day's end. They have exhibited that there is no connection between the autocorrelation capacities (ACFs) and partial autocorrelation capacities (PACFs) utilizing Q-insights and Correlation plots. For the non-fixed information, it was

made fixed with the assistance of differencing methods. It was reasoned towards the finish of the examination that the ARIMA model is exceptionally helpful for the transient forecast. [11]

12. A machine learning model for stock market prediction

In this paper, the proposed model is based on the study of historical data and technical data. It optimizes LS-SVM with a PSO algorithm for daily stock prediction. Levenberg-Marquardt (LM) algorithm is employed as a benchmark for comparison with LS-SVM and LS-SVM-PSO models. This proposed algorithm was tested for many companies belonging to the sector of information & technology, financials, health care, energy, communications, materials, and industries. Five indicators were calculated from the raw datasets such as Relative Strength Index (RSI), Money Flow Index (MFI), Exponential Moving Average (EMA), Stochastic Oscillator (SO), and Moving Average Convergence/Divergence (MACD). LS-SVM-PSO has a better performance compared to other algorithms like LS-SVM and ANN while the ANN-BP algorithm has the highest error value. [12]

13. Stock Market Prediction Using Machine Learning

In this paper, the proposed model is based on the study of historical data and technical data. It streamlines LS-SVM with a PSO calculation for the everyday stock forecasts. Levenberg-Marquardt (LM) calculation is utilized as a benchmark for correlation with LS-SVM and LS-SVM-PSO models. This proposed calculation was tried for some, organizations having a place in the area of data and innovation, financials, medical services, energy, communications, materials, and businesses. Five markers were determined from the crude datasets like Relative Strength Index (RSI), Money Flow Index (MFI), Exponential Moving Average (EMA), Stochastic Oscillator (SO), and Moving Average Convergence/Divergence (MACD). LS-SVM-PSO has a superior exhibition contrasted with different algorithms like LS-SVM and ANN while the ANN-BP algorithm has the highest error value. [13]

14. Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis

This paper proposed a comparative study of nine machine learning models, namely Random Forest, Decision Tree, eXtreme Gradient Boosting, Adaptive Boosting, K-Nearest Neighbors, Support Vector Classifier, Naïve Bayes, Artificial Neural Network and Logistic Regression. Along with this Long Short-Term Memory and Recurrent Neural Networks, the deep learning methods were additionally studied. To all these models, a total of ten technical indicators were provided as input values. Two totally different approaches for inputs were inclosed ie. continuous data and binary data. This investigated the effect of reprocessing. The continuous data inputs used the stock trading data such as open price, close price, high value and low value. While the binary data inputs provided a preprocessing step to convert the previous continuous data into the binary form. [14]

15. NSE Stock Market Prediction Using Deep-Learning Models

In this paper, the authors have used four types of deep learning architectures which are Recurrent Neural Networks, Multilayer Perceptron, Convolutional Neural Network and Long Short-Term Memory. All these architectures were used for predicting the future value of the stock of a company from the historical dataset obtained. The results procured were compared with the ARIMA model. Observations were made that the ARIMA model failed to perform higher when put next with the neural networks. [15]

16. Stock Market Forecasting using Machine Learning: Today and Tomorrow

In this paper, various stock market prediction techniques such as classification, regression, ensemble, evolutionary, deep learning, hybrid, and ABS have been compared. From the results, Random Forest has the highest accuracy of 80.8% for the Zagreb Stock Exchange whereas deep learning has the least MAPE value(0.0002) for the Chinese securities index 10(CSI 300). Each model has an edge over the other but also has some limits. These can be because of the datasets used or because of their evaluating factors. Among all, the models which predicted the results with higher precision were fusion or hybrid algorithm. This was because these algorithms incorporate the important features of various individual techniques that they are made up of and the computational time is also enhanced when compared to other techniques. [16]

Table – 1 Comparative Study

Sr. No.	Title	Authors	Year	Technology Stack
01.	Stock Market Prediction using Machine Learning Algorithms: A Classification Study	M. Misra, A. Yadav, H. Kaur	2018	Linear Regression, SVM, Principle Component Analysis
02.	Stock Market Prediction Using Machine Learning Techniques	N. Sirimevan, I. Mamalgaha, C. Jayasekara,	2019	Sentimental Analysis, Recurrent Neural Network ,LSTM
03.	Survey of stock market prediction using machine learning approach	A. Sharma, D. Bhuriya, U. Singh	2017	Linear Regression Algorithm
04.	Stock market prediction using data mining techniques	S. Maini, K. Govinda	2018	SVM and Random Forest

05.	Stock Closing Price Prediction using Machine Learning Techniques	M. Vijh, D. Chandola, V. Tikkiwal, A. Kumar	2020	Multi-layer Perceptron, SVM, Least square classifier
06.	Empirical Study on Stock Market Prediction Using Machine Learning	R. Sable, S. Goel, P. Chatterjee	2019	K-Nearest Neighbor,SVM
07.	DP-LSTM: Differential Privacy-inspired LSTM for Stock Prediction Using Financial News	Xinyi Li1 , Yinchuan Li2,1 , Hongyang Yang1 , et al	2019	DP-LSTM , ARMA
08.	Predicting NASDAQ and NSE Stocks using Machine Learning Algorithms: ARIMA, LSTM & Linear Regression	A. Narote, K. Jadhav , J. Barot , S. Sawant	2020	LSTM, Linear Regression, ARIMA
09.	Prediction of Stock Market using Machine Learning Algorithms	Neha Bhardwaj, MD Akil Ansari	2019	Logistic-Regression, K-NN, Random Forest, ARIMA
10.	Social Media Monitoring using Sentiment Analysis of Twitter Data	Rupawari Jadhav, M. S. Wakode	2017	Naive Bayes, Random Forest, SVM, Maximum Entropy
11.	Stock Price Prediction Using ARIMA Model	Ayodele A. Adebiyi., Aderemi O. Adewumi, Charles K. Ayo	2014	ARIMA
12.	A machine learning model for stock market prediction	Osman Hegazy , Omar S. Soliman and Mustafa Abdul Salam	2014	PSO, LS-SVM,
13.	Stock Market Prediction Using Machine Learning	Ishita Parmar,Ridam Arora,Sheirsh Saxena, et al	2018	LSTM, Regression Based Model
14.	Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis	M Nabipour, P Nayyeri, H Jabani, Shahab S, A Mosavi	2020	Random Forest, Decision Tree, eXtreme Gradient Boosting, Adaptive Boosting, K-Nearest Neighbors, Support Vector Classifier, Naïve Bayes, Artificial Neural Network, Logistic Regression, Long Short-Term Memory and Recurrent Neural Networks
15.	NSE Stock Market Prediction Using Deep-Learning Models	Gopalakrishnan E A, Hiransha M, Vijay K M, Soman K P	2018	Multilayer Perceptron, Recurrent Neural Networks, Convolutional Neural Network, Long Short-Term Memory
16.	Stock Market Forecasting using Machine Learning: Today and Tomorrow	Sukhman Singh, Tarun Kumar Madan, Jitendra Kumar, Ashutosh Kumar Singh	2019	Classification, Deep learning, hybrid, and ABS Regression, Ensemble, Evolutionary

III. PROPOSED METHODOLOGY

In the proposed methodology we will be working with Google Colab for data analysis. We will work with Open, Close, High, Low, Adj. Close and Traded Volume Parameters. The historical data will be collected from Yahoo Finance API.
Selected Algorithms:
1. Linear Regression
    Linear Regression is a supervised machine learning algorithm. It is used to predict a correlation amongst more than one independent variable.

    $Y0 = b0 + b1*x1$
            Where Y = Dependent variable

X = Independent variable

2. LSTM
    LSTM is an acronym that means Long Short Term Memory networks. This is a special kind of RNN that is capable of learning long-term dependencies. LSTMs are explicitly designed to evade the long-term dependency problem. Remembering information for long periods is peculiarly their default behavior. All recurrent neural networks have the shape of a sequence of repeating modules of neural networks. In standard RNNs, this repeating module will have a really simple structure, like one tanh layer.

3. ARIMA

ARIMA is an acronym that means AutoRegressive Integrated Moving Average, where

AutoRegressive = the model takes the edge of the connection between a predefined number of lagged observations and therefore the present one.

Integrated = differencing between raw observations (eg. subtracting observations at different time steps).

Moving Average = the model takes advantage of the connection between the residual error and thus the observations.

It use of three main parameters (p,d,q). These are:

p = number of lag observations.

d = the degree of differencing.

q = the dimensions of the moving average window.

IV. CONCLUSION

Stock investments are of interest to several investors around the world. However, making a choice may be a complex task as numerous factors are involved. For successful investment, investors are keen to forecast the longer-term situation of the securities market. Even the slightest improvements of predictive efficiency will be very profitable. a top-quality prediction system will help investors make investments more accurate which is more profitable by providing supportive information like the long-run direction of stock prices. This paper discussed and compared the ML algorithms and techniques that are utilized in the stock price prediction. Many studies have proven that the sentiment (mood) analysis features a high impact on future prices. Thus, a mixture of technical and fundamental analyses can produce a highly efficient prediction.

V. REFERENCES

[1] Misra, Meghna, et al. "Stock Market Prediction Using Machine Learning Algorithms: A Classification Study." *2018 International Conference on Recent Innovations in Electrical, Electronics and Communication Engineering*, *ICRIEECE 2018*, IEEE, 2018, pp. 2475–78.

[2] Sirimevan, Naadun, et al. "Stock Market Prediction Using Machine Learning Techniques." *2019 International Conference on Advancements in Computing, ICAC 2019*, no. 4, 2019, pp. 192–97.

[3] Sharma, Ashish, et al. "Survey of Stock Market Prediction Using Machine Learning Approach." *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology*, *ICECA 2017*, vol. 2017-Janua, 2017, pp. 506–09.

[4] Maini, Sahaj Singh, and K. Govinda. "Stock Market Prediction Using Data Mining Techniques." *Proceedings of the International Conference on Intelligent Sustainable Systems*, *ICISS 2017*, no. Iciss, IEEE, 2018, pp. 654–61.

[5] Vijh, Mehar, et al. "Stock Closing Price Prediction Using Machine Learning Techniques." *Procedia Computer Science*, vol. 167, 2020, pp. 599–606.

[6] Sable, Rachna, et al. "Empirical Study on Stock Market Prediction Using Machine Learning." *2019 6th IEEE International Conference on Advances in Computing, Communication and Control, ICAC3 2019*, IEEE, 2019, pp. 1–5.

[7] Li, Xinyi, et al. "*DP-LSTM: Differential Privacy-Inspired LSTM for Stock Prediction Using Financial News.*" *ArXiv, no. NeurIPS, 2019*, pp. 1–9.

[8] Narote, Prof Amit, et al. *Predicting NASDAQ and NSE Stocks Using Machine Learning Algorithms : ARIMA , LSTM &.* no. Iii, 2020, pp. 39–46.

[9] Bhardwaj, Neha, and Akil Ansari. "Prediction of Stock Market Using Machine Learning Algorithms." *International Research Journal of Engineering and Technology*, vol. 5994, no. May, 2019, pp. 5994–6005.

[10] Jadhav, Rupawari, and Wakode M. S. "*Survey: Sentiment Analysis of Twitter Data for Stock Market Prediction.*" *Ijarcce*, vol. 6, no. 3, 2017, pp. 558–62.

[11] Adebiyi, Ayodele A., et al. "Stock Price Prediction Using the ARIMA Model." *Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014*, 2014, pp. 106–12.

[12] Hegazy, Osman, et al. "A Machine Learning Model for Stock Market Prediction" International Journal of Computer Science and Telecommunications, Volume 4, Issue 12. [Online]. Available: http://arxiv.org/abs/1402.7351

[13] Parmar, Ishita, et al. "Stock Market Prediction Using Machine Learning." *ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications, IEEE, 2018*, pp. 574–76.

[14] Nabipour, Mojtaba, et al. "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis." IEEE Access, vol. 8, 2020, pp. 150199–212.

[15] Hiransha, M., et al. "NSE Stock Market Prediction Using Deep-Learning Models." *Procedia Computer Science,* vol. 132, no. Iccids, Elsevier B.V., 2018, pp. 1351–62.

[16] Singh, Sukhman, et al. "Stock Market Forecasting Using Machine Learning: Today and Tomorrow." 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2019, 2019, pp. 738–45.