Study and Analysis Variety of Approaches for Stock Market Prediction

Shishir Kumar Sharma[1], Dr. Pooja Sharma[2]

[1]Research Scholar Kalinga University Nawa Raipur(C.G.)
[2]Associate Professor Kalinga University Nawa Raipur(C.G.)

Abstract : Stock Market Prediction (SMP) is one of the most complex task and its research has been of utmost importance in recent years. People who involved in the Stock Market can invest based on some kind of prediction they do not invest randomly. Employing traditional methods like fundamental and technical analysis might not make sure the reliability of the prediction. No one can never be sure of the rise and fall of the Market, predicting it with some accuracy is very much possible using the modern techniques like Machine Learning (ML), Data Mining and Deep Learning. In this paper, we survey various approaches including Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forests (RF), Regression and some fusion models. These modern approaches which involve ML, Data processing and Deep Learning have proven to offer more reliable results than traditional methods of Stock Prediction and have great possibility of advancement in the future.

Keywords **: Stock Prediction, Data Mining, Deep Learning, Machine Learning, Artificial Neural Network, Random Forests, Support Vector Machine.**

1. INTRODUCTION

Stock Market Prediction is mainly determining the worth of a company's stock or some other financial tool traded on an exchange. If in someway one could predict the Stock Market, it would result in substantial profits for investors. SMP is very challenging task in itself and has attracted many researchers. Traditional Prediction methods are not sufficient for investors anymore. SMP is categorized into two parts viz, fundamental analysis or Time Series Forecasting (TSF) and technical analysis or Trend Prediction Model[1]. The Trend Prediction model prescribes the market by constructing a connection amongst numerous technical indicators and fluctuations of the share prices while TSF involves predicting future stock prices by analyzing the stock's past return. We have focused our study on Technical Analysis. In this paper, we survey various approaches including Regression, SVM, ANN, RF and some fusion models. These modern methods which involve ML [2], Data Mining [3]and Deep Learning [4] have proved to give more reliable results than the traditional methods of Stock Prediction and have great possibility of advancement in the future.

2. LITERATURE REVIEW

Siew and Nordin [5], discuss the principles and applications of the regression approach to predict the stock prices by converting the figures into an ordinal data type. The chief design element is constructed on regression analysis from WEKA Machine Learning Software. The authors emphasize the fact that converting data into ordinal data type reduces error by a huge margin. The stock pricing dataset was extracted from various firms in BURSA Malaysia and was used for training multiple regression models. Significant attributes viz. Asset Turnover, Altman Z- Score, Debt to Equity ratio, Liquid Assets, and Net Tangible Asset were selected for training purposes. The selected attributes were used to predict the categorical type ordinal values of the stock prices. The primary aim of converting the data to the ordinal type was to create more organized data. The following figure outlines the overall methodology of evaluating the result.

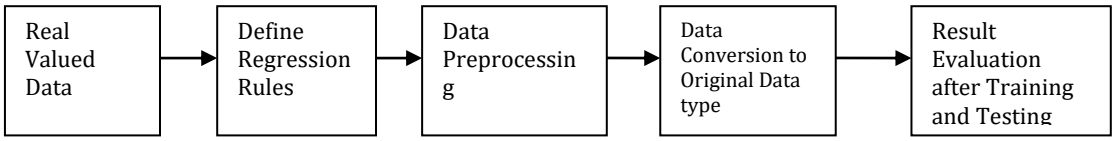

Figure 1

The researchers formulated two datasets. First being with Real Values while the second dataset with converted data values. SMO regression method surpasses others with an RMSE value of 0.8164 for the 2nd dataset. This was followed by Linear Regression, Simple Linear Regression and Regression by Discretization in the order mentioned. It was observed that by converting the data, RMSE values decreased by an average of 19.2835. Hu et al. [6] discuss a conjectural and heuristic structure to predict stock values using Support Vector Machine (SVM). The researchers took a total of 10 parameters of which some were company-specific and others were macroeconomic. 10 significant features were selected for training and analyzing purposes. 04 of them were company-based factors and the remaining 06 were macroeconomic factors having a great impact on the stock trend. The company-specific factors are namely Diluted EPS, P/E Ratio, Net Income, and Net Revenue. The macroeconomic factors are namely Dow Jones Industrial Average, Inflation Rate, Consumer Investment, Unemployment Rate, Consumer Spending, and Federal Funds Rate. The data used for training purposes was from 15 arbitrarily selected companies spread over 7 years. The data matrix contained a total of 88 entries of which 10 values were selected at random for testing purposes while the remaining were used for training purposes. SVM was then applied and the model trained by it achieved an extremely high accuracy of 97.44% with the training dataset, while a mere 70% accuracy was achieved with the testing set.

Ouahilal et al. [7] discuss a fusion model that incorporates Support Vector Regression (SVR) and Hodrick-Prescott (HP) filter. The HP filter is a data smoothening method primarily used in macroeconomics, to eliminate short term fluctuation in the business cycle. The aim of the HP filter for the proposed model is to filter the noise and to perform data value standardization on each feature distinctively. To compute the efficiency of the presented model the authors conducted numerous assessments. The figures range from 2004 to 2016. The data collected has 6 features viz. open, close, low, high, volume and date and has 2840 entries, thereby predicting the value of the closing price of the stock. The entries from 2004 to 2012 were used as input for training purposes while the remaining entries were employed for testing purposes. The testing data was divided into 4 sets according to the years. The first set contained entries from 2013, the second one contained entries from 2014 and so on. The result is then computed and the error is in the form of Mean Average Percentage Error (MAPE). The results show that the MAPE values decreased by an average of 0.185 when the HP filter was employed compared to when it was not in addition to the SVR. Naik et al. [8] discuss about SMP using an Artificial Neural Network (ANN) Regression model and employing Boruta Feature Selection (BFS) method for feature selection. The data used was gathered from the NSE, India and contained information about the opening price, closing price, highest and lowest price of ICICI and SBI stocks. The authors considered 33 combinations of technical indicators and applied BFS to choose the ideal feature. ANN which has 3 layers is used for SMP and the sigmoid function is employed in the regression-based ANN model. The selected attributes of the technical indicators were provided as input for the regression-based ANN mode. The efficiency of the model is evaluated by RMSE and MAE values. The MAE values decreased by an average of 11.1206 and the RMSE values decreased by an average of 14.7124 upon the application of the proposed model. The MAE value of ICICI was evaluated to be 15.1221 and of that of SBI were 17.4341. Nivetha et al. [9] devised a framework to forecast and estimate the monthly and daily stock market price. Multiple Linear Regression, Artificial Neural Network, and SVM were selected for building and then comparing to find the best one. Data of various firms for the few past months was collected from yahoo finance and used for monthly prediction whereas, for everyday forecast, sentimental analysis data containing the NEWS and reviews were collected from twitter and Economic Times. The significant key features were selected from NSE and BSE data which was extracted from yahoo financial data. For sentimental data analysis, each word by the reviewers was inspected and its relation with the stock price is found. ANN was found to be the best prediction model among the three chosen models.

Manojlovic and stajduhar [10] have considered 5 and 10 days ahead SP models using the RF algorithm. The data is from numerous companies listed on ZAGREB Stock Exchange and the models are constructed upon the CROBEX index. The figures range from 2008 to 2013, consisting of 4 features viz. open, close, low and high price. A total of 12 technical indicators were used as model inputs viz, CCI, Stochastic %K and %D, RSI, MACD, 10 days Momentum and Rate of Change, 5 days Standard Deviation, 5- and 10-days Moving Average, Weighted Moving Average, and Disparity. The authors used the RF algorithm and executed in WEKA toolkit for the classification purpose. Ideal results were obtained by setting the number of attributes for random selection to 5, the number of trees to 100, and unbounded tree depth. Average accuracy for 5 and 10 days ahead models were evaluated to be 76.5% and 80.8% respectively. Shahabuddin et al. [11] discuss data mining algorithms particularly employing Naïve Bayes to enhance stock price prediction. The proposed model categorizes text data associated with FTSE100 on the specified websites while the trained classifier notes the fluctuations in the stock values. The authors employed Naïve Bayes because of its ability to handle numerous features. The figures primarily contained reports on the stock price of FTSE100 which were gathered from thisismoney.com. The data was cleansed in 4 steps and the entire system was constructed in MATLAB. Following could be a brief overview of the steps involved.

1. Number omission from the text
2. Special Character omission
3. Monosyllabic characters were then removed
4. A list of frequent words is then analogized with the left-over words and crossed out if they match.

The researchers took the data sets from April to November 2018 for training and testing purposes. A total of 175 sets were taken, of which 140 were employed for training purposes while the remaining were employed for testing purposes. By adding the heuristics, the authors mentioned testing accuracy increased by an average of 7.3% with the highest achieved testing accuracy of 87.23%.

The following Table gives a brief representation of all the reviewed papers. Author's Name, Machine Learning Technique used. Criterion Used to Evaluate the result and the publishing year is stated in the table given below :

RESULT DISCUSSION	EVALUATION CRITERION	ML TECHNIQUE USED	YEAR	AUTHOR
Predicting next day stock price	Accuracy Rate	Naive Bayes	2010	Shihavuddin et al.[11]
Ranking of the predicted price trend in categorical ordinal value.	Correlation Coefficient, MAE, RMSE	SMO, Linear Regression, Regression by Discretization	2012	Siew and Nordin[5]
Classified Stock as Good or Poor investment	Accuracy Rate	Support Vector Machine	2013	Hu et al.[6]
Stock Market Trend Prediction	Accuracy Rate, F Score	Random Forest	2015	Manojlović and Štajduhar [10]
Predicting the value of the closing price of the stock	MAPE	Support Vector Regression with Hodrick-Prescott filter	2017	Ouahilal et al.[7]

Estimates the open value of subsequent day within the market	Accuracy Rate	Multiple Linear Regression, SVM, ANN	2017	Nivetha et al.[9]
Stock Market Trend	Root Mean Square Error	Linear Regression Model with PCA	2017	Waqar et al.[12]
Predictions of the Stock Prices	MSE, MAPE	Support Vector Regression with IFOA	2018	Hou et al.[13]
Stock Price Prediction	MAE, RMSE	ANN Regression	2019	Naik et al.[8]

Waqar et al.[12] designed a combination of Linear Regression (LR) and Principal Component Analysis (PCA) to handle the high dimensionality of the dataset for predicting the trends in the stock market. PCA enhanced the performance of the ML model by selecting the most significant and un-correlated features while removing unnecessary data and LR was used for classification. Data was collected from Karachi, London, and NY Stock Exchange. RMSE was used as a determining criterion for comparing the dataset before and after applying PCA. Applying PCA for feature selection tremendously reduced the RMSE value from 36 to 1.00 for NYSE and from 16.43 to 1.4 in the case of LSE. Hou et al. [13] discuss an SVR based prediction model which was enhanced by an algorithm viz, improved fruit fly algorithm (IFOA) for SMP. The authors employed Grey Relational Analysis (GRA) for choosing technical indicators followed by the application of IFOA to enhance the attributes of the SVR. The authors used the libsvm toolkit in their experiments. After filtering out from 9 experimental indicators using GRA, the authors found that the best possible results will be obtained when the indicators selected are MACD, highest, lowest, closing, and opening price of the stock. After providing input for the training and testing data set, standardization of the training and testing set was performed. Following this IFOA was employed to enhance the attributes viz, penalty and kernel function parameter. This was repeated for 20 times to finally calculate the MSE and MAPE values. The MSE value when IFOA was used was found to be 0.1623 and the MAPE value was found to be 0.0321. Although employing the improved FOA resulted only in a slight decrease in error values was noted.

3. DISCUSSION AND FUTURE SCOPE

After analyzed different approaches for predicting stock market prices and concluded that ML and the regression model helps to attain valuable results with high accuracies. The majority of ML models used SVM as a classifier for prediction. The usage of SVM along with HP filter decreased the mean calculated error significantly. Among all the regression techniques, SMO outstrip others giving commendable results. Using an improved fruit fly algorithm in addition to SVR, resulted in significantly fewer MAPE values. One of the models which used RF achieved an accuracy of 80.8%. One of the models discussed above used PCA in addition to LR which significantly decreased the RMSE values. Choosing valuable and distinct attributes with a well-distributed and balanced dataset can result in highly remarkable and considerable accuracies. There are high chances of improving the accuracy rate by selecting appropriate valuable attributes and optimizing them to reduce the error as much as possible. Predicting stock trends will help the investors to choose wisely and make out the best profits by analyzing the previous stock exchange data. The combination of ML with Deep Learning Models can result in better training and implementation of prediction model. Big data tools can also be used for handling large data and refining them to get more useful features and train the model.

4. CONCLUSION

There are lots of approaches for stock prediction, I have considered a some variety of approaches and after a deep study, conclude that it is possible to create a new and better cross method to predict the future fluctuations of the stock market. There are existing fusion models but they are of relatively low accuracy. Accuracy is one of the most important factor in prediction

and Accuracy can be improved by selecting more significant attributes and optimizing dimensionality reduction methods. Forming a balanced and evenly distributed dataset improves the reliability of the training model. We believe that there will be further enhanced and better models that employ the use of ML, Data Mining, and Deep Learning.

REFERENCES

[1] "Investopedia." [Online]. Available: investopedia.com.

[2] K. C. Rasekhschaffe and R. C. Jones, "Machine Learning for Stock Selection," *Financ. Anal. J.*, vol. 75, no. 3, pp. 70–88, Jul. 2019.

[3] C. Y. Huang and P. K. P. Lin, "Application of integrated data processing techniques available for market forecasting," *Cogent Econ. Financ.,* vol. 2, no. 1, 2014.

[4] G. Ruxanda and L. M. Badea, "Configuring artificial neural networks for stock exchange predictions," *Technol. Econ. Dev. Econ*., vol. 20, no.1, pp. 116–132, 2014.

[5] H. L. Siew and M. J. Nordin, "Regression techniques for the predic- tion of stock price trend," ICSSBE 2012 - Proceedings, 2012 Int. Conf. Stat. Sci. Bus. Eng. "Empowering Decis. Mak. with Stat. Sci., pp. 99– 103, 2012.

[6] Z. Hu, J. Zhu, and K. Tse, "Stocks market prediction using Support Vector Machine," Proc. 2013 6th Int. Conf. Inf. Manag. Innov. Manag. Ind. Eng. ICIII 2013, vol. 2, no. 3, pp. 115–118, 2013.

[7] M. Ouahilal, M. El Mohajir, M. Chahhou, and B. E. El Mohajir, "Opti- mizing stock market price prediction using a hybrid approach based on HP filter and support vector regression," Colloq. Inf. Sci. Technol. Cist, vol. 0, pp. 290–294, 2017.

[8] N. Naik and B. R. Mohan, "Optimal Feature Selection of Technical Indicator and Stock Prediction Using Machine Learning Technique," vol. 985, A. K. Somani, S. Ramakrishna, A. Chaudhary, C. Choud- hary, and B. Agarwal, Eds. Singapore: Springer Singapore, 2019, pp. 261–268.

[9] R. Y. Nivetha and C. Dhaya, "Developing a Prediction Model for Stock Analysis," Proc. - 2017 Int. Conf. Tech. Adv. Comput. Com- mun. ICTACC 2017, vol. 2017-October, pp. 1–3, 2017.

[10] T. Manojlović and I. Štajduhar, "Predicting stock exchange trends using random forests: A sample of the Zagreb stock exchange," 2015 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2015 - Proc., no. May, pp. 1189–1193, 2015.

[11] A. S. M. Shihavuddin, M. N. Ambia, M. M. N. Ardin, M. M. Hossain, and A. Anwar, "Prediction of stock price analyzing the web financial news using Naive Bayes classifier and native economic trends," ICACTE 2010 - 2010 3rd Int. Conf. Adv. Comput. Theory Eng. Proc., vol. 4, pp. 22–26, 2010.

[12] M. Waqar, H. Dawood, M. B. Shahnawaz, M. A. Ghazanfar, and P. Guo, "Prediction of Stock market by Principal Component Analysis," Proc. - 13th Int. Conf. Comput. Intell. Secur. CIS 2017, vol. 2018-Janu- ary, pp. 599–602, 2018.

[13] Homg-Jinh Chang & Mei-Chi Liu (1998) A statistical fitting model in Taiwan's stock price, *Journal of Information and Optimization Sciences*, 19:3, 353-371

[14] H.S. Hota and R. Handa, "Hybridization of Wavelet Transform and ANN Techniques as Stock Prediction Model", International Journal of Innovations in Engineering and Technology (IJIET) Vol. 7(2), 2016, pp.227-232.

[15] S. Shrivastava and A. K.Shrivas, "Performance Analysis of Data Mining Techniques for Stock Price Forecasting", International Journal of Engineering Research and Technology (IJERT) Vol. 6(5), 2017, pp. 861-864.