MACHINE LEARNING MODELS TO PREDICT MARKET MOVEMENTS BASED ON HISTORICAL PRICE DATA AND ECONOMIC INDICATORS

Arif Hossen
MBA in Business Analytics, International American University
Email: arifhossen4295@gmail.com, ORCID: https://orcid.org/0009-0005-4024-8611

Md Fokrul Islam Khan
M.Sc. in Management Information System, International American University
Email: fokrulkhan837@gmail.com, ORCID: https://orcid.org/0009-0001-8174-4831

Yeasin Arafat
MBA in IT Management, Westcliff University
Email: y.arafat.570@westcliff.edu

Farhad Uddin Mahmud
Master's in Management Information System
Email: mahmudfarhad89@gmail.com

ABSTRACT

In the ongoing economic environment, predicting stock market prices is a significant point. Accordingly, researchers are presently more leaned to look for new chances to conjecture the stock market. We examine machine learning methods utilized in stock market estimating. Determining the bearing of the stock market's price advancement could yield critical benefits. To predict the future price of the stock, sellers utilize particular request, which is the investigation of price by looking at the past prices. A particular investigation instrument called a moving average assists the broker with distinguishing examples and pinpoint significant price focuses for stock exchanges. The moving normal is an incidental effect as well as a pattern marker. Reactive result is a financial indicator that is only provided following a significant price shift. The purpose of this work is to apply machine learning techniques to a specialized pointer. The suggested model will use relapse on moving averages to reduce the exchange signal's idle time and eventually overcome its drawback. By forecasting the exchange signal provided by the moving averages, the model is able to predict the pattern's inversion.

Keywords: Machine Learning Models, Predict Market, Movements, Historical Price Data, Economic Indicators

1. INTRODUCTION

The demonstration of assessing the future worth of an organization's stock or some other monetary instrument exchanged on a market is known as stock market prediction and examination. The stock market is an imperative part of the national economy and assumes a key part in the development of exchange and business the country, which thus influences the national economy. The business and the two monetary backers are keen on the stock market and want to find out whether certain stocks will rise or decline over a given timeframe. The stock market is an imperative center for any organization looking to raise capital for development. It is subject to the ideas of organic market. At the point when there is more interest for an association's stock, its portion price rises; on the other hand, when there is less interest for the association's stock, its portion price falls.

Because of the blend of various ventures and associations, it involves unbelievably huge data structures from which it is hard to separate individual data and actually take apart their work designs. Exploration and investigation of the stock market can uncover market examples and gauge when buying stocks will be ideal. A stock's future price prediction that demonstrates precise could yield gigantic prizes. In doing as such, it is affirmed that time series plans offer a really huge predictive potential for high chance of supportive exchanges and high useful returns for the serious business adventure. This is finished by utilizing tremendous significant market data to deal with changing circumstances. Machine learning is a data examination approach that robotizes intelligent model development. Machine learning, which makes utilization of calculations that iteratively gain from data, empowers PCs to find stowed away information without being explicitly customized where to look.

The use of machine learning algorithms is basically centered around particular examination; in any case, coordinating the ideas of focal examination concerning machine learning can be gainful. This undertaking portrays the many methodologies that have been utilized to apply machine learning to stock expecting and furthermore recommends novel, intense thoughts that warrant future examination. As of late, a ton of energizing work has been finished in the space of utilizing machine learning estimations to examine price plans and conjecture changes to documents and stock qualities. Nowadays, most of stock dealers depend on Wise Trading Frameworks, which help in price prediction considering different situations and settings, empowering them to make quick business choices.

2. LITERATURE REVIEW

Gurrib's (2014) The adequacy of the Moving Average Crossover (MAC) strategy in guaging the movements of the S&P 500 market record is researched in this review. To distinguish exchanging signals, the MAC approach looks at both the short-term and long-term moving averages of stock prices. Gurrib assesses the introduction of the cycle utilizing back testing and careful assessment. That's what the discoveries propose albeit the MAC strategy might yield benefits under specific market conditions, its appropriateness differs with time and is affected by elements like example length and market unpredictability. As a rule, the evaluation stresses that it is so essential to consider market factors and change exchanging methods likewise.

Hegazy et al. (2013) give a machine learning model that makes utilization of computational knowledge strategies to gauge stock market movements. The survey centers around upgrading a prediction model utilizing past stock data, particular indicators, and an investigation of market opinion. The predictive model is ready and tried by the designers utilizing different machine learning calculations, for example, support vector machines (SVM) and artificial neural networks (ANN). The accuracy of the stock price vacillations predicted by the outcomes is great, exhibiting the way that machine learning methods could improve trading systems. Notwithstanding, the concentrate likewise acknowledges difficulties connected with data pre-taking care of, for example, independent direction and model guess, and gives roads to additional examination and improvement.

Jui's (2012) Study dives into the use of computational knowledge gets close, like variety understanding, feathery rationale, and innate calculations, for stock price assessment. To further develop prediction accuracy, the audit underscores the benefit of joining a few data sources, like past price data, central indicators, and market assessment examination. Jui desires to settle the innate weaknesses and intricacies of monetary markets by using the versatile and learning abilities of computational knowledge strategies. As indicated by the discoveries, crossbreed models that consolidate a few computational knowledge approaches can create more exact execution when contrasted with individual strategies. The paper likewise accentuates the meaning of model progression and limit tuning in improving predicted accuracy and robustness.

Ling et al. (2012) recommend a mutt technique for stock rundown assessment that joins wavelet-based feature extraction with Support Vector Regression (SVR) and Multivariate Adaptive Regression Splines (MARS). By utilizing wavelet change, the survey endeavors to evaluate the capacity to catch both transient and repetitive region data in stock market data. Then, non-direct connections between's the isolated features and stock document movements are distinguished utilizing MARS and SVR models. The outcomes show further developed assessment accuracy when contrasted with regular methodologies, featuring the appropriateness of consolidating cutting edge regression strategies with wavelet-based feature extraction. One way or the other, the evaluation underscores that thorough execution across a scope of market circumstances relies upon cautious limit tweaking and model acknowledgment.

M et al. (2018) Look at whether significant learning models might be utilized to estimate market patterns for the National Stock Exchange (NSE). Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are utilized in the survey to examine historical market data and spotlight on critical cases. Through the use of moderate features produced by profound learning designs, the recommended models exhibit better prediction execution analyzed than regular strategies. The concentrate additionally looks into how other information features, like particular indicators and market feeling examination, influence prediction accuracy. Albeit profound learning models yield empowering results, difficulties like restricted data, interpretability of the model, and computational intricacy are still regions that require further innovative work.

Naved and Srivastava (2015) Look at the exhibition of five famous moving average variations north of a ten-year time frame on the Indian market record S&P CNX Shrewd 50. The capacity of simple moving averages (SMA), weighted moving averages (WMA), triangular moving averages (TMA), exponential moving averages (EMA), and adaptive moving averages (AMA) to create beneficial trading signals is analyzed in this paper. That's what the discoveries propose albeit a wide range of moving averages show some degree of efficiency, the presentation shifts relying upon limit settings and market conditions. The examination features the significance of choosing fitting moving average classes and limits considering the fundamental market parts for compelling exchanging procedures.

3. A HIGHLIGHT OF THE INFORMATION AND MACHINE LEARNING METHODS APPLIED TO STOCK MARKET PREDICTION

3.1. Data sources

As per the information accessible, the most powerful components influencing changes in stock prices are those connected with cash, macroeconomic circumstances, and technical signs. Regardless, since there isn't a lot of agreement on essentially the parts that are all huge for stock market anticipating, these investigations consolidated unique plans of variables as information data for their prediction models. As delineated in Fig. 1, we ordered every one of the components in the chose text that were utilized as information data in this survey into four principal types and a couple of subcategories.

Since technical indicators assume a urgent part in stock exchange signals, they are normally utilized as information parts in prediction studies. Technical indicators were frequently organized into two classes: "other technical indicators" and "fundamental technical indicators." A portrayal and examination of every kind and subtype are given, which examines the variables and stock market records examined in the article. While many examinations zeroed in on different other technical indicators, a couple of examinations involved significant technical signs as the information data for precisely determining stock markets.

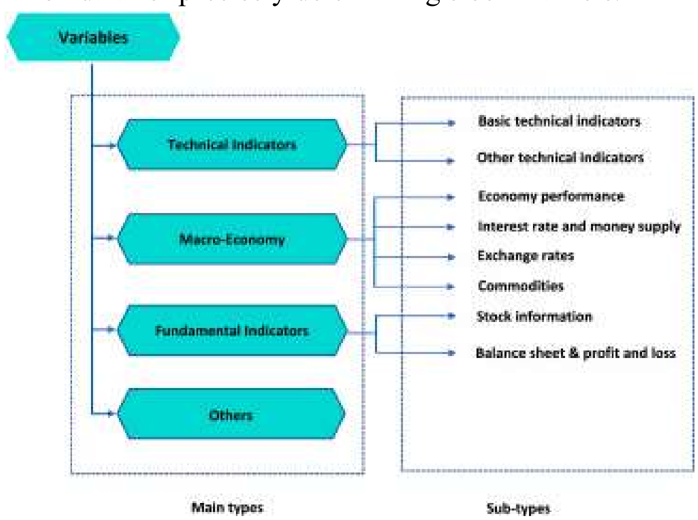

Figure 1: Predictions of stock price and return are subject to variable classifications.

Besides, macroeconomic variables are eminent in research on the stock market. For this order, we characterized four subcategories: "exchange rates," "items," "economic execution," and "credit expense and money supply."

Extra part related to pay in stock market prediction concentrates on has been focal pointer related factors. These variables are likewise looked at in two subclasses: (1) "Stock information variables" allude to variables that are linked to or reliant upon a specific association's stocks exchanged on a public stock market; (2) "balance sheet and profit and loss statement variables" allude to variables that are demonstrative of monetary information.

The excess variables are, eventually, alluded to as "various variables" and have not very many sub-divisions. A couple of studies, for instance, determined the price of different records or variables removed from monetary news, unconstrained statements, email data, and tweets, comparable to a particular stock market.

3.2. Machine learning

Machine learning depends on extricating information from data. The most broadly involved machine learning method for stock market prediction is overseen learning. The overall workflow of a coordinated learning-based methodology utilized for stock market prediction is displayed in Fig. 2.

Utilizing time-series data (stock price as well as return) or potentially significant information (financial news, for instance) from a particular time span begins the cycle. In the unlikely occasion that the task includes gathering, the objective class is either known or should be expected.

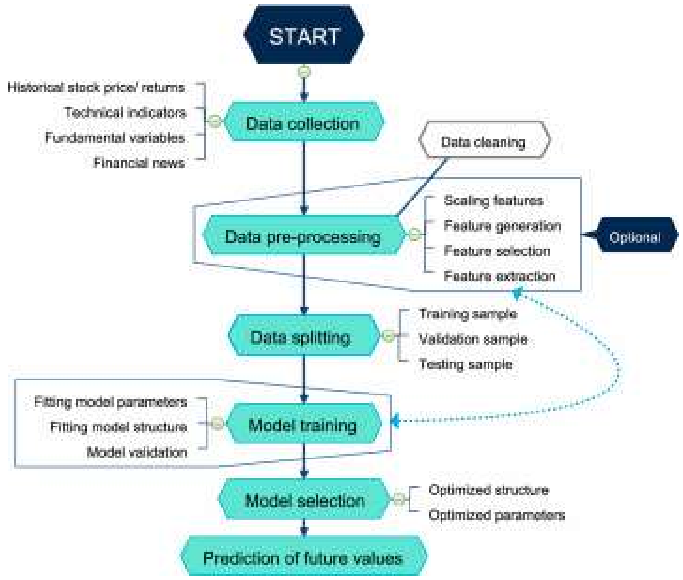

Figure 2: Workflow of a supervised learning stock market prediction model.

The related data should initially be pre-taken care of, which incorporates removing and clearing up any fragmented or clearly good for nothing data (like identifiers). From that point onward, technical signs can be distinguished utilizing the major time-series data, like close price data. Following the securing of the cleaned data with technical signs, the data are further pre-taken care of by scaling and dimensionality diminishes (i.e., feature determination, feature extraction, and component age) to remove pertinent variables and channel out immaterial ones. Making utilization of pre-handled data frequently brings about exact figures. Regardless, this stage is displayed as discretionary in Fig. 2 since it as a rule relies upon the chose region and is likewise passed on to the maker's judgment. Choosing an inventive or state of the art machine learning procedure to figure the objective variable is the task after the information data are ready. Along these lines, input data are generally partitioned into three classifications: arrangement (to make the model and put down its stopping points), endorsement (to assess each pre-arranged model and pick the best model plan and limits), and test (to assess the keep going model's execution in view of discernments that it has not experienced prior to during readiness and endorsement). Feature determination is recorded as a data pre-taking care of stage since it tends to be openly applied to the learning calculation in its most fundamental structure (i.e., channel techniques). Notwithstanding, it very well might be connected with model readiness by performing feature determination utilizing the learning estimation's presentation (covering approach) or it could be incorporated into the model advancement process itself (embedded strategy). The readiness step is associated with the pre-taking care of step with a ran line in the stream diagram to show this likely connection between feature determination and model planning. Eventually, prediction is done utilizing the created regression model or the pre-arranged gathering model.

A couple of machine learning strategy variations have been made in this paper and used to stock market predictions. The most broadly utilized strategies among these are support vector machines (SVMs), artificial neural networks (ANNs), and their variations, as they have exhibited empowering prediction results. An enormous collection of writing dissecting stock market measuring models considering cushioned speculation has been scattered starting from the introduction of smart frameworks in cushy theory that coordinate with weakness in data. These models incorporate cushy time-series, cushioned construing frameworks in light of adaptive networks, fleecy frameworks of the Takagi-Expert Kang (TSK) type, and a few varieties.

Moreover, as the article refers to, extra machine learning procedures including irregular timberlands, k-nearest neighbour (KNN) classifiers, and Bayesian networks have likewise been utilized a ton. In any case, considering that the extraordinary greater part of the strategies referenced enjoy benefits and drawbacks of their own, a few researchers would by and large further develop the assessing accuracy of those techniques. Subsequently, blends of different procedures, like ANN + SVM, KNN + SVM, and others, have been examined for their capability to predict stock returns or costs. Besides, by utilizing feature determination methods, the guess accuracy of the recently referenced systems has been additionally improved. These methods incorporate, yet are not restricted to, extraction methodologies, for example, head part assessment (PCA), extraordinary algorithms like genetic algorithms (GA), Wavelet changes, and atom swarm enhancements. Moreover, the capacity of collection as an independent methodology was likewise inspected at assessing stock expenses, instead of the as of late examined regulated learning systems.

Constant headways in the space of stock market estimating have prompted a restoration of profits in profound learning systems. Profound learning is a part of machine learning that doesn't depend on human skill or economical suppositions to distinguish stowed away nonlinear connections and distil significant features from perplexing and boisterous data. Long short-term memory networks (LSTM), convolutional neural networks, and profound neural networks have all been generally utilized in stock market price and bring assessing back, depending on the situation.

4. METHODOLOGY

1. Moving average

The average price over the previous 'n' days is the moving average for that day.

$$\mathbf{MA} = \frac{\mathbf{A_1} + \mathbf{A_2} + \ldots + \mathbf{A_n}}{\mathbf{n}} \qquad (\mathbf{1})$$

2. Confusion Matrix:

A confusion matrix is used to evaluate how well a classifier is presented. The actual class is represented by each column in the confusion matrix. However, each segment corresponds to the class that the classifier predicted. The number of modified organized results of the models indicates the convergence of the real and the projected class. Therefore, each True Positive is represented by the matrix's slant.

3. Accuracy:

One statistic used to evaluate order models is accuracy. The ratio of accurate predictions to total predictions is what determines its value.

$$\mathbf{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{n}} \qquad (2)$$

4. Precision:

Precision is the classifier's positive prediction accuracy.

$$\mathbf{Precision} = \frac{\mathbf{TP}}{\mathbf{TP + FP}} \qquad (3)$$

5. Recall:

Reaction time or true sureness rate are different terms for recall. The level of positive cases that the classifier accurately perceives is known as recall.

$$\mathbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP + FN}} \qquad (4)$$

6. F1 score:

The F1 score is a solitary evaluation that consolidates recall and precision. It is the amicable average of recall and precision.

$$\mathbf{F1} = 2 \times \frac{\mathbf{precision \times recall}}{\mathbf{precision + recall}} \qquad (5)$$

7. Regression

Regression is used to forecast when the two moving averages will crossover and to get ready for it to happen.

$$\widehat{y_t} = \widehat{\beta_0} + \widehat{\beta_1} x_{1,t} + \widehat{\beta_2} x_{2,t} + \dots + \widehat{\beta_k} x_{k,t} \qquad (6)$$

5. RESULTS

Table 1 displays the model's confusion matrix for the IBM stock. The IBM stock confusion matrix provides tidbits of information on how a predictive model is displayed when ordering stock activities (SELL, NO Activity, Purchase). The numbers in the slanted cells correspond to correct predictions, but the off-askew cells show incorrect classifications. 171 SELL activities, 3856 NO Activity cases, and 172 Purchase activities were all correctly predicted by the model. However, it misclassified a few examples, particularly in the areas of forecasting NO Activity when it was actually SELL (196 occurrences), projecting SELL when it was actually NO Activity (356 cases), and predicting Purchase when it was actually NO Activity (142 occasions). These misclassifications highlight areas where the model's viability and accuracy in predicting IBM stock activity should be improved.

Table 1: IBM stock confusion matrix

	Actual SELL	Actual NO ACTION	Actual BUY
Predicted SELL	171	356	7
Predicted NO ACTION	196	3856	142
Predicted BUY	10	366	172

The model's confusion matrix for the GOOGL stock is displayed in Table 2. The GOOGL stock confusion matrix shows how a prediction model is shown when stock actions (SELL, NO Activity, Purchase) are grouped together. While off-corner to corner cells display misclassifications, slanting cells address exact predictions. 147 SELL events, 2994 NO Activity instances, and 126 Purchase activities were all correctly predicted by the model.

Whatever the case, it incorrectly identified several examples, most notably 237 instances of SELL when the true activity was NO Activity, 181 instances of NO Activity when it was actually SELL, and 99 instances of Purchase when it was NO Activity. These misclassifications highlight areas where the model's accuracy and consistency in predicting GOOGL stock activity need to be strengthened.

Table 2: Confusing matrix for the stock of GOOGL

	Actual SELL	Actual NO ACTION	Actual BUY
Predicted SELL	147	237	5
Predicted NO ACTION	181	2994	99
Predicted BUY	11	294	126

The model's confusion matrix for the AAPL stock is displayed in Table 3. The presentation evaluation of a prediction model in characterizing stock activities (SELL, NO Activity, Purchase) is presented in the confusion matrix for AAPL stock. Cells with a slant indicate accurate predictions, while those without a slant indicate incorrect classifications. 160 Purchase actions, 3918 NO Activity occurrences, and 161 SELL activities were all accurately predicted by the model. Regardless, it demonstrated misclassifications in several scenarios, particularly in predicting SELL when the true activity was NO Activity (357 instances), predicting NO Activity when it was actually SELL (203 instances), and predicting Purchase when it was NO Activity (106 times). These misclassifications highlight areas where the model's accuracy and suitability for predicting AAPL stock activity need to be improved.

Table 3: Matrix of confusion for AAPL stocks

	Actual SELL	Actual NO ACTION	Actual BUY
Predicted SELL	161	357	8
Predicted NO ACTION	203	3918	106
Predicted BUY	9	354	160

The general display of the suggested model on IBM, GOOGL, and AAPL is displayed in Table 4. Table 4 displays the general presentation metrics of a suggested predictive model for each of the three distinct companies—AAPL, GOOGL, and IBM. With values of approximately 0.8 for each of the three organizations, accuracy indicates the proportion of correctly classified cases among all occurrences and suggests a significantly enhanced level of overall correctness in forecasts. Precision estimates the number of true sure predictions there that are among all certain predictions, mirroring the model's capacity to keep away from misleading positives. It increments from 0.775 to 0.788, showing predictable precision across the associations. Recall estimates the number of true sure predictions there that are among all genuine positive events, exhibiting the model's capacity to distinguish each certain model. It goes from 0.799 to 0.807, proposing a comparable degree of recall among the different associations. The F1 Score consolidates recall and precision into a solitary measurement. Values for all organizations range from 0.78 to 0.79, demonstrating a fair compromise among memory and precision. Generally, the recommended model performs well on all measurements, with specific accentuation on how well it predicts stock action for AAPL, GOOGL, and IBM.

Table 4: Overall effectiveness of the suggested model

	Accuracy	Precision	Recall	F1 Score
IBM	0.799	0.775	0.799	0.781
GOOGL	0.806	0.788	0.806	0.792
AAPL	0.807	0.782	0.807	0.787

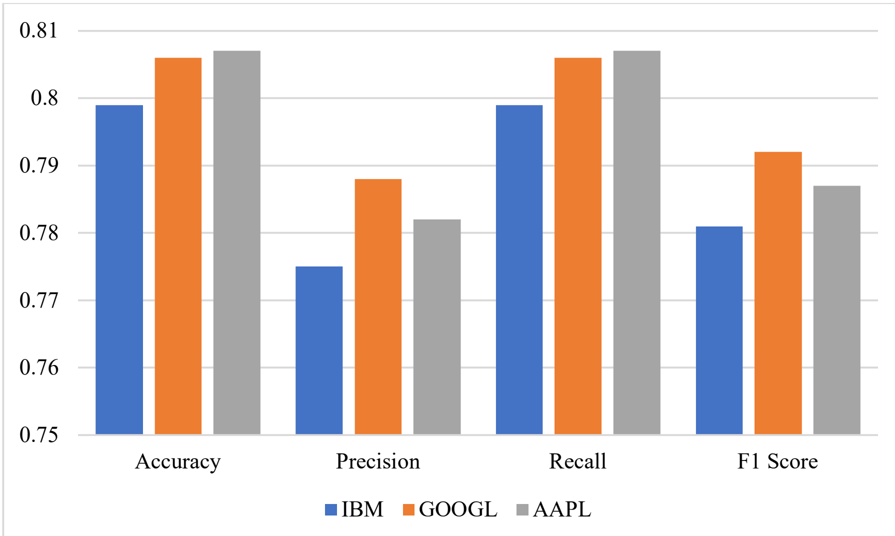

Figure 3: Graphical Representation of Overall effectiveness of the suggested model

6. CONCLUSION

A practical way for market examination and free heading is the utilization of machine learning models to gauge market movements utilizing past evaluating data and economic variables. Using refined algorithms, these models can reveal complicated designs inside enormous datasets, giving bits of knowledge on potential patterns and changes in the market. To conquer the constraints of the pointer-based trading strategy, this study proposes utilizing machine learning to technical indicators. This approach acknowledges that moving averages are latent and proposes a strategy to beat this weakness. To decrease lethargy and expect design reversal, the model predicts the negative and bullish intersection early. In view of the outcomes, obviously the recommended model defeats the moving average's drawbacks as well as empowers the broker to increment profit.

REFERENCES

1. Al-Radaideh, Qasem & Assaf, Adel & Alnagi, Eman. (2013). Predicting Stock Prices Using Data Mining Techniques.
2. Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A. and Weber, I. (2012). Web Search Queries Can Predict Stock Market Volumes. [online] Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.004 0014 [Accessed 3 Mar. 2019].
3. Dassanayake, W. and Jayawardena, C. (2017). Determinants of stock market index movements: Evidence from the New Zealand stock market. [online] https://www.researchgate.net.

4. Gurrib, Ikhlaas. (2014). The Moving Average Crossover Strategy: Does it Work for the S&P500 Market Index? SSRN Electronic Journal. 10.2139/ssrn.2578302.

5. Hegazy, Osman & Soliman, Omar S. & Abdul Salam, Mustafa. (2013). A Machine Learning Model for Stock Market Prediction. International Journal of Computer Science and Telecommunications. 4. 17-23.

6. Jui Y.: "Computational Intelligence Approaches for Stock Price Forecasting", IEEE International Symposium on Computer, Consumer and Control (IS3C), pp. 52 - 55, 2012.

7. Ling J., Chih C., Chi L, Chih H., "A hybrid approach by integrating wavelet-based feature extraction with MARS and SVR for stock index forecasting", ELSEVIER, Decision Support System, 2012.

8. M, H., E.A., G., Menon, V.K. and K.P., S. (2018). NSE Stock Market Prediction Using Deep-Learning Models. Procedia Computer Science, 132, pp.1351–1362

9. Naved, Mohd & Srivastava, Prabhat. (2015). The Profitability of Five Popular Variations of Moving Averages on Indian Market Index S & P CNX Nifty 50 During January 2004– December 2014. 2.

10. Oncharoen, P. and Vateekul, P. (2018). Deep Learning for Stock Market Prediction Using Event Embedding and Technical Indicators. 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA).

11. Patel, J., Shah, S., Thakkar, P. and Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications, 42(4), pp.2162–2172.

12. Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M. and Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. [online] https://www.researchgate.net.

13. S. Bharathi1*, A. Geetha1 and R. Sathiynarayanan1, "Sentiment Analysis of Twitter and RSS News Feeds and Its Impact on Stock Market Prediction", 2017. [Online]. Available: https://www.researchgate.net/publication/320694083_Sentiment_Ana lysis_of_Twitter_and_RSS_News_Feeds_and_Its_Impact_on_Stock_ Market Prediction. [Accessed: 05- Mar- 2019].

14. Usmani, M., Adil, S.H., Raza, K. and Ali, S.S.A. (2016). Stock market prediction using machine learning techniques. 2016 3rd International Conference on Computer and Information Sciences (ICCOINS).

15. Yi X., Jin X., Fengbin L. & Shouyang W: "Ensemble ANNs PSO-GA Approach for Day-ahead Stock E-exchange Prices Forecasting", International Journal of Computational Intelligence Systems, Volume 6, Issue 1, pp.96-114 2013.