STOCK PREDICTION USING FINANCIAL DATA AND NEWS SENTIMENTAL ANALYSIS

Bindeshwari Chandel *1, Suman Kumar Swarnkar *2

Department of Computer Science & Engineering
Chhatrapati Shivaji Institute of Technology, Drug [C.G.]

Predicting stock market prices has long been a subject of concern to investors as well as academics. Because of their extremely volatile existence, which depends on numerous political and economic influences, leadership transition, investor sentiment and several other influences, stock prices are hard to predict. Predicting stock prices based on either historical or textual data alone proved inadequate. Current studies in sentiment analysis have found a strong link between stock price movement and the publishing of news papers. Several studies of sentiment analysis were attempted at different stages, using algorithms such as supporting vector machines, naive regression of Bayes, and deep learning. The accuracy of deep learning algorithms depends on the amount of data given for the training. However, the amount of textual data obtained and analyzed during the past studies was inadequate and thus led to low accuracy predictions. In our paper, we are improving the accuracy of stock price forecasts by collecting and analyzing a large amount of time series data in relation to relevant news storey's, using deep learning models. For five years, the dataset we have compiled contains daily stock prices for S&P500 firms, along with more than 265,000 pieces of financial news related to those firms. Because of the large size of the dataset, we use cloud computing as an indispensable resource to train prediction models and perform inferences in real time for a given stock. Index Market prediction terms-stock, cloud, big data, machine learning, regression.

Keywords: Companies, models, Data ,Feature extraction, Facebook, Recurrent neural networks, Stock markets

I. INTRODUCTION

It is a established fact that in the longer period of time investing in the stock market yields more income, but that is not always the case. Choosing growing stock to invest in plays a crucial role in obtaining profit. Investing in a stock is nothing but committing our money to a particular stock-market-listed firm, with the hope of an optimised profit. Without prior diligence on the part of the investor a solid investment is not feasible. A proper analysis is an essential step, before deciding to invest in a given stock. Investment stocks chosen at random will result in adverse losses. Today, many newbie investors have no idea how the stock's future would look like. It's more like a show of hands for them, invest in any randomly picked stock, if its price goes up, lucky enough other lucky better next time! In fact this is not the way it works. The astrology of forecasting the future stock price is one of the most fashionable and debated subjects of all time related to various fields including statistics , economics, trading and computer science, its purpose certainly, to forecast the course of the future stock price so that they can be purchased and then sold at higher profit.

The price of the product depends on several factors, such as market forces, the relationship between supply and demand etc. If the consumer needs more than what is supplied for a given product, the product price will rise; otherwise the price will go down, i.e., if the supply is more than the demand. Many investors in the stock market are big believers too! Some assume that forecasting the stock price is an difficult task while some say; given a bit of graphical analysis and a few estimates from the past, it is entirely predictable. While the above category of investors are very correct in their convictions, it's basically far more than that.

The price of the product depends on several factors, such as market forces, the relationship between supply and demand etc. If the consumer needs more than what is supplied for a given product, the product price will rise; otherwise the price will go down, i.e., if the supply is more than the demand. Many investors in the stock market are big believers too! Some assume that forecasting the stock price is an difficult task while some say; given a bit of graphical analysis and a few estimates from the past, it is entirely predictable. While the above category of investors are very correct in their convictions, it's basically far more than that. Hence, by following the news articles regarding a particular financial set up over the available online media, the future stock price of that firm may be evaluated.

Again, due to the proliferation of news articles to be digested in less time span, such a monitoring is not feasible if it is to be performed manually. Computer Science's exhilarating area provides the most successful solution to the above issue. Some instruments such as Sentiment Analysis can accomplish the task of stock price prediction. Sentiment Analysis immediately gives in just a few fractions of seconds the overall feel of the news posts.

With the astonishing popularity of these methods, it makes the developments in the stock market trends visibly clearer and easily understandable, by giving decent yield with little or no effort! The stock market is all about dynamics, hence it is extremely important to accurately forecast further movements of the stock bids. The three factors that directly or indirectly influence the decision-making on stock market investment are, after a series of events, news articles related to the company, financial health of the company and movements of the company's stock prices. In the present study, all three main aspects were taken up to provide supporting evidence of the connexion between stock prices and news articles, to create a framework that provides a company's overall financial health, and to develop a system that would evaluate and forecast stock price fluctuations over a period based on the sequence of events. For reference, here are some numbers.

The Bombay Stock Exchange (BSE) was founded in 1875 in India and is located at Mumbai. Wikipedia reveals that as of 23 January 2015, Bombay Stock Exchange is listed as the 10th largest stock market in the world with a market capitalization of $1.7 trillion. In the early 90's the internet found its way to India, whose usage has grown exponentially over the years. The Internet user base in India is the world's third largest with more than 243,198,922 users as of 2014.Internet trading (online trading) was launched by the financial markets. This data clearly shows that the users of the internet surely look for the stock market predictors in the news articles and the historic prices of the company they are interested in.

II. RELATED WORKS

In 2019, Zhong and Eake introduced a method for forecasting a collection of stocks' regular return path. Deep neural networks (DNNs) and conventional ANNs are deployed in the pre-processed but non-transformed dataset along with two datasets transformed via main component analysis (PCA) to predict the regular course of future stock market index returns. As the number of hidden layers increases steadily from 12 to 1000, a trend for the classification accuracy of the DNNs is observed and demonstrated when controlling for overfitting. Results of simulation show that DNNs using two PCA-represented datasets offer significantly higher classification accuracy than those using the entire untransformed dataset or other learning algorithms of hybrid machines. Trading strategies based on PCA-represented data, driven by the DNN classification process, perform slightly better than the others evaluated, including a comparison against two standard benchmarks.

In 2018, Pierdzioch and Risse used an ML algorithm known as boosted regression trees (BRT) to carry out an orthogonality evaluation of aggregate stock market forecast rationality. The BRT algorithm endogenously selects the predictor variables that are used to proxy the forecaster information set to optimise the predictive capacity for the forecast error. The BRT algorithm also accounts for a possible non-linear dependency on the predictor variables of the forecast error and interdependence between the predictor variables. Their key finding is that the reasonable expectations hypothesis (REH) can not be dismissed for short-term forecasts, given the set of predictor variables used in this analysis, and there is evidence against the REH for longer-term forecasts. The key finding is corroborated by findings for three distinct classes of forecasters.

In 2017, Chong, Han and Park analyse deep learning networks for stock market analysis and prediction. Deep learning networks extract features from a large set of raw data without relying on prior knowledge of predictors which makes it useful for high frequency stock market prediction. They provide an objective assessment of both the advantages and drawbacks of deep learning algorithms for stock market analysis and prediction. Using high-frequency intraday stock returns as input data, they examine the effects of three unsupervised feature extraction methods—principal component analysis, autoencoder, and the restricted Boltzmann machine—on the network's overall ability to predict future market behaviour.

In 2016, Li et. al. presents the design and architecture for a trading signal mining platform that employs an extreme learning machine (ELM) to make stock price predictions based on two data sources concurrently. Experimental comparisons between ELM and support vector machines and backpropagation neural networks (BPNNs) are made based on the intra-day data of the H-share market (shares of companies incorporated in mainland China that are traded on the Hong Kong Stock Exchange) and contemporaneous news archives. The results show that (1) both RBF ELM and RBF SVM achieve higher prediction accuracy and faster prediction speed than BPNN, (2) the RBF ELM achieves similar accuracy with the RBF SVM, and (3) the RBF ELM has faster prediction speed than the RBF SVM.

In 2016, Dash and Dash introduce a novel decision support system using a computationally efficient functional link artificial neural network (CEFLANN) and a rule set to more effectively generate trading decisions. They view the stock trading decision as a classification problem with three possible values –buy, hold or sell. The CEFLANN network usedin the decision support system produces a set of continuous trading signals by analyzing the nonlinear relationship that exists between some popular technical indicators. The output trading signals are also used to track trends and to produce trading decisions based on that trend using trading rules. This is a novel approach focused on profitable stock trading decisions through integration of the learning ability of the CEFLANN neural network with the technical analysis rules. The model is compared against other machine learning techniques such as a SVM, a naive Bayesian model, a K nearest neighbor model, and a decision tree.

In 2015, the Indian stock market prediction models of Patel, Shah, Thakkar and Kotecha were compared: ANN, SVM, the Random and the Native Bays with two modell intake approaches. Input data is initially determined using stock market data for ten technical parameters (open, big, low and near prices) while in the second process, these technical parameters are provided in the form of trend-defining data. They test with both input approaches the accuracy of each of the prediction models. The results show that random forests are above the three other prediction models for the first input data method.

By surveying different model input parameters found in nine publications chosen in 2013, Chavan and Patil help us understand the ANN stock market prediction. They try to find the main input parameters which produce better prediction precision for the model. Based on their study, most ML techniques employ technical variables rather than simple variables for a certain stock price prediction, while microeconomic variables are primarily used to forecast bond index values. Moreover, when compared to using just a single input variable type, hybridised parameters produce better results.

Dai, Wu and Lu are proposed to forecast Asian stock markets in 2012, a prediction model in time series incorporating non-linear independent components (NLICA) research and neural networks. NLICA is a new technique for extracting featured data from observed nonlinear mixing data that is not possible in the case of data mixing mechanisms. In the proposed method the input space composed of original time series data is first transformed by NLICA into the function space composed by independent components representing information from the original data.

A number of neural network and hybrid models have been proposed in Guresen, Kavakutlu and Daim in 2011 as an effort to outperform conventional linear and nonlinear bond forecasting methods, although the most success of ANN in this field is subject to some limitations. Assess multiple-layer perceptron (MLP), artificial dynamic neural network (DAN2), and hybrid neural networks using GARCH for the extraction of new input variables; Evaluation of the effectiveness

In 2011, Yeh, Huang and Lee discuss problems with kernel function hyperparameters when using vector regression to forecast stock market values. A hyperparameter is usually a parameter that is set before the study process starts. The system will combine advantages from different settings of hyper-parameters and increase overall system efficiency. By integrating a sequence minimal optimization and the gradient projecting process, they are developing a two-stage multi-kernel learning algorithm.

The purpose of this study is to identify directions for potential stock market prediction machine learning research based on an analysis of current literature. Given the structures, problem contexts and conclusions mentioned above and the taxonomy categories provided in every selected paper, numerous conclusions can be drawn on our current research knowledge. First of all, there is a clear link between the ML methods and the prediction problems. This is like task-techniques, where the output of a device is calculated by the interaction between tasks and technologies. The best way to forecast numerical stock index values is to use neural artificial networks. Supporting vector machines better suits classification concerns, such as whether the overall stock market is expected to increase or decrease. In order to classify higher quality system inputs or predict growing stock to include in a portfolio, genetic algorithms employ an evolutionary solution approach to achieve the best returns. Although each analysis showed that the methods can be used efficiently, there are restrictions on individual method implementations. One approach that can overcome some of these drawbacks is hybrid machine learning techniques. The problem is that, at some point, the systems become so complex that they are not useful in practice. This is a theoretical and practical problem that can be addressed in future studies..

III. Methodology of proposed framework

The overall process for forecasting the market direction includes many steps including data collection, pre-processing and selection of features. The programming is important for the completion of the overall study. Java has been used in R language, Java python. The software packages were used to implement the algorithms proposed. There are two types of data required for our dilemma. The prices of historical stocks and the news storey have to be derived from. Our system is based on the streaming data and the static data, unlike other systems, which have used the static data. The crawler crawls on the given site and takes out the news items of the stated company for which it is important to forecast the future direction of the stock. Since stock prices must be aligned with the news items, news items and timestamps should be collected. These news articles are then used as an input for the framework for feeling analysis.

For many years , the researchers have researched ways of interpreting feelings, and they have developed several different algorithms for classifying the feelings of the text. There are some advantages and disadvantages for any algorithm. The choice of the feeling analysis algorithm can depend on the datasets, domain and previous experience available. One approach is to be selected among the methods, linguistic, lexical and machine learning. After selecting the solution, the right algorithm has to be decided.

In the first section, The Neural predictor network is used to take stock inputs and the sentiment of news items and to generate the anticipated prices. The artificial neural network is used. The module, combination of methodology and sentiment analysis incorporates both technological and sentiment analysis advantages and offers a whole company's overall health. The investor thus makes an informed decision as to whether or not to invest in the underlying firm. The figure below is Fig. The work flow demonstrates the technological and emotional analysis combined in the section.

The following Figure 4.1 shows work flow in the module neural network predictor.

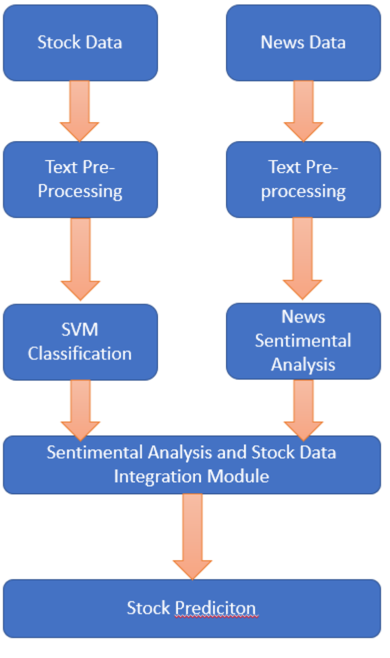

Figure 4.1: Work flow of Stock Price Predictor module

IV. TEXT PREPROCESSING

Choosing the right pre-processing technique will increase the accuracy of the Sentiment Analysis. This makes the pre-processing stage of the data very relevant. Some news article require unique pre-processing techniques, as the content created, for example, by Twitter, is generated by the user community. Data preprocessing greatly reduces the space of the term, but it is also possible to lose details.

The steps involved in data pre-processing are:

i. Tokenization: a fundamental technique might be to break the news items into no alphanumeric characters. There may be a risk of loss of knowledge; advanced techniques for text tokenization are required. This is a phase based on the domain.

ii. The deletion of the word that is common: the word may be marked with a no or little meaning of knowledge that is common (for example: a, is, be, etc.). The space for text will be greatly decreased and only the material that most relates to the recognition of feelings remains in the text.

iii. Standardisation: It is a process in which the equivalence groups of words are formed. INFOSYSS and INFY for instance.

iv. Lemmatization and Stemming: News articles may have different word forms.

V. SENTIMENT ANALYSIS ALGORITHM

The following move after the news articles are extracted is to draw the view that the article is negative, optimistic or neutral. The algorithm of sentiment analysis seeks to determine the position / sense of financial companies a news article may hold.

The steps in sentimental study, as shown in figure 4.2, are:

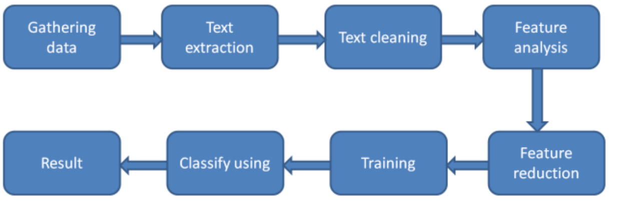

Figure 4.2: Steps in sentiment analysis

Step 1: Internet data collection is based solely on the (SOR) reference subject (e.g. ICICI bank). We use web mining techniques (e.g. crawlers) to collect all the SOR web pages.

Step 2: Text extraction can be performed using different techniques of mining or texting of data from basic "keyword matching" and "DOM structure mining" to the methods of "neural networks." The key problem is to have highly unstructured web documents and no single approach can provide 100 percent clean text extraction for all documents.

Stage 3: Text cleaning is heuristic and case specific in particular. This is to classify the unwanted parts of extracted content for the different types of web documents from phase 2, such as news articles , journals, reviews, micro-blogs and more, and then to write clear clean-up codes based on that information that deletes those undesirable portions with great precision.

Step 4: When we have the clean documentation data corpus from previous phases, the various processing engines are then moved to. This information may be evaluated according to the criteria for feature analysis or corporate analysis or market research or customer activity or consumer opinion analysis. For each such function, different techniques are used. For example, the TF IDF technology can be applied to collect a pool of different features of SOR, which can be used for the customer service of an ICICI bank, credit card, rehabilitation officer, customer loyalty, etc. The ICICI bank can obtain. This is known as an analysis function. From this pool one can determine things like how many consumers talk about recovery agent while talking about credit cards compared to how many consumers talk about customer satisfaction while talking about credit cards.

Step 4.1: The following feature analysis methodology inevitably brings up some unnecessary features into the pool that need to be excluded from final analysis. This can be achieved by mapping functions with keywords that reflect various well-known characteristics from the pool. The SOR can also describe these features beforehand.

Step 5: The text to be graded as Positive , Negative or Neutral is used to evaluate feelings. Sentiment analyses can be performed in two ways-manual rating or automatic evaluation of certain Web documents-on the cleanly extracted web documents. While manual rating is a nearly perfect way to do that, it is sluggish when web papers are too big. In comparison, automated systems are much faster, but are bound to be imprecise because they effectively machine learning and produce human feelings through content produced by users. The language barrier is also a big obstacle for the automated study of feelings. Nevertheless, comprehensive research on the natural language processing has well tackled these problems and fairly high-performance machine learning techniques that can interpret the feelings of web documents have developed. Naive Bayesian Classifier (NBC) and Maximum Entropy Support Vector Machines (SVM) are the most powerful techniques. In addition, there are also lexicon-based techniques. These learning algorithms need an apprenticeship, which then can lead you to feelings from web documents.

Step 5.1: NBC / SVM training is a fairly simple and well-studied technique where a manually classified corpus is created of several thousand web materials which classifies them as negative, positive or neutral for a particular SOR. The NBC / SVM motors then feed these corpusses into one of the three categories (negative , positive ou neutral), which generate several parameters for a given paper.

Phase 5.2: After training of NBC / SVM engines, other web documents can be classified using their generated parameters. The precision will certainly not be 100 percent, but many training and tuning layers will boost and optimise precision. When the feeling is taken out these values will be paired with the historical values of the stock market and the course is expected. As already mentioned, techniques such as Artificial Neural Networks and For the forecast, NBC is used. In an artificial neural network, five input neurons, two hidden layers and one output neuron are used. The historical price and the feeling score are the machine inputs and the training teaches the network. On the basis of the historical stocks and sentiment values, it produces the next day price. In NBC, the classifier is educated and news items categorized as positive, negative or neutral based on their likelihood. The classifier is taught.

VI. MEASURES OF EVALUATION

Accuracy - It may be defined as the number of correct predictions made by our ML model. We can easily calculate it by confusion matrix with the help of following formula –

Accuracy=

(TP+TN) / (TP+FP+FN+TN)……………….……….5.1

Precision - Precision, used in document retrievals, may be defined as the number of correct documents returned by our ML model. We can easily calculate it by confusion matrix with the help of following formula –

Precision=

TP / (TP+FP)…………………………………..5.2

Recall or Sensitivity - Recall may be defined as the number of positives returned by our ML model. We can easily calculate it by confusion matrix with the help of following formula −

Recall=

TP / (TP+FN)……………….……………………5.3

F- Measure or Specificity - Specificity, in contrast to recall, may be defined as the number of negatives returned by our ML model.

F- Measure =

TN / (TN+FP)…………………………………5.4

VII. RESULT

Figure 5.1: Results of Stock Market Analysis of Facebook Inc. (FB) Prices.

Figure 5.2: Results of Stock Market Analysis of Alphabet Inc. (GOOGL) Prices.

Table5.4: Comparison of classifiers

Model	Precision	Recall	F1 Sore	Accuracy
SVM	71.34%	72.11%	71.23%	74.3%
Auto Regressive Model	86.5%	85.3%	85.11%	87.6%
BRT	89.3%	88.8%	89.4%	90.34%
ANN/DNN	91.8%	90.3%	91.2%	92.5%
Proposed Model	95.5%	94.2%	95.3%	96.9%

CONCLUSION

Only businesses with negative news have been chosen to accomplish the third goal and the findings have shown that the stock price of the same company has declined immediately after the bad news. It shows that there is a further rise in trend after an occurrence. Empirical evidence from the work carried out in this study provides the investor with an educated decision to invest in a business safely and to achieve a high return. In achieving our three goals, the modern approach of sentiment analysis introduced in our work has been applied effectively, which can be generalised to real life.

References

[1] X. Zhong and D. Enke, “Predicting the daily return direction of the stock market using hybrid machine learning algorithms,” Financ. Innov., vol. 5, no. 1, p. 4, 2019.

[2] C. Pierdzioch and M. Risse, “A machine-learning analysis of the rationality of aggregate stock market forecasts,”, Int J Fin Econ, vol. 23, no. 4, 642-654, 2018 [doi:10.1002/ijfe.1641].

[3] F. Cabitza et al., “Machine learning in orthopedics: A literature review,” International Journal of Finance & Economics, Front. Bioeng. Biotechnol., vol. 6, p. 75, 2018 [doi:10.3389/fbioe.2018.00075].

[4] E. Chong et al., “Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies,” Expert Syst. Appl., vol. 83, pp. 187-205, 2017 [doi:10.1016/j.eswa.2017.04.030].

[5] X. Li et al., “Empirical analysis: Stock market prediction via extreme learning machine,” Neural Comput. Appl., vol. 27, no. 1, pp. 67-78, 2016 [doi:10.1007/s00521-014-1550-z].

[6] R. Dash and P. K. Dash, “A hybrid stock trading framework integrating technical analysis with machine learning techniques,” J. Fin. Data Sci., vol. 2, no. 1, pp. 42-57, 2016 [doi:10.1016/j.jfds.2016.03.002].

[7] J. Patel et al., “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques,” Expert Syst. Appl., vol. 42, no. 1, pp. 259-268, 2015 [doi:10.1016/j.eswa.2014.07.040].

[8] P. S. Chavan and S. T. Patil, “Parameters for stock market prediction,” Int. J. Comput. Technol. Appl., vol. 4, no. 2, p. 337, 2013.

[9] W. Dai et al., “Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes,” Expert Syst. Appl., vol. 39, no. 4, pp. 4444-4452, 2012 [doi:10.1016/j.eswa.2011.09.145].

[10] D. Y. Chiu and P. J. Chen, “Dynamically exploring internal mechanism of stock market by fuzzy-based support vector machines with high dimension input space and genetic algorithm,” Expert Syst. Appl., vol. 36, no. 2, pp. 1240-1248, 2009 [doi:10.1016/j.eswa.2007.11.022].

[11] S. P. Das and S. Padhy, “Support vector machines for prediction of futures prices in Indian stock market,” Int. J. Comput. Appl., vol. 41, no. 3, 2012.

[12] E. Guresen et al., 2011, Using Artificial Neural Network Models in Stock Market Index Prediction. Expert Systems with Enke.

[13] D. Enke and S. Thawornwong, “The use of data mining and neural networks for forecasting stock market returns” Expert Syst. Appl., vol. 29, no. 4, pp. 927-940, 2005 [doi:10.1016/j.eswa.2005.06.024].

[14] D. L. Goodhue and R. L. Thompson, “Task-technology fit and individual performance,” MIS Q., vol. 19, no. 2, pp. 213-236, 1995 [doi:10.2307/249689].

[15] J. H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, vol. 38, no. 8, pp. 10389-10397.1992.

[16] T. Jasic and D. Wood, “The profitability of daily stock market indices trades based on neural network predictions:

Case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965–1999" Appl. Financ. Econ., vol. 14, no. 4, pp. 285-297, 2004

[17] H. J. Kim and K. S. Shin, "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets," Appl. Soft Comput., vol. 7, no. 2, pp. 569-576, 2007 [doi:10.1016/j.asoc.2006.03.004].

[18] K. J. Kim and I. Han, "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index," Expert Syst. Appl., vol. 19, no. 2, pp. 125-132, 2000 [doi:10.1016/S0957-4174(00)00027-0].

[19] K. J. Kim and W. B. Lee, "Stock market prediction using artificial neural networks with optimal feature transformation," Neural Comput. Appl., vol. 13, no. 3, pp. 255-260, 2004 [doi:10.1007/s00521-004-0428-x].

[20] M. J. Kim et al., "An evolutionary approach to the combination of multiple classifiers to predict a stock price index," Expert Syst. Appl., vol. 31, no. 2, pp. 241-247, 2006 [doi:10.1016/j.eswa.2005.09.020].

[21] L. Kumar et al., "A hybrid machine learning system for stock market forecasting," J. Int. Technol. Inf. Manag., vol. 20, no. 1, p. 3, 2011.

[22] K. H. Lee and G. S. Jo, "Expert system for predicting stock market timing using a candlestick chart," Expert Syst. Appl., vol. 16, no. 4, pp. 357-364, 1999 [doi:10.1016/S0957-4174(99)00011-1].

[23] M. C. Lee, "Using support vector machine with a hybrid feature selection method to the stock trend prediction," Expert Syst. Appl., vol. 36, no. 8, pp. 10896-10904, 2009 [doi:10.1016/j.eswa.2009.02.038].

[24] Z. Liao and J. Wang, "Forecasting model of global stock index by stochastic time effective neural network," Expert Syst. Appl., vol. 37, no. 1, pp. 834-841, 2010 [doi:10.1016/j.eswa.2009.05.086].

[25] R. Malhotra, "A systematic review of machine learning techniques for software fault prediction," Appl. Soft Comput., vol. 27, pp. 504-518, 2015 [doi:10.1016/j.asoc.2014.11.023].

[26] J. O et al., "Adaptive stock trading with dynamic asset allocation using reinforcement learning," Inf. Sci., vol. 176, no. 15, pp. 2121-2147, 2006 [doi:10.1016/j.ins.2005.10.009].

[27] P. Ou and H. Wang, "Prediction of stock market index movement by ten data mining techniques," Mod. Appl. Sci., vol. 3, no. 12, pp. 28-42, 2009 [doi:10.5539/mas.v3n12p28].

[28] J. Pound, 2019, Dec. 24, "Global stock markets gained $17 trillion in value in 2019". Available at: https://www.cnbc.com/2019/12/24/global-stock-markets-gained-17-trillion-in-value-in-2019.html.

[29] B. Qian and K. Rasheed, "Stock market prediction with multiple classifiers," Appl. Intell., vol. 26, no. 1, pp. 25-33, 2007 [doi:10.1007/s10489-006-0001-7].

[30] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," ACM Trans. Inf. Syst. (TOIS), vol. 27, no. 2, p. 12, 2009.

[31] R. P. Schumaker and H. Chen, "A discrete stock price prediction engine based on financial news," Computer, vol. 43, no. 1, pp. 51-56, 2010 [doi:10.1109/MC.2010.2].

[32] J. Wen et al., "Systematic literature review of machine learning based software development effort estimation models," Inf. Softw. Technol., vol. 54, no. 1, pp. 41-59, 2012 [doi:10.1016/j.infsof.2011.09.002].

[33] L. Yu et al., "Evolving least squares support vector machines for stock market trend mining," IEEE Trans. Evol. Comput., vol. 13, no. 1, pp. 87-102.Innovation, 5(1), 4, 2008.