

HOW INVESTORS LEVERAGING GAIN IN STOCK MARKET INVESTMENTS: A PREDICTIVE ANALYSIS

Naliniprava Tripathy, Indian Institute of Management Shillong

ABSTRACT

The purpose of the study is to predict the movement of the direction of the Nifty Technology stock market price in India. The study has used Support Vector Machines, Random Forest, and Logistic Regression models, to the time series data to estimate the movement of the direction of the stock market price. The prediction value of all three models is compared to shed light on the performance efficiency of each algorithm. The results of the study indicate that the LR model provides an average of 60.73% accuracy for predicting the movement of the direction of the Nifty Technology stock market price. The outcomes are slightly encouraging results of estimating the direction. Nevertheless, the findings would help the investors, institutional investors, portfolio managers, and foreign investors to develop an efficient market trading strategy to make buy, hold, and sell decisions before making investment decisions. It will also facilitate the company in forecasting financial quotes and enable policymakers to understand the influential factors, which affect to stock market movement. This study will not only be helpful to the regulators but also be of great value to the Investors.

Keywords: Support Vector Machine, Random Forests, Logistic Regression, Stock Market Prediction.

JEL classifications: G1, C32, C53.

INTRODUCTION

Stock market data are nonlinear, and chaotic in nature. Many factors, such as physiological, rational, and irrational behavior, make share price volatility, leading difficult to predict the share price. Therefore, the movement of the direction of the stock price has always been an important issue and subject of interest for most of the professional analysts. Efficient Market Hypothesis debated that any information received by the market reaches to all investors concurrently and the current stock price fails to affect by its past price. Hence, investors cannot gain any profit by using this information. EMH subsists in three procedures: weak, semi-strong, and strong. In weak form, the current price only contains historical information, so it rules out predictions. The semi-strong form incorporated all current as well as past information into the stock price. The strong form proclaimed that all past, public, and private information, along with insider information, is fully emulated in the stock price, indicating that forecasting of the stock price is not possible. Therefore, forecasting the stock price is a highly challenging task. However, researchers have proposed Machines' learning models that can capture the nonlinear behavior of the stock market and predict the stock price. Therefore, the Support Vector Machines model employed to forecast the stock market price during the last decade. SVM model performs more efficiently than ANNs do. (Kim, 2003). Market direction is significant for investors or traders. Usually, the market is moving either upward or downward, and the nature of the market movement is binary. If the stock market price would be predicted, it will not only help the investors to make better investment decisions but also enable them to yield significant income.

Hence, the projection of the movement of the stock price is a topic of concern for investors, traders, researchers, and policymakers today. However, few research works have undertaken, especially after the financial crises period (2008), to forecast the movement of the direction of the stock market price in India. The present research study tries to fill this gap. The present study has used three models, such as the Support Vector Machine, Random Forest, and Logistic Regression model, to forecast the change of the direction of the stock market price of India. Secondly, the study assessed the superiority of prediction accuracy comparing three models to facilitate the investors to take better investment decisions in the market.

The remaining of the paper planned as follows: Section 2 describes the past reviews of study, and then the Machine Learning approach, data, and methodology are briefly explained in section 3. The results and analysis are exhibited in section 4. The concluding observations are provided in section five.

LITERATURE REVIEW

Support Vector Machines (SVMs) have actively applied in different areas and comprehensively researched in the machine learning community for the last decade. Kim (2003) predicted the direction of change in the daily Korean Composite Stock Price Index 200 using the SVM and ANN model. The results indicated that SVM is better than the ANN model in predicting the future direction of a stock market price. The SVM model delivered a prediction accuracy of 57.8 %, which was above the 50% threshold. Huang et al. (2004) used SVM and Back Propagation Neural Network (BNN) in the United States credit rating analysis and indicating the prediction accuracy around 80%. Huang et al. (2005) explored the probability of the weekly movement direction of Nihon Keizai Shimbun Index 225 (NIKKEI 225) using Linear Discriminate Analysis; Quadratic Discriminate Analysis and Elman Back-Propagation Neural Networks. The outcomes of the analysis exhibited that SVM is superior to other methods. Chen & Shih (2006) compared SVM and Back-Propagation (BP) Neural Networks when predicted the six major Asian stock markets. SVM model showed better performance in four out of six markets than the BP model. Shah (2007) used various machine-learning models to estimate the stock prediction and found that the best result achieved with SVM is 60 %. The study is in line with Kim's conclusion. Yakup et al. (2011) used the ANN and SVM model to forecast the change of direction of the Istanbul Stock Exchange (ISE) National 100 Index. The study signposted the average performance of the ANN model is 75.74%, and the SVM model is 71.52%, demonstrating that the performance of the ANN model is better than the SVM model.

Jigar, et al. (2015) studied predicting future values of the stock market price of India using Support Vector Regression (SVR) in the first stage. The study further used an Artificial Neural Network (ANN), Random Forest (RF) model for predicting the future values of the stock market price. The study compared hybrid SVR-ANN, SVR-RF, and SVR-SVR with SVR, ANN, and RF models. The outcomes of the study indicated that hybrid models are better prediction models than a single-stage model for forecasting the stock price of India. Madge (2015) predicted the stock prices of the NASDAQ-100 Technology Sector Index using the SVM model. The study found the prediction of accuracy price direction is 80 %. Makram and Amina (2016) used logistic model and artificial neural network model to forecast the Saudi Arabian stock market trends and indicated that both models deliver prediction accuracy 81%. Hakob (2016) forecasted the closing price of the Bucharest Stock Exchange using the SVM and SVM-ICA model. The study used independent component analysis (ICA) to select the input variables from the technical and fundamental analysis. The study found that the SVM-ICA model beats the SVM-based model in forecasting the stock price. Aparna et al. (2016) projected the stock market trends employing Boosted Decision Tree, Logistic Regression, and Support Vector Machine models and indicated 70% accuracy found in the daily prediction of the stock market price. Amin et al. (2016) forecasted the daily NASDAQ

stock using an artificial neural network model. The study used three hidden layers, and specified 20-40-20 neurons are the optimized networks for prediction. Xiao and David (2017) forecasted the daily direction of the S&P 500 Index ETF (SPY) employing principal component analysis (PCA), fuzzy robust principal component analysis (FRPCA), and kernel-based principal component analysis (KPCA) to simplify the original data structure. The artificial neural networks model applied to forecast the direction of the SPY for the next day. ANNs with the PCA provides superior accuracy than other methods.

Adam et al. (2018) investigated the impact of news-derived information on market volatility in the US market and signposted that the prediction accuracy of volatility in the arrival news is 56%. Bruno et al. (2018) used Support Vector Regression (SVR) to predict the stock prices of Brazilian, American, and Chinese stocks and specified that the American blue-chip stocks are correlated adversely with average return and predicted price. The findings of the study also demonstrated that there is a robust association exists between prediction errors and volatility, suggesting that smaller prediction errors create by using a fixed training set on daily price than using a linear kernel in the training set. Suryodayet al. (2019) deliberated on the direction of stock market price movements using random forests, and gradient boosted decision trees. The study considered stock price movements as a classification problem and used machine-learning techniques to predict the direction of the stock price. The model predicted the direction accuracy, which is 78%, and preferred the random forests to the decision trees model for forecasting the same. KiHwan and NohYoon(2019) predicted the stock price movement based on financial news using the machine learning model. The experimental results of the study indicated that the stock price directional changes could be forecasted without news effects on the firm.

METHODOLOGY

The required time series closing price data collected from www.nse.com for a period of five years from January 2013 to December 2018 of 23 stocks. The study has taken the Nifty Teck index, which comprises the technology sector, media, & publishing, telecommunications sectors (TMT). Since the development of the TMT sector as a significant force in the Indian economy today and the extraordinary growth of this sector reflected in the financial markets, the study treated Nifty Tech Sensex as a proxy for the entire market. The data alienated into two sets, such as the training set and test set. The first set comprises records required for the model training and the second set necessary for model testing. The five years of data have divided into training and validation data set based on random selection.

Number of companies 23 + 1

Number of years 5

Total data points 30,744

Volatility Calculations N: (5, 15, 90, 180)

Prediction days in future Days: (1, 7, 15, 30, 90, and 180)

As inputs, two csv files were loaded into the model. Format of the file is as under:

Date	TCS	Infosys	Wipro	Cyient	HCL Tech	Mindtree	Hexaware Technology
2-Sep-13	44.45	3077.35	484.2	185.1	1037.9	1052.95	128.9
3-Sep-13	43.2	3033.85	479.2	185.15	1022.9	1045	131.05
4-Sep-13	44.8	3105.9	481.55	186.4	1040.35	1046	128.25
5-Sep-13	44.95	3002.7	468.35	188.9	1016.95	1042.85	126.25

6-Sep-13	45.3	3020.85	468.4	187.55	1016.4	1034.85	125.45
----------	------	---------	-------	--------	--------	---------	--------

Machine learning aims to create an accurate model based on past data to predict future events. There are two significant types of machine learning, such as regression and classification. In this paper, the prediction of the stock market price problem treated as a classification problem. The classification model delivers a probabilistic view of analysis to predict the change of the trend. Hence, the study tries to use machine-learning models to learn from the market data to forecast the direction of stock price to determine the likely benefit leveraging in the market.

The Machine Learning Approach

Machine learning methods allow a machine to decide with explicit programming (Avrim & Langley, 1997). Machine learning techniques typically split into two sets – supervised learning and unsupervised learning. A set of training data supplied to the machine in supervised learning to learn them, whereas, in unsupervised learning, no training data provided. The unsupervised learning algorithm attempts to determine a similarity between the data without any given labels (Avrim & Langley, 1997). Supervised learning further split into classification and regression problems. A set number of outputs can be labeled in classification problems, whereas the output can take on continuous values in regression problems.

The study predicts the direction of the daily changes of the Nifty Teck index. This trend modeled as a two-class classification problem. Hence, the classes are leveled with 0 and 1. For each day, an upward and downward change is calculated. It means “0” indicates the daily closing price is lower than the closing of the previous day. ($U_t = 0, D_t = S_{t-1} - S_t$) i.e. fall in the stock price and “1” means “Up days” are characterized by the daily closing price S_t being higher than the closing of previous day S_{t-1} ($U_t = S_t - S_{t-1}, D_t = 0$) i.e. a rise in the stock price.

Overall, machine-learning algorithms used to develop three training models:

- Support Vector Machine
- Random Forest Algorithm
- Logistic Regression

Support Vector Machine

The primary objective of SVM is to identify the maximum margin hyperplane. Vapnik and co-workers developed support Vector Machine, a supervised machine-learning algorithm based on statistical learning theory in 1992. SVM is useful for data classification and regression analysis. It is also an effective method for pattern recognition and regression. SVM mostly used in classification problems such as classification of linear and nonlinear data. SVM creates a boundary, and the data points on either side of the border labeled differently (Keerthi et al. 2005). The edge in the multidimensional case called a hyperplane. The most critical elements in this technique are the data points closest to the hyperplane. The optimal hyperplane maximizes the distance of the hyperplane from the extreme points on either side of the hyperplane. The optimal hyperplane referred to as the maximum margin hyperplane (MMH). The hyperplane selection only based on these extreme points. These extreme data points called the support vectors and the maximum margin

hyperplane known as the Support Vector Classifier (SVC). Support vector regression is dissimilar from traditional regression. In ordinary least squares regression, the optimal line is the one for which the sum of error minimized.

In SVM, the SVC is the hyperplane for which the sum of the distance between the hyperplane and the support vector is the maximum. SVM provides not only linear boundaries but also models nonlinear hyperplane. SVM does not condense the empirical risk of making a few mistakes but pretends to form consistent models with future data called Structural Risk Minimization (SRM) in statistical learning theory. SVM uses the Structural Risk Minimization (SRM) principles and seeks to reduce an upper bound generalization error rather than training error to execute superior to conservative techniques. SVM is usually resilient to the over-fitting problem.

In this study, the SVM model with RBF Kernel function used for stock market price forecasting. The function represented as follows:

$$K(x_i, x_j) = \exp\left(-\frac{1}{\sigma^2} \sum_{j=1}^N (x_{ij} - x_k)^2\right) \dots \dots \dots (1)$$

Where σ is the bandwidth of the kernel function, a bandwidth of radial basis function is represented by σ . If the value of σ fall in-between 0.1 and 0.5, the SVR model achieves the best performance. In this paper, the σ value is determined as 0.5. The R software used to accomplish the experiment.

SVM provides not only linear boundaries but also models nonlinear hyperplane. The radial basis kernel function employed for transmuting the input space to the higher dimension space. The advantage of the radial kernel function is that it can output a nonlinear boundary function. It also matches the test data points to the training data points using a minimum Euclidean distance. It weights closer training data points heavily and then outputs the predicted label.

Random Forest Algorithm

Random Forest model is used for both classification and regression tasks. Leo Breiman (2001) first proposed the introduction of RF in classification and regression trees (CART) in a research paper. Random Forest creates the forest with some decision trees and merges them to improve the predictive accuracy and control over-fitting. The Random Forest algorithm used to identify the most critical features of the training dataset. The RF model prediction is the mean of the projections of all individual trees in the RF. The individual trees are different from one another. Hence there is no correlation exists between their predictions that lead to predicting the RF generalized better. When K trees are combined, the forecasted decision is gained as the average value over these K trees. Random Forest algorithm is used to identify a stock's behavior and minimize the risk of investment in the stock market.

Logistic Regression

Logistic regression falls under supervised learning and measures the relationship between dependent variables and independent variables by assessing probability using the sigmoid function. The study used logistic regression since it assumes that the relation of the variables used in the study is non-linear. However,

logistic regression is not used for machine learning regression problems to forecast the real value output; instead, it is a classification problem used to predict a binary outcome (1 or 0, -1 or 1) given a set of independent variables. The Logistic regression objective is to find the best fitting model to elucidate the association between the binary characteristic of interest and the collection of independent variables. Since the market is moving downward trend and upward, the nature of the movement of the market is binary. Logistic regression model fit a model using binary behavior and predict the market direction. Logistic regression assigns a probability to each event.

RESULTS AND ANALYSIS

The descriptive statistics of the closing price of the Nifty Teck index are presented in Table 1. The daily mean return of the stock market is positive. The stock market return exhibits negative skewness and kurtosis >3 , indicating that the returns are not normally distributed. The Jarque-Bera statistics are greater than zero, indicating that the series is leptokurtic, exhibits non-normality, and shows the presence of Heteroscedasticity. Further, the unit root test is evaluated to test the stationarity of the data series.

Statistics	Nifty Teck index
Mean	0.000736
Std. Dev.	0.010222
Skew ness	-0.429272
Kurtosis	5.392528
Jarque-Bera	338.6790(0.00)

The study has used the KPSS test and DF-GLS statistic to measure the stationarity of time series data. The Table-2 shows that the KPSS test and DF-GLS statistic for all variables are stationary at levels and rejected the null hypothesis. The lag values are selected by AIC and SIC 's criteria.

Variable	KPSS	DF-GLS
Nifty Teck index	0.241340	-2.910083
Asymptotic Critical values [*]		
1% level	0.739000	-2.566819
5% level	0.463000	-1.941078
10% level	0.347000	-1.616528

	Nifty Teck index closing price
BDS statistics	0.007718
z-Statistic	3.473384
Epsilon	0.014197
(P -value)	(0)

Table 3 presents the BDS test. The result of the BDS test indicates that there are significant BDS statistics found in the Indian stock market. Therefore, it suggests that nonlinearity presence in the stock market. This result justifies the use of the non-linear model in the stock market.

Table 4			
GARCH MODEL			
	Coefficient	z-Statistic	Prob.
intercept	0.000771	2.674553	0.0075*
Variance Equation			
intercept	5.35E-06	3.303863	0.0010*
ARCH (1) α	0.069392	6.357709	0.0000*
GARCH (1) β	0.883582	39.44166	0.0000*
Log likelihood	4052.745		
Akaike info criterion	-6.326164		
Schwarz criterion	-6.310056		
Durbin-Watson stat	1.742809		

The output of the GARCH model is illustrated in Table 4. The estimated parameters α and β capture the short-run dynamics of the volatility. It can be seen from Table 5 that the α coefficient indicates the latest news, and the β coefficient signifies past news that is statistically significant at 1% level. Therefore, the results of the study reporting that past and present news of the stock market influencing the volatility of the stock market. $\alpha + \beta$ measures the degree of persistence of volatility shocks.

The degree of persistence of volatility is 0.94, indicating that ARCH and GARCH coefficients are very close to one, postulating the presence of volatility clustering persistent in the Indian stock market. It specifies that any significant change in Indian stock market returns tends to be followed by large changes, and small changes in Indian stock market returns tend to be followed by small changes.

Further, the data alienated into training, validation, and test sets. The model building made using the training set and the validation set used for parameter optimization. For evaluating the model, the test set is used. The study adopted a Radial Basis Function (RBF) on sample data for SVM. The RBF kernel function involves parameter σ in addition to the regular parameters C and ϵ . The performance of the model evaluated after renormalizing the output generated by the models.

The market return used to calculate the direction of movement for one day - ahead of the market price (M=prediction accuracy). M takes the following values, such as \in (1, 7, 15, 30, 90, and 180).

The Table 5 summarizing the prediction accuracy of the three models at various input parameters:

Table 5					
PREDICTION ACCURACY OF THREE MODELS					
NCO*	NIN*	Pred-Days*	SVM	RF	LR
5	5	1	51.9%	50.3%	53.2%
5	5	7	51.7%	50.2%	50.8%
5	5	15	52.4%	50.8%	52.4%
5	5	30	54.0%	51.5%	52.8%
5	5	90	56.4%	52.4%	56.6%

5	5	180	60.1%	56.5%	61.4%
15	15	7	52.0%	50.7%	51.8%
15	15	180	56.5%	56.0%	59.2%
90	90	7	53.6%	52.3%	51.6%
90	90	180	55.7%	56.0%	61.3%
180	180	7	52.0%	50.6%	51.0%
180	180	180	57.5%	55.0%	61.0%

Table 5 depicts the number of days taken to calculate the volatility of the stock price of companies. It also exhibits the number of days in the future where the movement of the stock price is predicted in comparison to the present day. The above analysis also shows the accuracy over the validation data of all the companies by using the Support Vector Machine model, Random Forest Model, and Logistic Regression model. It is observed from the analysis that the accuracy increases when the time window increases for SVM, RF, and LR, respectively. The table shows that the prediction accuracy peaked at 60.1% in 180 days for SVM, 56.5% for RF, and 61.4% for LR.

Further, the Logistic Regression model gives the highest accuracy when the time window is 180 days (61.4%) in comparison to other models indicating the most stable model. The results suggest that the model can predict price direction with 61.4% accuracy. The study specifies that long-run stock direction predicted, and the price reacts not only to new information but also to existing news. This study violating the EMH since the research indicates that the mean price not only responded to new information transmitted to the market instantaneously but also to the past data.

The confusion matrices of the three models under different scenarios are presented in Table 6.

NCO PreDays		NIN	SVM		RF		LR	
			0	1	0	1	0	1
5	5	180	0	727 2047 26.21% 60 .10%	951 1823 34,28% 56.54%	701 2073 25.27% 60.10%		
			1	972 3821 79.72%	1465 3328 69.43%	846 3947 82.35%		
5	5	90	0	966 2479 28.04% 56.40%	1284 2161 37.27% 52.00%	867 2578 25.17% 56.6%		
			1	1095 3648 76.91%	1739 3004 63.34%	973 3770 79.49%		
5	5	45	0	1297 2660 32.78% 55.20%	1680 2277 42.46% 52.70%	1121 2836 28.33% 55.00%		
			1	1152 3401 74.7%	1752 2801 61.52%	992 3561 78.21%		
5	5	15	0	1882 2557 42.4% 52.00%	2087 2352 47.02% 50.80%	1685 2754 37.96% 52.40%		
			1	1595 2683 62.72%	1936 2342 54.75%	1393 2885 67.44%		

In the confusion matrices

0 s and 1s in in the second row depicts predicted downward and upward movement of stock price respectively.

0 s and 1s in in the fourth column depicts actual downward and upward movement of stock price respectively.

The table gives a summary of:

True Positives

False Positives

True Negatives

False Negatives

The Table 6 depicts that ‘0’ and ‘1’ indicating the predicted and actual downward and upward movement of the stock price, respectively. The accuracy measures the portion of all testing samples and indicating that the predicted downward and upward movement of the stock price is 28.04% and 76.91%. The actual downward and upward movement of the stock price is 42.4 % and 62.72%, respectively, under the SVM model. Similarly, the predicted downward and upward movement of stock price under RF is 32.27% and 63.34%. The actual downward and upward movement of the stock price is 47.02 % and 54.75%, respectively. It also noticed that the predicted downward and upward movement of stock price under the logistic regression model is 25.17% and 79.49%. The actual downward and upward movement of the stock price is 37.96 % and 67.44%, respectively. It is noticed that the prediction day in future increases as and when, overall accuracy of all the models tend to increase significantly. The analysis also indicating that accuracy of predicting positive movement of stock is much higher than random guessing while predicting downward movement is lower than random guessing. As prediction day in future increases, accuracy of predicting upward movement rises as high as 82%. Similarly, as prediction day in future increases, accuracy of predicting downward movement down as low as 25%. Table 7 uses the sensitivity measure to identify positive levels and adverse levels and gives a summary of True and False Positives, True and False Negatives.

Table 7			
RANKINGS OF THREE MODELS			
	SVM	RF	LR
Overall Average	54.49%	52.69%	55.26%
Average of Best in Ranking	57.45%	55.88%	60.73%
	2	3	1

Table 7 summarizes the rankings of the three models based on the accuracy of their prediction. The study compares the non-linear models and results shown in Table 7. It observed that the logistic regression model performs superior to the SVM and RF model. Therefore, it indicates that the LR model appears to be the most reliable, and prediction accuracy is 60.73%. It also suggests that as and when the prediction day in future increases, the overall efficiency of all the models tend to increase significantly. The results of the study proposed that long-term forecasting change of price direction probably depends more on the overall market trends and macro conditions of the economy.

Concluding Observations and Managerial Contribution

The study forecasts the change of the direction of the Indian stock market from January 2013 to December 2018 of 23 stocks taking time series data of daily Nifty Teck index, which comprises the technology sector, media, & publishing, telecommunications sectors (TMT) using SVM, Random forest, and the logistic regression model. The study also used Hit Ratio to find the accuracy of the model. The study used the GARCH model and the results indicating that the past and present news influencing the

volatility of the stock market and the presence of volatility clustering is persistent in the Indian stock market. The findings of the result indicate that the prediction accuracy peaked at 60.1% in 180 days for SVM, 56.5% for RF, and 61.4% for LR. The findings also suggested that the Logistic Regression model delivers the highest accuracy in comparison to another model. Different combinations of the feature set used to detect an efficient combination. Therefore, it indicates that the LR model appears to be the most reliable and presents the average prediction accuracy of 60.73%. The findings of our study support the findings of Kim (2003), Shah (2007) and Adam et al. (2018).

The outcomes of the study have a significant contribution to policymakers, investing community, and companies at large. The research findings are vital to the investing community since traders can predict the overall market return and risk, which can be used for the trading market index. Accurate forecasting of stock market movement direction is exceptionally significant for articulating the best market trading solutions. Many people have invested their money in the stock market also earn a profit. The findings of the study will help investors, institutional investors, portfolio managers, and foreign investors by predicting the returns of stock to make a proper investment decision to minimize their risk and maximize returns. This study will also help to develop an efficient market trading strategy and enable us to make buy, hold, and sell decisions before making investment decisions. The companies also benefited in forecasting financial quotes if stock price direction predicted accurately. Another contribution of the study is that it also provides information to the policymakers to understand the factors, which are most influential to the stock market movement.

The future study can consider other than tech companies to check the accuracy and ranking of different models. Further, in the present study, for logistic regression, 0.5 probability is taken as cut off. This value can be changed to improve the overall accuracy of the model also for improving the accuracy of detecting upward movement or downward movement. For random forest algorithm, number of trees is set to 500. It can be reconnoitered to arrive at better accuracy of the model. The other models such as like Naïve Bayes and KNN model can also be explored to check whether these models deliver better prediction accuracy. Further, other features such as the P/E ratio, market capitalization ratio, and Tobin's Q ratio can be considered for future research, which also profoundly influencing the stock market.

REFERENCES

- Adam, A., Mahesan, N., & Enrico, Gerding. (2018), "Financial news predicts stock market volatility better than close price", *The Journal of Finance and Data Science*, Vol.4, Issue.2, pp.120-137.
- Amin, H., Moghaddam, Moein, H.M. & Morteza, E. (2016) "Stock market index prediction using artificial neural network, *Journal of Economics, Finance and Administrative Science* 21(41), 89-93.
- Aparna, N., M.M. Manohara, P. & Radhika, M.P (2016), "Prediction Models for Indian Stock Market", *Procedia Computer Science*.89, 441- 449.
- Avrim, L., & Pat, L. (1997), "Selection of relevant features and examples in machine learning", *Artificial Intelligence*, .97(1-2), 245-271.
- Bruno, M.H., Vinicius, A.S., & Herbert, K. (2018), "Stock price prediction using support vector regression on daily and up to the minute prices", *The Journal of Finance and Data Science* .4, (3), 183-201.
- Chen, Wun-Hua., & Shih, J.Y. (2006), "Comparison of support vector machines and back propagation neural networks in forecasting the six major Asian stock markets", *International Journals Electronics Finance* 1(1) 49-67.

- Hakob, G. (2016), "A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)" *Database Systems Journal*, 7(1) 12-21.
- Huang, W., Nakamori, Y. & Wang, S.Y. (2005), "Forecasting Stock Market Movement Direction with Support Vector Machine", *Computers and Operations Research*, 32, (10) 2513–2522.
- Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., & Wu, S. (2004), "Credit rating analysis with support vector machines and neural networks: a market comparative study" *Decision support systems*, Vol.37, No.4, pp.543-558
- Hsu S.H., Hsieh J.J. P. A., Chih, T.C., & Hsu, K C. (2009), "A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector Regression" *Expert Systems with Applications*, Vol.36, No.4, pp.7947–7951
- Jigar, P., Sahil, S., Priyank, T.K. Kotecha. (2015), "Predicting stock market index using fusion of machine learning Techniques", *Expert Systems with Applications* Vol.42, No.4, pp.2162-2172
- KiHwan, Nam., & NohYoon, Seong. (2019) "Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market", *Engineering Applications of Artificial Intelligence* 82, .1-12.
- Keerthi, S.S. Duan, K.B., Shevade, S.K., Poo, A.N., (2005), "A Fast Dual Algorithm for Kernel Logistic Regression", *Machine Learning*, Vol.61, Issue.1–3, pp.151–165
- Kim, K. (2003), "Financial time series forecasting using support vector machines", *Neurocomputing*, 55, (1) 307-319.
- Leo, B. (2001), Statistics Department, University of California Berkeley, CA 94720. Random Forests.
- Madge, S. (2015), "Predicting Stock Price Direction using Support Vector Machines", Independent Work Report Spring, - pdfs.semanticscholar.org
- Makram, Z., & Amina, A. (2016), "Forecasting stock market trends by Logistic Regression and Neural Networks: evidence from KSA stock Market", *International Journal of Economics, Commerce and Management*, 4(6), 220-234.
- Suryoday, B., Saibal, K., Snehanshu, Saha., Luckyson, Khaidem., & Sudeepa, R.D. (2019), "Predicting the direction of stock market prices using tree-based classifiers", *North American Journal of Economics and Finance*, 47, 552-567.
- Shah, H.V. (2007) *Machine Learning Techniques for Stock Prediction*, Foundations of Machine Learning, Spring.
- Soni S. & Shrivastava, S., (2010), "Classification of Indian Stock Market Data Using Machine Learning Algorithms", *International Journal on Computer Science and Engineering*, 2(9), 2942-2946.
- Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. Springer, Heidelberg, Germany
- Xiao, Z., & David, E. (2017), "Forecasting daily stock market return using dimensionality reduction", *Expert Systems with Applications*, Vol.67, No. C, pp.126-139
- Yakup, K., Melek, A.B., Ömer, K.B (2011), "Predicting direction of stock price index movement using artificial neural networks and support vector machines", *Expert Systems with Applications: An International Journal*, 38(5), 5311-5319.