Analysing Stock Prices of Manufacturing Companies through Support Vector Regressor and Random Forest Algorithm

Chandan Kumar Bal

Doctoral Scholar, Department of Business Administration, Sambalpur University
Sambalpur, Odisha, PIN: 768019
Email Id: balchandan2@gmail.com

Dr. Rohita Kumar Mishra

Assistant Professor, Department of Business Administration, Sambalpur University
Sambalpur, Odisha, PIN: 768019
Email Id: rohitjrf@gmail.com

Abstract: *Stock market prediction is a challenging task. The purpose of the study is to analyze and predict stock prices of five companies using machine learning. These companies include three automotive manufacturers, one steel producer, and one aluminum manufacturer. Ratios have been used to analyze the stock for the given period and two machine learning algorithms named support vector machine and the random forest are used to predict the closing stock prices. The smoothed independent variables are feed to the models for prediction. The variables are smoothed using Hendrick-Prescott filter. The result of both the support vector regression and random forest are significant, performed well. Both Models has performed better than the hit ratio. Both models can be considered for strategy making but random forest Regressor surpassed another model in accuracy. A limited variable is taken for a short period of time. Data can be pre-processed more precisely to make the study more accurate by using feature engineering. The predicted results from the models along with financial performance analysis can be used for taking decisions on buying, selling and holding stocks, diversification of fund to different assets and maximize return with minimal risk. The paper tried to analyze the financial performance of five companies in India with the help of Support Vector Regressor-Hendrick-Prescott filter and Random Forest- Hendrick-Prescott filter to understand the nature of the stock data and efficient prediction of stocks.*

Key Words: *Support vector Regressor, Random Forest algorithm, stock price*

1. INTRODUCTION

Trade is a significant part of human history since Indus valley civilization (3000 BCE). The Modern stock trading is linked back to first half of seventeenth century with establishment of Amsterdam stock exchange in 1611 CE. India entered formal trading system in 1875 with the setup of Bombay Stock Exchange. Stock market is the key pillar in organized sector which is a significant part of every economy. Fictitious economy (stock market) reflects the real economy

(Zhou *et al*., 2010). Formation of stock trading or investment strategy involves a lot of complexity due to number of factors contributing stock price fluctuation. Stock price fluctuates due to the law of demand and supply of share. Share price will increase if demand for share goes up and share price will go down when there is less demand. Changes in share demand and supply is caused by lots of factors (Sindhu *et al*.,2014).

Stock price is influenced by lots of factors. Investors are considering factors for analysis as per the importance of the factors. Historical price and volume have been used from a very long period of time for stock behavior analysis. In last two-decade, sentiment analysis has been emerged as a prominent player in the field of stock analysis. Sentiment analysis has been undertaken from data collected from different social media platform, financial news article, companies' announcement and government notice(Cristescu *et al*., 2022). Some researchers are considering both sentiment analysis and historical data for making investment strategy(Charalampos M. Liapis *, 2023). Some researcher are considering macroeconomic variables factors like oil price, change in government policy, economic condition, inflation, exchange rate, interest rate(Sindhu *et al*., 2014), GDP growth (Hsing, 2011) politics in the region (joseph d. Piotroski, t. J. Wong, 2015) etc. for analyzing stock price.

Investors use the historical price data along with different financial ratio like liquid ratio, assets to turnover ratio, profitable ratio, growth ratio, assets structure ratio, debt ratio and solvency ratio for stock price analysis (Delen et al., 2013). Different technical analysis indicator based on historical data like moving average, exponential moving average, moving average convergence and divergence (MACD), relative strength index, support/resistance indicator and oscillator indicator etc. are used for analysis of stocks (Nti *et al*., 2020). These technical indicators are considered for deciding about where to invest or not and the timing of investment.

Prediction is a complex process due to non-linear nature of stock price movement. Stock market prediction involve with huge data in current world so manual prediction of stocks takes much more time. Technology is introduced as a medium to avoid that drawback in manual calculations. Technology based trading is very useful in stock market analysis. Technology based trading involves huge amount of data for analysis to get insights for profitable decision making (Rouf et al., 2021). Internet of things, machine learning, artificial neural network and deep learning are converting scratch data to meaningful one. Machine learning is being used in stock market for regression, classification, clustering(Shah & Isah, 2019). Regression used for prediction of future price behavior of stocks and indexes (Ataman & Kahraman, 2021). The classification technique used as tool to classify when the price of stocks will fall or rise (Karimuzzaman et al., 2021). Clustering used as tool to identify the community of similar stock trends (Chen et al., 2022). Moreover feature engineering is performed to find out the meaningful combination of features out of universe of feature to avoid the over-fitting of model (Yun *et al*., 2021). In this study we will consider two prediction algorithms of supervised machine learning namely support vector machine and random forest algorithm.

Manufacturing sector is a prominent sector in every economy. It provides employment to more than 30 million people in both organized and unorganized manufacturing sector in India. India has currently 5 million manufacturing establishment in rural India. India manufacturing companies are leading in sectors like pharmaceutical and textile. Manufacturing sector currently contributing 16-17 percent to Indian gross domestic product. India now going through industry 4.0. Indian manufacturing expecting to reach 1 trillion by 2025 (*Manufacturing Sector in India: Market Size, FDI, Govt Initiatives | IBEF*, 2022). The driver working behind the growth of the manufacturing sector are government incentives, domestic consumption, huge workforce pool, international investment and public private partnership.

Health security of citizen was the primary concern for every government. A number of safety protocol were imposed by world health organization, different government bodies which freezes some sectors in the economy. Manufacturing sector was victim of the pandemic. Governmental bodies eased restriction slowly due to active cases were coming down. Manufacturing sector started acceleration again after easing of covid norms. This paper aims at measuring performance of some manufacturing companies in last two years and make an effort for predicting the stock price of these companies for optimal portfolio making after easing of covid-19 restriction. This study is helpful for taking investment decision after covid 19. This study will be beneficial for investors for choosing fundamentally stable and greater predictable stocks. Here two machine learning model namely Random Forest Regressor and Support Vector Regressors have been used in the study. These models are chosen because of their higher performance in prediction as compared to traditional methods which is evident from the literature (Syukur & Istiawan, 2020).

2. PREVIOUS WORK

2.1. Covid-19 and Financial Performance

Covid-19 pandemic caused lockdown and shutdown globally. These lockdown and shutdown hamper industrial and production activity throughout the globe. (He *et al*., 2020) has analyzed the Industrial responses to covid-19 in chinese stock market using event study analysis approach and found mining, environmental, education industries are adversely affected. In the same period, the manufacturing as well research and consultancy industries are less affected. The financial performance is considered as the sole factor for the investors to invest in the stocks. The assessment of financial performance is a complex task due to the involvement of different interlinking factors. Delen et al, (2013), and Ginting, (2021) has used ratios like current ratio, quick ratio, debt to assets, debt to equity, activity ratio, inventory ratio, gross profit ratio, net profit margin, return on investment and return on equity.

2.2. Feature Selection

Features are significant part of every prediction model whether it is an application of econometric or machine learning. In order to improve prediction accuracy, it is essential to choose features having higher importance. Statistical features like variance, Sharpe ratio, compound annual growth rate and simple moving average has been used in prediction model (Christy Jackson *et al*., 2022). Fundamental and technical indicator have been taken into account for above purpose (Almeida and Neves, 2022:Gao *et al*., 2021:Houssein *et al*., 2022). Researchers also have employed macroeconomic variables to predict stock market (Ataman & Kahraman, 2021). (Goel & Singh, 2021) have put macroeconomic variables along with a global stock market factors for prediction of stock price. Researcher have used network variable. These network variables were calculated from social media as supplement to traditional variables (Keyan Liu et al.,2021). Investor sentiment from social media were also used in prediction of stock prices (li yelin *et al*.,2020).

2.3. Optimization of Hyperparameter

For avoiding model overfitting and improving model generalizability, researcher are using cross validation and forward validation in their training datasets. (X. Li et al., 2022) has used 7 fold cross validation in their research to improve the generalizability of model on unseen data and reduce overfitting in the model. (Sasi Kiran et al., 2024) has used XGBoost algorithm along with cross validation for predicting stock prices of the companies. (Candra et al., 2018) has employed C4.5 algorithm to predict the stock of good for business. It is found that stock prediction with cross validation improves the prediction accuracy by 60 percent. (M. Li et al., 2018) has found Support Vector Machine gives higher result with cross validation as comparison to without cross validation. (Jiao & Jakubowicz, 2017) has deployed cross validation, sequence validation and single validation along with machine learning models including random forest. Optimal parameters of regression models are required to decide to minimise the error and maximize the model accuracy. (Beniwal et al., 2023) has used Grid Search, Genetic Algorithm and Optimal Genetic Algorithm to decide the optimal parameters. (Bhandari et al., 2022) has used Adam, Adadelta and Nadam optimizer to set optimal parameter for models.

2.4. Prediction Models

Prediction models are significant aspect of prediction analysis. The model adoption depends upon nature of data whether it is linear or non-linear. (Christy Jackson *et al*., 2022) has employed a linear prediction model named as autoregressive integrated moving average (ARIMA). Researchers had proposed a hybrid model consisting of k-mean clustering and long-short term memory model for forecasting the price of stocks of chines commercial banks. K-mean clustering was adjusted with dynamic time wrapping for identifying stocks having similar kinds of trends and LSTM predicted stock price of similar trend companies. The study was able to capture satisfactory level of accuracy in both static and dynamic prediction(Chen *et al*., 2022). Another hybrid model was proposed consists of long-short term memory and random forest to predict the return from the stock market. Random forest was employed to decide upon the important parameters to avoid the overfitting of the model and LSTM was used for predicting the return. The hybrid model has been able to predict the return with greater accuracy than the baseline algorithm(Park *et al*., 2022). A multivariate intuitionistic fuzzy information system was introduced and sigma-pi neural network was used as inference tools. Non-membership value, membership value and lagged crisp values were taken as input for prediction purpose in Taiwan stock market. The proposed model was able to predict with R square more than 99 percent(Yolcu *et al*., 2022). A comparative study of different algorithm namely logistic regression, naïve bayes, linear discriminant analysis, backpropagation neural network, support vector machine, k-nearest neighbor, k-star, c.45, CART and random forest was undertaken to predict the Indonesian stock market index LQ45. Performance of the algorithms were measured on the parameter of accuracy, recall and precision. The Random forest algorithm topped the list in all criteria and support vector machine, LR, CART and C.45 performed well (Syukur & Istiawan, 2020). (Gao *et al*., 2021) have employed LSTM and GRU algorithm with refined feature for stock forecasting. Technical and financial data were feed to LASSO and PCA for feature reduction to improve the accuracy of the model. LSTM and GRU were equally competent in this aspect but LASSO outperform CPA. (Q. Liu et al., 2022) have put Deep learning neural network with technical indicator and stock fundamental for short term price prediction in chines stock prediction. Researchers have been making effort time to time to improve the accuracy of the prediction model. A hybrid model combined of Autoregressive integrated moving average and support vector regressor was proposed by researchers to remove the drawback of the ARIMA (linear regressor). The model was implanted for predicting daily and cumulative return of the stocks(Rubio & Alba, 2022). There are some kinds of issue with base line neural network like global minima and convergence issue. Metaheuristic based hybrid neural network was bought into picture to avoid the drawback of the ANN. The MHNN predicted the share price movement with 94.97 percent accuracy with optimal precision, F-measure and recall (Venkatanarayana & Satyanarayana, 2019). Another hybrid algorithm consists of support vector regressor with equilibrium optimizer was used with technical indicators for forecasting purpose. The proposed model surpassed other counterparts(Houssein *et al*., 2022). Researchers have made an effort to predict the prices of fifteen companies of different sectors by using real-time time series data sets. Holt-winter, recurrent neural network and recommendation system algorithm were used for the above purpose. The recommendation system performed well as comparison to other algorithms (Mohan M *et al*., 2023). (Ataman & Kahraman, 2021) have deployed linear regression-artificial neural was used for the stock prediction purpose. A very high score was found for overall BRICS nation with MAPE 71 percent. (Goel & Singh, 2021) worked upon predicting stock price by using artificial neural network with Scaled Conjugate Gradient Algorithm. The given model is able to achieve 93% accuracy in predicting the BSE Sensex closing prices. (Keyan Liu et al.,2021) LSTM, RNN and GRU applied to network variables for predictions. LSTM having slightly higher result than Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU). (li yelin *et al*.,2020) used Investor sentiment from social media were used in prediction of stock prices with the help of Long Short-Term Memory Model (LSTM), Logistic Regression, Support Vector Machine, and Naive Bayes Model and found LSTM has more predictive power. (Shynkevich Yauheniya *et al*., 2016) Another effort has been made to forecast the prices of health care stock by using different categories of news simultaneously having different degree of weightage with the help of Multiple Kernel Techniques. Performance increased as comparison to Support Vector Machine (SVM) and K-Nearest Neighbour (KNN) based on single news category. (Nguyen, Nguyet. 2018) (Nguyen, Nguyet. 2017)Forecasting based on Hidden Markov Model (HMM) has outperformed the historical average model. four criteria, including the Akaike Information, the Bayesian Information, the Hannan Quinn Information, and the Bozdogan Consistent Akaike information were used for deciding the number of states in HMM. Accuracy of HMM exceeds accuracy of Historical Average Model. (R. Liu & Vakharia, n.d.) has used LSTM for inventory prediction along with CNN for feature extraction. (Huang & Vakharia, n.d.) used reverse cross attention and bi-directional LSTM for stock prediction. (Shen & Shafiq, 2020) used recursive feature elimination for feature selection, PCA for feature reduction and LSTM for prediction.

3. METHODOLOGY

3.1. Data Collection

Data have been collected from yahoo finance on five manufacturing companies namely Tata Motors, Tata Steel, Maruti Suzuki, Mahindra & Mahindra and Hindalco. The entire dataset of the company is broken into two parts (a) Model training and model testing from 01/06/2020 to 9/3/2022 (b) data kept outside model for back testing from 10/3/2022 to 30/3/2022. The data were collected on stock close price, open price, high price, low price, adjusted close price and financial ratio.

Figure 1: Data splitting

The entire data set from 1st June 2020 to 30 march 2022

Model training and testing data ranges from 1st June 2020 to 10th March 2022. Training and testing ratio 80:20 respectively.	**Back Testing From 10/03/22 to 30/03/22**

The above figure showing model training, testing and back-testing split.

This study uses models to predict the out-sample from 10th March 2022 to 30th March 2022. The out-sample predicted will be compared with actual data kept outside for back testing to see whether our model works properly outside the training and testing data. This back testing will be performed on the stocks selected on the basis of highest R square.

3.2. Financial Ratios

Performance of companies were measured by using four ratio types namely liquidity, leverage, operating and profitability. Two ratios were selected from each head. Liquidity ratio talks about firms' ability to pay current liabilities. Firms' inability to pay off short term debt affects the credit rating and the credibility of firm. Repeated default of payment of current liabilities leads commercial bankruptcy and dissolution of firm. Leverage ratio measures the long-term stability and capital structure of the firm. It talks about the proportion of lender fund and owners fund in total capita mixture. Activity ratio tells the investors to how efficiently the operations and assets are being managed. Profitability ratio talks about profitability from operation of business and others.

Table 1: Ratios

Heads of ratio	Ratio	Formula
Liquidity ratio/short-term solvency	Current ratio	Current assets/current liabilities
	Quick ratio	Quick assets/current liabilities
Leverage ratio/long-term solvency	Debt to net assets	Total debt/ capital employed
	Debt to total assets	Total debt/total assets
Activity ratio	Total assets turnover ratio	Sales/total assets
	Capital turnover ratio	Sales/capital employed
Profitability ratio	Earnings per share	Profit for distribution/number of equity share
	Price earning share	Earnings per share/current price of share

3.3. Data Pre-processing

Hedrick - Prescott filter- Time series consist of trend, seasonality, cyclical and irregularities so data need to be filter before proceed to further analysis. There are different techniques for data cleaning and Hodrick – Prescott filter one of them. It is used for decomposing cyclical components from the time series data set. It is used to calculate the smooth curve representing the time series. It is more sensitive to long – run than short – run. The sensitive is adjusted to short – run fluctuation by modifying smoothing factor λ. Here λ value is 13,322,500(Hodrick, R. J., & Prescott, E. C., 1997).

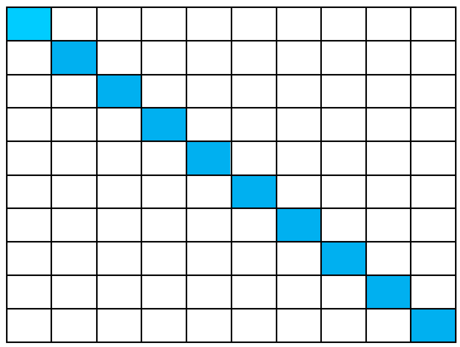
$$\min_{\tau}\left(\sum_{t=1}^{T}(y_t-\tau_t)^2+\lambda\sum_{t=2}^{T-1}[(\tau_{t+1}-\tau_t)-(\tau_t-\tau_{t-1})]^2\right)$$

These decomposing filters were applied to open price, high price, low price and adjusted close price to remove cyclical component from features. The trend data obtained after removing cyclical component are feed to machine learning model to predict the stock closing price.

3.4. Optimization of parameters

Generalizability of machine learning prediction is very much essential. So cross validation came into picture. Cross-validation is a critical instrument for evaluating the effectiveness of regression and classification techniques. Cross-validation is a widely used method for assessing the performance of a model on data that was not present during the training process. Cross-validation is the process of training and evaluating a model by dividing a dataset into multiple subsets or folds. One-fold is designated as a validation set for evaluation, while the remaining folds are employed for training. The performance metrics are averaged to estimate the performance of generalization on unseen and this procedure is repeated for each fold. The validation set is crucial for the optimization of hyper parameters and the prevention of over fitting of the model during the training process. The most prevalent method is K-fold cross-validation, which is a variation of cross-validation. Apart from cross validation, other optimal parameter must be decided to improve accuracy of the models. To optimize the parameters of Support Vector Regressor and Random Forest Regressor, GRID search is used in the study. The Grid search works by taking different combination of coefficient and intercept. The best combination with minimum errors is chosen for prediction. The grid search will decide about optimal parameters like n_estimators, max_depth, min_samples_split, min_samples_leaf of Random Forest regressors and kernel, C, gamma of Support Vector Regressor.

Figure 2: 10-fold cross validation table



In the given study, 10-fold cross validation are used to reduce the over fitting of the models. The 1- fold is kept aside for validation and 9-fold are trained. This process will be repeated until the last validation set is reached. The cross validation is carried upon training data sets only. The testing data set were used to measure the performance of the models.

3.5. Machine Learning models

Support Vector Regressor- Support vector machine is a powerful algorithm used for both prediction and classification. In this study, the computational power of support vector regressor used for prediction of stock price. Support vector regressor is a supervised machine learning used to predict the discrete value. It aims at obtaining the best fit line. The best fit line is called as the hyper plane having maximum number of points. SVR tries to draw the best fit line within the threshold values. The threshold value is the distance between boundary and hyper plane. Kernel function in SVR is used to transform the data from lower dimension to appropriate higher dimension to accurate prediction. The kernel functions are linear kernel, polynomial kernel and gaussian rbf kernel. Selecting kernel functions for SVR completely depends on the characteristics of the dataset. If the dataset is more linear, we can use the linear kernel, if the dataset is more non-linear, we can use polynomial, radial basis function, and sigmoid kernel.

Minimize

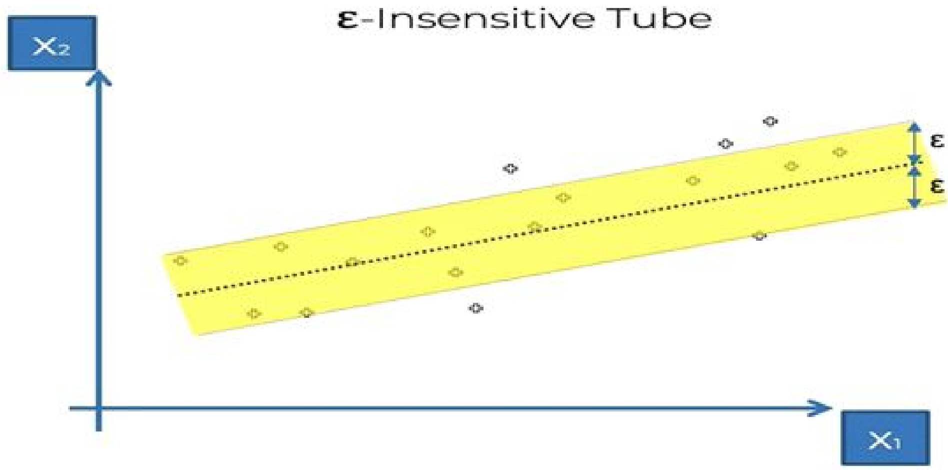
$\frac{1}{2}\omega^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$........... (I)

Constraints:

$y_i - wx_i - b \le \varepsilon + \xi_i$...........(ii)

$\omega x_i + b - y_i \le \varepsilon + \xi_i$...............(iii)

$\xi_i, \xi_i^* \ge 0$..........................(iv)

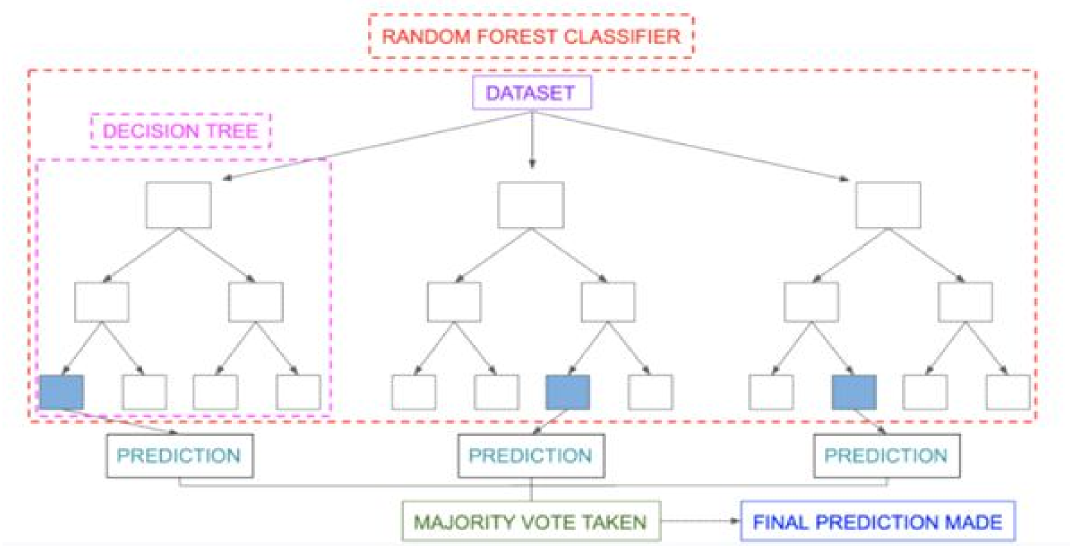


ε-Insensitive Tube on 2-D plot.

(Parashar, 2020) Support Vector Regression

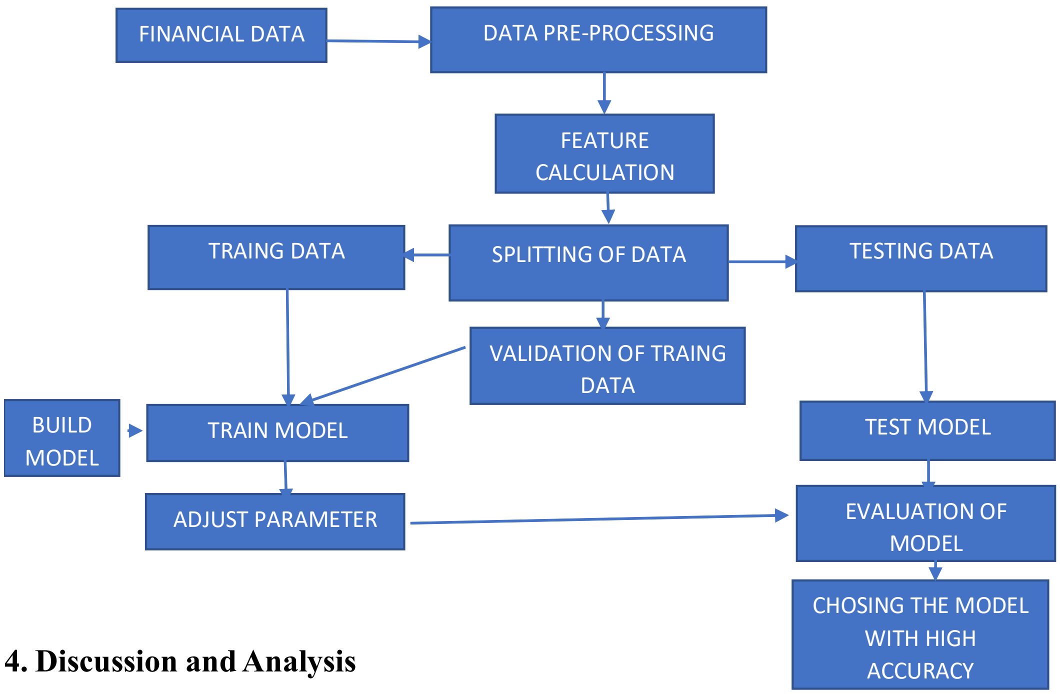
Random Forest Regressor - Random Forest Regressor is the combination of number of individual decision tree. Results from each decision tree are averaged to predict the values. Each decision tree is built using a random subset of the training data and a random subset of the features. This randomness helps to reduce overfitting and improve the generalization of the model. This averaging process helps to smooth out individual tree predictions and provides a more stable and reliable prediction.

Figure 3: Random Forest Regressor



(Random Forest Algorithm for Absolute Beginners in Data Science, 2021)

Figure 4: Flow chart of analysis



4. Discussion and Analysis

Ratios were calculated from the financial statement of different companies to display the relationship between different aspects of financial statement. The below table is showing all the ratios for measuring performance of the companies.

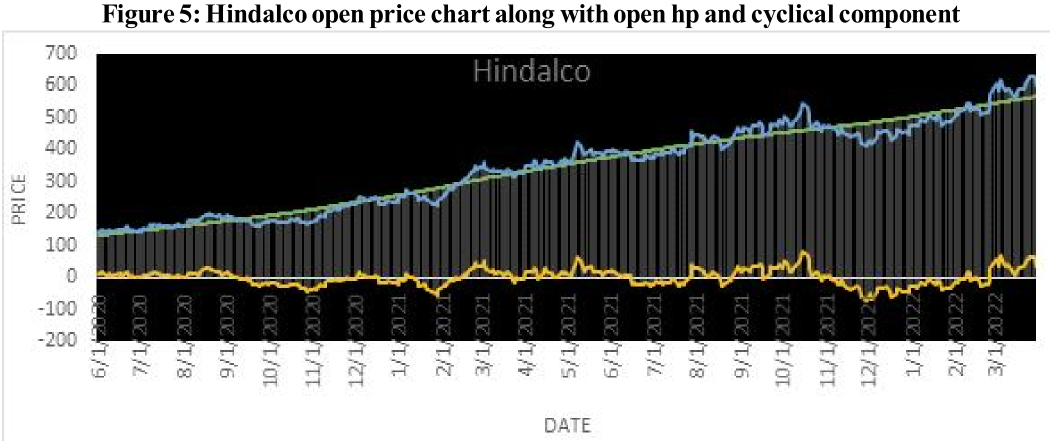
Table 2: Ratio of five companies for two years

Ratios	Year: 2021-2022					Year:2020-2021				
	T. S	T.M	M&M	M&S	HIND ALCO	T. S	T.M	M&M	M&S	HIND ALCO
CURRENT RATIO	0.97	0.58	1.38	0.99	1.37	0.58	0.6	1.34	1.15	1.48
QUICK RATIO	0.55	0.44	1.06	0.78	0.63	0.21	0.43	1.08	0.96	0.64
Debt to net assets	0.25	0.45	0.19	0.04	0.23	0.36	0.5	0.22	0.4	0.27
Debt to total assets	0.19	0.26	0.13	0.03	0.16	0.3	0.3	0.16	0.03	0.21
Total assets turnover ratio	0.58	0.73	0.85	1.21	0.68	0.46	0.46	0.75	1	0.48
Capital turnover ratio	0.76	1.28	1.18	1.56	0.95	0.55	0.77	1.01	1.3	0.62
Earnings per share	270.33	-3.63	41.28	124.68	24.76	145	-6.92	2.25	140	4.46
Price earning share	0.48	-7.75	3.83	63.85	242.2	0.55	-3.18	88	6771.6	72.9

Tata steel (TS) has performed good but it should work on its liquidity to avoid the short-term happening and it will also attract more investors. The company has given highest profit as comparison to other companies in both of the year. Long term investment in tata steel may risk the capital as it liquidity issue. Tata motors has been going through loses. Its operating activity quite good but it should try to reduce the operating expenses. Mahindra and Mahindra is performing overall good in all aspect. Investors can opt for investment in Mahindra and Mahindra because it gives a stable fundamental for investment. Maruti and Suzuki's capital mixture and operating activity doing well but it should look at its current paying ability and also at reducing operating expenses. Hindalco has performed the below average with very low profit margin.

Hendrick-Prescott filter is used to decompose the cyclical element from time series data. The figures below showing the trends after decomposing the cyclical element of different companies. All the companies are showing positive trends but with different degree. The HINDALCO, TATA MOTORS, and TATA STEEL are showing high positive trends as comparison to others. Here the open price charts are given below after applying the Hendrick-Prescott filter. The filter also applied for other variable in study.

Figure 5: Hindalco open price chart along with open hp and cyclical component

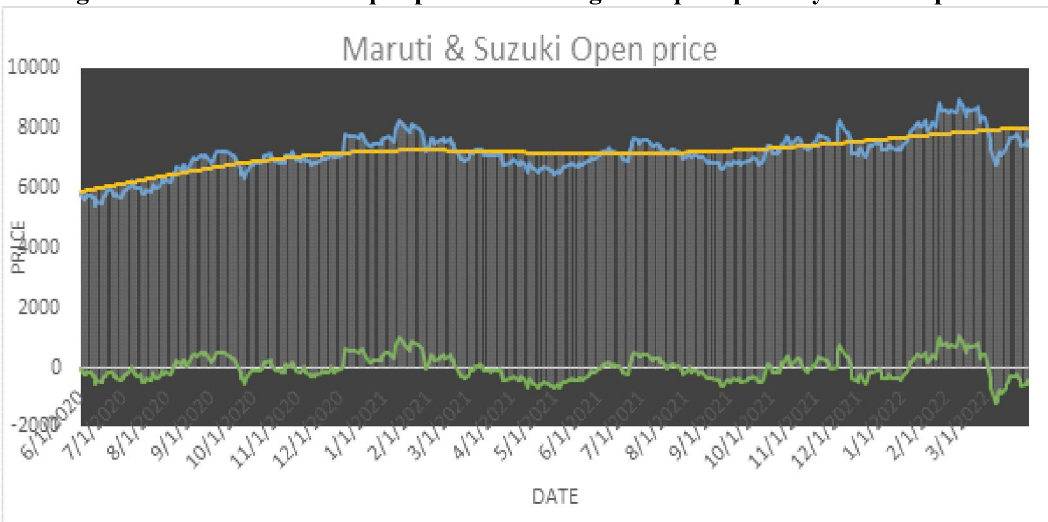


The above figure shows the open price chart of Hindalco company. The cyclical components were removed by using Hendrick-Prescott filter. The yellow line shows the cyclical component of time series data. The green line representing the open price trend

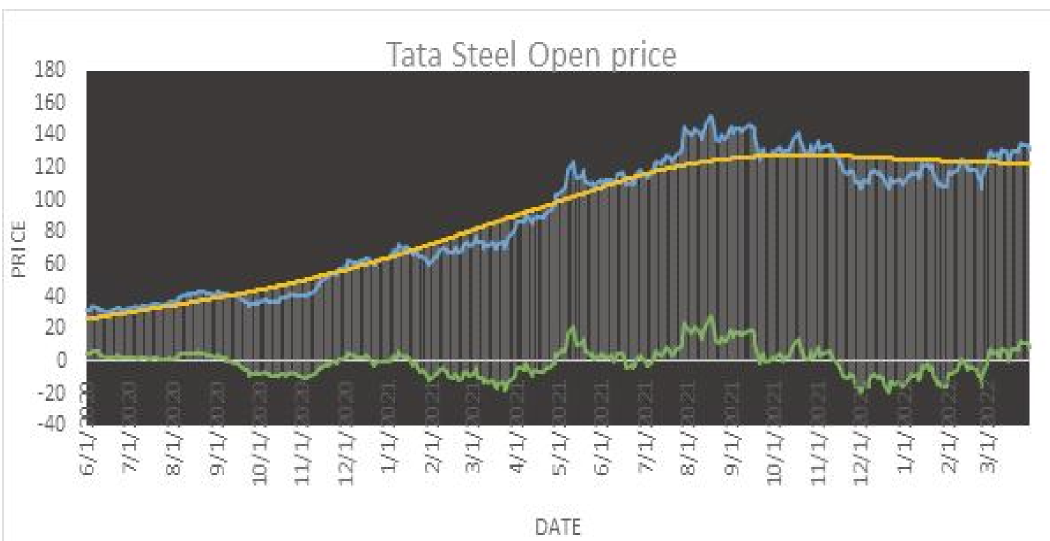
Figure 6: Mahindra & Mahindra open price chart along with open hp and cyclical component

The above figure shows the open price chart of M&M companies. The data are of time series in nature. The Hendrick-Prescott filter has been used to find the data trend after removing the cyclical components. The yellow line representing the cyclical components in time series data. The blue line presenting the open price trend.

Figure 7: Maruti & Suzuki open price chart along with open hp and cyclical component

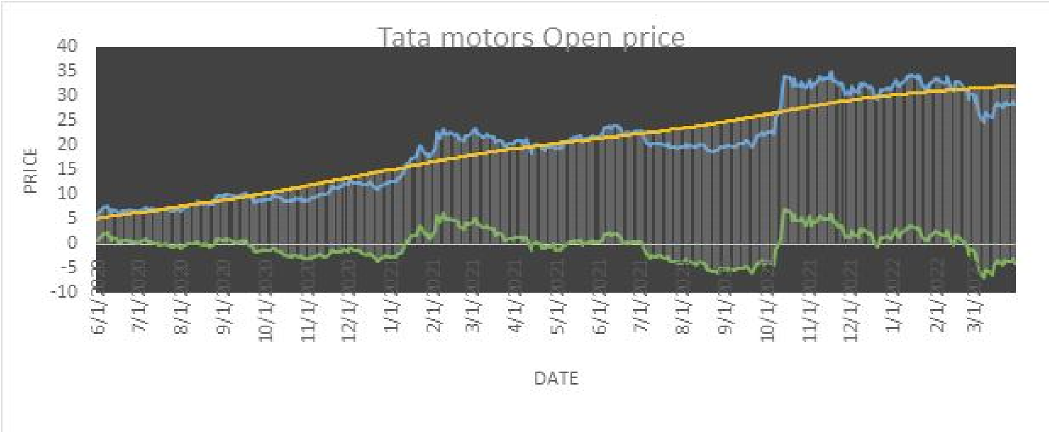


The above figure shows the open price chart of Maruti & Suzuki. The figure shows open price (blue line), open price after removing cyclical component (yellow line) and the green line representing the cyclical components. The cyclical component was removed by applying the Hendrick-Prescott filter.



The above figure shows open price chart of Tata Steel. The yellow line showing the open price trend of the company. The cyclical green line representing the cyclical components. The cyclical components were separated by using Hendrick-Prescott filter.

Figure 9: Tata motors open price chart along with open hp and cyclical component



The above figure talks about the open price chart of Tata motors and the yellow line representing the open price trend. The green line is the cyclical component line.

Figure 10:Tata steel SVR

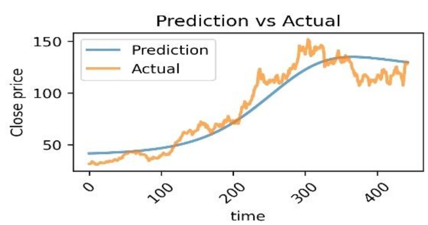


Figure 11: Tata steel RF

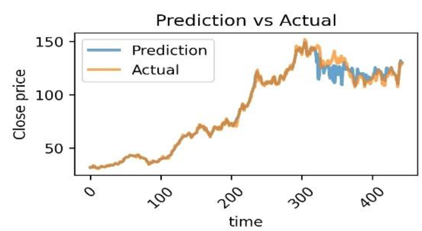


Figure 12: Mahindra &Mahindra SVR

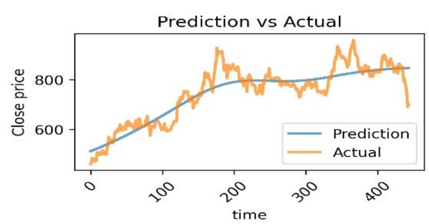


Figure 13: Mahindra & Mahindra RF

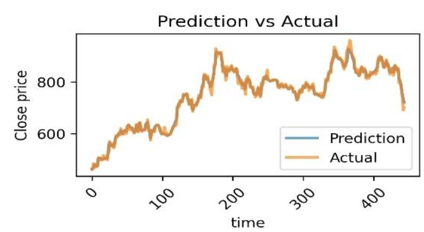


Figure 14: Hindalco SVR

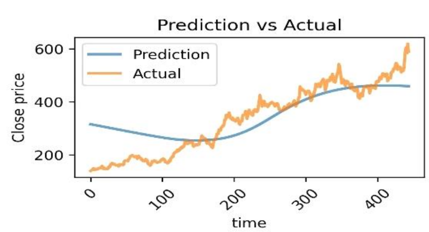


Figure 15: Hindalco RF

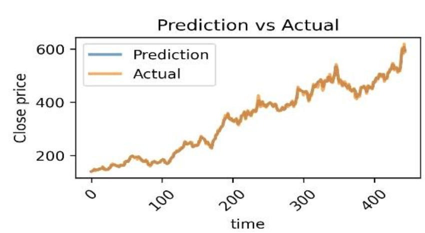


Figure 16: Maruti SVR

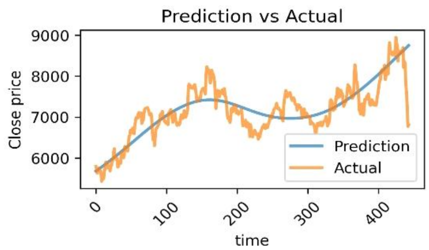


Figure 17: Maruti RF

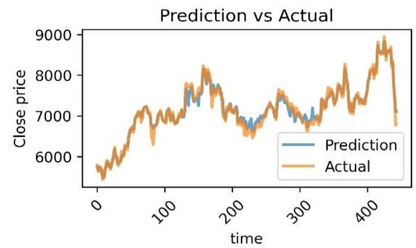
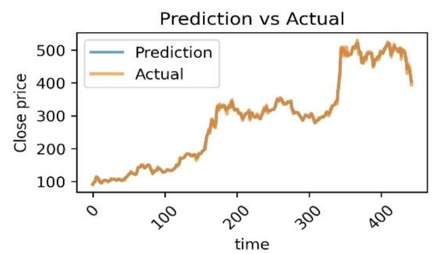


Figure 18: Tata motor SVR

Figure 19: Tata Motors RF



The above figures are showing model fitting on different companies' data. The models are trained with 80 percent data point and rest 20 percent were used in testing of the models. The GRID search has optimized different parameter of both Support vector regressor and Random Forest regressor on each run. The average R square of the Support Vector Regressor (SVR) is around 77 percent while the average R Square of Random Forest (RF) is 94 percent on test data. The RF model is able to achieve approx. 96, 95, 96 percent R square in Tata motors, Mahindra & Mahindra and Hindalco respectively. The average accuracy of the random forest regressor throughout different companies is higher than support vector regressor. Investor should look for random forest regressor for predicting prices of stocks. The R square in both models differs due the nature of model to approach the data.

Table 3: R square results

R square	RANDOM FOREST REGRESSOR				
	T. STEEL	T. MOTORS	M &M	M. & SUZUKI	HINDALCO
	91	96	95	92	96
	SUPPORT VECTOR REGRSSOR				
	T. STEEL	T. MOTORS	M &M	M. & SUZUKI	HINDALCO
	87	76	74	69	79

Back testing

This study found highest R square in three companies applying RF. So the study consider only three companies namely Mahindra & Mahindra, Hindalco and Tata motors for portfolio. The out-sample predicted data were compared with the actual data to see whether our model works in real trading or not. The Back testing is done on last 14 days data from the comparison, it is clear that our model is working properly on real market data. The difference between actual and out-sample prediction in case of Mahindra & Mahindra and Hindalco is very limited.

Figure 19: Hindalco Out-sample prediction vs Out-sample actual

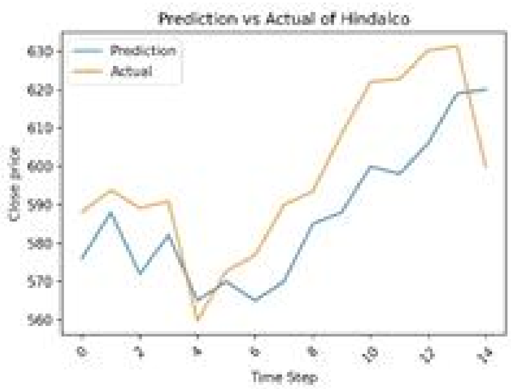


Figure 20: Tata motor Out-sample : Prediction vs Out-sample actual

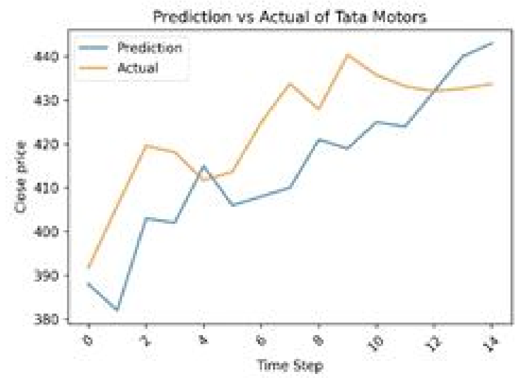
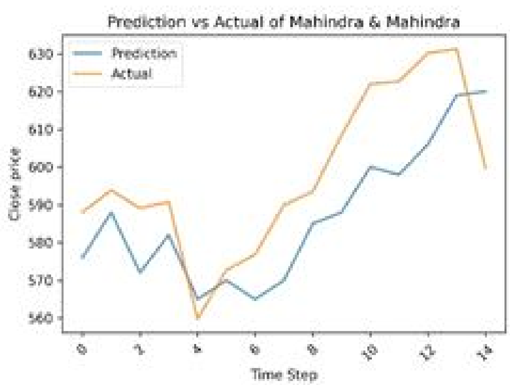


Figure 21: M & M Out-sample Prediction vs Out-sample actual



The above three figure showing out-sample prediction and actual stock price of the three companies. All three are based on Random Forest Regressor.

5. Conclusion and Future Directions

Here Ratio analysis has been used for analysing stocks of different companies. Different companies performed better in different parameters. Both of our machine learning model performed better than the Hit ratio. The random forest performed well as compared to the support vector regressor with more 94 percent R square. Investors most of the time aims to minimize risk and maximize the profit. It is essential for investors to check out both fundamental and prediction result from the models. This study found that Mahindra and Mahindra doing fundamentally well as comparison to Tata Motors, Tata Steel, Hindalco, and Maruti & Suzuki. It is also found that the prediction accuracy of Random Forest Regressor result is quite favourable for the three companies namely Tata Motors, Mahindra and Mahindra and Hindalco. The result also shows that the Mahindra and Mahindra is good destination of investment for short period of time. This study has some limitations also as it can consider other variable of both qualitative and quantitative nature. This study has taken data for two years only but more data can be considered for the further study. The paper has removed cyclical components but there are other noises which can reduce model accuracy like seasonality and irregularities. Study can be more accurate if data will be prepared more technically. Different technique of feature engineering can be applied to the study to get more meaningful insights. the study has considered only few ratios for the study. Other ratio can be used for further study.

References

Almeida, R. L. de, & Neves, R. F. (2022). Stock market prediction and portfolio composition using a hybrid approach combined with self-adaptive evolutionary algorithm. *Expert Systems with Applications*, *204*(April), 117478. https://doi.org/10.1016/j.eswa.2022.117478

Ataman, G., & kahraman, S. (2021). Stock market prediction in brics countries using linear regression and artificial neural network hybrid models. *The Singapore Economic Review*, *67*(02), 635–653. https://doi.org/10.1142/S0217590821500521

Beniwal, M., Singh, A., & Kumar, N. (2023). Forecasting long-term stock prices of global indices/ : A forward-validating Genetic Algorithm optimization approach for Support Vector Regression. *Applied Soft Computing*, *145*, 110566. https://doi.org/10.1016/j.asoc.2023.110566

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. C. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, *9*(May), 100320. https://doi.org/10.1016/j.mlwa.2022.100320

Candra, B. P., Saputra, E. R. S. H., Wicaksono, K., & Kusrini, K. (2018). Decision support system for stock prediction and supplier selection using least square and C4.5 algorithm. *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018*, 241–246. https://doi.org/10.1109/ICITISEE.2018.8721001

Charalampos M. Liapis *, A. K. * S. K. (2023). Investigating Deep Stock Market Forecasting with. *Entropy*. https://doi.org/https://doi.org/10.3390/25020219

Chen, Y., Wu, J., & Wu, Z. (2022). China's commercial bank stock price prediction using a novel K-means-LSTM hybrid approach. *Expert Systems with Applications*, *202*(April). https://doi.org/10.1016/j.eswa.2022.117370

Christy Jackson, J., Prassanna, J., Abdul Quadir, M., & Sivakumar, V. (2022). Stock market analysis and prediction using time series analysis. *Materials Today: Proceedings*, *xxxx*. https://doi.org/10.1016/j.matpr.2020.11.364

Cristescu, M. P., Nerisanu, R. A., Mara, D. A., & Oprea, S. (2022). Using Market News Sentiment Analysis for Stock Market Prediction. *Mathematics*. https://doi.org/https://doi.org/10.3390/math10224255

Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, *40*(10), 3970–3983. https://doi.org/10.1016/j.eswa.2013.01.012

Gao, Y., Wang, R., & Zhou, E. (2021). Stock Prediction Based on Optimized LSTM and GRU Models. *Scientific Programming*, *2021*. https://doi.org/10.1155/2021/4055281

He, P., Sun, Y., Zhang, Y., & Li, T. (2020). COVID–19's Impact on Stock Prices Across Different Sectors—An Event Study Based on the Chinese Stock Market. *Emerging Markets Finance and Trade*, *56*(10), 2198–2212. https://doi.org/10.1080/1540496X.2020.1785865

Houssein, E. H., Dirar, M., Abualigah, L., & Mohamed, W. M. (2022). An efficient equilibrium optimizer with support vector regression for stock market prediction. In *Neural Computing and Applications* (Vol. 34, Issue 4). Springer London. https://doi.org/10.1007/s00521-021-06580-9

Hsing, Y. (2011). Makroekonomske varijable i tržište dionica: Slučaj hrvatske. *Ekonomska Istrazivanja*, *24*(4), 41–50. https://doi.org/10.1080/1331677X.2011.11517479

Huang, Y., & Vakharia, V. (n.d.). Deep Learning-Based Stock Market Prediction and Investment Model for Financial Management. *Journal of Organizational and End User Computing*, *36*(1), 1–22. https://doi.org/10.4018/JOEUC.340383

Jiao, Y., & Jakubowicz, J. (2017). Predicting stock movement direction with machine learning: An extensive study on S&P 500 stocks. In N. J.-Y., O. Z., S. T., G. R., N. R., W. C., Z. H., B.-Y. R., B.-Y. R., H. X., K. J., C. A., T. J., & T. M. (Eds.), *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017* (Vols. 2018-January, pp. 4705–4713). Institute of

Electrical and Electronics Engineers Inc. https://doi.org/10.1109/BigData.2017.8258518

Joseph PIOTROSKI., T. J. WONG., T. Z. H. (2015). Political Incentives to Suppress Negative Information/ : Evidence. *Journal of Accounting Research*, *53*(2). https://doi.org/10.1111/1475-679X.12071

Karimuzzaman, M., Islam, N., Afroz, S., & Hossain, M. M. (2021). Predicting Stock Market Price of Bangladesh: A Comparative Study of Linear Classification Models. *Annals of Data Science*, *8*(1), 21–38. https://doi.org/10.1007/s40745-020-00318-5

Li, M., Puthal, D., Yang, C., Luo, Y., Zhang, J., & Li, J. (2018). Stock market analysis using social networks. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3167918.3167967

Li, X., Yu, Q., Tang, C., Lu, Z., & Yang, Y. (2022). Application of Feature Selection Based on Multilayer GA in Stock Prediction. *Symmetry*, *14*(1415). https://doi.org/https://doi.org/10.3390/sym14071415

Liu, Q., Tao, Z., Tse, Y., & Wang, C. (2022). Stock market prediction with deep learning: The case of China. *Finance Research Letters*, *46*(June), 102209. https://doi.org/10.1016/j.frl.2021.102209

Liu, R., & Vakharia, V. (n.d.). Optimizing Supply Chain Management Through BO-CNN-LSTM for Demand Forecasting and Inventory Management. *Journal of Organizational and End User Computing*, *36*(1), 1–25. https://doi.org/10.4018/JOEUC.335591

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. In *Artificial Intelligence Review* (Vol. 53, Issue 4). Springer Netherlands. https://doi.org/10.1007/s10462-019-09754-z

Park, H. J., Kim, Y., & Kim, H. Y. (2022). Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework. *Applied Soft Computing*, *114*, 108106. https://doi.org/10.1016/j.asoc.2021.108106

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. (2021). Stock Market Prediction Using Machine Learning Techniques/ : A Decade Survey on Methodologies , Recent Developments , and Future Directions. *Electronics*.

Rubio, L., & Alba, K. (2022). Forecasting Selected Colombian Shares Using a Hybrid ARIMA-SVR Model. *Mathematics*, *10*(13). https://doi.org/10.3390/math10132181

Sasi Kiran, J., Dhana Lakshmi, P., Sultana, N., Naga Rama Devi, G., Gothane, S., & Reddy Madhavi, K. (2024). Stock Market Price Prediction Using Sentiment Analysis. In V. Bhateja, P. S. R. Chowdary, W. Flores-Fuentes, S. Urooj, & R. Sankar Dhar (Eds.), *Evolution in Signal Processing and Telecommunication Networks* (pp. 261–268). Springer Nature Singapore.

Shah, D., & Isah, H. (2019). *Stock Market Analysis/ : A Review and Taxonomy of Prediction Techniques*. *ii*.

Shen, J., & Shafiq, M. O. (2020). Short term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*. https://doi.org/10.1186/s40537-020-00333-6

Sindhu, M. I., Bukhari, S. M. H., & Hussain, A. (2014). Macroeconomic Factors do influencing Stock Price/ : A Case Study on Karachi Stock Exchange. *Journal of Economics and Sustainble Development*, *5*(7), 114–125.

Syukur, A., & Istiawan, D. (2020). Prediction of LQ45 Index in Indonesia Stock Exchange: A Comparative Study of Machine Learning Techniques. *International Journal of Intelligent Engineering and Systems*, *14*(1), 453–463. https://doi.org/10.22266/IJIES2021.0228.42

Venkatanarayana, K., & Satyanarayana, B. (2019). Performance analysis on metaheuristic based hybrid neural network to predict the stock movement. *International Journal of Scientific and Technology Research*, *8*(8), 1686–1692.

Yolcu, O. C., Egrioglu, E., Bas, E., & Yolcu, U. (2022). Multivariate intuitionistic fuzzy inference system for stock market prediction: The cases of Istanbul and Taiwan. *Applied Soft Computing*, *116*, 108363. https://doi.org/10.1016/j.asoc.2021.108363

Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Systems with Applications*, *186*(March), 115716. https://doi.org/10.1016/j.eswa.2021.115716

Zhou, Y. Y., Liu, C. Z., & Tang, L. (2010). Empirical analysis of influences of fictitious economy on real economy in China. *2010 International Conference on Logistics Systems and Intelligent Management, ICLSIM 2010*, *1*(1), 305–309. https://doi.org/10.1109/ICLSIM.2010.5461414

Nguyen, Nguyet. (2018), Hidden Markov model for stock trading. International journal of financial studies, Vol. 6 No. 36, Doi:10.3390/ijfs6020036

Nguyen, Nguyet., (2017), An analysis and implementation of the hidden markov model to technology stock prediction, Risks, Vol. 5 No. 62. Doi:10.3390/risks5040062

Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: an empirical investigation. Journal of Money, credit, and Banking, 1-16.

Parashar, Priyanka. “Support Vector Regression and It’s Mathematical Implementation.” *The Startup*, 18 June 2020, medium.com/swlh/support-vector-regression-and-its-mathematical-implementation-4800456e4878. Accessed 11 Aug. 2023.

Random Forest Algorithm for Absolute Beginners in Data Science.” *Analytics Vidhya*, 19 Oct. 2021, www.analyticsvidhya.com/blog/2021/10/an-introduction-to-random-forest-algorithm-for-beginners/. Accessed 11 Aug. 2023.

Manufacturing Sector in India: Market Size, FDI, Govt Initiatives | IBEF. (2022). Www.ibef.org. https://www.ibef.org/industry/manufacturing-sector-india.