Determination of Factors that Influence Indian Stock Market using Machine Learning Algorithms

First Author: Shubha R

Co Authors: Naveen C, Manoj N S

Abstract:

This study presents a thorough investigation into the variables affecting price changes on the Indian Stock Market. The study aims to pinpoint the key elements that significantly influence the closing price of the stock market by utilizing three different datasets and three different machine learning algorithms. The results show a notable result: the efficiency of the algorithm did not significantly increase as a result of including economic variables in the analysis. Instead, adding these variables reduced the algorithm's capacity for prediction. As a result, the study contends that technical factors, including historical price patterns, trading volumes, and technical indicators, have a greater impact on price prediction in the Indian Stock Market. For investors and other market participants looking to predict future market conditions, the implications of this study are extremely important. Investors can make well-informed investment decisions that could result in higher returns while lowering risks by realizing the predictive value of technical variables. This study emphasizes how crucial it is to combine machine learning algorithms with sophisticated data analysis methods in order to uncover patterns and insights from market data. Investors can improve their ability to predict market trends and make better investment decisions by incorporating technical variables into predictive models. The study adds to the knowledge by highlighting the necessity of focusing on technical aspects rather than just economic data when forecasting prices on the Indian Stock Market. It implies that the market's supply and demand dynamics, investor attitude, and trading patterns are better reflected when technical factors are taken into consideration. This study highlights the potential advantages of making investment decisions, portfolio allocations, and trading choices proactively based on expected market fluctuations. Overall, the study provides insightful information that might improve the performance and accuracy of forecasting models in relation to the Indian Stock Market.

Keywords: Nifty 50, Technical variables, Economic variables, Machine Learning, Prediction, Investment Decisions

Introduction:

India is the second-largest nation by population and ranks seventh in terms of total area. According to the reports by mint genie, it is said that approximately 3% of the Indian population has invested in the Indian Stock Market which makes it one of the huge markets where trading takes place. A stock market is a place where people with savings will come down to buy shares of publicly held companies or sell their shareholdings in order to liquidate. In simpler words, a stock market is a place where buyers and sellers of securities will connect, communicate, and conduct business on the stock market. A certificate granting the possessor the right to become one of the company's owners is known as a share. Dividends are the most typical form of return on investment for these stockholders. If the existing shareholder wants to liquidate the shares, they can be sold back to another buyer on the stock market. Buyers and sellers can be confidence that they will receive a fair price, a high level of liquidity, and transparency since market participants compete for business. Financial instruments including derivatives, mutual funds, stocks, and bonds are bought and sold by traders. Trading is available to both individual investors and professionals working for businesses or financial organizations. (PS, 2013). In order to prevent and minimize stock market fraud, SEBI is an intermediary that regulates the Indian stock market.

All stock market investment has benefits and drawbacks of their own. Trading in a stock market will lead to higher returns for the investor, business expansion, and improved economic conditions in a country. In the stock business, stock market forecasting is crucial. (Naeini MP, 2010). If not, it can result in unintended losses, hence taking preventative measures to guard against potential dangers is highly desirable. One of the most crucial factors to take into account when investing is keeping an eye on recent market trends or following the Indian stock market.

According to a 2018 World Bank research (The World Bank, n.d.), the total market value of stocks on the planet has surpassed 68.654 trillion US dollars. Technical developments are partially responsible for the current rise in popularity of stock trading. Investors seek strategies and tools that could boost profits while reducing risk. The task of stock market prediction is not very straightforward, nevertheless, because it is stochastic, nonlinear, dynamic, and inaccurate. Analysts from a variety of disciplines, including economics, mathematics, material science, and computer science, have voiced concern about the financial market prediction. (Nusrat Rouf, 2021). Gaining profits by trading the stocks is an important reason behind the stock market prediction (Ali Khan, 2016).

Literature Survey:

Over the course of the last few years, there have been many highs and lows for the Indian stock market. Dealing with such volatility is now a major concern for economists, traders, and policymakers. The country's economic stability is impacted by stock market volatility (Ms N VivekaPriya, 2022). The Indian Stock Market's two stock exchanges see the lion's share of activity. These are the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). The same trading days, hours, and settlement each trading platform makes use of certain procedures. Both of these exchanges list all of the major companies. Compared to the National Stock Exchange, which has roughly 2000 listed companies, the Bombay Stock Exchange has about 5000 listed companies. Even though there are fewer companies listed on the NSE than the BSE, there is a significantly higher volume of share trading and a higher turnover value on the NSE. In this study, the NSE index Nifty 50 will be taken into account, and the influence of stock-specific traits and macroeconomic variables will be investigated using machine learning techniques.

Investors most typically utilize fundamental research, technical analysis, and quantitative analysis to assess the advantages and hazards of long-term stock market investing. Investors with a longer time horizon seek out investments that have a better potential of increasing their profits. The fundamental analysis defines the price at which one share of a firm should trade by determining the true value of the industry or sector of the company. If given enough time, it is assumed that the business will switch to a cost agreement. (2021; Nusrat Rouf). Fundamental analysts most frequently use the Price to Earnings ratio (P/E) and also the Price to Book ratio (P/B) to predict annual long-term price changes. The P/E ratio is one indicator.

Technical analysis is the examination of stock prices with the intention of boosting returns or enhancing investing decisions. (Zhu & Zhou, 2016) Technical analysis predicts how stock prices will move in the future based on historical data. Additionally, employing technical indicators, it facilitates the examination of financial time series data. (Nusrat Rouf, 2021).

Machine learning algorithms have become more widely used in recent years to examine a variety of industrial problems and predict future changes in the variable that is said to have been impacted by a number of other variables. In order to recognize the trends that the data has, machine learning is used in stock price prediction. (Mutalib, Rahman, Abdul-Rahman, & Rahman, 2017). It is very much possible to quickly analyze a complicated data that is heterogeneous and get more reliable results using machine learning algorithms. Stock market forecasts have been made using machine learning techniques.

A number of studies have been carried out to predict the Stock Market Index Nifty 50 using traditional stock market variables. This paper focuses on analyzing the variables hypothesized to have an impact on the Stock market being studied across multiple Machine Learning Algorithms. Considering the hybrid factors will train the model well to cover the times of exceptional situations like Covid. To the best of my knowledge, there are a number of models that are built using stock-specific factors like High, Low, Open, and Close. But this paper considers stock-specific and hybrid factors in training the model for predicting the Nifty 50 closing prices.

Objective:

The objectives of the paper are:

1. To study the factors influencing NSE index
2. To predict the market performance of Nifty 50 index using different machine learning algorithms
3. To evaluate the efficiency of machine learning algorithm in predicting monthly closing prices of Nifty 50 index using R-square and RMSE.

Research Methodology:

The method used for the collection of data is secondary. The data is collected at the monthly level over the years. The research uses factors that are quantitative in nature. Using test data, the model's effectiveness is evaluated after it has been trained using train data. The proportion of data split used for the study is in the ratio of 80:20. The supervised learning model's main tenet is to find patterns and relationships in the data of the training set and then duplicate them in the test set. The statistical tool used in the prediction is Linear Regression, Logistic Regression, and Neural Networks.

In order to model the relationship between two variables, the statistical method of linear regression includes fitting a linear equation to the observed data. The goal is to find the line of best fit that fits the given dependent variable properly illustrates the relationship between the variables.

Logistic regression machine learning algorithms are used most commonly in the supervised learning category. It is used to forecast the categorical dependent variable using a collection of predetermined set of categorical and continuous independent variables. The study uses logistic regression and converts the closing price into a categorical variable by coding the close price increase as 1 and the close price decline as 0.

Neural networks, commonly referred to as Artificial Neural Networks (ANNs) or Simulated Neural Networks (SNNs), are a type of machine learning that form the core of deep learning methodologies. They mimic the organisation and terminology of the human brain to mimic how real neurons communicate.

Each node layer in an artificial neural network consists of a node, an input layer, one or more hidden layers, and an output layer. Each node, is connected to the others and has a weight and a threshold that is calculated using gradient descent. Any node whose output rises above the specified threshold value turns on and starts sending the information to the top layer of the network. If not, no information is transmitted to the next layer of the network. The data from the study must first be normalised before being fed into the model to create the neural network model. There are 5 hidden layers used to stop the model from overfitting the data. The error.fct used is "CE" - cross entropy.When the dependent variable is a categorical variable, it is used. Log loss is the term most commonly used to describe the error function.

The analysis is carried by considering three different set of independent variables used in predicting Nifty 50 Index, they are:

- **Model I:** Stock Market Specific - Open, High, Low, Monthly Traded Volume, Volume
- **Model II:** Stock Market Specific + Monthly Closing Prices of three popular Exchange Rates - EURVsINR, GBPVsINR, USDVsINR
- **Model III:** Stock Market Specific + Exchange Rates + Macro Economic Variables - Trade Balance, Forex, FII, Bank rate, monthly closing prices of Gold and Silver.

DATA ANALYSIS:

Linear Regression:

Stock Market Specific Model:

```
Coefficients:
                     Estimate Std. Error    t value Pr(>|t|)
(Intercept)        -5.329e-12  4.830e-13 -1.103e+01  < 2e-16 ***
Month               4.295e-21  3.775e-22  1.138e+01  < 2e-16 ***
Open               -5.023e-17  2.886e-17 -1.741e+00  0.08612 .
High                1.172e-16  3.373e-17  3.476e+00  0.00088 ***
Low                 1.011e-16  3.028e-17  3.340e+00  0.00135 **
MonthlyTradedVolume  1.000e+00  2.838e-17  3.523e+16  < 2e-16 ***
Volume             -3.721e-22  3.196e-21 -1.160e-01  0.90766
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.783e-14 on 70 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 4.13e+34 on 6 and 70 DF,  p-value: < 2.2e-16
```

Stock Market Specific + Monthly Closing Prices of three popular Exchange Rates:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -6.393e+03  6.651e+03  -0.961 0.357105    
Month        3.618e-06  6.027e-06   0.600 0.560407    
Open        -7.569e-01  1.446e-01  -5.235 0.000279 ***
High         8.839e-01  1.873e-01   4.719 0.000630 ***
Low          7.559e-01  1.563e-01   4.835 0.000523 ***
Volume      -1.049e-05  4.211e-05  -0.249 0.807865    
EURVSINR     6.935e+00  2.601e+01   0.267 0.794689    
GBPVSINR     1.211e+01  1.017e+01   1.190 0.259080    
USDVISNR     8.485e+00  2.604e+01   0.326 0.750644    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 153.3 on 11 degrees of freedom
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9979 
F-statistic:  1115 on 8 and 11 DF,  p-value: 6.84e-15
```

Stock Market Specific + Exchange Rates + Macro Economic:

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -7.948e+03  6.285e+03  -1.265 0.210764    
Month         5.906e-06  4.640e-06   1.273 0.207884    
Open         -5.052e-01  1.403e-01  -3.601 0.000632 ***
High          8.298e-01  1.249e-01   6.644 8.84e-09 ***
Low           5.625e-01  1.405e-01   4.002 0.000170 ***
TradedVolume -3.517e-05  1.854e-05  -1.897 0.062460 .  
EURVSINR     -1.301e+01  1.399e+01  -0.930 0.355942    
GBPVRINR      1.403e+01  1.148e+01   1.222 0.226192    
USDVSINR      1.151e+01  2.388e+01   0.482 0.631612    
TradeBalance -1.479e-03  1.761e-03  -0.840 0.404251    
Forex        -2.490e-04  4.263e-04  -0.584 0.561252    
FII           1.701e-03  3.231e-03   0.526 0.600451    
Bankrate     -4.259e+01  1.367e+02  -0.312 0.756346    
Gold          3.341e-03  2.577e-02   0.130 0.897272    
Silver        5.688e-04  1.123e-02   0.051 0.959756    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 249.4 on 62 degrees of freedom
Multiple R-squared:  0.9953,    Adjusted R-squared:  0.9943 
F-statistic: 947.5 on 14 and 62 DF,  p-value: < 2.2e-16
```

Logistic Regression:

Stock Market Specific Model:

```
Call:
glm(formula = return10 ~ . - srno - Close, family = "binomial",
    data = train)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.95168  -0.07281   0.01214   0.24534   1.65956

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          1.762e+01  4.047e+01   0.435  0.66332
Month               -1.570e-08  3.165e-08  -0.496  0.61983
Open                -1.417e-02  4.600e-03  -3.079  0.00207 **
High                -5.990e-03  4.236e-03  -1.414  0.15728
Low                  6.928e-03  3.772e-03   1.837  0.06626 .
MonthlyTradedVolume  1.407e-02  4.683e-03   3.005  0.00266 **
Volume               1.475e-07  2.930e-07   0.503  0.61470
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 105.168  on 76  degrees of freedom
Residual deviance:  29.714  on 70  degrees of freedom
AIC: 43.714

Number of Fisher Scoring iterations: 8
```

Stock Market Specific + Monthly Closing Prices of three popular Exchange Rates:

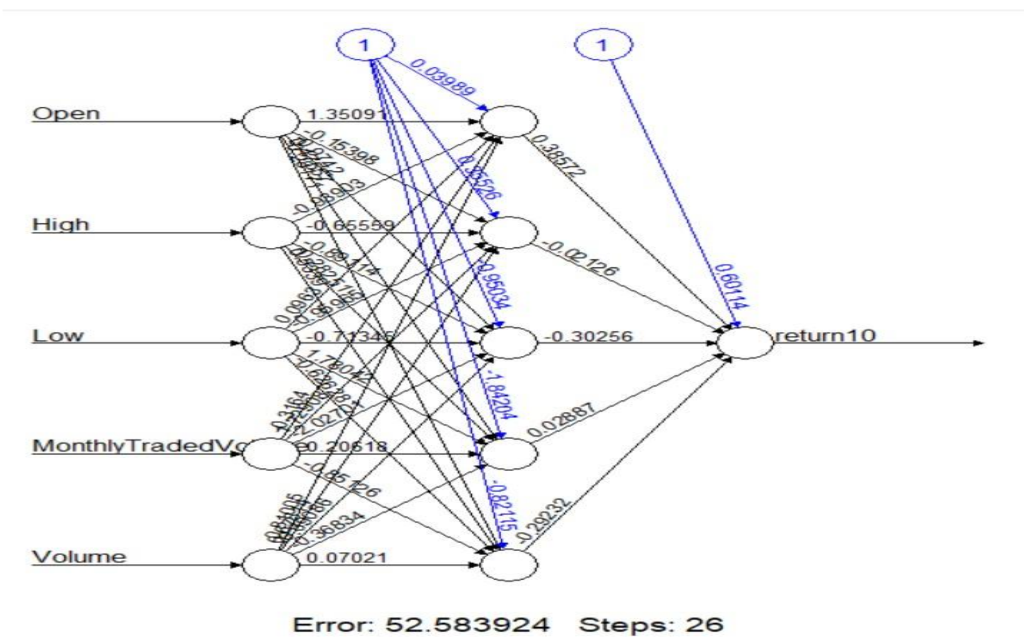
```
Call:
glm(formula = return10 ~ . - Close, family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7597  -0.4462   0.1640   0.5608   2.1077

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.199e+01  4.009e+01  -1.047  0.29492
Month        2.908e-08  3.635e-08   0.800  0.42379
Open        -1.022e-02  2.538e-03  -4.028 5.62e-05 ***
High         1.799e-03  1.517e-03   1.185  0.23584
Low          7.988e-03  2.302e-03   3.470  0.00052 ***
Volume      -3.516e-08  1.693e-07  -0.208  0.83541
EURVSINR    -2.191e-01  1.632e-01  -1.343  0.17935
GBPVSINR     1.655e-01  1.047e-01   1.581  0.11388
USDVISNR     9.562e-02  2.278e-01   0.420  0.67463
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 105.17  on 76  degrees of freedom
Residual deviance:  55.88  on 68  degrees of freedom
AIC: 73.88

Number of Fisher Scoring iterations: 6
```

Stock Market Specific + Exchange Rates + Macro Economic:

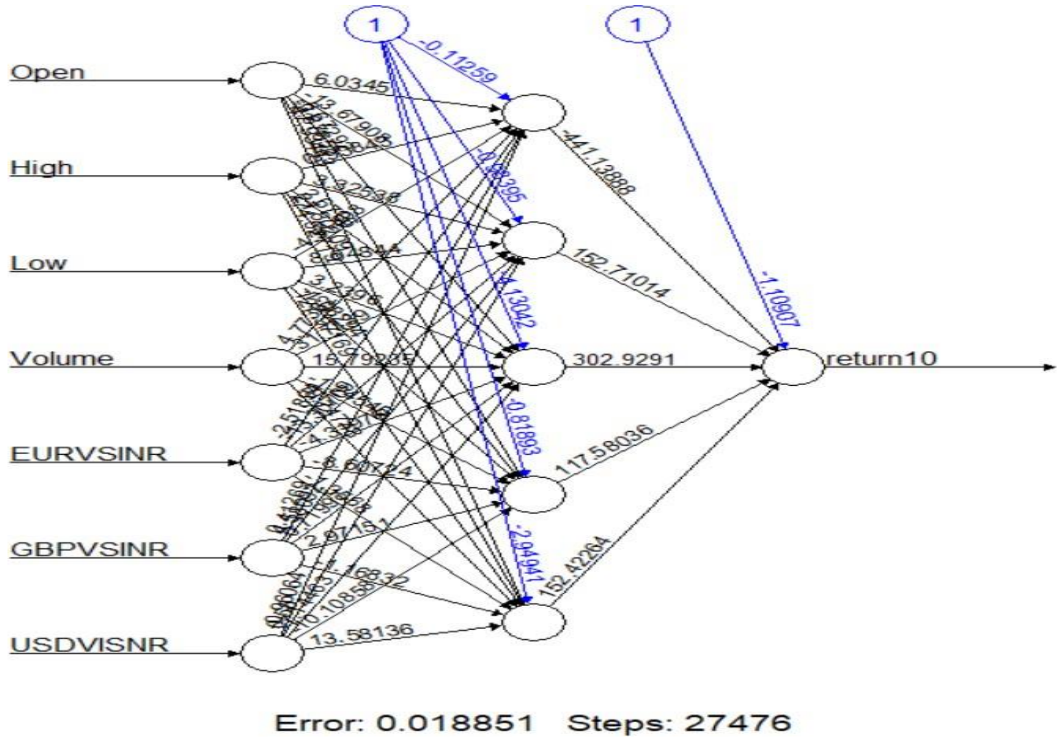
```
Call:
glm(formula = return10 ~ . - Close, family = "binomial", data = train)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.97405  -0.27410   0.08878   0.46869   2.14291

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.686e+01  1.053e+02  -0.445 0.656169
Month         3.143e-08  7.662e-08   0.410 0.681639
Open         -1.143e-02  3.245e-03  -3.524 0.000425 ***
High          4.235e-03  2.150e-03   1.970 0.048877 *
Low           7.078e-03  2.602e-03   2.720 0.006520 **
TradedVolume  7.037e-08  2.519e-07   0.279 0.779943
EURVSINR     -3.360e-01  1.939e-01  -1.733 0.083049 .
GBPVRINR      2.006e-01  1.404e-01   1.429 0.153049
USDVSINR      4.557e-01  3.506e-01   1.300 0.193663
TradeBalance  1.730e-05  2.727e-05   0.635 0.525701
Forex         2.899e-07  5.034e-06   0.058 0.954085
FII           3.088e-05  5.504e-05   0.561 0.574711
Bankrate     -1.105e+00  1.855e+00  -0.596 0.551391
Gold         -6.861e-04  3.913e-04  -1.754 0.079512 .
Silver        1.854e-04  1.858e-04   0.998 0.318457
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 105.168  on 76  degrees of freedom
Residual deviance:  50.567  on 62  degrees of freedom
AIC: 80.567

Number of Fisher Scoring iterations: 6
```

Neural Network:

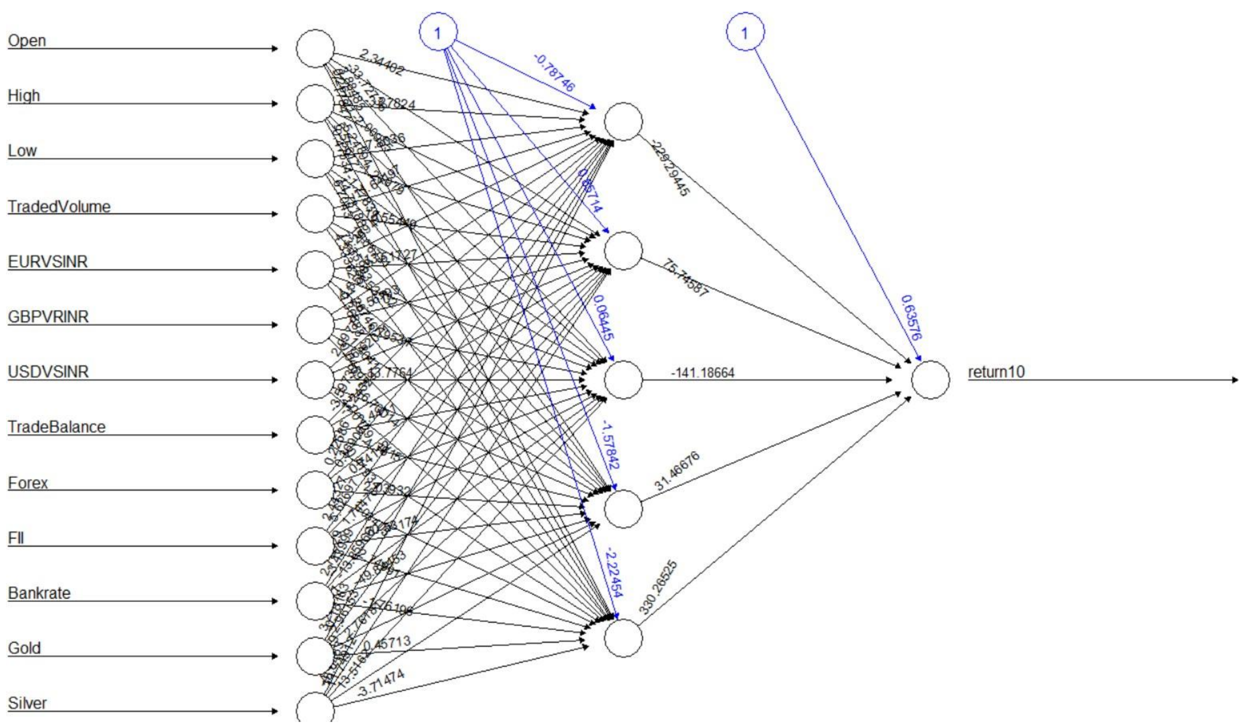
Stock Market Specific Model:



Stock Market Specific + Monthly Closing Prices of three popular Exchange Rates:



Stock Market Specific + Exchange Rates + Macro Economic:



The following table summarizes the results of the analysis:

Independent Variables	Linear Regression			Logistic Regression			Neural Network	
	Dependent Variable	**R2**	**Significant Variables**	**Dependent Variable**	**Error**	**Significant Variables**	**Dependent Variable**	**Error**
Model I	Close	1	Month, Open, Low, Monthly Traded Volume	Return10	6.49%	Open, Monthly Traded Volume	Return10	45%
Model II	Close	99.88 %	Open, High, Low	Return10	22%	Open, Low	Return10	35%
Model III	Close	99.53 %	Open, High, Low	Return10	16.8%	Open, High, Low	Return10	50%

As seen in the above table, three separate sets of variables were used for the analysis, and models were created for each of these sets using one of three machine learning algorithms: linear regression, logistic regression, or neural networks. According to the analysis's findings, each of the algorithms employed in the model that takes into account the variables unique to the stock market has a very low error percentage.

The significant variables identified by each of these models mostly include Stock market specific variables rather than the macro economic variables.

Interpretation:

Linear Regression: Using Linear Regression, we can observe that the R2 is a one in Model I which considers only the stock specific variables. As and when the independent variables are added the R2 value is reducing. Which means that the Linear Regression proposes the use of stock specific variables only.

Logistic Regression: From the analysis of Logistic Regression outputs across three models, we can observe that the error is least with the stock specific variables only. This is also evident from the AIC values of each of these models. Since Model I has the least AIC value which speaks about the goodness of fit and complexity, it can be summarized that stock specific variables are best to explain the variation in the stock closing prices (In Log regression, the increase is taken as a 1 and decrease from the previous value is taken to be zero)

Neural Networks: From the output of neural network we can observe that the error is at 45% if stock specific variables are used. When the exchange rates are used, the error is reducing to 35%. But the error increases when economic indicators are used.

Conclusion:

In conclusion, this study examined the variables influencing stock market prices in the Indian Stock Market. Through the use of distinct datasets and machine learning algorithms, the research aimed to identify the key drivers in predicting stock market closing prices. Interestingly, the findings indicate that incorporating economic factors into the analysis did not significantly enhance the efficiency of the algorithms.

Instead, the study highlights the importance of technical factors such as Open, Low, and High prices in predicting changes in stock prices. This suggests that investors can rely on historical performance and technical indicators to anticipate price fluctuations and make informed investment decisions.

By focusing on the entire market environment and leveraging machine learning techniques, investors can gain valuable insights into the Indian stock market. This enables them to take appropriate actions and maximize the success of their investment selections.

References

1. Ali Khan, J. (2016). Predicting Trend in Stock Market Exchange Using Machine Learning Classifiers. *Google scholar*.

2. Ballings, M., Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Google Scholar*, 7046-7056.

3. Ms N VivekaPriya, D. S. (2022). STOCK PREDICTION USING MACHINE LEARNING TECHNIQUES. *IJCRT*, 48-60.

4. Naeini MP, T. H. (2010). Stock market value prediction using neural networks. *IEEE*.

5. Nusrat Rouf, M. B.-C. (2021). Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. *Electronics*.

6. Polamuri, D. &. (2020). A Survey on Stock Market Prediction Using Machine Learning Techniques. *Researchgate* .

7. PS, C. (2013). Parameters for stock market prediction. Int J Comput Technol Appl. *IJCTA*.

8. Rahman, A., Abdul-Rahman, S., & Mutalib, S. (2017). Mining Textual Terms for Stock Market Prediction Analysis Using Financial News. *Google Scholar*, 293-305.

9. Tan, T., Quek, C., & Ng, G. (2007). Biological Brain-Inspired Genetic Complementary Learning for Stock Market and Bank Failure Prediction. *Google Scholar*, 236-261.

10. *The World Bank*. (n.d.). Retrieved from The World Bank website: https://data.worldbank.org/indicator/CM.MKT.LCAP.CD

11. Upadhyay, A., & Bandyopadhyay, G. (2012). Forecasting Stock Performance in Indian Market using Multinomial Logistic Regression. *Google Scholar*, 16-39.

12. Venkatesh, C., & Tyagi, M. (2011). Fundamental Analysis as a Method of Share Valuation in Comparison with Technical Analysis. *Google Scholar*, 167-174.

13. Zhu, Y., & Zhou, G. (2016). Technical analysis: An asset allocation perspective on the use of moving averages. *Google Scholar*, 519-544.