International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

# Statistical Analysis for Indian Stock Market Prediction

**Aniruddha Navale[1], Ashish Khanapure[2], Narendra Waval[3], Nitin Sakhare [4]**

*[123]BTech Student, Savitribai Phule Pune University, Pune, Maharashtra, India*
*[3] Professor, Dept. of Computer Engineering, VIIT, Pune University, Maharashtra, India*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** *Stock markets are complex and difficult to predict due to their dynamic, non-linear, and chaotic nature. Traditional methods like statistical models struggle with noisy and unstructured data, which is common in stock markets.*

*Neural networks, on the other hand, excel at pattern recognition and can handle complex relationships in data. This makes them well-suited for stock market prediction, even with noisy and unstructured data.*

*Therefore, this research will explore the use of artificial neural networks (ANNs) for stock market prediction. Specifically, we will use standard supervised backpropagation neural network learning methodologies like Boltzmann machines, ADALIN, and MADALIN.*

***Key Words***: *Technical Analysis, Fundamental Analysis, Supervised Back-propagation Neural Network Learning, Boltzmann Machine, ADALIN, MADALIN.*

## 1. INTRODUCTION

While perfectly predicting the stock market remains elusive, it's not without patterns and recurring trends. Investors must carefully consider three crucial factors: the current and projected costs of their desired assets, as well as potential future bid prices. Despite these considerations, many traders rely on historical price data to guide their decisions, leading to diverse approaches: some avoid seemingly overvalued assets expecting a correction, while others hesitate towards declining stocks fearing further losses.

### 1.1 Understanding the Stock Market Ecosystem

Primary Market: Companies raise capital by issuing shares during an Initial Public Offering (IPO).
Secondary Market: After initial purchase, investors trade existing shares amongst themselves at the prevailing market price or a mutually agreed upon price. This secondary market, also known as the stock exchange, is regulated by authorities like SEBI in India.
Stock Brokers: These agents facilitate the buying and selling of shares and securities on behalf of investors. Only listed stocks can be traded on exchanges. The Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE) are India's leading exchanges.

### 1.2 Prediction Strategies:

This paper explores predicting Indian stock market indices using statistical tools and a Madaline neural network. Given the vast amount of daily non-linear data in the market, with possible outcomes of up, down, or unchanged, precise prediction remains a challenge. Researchers have developed prediction models based on regression analysis, emphasizing the importance of selecting effective input variables and identifying their unique patterns. To this end, this study performs a sensitivity analysis of input variables to optimize the prediction model.

In essence, this paper delves into the complexities of predicting the Indian stock market, highlighting the challenges, and exploring potential approaches using statistical tools and neural networks. purposes.

## 2. REVIEW OF LITERATURE:

Egeli, Ozturan, and Badur propose a method to predict Istanbul Stock Exchange (ISE) index values using Artificial Neural Networks (ANNs). Their model uses previous days' index cost, exchange rates, and interest rates as input, and is trained to predict future values based on these patterns. The data for their model was collected from the Central Bank of Turkey for a period of 417 days, starting from July 2, 2001, to February 28, 2003. They used 90% of the data for training and 10% for testing. Their ANN-based model was more accurate than traditional Moving Average models in predicting ISE index values.

In 2005, Qing Cao, Karyl B. Leggio, and Marc J. Schniederjans employed artificial neural networks (ANNs) to predict inventory rate movements (specifically, charge returns) for firms listed on the Shanghai Stock Exchange. Their study compared the predictive accuracy of univariate and multivariate neural network models against linear models. The results demonstrated statistically significant outperformance of ANNs compared to linear models across the sample companies, suggesting the potential of ANNs as a valuable tool for stock price prediction in emerging markets like China.

In 2012, Dastgir and Enghiad investigated the effectiveness of Artificial Neural Networks (ANNs) in predicting the Tehran Stock Exchange Trade Charge Index, the main benchmark for the Iranian stock market. Their study employed ANNs with two hidden layers, varying the number of neurons in each layer between 1 and 16. The analysis revealed that an ANN configuration with 3 neurons in the first hidden layer and 4 neurons in the second achieved the best performance in forecasting the index movement.

In 2006, Datta et al. investigated the use of Artificial Neural Networks (ANNs) in modeling Indian stock market data. Their study focused on the Bombay Stock Exchange (BSE) and its Sensitive Index (Sensex). Similarly, Panda and Narasimhan (2006) employed ANNs to forecast daily Sensex returns. They compared the ANN's performance against random walk and linear autoregressive models using six different metrics.

The development of a model capable of accurately predicting trends in the dynamic stock market could significantly impact both investors and regulators. Such a model would not only facilitate growth in market investments but also empower regulators to implement timely corrective measures when necessary. Backpropagation, a learning approach used in training neural networks, presents a promising avenue for constructing such a model. By applying this technique, neural networks can be trained to anticipate stock market movements, leading to more informed investment decisions and the creation of improved risk management tools.

**Objectives:**

1. Evaluate the reliability and robustness of the chosen data sample for stock market analysis.

2. Develop a multidimensional regression model to predict the future values of Indian stock market indices and individual share prices.

3. Assess the accuracy of the multidimensional regression model using appropriate statistical techniques such as ANOVA.

4. Measure the effectiveness of the model's predictions for the next day's CNX S&P Nifty Index using the Sign Correctness Percentage (SCP) metric. Conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

**Research Method:**

**Multidimensional Regression Analysis**

Regression analysis is a powerful tool for understanding how one variable (dependent) is influenced by changes in other variables (independent). It essentially builds a model to predict the dependent variable based on the independent ones.

Types of regression:

Simple regression: This uses only one independent variable to explain the dependent variable. Think of it as studying how one factor affects another.

Multiple regression: This uses multiple independent variables to explain the dependent variable. Imagine investigating how several factors jointly influence something.

Finding the right fit:

There are two main approaches to building a regression model:

Hit and trial: This involves testing different combinations of independent variables to see which ones best explain the dependent variable. It's like finding the best recipe by experimenting with different ingredients.

Pre-conceived model: This uses prior knowledge to select specific independent variables believed to be relevant. It's like following a proven recipe, but adjusting it based on your preferences.

The math behind the magic:

The general formula for a regression model looks like this:

$Y = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta kXk + \varepsilon$

Y: The dependent variable you're trying to predict.

β0: The intercept, representing the average value of Y when all X's are zero.

βi: The partial regression coefficients, indicating how much each independent variable (Xi) influences Y, holding other variables constant.

Xi: The independent variables you're using to explain Y.

ε: The error term, accounting for any unexplained variation in Y.

Estimating the model:

We use statistical techniques to estimate the β values and create a model that best fits the data. This estimated model can then be used to predict future values of Y based on new values of the independent variables.

**Statistics Associated with Multiple Regressions:**

Regression models aim to explain how a dependent variable (Y) changes with several independent variables (X1, X2, ...). Two key measures assess how well the model fits the data:

1. Adjusted R-squared:

This statistic reflects the portion of Y's variation explained by the model, considering both the number of independent variables and the sample size.

Adding too many variables initially increases R-squared, but the improvement slows down after a certain point. Adjusted R-squared penalizes unnecessary variables, giving a more accurate picture of the model's fit.

2. F-test:

This test checks if the model significantly explains Y's variation.

It compares the amount of explained variation to the unexplained variation.

A low p-value (usually below 0.05) indicates a statistically significant model.

Importance of individual variables:

Even in a significant model, not all variables may contribute equally.

Partial F-test: This test assesses the individual significance of each variable (Xi) after controlling for the others.

It shows if including Xi in the model significantly improves its explanatory power.

Partial regression coefficient:

This value represents the average change in the predicted value of Y (Ŷ) for a one-unit increase in Xi, holding all other variables constant.

It captures the unique influence of each variable on Y, considering its interactions with the others.

**Data Analysis:**

1) NIFTY50
https://www.nseindia.com/products/content/equities/indices/historical_pepb.htm
2) Dow Jones Industrial Average Historical Data from https://in.investing.com/indices/us-30-historical data
3)USR/INR
https://in.investing.com/currencies/usd-inr-historical-data
4) HANG SENG INDEX (^HSI)
https://finance.yahoo.com/quote/%5EHSI/history?period1=1167849000&period2=1531333800&interval=1d&filter=history&frequency=1d

(Descriptive Statistics)

| | N | Minimum | Maximum | Mean | | Std. Deviation | Skewness | | Kurtosis | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| Nifty | 2560 | 2524.20 | 11130.40 | 6315.2458 | 38.18153 | 1931.84944 | .447 | .048 | -.644 | .097 |
| P/E | 2560 | 10.68 | 28.29 | 20.6984 | .06149 | 3.11119 | -.268 | .048 | .172 | .097 |
| DY | 2560 | .82 | 2.24 | 1.2933 | .00481 | .24315 | .702 | .048 | .939 | .097 |
| DJ | 2560 | 6547.05 | 26439.48 | 14600.5382 | 79.93968 | 4044.66318 | .587 | .048 | -.116 | .097 |
| HSK | 2560 | .00 | 32966.89 | 22305.2862 | 69.72077 | 3527.62280 | -.467 | .048 | 3.832 | .097 |
| Rs/$ | 2560 | 39.08 | 68.81 | 54.2498 | .18462 | 9.34103 | -.016 | .048 | -1.489 | .097 |
| Valid N (listwise) | 2560 | | | | | | | | | |

**Reliability Analysis:**
Reliability analysis is done to check whether the Five Independent variables i.e. Rs/$, P/E, HSK, DY, DJ and one dependent variable Nifty would produce consistent result or not.

**Case Processing Summary**

| | | N | % |
|---|---|---|---|
| Cases | Valid | 2560 | 100.0 |
| | Excluded[a] | 0 | .0 |
| | Total | 2560 | 100.0 |

**Reliability Statistics**

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | No. of Items |
|---|---|---|
| .685 | .740 | 6 |

The Nifty measurement tool's reliability was assessed using Cronbach's Alpha. While the overall value (0.685) falls slightly below the "good" threshold (0.7), the standardized item Alpha (0.740) exceeds it. Therefore, most factors measuring the Nifty appear reliable.

Fundamental analysis often uses multiple regression, but this assumes linear relationships. The stock market's non-linear nature challenges that assumption, potentially leading to inaccurate results. Consider alternative analysis methods to capture the market's complexity.

**Variables Entered/Removed[a]**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | Rs/$, P/E, HSK, DY, DJ[b] | | Enter |

**Model Summary[b]**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | | Durbin-Watson |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change | |
| 1 | .986[a] | .972 | .972 | 324.49189 | .972 | 17629.278 | 5 | 2554 | .000 | .042 |

ANOVA[a]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 9281372768.403 | 5 | 1856274553.681 | 17629.278 | .000[b] |
| | Residual | 268923388.144 | 2554 | 105294.984 | | |
| | Total | 9550296156.547 | 2559 | | | |

**Model Fit:**

**R-squared: 97.2% of variations in the dependent variable (Nifty) are explained by the 5 independent variables.**
**Adjusted R-squared: Identical to R-squared, indicating no significant multicollinearity (redundant variables).**
**Autocorrelation:**
**Durbin-Watson test: Value of 0.042 signifies no significant serial correlation (autocorrelation) in the data.**
**Overall Model Significance:**
**F-test: Highly significant p-value (0.000) confirms the model is not likely random, meaning the 5 variables together impact Nifty.**

Residual Statistics[a]

| | Minimum | Maximum | Mean | Std. Deviation | N |
|---|---|---|---|---|---|
| Predicted Value | 2229.8762 | 11135.4639 | 6315.2458 | 1904.45609 | 2560 |
| Residual | -1176.45935 | 1160.29077 | .00000 | 324.17472 | 2560 |
| Standard Predicted Value | -2.145 | 2.531 | .000 | 1.000 | 2560 |
| Standard Residual Value | -3.626 | 3.576 | .000 | .999 | 2560 |

**Coeffecients:**

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig | Co-Linearity Statistics | |
|---|---|---|---|---|---|---|---|
| | B | Std. Error | Beta | | | Tolerance | VIF |
| (Constant) | -7431.296 | 158.289 | | -46.952 | .000 | | |
| P/E | 262.093 | 5.236 | .422 | 50.058 | .000 | .155 | 6.449 |
| DY | 365.691 | 64.090 | .046 | 5.706 | .000 | .169 | 5.902 |
| DJ | .164 | .004 | .336 | 36.482 | .000 | .130 | 7.682 |
| HSK | .34 | .004 | .062 | 9.531 | .000 | .259 | 3.867 |
| Rs/S | 87.516 | 1.624 | .423 | 53.878 | .000 | .179 | 5.595 |

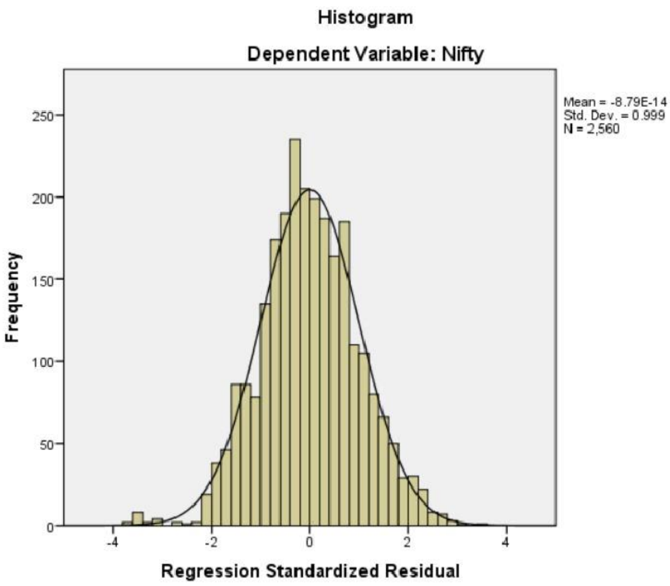
This analysis confirms that all five independent variables significantly influence the Nifty index:

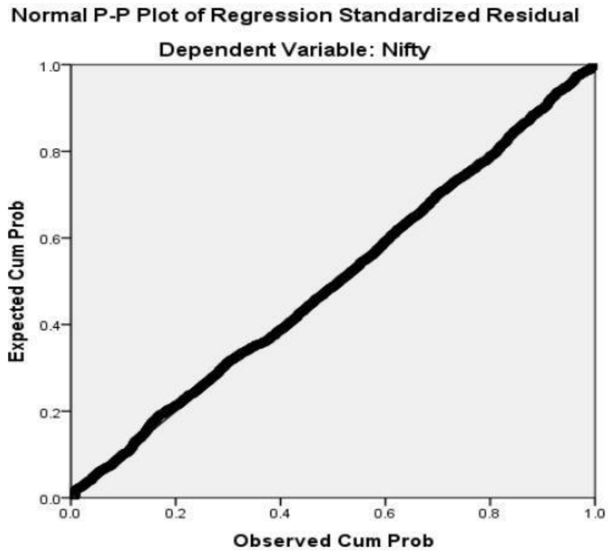
P/E Ratio (X1): Has the strongest positive impact (0.422) on Nifty.
Dividend Yield (X2): Has a positive impact (0.046) but is less influential than P/E.
Dow Jones (X3): Has a significant positive impact (0.336) on Nifty.
Hang Seng (X4): Has a weaker positive impact (0.062) on Nifty.
Rs./Dollar (X5): Also has a strong positive impact (0.423) on Nifty.

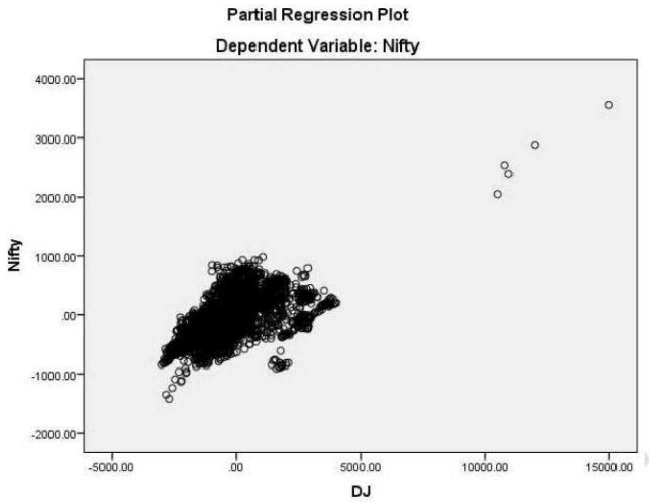
Regression equation obtained from the Analysis as per standardized coefficient Beta is:
Nifty = 0.422 (P/E) + 0.046 (DY) + 0.336 (DJ) + 0.062 (HSK) + 0.423 (Rs. /$)
Y = 0.422 X1 + 0.046 X2 + 0.336 X3 + 0.062 X4 + 0.423 X5
Regression equation obtained from the Analysis is:
Nifty = -7431.926 + 262.093 (P/E) + 365.691 (DY) + 0.160 (DJ) + 0.034 (HSK) +87.516 (Rs. /$)
Y = -7431.926 + 262.093 X1 + 365.691 X2 + 0.160 X3 + 0.034 X4 +87.516 X5

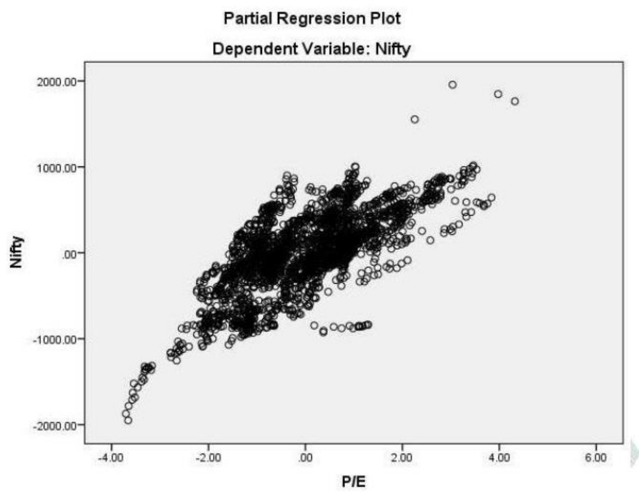
Correlations

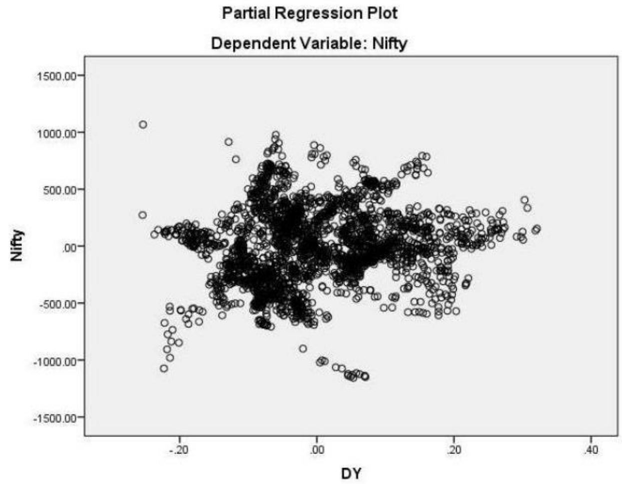
| | | Nifty | P/E | DY | DJ | HSK | Rs/$ |
|---|---|---|---|---|---|---|---|
| Pearson Correlation | Nifty | 1.000 | .694 | -.243 | .936 | .680 | .788 |
| | P/E | .694 | 1.000 | -.792 | .556 | .753 | .177 |
| | DY | -.243 | -.792 | 1.000 | -.138 | -.576 | .302 |
| | DJ | .936 | .556 | -.138 | 1.000 | .675 | .780 |
| | HSK | .680 | .753 | -.576 | .675 | 1.000 | .235 |
| | Rs/$ | .788 | .177 | .302 | .780 | .235 | 1.000 |

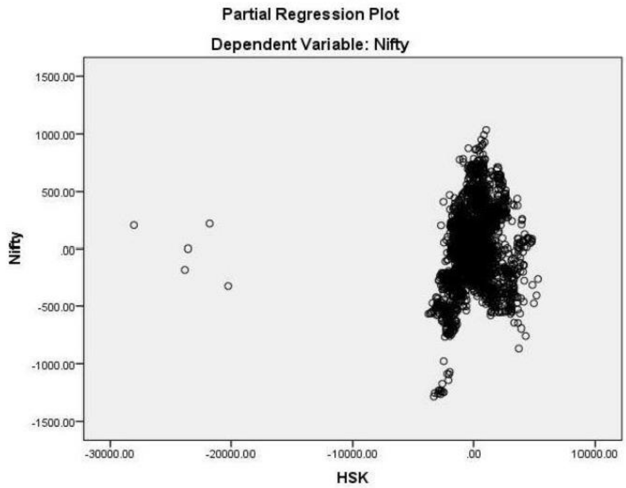


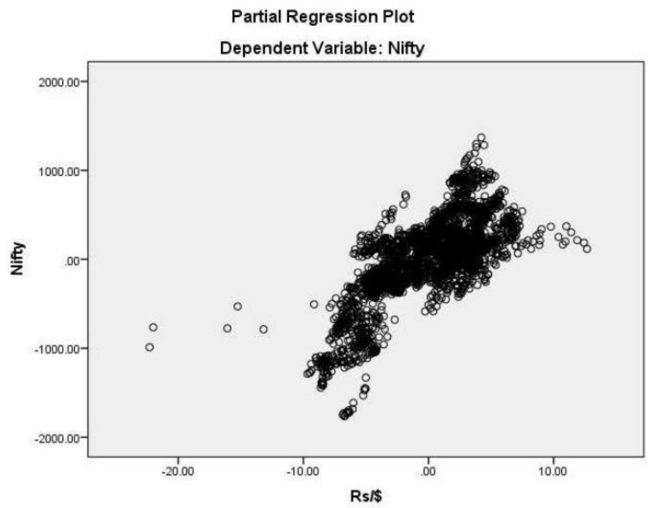



## 3. CONCLUSIONS

Accurately predicting the ever-evolving dynamics of the stock market remains a captivating yet elusive goal for investors and researchers alike. This paper delves into the potential of utilizing Artificial Neural Networks (ANNs), specifically Supervised Back-propagation Learning

Methodology, as a novel approach to enhance stock market prediction accuracy.

Current limitations in forecasting models stem from the inherently noisy, unstructured, and multifaceted nature of stock market data. Traditional regression analyses, while valuable, fall short in comprehensively capturing the interplay of known and unknown influences on stock prices. This paper posits that ANNs, with their unique ability to identify complex patterns in intricate datasets, may offer a breakthrough solution.

Drawing inspiration from the parallel processing architecture of the human brain, ANNs can learn from vast amounts of historical data, uncovering subtle patterns and associations beyond the grasp of conventional models. This ability to discern hidden relationships offers the potential to significantly improve our understanding of factors driving stock market movements.

While leveraging ANNs in this domain holds immense promise for both investors and agricultural researchers, a realistic approach acknowledges the inherent complexities involved. The unpredictable nature of financial markets and external factors necessitates a cautious interpretation of any model's results.

The proposed research utilizes Boltzmann Machines, ADALINs, and MADALINs within the Supervised Back-propagation Learning Methodology. This framework aims to identify hidden patterns within stock market data, potentially leading to more informed investment decisions and optimized crop yields.

Beyond the specific application to stock market prediction, exploring the potential of ANNs opens exciting avenues for research in diverse fields. For instance, understanding the combined effect of various factors on rice yield per acre, as mentioned in the original text, could be significantly benefitted by employing this methodology.

This paper advocates for a comprehensive exploration of ANNs as a promising avenue to unlock the secrets of market behavior and guide informed decision-making across various domains. It acknowledges the challenges and limitations inherent in financial forecasting, while emphasizing the potential rewards of pushing the boundaries of predictive methodologies.

## REFERENCES

[1]  Schoeneburg,E, Stock Price Prediction Using Neural Networks, A Project Report,

[2] Neurocomputing,Vol.2,1990,pp.17-27.

[3]  Wong, B.K. Bonvovich,T.A. and Selvi,Y., neural Network Applications in Business: A Review and

[4] Analysis of the Literature(1988-95),Decision Support System, Vol.19,1997,pp.301-320.

[5]  Z. Tang and P.A.Fiswick, "Backpropagation Neural Networks as Models For Time

[6] SeriesForecasting"QRSA Journal on computing Vol.5,No.4, pp.374-384,1993.

[7]  J.H Wang And J.V.Leu, "Stock Market Trend Prediction Using ARIMA Based Neural Network",Proc.

[8] Of IEEE Conference on neural Networks, Vol.4,pp.2160-2165,1996.

[9]  Haykin S.(1994) Neural Network:A Comprehensive Foundation Macmilan College Publication, New

## BIOGRAPHIES :

Aniruddha Navale
(BTech 4th year)

Ashish Khanapure
(BTech 4th year)

Narendra Wavhal
(BTech 4th year)

Dr. Nitin Sakhare
(Assistant Professor- Department of Computer Engineering, VIIT Pune)