Identifying Significant Macroeconomic Indicators for Indian Stock Markets

PRAKASH K. AITHAL1, (Member, IEEE), DINESH ACHARYA U.2, AND GEETHA M.3,

1Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India (e-mail: prakash.aithal@manipal.edu)
2Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India (e-mail: dinesh.acharya@manipal.edu)
3Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India (e-mail: geetha.maiya@manipal.edu)

Corresponding author: Geetha M. (e-mail: geetha.maiya@manipal.edu).

The authors would like to thank Enago (www.enago.com) for the English language review.

ABSTRACT The macroeconomic indicators play a major role in all the stock markets, and they vary from nation to nation. This paper identifies the influence of macroeconomic indicators on the National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE) of India. Total of forty-four macroeconomic indicators for eight years from the year 2011 to 2018 are considered in this study. The macroeconomic factors are aggregated and considered in average monthly form. The proposed method finds the correlation matrix of all considered macroeconomic indicators. The need for dimensionality reduction and the existence of multicollinearity are proven using validation techniques such as the Kaiser-Meyer-Olkin and Bartlett tests. The Principal Component Analysis (PCA) method is used to reduce the dimensionality to seven factors and then PCA with the varimax rotation method is applied to find factors with maximum variation. In addition, the influence of these seven factors on the NSE Nifty and BSE SENSEX indices are analyzed using regression. Finally, an Artificial Neural Network is used to predict stock market movement with the help of macroeconomic indicators. Accuracy of 92% and 87% are obtained on NSE NIFTY and BSE SENSEX respectively.

INDEX TERMS Decision Support Systems, Knowledge Discovery, Macroeconomic Indicators, Principal Component Analysis, Artificial Neural Network, Data Mining

I. INTRODUCTION

THERE are many fundamental theories in finance. For example, the Markowitz risk-return theory, Sharpe’s Capital Asset Pricing Model (CAPM), and Arbitrage Price Theory (APT) are prominent among them. The Markowitz model proposes that one must build a portfolio i.e., a collection of stocks, rather than investing in individual stocks. The portfolio is collection of different stocks. The portfolio risk will always be less than the individual stocks forming the portfolio, and return will be an average of returns of the constituting stocks of the portfolio. The CAPM model is an extension of the Markowitz risk-return theory. In the CAPM market risk, or systematic risk is the only risk that affects the stock market, and company-specific risk is reduced to zero by diversification. There are two types in the APT model they are single factor and multi-factor APT. Unlike the CAPM, in a single factor APT, the APT user can select the factor. In multi-factor APT, more than one macroeconomic or financial indicators are considered for study. The Fama French model is a three-factor model. The general APT model does not require returns to be normally distributed, and investors need not be mean-variance optimizers. The APT proposes that the stock market is dependent on macroeconomic indicators.

The influence of macroeconomic indicators on the stock market is inevitable. However, macroeconomic factors affecting one nation need not affect the another nation because different macroeconomic indicators capture different information, some macroeconomic indicators are global in scope, and some are specific to a given nation, and others are unique to specific industries. If a nation is not connected economically with the world, then the influence of macroeconomic indicators with the global scope is either minimal or nonexistent. The proposed paper utilizes the following techniques

- Knowledge Discovery from Data
- Pre-Dimensionality Reduction methods such as
- Correlation Matrix
- Bartlett Test
- Kaiser Meyer Olkin (KMO) Test

- Dimensionality Reduction methods such as
  - PCA
  - K1-Kaiser Method
  - Scree Test
  - PCA with varimax rotation

- Unit Root Test and measure such as
  - Augmented Dickey-Fuller (ADF) Test
  - Vector Error Correction Model (VECM)

- Data Mining techniques such as
  - Regression Analysis
  - ANN

The above techniques are explained in the following subsection.

A. KNOWLEDGE DISCOVERY FROM DATA

Knowledge Discovery from Data (KDD) as shown in Figure 1. includes data collection, data cleaning, data transformation, data mining, result evaluation, result analysis, and result presentation. Data Mining is a tool to analyze data to extract nontrivial information. The data are collected from the different sources. Data cleaning steps include missing value imputation and outlier removal, and data transformation involves converting data to make it suitable for a given data mining task.

B. PRE-DIMENSIONALITY REDUCTION

The need for dimensionality reduction is identified by applying the correlation matrix, the Bartlett test, and the KMO test.

1) Correlation Matrix

Correlation matrix is obtained by computing the Pearson correlation coefficient formula given in Equation 1.

\[ \rho_{xy} = \frac{S_{xy}}{S_x \times S_y} \] (1)

Here \( S_{xy} \) is the sample covariance of \( x \) and \( y \) and \( S_x \) and \( S_y \) are sample standard deviations of variables \( x \) and \( y \) respectively.

The correlation coefficient value varies between -1 to 1. The value -1 signifies that if the value of \( X \) increases, then the value of \( Y \) decreases which is a perfect negative correlation. In contrast, a value of 1 signifies if the value of \( X \) increases, then the value of \( Y \) also increases which is a perfect positive correlation. 0 signifies there is no correlation. Any intermediate value between -1 and 1 indicates that the \( X \) and \( Y \) variables vary proportionately. If there are many highly positive or highly negative values are present in the correlation matrix, then there are chances of the existence of multicollinearity. Multicollinearity is a situation in which multiple variables are interdependent. Bartlett and KMO tests are used to identify multicollinearity. If multicollinearity is present in the system of variables, then dimensionality reduction technique such as PCA can be applied.

2) Bartlett Test

The Bartlett test is used to determine if 'k' variables have the same variance. Equation 2 depicts the Bartlett formula.

\[ B = \frac{(N - k) \ln(\frac{\sum_{i=1}^{k} (n_i - 1) s_i^2}{(N-k)}) - \sum_{i=1}^{k} (n_i - 1) \ln(s_i^2)}{1 + \frac{1}{3k(k-1)} \left( \frac{\sum_{i=1}^{k} (n_i - 1)}{N-k} - \frac{1}{N-k} \right)} \] (2)

Here \( N \) is the sum of all sample sizes, \( k \) is the number of normally distributed populations, \( s_i^2 \) is the variance, and \( n_i \) the number of sample items.

Note that the factor analysis should not be applied if Bartlett test is proved insignificant. Bartlett test utilizes a chi-square test, which is well suited for normal distributions or a large number of variables.

3) KMO Test

The KMO test validates the adequacy of the sample. Equation 3 depicts the KMO formula.

\[ KMO_j = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u_{ij}} \] (3)
Here $r_{ij}$ is the correlation matrix and $u_{ij}$ is partial covariance matrix.

Researchers have different opinions about the significant value of KMO. One school of thought proposes that the KMO value should be greater than 0.5 while others take the opinion that the KMO value should be greater than 0.6 or 0.7. The KMO test cannot be applied if there are variables that can deduce other variables perfectly. In other words, KMO test cannot be applied if two variables are perfectly correlated, which occurs when the correlation matrix is nonpositive definite.

**C. DIMENSIONALITY REDUCTION**

The dimensionality is reduced using the PCA algorithm and related macroeconomic indicators are grouped together into factors. The number of factors to be retained are determined through K1-Kaiser and Scree tests. The prominence of each macroeconomic indicator within a factor is identified using PCA with varimax.

1) **PCA**

PCA, which is an unsupervised factor analysis method, is a dimensionality reduction technique that takes an input correlation matrix and produces eigenvalue and eigenvector. There are two types of factor analysis i.e. confirmatory and exploratory factor analysis. In confirmatory factor analysis number of factors is known. In contrast, the number of factors is unknown in exploratory factor analysis. In other words, confirmatory factor analysis is supervised factor analysis, and exploratory factor analysis is unsupervised factor analysis. A detailed PCA tutorial is given in literature [1].

2) **K1-Kaiser**

The K1-Kaiser method is used to retain the factors. According to K1-Kaiser method, one must retain the factors that have an eigenvalue greater than one. It is the simplest of the methods to retain the factors.

3) **Scree Test**

The scree test is another method used to retain the factors. According to this test, one must retain factors until there is a long tail in the graphical scree plot. Note that there will be breakpoints in the scree plot until there is a long tail, and these breakpoints distinguish between the stronger factors and less relevant factors.

4) **PCA with Varimax Rotation**

PCA cannot tackle the multicollinearity problem, Thus, PCA with varimax is utilized. In varimax, the rotation axis is rotated orthogonally to solve the multicollinearity problem.

**D. UNIT ROOT TEST AND VECM**

The unit root test is performed to determine which method to apply to the data. The ADF test is a widely used representative unit root test.

1) **ADF Test**

The ADF test finds the autocorrelation of a variable. If the p-value of the ADF test is significant then we say that the variable under study is stationary else non-stationary. If the dependent and independent variables are stationary then simple regression can be performed. If the dependent and independent variables are stationary at first difference and residual is stationary at an initial level, then cointegration between variables exists. If there is only one endogenous variable, an error correction model should be utilized; otherwise, a VECM should be implemented.

2) **VECM**

VECM is applied when multiple endogenous variables are nonstationary at an initial level when the residual is stationary. They estimate the short and long term effects of one time series over the other.

**E. DATA MINING**

Data mining is an important step in KDD. Several data mining techniques are available for classification, association, and clustering. The regression analysis and ANNs are considered classification techniques.

1) **Regression Analysis**

Regression determines the relationship between a dependent variable and independent variable under the assumption that extraneous variables do not affect the equation.

2) **Artificial Neural Network**

An ANN is a model that mimics the human brain. The ANN consists of an input layer, hidden layers, and an output layer. The input layer passes the input values to the hidden layers, and the hidden layers pass the result to the output layer. All neurons in an ANN are connected by at least one connection. Neurons are activated by an activation function, which can be a binary step, linear, or nonlinear function.  

Note that the binary step function can take only two values: yes or no. A linear function makes the ANN equivalent to a linear regression model: However, a linear function does not allow backpropagation.

The nonlinear activation functions include the sigmoid, hyperbolic tangent, Rectified Linear Unit(ReLU), softmax, and swish functions. The sigmoid function is heavy, while the ReLU function is efficient. The sigmoid function allows all types of integers, while the basic ReLU function fails for negative integers: However, there are ReLU variants that can work with negative values.

Rest of the paper is organized as follows A literature review is provided in, Section II, the methodology is described in Section III. The results are discussed in Section IV, and Section V concludes the paper and gives the future direction.

**II. LITERATURE REVIEW**

Previous studies have investigated the effect of macroeconomic indicators on the performance of the index. These
studies considered the stock market of various nations and a variety of macroeconomic indicators.

For example, T. J. Flavin [2] studied the influence of macroeconomic factor retail price index on the stock market in the United Kingdom. He observed that Henry Markowitz’s mean-variance analysis model and Sharpe’s capital asset pricing model assume a constant portfolio frontier. In reality, the portfolio frontier varies with time due to the effect of macroeconomic factors on stock market performance. The asset pricing theory assumes that stock market performance is influenced by multiple factors referred to as macroeconomic and microeconomic indicators.

Zarina [3] studied the influence of Unexpected Gross Domestic Product(UGDP), Unexpected Total Trade (UTT), and Unexpected Market Returns (UMR) relative to the stock markets in Malaysia, Singapore, Thailand, Indonesia, and the Philippines. She found that the UGDP, UTT, and UMR have very little influence on these stock markets.

Xue CHEN and Xuejun JIN [4] studied the Chinese stock market and following macroeconomic indicators risk premium, credit supply, demand factor, economic policy measures and they found that the considered macroeconomic factors influence the Chinese stock market. Note that the study undertaken is not on the pure macroeconomic indicators.

Shilpa Sudhakaran and Dr. P. Balasubramanian [5] considered the macroeconomic indicators inflation rate, Foreign Direct Investment(FDI), money supply, Foreign Exchange(Forex), industrial production index, and foreign portfolio investment. They selected these indicators because it has not been considered in the previous studies. They identified that FDI and Forex influence the Indian stock market. The macroeconomic indicators were selected randomly.

Stijn Claessens and M Ayhan Kose [6] surveyed the influence of macroeconomic indicators exchange rates and interest rates on the international stock markets. They concluded that some researchers think that, due to limited knowledge of macroeconomic indicators the stock markets vary randomly while some others argue that humans have thorough understanding of the influences of macroeconomic indicators on stock markets.

Qi Jessica Zhang et al. [7] considered the macroeconomic indicators Gross Domestic Product(GDP), inflation, ten-year bond yield, credit spread, and short term interest rate and studied their influence on stock markets in United States of America and United Kingdom. They concluded that macroeconomic factors influence these stock markets.

Gurloveleen K and Bhatia BS [8] have undertaken India BSE 500 stock index for 150 manufacturing firms. They studied the effect of ten macroeconomic indicators namely, crude oil price, call money rate, exchange rate, Forex, broad money, foreign institutional investors, gross fiscal deficit, industrial production index, and inflation rate on the manufacturing firms stock index BSE 500 and conclude that the stock market is weak-form efficient and macroeconomic indicators do not influence the stock market.

Chetana Parmar [9] explored the possible influence of seven macroeconomic indicators namely, reserve rate, repo rate, industrial production index, exchange rate, consumer price index, gold rate, and oil rate on the BSE SENSEX the Indian stock index. She concluded that, local macroeconomic indicators have greater influence than the global macroeconomic indicators on the Indian stock index BSE SENSEX in the long run.

Andry Alamsyah and Muhammad Fahmi Permana [10] evaluated the effect of five macroeconomic indicators, namely, GDP, investment, import, export, and total expenditure on the Indonesian stock market. They concluded that it is possible to predict the stock market using macroeconomic indicators.

Zhihao PENG and Xucheng LI [11] proposed that real GDP, inflation, and unemployment are independent of each other, and these macroeconomic indicators can be used to predict the stock market fairly efficiently.

Shreya and Sangeetha [12] examined the effect of ten macroeconomic indicators namely, gold prices, industrial production, silver prices, exchange rate, interest rate, money supply, oil prices, inflation, forex reserve, and trade balance on the Indian stock market cnx NIFTY. They found a strong correlation between different macroeconomic indicators.

Pooja Misra [13] studied the influence of industrial production index, inflation, rate of interest, price of gold, rate of exchange, FII, and the supply of money in the Indian stock market. They identified a long term relationship between macroeconomic indicators and the BSE SENSEX, they also found that inflation and the money supply have no short-term influence on the BSE SENSEX.

Khanyisa Ntsangase et al. [14] found that there exists a strong influence between the macroeconomic factors namely, inflation, government expenditure, money supply, interest rate, and exchange rate and the South African stock market.

 Cordelia Onyinyechi Omodero and Sunday Mlanga [15] found that the interest and exchange rates do not influence the Nigerian stock market. They also found that inflation negatively influences the stock market and GDP positively influences the Nigerian stock market.

Yaguang Wang et al. [16] studied the effect of four macroeconomic indicators, namely, industrial production index, consumer price index, unemployment rate, and federal funds rate on the stock market in the United States of America. They found that these macroeconomic indicators have a very strong influence on the stock market of USA.

Peiwen Li [17] proved that macroeconomic indicators GDP, CPI, and composite stock index influence the housing prices in China; thus, housing prices and the stock index are interrelated.

Amith Vikram Megaravall and Gabriele Sampagnaro [18] considered the exchange rate and CPI of India, China, and Japan and found their influence on developing economies India and China and developed economy Japan. The results indicate that the exchange rate and CPI influence these three
the stock market when an analysis was performed using a model, while all the macroeconomic indicators influence stock market.

while banking sector development, economic growth, and indicators on the South African stock market. He found employment rate have a negative influence on the Indian stock market. Their result indicates that GDP, inflation, and exchange rate have a positive influence on the stock market, while the crude oil price has a negative influence on the Indian stock market.

Alina Kvietkauskiene and Modestas Plakys [21] surveyed the effects of macroeconomic factors on stock market indicators. Through survey they found that GDP, 10-year government benchmark yields, pension funds market size, investment funds market size, employment rate, economic sentiment indicator, purchasing managers indicator, and ZEW have positive influence on the stock market, while gold price, silver price, oil price, country credit rating, unemployment rate, and P/E ratio have negative influence on the stock market. In addition, the consumer price index, inflation, government debt level, and exchange rate have a mixed influence.

Sin-Yu Ho and N.M. Odhiambo [22] empirically studied the impact of trade openness, the banking sector, and exchange rate on the Philippines stock market. They found that trade openness has a negative impact, while banking sector development and exchange rate have a positive impact on the Philippines stock market.

Willem Thorbecke [23] found that oil price has a strong relationship with the USA stock market. He concluded that after 2010 due to soaring oil production by the USA, the oil prices are beneficial to the USA stock market.

Keshav Garg and Rosy Kalra [24] studied the effect of six macroeconomic indicators, namely, GDP, gold prices, exchange rate, foreign exchange rate, average inflation, and unemployment rate on the Indian stock market. They found that GDP, gold prices, exchange rate, and foreign exchange rate have a positive impact, and average inflation and unemployment rate have a negative influence on the Indian stock market.

Sin-Yu Ho [25] studied of the influence of macroeconomic indicators on the South African stock market. He found that trade openness and interest rate have a negative impact, while banking sector development, economic growth, and exchange rate have a positive impact on the South African stock market.

Geul Lee and Doojin Ryu [26] found that the macroeconomic indicators do not affect when analyzed with the linear model, while all the macroeconomic indicators influence the stock market when an analysis was performed using a nonlinear model.

H.M. Nijam et al. [27] empirically proved that GDP, exchange rate, and interest rate has a positive impact. Inflation has a negative impact, and balance of payment has no impact on Sri Lankan stock exchange.

Hassan Heidari et al. [28] considered the effect of inflation in health care, oil price shocks, and monetary policy on Iranian pharmaceutical companies. They found that these macroeconomic factors influence the Iranian stock market.

Jawad Khan and Imran Khan [29] analyzed the Karachi stock exchange of Pakistan and found that money supply, exchange rate, and interest rate have a strong influence on the stock market in the long run. They also found that exchange rate has a negative impact in the short term.

Waqar Ahmed et al. [30] studied the effect of three macroeconomic indicators, namely, inflation, interest rate, and exchange rate on Karachi stock exchange-100 and concluded that these macroeconomic indicators were insignificant and does not influence the stock market in Pakistan.

Ghulam Abbas et al. [31] analyzed the influence of macroeconomic factors on stock exchanges in the USA, UK, France, Italy, Canada, Japan, and Germany. They identified that industrial growth and oil prices as the most significant influencing factors affecting the stock markets of the G7 nations.

Boldwin Ndlovu et al. [32] explored the effect of macroeconomic indicators on the Johannesburg stock exchange. They concluded that inflation, interest rate, and money supply have a positive impact on stock prices, while, the exchange rate influences stock prices negatively.

Jung Wan Lee and Tantatape Brahmasrene [33] studied the effect of macroeconomic factors on the Korean stock market. They concluded that the stock markets are affected by macroeconomic indicators in the long and short term. The influencers were industrial production index, exchange rate, and inflation. They found that money supply does not affect the Korean stock market.

Anup Chowdhury et al. [34] found that changes in money supply and government expenditure affect liquidity in emerging Asian stock markets.

Phan Thi Quoc Huong [35] identified that inflation has positive influence and corporate tax rate has a negative influence on the corporate capital structure of the stock market in Vietnam.

The result of these studies vary from macroeconomic indicators not influencing the stock market performance to the very strong influence on performance. From this literature review, it is evident that the selection of macroeconomic indicators greatly influences whether any given indicator will, in turn, affect the stock market. The reviewed research has considered a one to ten macroeconomic indicators and studies their effect on stock market indicators. The influence of macroeconomic indicators on stock market in the USA, UK, France, Germany, Italy, China, Japan, South Africa, Nigeria, Vietnam, Canada, Pakistan, Bangladesh, Malaysia, Philippines, Singapore, and India were explored. The selection of the macroeconomic indicator varies from nation to
nation, and depends on how well the nation is connected economically with the rest of the world. The following methods were employed in most of the reviewed papers

- Auto Regressive Distributed Lag (ARDL)
- ADF test
- Multi-regression
- Granger causality test

They identify whether there is exogenous variable that affects the stock market through these tests. In some of the cases, the result suggest exogenous variables do affect the stock market performance.

III. METHODOLOGY

The premium dataset trading economics with code SGE offered through quandl.com is considered for analysis, as well as data from the World Gold Council (WGC). Note that, the missing values were substituted by their means in a data cleaning process. The data was present in different granularity. For example, GDP was recorded yearly, change in inventories was recorded quarterly, the balance of trade was recorded monthly, foreign exchange reserve was recorded weekly, and treasury bill 91 days yield was recorded daily. All the macroeconomic indicators are transformed into monthly observations. The forty-four macroeconomic indicators for a period of eight years from January 2011 to December 2018 were considered. Thus, the dataset consists of forty-four columns and ninety-six rows.

A Correlation matrix of all the forty-four macroeconomic indicators was found. KMO and Bartlett tests were applied to identify the suitability of applying a dimensionality reduction algorithm [36]. The Bartlett B and KMO K values were calculated using equations 2 and 3 respectively. The PCA algorithm was applied [37], and the number of factors to be retained was obtained by applying the K1-Kaiser method and scree test [38]. According to the K1-Kaiser method all factors whose eigenvalue is greater than 1 are retained. In a scree plot, if there is a break between two factors, then those factors are retained. In addition, in a scree plot, user select the threshold. After identifying the number of factors, PCA with varimax rotation was utilized to identify the belongingness of each macroeconomic indicator to each factors. In addition, ADF test was applied to identify the autocorrelation of prominent macroeconomic indicators and stock indices. It was observed that both endogenous and exogenous variables were initially non-stationary and the residual was stationary. The endogenous and exogenous variables are stationary at first difference which implies that there is cointegration between them; thus VECM should be performed. VECM was performed, and correction was applied. Regression analysis was applied with NSE Nifty as the dependent variable and prominent macroeconomic indicators from each of the factors as independent variables. Regression analysis was repeated with BSE SENSEX as the dependent variable. The purpose of applying regression analysis was to identify which factors influence the stock market positively and which factors have a negative impact. The Figure 2 depicts the block diagram of the process. The approach is similar to that of KDD (Figure 1).

Finally, a feedforward neural network with backpropagation with sigmoid and ReLU as activation functions were applied to predict the stock index prices from prominent macroeconomic indicators belonging to each of the factors.
which depends on balance-of-trade, are dropped. The Bartlett test is conducted with all the forty-four macroeconomic indicators as it does not have the constraint of dependency between variables. The result of the Bartlett test is presented in Table3. As p-value is zero for the Bartlett test, which is less than 0.05, and KMO value is 0.79, which is greater than 0.5, one can consider reducing the dimensionality of the problem.

The PCA algorithm results are shown in Table4. According to the K1-Kaiser method, seven factors whose eigenvalue was greater than one were considered. The number of factors identified through the scree plot matches that obtained by the K1-Kaiser method. The scree plot is depicted in Figure.10. PCA with varimax rotation was applied with seven factors. The PCA with varimax gives the belongingness of each of the macroeconomic indicator for each factor. Table5 depicts the belongingness of prominent macroeconomic indicators in each of the seven factors obtained from PCA with varimax. The macroeconomic indicators fall to the factor for which its belongingness is highest. The macroeconomic indicators under each of the factors are given in Tables7,8,9,10,11,12, and 13. The seven factors can explain 80% of the variance among the macroeconomic indicators. Table6 depicts the cumulative variance explained by each factor. The prominent macroeconomic indicators from each of the factors were considered for regression analysis. A list of prominent and representative macroeconomic indicators is given in Table14.

Regression analysis was performed once for the NSE Nifty index and once for the BSE SENSEX index. The regression equation for the NSE Nifty index is given in Table14. The regression equation for the BSE SENSEX index and once for the BSE SENSEX index. The regression equation for the NSE Nifty index is presented in Equation-5. As p-value is zero for the Bartlett test, which is less than 0.05, and KMO value is 0.79, which is greater than 0.5, one can consider reducing the dimensionality of the problem.

The ANOVA result are presented in Tables17 and 18 respectively. The ANOVA result are presented in Tables17 and 18 respectively. The regression equation for the NSE Nifty index is presented in Equation-5.

\[ \text{NIFTY} = -7951.43579275191 + 159.713597784824 \times \text{CORPORATETAX RATE} + 214.932823343983 \times \text{CURRENTACCOUNTTOGDP} + 4.93315919740711 \times \text{GDP} + 0.0346532627892473 \times \text{IMPORT} - 14.3079912478492 \times \text{INDUSTRIALPRODUCTION} - 118.4865379 \times \text{PRIMELENDINGRATE} - 6.56555 \times 0.05 \times \text{TOURISTARRIVALS} \]

\[ \text{SENSEX} = -52808.87195 + 748.3161561 \times \text{CORPORATETAX RATE} + 380.2482201 \times \text{CURRENTACCOUNTTOGDP} + 18.88783164 \times \text{GDP} - 0.063282173 \times \text{IMPORT} - 31.0047566 \times \text{INDUSTRIALPRODUCTION} + 1683.365331 \times \text{PRIMELENDINGRATE} - 0.002197158 \times \text{TOURISTARRIVALS} \]

It is evident from the results that even within the nation, different indices are affected by different macroeconomic indicators differently. From equation4, it is evident that the macroeconomic indicators belonging to factors 1, 2, 4, and 5 have a positive relationship with the NSE Nifty index, and the macroeconomic indicators belonging to factors 3, 6, and 7 have a negative relation with NSE Nifty. From Equation 5 it is clear that macroeconomic indicators belonging to factors 1, 4, 5, and 7 have a positive relationship with the BSE SENSEX while macroeconomic indicators belonging to factors 2, 3, and 6 have a negative relation.

A nonlinear ANN with the following configuration was constructed:

1) ReLu activation function
2) Alpha value of 0.0001
3) Beta-1 value of 0.9; Beta-2 value of 0.999
4) Early stopping was set to false
5) Epsilon value of 1e-08
6) Maxiteration set to 1000
7) Validation fraction was set to 10%

The loss function of the feedforward neural network with backpropagation and ReLu activation function for the Nifty index as the output variable is shown in Figure.3, and the same for the SENSEX as the output variable is shown in Figure.4. It is observed that with the ReLU activation function the optimizer would not converge even after thousand epochs.

\[ \text{FIGURE 3: Loss function of ANN with ReLU for Nifty as the target variable} \]

The feedforward neural network with backpropagation and sigmoid activation function demonstrated better performance and accuracy. The feedforward neural network was constructed with three layers. There were seven neurons in the input layer corresponding to the seven factors. The optimal number of hidden layer neurons was obtained by trial and error. Ten neurons were present in the hidden layer, and the output layer consists of one neuron corresponding to the stock market index. During the training phase, 60% of the data was allotted for training, 20% for validation, and 20% for testing. The built ANN is depicted in Figure.5. R plots for ANN with Nifty and SENSEX as target variables are
FIGURE 4: Loss function of ANN with ReLU for SENSEX as the target variable

depicted in Figures 6 and 7, respectively. The performance plots for Nifty and SENSEX as the target variables are presented in Figures 8 and 9, respectively.

FIGURE 5: ANN with three layers

FIGURE 6: R plot for ANN with Nifty as target variable

The test data span from January 2019 to August 2019. After testing the ANN for test data, 92% accuracy was obtained with NSE Nifty as the output variable, and 87% accuracy was obtained with BSE SENSEX as the output variable.

V. CONCLUSION AND FUTURE WORK

This paper provides a scientific way of identifying significant macroeconomic indicators influencing stock markets. The study can be applied to any global stock market. The case of Indian stock markets NSE Nifty and BSE SENSEX was taken as sample indices to demonstrate the methodology. KMO and Bartlett tests substantiated the necessity of dimensionality reduction. PCA algorithm was applied on the macroeconomic indicators, and with K1-Kaiser and scree plot, seven factors were identified. PCA with varimax ro-
FIGURE 9: Performance plot for ANN with SENSEX as target variable

TABLE 1: List of Macroeconomic Indicators

<table>
<thead>
<tr>
<th>The Macroeconomic Indicators</th>
<th>CAR SALES</th>
<th>CURRENT ACCOUNT</th>
<th>EXPORT</th>
<th>FOOD INFLATION</th>
<th>FDI</th>
<th>IMPORT</th>
<th>REMITTANCES</th>
<th>VAT</th>
<th>WHOLE SALE INDEX SALE</th>
<th>GOLD PRICES</th>
<th>MONEY SUPPLY M1</th>
<th>MONEY SUPPLY M2</th>
<th>PRODUCER PRICES</th>
<th>CENTRAL-GOV-BUDGET</th>
<th>GDP</th>
<th>GOVT SPENDING</th>
<th>GOVT BUDGET VALUE</th>
<th>SERVICE PMI</th>
<th>TOURIST ARRIVALS</th>
<th>CORPORATE TAX RATE</th>
<th>UNEMPLOYMENT RATE</th>
<th>INTEREST RATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BALANCE-OF-TRADE</td>
<td>CENTRAL-BANK-BALANCE-SHEET</td>
<td>INDUSTRIAL PRODUCTION</td>
<td>INFLATION RATE</td>
<td>INFRASTRUCTURE OUTPUT</td>
<td>MANUFACTURING PMI</td>
<td>PRIME LENDING RATE</td>
<td>TOTAL EXTERNAL DEBT</td>
<td>EXPORT PRICES</td>
<td>MONEY SUPPLY M1</td>
<td>MONEY SUPPLY M2</td>
<td>CENTRAL-BANK-BALANCE-SHEET</td>
<td>CHANGE-IN-INVENTORIES</td>
<td>CONSUMER PRICE INDEX</td>
<td>TREASURY BILL 91 DAYS YIELD</td>
<td>CURRENT-ACCOUNT-TO-GDP</td>
<td>GOVT BOND 10 YEARS YIELD</td>
<td>GROSS FIXED CAPITAL FORMATION</td>
<td>TOTAL DISPOSABLE INCOME</td>
<td>GOLD RESERVE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 2: The Correlation of Balance-of-Trade with other Macroeconomic Indicators

| BALANCE-OF-TRADE | 1 |
| CAR SALES | 0.032689557 |
| CENTRAL-BANK-BALANCE-SHEET | 0.186131577 |
| CENTRAL-GOV-BUDGET | 0.33853302 |
| CHANGE-IN-INVENTORIES | 0.454132671 |
| Competitive Index | -0.276301522 |
| Consumer Price Index | 0.183039341 |
| Corporate Tax Rate | 0.165509306 |
| CRUDE OIL PRODUCTION | 0.028258073 |
| Current Account | 0.824001414 |
| CurrentAccount-To-GDP | 0.52776772 |
| Export | -0.219090857 |
| Export Prices | 0.206546739 |
| FOOD INFLATION | -0.07100893 |
| FDI | 0.138589187 |
| Foreign Exchange Reserve | -0.08392864 |
| GDP | 0.036090317 |
| Gold Prices | -0.473007161 |
| Gold Reserve | -0.22992699 |
| GOVT SPENDING | -0.079906896 |
| Govt Bond 10 Years | -0.127055361 |
| Govt Budget Value | -0.088137462 |
| Gross Fixed Capital Formation | -0.08392864 |
| Import | -0.08392864 |
| Industrial Production | 0.18052523 |
| Inflation Rate | -0.175680471 |
| Infrastructure Output | -0.009809346 |
| Interest Rate | -0.21479898 |
| Manufacturing PMI | -0.32361362 |
| Manufacturing Production | 0.074896552 |
| Money Supply M1 | 0.909149284 |
| Money Supply M2 | 0.077493738 |
| Money Supply M3 | 0.131569445 |
| Prime Lending Rate | -0.31470875 |
| Producer Prices | 0.006039405 |
| REMITTANCES | -0.087347149 |
| VAT | 0.039028954 |
| SERVICE PMI | -0.336477072 |
| Total Disposable Personal Income | 0.106999895 |
| Total External Debt | 0.178491233 |
| Tourist Arrivals | 0.081526841 |
| Treasury Bill 91 Days Yield | -0.23544216 |
| Unemployment Rate | -0.313810984 |
| Whole Sale Price Index Sale | -0.421584221 |

TABLE 3: Result of the Bartlett Test

<table>
<thead>
<tr>
<th>Bartlett Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>13133114641070972</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The variance explained approximately 80% of the variance among the macroeconomic indicators. The ADF test proved the existence of cointegration between the exogenous and endogenous variables. The regression equation uncovered the fact that within a nation, different stock exchanges are influenced by different macroeconomic indicators differently. ANNs with ReLu and Sigmoid activation functions were employed to predict the stock market movement using the prominent macroeconomic indicators from each factor. An accuracy of 92% in case of Nifty and 87% in case of SENSEX were obtained. Even ANN results prove that macroeconomic indicators affect different stock markets differently with the same
nation. The proposed methodology can be applied to the stock market of any nation to identify the significant financial indicators influencing it. In future, this work can be extended to identify qualitative data influencing stock markets.

REFERENCES

### TABLE 5: Belongingness of prominent macroeconomic indicators to each of the factors

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
<th>Factor4</th>
<th>Factor5</th>
<th>Factor6</th>
<th>Factor7</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>0.972948</td>
<td>0.020884</td>
<td>0.051597</td>
<td>-0.09892</td>
<td>0.106961</td>
<td>0.087275</td>
<td>-0.081091</td>
</tr>
<tr>
<td>IMPORT</td>
<td>0.019677</td>
<td>0.934982</td>
<td>0.024314</td>
<td>-0.03348</td>
<td>-0.18479</td>
<td>-0.00678</td>
<td>0.227413</td>
</tr>
<tr>
<td>INDUSTRIAL PRODUCTION</td>
<td>0.077752</td>
<td>-0.01995</td>
<td>0.990015</td>
<td>0.071127</td>
<td>0.076658</td>
<td>0.032767</td>
<td>-0.02155</td>
</tr>
<tr>
<td>CORPORATE TAX RATE</td>
<td>0.017527</td>
<td>-0.12688</td>
<td>0.078537</td>
<td>0.91941</td>
<td>0.08866</td>
<td>-0.02634</td>
<td>0.024358</td>
</tr>
<tr>
<td>CURRENT ACCOUNT-TO-GDP</td>
<td>0.102613</td>
<td>-0.51561</td>
<td>0.060485</td>
<td>0.283664</td>
<td>0.873045</td>
<td>0.046151</td>
<td>0.257651</td>
</tr>
<tr>
<td>TOURIST ARRIVALS</td>
<td>0.605483</td>
<td>-0.04064</td>
<td>0.038181</td>
<td>-0.02347</td>
<td>0.043013</td>
<td>0.861472</td>
<td>-0.09876</td>
</tr>
<tr>
<td>PRIME LENDING RATE</td>
<td>-0.59731</td>
<td>0.139213</td>
<td>-0.2293</td>
<td>-0.02538</td>
<td>-0.3383</td>
<td>-0.03808</td>
<td>0.933821</td>
</tr>
</tbody>
</table>

### TABLE 6: Cumulative Variance of each Factor

<table>
<thead>
<tr>
<th>Factor</th>
<th>SS Loadings</th>
<th>Proportion Var</th>
<th>Cumulative Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor1</td>
<td>18.937795</td>
<td>0.430404</td>
<td>0.430404</td>
</tr>
<tr>
<td>Factor2</td>
<td>5.096869</td>
<td>0.115838</td>
<td>0.546242</td>
</tr>
<tr>
<td>Factor3</td>
<td>1.64201</td>
<td>0.037414</td>
<td>1.233219</td>
</tr>
<tr>
<td>Factor4</td>
<td>1.233219</td>
<td>0.033028</td>
<td>1.233219</td>
</tr>
</tbody>
</table>

### TABLE 7: The macroeconomic indicators belonging to factor-1

- CAR SALES
- CENTRAL-BANK-BALANCE-SHEET
- CENTRAL-GOVT-BUDGET
- COMPETITIVE INDEX
- CONSUMER PRICE INDEX
- EXPORT PRICES
- FDI
- FOREX
- GDP
- GOLD RESERVE
- GOVT SPENDING
- GOVT BUDGET VALUE
- GROSS FIXED CAPITAL FORMATION
- MONEY SUPPLY M2
- MONEY SUPPLY M3
- PRODUCER PRICES
- REMITTANCES
- VAT
- TOTAL DISPOSABLE INCOME
- TOTAL EXTERNAL DEBT
- UNEMPLOYMENT RATE

### TABLE 8: The macroeconomic indicators belonging to factor-2

- EXPORT
- FOOD INFLATION
- GOLD PRICES
- IMPORT
- INFLATION RATE
- MANUFACTURING PMI
- WHOLESALE PRICE INDEX

### TABLE 9: The macroeconomic indicators belonging to factor-3

- INDUSTRIAL PRODUCTION
- INFRASTRUCTURE OUTPUT
- MANUFACTURING PRODUCTION
- SERVICE PMI

### TABLE 10: The macroeconomic indicators belonging to factor-4

- CORPORATE TAX RATE
- GOVT BOND 10 YEARS

### TABLE 11: The macroeconomic indicators belonging to factor-5

- BALANCE-OF-TRADE
- CURRENT ACCOUNT
- CURRENT ACCOUNT-TO-GDP

### TABLE 12: The macroeconomic indicators belonging to factor-6

- TOURIST ARRIVALS

### TABLE 13: The macroeconomic indicators belonging to factor-7

- CHANGE-IN-INVENTORIES
- CRUDE OIL PRODUCTION
- INTEREST RATE
- PRIME LENDING RATE
- TREASURY BILL 91 DAYS YIELD

### TABLE 14: The prominent macroeconomic indicator belonging to each category

<table>
<thead>
<tr>
<th>FACTOR NUMBER</th>
<th>FACTOR NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACTOR 1</td>
<td>GDP</td>
</tr>
<tr>
<td>FACTOR 2</td>
<td>IMPORT</td>
</tr>
<tr>
<td>FACTOR 3</td>
<td>INDUSTRIAL PRODUCTION</td>
</tr>
<tr>
<td>FACTOR 4</td>
<td>CORPORATE TAX RATE</td>
</tr>
<tr>
<td>FACTOR 5</td>
<td>CURRENT ACCOUNT TO GDP</td>
</tr>
<tr>
<td>FACTOR 6</td>
<td>TOURIST ARRIVALS</td>
</tr>
<tr>
<td>FACTOR 7</td>
<td>PRIME LENDING RATE</td>
</tr>
</tbody>
</table>

### TABLE 15: The regression statistics for NSE NIFTY index

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.961408563</td>
<td></td>
</tr>
<tr>
<td>R Square</td>
<td>0.924306425</td>
<td></td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.918285345</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>549.7053518</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>96</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 16: The regression statistics for BSE SENSEX index

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.961883412</td>
</tr>
<tr>
<td>R Square</td>
<td>0.9251219697</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.919271264</td>
</tr>
<tr>
<td>Standard Error</td>
<td>1754.668984</td>
</tr>
<tr>
<td>Observations</td>
<td>96</td>
</tr>
</tbody>
</table>

TABLE 17: The ANOVA test results for NSE Nifty index

<table>
<thead>
<tr>
<th>ANOVA test results</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>SS</td>
</tr>
<tr>
<td>Regression 7</td>
<td>324712911.8</td>
</tr>
<tr>
<td>Residual 88</td>
<td>26591485.7</td>
</tr>
<tr>
<td>Total 95</td>
<td>351304397.5</td>
</tr>
</tbody>
</table>

TABLE 18: The ANOVA test results for BSE SENSEX index

<table>
<thead>
<tr>
<th>ANOVA test results</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>SS</td>
</tr>
<tr>
<td>Regression 7</td>
<td>3352206189</td>
</tr>
<tr>
<td>Residual 88</td>
<td>270939965.5</td>
</tr>
<tr>
<td>Total 95</td>
<td>3623146154</td>
</tr>
</tbody>
</table>


PRAKASH K. AITHAL is an Assistant Professor in the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal India. His current research includes Data Analytics in Financial domain. He has presented several papers in national and international conferences, and his work has been published in several international journals.

DINESH ACHARYA U. is a Professor in Dept. of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal India. He obtained his Ph.D. from Manipal Academy of Higher Education in 2008. His current research includes Data Analytics in Healthcare, Agriculture, and Financial sectors. He has presented several papers in national and international conferences, and his work has been published in several international journals.

GEETHA M. is a Professor in Dept. of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal India. She obtained her Ph.D. from NITK, Surathkal in 2010. Her current research includes Data Mining, Text Mining in Healthcare and Financial sectors. She has presented several papers in national and international conferences, and her work has been published in several international journals.