IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 7, July 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207139 |

Comparative Study of Stock Market Prediction Using Random Forest, Decision Tree and Linear Regression

Afrina Beevi M[1], Babu Renga Rajan S [2]

[1]PG Student, Department of Computer Science and Engineering, PET Engineering College, Tirunelveli, India

[2]Head of the Department, Department of Computer Science and Engineering, PET Engineering College, Tirunelveli, India

ABSTRACT- Stock market is an volatile place for predicting since there are no significant rules to estimate or predict the price of a share in the share market. Methods like, fundamental analysis, time series analysis are used to predict the price in the share market but none of these methods are proved as a consistently acceptable prediction tool. Here random forest ,decision tree,and linear regression are implemented to predict stock market prices. The aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. In this paper we aims at proving the efficiency of Random forest,decision tree and linear regression for predicting the stock prices.

KEYWORDS: Random Forest,linear regression,decision tree,stock market prediction,time series analysis,fundamental analysis.

I. INTRODUCTION

Stock market prediction has been an area of interest for investors as well as researchers for many years due to its volatile, complex and regular changing nature, making it difficult for reliable predictions. Stock market prediction is the act of trying to determine the future value of company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield a significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable[1]. Others disagree and those with this view point possess several methods and technologies which supposedly allow them to gain future price information. Predicting how the stock market will perform is one of the most difficult things to do. Intrinsic volatility in the stock market across the globe makes the task of prediction challenging. There are so many factors involved in the prediction – physical factors vs. technical, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy. Using features like the latest announcements about an organization, their quarterly revenue results, etc., machine learning techniques have the potential to unearth patterns and insights we didn‟t see before, and these can be used to make unerringly accurate predictions. Stock Price Prediction Due to the high profit of the stock market, it is one of the most popular investments. People investigated for methods and tools that would increase their gains while minimizing the risk, as the level of trading and investing grew. Two stock exchanges namely- the National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE), which are the most of the trading in Indian Stock Market takes, place. Sensex and Nifty are the two prominent Indian Market Indexes. Since the prices in the stock market are dynamic, the stock market prediction is complicated. From gradually the very past years some forecasting models are developed for this kind of purpose and they had been applied to money market prediction. Generally, this classification is done by Time series analysis and Fundamental analysis[2].

Time series analysis

The definition of forecasting can be like this the valuation of some upcoming result or results by analyzing the past data. It extents different areas like industry and business, economics and finance, environmental science. Forecasting problems can be classified as follows:

Long term forecasting (estimation beyond 2 years)

Medium-term forecasting (estimation for 1 to 2 years)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207139 |

Short term forecasting (estimation for weeks or months, days, minutes, few seconds)
The analysis of time consist of several forecasting problems.The designation of a time series is a linear classification of observations for a selected variable. The variable of the stock price in our case. Which can weather multivariate or univariate[3]? Only particular stock is included in the univariate data while more than one company for various instances of time is added in multivariate. For investigating trends, patterns and cycle or periods the analysis of time series advantages in the present data. In spending money wisely an early data of the bullish or bearish in the case of the stock market. Also, for categorizing the best-performing companies the analysis of patterns plays its role for a specific period. This makes forecasting as well as time series analysis an important research area.

Fundamental analysis

Fundamental Analysts are concerned with the business that reasons the stock itself. They assess a company's historical performance as well as the reliability of its accounts. Different performance shares are created that aid the fundamental forecaster with calculating the validity of a stock, such as the P/E ratio. Warren Buffett is probably the foremost renowned of all Fundamental Analysts. What fundamental analysis within the stock market is making an attempt to reach, is organizing the true value of a stock, that then will be matched with the worth it is being listed on stock markets and so finding out whether or not the stock on the market is undervalued or not. Find out the correct value will be completed by numerous strategies with primarily a similar principle. The principle is that an organization is price all of its future profits [4]. Those future profits has to be discounted to their current value. This principle goes on the theory that a business is all about profits and nothing else. Differing to technical analysis, the fundamental analysis is assumed as further as a long approach. Fundamental analysis is created on conviction that hominoid society desires capital to make progress and if the company works well, than it should be rewarded with an additional capital and outcome in a surge in stock price. Fundamental analysis is usually used by the fund managers as it is the maximum sensible, objective and prepared from openly existing data like financial statement analysis. One more meaning of fundamental analysis is on the far side bottom-up business analysis, it discusses the top-down analysis since initial analysing the world economy, followed by country analysis and also sector analysis, and last the company level analysis.

II. LITERATURE SURVEY

Publication Year: 2018 Author: Jai Jagwani, Hardik Sachdeva, Manav Gupta, Alka Singhal Journal Name: 2018 IEEE Summary: To identify the relationship between different existing time series algorithms namely ARIMA and Holt Winter and the stock prices is the main objective of the proposed work, for the investments a good risk-free range of stock prices are analyzed and therefore better accuracy of the model can be seen. To find distinguished results for shares in the stock market, the combination of two different time series analysis models is opted by producing a range of prices to the consumer of the stocks. Not complex in nature and estimation of values which are purely based on the past stock prices for non-seasonal or seasonal is the main advantage of these models. In this experiment, some limitations are, the work that never takes into consideration and other circumstances like news about any new market strategy or media release relevant to any company which may get affected by the prices of stocks.

Stock Market Prediction Using Machine Learning: Publication Year: 2018 Author: Ishita Parmar, Ridam Arora, Lokesh Chouhan, Navanshu Agarwal, Shikhin Gupta, Sheirsh Saxena, Himanshu Dhiman Journal Name: 2018 IEEE Summary:In this paper studies, the use of Regression and LSTM based Machine learning to forecast stock prices. Factors measured are open, close, low, high and volume [5]. This paper was an attempt to determine the future prices of the stocks of a company with improved accuracy and reliability using machine learning techniques. LSTM algorithm resulted in a positive outcome with more accuracy in predicting stock prices.

Multi-Category Events Driven Stock Price Trends Prediction: Publication Year: 2018 Author: Youxun Lei, Kaiyue Zhou, Yuchen Liu Journal Name: 2018 IEEE Summary: In this paper, multi-category news events are used as features to develop stock price trend prediction, model. The multi-category events are based on already defined feature word dictionary. And we have employed both neural networks and SVM models to analyse the relationship between stock price movements and specific multi-category news. Experimental results showed that the predefined multi-category news events are more improved than the baseline bag-of-words feature to predict stock price trend. As compared to long term prediction, short term prediction is better based on this study[6].

AUTHOR:[1] S. A. R. Nai-Fu Chen and Richard Roll, TITLE:Economic Forces and the Stock Market, YEAR:1986.This paper tests whether innovations in macroeconomic variables are risks that are rewarded in the stock market. Financial theory suggests that the following macroeconomic variables should systematically affect stock market returns: the spread between long and short interest rates, expected and unexpected inflation, industrial

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207139 |

production, and the spread between high- and low-grade bonds. We find that these sources of risk are significantly priced. Furthermore, neither the market portfolio nor aggregate consumption are priced separately. We also find that oil price risk is not separately rewarded in the stock market.

AUTHOR: E. F. Fama, Random Walks in Stock Market Prices, TITLE:Financial Analysts Journal,vol.51,no.1,pp.75–80,YEAR:(1995).[Online].Available:http://www.jstor.org/stable/4479810.MANY YEARS cconomists, Statisticians, and teachers of finance have been interested in developing and testing models of stock price behavior. One important model that has evolved from this research is the theory of random walks. This theory casts serious doubt on many other methods for describing and predicting stock price behavior — methods that have considerable popularity outside the academic world. For example, we shall see later that if the random walk theory is an accurate description of reality, then the various "technical" or "chartist" procedures for predicting stock prices are completely without value. In general the theory of random walks raises challenging questions for anyone who has more than a passing interest in understanding the behavior of stock prices. Unfortunately, however, most discussions of the theory have appeared in technical academic journals and in a form which the non-mathematician would usually find incomprehensible. This article describes, briefly and simply, the theory of random walks and some of the important issues it raises concerning the work of market analysts. To preserve brevity some aspects of the theory and its implications are omitted. More complete (and also more technical) discussions of the theory of random walks are available elsewhere; hopefully the introduction provided here will encourage the reader to examine one of the more rigorous and lengthy works listed at the end of this article[7].

AUTHOR: S. J. Grossman and R. J. Shiller, TITLER:The Determinants of the Variability of Stock Market Prices, National Bureau of Economic Research, Working Paper 564, October YEAR:(1980)[Online].Available:http://www.nber.org/papers/w0564.The most familiar interpretation for the large and unpredictable swings that characterize common stock price indices is that price changes represent the efficient discounting of "new information" It is remarkable given the popularity of this interpretation that it has never been established what this information is about. Recent work by Shiller, and Stephen LeRoy and Richard Porter, has shown evidence that the variability of stock price indices cannot be accounted for by information regarding future dividends since dividends just do not seem to vary enough to justify the price movement. These studies assume a constant discount factor. In this paper, we consider whether the variability of stock prices can be attributed to information regarding discount factors (i.e., real interest rates), which are in turn related to current and future levels of economic activity. The appropriate discount factor to be applied to dividends which are received k years from today is the marginal rate of substitution between consumption today and consumption k periods from today, We use historical data on per capita consumption from 1890-1979 to estimate the realized value of these marginal rates of substitution. Theoretically, as LeRoy and C. J. La Civita have also noted independently of us, consumption variability may induce stock price variability whose magnitude depends on the degree of risk aversion.

AUTHOR: A. W. Lo and A. C. MacKinlay, TITLE:Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test, Review of Financial Studies, vol. 1, no. 1, pp. 41–66, YEAR:(1988). In this paper, we test the random walk hypothesis for weekly stock market returns by comparing variance estimators derived from data sampled at different frequencies. The random walk model is strongly rejected for the entire sample period (1962-1985) and for all sub-periods for a variety of aggregate returns indexes and size-sorted portfolios. Although the rejections are largely due to the behavior of small stocks, they cannot be ascribed to either the effects of infrequent trading or time-varying volatilities. Moreover, the rejection of the random walk cannot be interpreted as supporting a mean-reverting stationary model of asset prices, but is more consistent with a specific nonstationary alternative hypothesis.

AUTHOR: P. P¨a¨akk ¨onen and D. Pakkala, TITLE:Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems,Big Data Research, vol. 2, no. 4, pp.166–186,YEAR:(2015).[Online]. Many business cases exploiting big data have been realized in recent years; Twitter, LinkedIn, and Facebook are examples of companies in the social networking domain. Other big data use cases have focused on capturing of value from streaming of movies (Netflix), monitoring of network traffic, or improvement of processes in the manufacturing industry. Also, implementation architectures of the use cases have been published. However, conceptual work integrating the approaches into one coherent reference architecture has been limited.The contribution of this paper is technology independent reference architecture for big data systems, which is based on analysis of published implementation architectures of big data use cases. An additional contribution is classification of related implementation technologies and products/services, which is based on analysis of the published use cases and survey of related work.The reference architecture and associated classification are aimed for facilitating architecture design and selection of technologies or commercial solutions, when constructing big data systems.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207139 |

III. EXISTING SYSTEM

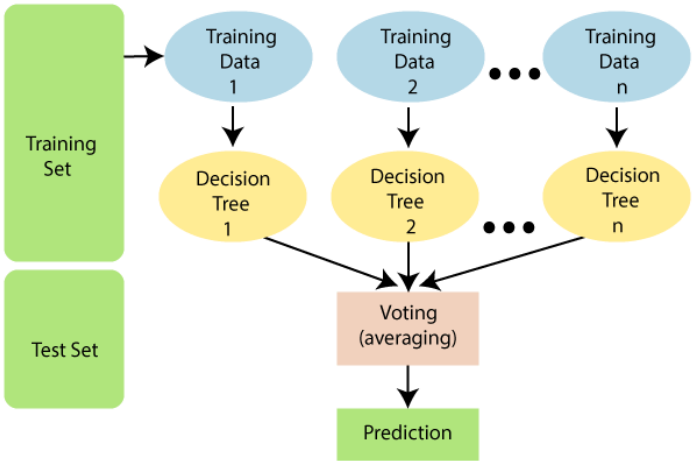
• The existing system fails when there are rare outcomes or predictors, as the algorithm is based on bootstrap sampling.
• The previous results indicate that the stock price is unpredictable when the traditional classifier is used.
• The existence system reported highly predictive values, by selecting an appropriate time period for their experiment to obtain highly predictive scores.
• The existing system does not perform well when there is a change in the operating environment.
• It doesn't focus on external events in the environment, like news events or social media[8].
• The existing system needs some form of input interpretation, thus need of scaling.
• It doesn't exploit data pre-processing techniques to remove inconsistency and incompleteness of the data.

IV.PROPOSED SYSTEM

In this proposed system, we focus on predicting the stock values using machine learning algorithms like Random Forest ,linear regression, decision table. We proposed the system "Stock market price prediction" we have predicted the stock market price using the random forest algorithm. In this proposed system, we were able to train the machine from the various data points from the past to make a future prediction. We took data from the historical year stocks to train the model. We majorly used two machine-learning libraries to solve the problem. The first one was numpy, which was used to clean and manipulate the data, and getting it into a form ready for analysis[8]. The other was scikit, which was used for real analysis and prediction. The data set we used was from the previous years stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data. The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data. We used the python pandas library for data processing which combined different datasets into a data frame. The tuned up data frame allowed us to prepare the data for feature extraction[10]. The data frame features were date and the closing price for a particular day. We used all these features to train the machine on random forest model and predicted the object variable, which is the price for a given day. We also quantified the accuracy by using the predictions for the test set and the actual values. The proposed system touches different areas of research including data pre-processing, random forest, and so on.

RANDOM FOREST

"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.



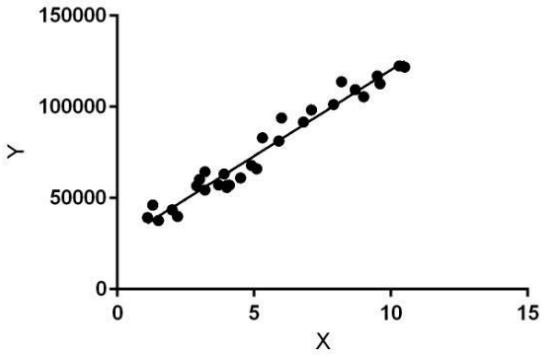
LINEAR REGRESSION

Linear regression is a type of supervised machine learning algorithm that computes the linear relationship between a dependent variable and one or more independent features. When the number of the independent feature, is 1 then it is known as Univariate Linear regression, and in the case of more than one feature, it is known as multivariate linear regression. The goal of the algorithm is to find the best linear equation that can predict the value of the dependent variable based on the independent variables. The equation provides a straight line that represents the relationship between the dependent and independent variables.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

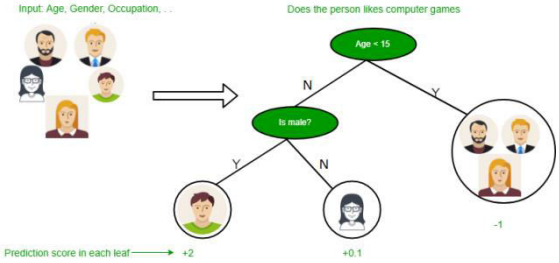
| DOI:10.15680/IJIRSET.2023.1207139 |



Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x)). Hence, the name is Linear Regression. In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best-fit line for our model.

DECISION TREE

A decision tree is a type of supervised learning algorithm that is commonly used in machine learning to model and predict outcomes based on input data.



V.SYSTEM ARCHITECTURE

In this sub ,pictorial representation of system architecture is shown in fig.1

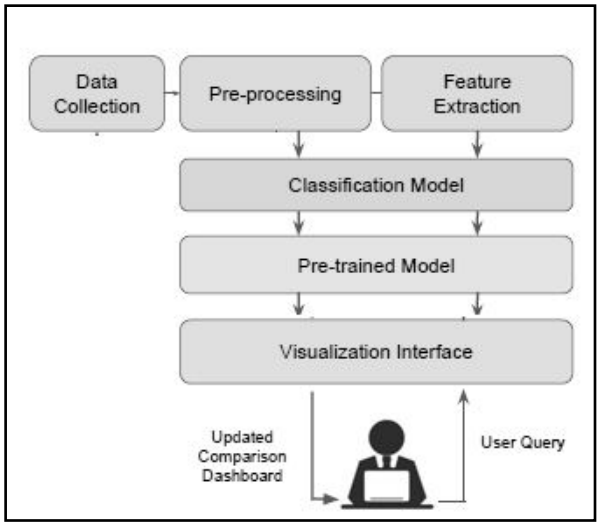


Fig :1 system architecture

VI.MODULES

- Data Collection
- Data Preparation
- Model Selection

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207139 |

- ❖ Analyze and Prediction
- ❖ Accuracy on test set
- ❖ Saving the Trained Model

MODULES DESCSRIPTION:

Data Collection:

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.

There are several techniques to collect the data, like web scraping, manual interventions and etc.

Data Preparation:

We will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain .Next we drop or remove all columns except for the columns that we want to retain. Finally we drop or remove the rows that have missing values from the data set.

Model Selection:

We used Random forest algorithm but in base paper neural network algorithm is given, The Passive-Aggressive algorithms are a family of Machine learning algorithms that are not very well known by beginners and even intermediate Machine Learning enthusiasts. However, they can be very useful and efficient for certain applications so we applied[11].

Analyze and Prediction:

In the actual dataset, we chose only 2features

1. Text: the text of the article; could be incomplete
2. Label: a label that marks the article as potentially unreliable

Accuracy on test set:

We got a accuracy of 70.2% on test set.

Saving the Trained Model:

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle .Make sure you have pickle installed in your environment. .Next, let's import the module and dump the model into .pkl file

VII.RESULTS

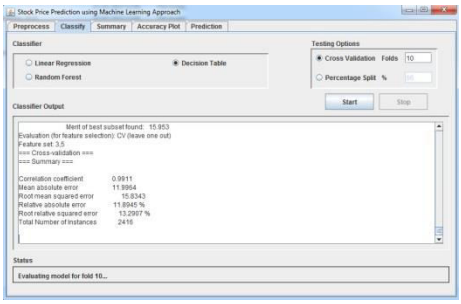


Fig 2.Screenshot shows preprocessor method

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207139 |

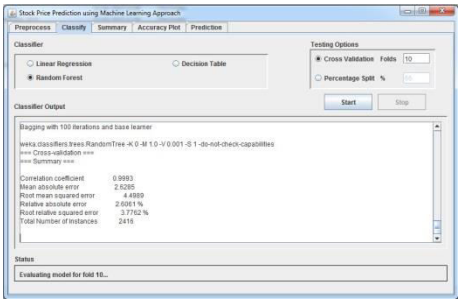


Fig 3. Screenshot show the accurancy plot method

VIII.CONCLUSION AND FUTURE ENCLUSION

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. With the ability to predict the stock prices or the trends alone, one is now able to make a very huge profit by investing in the stock market.

REFERENCES

[1] “Apache Spark - Lightning-Fast Unified Analytics Engine,” 2020. Available: https://spark.apache.org/
[2] V. Atanasov, C. Pirinsky, and Q. Wang, “The efficient market hypothesis and investor behavior,” 2018.
[3] S. Agrawal, D. Thakkar, D. Soni, K. Bhimani, and C. Patel, “Stock market prediction using machine learning techniques,” International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2019.
[4] M. Afeef, A. Ihsan, and H. Zada, “Forecasting stock prices through univariate arima modeling,” 2018.
[5] P.-F. Pai and C.-S. Lin, “A hybrid arima and support vector machines model in stock price forecasting,” Omega, vol. 33, no.
6, pp. 497–505, 2018. 6. Tae Kyun Lee et al, ”Global Stock Market Investment Strategies Based On Financial Network Indicators Using Machine Learning Techniques”, Volume: 117, Issue: 1, Date: 2019.
7. Kang Zhang et al, ”Stock Market Prediction Based on Generative Adversarial Network”, Volume:147, Issue:2, Date:2019.
8. Yi-Fan Wang,” On-Demand Forecasting of Stock Prices Using a Real-Time Predictor”, Volume: 15, Issue: 4, Date: 2003.
9. Pei-Yuan Zhou, et al, “Corporate Communication Network and Stock Price Movements: Insights from Data Mining”, Volume:5, Issue: 2, Date: 2018.
10. Pei-Chann Chang et al,” Integrating a Piecewise Linear Representation Method and a Neural Network Model for Stock Trading Points Prediction”, Volume: 39, Issue:1, Date:2009.
11. Min-Wen et al,” Stock Market Trend Prediction Using High-Order Information of Time Series”, Volume: 7, Issue: 4,Date:2019.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details