SOME IMPROVED FORECASTING MODELS AND THEIR APPLICATIONS IN DIFFERENT FIELDS

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

MATHEMATICS

By

ARUN KUMAR GAROV

Registration Number: 11919171

Supervised By

Prof. (Dr.) A.K. Awasthi (25155)

Professor, Department of Mathematics

School of Chemical Engineering and Physical Sciences

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2023

DECLARATION

I, hereby declared that the presented work in the thesis entitled **"Some Improved Forecasting Models and Their Applications in Different Fields"** in fulfilment of degree of **Doctor of Philosophy (Ph. D.)** is outcome of research work carried out by me under the supervision **Dr. A.K. Awasthi**, working as **Professor**, in the **Department of Mathematics, School of Chemical Engineering and Physical Sciences** of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: **Arun Kumar Garov**

Registration No.: **11919171**

Department/school: **Department of Mathematics, School of Chemical Engineering and Physical Sciences,**

Lovely Professional University,

Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph. D. thesis entitled **"Some Improved Forecasting Models and Their Applications in Different Fields"** submitted in fulfillment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the **Department of Mathematics, School of Chemical Engineering and Physical Sciences** is a research work carried out by **Arun Kumar Garov, 11919171**, is bonafide record of his original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor)

Name of supervisor: **Prof. (Dr.) A.K. Awasthi**

Designation: **Professor**

Department/school: **Department of Mathematics, School of Chemical Engineering and Physical Sciences**

University: **Lovely Professional University, Punjab, India**

Abstract

Forecasting models are important tools for predicting future events and guiding decisions. In recent years, the development of new forecasting models has allowed for greater accuracy and precision in forecasting for a variety of applications. This thesis explores the potential of improved forecasting models and their applications in different fields with technical and non-technical models. Technical and non-technical models combined use to improve forecasting. Specifically, the thesis focuses on improving forecasting models for application in finance, agriculture, and healthcare. Time series models such as the Autoregressive Integrated Moving Average (ARIMA) model, the thesis seeks to explore more advanced machine learning models and their applications to forecasting. The thesis described the forecasting description as stages, limitations, advantages, classification, and application, followed by a discussion of their respective strengths and weaknesses. And work reviews the existing literature in the fields of time series forecasting with several methods. This examines existing applications of models in various fields, such as finance, economics, and agriculture, healthcare forecasting.

Time series data consider for forecasting as TAIEX, Dow Jones, NASDAQ, and NATCO PHARMA indexes of the stock and production of crops and price of the crops are considered for forecasting. For forecasting, time series used methods and applications on a different data set. In Quantity-based time series forecasting (QBFTS), the main role of a quantity of the considered data set. For that used fuzzy time series model, uses a fuzzy set, fuzzy logical relationship of the data set, fuzzy logical relationships (FLR), and fuzzy logical relationship groups (FLRGs) with statistical weighted system and training and testing processes used for forecasting that is a combination of technical and non-technical models of forecasting.

In addition, the thesis explores the potential of new machine-learning techniques and their role in improving forecasting models. Some basic forecasting models such as Naïve, moving average, auto-regression, auto-regressive integrated moving average model, etc., are discussed and some of them are applied to different types of time series data set by machine learning. For the analysis of the forecasting used some software's as python 3 in Zupyter notebook, SPSS, etc., for analysis, and a Hybrid model as RF-DT for the forecasting of the covid-19 data. Hybridization of two models as fuzzy and ARIMA used to forecast crop price time series.

ACKNOWLEDGMENT

At the outset, I'd like to express my sincere gratitude to the almighty God and my supervisor, ***Prof. (Dr.) A.K. Awasthi****, Professor, School of Chemical Engineering and Physical Sciences, Lovely Professional University, Punjab, for his interest, excellent rapport, untiring cooperation, invaluable advice, and encouragement throughout my Ph.D. Degree. Without his unfailing support and belief in me, this thesis would not have been possible. I indeed feel short of words to express my sense of gratitude towards him. Besides being a scholar par eminence, he is a person par excellence a perfect embodiment of dedication, intelligence, and humanity.*

I acknowledge with pleasure the cooperation from the esteemed faculty of the University for their enthusiastic support and encouragement throughout the research work.

I thank all my friends and research scholars for their help in this duration of research.

I feel very much obliged to My Supervisor, Parents, and Fiancée for what I have received from them in the form of motivation, encouragement, and moral support. Without them, it would have been impossible to carry out this task. Their constant motivation kept me going even in the most challenging times.

Arun Kumar Garov

Dedicated

to

Dream of My Father

&

My God

&

My Parents, My Wife,

and Friends,…!

Content

List of Tables

List of Figures

CHAPTER

I

Introduction

1.1 GENERAL INTRODUCTION

Forecasting is a current topic of growing important in business and economic analysis. It is an effort to predict the future by looking back at the past. Based on information and experience from the past and present, it entails producing objective estimations of the outcome of some variable in the future. Generally, a forecast may be considered as a statement about future. There are many ways of making forecasts. In economics forecasting methods include: (i) expert judgment (ii)guessing, rule of thumb or informal models (iii) extrapolation (iv) leading indicators (v) surveys (vi) economic systems and (vii) time series models. It is difficult to judge the accuracy of any forecast without knowing the circumstances in which it was made.

Basically, forecasting process involves the following components: (i) Determining the objective (ii) Developing a model (iii) Testing the model (iv) Applying the model (v) Evaluating and revising the model.

Since three decades a considerable amount of research has been conducted in the field of forecasting. Attempts have been made to make forecasting as scientific as possible. Statistics is the base of scientific forecasting. The problem of forecasting is essentially statistical in nature. There are mainly two aspects of scientific forecasting. The first is the analysis of past conditions and the second is the analysis of current conditions in relation to a prospective future trend. Scientific forecasting requires detailed information about past movements and special factors affecting the movement.

In statistics, forecasting refers to extending or projecting time series data into the future to predict future outcomes. Time series data consists of observations collected sequentially over time, such as monthly sales or daily temperature readings. Forecasting methods have been developed by statisticians and economists in order to make predictions about future values of a given variable.

The three terms prediction, projection, and forecast can be distinguished separately.

Prediction: A prediction is a statement or claim that a certain event will happen in the future. It is based on knowledge, experience, or instinct about what might happen.

Projection: A projection is an educated guess of what might happen in the future based on current trends and data. It is a calculated estimation of what could happen in the future based on the present.

Forecast: A forecast is an estimate of what will happen in the future based on past data and trends. It is a prediction based on existing evidence and can be used to plan for the future.

With the development of more sophisticated forecasting techniques, along with the advent of computers, forecasting has received more and more attention. Forecasts are needed in finance, marketing, business, economics, production, government, and non-governmental organizations, etc. Forecasts are used to predict future trends and outcomes, to help inform decisions and strategies, and to measure the impact of certain actions and events. In finance, forecasts are used to predict future stock prices and other market indicators, as well as to inform investment decisions. In marketing, forecasts are used to predict customer demand, to inform pricing and promotional strategies, and to measure the impact of marketing campaigns. In business, forecasts are used to predict future sales and profits, and to inform operational and strategic decisions. In economics, forecasts are used to predict future economic indicators, such as GDP and inflation, and to inform economic policy decisions. In production, forecasts are used to plan future production levels, determine inventory levels, and determine the capacity needed to meet demand. By using forecasts in this way, companies can better manage their resources and maximize their efficiency. In government and non-governmental organizations, forecasts are used to create policies and programs that are designed to address the needs of the population. Forecasts can also be used to assess the effectiveness of these policies and programs.

Finally, in private business, forecasts are used to make long-term decisions, such as deciding what products or services to offer and when. Forecasts can also be used to set pricing and marketing strategies, plan for changes in the industry, and assess the competitive landscape. The main point to remember is that a forecast must result in present action to improve the future.

Forecasting is not for only one sector and one situation, it for many fields at different situation. Some fields are given below where can used forecasting;

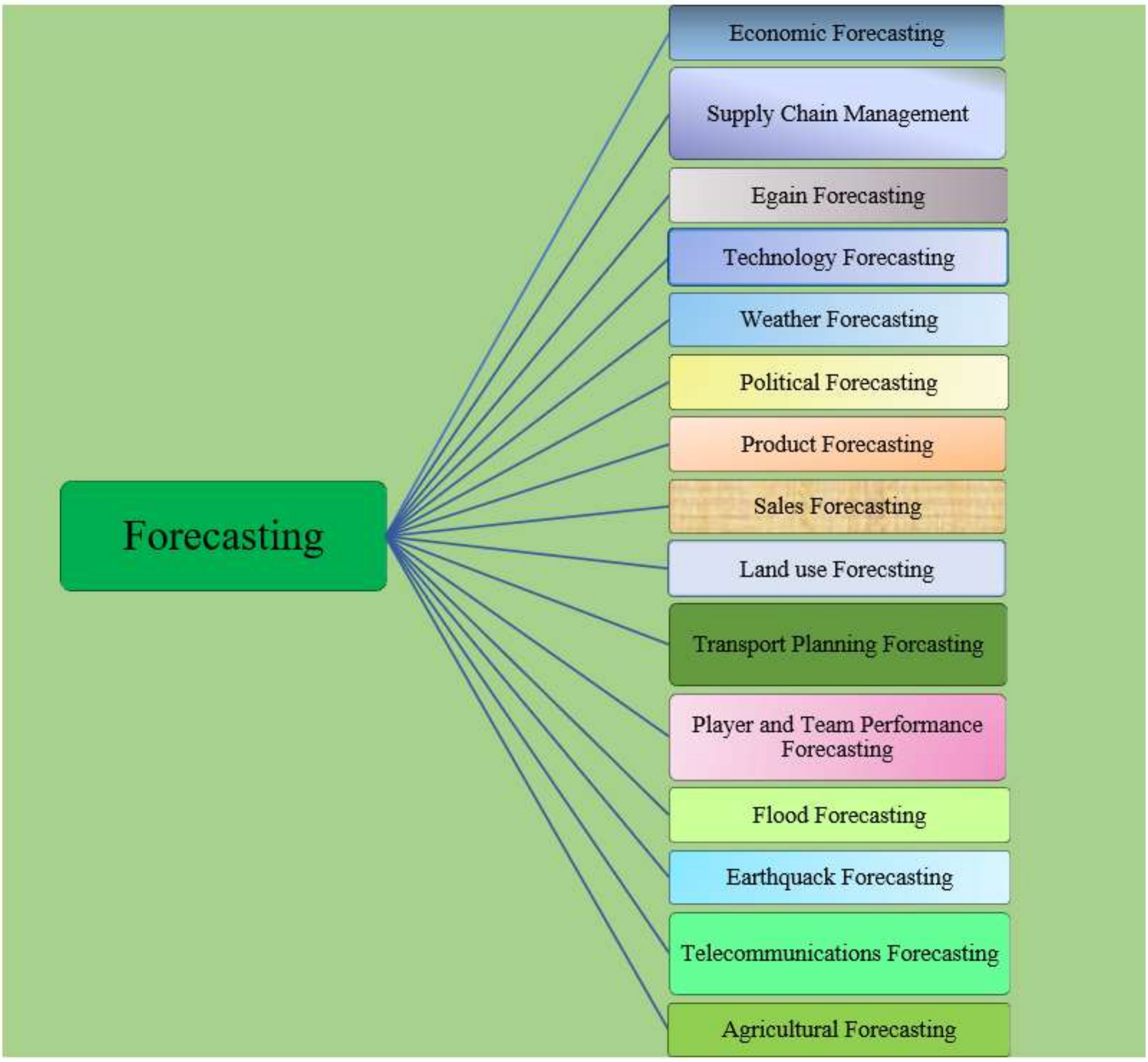


FIGURE 1.1. Forecasting fields

Now understand the different stages for predictive analysis and forecasting.

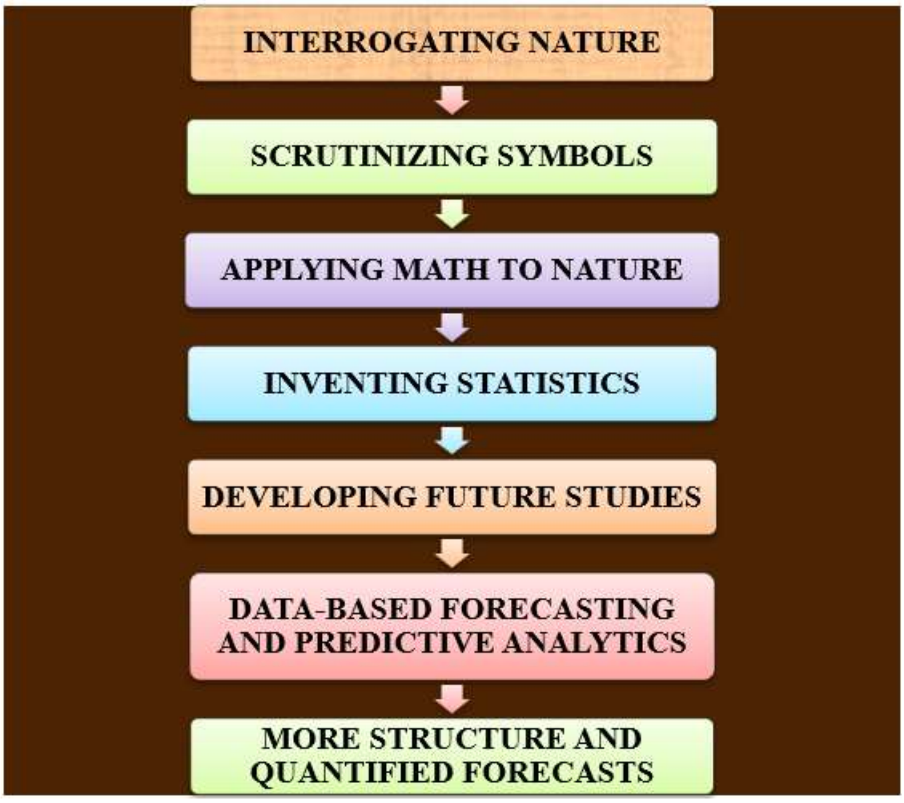


FIGURE 1.2. Stages of Predictive Analysis and Forecasting

As accordance above stages, we can link all stages and primary era by the following way,

FIGURE 1.3. Classification of Primary Era

Relationship of Primary Era Types with Forecasting Stages;

Stages No.	Stages	Primary Era
1.	INTERROGATING NATURE	Prehistory to Classical
2.	SCRUTINIZINGSYMBOLS	Classical to Medieval
3.	APPLYING MATH TO NATURE	Revitalization
4.	INVENTING STATISTICS	Later 18th to Early 20th Century
5.	DEVELOPING FUTURE STUDIES	Post World War II
6.	DATA BASED FORECASTING AND PREDICTIVE ANALYTICS	20th - 21st Century
7.	QUANTIFIED FORECASTING AND MORE STRUCTURE	Early 21st Century

FIGURE 1.4. Relationship between Primary Era and Forecasting Stages

1.2 STATEMENT OF THE PRESENT RESEARCH PROBLEM

There have been many developments in estimation and forecasting over the past decades that have direct relevance and applicability to organizational forecasting. The forecasting literature is just now starting to focus on translating the theoretically possible and computationally feasible into a form that can be easily understood and applied.

Several forecasting techniques may be divided into two major categories, namely quantitative and qualitative or technological methods.

(i). Qualitative Analysis

Qualitative analysis is a method of forecasting that uses subjective judgment based on intangible factors such as consumer trends, market sentiment, and competitive positioning. It is used to forecast consumer demand for a product or service by analyzing data such as customer surveys, focus groups, and interviews. Qualitative analysis is often used to make decisions about product design, pricing, and marketing strategies.

Qualitative techniques can be divided into series and causal methods. Series methods are used to describe and analyze the characteristics of a population over a period of time. They involve the collection and analysis of data in order to identify patterns and trends. Causal methods are used to identify the cause-and-effect relationships between two or more variables. They involve the use of mathematical models and statistical methods to identify relationships between variables.

(ii). Quantitative Analysis

Quantitative analysis is a method of forecasting that uses mathematical models to predict future outcomes. It is used to forecast consumer demand for a product or service by analyzing data such as sales trends, financial data, and market trends. Quantitative analysis is often used to make decisions about product development, production planning, and pricing strategies.

Quantitative forecasting can apply when data points are available to generate a model of the future trend. This type of forecasting is often used in financial markets and for demand planning. It involves looking at historical data and using mathematical techniques such as time series analysis, regression analysis, and other analytical techniques to predict future trends.

It can be divided into explanatory and normative methods. Explanatory methods involve the use of interviews, surveys, focus groups, and other techniques to gain a better understanding of a particular topic or issue. These techniques are used to uncover the motivations and experiences of participants in order to gain insight into the underlying causes of their behavior. Normative methods involve the analysis of data and the identification of potential solutions to a problem. This can involve the development of models, simulations, and other tools to test the effectiveness of potential solutions. It can also involve the use of qualitative data to inform decision making.

- **Qualitative Analysis and Quantitative Analysis for forecasting**

Qualitative analysis is a method of forecasting that is based on judgment and opinion, rather than on hard data and statistics. It involves the use of subjective information such as customer surveys, interviews, and industry trends to predict future outcomes. By contrast, quantitative analysis uses hard data and statistics to predict future outcomes. It relies on mathematical models and algorithms to analyze data sets and make predictions. Both qualitative and quantitative analysis can be used to forecast future trends, but qualitative analysis is often used to supplement quantitative analysis and to provide a more complete picture of the situation.

In causal forecasting, the goal is to understand the cause-effect relationships between different events and processes and make predictions about their future behavior. It is based on the idea that the past can be used to predict the future, and that the effects of different factors can be identified and used to make more accurate forecasts. Causal forecasting involves using statistical models and machine learning algorithms to identify relationships between different factors and to make predictions about future behavior. It is a powerful tool for businesses to improve their decision-making and to make more informed decisions.

In time series analysis forecasting, a statistical analysis on past demands is used to generate the forecasts. This involves using historical data, such as sales figures, and analyzing it to identify patterns and trends. Statistical methods such as linear regression, exponential smoothing, and time series decomposition can be used to generate forecasts. These forecasts can then be used to plan for future demand, such as inventory levels. Forecasts can also be used to inform decisions around pricing, marketing, and strategy.

Econometric models for forecasting use a variety of techniques to estimate the values of the parameters in the equations. These models are systems of relationships between the variables under investigation. Their equations are then estimated from the available data, mainly aggregate time series. A forecast made from an econometric model may not be exact, because the estimates of the parameters are subject to sampling errors. The forecast error, or the deviation of the forecast from the actual value, can be minimized by using more data, more equations, and more sophisticated estimation techniques.

1.2.1 Classification of forecasting methods

Forecasting methods which are classified according to various criteria like Time Horizon, form of data used, type of data used, etc. can see in the literature. Described here some of the important classifications of forecasting methods;

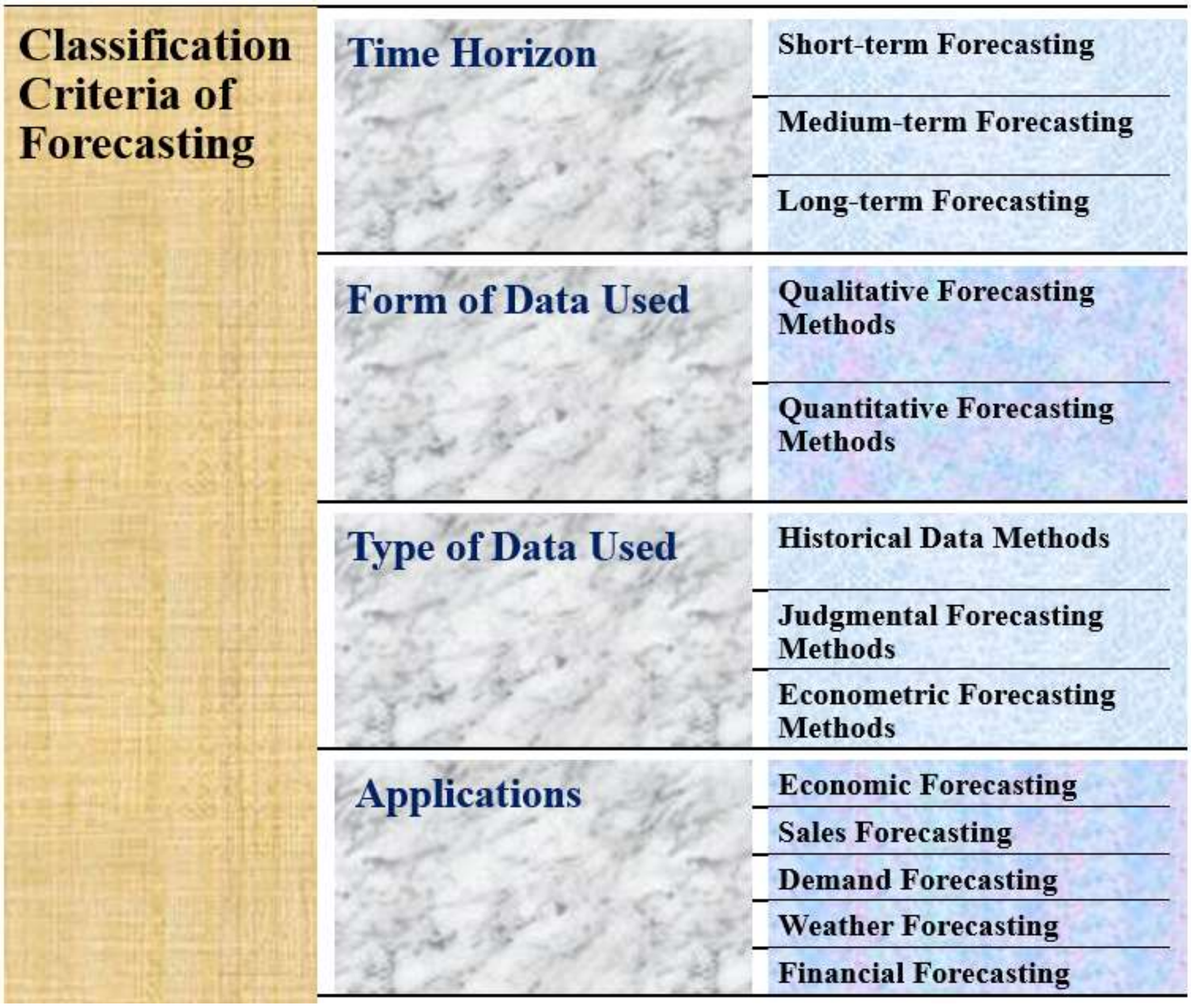


FIGURE 1.5. Forecasting Classification Criteria

1.2.1.1 Classification according to Time Horizon

Forecasting methods can be broadly classified according to various criteria such as the time horizon of the forecast. Forecasts can be classified as short-term, medium-term, and long-term, depending on the period being forecasted.

i) Short-term forecasting method:

Short-term forecasting methods typically involve using statistical techniques to project future outcomes based on existing data and trends. These methods often involve using linear regression, time series analysis, and other machine learning techniques to analyze historical data and predict future outcomes. Additionally, forecasting models such as ARIMA and Holt-Winters can be used to develop short-term forecasts for a variety of different time horizons.

Short-term forecasting methods include exponential smoothing, time series analysis, autoregressive integrated moving average models, and regression analysis. Time series analysis is used to identify patterns in a given data set to make predictions about future data points. ARIMA models use past data points to predict future values by taking into account the autocorrelation between the data points. Exponential smoothing uses weighted averages to make predictions based on past data. Regression analysis is used to examine relationships between variables in a given data set to make predictions about future outcomes.

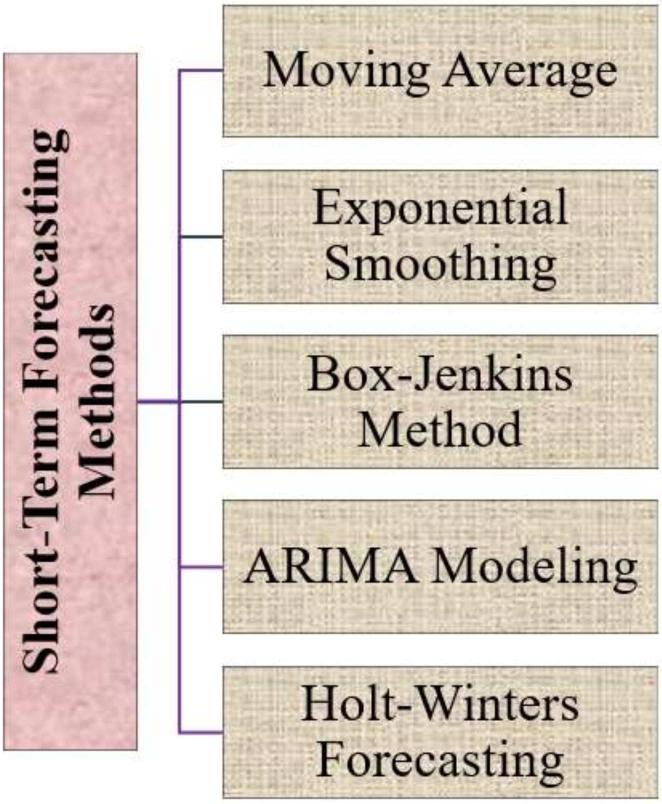


FIGURE 1.6. Methods of Short-Term Forecasting

Time series analysis uses statistical methods to analyze historical data and make predictions about future events and ARIMA models use mathematical equations to analyze trends in data and make predictions about future values.

Short-term forecasting **in the finance sector** can include forecasting stock prices, currency exchange rates, interest rates, and economic indicators such as GDP and inflation. Forecasting these variables helps investors and financial institutions plan for future market conditions. Analysts use a variety of methods to forecast these variables, including fundamental and technical analysis, econometric models, and machine learning algorithms. Additionally, information from news sources and economic reports can be used to inform forecasts. We can find daily forecasts happening for stock market indices such as Nasdaq, Dow Jones, TAIEX. These forecasts are designed to provide investors with an idea of what the markets may do on a given day. They may provide information on which stocks to buy or sell, or when to invest or divest. In addition, they can provide insight on the health of the economy and the overall market. Forecasts may be based on a variety of factors, including economic data, company news, and technical analysis.

Short-term forecasting **in the healthcare system** can be used to help anticipate changes in demand for healthcare services, predict future capacity needs, and identify areas for cost savings. This type of forecasting can include methods such as trend analysis, time-series forecasting, and regression analysis. Forecasting can help healthcare organizations prepare for a variety of scenarios, such as seasonal variations in demand, changes in patient population, and fluctuations in the availability of resources. It can also be used to better understand the impact of new technologies, regulations, and healthcare policies on the system.

ii) Medium term forecasting:

This forecasting involves looking at the trends of a business over a period of several months to several years. This type of forecasting can be used to make decisions about future strategies, investments, and resource allocation. Factors such as economic growth, customer demand, and technological advances are taken into consideration when making medium-term forecasts. Companies use this information to make decisions about where to focus their resources and investments and to make more informed decisions about the future.

Forecasting error in Medium-term forecasting is mainly caused by,

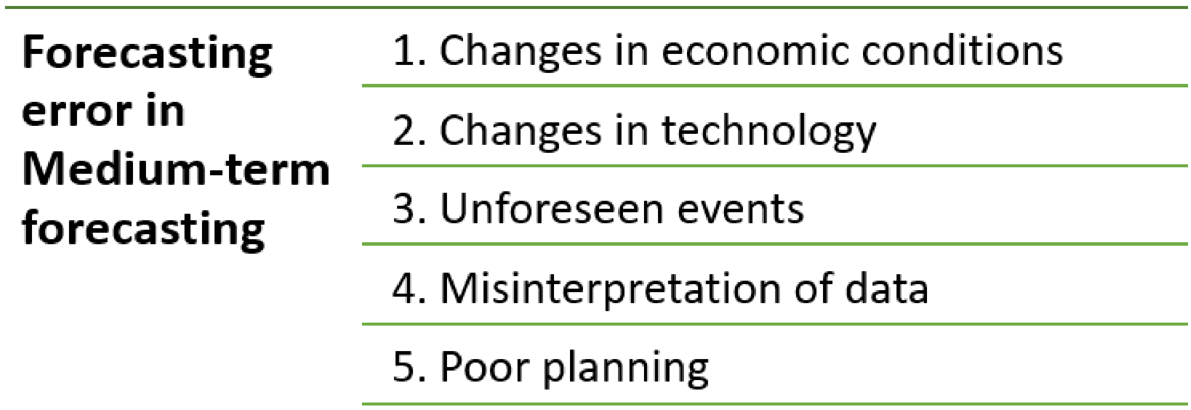


FIGURE 1.7. Forecasting error in medium-term forecasting

1. Changes in economic conditions: Changes in economic conditions, such as inflation, economic growth, and consumer spending, can have a major impact on the accuracy of medium-term forecasts.

2. Changes in technology: Rapidly changing technology can cause medium-term forecasting accuracy to suffer due to a lack of accurate data.

3. Unforeseen events: Unforeseen events, such as natural disasters, political unrest, or pandemics, can have a large impact on medium-term forecasts.

4. Misinterpretation of data: Misinterpreting data can lead to inaccurate assumptions and forecasts.

5. Poor planning: Poorly executed plans and strategies can lead to inaccurate medium-term forecasts.

Forecasting error in Medium term forecasting is typically higher than short-term forecasting because of the increased complexity of the data being used and the greater number of variables that need to be taken into account.

iii) Long term forecasting method:

This forecasting method includes econometric models, time-series analysis, and trend analysis. Time-series analysis involves analyzing historical data to determine patterns and trends that can be used to predict future outcomes. Econometric models use economic variables such as inflation, employment, and gross domestic product to predict future economic conditions. Trend analysis involves studying the behavior of markets to identify trends that may be used to forecast future market performance.

This method used to predict future trends and events based on past data. This method can be used to make predictions about sales, economic cycles, population growth, and much more. There are a variety of methods used for long-term forecasting including time series analysis, regression analysis, trend analysis, and econometric models. Each of these methods has its own strengths and weaknesses and should be chosen based on the specific forecasting needs of the organization.

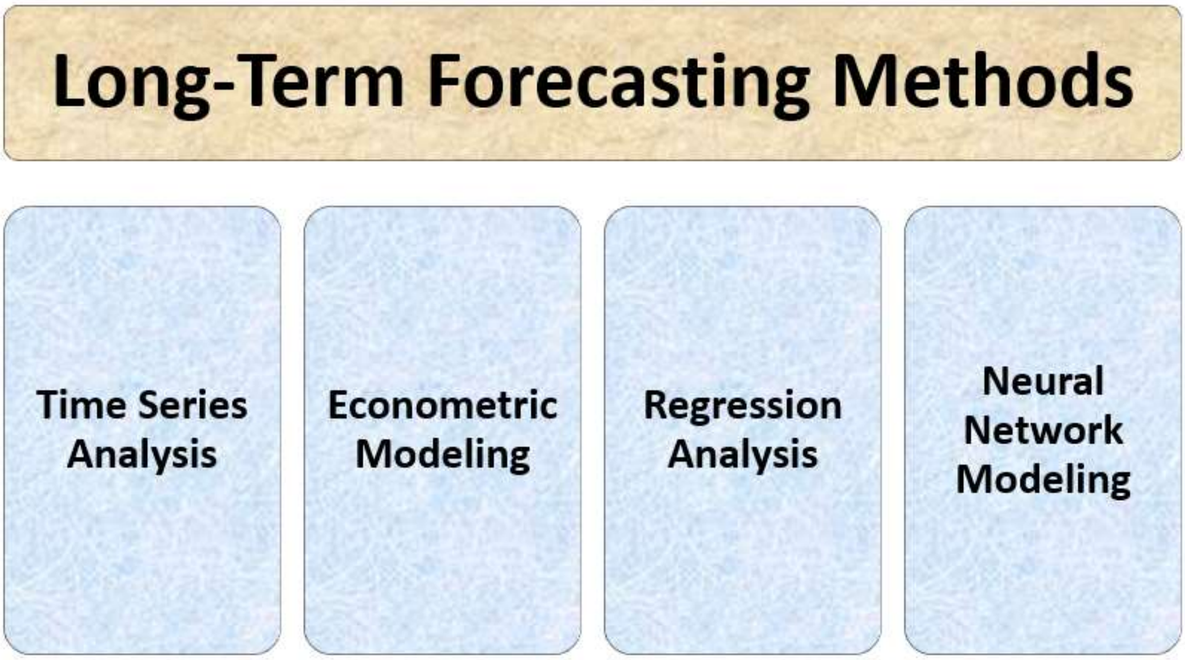


FIGURE 1.8. Methods of Long-Term Forecasting

Long-term forecasting methods include qualitative methods such as Delphi technique, consumer surveys, and expert opinion, as well as quantitative methods such as linear regression, time series analysis, and econometric models. Qualitative methods are best suited for forecasting trends in consumer preferences and industry trends, while quantitative methods are best suited for forecasting future market sizes, sales volumes, and financial performance.

Forecasting error in Long term forecasting is higher than short-term forecasting because it is more difficult to accurately predict future events that are further away in time. Long-term forecasting is subject to more challenges due to the changing environment, which makes it difficult to make accurate predictions. The accuracy of the forecast decreases as the time frame increases, as there are more uncertainties when attempting to predict events in the future.

1.2.1.2 Classification according to Application

For each circumstance and application area, there are several forecasting methodologies. For instance, we won't use time series forecasting for data if there isn't a time series record. Each application field has a unique model and set of steps to choose the best model from among them.

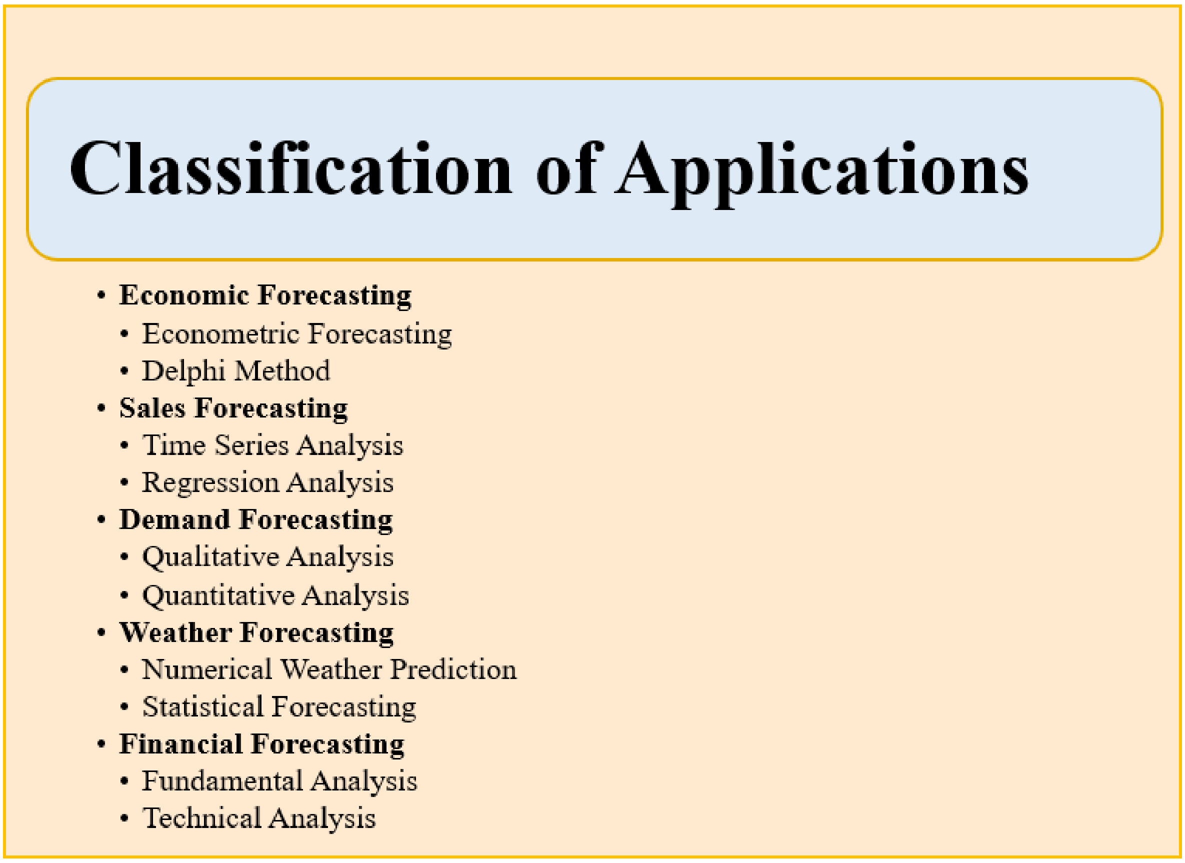


FIGURE 1.9. Application’s Classification

In each application field, the best forecasting methodology should be chosen based on the data and the context of the problem. It is important to consider the accuracy, cost, scalability, and other factors when selecting the right forecasting methodology.

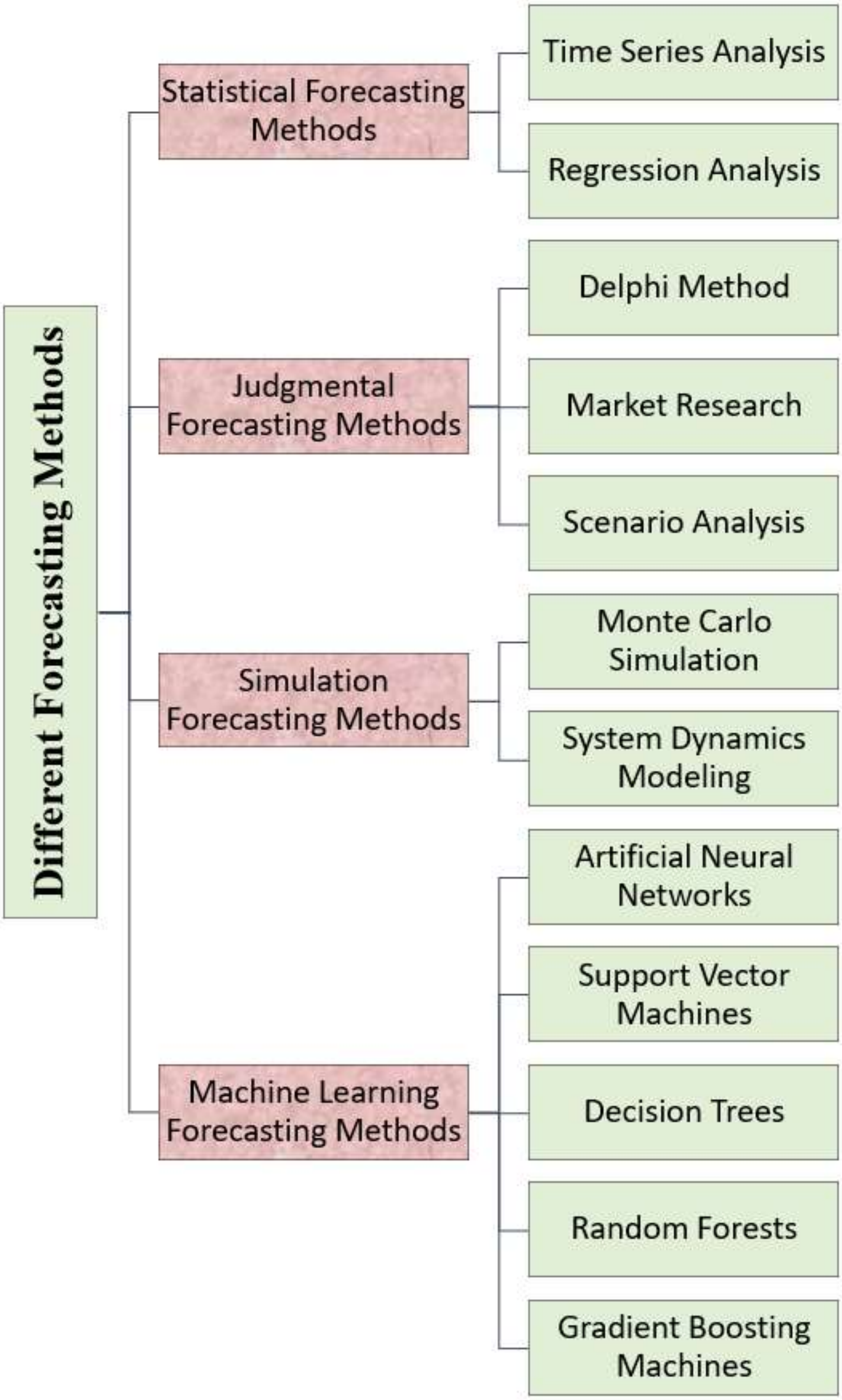


FIGURE 1.10. Forecasting Methods

We have discussed the forecasting techniques that are widely used in some application fields.

i) Multiple equation forecasting method:

Multi-equation forecasting methods are approaches to forecasting that use multiple equations to make predictions. These methods are typically used when data is complex and includes many variables. They are also used when there is a need to account for different types of relationships between variables. Examples of multi-equation forecasting methods include structural equation modeling, vector autoregression, and artificial neural networks.

Multiple equation forecasting is a type of forecasting that uses equations to predict a set of related variables. It is used to forecast multiple related variables at the same time, using a combination of statistical techniques. This type of forecasting is useful for understanding the relationships between different variables and for predicting their future values. It is often used in economics and finance to predict the future values of a range of economic variables, such as inflation, interest rates, economic growth, and unemployment.

Multiple equation forecasting is a method used in economics and finance to forecast future values of multiple variables. It involves using multiple equations which represent the relationship between the variables to predict their future values. This method can be used when the relationships between the variables are complex and cannot be captured by a single equation. It is most often used in econometric models, which are used to analyze economic and financial data. The equations used in multiple equation forecasting are typically derived from economic theory and can incorporate both qualitative and quantitative information.

ii) Time series forecasting method:

This method is used to predict future values of a series of data points based on past values. This method is used in a variety of industries, such as finance, economics, and weather forecasting. It is based on the assumption that patterns that occurred in the past will also occur in the future. Common techniques used in time series forecasting include exponential smoothing, autoregressive integrated moving average models, and the Holt-Winters technique.

Common methods include ARIMA models, exponential smoothing, and neural networks. Each of these methods uses historical data to develop a forecast of future values. The choice of which method to use depends on the type of data being analyzed, the time period being forecasted, and the accuracy required.

When time series observations are correlated, it means that the value of the series at any given point in time is dependent on the value of the series at previous points in time. This is often referred to as autocorrelation, and it can be measured using various statistical tests, such as the autocorrelation function or the Durbin-Watson test. Correlated time series can be used to make predictions about future values, as they provide information about the underlying trends and patterns in the data. In time series analysis, primary aim is to identify patterns in the data and make predictions based on them.

Time series is an observation that changes over time/according to time, and that time is not certain, it changes in units of a second, a minute, an hour, a day, a month, a year, etc. and the time series changes in necessary units. Nothing is certain in the universe, everything changes according to time, for example; temperature, price, demand, supply, birth, death, etc.

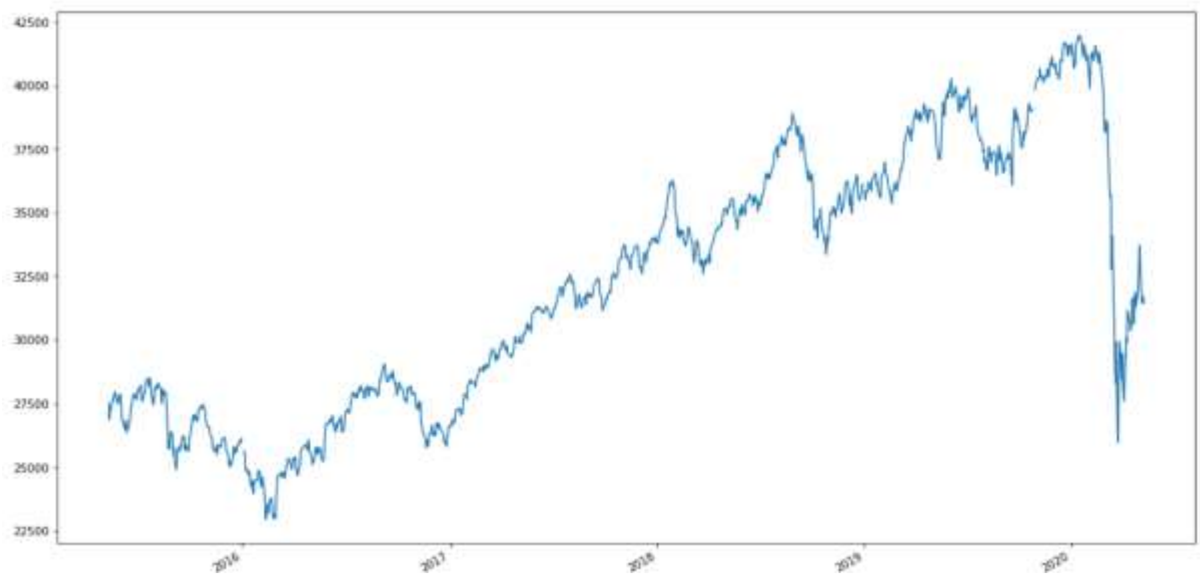


FIGURE 1.11. Time series of BSE Stock index Close data from April 2015 to April 2020

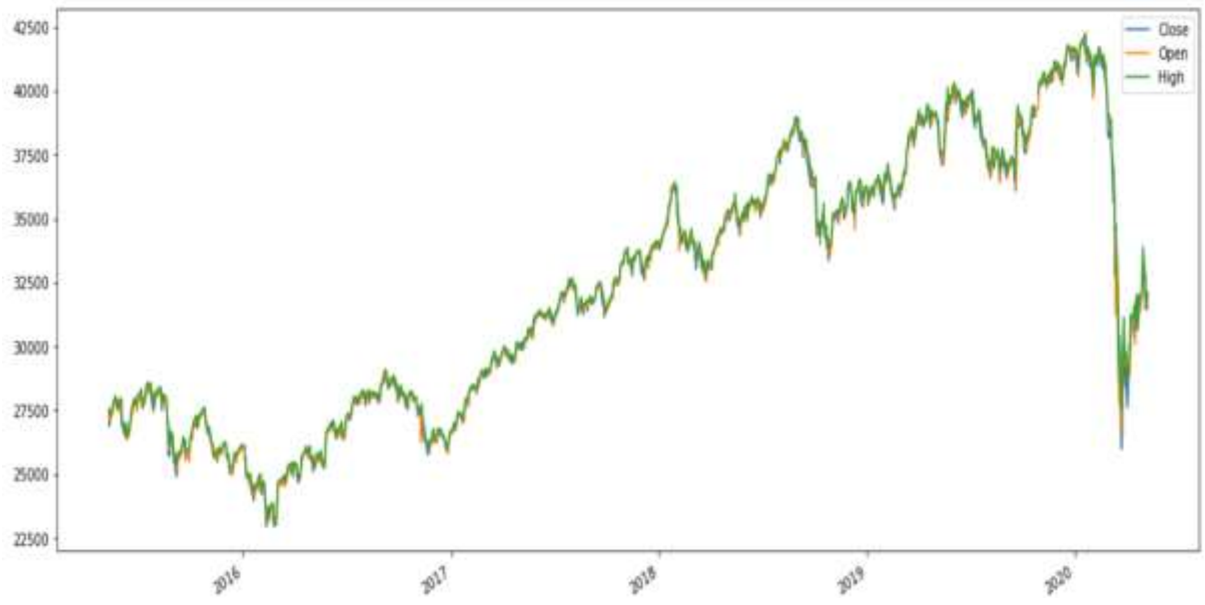


FIGURE 1.12. Time series of BSE Stock index from April 2015 to April 2020

iii) Composite methods of forecasting:

The composite method of forecasting is a combination of two or more forecasting methods in order to increase the accuracy of the forecast. This method is typically used when the data set is complex and contains multiple trends. It involves combining different forecasting techniques to come up with a more accurate forecast. This can involve combining qualitative and quantitative methods, as well as combining statistical methods with more traditional forecasting methods. The method seeks to combine the strengths of different forecasting methods to create a more accurate forecast than any single method could provide.

The method is used to reduce the potential error of one single forecast technique by taking into account different views of the same problem. This method is often used to minimize risk and to ensure that the most accurate prediction is made. The individual forecasts are weighted according to their accuracy and combined into a single consensus forecast. This method is often used in long-term forecasting as it can help account for the uncertainty of future events and trends. The best example of a composite method of forecasting is the Delphi method. This method combines the insights of a panel of experts who are asked to anonymously provide estimates of the future. The estimates are then collated, averaged, and discussed in depth to reach a consensus forecast. This method can be used to forecast a variety of outcomes, including economic indicators, population growth, and consumer trends.

iv) Simulation modeling methods:

Simulation modeling is a method of constructing a model of a real-world process using computer software. It is used to analyze the behavior of a system over time and to predict the outcome of different scenarios. Simulation models are used in many different fields, such as engineering, finance, logistics, and healthcare. They allow researchers to identify problems, develop solutions, and test hypotheses. Simulation models are used to study complex systems that are difficult to understand without a computer-based model. Examples of simulation modeling methods include discrete-event simulation, system dynamics, agent-based simulation, and Monte Carlo simulation.

It is an approach to modeling systems in which a model is developed by simulating the system's components and their interactions. Simulation models are used to study complex systems, such as traffic flow, supply-chain networks, and economies, as well as to predict the behavior of those systems. A possible drawback of Simulation modeling method is that it can be very expensive and time-consuming to create a simulation model. Additionally, it is difficult to accurately predict the results of the simulation, so the results may not be accurate.

Simulation modeling method is based on the concept of creating a model of reality in order to simulate the behavior of a system. This method is used to analyze complex systems and processes, such as natural phenomena, business processes, and social systems. It involves creating a mathematical representation of the system, which can then be used to simulate its behavior over time. The behavior of the system is then studied and analyzed to gain insights into its dynamics, such as how various factors affect its performance, as well as to predict its future behavior. Simulation modeling is often used in fields such as economics, engineering, biology, and other sciences.

Simulation modeling allows for the testing of a system without having to build it in real life. It also allows for experimentation with various scenarios, which can provide valuable insight into the system's behavior and performance. Additionally, simulation modeling can reduce the costs associated with making expensive design changes, since these changes can be tested virtually before they are implemented in the real world.

v) Cross-impact matrix method:

The cross-impact matrix method is a forecasting technique that uses a two-dimensional matrix to represent the interrelationships between variables in a system. It is based on the assumption that the relationships between variables are not linear and that the impact of one variable on another is not necessarily constant over time. The matrix is filled with numerical estimates of the relative strength of the influence between pairs of variables. Analysts use the matrix to assess the potential impacts of changes to one or more variables on the system as a whole. This technique is useful for forecasting the effects of decisions or events that may have multiple, interconnected effects on a system.

The cross-impact matrix method is a method used to investigate the interactions between different elements of a system. It is a qualitative approach used to identify the potential impacts between different elements, or variables, of a system. The cross-impact matrix method is used to forecast the effects of changes in one element on the other elements of the system. This method relies on expert opinion to assess the degree of influence the different elements have on each other, which is then represented in the form of a matrix. The matrix contains the numerical scores of the influence between the different elements. The scores range from -3 to +3, with negative scores indicating a negative influence and positive scores indicating a positive influence. This method is used by a wide range of organizations and can be used to analyze both short and long-term impacts. It is also used to identify potential risks and opportunities.

The cross-impact matrix forecasting method is a technique used to predict the impacts of future events and changes on a variety of different variables. It is a type of system dynamics modeling that uses a matrix to visualize the relationships between the variables, and the impacts of changes on them. The matrix is filled with values that represent the amount of impact each variable has on the other variables. This allows the user to identify which variables are most important and how changes in one variable will affect the others. The matrix can then be used to forecast the likely outcomes of different scenarios.

The advantage of Cross-impact matrix method allows for quick decision making by presenting all the factors in an organized manner. And also allows for a more comprehensive assessment of the impacts of various factors since it looks at the interactions between different variables. It is also adaptable to different situations since it can be used to look at both long-term and short-term impacts. The method also offers a degree of flexibility since it can be used to analyze both quantitative and qualitative data. Lastly, the cross-impact matrix method is relatively simple to use and understand, making it an ideal tool for decision-makers.

vi) Regression forecasting method:

Regression forecasting is a method of forecasting that uses historical data and existing trends to estimate future values. It is a quantitative approach that uses mathematical models to predict future outcomes. The models used in regression forecasting typically involve linear or non-linear relationships between variables, such as time-series data and other factors. The goal is to identify patterns and trends in the data and use them to accurately forecast future values. Regression forecasting can be used to predict future sales, revenue, or other business outcomes. This technique is commonly used in business forecasting, where it is assumed that the past behavior of a given variable can be used to predict its future behavior. The regression method typically involves fitting a mathematical model to a set of historical data points in order to find the best fit. This model is then used to predict future values of the dependent variable.

And many more techniques are used to make forecasts in many different areas. The choice of which technique to use will depend on the specific forecasting problem and the data available. Additionally, the choice of model should depend on the availability of data and the complexity of the problem. The accuracy of the model should be evaluated using various metrics such as Root Mean Squared Error, Mean Absolute Error, mean absolute percentage error (MAPE), and mean absolute deviation (MAD), are commonly used.

1.2.2 Characteristics of forecasting models

Forecasting models are mathematical or statistical techniques used to predict future events or patterns based on historical data. They can help organizations anticipate sudden changes in the market, understand customer behavior, and develop strategies for the future. Forecasting models vary in complexity, from simple linear regression models to more sophisticated ensemble models. These models are used in a variety of areas including finance, economics, marketing, and operations. Forecasting models are designed to provide insight into the future and can be used to make more informed decisions. They are essential for businesses, as they can help them plan for the future, anticipate risks, and make sure their resources are allocated properly. There are many different types of forecasting models, each with its own unique characteristics.

1.2.2.1 Functional form of relationship between variables

This describes how the change in one variable affects the other. This relationship can be expressed using an equation, such as a linear equation, an exponential equation, a logarithmic equation, or a polynomial equation. The specific equation used will depend on the type of data and the pattern of the relation between the two variables. For example, the linear relationship between $\mathcal{X}$ and $\mathcal{Y}$ (two variables) can be expressed as

$$\mathcal{Y} = n\mathcal{X} + d$$

where n is the slope and d is the $\mathcal{Y}$-intercept. Other examples of functional forms include exponential, logarithmic, and polynomial equations.

1.2.2.2 Dependent variable or response variable

The dependent variable or response variable for forecasting is the variable that is being predicted or estimated based on the independent variables. For example, in a forecasting model for predicting sales, the dependent variable would be the predicted sales.

1.2.2.3 Independent variable or explanatory variable

The independent or explanatory variable used for forecasting is the predictor variable, which is the factor that is used to predict the outcome of the forecast. Examples of predictor variables could include the current economic conditions, the number of customers, the amount of sales, or any other relevant factors that can be used to predict the outcome.

1.2.2.4 Horizon of forecasting

Forecasting horizons are the time frames within which a forecast is made. The forecasting horizon can vary depending on the type of forecast being made and the purpose of the forecast. In general, a short-term forecast might be made for the next quarter or year, whereas a long-term forecast might be made for five or more years. The length of the forecasting horizon depends on the accuracy of the data and the type of forecast being made. The horizon of forecasting can range from a few days to many years and is often based on the type of forecast being made and the purpose of the forecast. Forecasting horizons can be used to plan for longer-term trends in the economy, to forecast demand for products and services, to predict future sales and marketing trends, and to anticipate the impact of external events on a business.

1.2.2.5 Defined assumptions and limitations

Assumptions:

1. The data used in the model is accurate and reliable.
2. The forecasts are based on past trends and the relationships between variables used in the model.
3. The model is able to capture long-term trends and short-term movement in the data.
4. The model is able to take into account the effect of external factors such as seasonality, economic conditions, etc.

Limitations:

1. The model is based on assumptions and can be affected by changes in the data.
2. The model is limited in its ability to capture non-linear relationships or sudden changes in the data.
3. The model is limited by the number of observations used and the sampling technique used.
4. The model is limited by the complexity of the underlying algorithms and the accuracy of the parameters used.
5. The model is limited by the assumptions made about the underlying data and the assumptions made about the relationships between the variables.

1.2.2.6 Scaling of forecasting model

All the fitted models are not allowable for scaling since in the case of which we need to fit the model again. Scaling activity involves reuse of model for an extended period, the facility to add more variables into the model, etc.

1.3 ASSUMPTIONS AND LIMITATIONS OF FORECASTING METHODS

Every forecasting assignment may have certain assumptions and limitations. The number of these items will always depend on the size and magnitude of the model's data. Usually, a model with huge data may involve a lot of assumptions and limitations over its data.

1.3.1 Assumptions of forecasting

Following assumptions are considered for the forecasting;

1. Historical Data: One of the major assumptions in forecasting is that historical data holds predictive value for the future. This means that the data reflects the underlying behavior of the process and can be used to generate future forecasts.

2. Constant Factors: Another assumption is that the factors driving the process remain constant over time. This means that the underlying relationships between the different variables remain the same over time and thus the historical data is indicative of future behavior.

3. Linear Relationships: Many forecasting models assume that the relationships between the variables are linear. This means that the effects of each variable on the outcome can be accurately modeled using a linear equation.

4. Unbiased Estimates: Another assumption is that the estimates generated by the models are unbiased. This means that the estimates are not affected by external factors such as bias or noise.

5. Stable System: A final assumption is that the system being modeled is stable. This means that the underlying relationships between the variables remain the same over time and that the estimates generated by the models are reliable and not affected by external changes.

1.3.2 Limitations of forecasting

Forecasting is a valuable tool for understanding and predicting future events and trends. However, it is important to note that forecasting is not an exact science, and there are several limitations associated with it. Forecasts are based on assumptions and models, which can be subject to errors and bias. Forecasts can also be affected by unforeseen events and changing circumstances, making them inherently uncertain. Additionally, forecasts are often limited by the availability of data and the amount of time allocated for analysis. Forecasting is based on past data and trends, which can be unreliable since past events may not accurately predict future outcomes. Additionally, forecasting does not take into account any potential external factors that could affect the outcome, such as changes in the economy, new technologies, or other unforeseen events. Furthermore, forecasts often include a high degree of uncertainty, as they rely on assumptions about the future that may not be accurate. Finally, forecasting models can be complex and time-consuming to build and maintain, and they may require significant resources to implement.

1.4 ADVANTAGES OF FORECASTING

Some significance or important points of Forecasting are expressed underneath.

1. Forecasting helps organizations to make better decisions by providing insight into future trends.

2. Forecasting helps organizations to make more informed decisions about future investments and resources.

3. Forecasting allows for better planning and budgeting.

4. Forecasting allows organizations to anticipate potential risks and take preventive measures to manage them.

5. Forecasting helps to identify opportunities and plan accordingly.

6. Forecasting can help organizations to better manage their resources and maximize profits.

7. Forecasting can help to reduce the uncertainty associated with planning and decision-making.

8. Forecasting can help to improve customer service by providing insights into customer needs.

1.5 DISADVANTAGES OF FORECASTING

1. Forecasting is not always accurate. Forecasting can be based on educated guesses, but it's still a guess, and there is no guarantee that it will be correct.

2. Forecasting can be expensive. Companies must pay for the services of experts in order to get accurate forecasts, and this can be costly.

3. Forecasting can be time-consuming. Analyzing data and making predictions can take a lot of time, which can be a major drain on resources.

4. Forecasting can lead to false assumptions. Even if the forecast is accurate, it may be based on assumptions that are not true, which can lead to bad decisions.

1.6 DIFFICULTIES OF FORECASTING

1. Uncertainty: Forecasting involves examining past events and trends in order to make predictions about the future. However, the future is inherently uncertain and it is impossible to predict the exact outcome of any event.

2. Data Quality: The quality of the data used for forecasting can have a significant impact on the accuracy of the predictions. Poor data quality can lead to incorrect forecasts and inaccurate predictions.

3. Complexity: Many forecasting methods involve complex models and algorithms which can be difficult to understand and interpret.

4. Changing Conditions: The future is constantly changing and forecasts may become outdated quickly. It is important to regularly update forecasts to take into account any changes in the environment.

5. Subjectivity: Forecasting involves making assumptions and judgments which can be subjective and open to interpretation. This can lead to different predictions from different people.

6. Human error: Forecasting involves human judgement and decision making, meaning there is always the potential for human error. It is important to have a system in place to track and address errors when they occur.

1.7 FORECASTING – MATHEMATICS - STATISTICS – COMPUTATIONAL TECHNIQUES: RELATIONSHIP AND DEPENDENCY

Forecasting, mathematics, statistics and computational techniques are all related and dependent on each other. Forecasting involves using mathematics, statistics and computational techniques to predict future events. Mathematics provides the underlying principles used in forecasting, such as mathematical models and equations. Statistics is used to analyze data and provide insights into trends and patterns. Computational techniques are used to process large amounts of data quickly and efficiently. All three of these fields are heavily intertwined and rely on each other in order to create accurate forecasts.

Forecasting Terminology with Statistics

- Trend
- Seasonality
- Autocorrelation
- Moving Average
- Regression Analysis
- Forecast Error
- Forecast Bias

FIGURE 1.13. Terminology with statistics

1. **Trend:** This is a long-term direction or pattern in a time series.

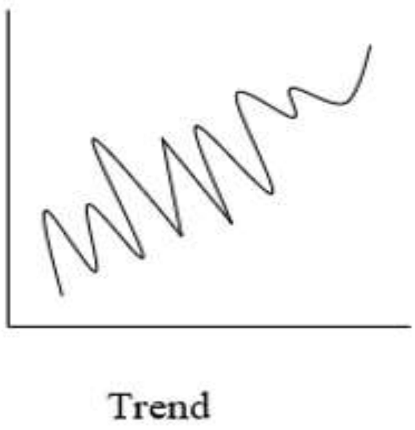


FIGURE 1.14. Trend

2. **Seasonality:** This refers to the presence of recurring patterns in a time series that are related to seasonal changes.

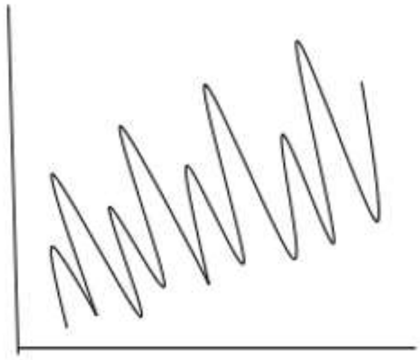


FIGURE 1.15. Seasonality

3. Autocorrelation: This is the degree of similarity between current and past values of a time series.

4. Moving Average: This is a method of smoothing out fluctuations in a time series by taking the average of a set of data points over a certain period of time.

5. Regression Analysis: This is a statistical technique used to identify relationships between variables, often used for forecasting.

6. Forecast Error: This is the difference between the actual value and the forecasted value of a time series.

7. Forecast Bias: This is the tendency of a forecast to be either too high or too low.

Forecasting with mathematics and statistic can apply with computational techniques. Computational methods of forecasting based on intuitionistic fuzzy set and fuzzy time series. A number of techniques using time series data that use fuzzy time series forecasting to predict future values and have linguistic meanings. The statistical weighted system also applied in for computational foresting.

1.8 SOME BASIC FORECASTING TECHNIQUES

Several forecasting techniques have been proposed differently by different Statisticians and Econometricians from time to time. Some important forecasting methods existing in the literature arc:

1. Naive Method
2. Moving Average Methods

3. Exponential Smoothing Models
4. Holt's Two-Parameter Method for Exponential Smoothing
5. Regression Forecasting Model
6. Trigonometric Forecasting Models
7. Autoregressive Moving Average Models
8. Autoregressive Integrated Moving Average Forecasting Model
9. Box-Jenkins Forecasting Models
10. Neural Network
11. Fuzzy Time series model

1.9 QUANTITY BASED STRATEGIES TO IMPROVED FORECASTING MODELS

Forecasting of the data it may be accurate or not, but if we have past data/ Historical data quantity then can understand the patterns of the past data or seasonal patterns of the past data which represent own role for accurate forecasting. Cannot say that forecasting is always right but it may be. It is like a probability, where no possibility to achieve something or not. When we toss a coin then we don't know about what will come but know about an option like head or tails/ 0 or 1. Forecasting always lies under 0 to 1.

1. If quantity of the data is large, then we know about more patterns of past data.
2. If quantity of the data is small, then we know about less amounts of patterns of the past data.

Both of the patterns effective on forecasting.

1. Seasonal Adjustment: Applying a seasonal adjustment factor to the historical data can help to reduce the effect of seasonal patterns in the data and improve the accuracy of the forecasting model.

2. Moving Averages: Using a moving average to smooth out the data can help to reduce the amount of variability in the data and improve the accuracy of the forecasting model.

3. Weighted Averages: Utilizing weighted averages, where more recent data is given more weight, can help to reduce the effect of outliers and improve the accuracy of the forecasting model.

4. Exponential Smoothing: Applying an exponential smoothing technique can help to reduce the effect of random fluctuations in the data and improve the accuracy of the forecasting model.

5. Hierarchical Forecasting: Utilizing a hierarchical forecasting approach, which takes into account both the aggregate and individual level data, can help to improve the accuracy of the forecasting model.

6. Data Mining: Using data mining techniques such as association rules and cluster analysis can help to uncover patterns in the data and improve the accuracy of the forecasting model.

7. Utilizing Historical Data: Analyzing historical data to identify patterns and trends in demand can help inform forecasting models. This includes studying past sales, customer behavior, and seasonal fluctuations.

8. Leveraging Automation: Automating parts of the forecasting process can help to streamline the process and reduce human error. Automated forecasting models can quickly process and analyze large amounts of data to generate accurate forecasts.

9. Utilizing Machine Learning: Machine learning algorithms can be used to identify patterns and trends in data that may not be immediately obvious. By leveraging machine learning, forecasting models can be more accurate and reliable.

10. Utilizing AI: AI can be used to identify opportunities for improvement in forecasting models. AI can process large and complex datasets to identify patterns and trends in data that may be difficult to identify with manual analysis.

11. Utilizing Real-Time Data: Utilizing real-time data in forecasting models can help to provide more accurate forecasts. Real-time data can provide insight into customer behavior, demand fluctuations, and changes in the market.

12. Utilizing Statistical Modeling: Statistical modeling techniques can be used to identify patterns and trends in data to improve forecasting models. This includes methods such as linear and logistic regression, decision trees, and Bayesian networks. Statistical modeling can be used to better understand the relationships between different variables and build more accurate predictive models. By understanding the underlying relationships between different variables, organizations can make better decisions, improve customer service, and optimize their operations.

13. Feature Engineering: Feature engineering is one of the most important aspects of creating a successful forecasting model. Adding meaningful features to the model can help improve its accuracy. This can include incorporating external data sources, transforming existing features, or identifying new features from existing data.

14. Utilize Statistical and Mathematical Techniques: Statistical and mathematical techniques, such as linear regression and exponential smoothing, can be used to improve forecasting models. These techniques can help to provide more accurate predictions and allow businesses to more effectively plan for future demand.

15. Use Machine Learning Algorithms: This algorithms can be used to improve forecasting models by automatically finding patterns in data and making more accurate predictions.

1.10 OBJECTIVE OF THE PROPOSED WORK

1. Study of stock market index by using Quantity based fuzzy time series (QBFTS).
2. Analysis of QBFTS to enrich the role of quantity based fuzzy logical groups and statistic in QBFTS algorithms.
3. Applications of Machine learning and fuzzy time series combination in Agricultural crops forecasting.
4. Application of fuzzy time series with different methods and technical analysis in different fields.

1.11 ORGANIZATION OF THESIS

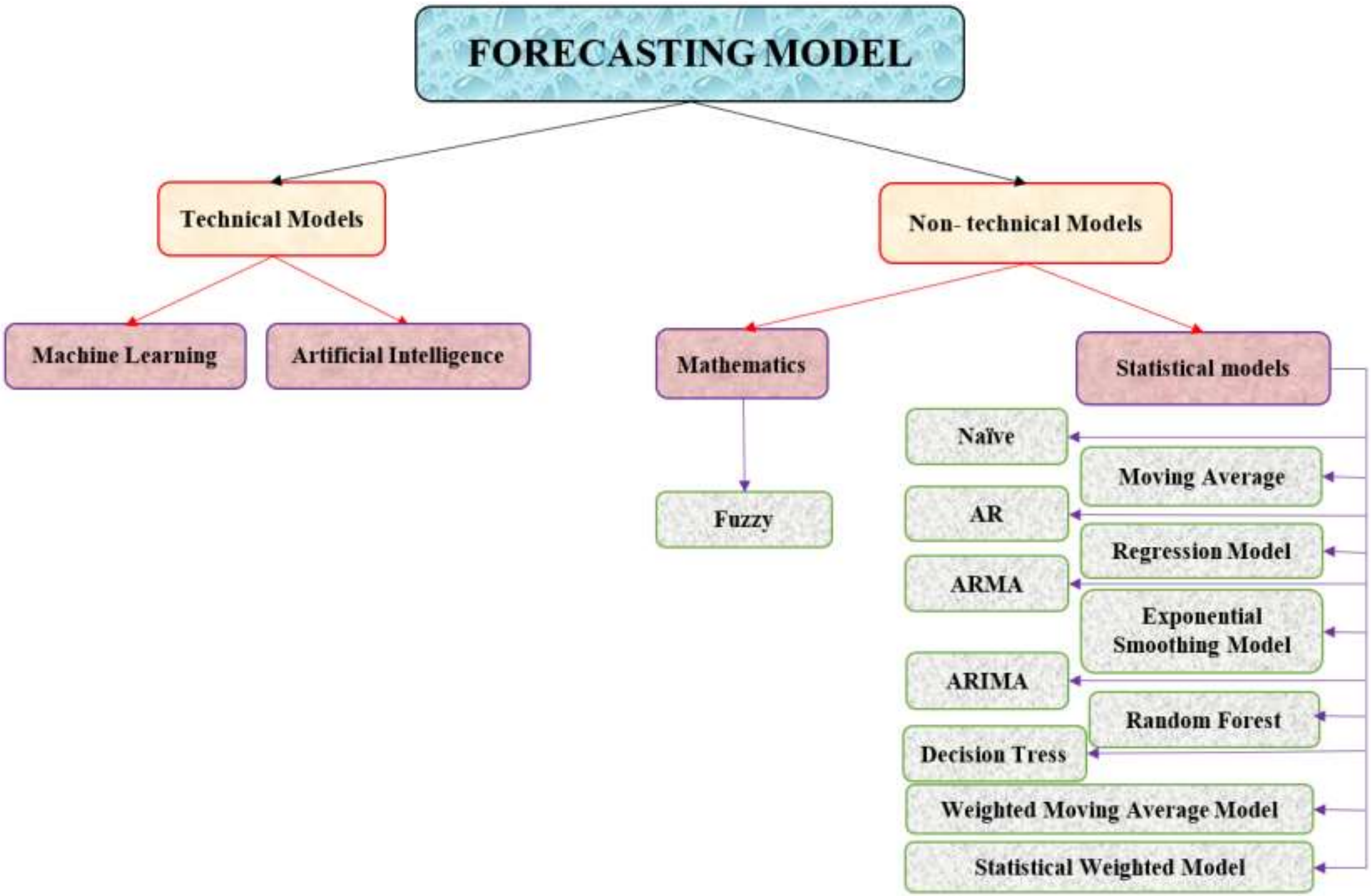


FIGURE 1.16. Flow Chart of Model

The present/ suggested workflow is for improved models to fulfill the objective of research work for the thesis title **“Some Improved Forecasting Models And Their Applications In Different Fields”**. The workflow is consisting of forecasting models enabled with different applications, figure 1.16 shows the technical and non-technical models of forecasting, a combination of these models used for improved forecasting.

Chapter one defines a general introduction to forecasting, with classification, and characteristics. Advantages, disadvantages, assumptions, limitations of the forecasting, methods, techniques, strategies to improve forecasting models, with the objective of the proposed thesis, etc., to fulfill the chapter-wise summary of the thesis.

Chapter two discusses the review of forecasting research work done in form of the era wise by many researchers. Researchers have taken different kinds of data to forecast with many applications and several methods used to compare current works with other researchers' work, which helps to justify the proposed work. several researchers gave conceptual/ pure research on the forecast, and some others have worked on these in the form of numerical analysis.

Chapter three is based on the application field as the TAIEX stock exchange index for improved forecasting combination has been considered of a fuzzy & statistical weights system and training-testing this is also a combination of technical & non-technical models and discussed the three new models which represent the technical and non-technical combination for improving forecasting.

Chapter four shows the application field as machine learning with applications in agriculture and healthcare like as covid-19 for improved forecasting as the combination of technical and non-technical models again. Two new models are discussed, first model of this chapter is based on an agriculture study with the use of random forest, SVM, NN, and the second model discussed the hybridization of the RF-DT on covid-19.

Chapter five gives the idea of the application field as crop production and crop price for improved forecasting considering the combination of ARIMA, machine learning, naïve forecast, and non-stationary time series forecasting by ARIMA model as a conjunction of technical and non-technical models.

Chapter six discusses the production of crops with the use of the statistical model as linear regression with technical uses, which represents a work combination of technical and non-technical models for improved forecasting. India's crop production data is taken for forecasting analysis.

A detailed discussion of the research work described in the six chapters of this dissertation has been done in the chapters.

End of this thesis with a list of relevant references of this workflow/chapters. At last, the work of the present thesis is one of the ways to improve the forecasting for the future, it is not the beginning nor the end, it is the part after the beginning, which has no end.

"Forecast is not certain, improvement gives better forecast, it will be close or similar to perfect but never perfect."

CHAPTER

II

Review on Forecasted Work

Think about the next time, next day, or next year, or can say that think about the future as a forecast of the future is a human habit. It did not just start, it started from when humans came to the earth or human lies. So, work of the future forecast is going on them, where humans forecast in every field, which it's known such as; physics, mathematics, chemistry, economics, biology, sociology, and so on.

This study is on mathematical-forecast related work in different sectors, where lots of researchers working from the Era's. Here, some work of the researchers is covered in this study. Due to the importance of the study of time series, several researchers took interest in prediction and applied several methods. Joshi and Kumar [1] presented a method to forecast the SBI share market price at the Bombay Stock Index, India with help of a hybrid fuzzy time series model. Also calculated the mean square error and compare some previous work. Yabuuchi et al. [2] compared two fuzzy time series models on the Nikkei Stock Average. Compared models are fuzzy autoregressive and fuzzy autocorrelation. Chen et al. [3] proposed the fuzzy time series and fuzzy variation group method to forecast of TAIEX. Ratio-based length of the interval in fuzzy time series for forecasting by Huarng and Yu [4]. Multivariate heuristic functions integrate with fuzzy time series model use of multi-heuristic variables for stock market index forecasting by Huarng and Yu [5] and these heuristic functions are extended and integrated, which is improve forecasting results. Neural network model in a fuzzy time series for forecasting performance between BSE1000 and NIFTY MIDCAP50 stock market index by Kumar & Murugan [6].

In [7], Chen et al. discussed fuzzy time series and generated weights of multiple factor based forecasting of Taiwan Stock Exchange Capitalization Weights Stock Index (TAIEX). Here using the variation and fuzzy variation groups of the factors as NASDAQ, TAIEX, Dow Jones, M1b). In [8], Basset et al. discussed about the neutrosophic time series with neutrosophic logical relationship groups and first and higher-order neutrosophic time series to forecast on student enrollment. Das et al. [9] studied three companies of BSE/NSE with use of fuzzy membership function and fuzzy intervals. Singh [10] discuss fuzzy time series and fuzzy time series forecasting with the help of soft computing (SC) techniques. In [11] Chou applied prediction on (Taiwan STI) Taiwan shipping and Transportation Index use of fuzzy time series with long–term significance level analysis. In [12] Garg and Garg presented a ordered weighted aggregation (OWA) to forecast of fuzzy time series with statistical concepts on TAIEX and University of Alabama enrollment data. Liu et al. [13] applied the fuzzy time series forecast with help of time- variation. In [14] Krollner et al. studied the forecast stock market movements with use of machine learning techniques and artificial intelligence. The feed-forward neural network (FFNN) model with forecasting non-linear time series for Canadian lynx data set by Kajitani [15].

In [16] Tsaur used a fuzzy time series with a Markov chain model for forecast the exchange rate. Bianco et al. [17] applied linear regression models for forecasting of electricity consumption in Italy. The intutionistic fuzzy logic (IFC) and neutrosophic logic generalized by Smarandache [18]. In [19] Susruth studied three methods for forecast the stock price of the Indian Stock Market. These methods are ARIMA time series method, moving average method and hall & winter exponential method. For Forecasting, applied the Neural network on a fuzzy time series moreover, used a Bivariate model for the forecasting performance of TAIEX by Yu and Haurang [20]. Yu [21] proposed a weighted model for fuzzy time series forecasting with two issues as recurrence and weighting, and compared with local regression models on Taiwan stock index forecasting. Forecast the regime switches use of autoregressive relationship by Huarng et al. [22].

Zhang et al. [23] applied a novel method to accurate prediction and used fuzzy logic in forecasting time series. Eyoh et al. [24] applied a type-2 and type-1 fuzzy model to forecast with membership grade and membership function on fuzzy time series. Song and Chissom [25] proposed the fuzzy time series, fuzzy relational equations. Huarng et al. [26] worked on a handle non-linear problem used by fuzzy time series model and generate the non-linear arrangement of the neural network for forecasting. Applied linear regression for forecasting behavior of TCS data set by Bhuriya [27]. Lv et al. [28] ARIMA & Long short-term memory model used for stationary & unstable time series forecasting, respectively and with adaptive noise complete ensemble empirical mode decomposition used for time series of stock index. Herawati et al. [29] discussed and compared of two forecasting models as higher-order Chen fuzzy time series & feed-forward back propagation natural network on composite stock price index. Abbasimehr and Paki [30] discussed two deep learning methods as multi-head attention and long short-term memory (LSTM) to improve time series forecasting.

Fuzzy time series of TAIEX index was used for the forecasting with a single multiplicative neuron to identify fuzzy relation with particle swarm optimization by Yoka & Alpaslan [31]. Forecast stock market time series such as BSE, NYSE, TAIEX of four months by fuzzy transfer learning by Pal and Kar [32]. Bisht & Kumar [33] discussed a fuzzy time series forecasting used of hesitant fuzzy sets for forecasting at the price of State Bank of India (SBI) and enrolment of the university of Alabama, and also used an aggregation operator for hesitant information. Pavlyshenko [34] applied a Machine Learning model for sale forecasting for the case of a new product or store launched calculated by jupyter notebook. Sousa et al. [35] discussed Hierarchical temporal memory (HTM) theory for time series forecasting in stock market. Torres et al. [36] applied deep learning techniques for time series forecasting.

Lim & Zohren [37] used one-step-ahead and multi horizon time series forecasting with deep learning models. Artificial intelligence technique used for forecasting of export sales by Sohrabpour et al. [38]. Anufriev et al. [39] discussed a simple heuristic forecasting model and Genetic Algorithm optimization procedure to forecast of prices at simple linear first-order. Concept of FTS (Fundamental-Technical-Speculative) analysis to a prediction by Miciula [40] for the changes of the exchange rate in the forex market. Tozan et al. [41] evaluated the forecasting models as fuzzy time series, fuzzy linear regression, and fuzzy grey GM(1,1) on performance of supply chain. Javedani Sadaei and Lee [42] proposed a model with multilayer to forecast of the stock index included with five layers, which are logical significant on considered five stock market index data. Dai et al. [43] discussed the combination of Markov chain and improved back-propagation (BP) neural network, with the modeling and computational techniques to forecast. Zadeh [44] proposed a novel fuzzy time series forecasting model with linguistic variables, fuzzy algorithms, and time series clustering for market prices. A mathematical model for securities of stock market described by Gonchar [45].

Couts et al. [46] studied the prediction of non-stationary and non-deterministic series processes with seasonality. Young [47] discussed the techniques in the non-stationary tie series analysis and also discussed the forecasting of nonstationary time series. Coulibaly and Baldwin [48] proposed RNN approach to forecast on different non-stationary time series of the water resource system. Online learning algorithms for estimate used of ARIMA models on time series by Liu et al. [49]. Adebiyi [50] used the ARIMA models for prediction of the stock index price data as the short-term prediction. Abhishekh et al. [51] discussed the intuitionistic FTS for forecasting and used fuzzify at historical time series data and created intutionistic fuzzy logical relationships on fuzzified data of intuitionistic. Wilinski [52] used Markov chains model at two different financial time series for prediction with fixed lengths. Tay and Cao [53] discussed the application part of the novel NN technique, a Support vector machine for forecasting of time series data of five stock indexes. Sah and Degtiarev [54] used time-invariant in fuzzy time series to forecasting at the performance of enrolment data. Chen et al. [55] presented Fibonacci sequence and weighted method to forecast of stock index at different period. Wang et al. [56]

proposed Wavelet De-noising-based back propagation neural network on stock index for forecasting and compared with single back propagation neural network. Wong et al. [57] applied Traditional and fuzzy time series methods for forecasting of Taiwan data. Chu et al. [58] proposed the fuzzy time-series model modified with a dual factor for forecasting on a stock index. Chen et al. [59] used comprehensive fuzzy time series and fuzzy logical relationships for the forecasting of stock prices.

Efendi et al. [60] calculated forecasting on the electricity load of all days data from TNB by using numerical time series and linguistics time series forecasting. Weron [61] applied autoregression (AR) model for short-term forecasting of electricity prices, California market, before market crash of winter in 2000/2001. Forecasting of electric load used to combined classic methods of principle component analysis (PCA) and autoregressive (AR) model, as well as used orthogonal principle least square (OPLS) by Gordillo-Orquera et al. [62]. Delima [63] applied ARIMA model for forecasting of Philippines electric consumption. Mazengia & Tuan [64] tested a multiple linear regression model on time series data of Canadian electricity market and Nordic electricity market for electricity price forecasting. ARMA & ARMAX model apply on the California power market system as loads and prices forecasting by Weron & Misiorek [65]. Cai et al. [66] used a ACO (Ant Colony Optimization) and Auto regression combined with time series for forecasting of TAIEX close price of three months from November 2012 to January 2013. Junior et al. [67] used ARIMA model for possible order to forecasting of time series data of Bovespa stock index. Support vector machine, and particle swarm optimization used to forecast of stock market index price movements by Zhiqiang et al. [68]. Tanuwijaya and Chen [69] presented a clustering and fuzzy time series at different interval lengths on data. Vovan [70] described improved FTS model to forecast of used variations of data. Atsalakis and Valavanis [71] used Adaptive Neuro fuzzy Inference System to forecast stock market as short-term. Bisht and Kumar [72]. Simple computational based on intuitionistic fuzzy set used to forecast of stock indexes. Tsai et al. [73] proposed novel multifactor FTS model for forecasting of a stock index at three factors. Hassan et al. [74] used hybrid model of fuzzy time series with ARIMA and Interval type-2 FLS to forecast stock index. Ballapragada et al. [75] forecast of exchange trade fund use of Box-Jenkins model (ARIMA) on past data and also used regression model. Mehtab and Sen [76] applied hybrid model of machine learning and deep learning as CNN to stock price forecasting. In the book, Montgomery et al. [77] described the concept with examples of time series analysis forecasting with several methods. Mathur [78] described the machine learning and application of ML used in Python for several sector for analysis of forecasting. Zimmermann [79] described the basics to advance part of the fuzzy set theory for forecasting analysis in several fields also discussed the application part with fuzzy.

Anderson [80] discussed export forecasting in the domestic field. For the forecasting of cancer trends used optimization control strategies by Janerich [81]. Grabowski & Vernon [82] Discussed the returns and disk to R& D for drugs into United States. Weiss et al. [83] Discussed univariate time series techniques as ARIMA, linear regression, exponential smoothing for the forecasting of area wide hospital. Kosirog et al. [84] used compartment model and Bayesian forecasting for aminoglycoside in patient of cancer. Goldman [85] used wild cards hypothesis techniques for healthcare forecasting. Abidi and Goh [86] described a backpropagation neural network model for the forecasting of bacteria- antibiotic interactions for infectious disease control. Grover et al. [87] estimated forecast for long-term benefits and cost effectiveness of lipid modification in secondary prevention of cardiovascular diseases.

Hulme & Xu [88] applied neural network configuration of optimization for stock market forecasting. Xue et al. [89] used stepwise autoregressive method and exponential smoothing models for forecast of End-stage renal disease (ESRD) patient in United States. Bae et al. [90] used time series model and autoregressive model for information of cancer death in Korea for upcoming years. Hao et al. [91] applied multiple regression analysis of time series for forecast of tourism demand from markets. Enke & Thawornwong [92] used the data mining and neural networks technique for forecasting of stock market returns. Lakshminarayanan et al. [93] used artificial neural network (ANN) model for stock market forecasting. Rodriguez [94] applied forecasting for the analysis of attribution returns. Masursky et al. [95] used time series method for forecasting to future demand of Anesthesia workload.

In the field of medical, Sun et al. [96] Applied ARIMA model for forecasting in emergency department to aid planning and analysis carried out by SPSS. Kam et al. [97] evaluated time series model for prediction of daily patients visiting in the emergency department (ED) of Korean hospital. Froelich et al. [98] for the prediction of prostate cancer used fuzzy cognitive maps (FCM) model by authors. Garg et al. [99] applied fuzzy time series for the forecasting of outpatient visits in hospitals.

Time series sequence techniques applied to forecast for cancer treatment clinic by Claudio et al. [100]. Bro et al. [101] discussed the forecasting of breast cancer by plasma metabolic and bio-contours with new possibilities and individual cancer risk. Golmohammadi & Zaiane [102] proposed contextual anomaly detection (CAD) method on time series for detecting market manipulation in stock market. In [103] Dang et al. applied Grey model and Lotka Volterra (LV) model for the forecasting analysis in healthcare traveling industry. Ganguly and Nandi [104] Applied a statistical technique, ARIMA for optimize staff scheduling in healthcare organization. Aripin et al. [105] used fuzzy time series for prediction of pollutant PM10 concentration in air. In [106] Calegari et al. for forecasting of demand in medical care in emergency department (ED) calculated by SARIMA model.

Iqelan [107] discussed grey prediction model GM (1, 1) and ARIMA model to forecast of female breast cancer. Kaushik et al. [108] proposed statistical and neural methods for time series forecasting in healthcare. Harrou et al. [109] proposed an effective method to forecast daily and hourly visits at an emergency department (ED) used VAE (Variational Auto Encoder) algorithm on time series data. Firmino et al. [110] used a non-central beta (NCB) probability density function to forecast on pandemic time series. Budiharto [111] discussed long-short term memory methods for stock price forecasting of Indonesian exchange as used index data of Bank of Central Asia & Bank of Mandiri.

In the field of agriculture, the purpose of forecasting must differ work on many crops at crop price, production, demand, etc., implementation of the mathematical applications, mathematical algorithms, statistical analysis, and computational processes by software. Discusses some related previous work done, Zhang et al. [112] applied single-variable regression in rice grain/vegetables versus natural log-transformed concentration in soil on cropland of China. Hare [113] discussed the impact of defoliation on Potato yield and the applied experiment was repeated seven times in intervals, where one interval is equal to two weeks. Risk-neutral and risk-averse formulations are applied on northeastern Oregon farms data by Nazer et al. [114]. Regression-based model was applied to estimate the impact of multiple pets on crop productivity by Johnson [115]. Gandhi et al. [116] applied a neural network to predict the rice production of districts of Maharashtra. Jadhav et al. [117] forecasted the price of Paddy, Ragi, and Maize in the Indian state of Karnataka for the year 2016 on time series data from 2002 to 2016 by the ARIMA model for the forecasting up to the year 2020. Biswas and Bhattacharyya [118] applied ARIMA (p,d,q) model for the forecasting of the area and production of rice in West Bengal. Elsamie et al. [119] used the ARIMA models on the time series of the data set for the forecasting of cultivated area, productivity, and production of cotton crops. Ho et al. [120] used neural networks and Box-Jenkins ARIMA on time series forecasting for failures of repairable systems. Box-Jenkins ARIMA and ANN used to forecast the behavior of data for production and used a hybrid approach on the time series data of wheat production in Haryana by Devi et al. [121]. Purohit et al. [122] used hybrid methods for the forecasting of the price of agricultural products such as tomatoes, Onions, and Potatoes. Awal and Siddique [123] used the ARIMA model to forecast rice production in Bangladesh. Singh et al. [124] developed a regression model to forecast of potato yield from farmers' fields in Manipur.

Forecasting of time series data in other fields related to air, tourism, production, water, etc., are related to research work done by researchers discussed there. Fuzzy time series and fuzzy time series forecasting techniques used to forecast for air pollution and air quality forecasting in China by Wang et al. [125]. Wind time series forecasting by deep neural network-based approach as NARX to wind speed forecasting by Rahman [126]. Gupta et al. [127] studied on day-ahead and intra-day wind power forecasting of actual and error wind power use of wavelet decomposition. For air pollution forecasting used a fuzzy time series model on air pollution index data and also applied Markov weighted fuzzy time series model by Alyousifi et al. [128]. Huarng et al. [129] forecast on tourism demand of Taiwan by fuzzy time series model. Xiao et al. [130] discussed a neuro-fuzzy combination model based forecasting of air transport demand with artificial intelligence technologies. Forecasting of tactical sales in the Tire industry by Sagaert et al. [131].

Haque et al. [132] described ARIMA model for forecasting as Marine, inland, and total production of fish in Bangladesh for five months of year 2005. For the water quality parameter prediction for true value, global optimization and generalization used a Least square support vector machine method on river water by Tan et al. [133]. Maier and Dandy [134] used Artificial Neural Networks (ANNs) for the forecast of water resource variables. LSSVM (Least squares support vector machine) model was used for forecasting of lake water pollution time series data with the help of kernel principal component analysis on Taihu Lake by Ni et al. [135]. Holt-Winter model was used to forecast of water pollution caused by the textile industry of Poland and Romania Paraschiv et al. [136]. Heuristic Gaussian cloud transformation-based approach to forecast the water quality time series with multi-factor by Deng et al [137]. Kogekar et al. [138] used three models of time series to forecast on water quality of the Ganga River. Dutta et al. [139] applied an ARIMA model for the period 1967 to 2015 to death forecasting in road accidents in India for the upcoming 10 years.

CHAPTER

III

Quantity-Based Fuzzy Time Series (QBFTS) on Stock Market Index

This Chapter, QBFTS and the application of FTS applied on different stock market indexes and also connect with statistical weights to quantity-based fuzzy logical relationship groups. Here, considered historical data is divided into two parts as training and testing. Further, discussed two types of factors as primary and secondary factors to different indexes. Primary index factor is a target index of forecasting but Secondary index factor is a supported index of primary index forecasting.

3.1. INTRODUCTION

Time series forecasting is an important problem in data analysis and predictive analytics. It is used to forecast the future values of a series based on past values. Traditional forecasting techniques such as ARIMA and NN have been widely used in stock market index forecasting. However, these methods are not suitable for fuzzy time series (FTS) due to the presence of uncertainty and imprecise data. To address this issue, the concept of **Quantity-Based Fuzzy Time Series (QBFTS)** proposes.

3.1.1. Quantity-Based Fuzzy Time Series (QBFTS)

QBFTS is a fuzzy method that combines quantitative and qualitative information to generate a forecast. It considers both the past and current data points of a series and evaluates the degree of membership of each data point in the series. Based on this, it computes the fuzzy membership of each data point and then uses it to generate the forecast. QBFTS has been successfully used in stock market index forecasting. It has been shown to outperform traditional forecasting methods in terms of accuracy and robustness. Additionally, QBFTS has been used to generate reliable forecasts of financial indices in different countries.

Overall, QBFTS is a promising approach for stock market index forecasting. It can be used to generate accurate and reliable forecasts in the presence of fuzzy and imprecise data. It is also more robust than traditional forecasting techniques and can be used to generate forecasts for different countries.

Quantity-based fuzzy time series is a forecasting technique that utilizes fuzzy logic to create a more accurate prediction of future trends. This method is based on the idea that future trends can be more accurately predicted by taking into account the **"quantity"** of past data points, rather than just their individual values. Fuzzy time series uses fuzzy logic to create a **"fuzzy set"** of values, allowing the forecasting algorithm to better approximate the behavior of future data points. This method is useful in many forecasting applications, such as predicting stock prices, predicting weather patterns, and predicting consumer demand.

QBFTS can be used to identify the major trends in the stock market by using fuzzy set theory. Theory of fuzzy set can be used to identify the major trends in the stock market by using fuzzy membership functions. The fuzzy membership functions can be used to identify the major trends in the stock market by assigning a degree of membership to each stock in the index. The degree of membership is based on the stock price movement. The fuzzy membership functions can then be used to identify the major trends in the stock market by determining the most likely direction of the stock price movement.

QBFTS can also be used to forecast future stock prices. This is done by taking into account the current trend in the stock market and by using fuzzy logic to determine the most likely future stock price. QBFTS can be used to forecast the future stock prices of individual stocks as well as the overall index. The forecasts can be used to identify the major trends in the stock market as well as to determine when to buy and sell stocks.

The application of the Quantity-Based Fuzzy Time Series (QBFTS) model in stock market index forecasting has been studied extensively in recent years. The purpose of this research is to assess the efficacy of this approach for predicting stock price movements. Specifically, this research will analyze whether the QBFTS model can provide reliable forecasts for stock indices. This study will focus on the accuracy of QBFTS predictions in comparison to traditional forecasting methods, such as linear regression and moving average. The study will also evaluate the effect of various parameters, such as the number of fuzzy sets and the weight of each set, on the accuracy of the forecasts. In addition, the research also investigates the impact of the presence of outliers and other noise elements in the data on the accuracy of the QBFTS model. Finally, the research will assess the scalability of the QBFTS model and its ability to be applied to large datasets.

3.1.2. Fuzzy Logic

Fuzzy logic is based on the idea of making decisions based on the degree of truth of a proposition. In fuzzy logic, propositions can have degrees of truth between 0 and 1, rather than being either true (1) or false (0). This allows for more flexibility and complexity in decision making.

Definition I: Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1. It is employed to handle the concept of partial truth, where the truth value may range between completely true and completely false.

Fuzzy logic is based on the mathematics of fuzzy set theory and fuzzy logic operations. Fuzzy set theory is used to define the set of values a variable can take, and fuzzy logic operations are used to manipulate these values.

Fuzzy logic has been extended to handle the concept of linguistic variables, which allows human language to be used in the formulation of complex rules and fuzzy if–then control statements. Fuzzy logic is used in many industrial and consumer applications such as medical diagnosis, artificial intelligence, control systems, and others.

In fuzzy logic, a set of rules is used to map an input to an output. These rules are written in a language that is similar to natural language, so that the user can understand them. The output is then determined by combining the results of the rules with a set of weights assigned to each rule.

3.1.3. Fuzzy Set

This is a set whose elements have degrees of membership. It is a generalization of the classical set theory in which the elements have two-valued membership.

The basic concept of a fuzzy set is that it allows objects to have partial membership in a set, where partial membership is indicated by a membership grade. This membership grade is usually a real number between 0 and 1, with 0 indicating that the object has no membership in the set, and 1 indicating that the object has full membership in the set. Fuzzy sets are an important tool used in many branches of mathematics. They are used in fuzzy logic, fuzzy optimization, fuzzy control, fuzzy clustering, and fuzzy pattern recognition.

3.1.4. Fuzzy Time Series (FTS)

Fuzzy time series is a type of time series analysis that uses fuzzy logic to analyze data. It is used to forecast future values and identify patterns in data that are too complex to detect using traditional statistical methods.

Fuzzy logic and time series are two different concepts. Fuzzy logic is a type of logic that deals with mathematical logic in terms of degrees of truth. It is used to describe situations in which the information is not precise or exact, and the rules of traditional logic do not apply. Time series, on the other hand, is a type of data that is collected over a period of time. It is used to measure the changes in a certain variable over a period of time, and can be used to forecast future values of the variable.

Fuzzy logic and time series analysis can be used together to improve the accuracy of forecasting models. Fuzzy logic can be used to identify patterns and trends in the time series data that may not be easily identified by traditional methods. This can then be used to create more accurate forecasting models. For example, fuzzy logic can be used to provide more accurate estimates of future demand and sales, making it easier to predict future market trends. In addition, fuzzy logic can be used to identify outliers in the data and adjust the forecasts accordingly. By combining fuzzy logic and time series analysis, organizations can better anticipate customer needs and make more informed decisions.

The fuzzy time series method and application of fuzzy set may be a dynamic process for multiple values of observations of problems. In the stock market has differ readings as open, low, high, close, etc, and constructed into a time series. Fuzzy time series is a method of forecasting future values by using fuzzy logic. This method is used in many fields such as engineering, economics, and finance. The process of fuzzy time series starts by constructing a fuzzy set. The fuzzy set is a range of values that can be assigned to a particular attribute. From many Era's, researchers are working on forecasting with fuzzy time series and perform as computations to generate forecasts.

3.1.5. Fuzzy Relationship

Fuzzy relationships are based on the idea that values exist on a continuum, rather than having definitive boundaries. For example, instead of assigning an object a specific temperature, a fuzzy relationship might assign a range of temperatures to the object. The same concept applies to other types of relationships, such as those between two people or two objects.

The mathematics of fuzzy relationships uses fuzzy logic principles to represent relationships in terms of fuzzy sets and fuzzy variables. A fuzzy set is a collection of values that have some degree of similarity, while a fuzzy variable is a numerical representation of the degree of similarity between two objects or values. Fuzzy inference is the process of making decisions based on the values present in a fuzzy set.

Fuzzy relationships are often used in artificial intelligence applications, such as robotics and autonomous vehicles. They can also be used in decision-making systems to make decisions based on uncertain or incomplete data. Fuzzy relationships can also be used to represent relationships between people, such as in social networks.

3.1.6. Membership Functions of Fuzzy Sets

A membership function is a mathematical representation of a fuzzy set that maps the universe of discourse to a value between zero and one. It is used to determine the degree to which a given element belongs to a given set. Membership functions can be used to represent fuzzy sets, fuzzy logic, and other forms of uncertain reasoning. Membership functions are typically defined by a bell-shaped curve, but can also be defined using other shapes such as a triangle, Gaussian, or any other arbitrary shape. Membership functions can be used to represent various aspects of a fuzzy set, such as its degree of membership or its degree of similarity to other sets.

Definition II: [42,44] Let Z be a universe (i.e., an arbitrary set). A fuzzy subset A of Z, is characterized by a function $A : Z \rightarrow [0, 1]$, which is called the membership function. For every $z \in Z$, the value A(z) is called a degree to which element z belongs to the fuzzy subset A.

Otherwise put, a fuzzy subset is a set of elements (e.g. a group of people, animals, objects, etc.) that have a degree of belonging to the set. This degree is expressed as a number between 0 and 1, where 0 indicates that the element does not belong to the set, and 1 indicates that the element does belong to the set.

Definition III: [33] A membership function for a fuzzy set A can be defined as $\mu_A(z):Z\rightarrow[0,1]$, where Z is the universe of discourse. This means that for any element z in the universe of discourse Z, the membership function maps z to a number between 0 and 1. The higher the membership value, the more likely it is that z belongs to the fuzzy set A.

Mathematically, the membership function of a fuzzy set A can be defined as:

$$\mu_A(z) = \begin{cases} 1 \ if \ z \in A \\ 0 \ if \ z \notin A \end{cases}$$

In other words, the membership function of a fuzzy set A is a function that takes an element z and returns 1 if it is a member of the fuzzy set A and 0 if it is not a member of A.

For example, the membership function of a fuzzy set A can be defined as:

$$\mu_A(z) = \begin{cases} 1 \ if \ z \geq 0 \\ 0 \ if \ z < 0 \end{cases}$$

This membership function will map any input z to 1 if z is greater than or equal to 0 and to 0 if z is less than 0.

3.2. TYPE OF MEMBERSHIP FUNCTION

A membership function is a type of mathematical function used to define the fuzzy set membership grade of an element. It is used in fuzzy logic to represent the degree of truth of a statement based on the given input. The membership functions help to determine the input and output variables of a system, how those variables interact, and how they can be manipulated to achieve a desired result.

FIGURE 3.1. Types of Membership Function

Each type of membership function is used to reflect different levels of fuzziness, or grades of membership, in a fuzzy set. Triangular membership functions are the most commonly used, as they are easy to understand and use. Other types of membership functions may be used to represent more complex fuzzy sets.

3.2.1. Trapezoidal Membership Function

A trapezoidal membership function has some parameters such as a, m, n, d. which is defined by;

$$\text{Trapezoidal}(x; a, m, n, d) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{m-a}, & a \le x \le m \\ 1, & m \le x \le n \\ \dfrac{d-x}{d-n}, & n \le x \le d \\ 0, & d \le x \end{cases}$$

3.2.2. Triangular Membership Function

$$\text{Triangular}(x: a, b, m) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ \dfrac{m-x}{m-b}, & b \le x \le m \\ 0, & m \le x \end{cases}$$

3.2.3. Gaussian Membership Function

$$\text{Generalized}(x; \sigma, m) = e^{\frac{-1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

3.2.4. Generalized-Bell Membership Function

$$\text{Generalized} - \text{Bell}(x; a, b, c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}}$$

3.2.5. Sigmoid Membership Function

$$\text{Sigmoid }(\text{x}; \text{a}, \text{c}) = \frac{1}{1 + e^{[-a(x-c)^2]}}$$

3.2.6. Left-Right (L-R) Membership Function

In this membership function have two parts as left and right curves, meet at point (c,1).

$$Left - Right(x; c, \alpha, \beta) = \begin{cases} F_L\left(\dfrac{c-x}{\alpha}\right), & x \le c, \alpha > 0 \\ F_R\left(\dfrac{x-c}{\beta}\right), & x \ge c, \beta > 0 \end{cases}$$

3.3. METHOD 1

Definition IV. (Fuzzy set): [16,29,33,41,42,58] Let the universe of the discourse $\mathcal{Z}, where\ \mathcal{Z} = z_1, z_2, z_3, \dots, z_m$ of considered historical data and the universe of discourse,

$$\mathcal{Z} = [D_{min} - D_1, D_{max} + D_2]$$

where, D_1 and D_2 are proper positive real values to partition the universe of discourse $\mathcal{Z}$ into m intervals $z_1, z_2, z_3, \ldots,$ and z_m are equal length and denoted D_{max} for maximum value and D_{min} for minimum value. [7, 16, 20, 21, 29, 33,42] A fuzzy set β_τ in $\mathcal{Z}$ is defined as follows;

$$\beta_\tau = \frac{f_{\beta_\tau}(z_1)}{z_1} + \frac{f_{\beta_\tau}(z_2)}{z_2} + \frac{f_{\beta_\tau}(z_3)}{z_3} + \cdots + \frac{f_{\beta_\tau}(z_m)}{z_m}$$

Where, f_{β_τ} is relationship function of the fuzzy set β_τ and $f_{\beta_\tau}(z_j)$ represents the degree of connection of (z_j) belonging to the fuzzy set β_τ.

$$f_{\beta_\tau}(z_j) \in [0,1], \quad 1 \le j \le m$$

Definition V: [11,12,16,21,31] Suppose M(t), t = …,-2,-1,0,1,2,3,… be the universe of discourse and be a subset of R (Real numbers). Let $f_\tau(t)$, where τ = 1,2, 3, be the fuzzy set, which is defined in the universe of discourse M(t) and let F(t) be a set of $f_\tau(t)$ then the set F(t) is called a Fuzzy time series of M(t).

Definition VI: [7,11,16,31,42] If a fuzzy relationship R(t, t+1) exists, such that $F(t+1) = F(t) \circ R(t, t+1)$ where "∘" denote the max-min composition operator, then F(t+1) is called the caused by F(t) and fuzzy relationship is ,

$$F(t) \rightarrow F(t+1)$$

Both F(t) and F(t+1) are fuzzy sets.

Definition VII: [12, 16] Let F(t) and F(t+1) are β_τ and β_j, respectively. Relation between these two fuzzy sets are represent in the form "$\beta_\tau \rightarrow \beta_j$". β_τ and β_j are left-hand (L-H) side and right-hand (R-H) side of a fuzzy logical relationship. FLR makes a FLRGs as follows;

$$\beta_\tau \rightarrow \beta_{j1}, \beta_{j2}, \ldots, \beta_{jm}$$

The calculation for Var_{t+1} i.e., for t+1 days will be

$$Var_{t+1} = \frac{Close_{\ t+1} - Close_t}{Close_t} \times 100$$

and represented by Var_{t+1}. Here, $Close_{t+1}$ is closing data value of day t+1 and $Close_t$ is closing data value of day t [7].

Now the next processes will be for computation of total difference var ($\text{difference}_{(\text{SQBF})_\text{i}}$) between the main quantity-based factor (MQBF) for t days and the elementary secondary quantity-based factor (SQBF) for t-1 days is,

$$difference_{(SQBF)_i} = \sum_{t=2}^{s} \left| Var_{(SQBF)_{t-1}} - Var_{(MQBF)_t} \right|$$

Where, number of days is s.

Calculate the weighted variation $WV_{(QBSF)_1}, WV_{(QBSF)_2}, \dots, WV_{(QBSF)_m}$ of secondary quantity-based factor $(QBSF)_1, (QBSF)_2, \dots, (QBSF)_m$, respectively.

$$WV_{(QBSF)_\tau} = \frac{difference_{(SQBF)_1} + difference_{(SQBF)_2} + \cdots + difference_{(SQBF)_m}}{difference_{(SQBF)_\tau}}, 1 \leq \tau \leq m$$

Calculate the normalized weighted variation $WV_{(SQBF)_1}, WV_{(SQBF)_2}, \dots, WV_{(SQBF)_m}$ of secondary quantity-based factor $(SQBF)_1, (SQBF)_2, \dots, (SQBF)_m$, respectively.

$$WV_{(QBSF)_\tau} = \frac{WV_{(QBSF)_\tau}}{WV_{(SQBF)_1} + WV_{(SQBF)_2} + \cdots + WV_{(SQBF)_m}}, \quad 1 \leq \tau \leq m$$

The secondary quantity-based factor's variation Var_t on t day is calculated as;

$$Var_t = Var_{((SQBF)_1)_t} \times WV_{(QBSF)_1} + Var_{((SQBF)_2)_t} \times WV_{(QBSF)_2} + \dots + Var_{((SQBF)_m)_t} \times WV_{(QBSF)_m}$$

The variation is done into fuzzy variation i.e., fuzzified and it is a fuzzy set. Let X_r be the variation into the fuzzy form of main quantity based factor for t days and X_s be the variation into the fuzzy form of secondary quantity-based factor for t-1 days then the fuzzy variation "X_r" of the main quantity-based factor (MQBF) of t day into the fuzzy variation "Group X_s", $1 \leq s \leq m$ and m is a number of fuzzy sets.

In "Group X_s" have different fuzzy variations and it makes a three condition are M< L, M > L, and M = L. Then, we can calculate $X_{L,k}$ with the help of these three conditions, where $1 \leq k \leq 3$.

Calculate the weights of $X_{L,1}, X_{L,2}, X_{L,3}$ is,

$$W_{X_{L,k}} = \frac{X_{L,k}}{X_{L,1} + X_{L,2} + X_{L,3}}\ , \qquad 1 \le k \le 3,\ \ 1 \le L \le 14$$

There two conditions for forecasting;

1. If the main quantity-based factor testing data of day t is fuzzified into fuzzy set β_τ and Variation of the secondary quantity-based factor on day t is fuzzified in X_j and quantity based FLR exist in the QBFLG “Group X_j”, then the forecasted value of the main quantity-based factor testing data of day t is calculated as;

$$z_i^* = z_i^L \times W_{X_{L,1}} + z_i^M \times W_{X_{L,2}} + z_i^H \times W_{X_{L,3}}$$

Where, $W_{X_{L,3}}$ is the weight and $z_i^L, z_i^M and\ z_i^H$ are lower value, mid-value and higher value of the interval.

2. If the main based factor testing data of day t is fuzzified into the fuzzy set β_τ and there are no quantity-based FLR in QBFLRG “Group X_j”, hence forecasted value of the main quantity-based factor of testing data for t days is equal to mid-value (z_i^M) of the interval z_j.

$$z_i^* = z_i^M$$

Forecasted value of the main quantity-based factor for t testing day is

$$\text{Forecast value } (z_i^*) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{e}} z_i^*}{\mathrm{e}}$$

3.4. MODEL: FORECASTING OF TAIEX STOCK INDEX

3.4.1. Data/ Data Source

Taiwan Capitalization Weighted Stock Index (TAIEX) is a Chinese stock market index, which was published in 1967 and 100 value as considered as base value of year 1966. (Details of TAIEX). Motivation of consideration of TAIEX stock index is that, in the field of the finance/ stock market index forecasting almost research work processes completed on TAIEX stock market index, which is useful for compare to proposed work, but past work data consider till to year 2004 and proposed work at the resent pass year as 2018. Daily price at the close of the day data of TAIEX from 1Jan 2018 to 31Dec 2018 obtained from Chinese stock market index.

3.4.2. Computation Process

The implementation of the quantity based FTS algorithms for forecasting on the TAIEX stock index forecasting, where considered quantity-based time series data value fuzzified into fuzzy sets and generate a quantity based FLR group and statistical weights on quantity based FTS algorithms.

Quantity-based time series factual data of the TAIEX Stock Index, NASDAQ, and Dow Jones consider for the forecasting during the period year 2018 from Jan–December [140,141,142]. Considered data divided into two parts training (Jan–Oct) and testing (Nov–Dec). And applied two types of factors i.e., main quantity-based factor and secondary quantity-based factor. The main quantity-based factor is TAIEX Stock Index and the secondary quantity-based factor is NASDAQ and Dow Jones.

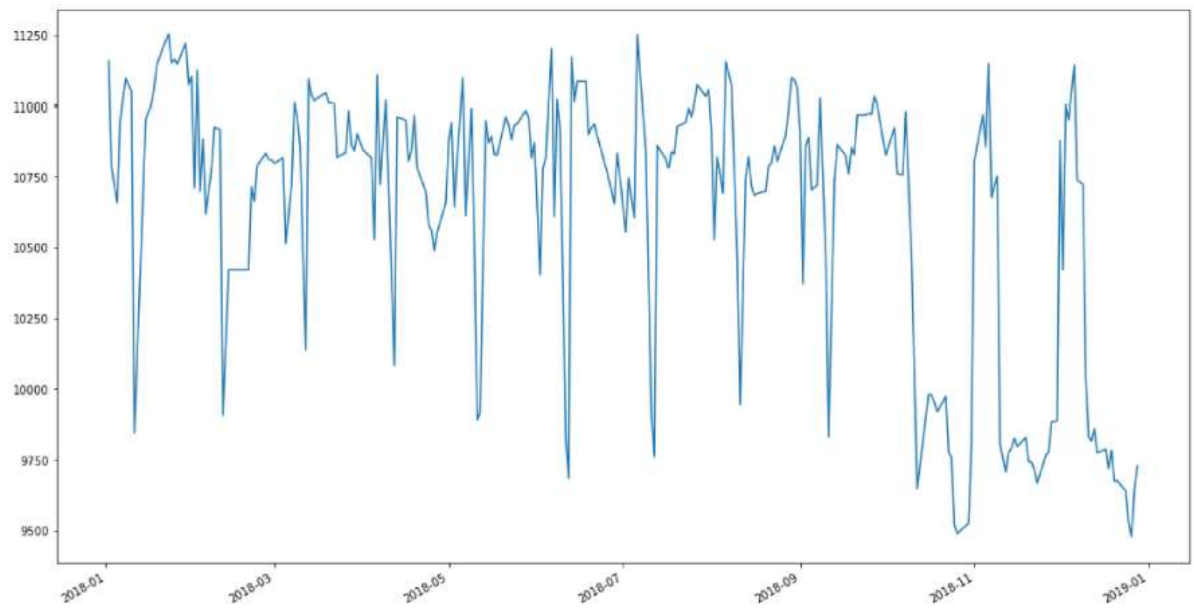


FIGURE 3.2. TAIEX stock market index actual close value of 2018.

Define the universe of discourse $\mathcal{Z}, \mathcal{Z} = [D_{min} - D_1, D_{max} + D_2]$, where, D_{max} = maximum value and D_{min} = minimum value, for quantity-based TAIEX stock index factual data. Where D_1 and D_2 are proper positive real values to partition the universe of discourse $\mathcal{Z}$ into m intervals $z_1, z_2, z_3, \ldots,$ and z_m are equal in length [16, 29,33,41, 58].

From historical data of the year 2018, maximum and minimum value will be, respectively. 11253.11035 and 9489.17968 & accordingly $\mathrm{D}_1 = 89.179688$ and $\mathrm{D}_2 = 46.88965$ the universe of discourse $\mathcal{Z}$ = [9400, 11300] partitioned by length of each interval is 100. Hence discourse $\mathcal{Z}$ divided into 19 intervals $\mathcal{Z} = z_1, z_2, z_3, \ldots,$ and z_{19} i.e.,

$$z_\tau = [9400 + (\tau - 1) * 100,\ \ 9400 + \tau * 100]\ , \qquad \tau = 1,2,3,\ldots,19$$

So newly framed will be a fuzzy set β_τ(τ =1,2,…,19).

$$\beta_1 = \frac{1}{z_1} + \frac{0.5}{z_2} + \frac{0}{z_3} + \frac{0}{z_4} + \frac{0}{z_5} + \frac{0}{z_6} + \cdots + \frac{0}{z_{18}} + \frac{0}{z_{19}}$$

$$\beta_2 = \frac{0.5}{z_1} + \frac{1}{z_2} + \frac{0.5}{z_3} + \frac{0}{z_4} + \frac{0}{z_5} + \frac{0}{z_6} + \cdots + \frac{0}{z_{18}} + \frac{0}{z_{19}}$$

$$\beta_3 = \frac{0}{z_1} + \frac{0.5}{z_2} + \frac{1}{z_3} + \frac{0.5}{z_4} + \frac{0}{z_5} + \frac{0}{z_6} + \cdots + \frac{0}{z_{18}} + \frac{0}{z_{19}}$$

$$\vdots$$

$$\beta_{19} = \frac{0}{z_1} + \frac{0}{z_2} + \frac{0}{z_3} + \frac{0}{z_4} + \frac{0}{z_5} + \frac{0}{z_6} + \cdots + \frac{0.5}{z_{18}} + \frac{1}{z_{19}}$$

Fuzzifying daily factual data of main quantity-based factor into a fuzzy set. The data of TAIEX 02-01-2018 is 10710.73047 fuzzified into the fuzzy set β_{14}. Similarly, data of TAIEX for 03-01-2018 is 10801.57031 and "10801.57031" is fuzzify into the fuzzy set β_{15}.

Constructing quantity based fuzzy logical relationship between 02-01-2018 to 03-01-2018 i.e.

$$\beta_{14} \rightarrow \beta_{15}$$

While data of TAIEX for 03-01-2018 is β_{15} and for 04-01-2018 is β_{15}. Also constructing again quantity based fuzzy logical relationship between 03-01-2018 to 04-01-2018 i.e.

$$\beta_{15} \rightarrow \beta_{15}$$

Similarly, we can find out first order QBFLR. i.e., shown in **Table 3.1**.

TABLE 3.1. Quantity Based Fuzzy logical relationship for 2018

Date	**Quantity Based Fuzzy logical relationship**
02-01-2018 → 03-01-2018	$\beta_{14} \rightarrow \beta_{15}$
03-01-2018 → 04-01-2018	$\beta_{15} \rightarrow \beta_{15}$
04-01-2018 → 05-01-2018	$\beta_{15} \rightarrow \beta_{15}$
05-01-2018 → 08-01-2018	$\beta_{15} \rightarrow \beta_{16}$
08-01-2018 → 09-01-2018	$\beta_{16} \rightarrow \beta_{16}$
⋮	⋮
25-10-2018 → 26-10-2018	$\beta_2 \rightarrow \beta_1$
26-10-2018 → 29-10-2018	$\beta_1 \rightarrow \beta_2$

29-10-2018 → 30-10-2018	$\beta_2 \rightarrow \beta_2$
30-10-2018 → 31-10-2018	$\beta_2 \rightarrow \beta_5$

Now, the factual data variation of the main quantity-based factor will be

$$\frac{10801.57031 - 10710.73047}{10710.73047} \times 100 = 0.84812\%$$

for day 02-01-2018 & 03-01-2018.

While the variation of the TAIEX for 04-01-2018 is,

$$\frac{10848.62988 - 10801.57031}{10801.57031} \times 100 = 0.435673\%$$

Similarly, we can obtain the variation for all day training data of the TAIEX, Dow Jones, and the NASDAQ in **Table 3.2**.

TABLE 3.2. Variation of uses Stock Indexes for 2018

Date	Variation of TAIEX	Variation of NASDAQ	Variation of Dow Jones
03-01-2018	0.84812%	0.836745%	0.397478%
04-01-2018	0.435673%	0.175222%	0.611697%
05-01-2018	0.287317%	0.828633%	0.880308%
08-01-2018	0.330431%	0.291878%	-0.05087%
⋮	⋮	⋮	⋮
25-10-2018	-0.16163%	2.953406%	1.631713%
26-10-2018	-2.44493%	-2.06508%	-1.18569%
29-10-2018	-0.33201%	-1.63132%	-0.99395%
30-10-2018	0.286017%	1.579508%	1.76624%

Defining the universe of discourse W, based on the minimum and maximum variation of the TAIEX each day i.e., -7% and 7% respectively, while the variations of the NASDAQ and Dow Jones do not have such limitations.

Let W be the universe of discourse, defined as (-∞, ∞) and each interval length between [-6%, 6%] be equal to 1%. W can be divided into 14 intervals $w_1, w_2, w_3, \dots, w_{14}$. Which are

$(-\infty,-6)$, $[-6,-5)$, $[-5,-4)$, $[-4,-3)$, $[-3,-2)$, $[-2,-1)$, $[-1,0)$, $[0,1)$, $[1,2)$, $[2,3)$, $[3,4)$, $[4,5)$, $[5,6)$, $[6,\infty)$.

Universe of discourse W represents a linguistic terms represented by fuzzy sets. We can define the linguistic terms X_j, j =1,2,3,…,14, represented by fuzzy sets, which are;

$$X_1 = \frac{1}{w_1} + \frac{0.5}{w_2} + \frac{0}{w_3} + \frac{0}{w_4} + \frac{0}{w_5} + \cdots + \frac{0}{w_{14}}$$

$$X_2 = \frac{0.5}{w_1} + \frac{1}{w_2} + \frac{0.5}{w_3} + \frac{0}{w_4} + \frac{0}{w_5} + \cdots + \frac{0}{w_{14}}$$

$$X_3 = \frac{0}{w_1} + \frac{0.5}{w_2} + \frac{1}{w_3} + \frac{0.5}{w_4} + \frac{0}{w_5} + \cdots + \frac{0}{w_{14}}$$

$$\vdots$$

$$X_{14} = \frac{0}{w_1} + \frac{0}{w_2} + \frac{0}{w_3} + \frac{0}{w_4} + \frac{0}{w_5} + \cdots + \frac{1}{w_{14}}$$

Where $w_1, w_2, w_3, \ldots$ and w_{14} are intervals.

Fuzzify each day variation of the QBMF into a quantity-based fuzzy variation represented by a fuzzy set. Considered example; TAIEX index variation for day 03-01-2018 is 0.84812% based on fuzzy set, and it is fuzzified into fuzzy variation X_8 and the variation for day 04-01-2018 is 0.435673% and this variation is based fuzzy variation X_8. Similarly, we can fuzzify each day variation of QBMF provided in **Table 3.3**.

TABLE 3.3. Fuzzify the variation of quantity based main factor (QBMF)

Date	Fuzzy Variations
03-01-2018	X_8
04-01-2018	X_8
05-01-2018	X_8
08-01-2018	X_8
⋮	⋮
25-10-2018	X_7
26-10-2018	X_5
29-10-2018	X_7
30-10-2018	X_8

Grouping of QBFLR having the same fuzzy variation at the LHS into a QBFLRG. In the Table 3.1, we can see the FLR from day 02-01-2018 to 03-01-2018 is $\beta_{14} \rightarrow \beta_{15}$ and Table 3.4 fuzzy variation of day 03-01-2018 is X_8 then we can group the QBFLR, $\beta_{14} \rightarrow \beta_{15}$ into the quantity-based FLRG "Group X_8". In the same way we can show the QBFLRGs with respect to different fuzzy variations in **Table 3.4**.

TABLE 3.4. QBFLRGs with respect to different fuzzy variations

Groups	**Quantity Based Fuzzy Logical Relationships**
Group X_1	$\beta_{11} \rightarrow \beta_5$
Group X_3	$\beta_{16} \rightarrow \beta_{11}$
Group X_5	$\beta_{16} \rightarrow \beta_{16}$ $\beta_4 \rightarrow \beta_2$
Group X_6	$\beta_6 \rightarrow \beta_4$ $\beta_7 \rightarrow \beta_6$ $\beta_{12} \rightarrow \beta_{10}$ $\beta_{13} \rightarrow \beta_{12}$ $\beta_{14} \rightarrow \beta_{13}, \beta_{11}$ $\beta_{15} \rightarrow \beta_{14}, \beta_{14}, \beta_{14}$ $\beta_{16} \rightarrow \beta_{15}, \beta_{15}, \beta_{14}, \beta_{15}$ $\beta_{17} \rightarrow \beta_{16}, \beta_{16}, \beta_{16}, \beta_{15}$ $\beta_{18} \rightarrow \beta_{17}, \beta_{16}$ $\beta_{19} \rightarrow \beta_{17}$
Group X_7	$\beta_2 \rightarrow \beta_1$ $\beta_4 \rightarrow \beta_4$ $\beta_6 \rightarrow \beta_6, \beta_6, \beta_6$ $\beta_{11} \rightarrow \beta_{10}$ $\beta_{12} \rightarrow \beta_{11}, \beta_{12}, \beta_{12}$ $\beta_{13} \rightarrow \beta_{12}, \beta_{13}, \beta_{13}, \beta_{13}$ $\beta_{14} \rightarrow \beta_{13}, \beta_{13}, \beta_{14}, \beta_{13}, \beta_{14}, \beta_{14}, \beta_{13}, \beta_{14}, \beta_{13}, \beta_{13}, \beta_{14}$ $\beta_{15} \rightarrow \beta_{14}, \beta_{15}, \beta_{15}, \beta_{15}, \beta_{14}, \beta_{15}, \beta_{15}, \beta_{15}, \beta_{15}, \beta_{14}, \beta_{15}, \beta_{14}, \beta_{15}, \beta_{15}, \beta_{14}, \beta_{15}$ $\beta_{16} \rightarrow \beta_{16}, \beta_{15}, \beta_{15}, \beta_{16}, \beta_{15}, \beta_{15}, \beta_{16}, \beta_{15}, \beta_{16}, \beta_{15}, \beta_{16}, \beta_{16}, \beta_{15}, \beta_{16}, \beta_{17}, \beta_{16}$ $\beta_{17} \rightarrow \beta_{16}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{16}, \beta_{17}, \beta_{16}, \beta_{17}, \beta_{17}, \beta_{16}, \beta_{17}$ $\beta_{18} \rightarrow \beta_{18}, \beta_{18}, \beta_{18}, \beta_{18}, \beta_{18}$
Group X_8	$\beta_1 \rightarrow \beta_2$

	$\beta_2 \rightarrow \beta_2$ $\beta_6 \rightarrow \beta_6, \beta_6$ $\beta_{10} \rightarrow \beta_{11}, \beta_{10}, \beta_{11}$ $\beta_{11} \rightarrow \beta_{12}, \beta_{11}, \beta_{11}, \beta_{11}, \beta_{11}, \beta_{11}$ $\beta_{12} \rightarrow \beta_{13}, \beta_{12}, \beta_{13}$ $\beta_{13} \rightarrow \beta_{13}, \beta_{13}, \beta_{14}, \beta_{14}, \beta_{13}, \beta_{14}, \beta_{13}$ $\beta_{14} \rightarrow \beta_{15}, \beta_{15}, \beta_{15}, \beta_{14}, \beta_{15}, \beta_{14}, \beta_{14}, \beta_{15}, \beta_{15}, \beta_{14}, \beta_{14}, \beta_{15}$ $\beta_{15} \rightarrow \beta_{15}, \beta_{15}, \beta_{16}, \beta_{16}, \beta_{16}, \beta_{15}, \beta_{15}, \beta_{15}, \beta_{16}, \beta_{15}, \beta_{15}, \beta_{15}, \beta_{15}, \beta_{16}, \beta_{15}, \beta_{16}, \beta_{15}, \beta_{16}, \beta_{15}, \beta_{15}, \beta_{16}, \beta_{16}, \beta_{15}$ $\beta_{16} \rightarrow \beta_{17}, \beta_{16}, \beta_{16}, \beta_{16}, \beta_{17}, \beta_{16}, \beta_{16}, \beta_{16}, \beta_{16}, \beta_{16}, \beta_{16}, \beta_{17}, \beta_{17}, \beta_{15}, \beta_{16}, \beta_{17}, \beta_{16}, \beta_{16}, \beta_{16}$ $\beta_{17} \rightarrow \beta_{18}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{18}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{17}, \beta_{17}$ $\beta_{18} \rightarrow \beta_{19}, \beta_{19}, \beta_{19}, \beta_{18}, \beta_{18}, \beta_{18}$ $\beta_{19} \rightarrow \beta_{19}, \beta_{19}$
Group X_9	$\beta_{11} \rightarrow \beta_{12}$ $\beta_{13} \rightarrow \beta_{14}, \beta_{14}, \beta_{15}, \beta_{14}$ $\beta_{14} \rightarrow \beta_{15}, \beta_{15}$ $\beta_{15} \rightarrow \beta_{17}, \beta_{16}, \beta_{16}, \beta_{16}, \beta_{16}$ $\beta_{16} \rightarrow \beta_{18}, \beta_{17}$
Group X_{10}	$\beta_5 \rightarrow \beta_7$ $\beta_{11} \rightarrow \beta_{14}$

Now, estimate total difference var between QBMF as TAIEX and the elementary SQBF as NASDAQ from 03-01-2018 to 30-10-2018. i.e.

$$|0.836745 - 0.435673| + |0.175222 - 0.287317| + \cdots + |(-1.63132) - 0.102876| = 176.942$$

Similarly, estimate total difference var between the QBMF as TAIEX and the elementary QBSF as Dow Jones from 03-01-2018 to 30-10-2018. i.e.

$$|0.397478 - 0.435673| + |0.611697 - 0.287317| + \cdots + |(-0.99395) - 2.89750| = 167.9273$$

The secondary quantity-based factor NASDAQ weight is,

$$\frac{176.942 + 167.9273}{176.942} = 1.949052$$

And the secondary quantity-based factor Dow Jones weight is,

$$\frac{176.942 + 167.9273}{167.9273} = 2.053683$$

The elementary secondary quantity-based factor NASDAQ normalized weight is;

$$\frac{1.949052}{1.949052 + 2.053683} = 0.48693$$

And the elementary secondary quantity-based factor Dow Jones normalized weight is;

$$\frac{2.053683}{1.949052 + 2.053683} = 0.51307$$

The fuzzify variations of SQBFs NASDAQ and Dow Jones by Table 3.2. Variation of the NASDAQ of day 03-01-2018 is 0.836745% and variation of the Dow -Jones of day 03-01-2018 is 0.397478%. Through the elementary quantity-based factors NASDAQ and Dow Jones normalized weights 0.48693 and 0.51307 respectively. The Variation ($\mathrm{Var_s}$) of SQBF the NASDAQ and Dow Jones of day 03-01-2018 will be

$$\mathrm{Var_s} = \mathrm{Var_{NASDAQ}} \times 0.48693 + \mathrm{Var_{Dow\ Jones}} \times 0.51307$$

$$\mathrm{Var_s} = 0.836745\% \times 0.48693 + 0.397478\% \times 0.51307$$

$$\mathrm{Var_s} = 0.61137\%$$

Similarly, the variation of the NASDAQ for day 04-01-2018 is 0.175222% and variation of the Dow Jones for day 04-01-2018 is 0.611697%. Through the normalized weights 0.48693 and 0.51307 of the elementary quantity-based factors NASDAQ and Dow Jones respectively will before calculating of SQBFs as NASDAQ and Dow Jones of day 04-01-2018 variation ($\mathrm{Var_s}$) s.t.,

$$\mathrm{Var_s} = \mathrm{Var_{NASDAQ}} \times 0.48693 + \mathrm{Var_{Dow\ Jones}} \times 0.51307$$

$$\mathrm{Var_s} = 0.175222\% \times 0.48693 + 0.611697\% \times 0.51307$$

$$\mathrm{Var_s} = 0.399164\%$$

On applying a similar process as above for the calculation of ($\mathrm{Var_s}$) of each day to the secondary quantity-based factor variation see **Table 3.5**.

TABLE 3.5. The variation ($\mathrm{Var_t}$) of each day of secondary quantity-based factor (SQBF)

Date (t)	Variations (Var_t)
03-01-2018	0.61137%
04-01-2018	0.399164%
05-01-2018	0.855146%
⋮	⋮
29-10-2018	-1.30431%
30-10-2018	1.675315%
31-10-2018	1.478113%

The quantity-based secondary factor the NASDAQ and Dow Jones variation of 03-01-2018 is 0.616737% and this variation 0.616737% into the fuzzy variation i.e. fuzzified and it is a fuzzy set X_8. Similarly, the variation of 04-01-2018 is 0.39383% into the fuzzy variation i.e. fuzzified and it is a fuzzy set X_8. Table 3.6 is a list the fuzzy variation i.e. fuzzified of everyday QBSFs the NASDAQ and Dow Jones variation denoted by fuzzy set.

TABLE 3.6. Fuzzify into fuzzy variation denoted as fuzzy set

Date	Fuzzy Variations
03-01-2018	X_8
04-01/2018	X_8
05-01-2018	X_8
⋮	⋮
29-10-2018	X_6
30-10-2018	X_9
31-10-2018	X_9

If X_r is the fuzzy var of the QBMF at t days and X_s is the fuzzy var of the QBSF at t-1 days, then put the fuzzy variation "X_r" of the quantity based main factor of t day into the fuzzy var "Group X_s".

The fuzzy variation of the QBSF for day 03-01-2018 is X_8 and that the fuzzy variation of the main factor for day 04-01-2018 is X_8. Then it is located to the fuzzy variation X_8 of the QBMF for day 04-01-2018 into the fuzzy var group "Group X_8". See **Table 3.7**, i.e., of fuzzy var groups with respect to different fuzzy var of the QBSF.

TABLE 3.7. Fuzzy variations groups

Groups	**Fuzzy Variations**
Group X_3	X_3,X_6
Group X_4	X_8,X_7
Group X_5	X_6,X_6,X_7,X_6,X_6,X_8
Group X_6	$X_8,X_7,X_8,X_1,X_8,X_7,X_7,X_7,X_7,X_7,X_6,X_7,X_7X_6,X_7$
Group X_7	$X_7,X_8,X_8,X_8,X_6,X_7,X_7,X_8,X_{10},X_7,X_8,X_8,X_7,X_7,X_7,X_6,X_6,X_7,X_7,X_7,X_8,X_8,X_7,X_8,$ $X_8,X_7,X_9,X_7,X_8,X_8X_8,X_7,X_6,X_8,X_6,X_8,X_7,X_8,X_8,X_7,X_8,X_8,X_8,X_7,X_8,X_7,X_7,X_5,$ $X_8,X_7,X_7,X_7,X_8,X_7,X_7,X_7,X_7,X_9,X_7,$ $X_9,X_8,X_8,X_7,X_6,X_6,X_7,X_8,$ $X_6,X_7,$ $X_7,X_8,$ X_6
Group X_8	$X_8,X_8,X_8,X_7,X_7,X_8,X_8,X_8,X_8,X_8,X_7,X_7,X_8,X_8,X_9,X_7,X_7,X_8,X_8,X_8,X_8,X_7,X_8,X_8,$ $X_6,X_9,X_7,X_7,X_8,X_8,X_8,X_8,X_8,X_7,X_7,X_7,X_8,X_7,X_7,X_8,X_7,X_7,X_8,X_8,X_8,X_6,X_7,X_9,$ $X_7,X_7,X_6,X_7,X_9,X_7,X_7,X_8,X_7,X_8,X_7,X_8,X_8,X_6,X_8,X_8,X_7,X_7,X_8,X_8,X_8,X_8,X_8,$ $X_8,X_8,X_9,X_7,X_6,X_7,X_8,X_8,X_7,X_7,X_8,X_8,X_7,X_7,X_8$
Group X_9	$X_9,X_9,X_9,X_9,X_8,X_8,X_{10},X_8,X_7,X_8,X_8,X_8,X_8,X_8,X_8,X_8,X_8,X_8,X_8,X_8,X_8,X_8,$ X_8,X_8,X_8,X_8
Group X_{10}	X_5,X_8,X_9
Group X_{11}	X_9

In above Table 3.7, fuzzy variation group "Group X_7" are $X_7,X_8,X_{10},X_8,X_6,X_7,X_7,X_8,X_8,$ $X_7,$ $X_8,X_8,X_7,X_7,X_7,X_7,X_7,X_7,X_6,X_6,X_8,X_8,X_7,X_8X_8,X_7,X_9,X_7,X_8,X_8,X_8,X_7,X_6,X_8,X_6,X_8,X_7,X_8,X_8,$ $X_7,X_8,X_8,X_8,X_7,X_8,X_7,X_7,X_5,X_8,X_7,X_7,X_7,X_8,X_7,X_7,X_7,X_7,X_9,X_7,X_9,X_8,X_8,X_7,X_6,X_6,X_7,X_8,$ $X_6,X_7,$ $X_7,$ X_8,X_6. In the Group X_7, we can see $\mathcal{L} = 7$. The fuzzy variance appearing in the Group X_7. In the fuzzy variations X_M, when M is less than $\mathcal{L}$ then there are total number of fuzzy variations will be 10 (i.e., $X_{7,1} = 10$); the fuzzy variations X_M when $M = \mathcal{L}$ then total no. of fuzzy variations will be 31 (i.e., $X_{7,2} = 31$) and the fuzzy variation X_M when M is greater than $\mathcal{L}$ then total no. of fuzzy variations will be 31 (i.e., $X_{7,3} = 31$). Statistics of the fuzzy var appearing in every fuzzy var group in **Table 3.8**.

TABLE 3.8. Statistics of $X_{\mathcal{L},1}, X_{\mathcal{L},2}, X_{\mathcal{L},3}$ with respect to fuzzy variations appearing in each fuzzy variation Group X

	$\boldsymbol{X_{\mathcal{L},1}}$	$\boldsymbol{X_{\mathcal{L},2}}$	$\boldsymbol{X_{\mathcal{L},3}}$
$\mathcal{L} = 1$	-	-	-
$\mathcal{L} = 2$	-	-	-

$\mathcal{L} = 3$	-	1	1
$\mathcal{L} = 4$	-	-	2
$\mathcal{L} = 5$	-	-	6
$\mathcal{L} = 6$	1	2	12
$\mathcal{L} = 7$	10	31	31
$\mathcal{L} = 8$	37	44	5
$\mathcal{L} = 9$	21	4	1
$\mathcal{L} = 10$	3	-	-
$\mathcal{L} = 11$	1	-	-
$\mathcal{L} = 12$	-	-	-
$\mathcal{L} = 13$	-	-	-
$\mathcal{L} = 14$	-	-	-

Let if $W_{X_{7,1}}, W_{X_{7,2}}, W_{X_{7,3}}$ be the weights of $X_{\mathcal{L},1}, X_{\mathcal{L},2}, X_{\mathcal{L},3}$, respectively. Here we can see in a by **Table 3.8** $X_{7,1}$ =10, $X_{7,2}$ = 31, $X_{7,3}$ = 31

Weight $W_{X_{7,1}}$ will be,

$$\frac{10}{10+31+31} = 0.138889$$

Weight $W_{X_{7,2}}$ will be,

$$\frac{31}{10+31+31} = 0.430556$$

Weight $W_{X_{7,3}}$ will be,

$$\frac{31}{10+31+31} = 0.430556$$

See **Table 3.9** for the weights $W_{X_{\mathcal{L},1}}, W_{X_{\mathcal{L},2}}, W_{X_{\mathcal{L},3}}$ of $X_{\mathcal{L},1}, X_{\mathcal{L},2}, X_{\mathcal{L},3}$ under condition $1 \leq \mathcal{L} \leq 14$.

TABLE 3.9. Weights $W_{X_{\mathcal{L},1}}, W_{X_{\mathcal{L},2}}, W_{X_{\mathcal{L},3}}$ of $X_{\mathcal{L},1}, X_{\mathcal{L},2}, X_{\mathcal{L},3}$ respectively, under-condition $1 \leq \mathcal{L} \leq 14$

	k=1	**k=2**	**k=3**

$W_{X_{1,k}}$	0.0	0.0	0.0
$W_{X_{2,k}}$	0.0	0.0	0.0
$W_{X_{3,k}}$	0.0	0.5	0.5
$W_{X_{4,k}}$	0.0	0.0	1
$W_{X_{5,k}}$	0.0	0.0	1
$W_{X_{6,k}}$	0.066667	0.133333	0.8
$W_{X_{7,k}}$	0.138889	0.430556	0.430556
$W_{X_{8,k}}$	0.430233	0.511628	0.05814
$W_{X_{9,k}}$	0.807692	0.153846	0.038462
$W_{X_{10,k}}$	1	0.0	0.0
$W_{X_{11,k}}$	1	0.0	0.0
$W_{X_{12,k}}$	0.0	0.0	0.0
$W_{X_{13,k}}$	0.0	0.0	0.0
$W_{X_{14,k}}$	0.0	0.0	0.0

3.4.3. Forecasting Performance

For forecasting performance of trading days 01-11-2018 for TAIEX with the help of first order fuzzy logical relationships. We have an actual trading value is 9844.74 belongs to fuzzified into the fuzzy set β_5. Then the variation of NASDAQ and Dow Jones of trading day 01-11-2018 will be 1.754201% and 1.055037% respectively. The SQBF variation of NASDAQ and Dow Jones of 01-11-2018 will be,

$$\mathrm{Var_s} = \mathrm{Var_{NASDAQ}} \times 0.48693 + \mathrm{Var_{Dow\ Jones}} \times 0.51307$$

$$\mathrm{Var_s} = 1.754201\% \times 0.48693 + 1.055037\% \times 0.51307$$

$$\mathrm{Var_s} = 0.854173 + 0.541308 = 1.395481\%$$

The secondary quantity-based factor variation of NASDAQ and Dow Jones of 01-11-2018 is 1.395481% which belongs into fuzzy variation i.e., fuzzified as it is fuzzy set X_9. Therefore, select the fuzzy variation group "Group X_9". The testing data of the QBMF for day 01-11-2018 is fuzzified into the fuzzy set β_5 and there are no FLR appearing in the QBFLRG "Group X_9".

After that the forecasting value of the QBMF of testing data on the day 01-11-2018 will be 9850.

The secondary quantity-based factor variation of NASDAQ and Dow Jones of 02-11-2018 is -0.72699% fuzzified into fuzzy set X_7 and the testing data of the main quantity-based factor of day 02-11-2018 is fuzzified into the fuzzy set β_5. Then we select the fuzzy var group "Group X_7". In the Group X_7, we taken the quantity based fuzzy logical relationship "$\beta_6 \rightarrow \beta_6, \beta_6, \beta_6$" and we can see the weights $X_{7,1}$ = 0.138889, $X_{7,2}$ = 0.430556, $X_{7,3}$ = 0.430556. Because the min value and mid value and max value of the interval z_7 are 9900, 9950 and 10000 considered. Calculated weighted value z_7^* will be,

$$9900 \times 0.138889 + 9950 \times 0.430556 + 10000 \times 0.430556 = 9964.593$$

Finally, we can find the forecasted TAIEX of day 02-11-2018

$$\frac{9964.593 + 9964.593 + 9964.593}{3} = 9964.593$$

Similarly, we can find the forecast of TAIEX for each testing days of 2018. See **Table 3.10.** To actual TAIEX and the forecasted TAIEX of Nov-Dec, 2018 by using the quantity based secondary factor of NASDAQ and Dow Jones to assist the forecasting.

TABLE 3.10. Actual and forecasted TAIEX of Nov-Dec, 2018

Date	Actual Index	Forecast Index	Date	Actual Index	Forecast Index
01-11-2018	9844.74	9850.000	03-12-2018	10137.87	10150.00
02-11-2018	9906.59	9964.593	04-12-2018	10083.54	10050.00
05-11-2018	9889.81	9850.000	06-12-2018	9684.72	9650.000
06-11-2018	9824.95	9850.000	07-12-2018	9760.88	9800.000
07-11-2018	9908.35	9787.667	10-12-2018	9647.54	9650.000
08-11-2018	9945.31	9964.31	11-12-2018	9707.04	9764.593
09-11-2018	9830.01	9850.000	12-12-2018	9816.45	9850.000
12-11-2018	9831.21	9850.000	13-12-2018	9858.76	9764.593
13-11-2018	9775.84	9764.593	14-12-2018	9774.16	9800.000
14-11-2018	9791.88	9764.593	17-12-2018	9787.53	9800.000
15-11-2018	9826.46	9850.000	18-12-2018	9718.82	9764.593
16-11-2018	9797.09	9764.593	19-12-2018	9783.21	9764.593
19-11-2018	9828.69	9850.000	20-12-2018	9674.52	9650.000
20-11-2018	9743.99	9764.593	21-12-2018	9676.67	9650.000

21-11-2018	9741.52	9764.593	24-12-2018	9639.7	9650.000
23-11-2018	9667.3	9650.000	26-12-2018	9478.99	9531.405
26-11-2018	9765.36	9764.593	27-12-2018	9641.56	9650.000
27-11-2018	9778.62	9764.593	28-12-2018	9727.41	9764.593
28-11-2018	9884.31	9850.000	RMSE		37.80826
29-11-2018	9885.36	9850.000			
30-11-2018	9888.03	9850.000			

Graphical representation of **Table 3.10** for the actual and forecasted TAIEX of Nov-Dec, 2018 by using the quantity based secondary factor of NASDAQ and Dow Jones to assist the forecasting, **Figure 3.3**.

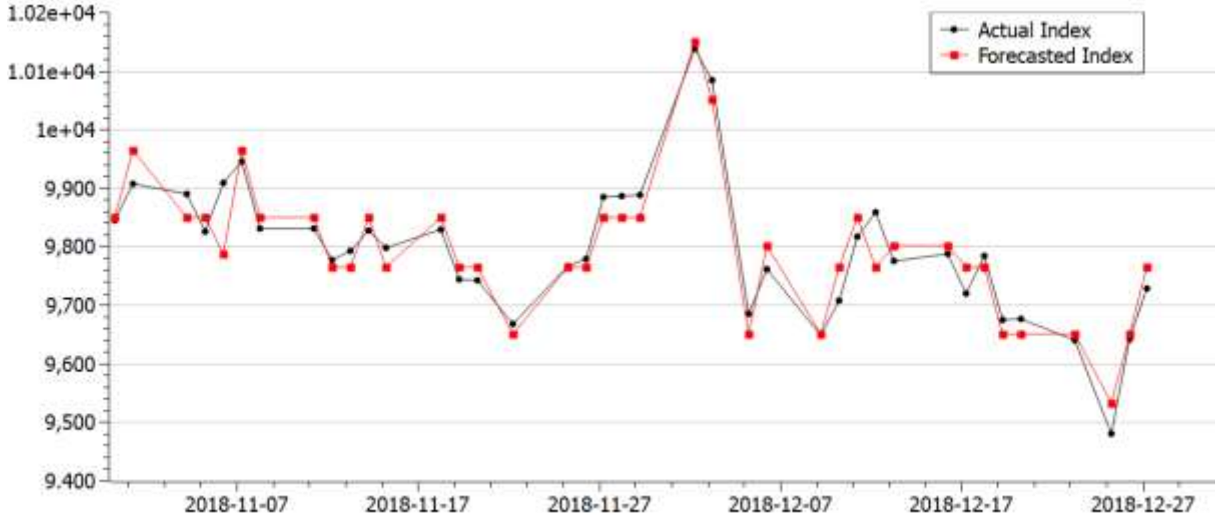


FIGURE 3.3. Representation of Actual TAIEX and the forecasted TAIEX, Nov- Dec, 2018

3.4.4. Result and Conclusion

FIGURE 3.4. Actual TAIEX index from 2004 to 2018

This model proposed a method to forecast of TAIEX, 2018, Nov-Dec, data. Figure 3.4 represent actual index from yahoo finance and also, we can see the range of year 2018 (proposed) and year 2004 ([3], [20]). It is clear that year 2018 range lies between the interval [9600-10200] and year 2004 range between intervals [5700-6200]. RMSE value depends on the number of days. If the number of days is increased then RMSE is decreases. For the RMSE of the proposed method apply;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(forecasted\ value_i - actual\ value_i\ )^2}{n}}$$

Where, n denote the number of days taken for forecasted. Where 2018 interval range is large compare to year 2004 interval range and volume of data is same. But calculated RMSE (Root mean square error) value of year 2018 is 37.8, smaller than year 2014 RMSE value, shown is table 3.11.

TABLE 3.11. Comparison of RMSE and Actual index reflection

Methods	**Range between intervals**	**RMSE**	**Years**
Proposed	9600-10200	37.8	**2018**
Chen [3]	5700-6200	57.73	**2004**
U_FTS Model by Huarng [20]	5700-6200	84	**2004**

U_R Model by Huarng [20]	5700-6200	146	**2004**
U_NN Model by Huarng [20]	5700-6200	60	**2004**
U_NN_FTS Model by Huarng [20]	5700-6200	116	**2004**
U_NN_FTS_S Model by Huarng [20]	5700-6200	116	**2004**
B_R Model by Huarng [20]	5700-6200	85	**2004**
B_NN Model by Huarng [20]	5700-6200	61	**2004**
B_NN_FTS Model by Huarng [20]	5700-6200	67	**2004**
B_NN_FTS_S Model by Huarng [20]	5700-6200	67	**2004**

3.5. METHOD 2

Let $\mathcal{Z} = z_1, z_2, z_3, \dots, z_m$ (Discourse the universe) and [7,16,20,21,29,33] fuzzy set β_τ in $\mathcal{Z}$ is defined as follows;

$$\beta_\tau = \frac{f_{\beta_\tau}(z_1)}{z_1} + \frac{f_{\beta_\tau}(z_2)}{z_2} + \frac{f_{\beta_\tau}(z_3)}{z_3} + \cdots + \frac{f_{\beta_\tau}(z_m)}{z_m}$$

Where f_{β_τ} is relationship function of β_τ (fuzzy set) and $f_{\beta_\tau}(z_j)$ represents the degree of connection of (z_j) belonging to β_τ.

$$f_{\beta_\tau}(z_j) \in [0,1], \quad 1 \le j \le m$$

Suppose, Y(t), t = …,-2,-1,0,1,2,3,… and be a subset of R. Let $f_\tau(t)$, where $\tau = 1,2,3,\dots$ be the fuzzy set, which is defined in the universe of discourse Y(t) and let F(t) be a set of $f_\tau(t)$ then the set F(t) is called FTS of Y(t) [11, 12, 16].

Definition VIII: [7,11,16] If a fuzzy relationship R(t, t+1) exists, such that $F(t+1) = F(t) \circ R(t, t+1)$ where " $\circ$ " denote the max- min composition operator, then F(t+1) is called the caused by F(t) and fuzzy relationship is,

$$F(t) \to F(t+1)$$

Both F(t+1), F(t) are fuzzy sets.

[21] Suppose F(t) and F(t+1) are β_τ and β_j, respectively. Relation between these two fuzzy sets is represented in the form "$\beta_\tau \to \beta_j$". β_τ and β_j are left-hand side and right-hand side of a FLR. FLR makes a FLRGs as follows;

$$\beta_\tau \to \beta_{j1}, \beta_{j2}, \dots, \beta_{jm}$$

Forecasting: [21] Suppose $F(t) = \beta_\tau$, and forecasting of F(t+1) is calculated by using the following rules;

Rule 1. if $\beta_\tau \rightarrow \beta_\tau$, then forecast of F(t+1) equal to β_τ.

Rule 2. if $\beta_\tau \rightarrow \beta_{j1}, \beta_{j2}, \dots, \beta_{jm}$, then forecast of F(t+1) equal to $\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}$.

Defuzzified matrix: [21] Defuzzified matrix for forecasted rule 1 is equal to the mid-point value of β_τ. i.e.

$$M(t) = [m_\tau]$$

and defuzzified matrix for forecasted rule 2 is equal to mid-point value of $\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}$. i.e.

$$M(t) = [m_{j1}, m_{j2}, \dots, m_{jm}]$$

Generate weight matrix: [21,42,55,58] Suppose the forecast of F(t+1) is $\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}$ then corresponding weights of these are $w_1, w_2, \dots, w_m$, respectively. Formulation of weight matrix W(t) for these weights $w_1, w_2, \dots, w_m$. i.e. $W(t) = [w_1', w_2', \dots, w_m']$ with satisfy a condition, $\sum_{h=1}^{m} w_h' = 1$. Other form i.e.

$$W(t) = [w_1', w_2', \dots, w_m'] = \left[\frac{w_1}{\sum_{h=1}^{m} w_h}, \frac{w_2}{\sum_{h=1}^{m} w_h}, \dots, \frac{w_m}{\sum_{h=1}^{m} w_h}\right]$$

$$= \left[\frac{1}{\sum_{h=1}^{m} h}, \frac{2}{\sum_{h=1}^{m} h}, \dots, \frac{m}{\sum_{h=1}^{m} h}\right]$$

Result calculation: [21,42,55] The final forecasted value F(t) is equal to the product of defuzzified matrix and transpose of the weight matrix:

$$F(t) = M(t) \times W(t)^T = [\mathrm{m}_{\mathrm{j1}}, \mathrm{m}_{\mathrm{j2}}, \dots, m_{\mathrm{jm}}] \times [w_1', w_2', \dots, w_m']^T$$

$$= [m_{j1}, m_{j2}, \dots, m_{jm}] \times \left[\frac{1}{\sum_{h=1}^{m} h}, \frac{2}{\sum_{h=1}^{m} h}, \dots, \frac{m}{\sum_{h=1}^{m} h}\right]^T$$

3.6. MODEL: FORECASTING OF TAIEX FOR YEAR 2018 AND 2019

Perform the QBFTS forecasting on quantity based fuzzy time series data/ Historical data of TAIEX [140]. Data of TAIEX divided into two parts i. e. training and perform. Training data is Jan to Oct and Perform data is Nov-Dec.

Universe of discourse $\mathcal{J}$, such as [9400, 11300] partitioned by length of each interval is 100 according to TAIEX data for year 2018, calculated by $\mathcal{J} = [D_{min} - D_1, D_{max} + D_2]$, $D_{max} =$ maximum value and $D_{min} =$ minimum value, for quantity-based TAIEX stock index historical data. Where D_1 and D_2 are proper positive real values to partition the universe of discourse $\mathcal{J}$ into m intervals $j_1, j_2, j_3, \ldots$, and j_m are equal in length [16]. Hence $\mathcal{J}$ divided into 19th intervals $\mathcal{J} = j_1, j_2, j_3, \ldots$, and j_{19}. i.e.

$$j_\tau = [9400 + (\tau - 1) * 100,\ \ 9400 + \tau * 100],\quad \tau = 1,2,3,\ldots,19$$

Frame the fuzzy set $\beta_\tau(\tau = 1,2,3,..,19)$ by the intervals, as follows;

$$\beta_1 = \frac{1}{j_1} + \frac{0.5}{j_2} + \frac{0}{j_3} + \frac{0}{j_4} + \frac{0}{j_5} + \frac{0}{j_6} + \frac{0}{j_7} + \cdots + \frac{0}{j_{17}} + \frac{0}{j_{18}} + \frac{0}{j_{19}}$$

$$\beta_2 = \frac{0.5}{j_1} + \frac{1}{j_2} + \frac{0.5}{j_3} + \frac{0}{j_4} + \frac{0}{j_5} + \frac{0}{j_6} + \frac{0}{j_7} + \cdots + \frac{0}{j_{17}} + \frac{0}{j_{18}} + \frac{0}{j_{19}}$$

$$\beta_3 = \frac{0}{j_1} + \frac{0.5}{j_2} + \frac{1}{j_3} + \frac{0.5}{j_4} + \frac{0}{j_5} + \frac{0}{j_6} + \frac{0}{j_7} + \cdots + \frac{0}{j_{17}} + \frac{0}{j_{18}} + \frac{0}{j_{19}}$$

$$\vdots$$

$$\beta_{18} = \frac{0}{j_1} + \frac{0}{j_2} + \frac{0}{j_3} + \frac{0}{j_4} + \frac{0}{j_5} + \frac{0}{j_6} + \frac{0}{j_7} + \cdots + \frac{0.5}{j_{17}} + \frac{1}{j_{18}} + \frac{0.5}{j_{19}}$$

$$\beta_{19} = \frac{0}{j_1} + \frac{0}{j_2} + \frac{0}{j_3} + \frac{0}{j_4} + \frac{0}{j_5} + \frac{0}{j_6} + \frac{0}{j_7} + \cdots + \frac{0}{j_{17}} + \frac{0.5}{j_{18}} + \frac{1}{j_{19}}$$

Daily historical data fuzzified into fuzzy sets. Fuzzy sets conduct the relation between every day that relation is known as quantity based fuzzy logical relationship (QBFLR). 2jan 2018 and 3jan 2018 data is fuzzified into β_{14} and β_{15}, respectively and quantity based fuzzy logical relationship is $\beta_{14} \rightarrow \beta_{15}$ between 2jan and 3jan. Similarly, quantity based fuzzy logical relationship between every day are shown in table 3.12.

Table 3.12. Quantity Based Fuzzy logical relationship Jan- Nov

Date	Quantity Based Fuzzy Logical Relationship
02 Jan → 03 Jan	$\beta_{14} \rightarrow \beta_{15}$
03 Jan → 04 Jan	$\beta_{15} \rightarrow \beta_{15}$
04 Jan → 05 Jan	$\beta_{15} \rightarrow \beta_{15}$
05 Jan → 08 Jan	$\beta_{15} \rightarrow \beta_{16}$
08 Jan → 09 Jan	$\beta_{16} \rightarrow \beta_{16}$
09 Jan → 10 Jan	$\beta_{16} \rightarrow \beta_{15}$
⋮	⋮

25 Oct → 26 Oct	$\beta_2 \rightarrow \beta_1$
26 Oct → 29 Oct	$\beta_1 \rightarrow \beta_2$
29 Oct → 30 Oct	$\beta_2 \rightarrow \beta_2$
30 Oct → 31 Oct	$\beta_2 \rightarrow \beta_5$

Quantity Based Fuzzy logical relationships are used to establish Quantity Based Fuzzy logical relationship groups (QBFLRGs) of every day data shown in table 3.13.

TABLE 3.13. Quantity Based Fuzzy Logical relationship groups

Group	**Quantity Based Fuzzy Logical Relationships**
Group β_1	$\beta_2 = 1 times$
Group β_2	$\beta_5 = 1 times, \beta_2 = 1 times\ , \beta_1 = 1 times$
Group β_4	$\beta_4 = 1 times\ , \beta_2 = 1 times$
Group β_6	$\beta_6 = 5 times\ , \beta_4 = 1 times$
Group β_7	$\beta_6 = 1 times$
Group β_{10}	$\beta_{11} = 1 times$
Group β_{11}	$\beta_{14} = 1 times, \beta_{12} = 2\ times, \beta_{11} = 1 times, \beta_5 = 1 times$
Group β_{12}	$\beta_{13} = 2 times, \beta_{12} = 3 times, \beta_{11} = 2 times, \beta_{10} = 1 times$
Group β_{13}	$\beta_{15} = 1 times, \beta_{14} = 6 times, \beta_{13} = 6\ times, \beta_{12} = 2 times$
Group β_{14}	$\beta_{15} = 10\ times, \beta_{14} = 10 times, \beta_{13} = 7\ times, \beta_{12} = 1 times$
Group β_{15}	$\beta_{17} = 1 times, \beta_{16} = 13 times, \beta_{15} = 21 times, \beta_{14} = 8 times$
Group β_{16}	$\beta_{18} = 1 times, \beta_{17} = 7 times, \beta_{16} = 19 times, \beta_{15} = 10 times, \beta_{14} = 2 times$
Group β_{17}	$\beta_{18} = 2 times, \beta_{17} = 19 times, \beta_{16} = 7 times, \beta_{15} = 1 times$
Group β_{18}	$\beta_{19} = 3 times, \beta_{18} = 8 times, \beta_{17} = 1 times, \beta_{16} = 1 times$
Group β_{19}	$\beta_{19} = 2 times, \beta_{18} = 2 times, \beta_{17} = 1 times$

From table 3.13, Forecasting of Group β_2 is $\beta_1, \beta_2, \beta_5$ and corresponding defuzzified forecast is $\mathrm{M(t)} = [9450, 9550, 9850]$. For that group weights matrix i.e.

$$W(t) = \left[\frac{1}{1+2+3}, \frac{2}{1+2+3}, \frac{3}{1+2+3}\right]$$

$$W(t) = \left[\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right]$$

Calculate the final forecast value. i.e.

$$Final\ (t) = M(t) \times W(t)^T$$

$$= [9450, 9550, 9850] \times \left[\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right]^T$$

$$= 9683.333$$

Similarly, we can find the forecast value of every day. i.e., represented in figure 3.5 with compression of actual and proposed value.

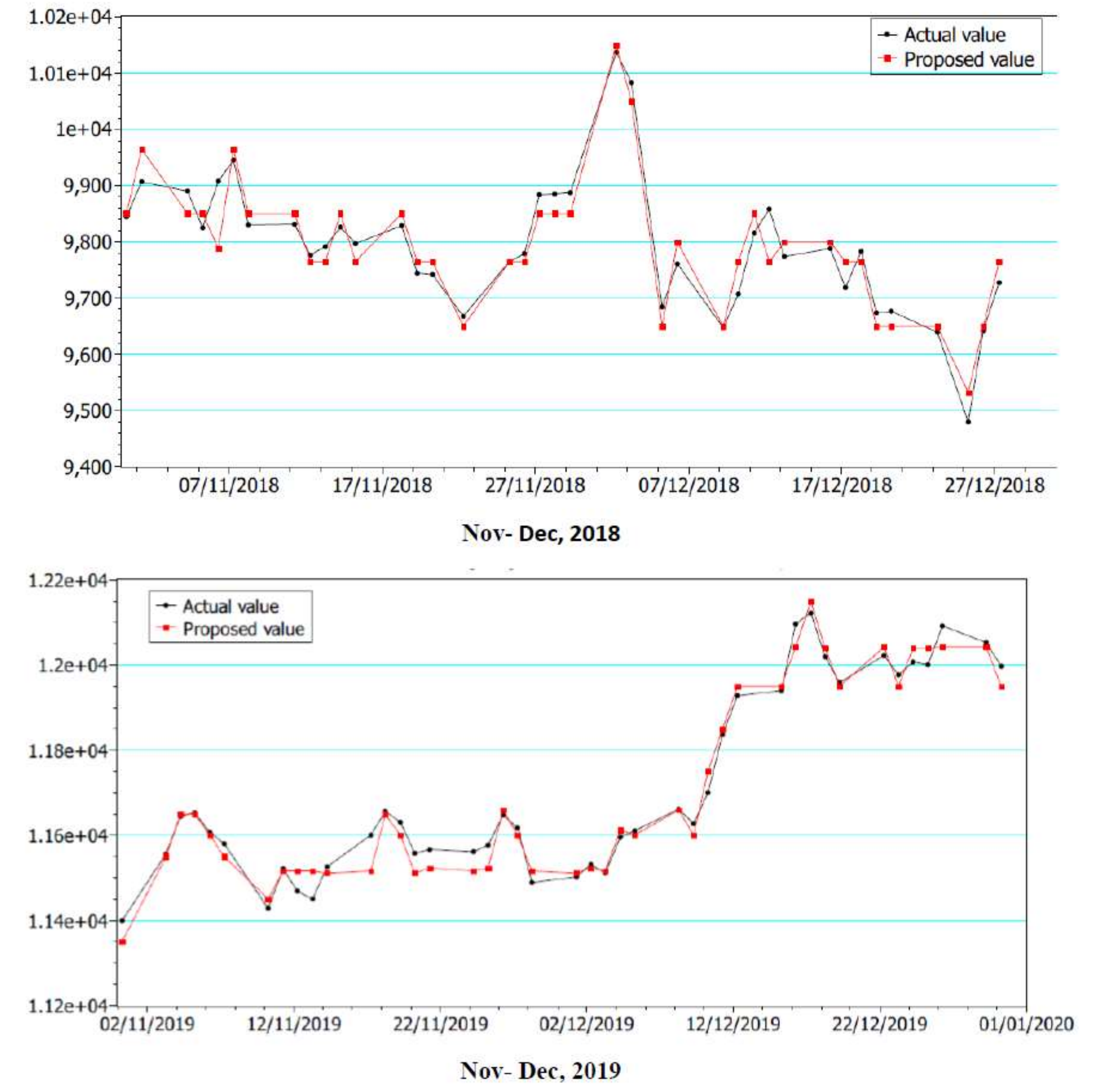


FIGURE 3.5. Actual and Proposed value comparison graph

3.6.1. Result and Conclusion

In consideration of TAIEX and funded the forecasted stock index value use of QBFTS, statistical weights and covered period 2018 and 2019. Also, discussed about RMSE of 2018 and 2019. Almost works covered the period 1999 to 2004 of TAIEX and get a higher RMSE value compare to proposed work see table 3.14 and figure 3.7. In figure 3.6 show the range of intervals period 2018 to 2019 between [9382.510, 12091.590] and period 1999 to 2004 between [3636.940, 10128.670]. According to that, the range is lower and RMSE is large value and in proposed work, the range is larger but RMSE is small value. Table 3.14. Shown that Chai et al. [66] used ACO and AR model for forecast then get the large RMSE value and proposed method get the min RMSE value. For measure performance used the RMSE,

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i=1}}^{\mathrm{n}}(\text{actual value}_i - \text{forecasted value}_i\,)^2}{\mathrm{n}}}$$

Where, n denote the number of days needed for the forecasted.

FIGURE 3.6. Actual Index from 1999 to 2019

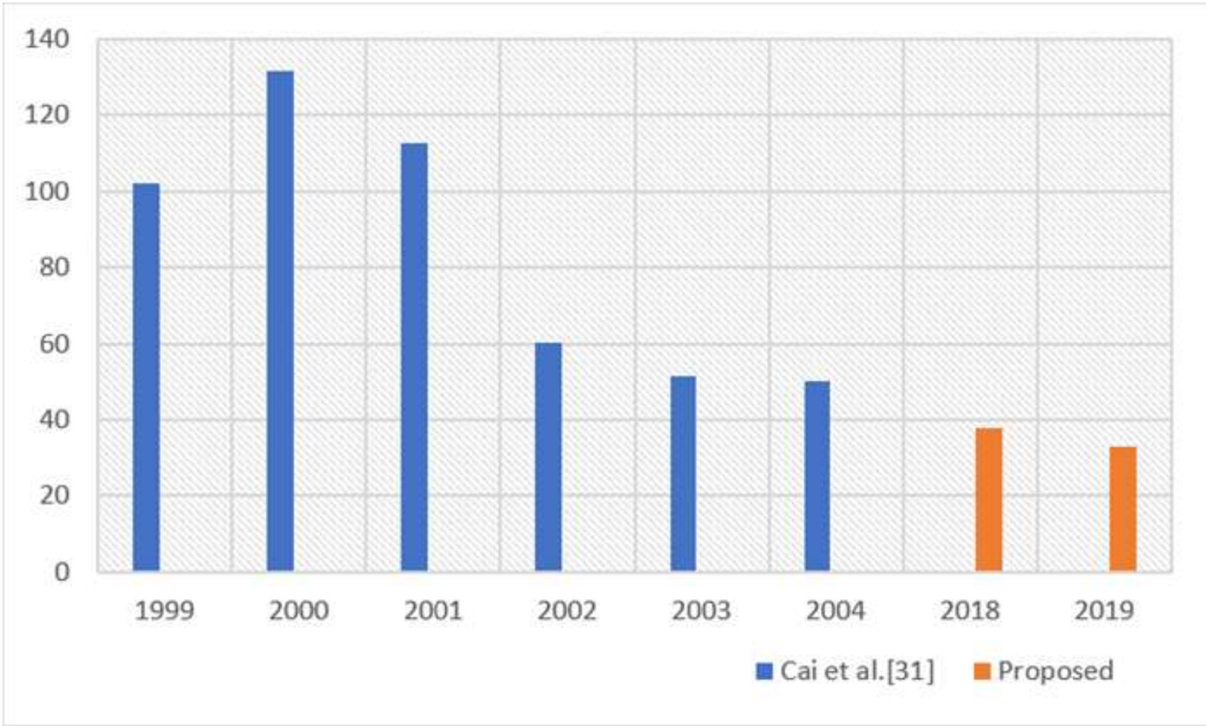


FIGURE 3.7. RMSE comparison graph

TABLE 3.14. Comparison of RMSE

	1999	2000	2001	2002	2003	2004	2018	2019
Cai et al. [66] used ACO & AR	102.22	131.53	112.59	60.33	51.54	50.33	-	-
Proposed	-	-	-	-	-	-	**37.80826**	**32.7927**

3.7. MODEL: FORECASTING OF INDIAN STOCK MARKET INDEX

Asian one of the most oldest stock market is Indian stock markets. BSE (Bombay Stock Market) Natco Pharma Limited (NATCOPHARM. BO) is one pharma index of BSE stock market index. In past years, many types of work done on stock index of different categories, where Pharma is one category of them. Natco pharma limited comes under this category. Time series discussed by several researchers, who are interested in prediction and applied several methods on pharma's stock index and other stock indexes.

In the proposed model, Quantity based Pharma stock market index (BSE) used for forecasting model purposes is Natco Pharma Limited (NATCOPHARM. BO), BSE stock market index done by quantity-based fuzzy time series. Use of different kinds of interval lengths for the universal set of stock market index data with statistical weighted system.

In pursuance of Quantity based close data of NATCOPHARM(BSE), [143] 2018-19 for forecasting model of NATCOPHARM 2019 (Nov-Dec). For these forecasting model processes, close data is divided into two partitions as training and testing. Close data partition as Jan 2018 to Oct 2019 and Nov- Dec, 2019 respectively training and testing.

For forecasting model, define universe $\mathcal{Z}$, as [450, 1000] with the length 50 of every interval, which are $z_1, z_2, z_3, \dots,$ and z_{11}. So, intervals are framed into fuzzy set $\beta_\tau(\tau = 1,2,3,..,11)$, as follows;

$$\beta_1 = \frac{1}{z_1} + \frac{0.5}{z_2} + \frac{0}{z_3} + \frac{0}{z_4} + \frac{0}{z_5} + \frac{0}{z_6} + \cdots + \frac{0}{z_{11}}$$

$$\beta_2 = \frac{0.5}{z_1} + \frac{1}{z_2} + \frac{0.5}{z_3} + \frac{0}{z_4} + \frac{0}{z_5} + \frac{0}{z_6} + \cdots + \frac{0}{z_{11}}$$

$$\beta_3 = \frac{0}{z_1} + \frac{0.5}{z_2} + \frac{1}{z_3} + \frac{0.5}{z_4} + \frac{0}{z_5} + \frac{0}{z_6} + \cdots + \frac{0}{z_{11}}$$

$$\vdots$$

$$\beta_{11} = \frac{0}{z_1} + \frac{0}{z_2} + \frac{0}{z_3} + \frac{0}{z_4} + \frac{0}{z_5} + \frac{0}{z_6} + \cdots + \frac{1}{z_{11}}$$

Every quantity-based training data is fuzzified into the fuzzy set. The NATCOPHARM 1jan 2018 close data value 948.5886 fuzzified into β_{10}. Similarly, 2jan 2018 close data value 929.1434 fuzzified into β_{10}.

In model, Fuzzified sets of close data build a relationship known as quantity-based fuzzy logical relationship. Build a quantity-based fuzzy logical relationship is $\beta_{10} \rightarrow \beta_{10}$ between 2jan and 3jan of 2018. In the same way, build a quantity-based fuzzy logical relationship of every-day training data, See table 3.15.

TABLE 3.15. Quantity based fuzzy logical relationship of every-day training data

Date	Quantity based Fuzzy Logical Relationship
01 Jan → 02 Jan	$\beta_{10} \rightarrow \beta_{10}$
02 Jan → 03 Jan	$\beta_{10} \rightarrow \beta_{10}$
03 Jan → 04 Jan	$\beta_{10} \rightarrow \beta_{10}$

⋮	⋮
29 Jan → 30 Oct	$\beta_3 \rightarrow \beta_3$
30 Jan → 31 Oct	$\beta_3 \rightarrow \beta_3$

Generate a quantity-based fuzzy logical relationship group of the quantity-based fuzzy logical relationship of every day training data shown in **Table 3.16**.

TABLE 3.16. Groups for quantity-based fuzzy logical relationships of every-day data

Group	**Quantity based fuzzy logical relationships**
Group β_1	$\beta_1 = 3times, \beta_2 = 3times$
Group β_2	$\beta_1 = 3times, \beta_2 = 99times, \beta_3 = 5times$
Group β_3	$\beta_2 = 5times, \beta_3 = 60times$
Group β_4	$\beta_3 = 1times, \beta_4 = 1times, \beta_5 = 3times$
Group β_5	$\beta_4 = 4times, \beta_5 = 54times, \beta_6 = 5times$
Group β_6	$\beta_5 = 6times, \beta_6 = 43times, \beta_7 = 12times$
Group β_7	$\beta_6 = 13times, \beta_7 = 100times, \beta_8 = 2times$
Group β_8	$\beta_7 = 2times, \beta_8 = 4times$
Group β_9	$\beta_7 = 1times, \beta_9 = 3times$
Group β_{10}	$\beta_9 = 1times, \beta_{10} = 6times, \beta_{11} = 2times$
Group β_{11}	$\beta_{10} = 2times, \beta_8 = 4times$

Weighted method

In a Group $\beta_S, 1 \leq S \leq m$ have different type quantity-based fuzzy logical relationship and it makes some conditions, i.e. M<S, M>S, and M=S. These conditions help to calculate $\beta_{S,k}$ where $1 \leq k \leq 3$.

Calculate the weights of $\beta_{S,1}, \beta_{S,2}, \beta_{S,3}$ is,

$$W_{\beta_{S,k}} = \frac{\beta_{S,k}}{\beta_{S,1} + \beta_{S,2} + \beta_{S,3}}, 1 \leq k \leq 3, 1 \leq S \leq 11$$

There are two conditions for forecasting model;

1. If the quantity-based factor testing data of t day is fuzzified into β_τ fuzzy set and QBFLRs exist in the QBFLG "Group β_j", then the forecasted value of the quantity-based factor testing data of t day is calculated as;

$$z_i^* = z_i^L \times W_{\beta_{S,1}} + z_i^M \times W_{\beta_{S,2}} + z_i^H \times W_{\beta_{S,3}}$$

Where, $W_{\beta_{L,3}}$ is the weight and $z_i^L, z_i^M and\ z_i^H$ are lower value, mid-value and higher value of the interval.

2. If quantity-based factor testing data of t day is fuzzified into β_τ fuzzy set and there is no QBFLRs in QBFLRG "Group β_j", then the forecasted value of the main quantity-based factor of testing data on t day is equal to mid-value (z_i^M) of the interval z_j.

$$z_i^* = z_i^M$$

The Forecasted model value of the quantity-based factor of testing t day is,

$$\text{Forecast value } (z_i^*) = \frac{\sum_{i=1}^{e} z_i^*}{e}$$

From the above **table 3.16**, Group β_3 $(S = 3)$ are 5times at 2 and 60times at 3. In variation β_τ when τ is less than S then quantity-based fuzzy logical relationship variation ($\beta_{3,1} = 5$), when τ is equal to S then quantity-based fuzzy logical relationship variation ($\beta_{3,2} = 60$), when τ is greater than S then quantity-based fuzzy logical relationship variation ($\beta_{3,1} = 5$).

TABLE 3.17. Statistics of quantity-based logical relationship in Groups

	$\beta_{S,1}$	$\beta_{S,2}$	$\beta_{S,3}$
S=1	-	3	3
S=2	3	99	5
S=3	5	60	-
S=4	1	1	3
S=5	4	54	5
S=6	6	43	12
S=7	13	100	2
S=8	2	43	-
S=9	1	3	-
S=10	1	6	2
S=11	2	11	-

With the help of Table 3.17. Weight $W_{\beta_{3,1}}$ will be,

$$\frac{5}{5+60+0} = 0.0769$$

Weight $W_{\beta_{3,2}}$ will be,

$$\frac{60}{5+60+0} = 0.9230$$

Weight $W_{\beta_{3,3}}$ will be,

$$\frac{0}{5+60+0} = 0$$

For the weights $W_{\beta_{S,k}}$ of $\beta_{S,k}$ under condition $1 \leq k \leq 3, 1 \leq S \leq 11$ see table 3.18.

TABLE 3.18. Weights $W_{\beta_{S,k}}$ of $\beta_{S,k}$ under condition $1 \leq k \leq 3, 1 \leq S \leq 11$

	k=1	**k=2**	**k=3**
$W_{\beta_{1,k}}$	-	0.5	0.5
$W_{\beta_{2,k}}$	0.0280	0.9252	0.0467
$W_{\beta_{3,k}}$	0.0769	0.9230	-
$W_{\beta_{4,k}}$	0.2	0.2	0.6
$W_{\beta_{5,k}}$	0.0634	0.8571	0.0793
$W_{\beta_{6,k}}$	0.0983	0.7049	0.1967
$W_{\beta_{7,k}}$	0.1130	0.8695	0.0173
$W_{\beta_{8,k}}$	0.3333	0.6666	-
$W_{\beta_{9,k}}$	0.25	0.75	-
$W_{\beta_{10,k}}$	0.1111	0.6666	0.2222
$W_{\beta_{11,k}}$	0.1538	0.8461	-

From table 3.18 weights $\beta_{3,1} = 0.0769, \beta_{3,1} = 0.9230, \beta_{3,1} = 0$ for Group β_3 and minimum, mid and maximum values of the interval z_3 are 550, 575, and 600 considered respectively. Calculated weights value z_3^* will be,

$$550 \times 0.0769 + 575 \times 0.9230 + 600 \times 0 = 573.943$$

The forecasted model value of NATCOPHARM. BO (BSE) of 1Nov, 2019 is,

$$\frac{573.943 + 573.943 + 573.943}{3} = 573.02$$

Similarly, find the forecasted model value of NATCOPHARM. BO (BSE) of Nov-Dec, 2019.

3.7.1. Result and Discussion

In the proposed work, applied quantity-based data to forecast model of the Natco Pharma Limited (NATCOPHARM. BO), BSE stock market index by quantity-based fuzzy time series, with different intervals lengths and weights method for that processes result are graphically represented of Nov-Dec, 2019 in **figure 3.8.**

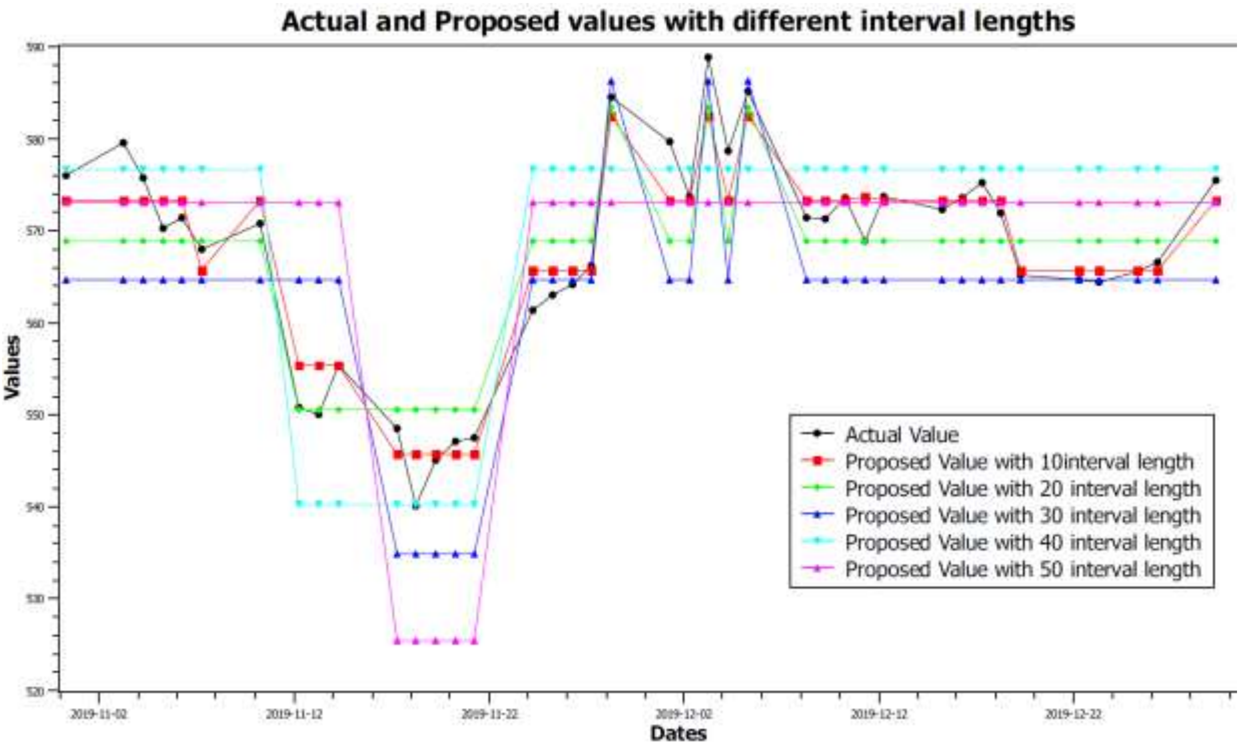


FIGURE 3.8. Actual and proposed value with different interval lengths by model.

For the accuracy of the graphically represented comparison of proposed work calculated RMSE and MSE values by,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(forecasted\ value - actual\ value)^2}{n}}$$

$$MSE = \sum_{i=1}^{n}(forecasted\ value - actual\ value)^2$$

This is shown in **table 3.19** for different length of intervals with statistical weighted methods, also assigned a rank for proposed work mean square error and RMSE into increasing order.

TABLE 3.19. RMSE and MSE values of proposed forecasting model

Intervals length	**No. of Intervals**	**RMSE**	**MSE**	**Rank**
10	51	3.014	9.0845	1
20	26	5.145	26.4756	2
30	17	8.581	73.6449	4
40	13	8.141	66.2759	3
50	11	10.998	120.9749	5

Proposed forecasting model gives **3.014** RMSE value when the length of the intervals is **10**. When we increase the length of the intervals then the RMSE value will also increase. Finally, we can conclude that, a greater number of intervals gave an accuracy and less Error which depends on lengths of intervals being lowest.

CHAPTER

IV

Application of Machine Learning in Fuzzy Time Series (FTS) Forecasting

In this chapter, some basic forecasting models such as Naïve, moving average, auto-regression model, ARIMA model, etc., are discussed and some of them are applied on different types of time series data set by machine learning. For the analysis of the forecasting used some software's as python 3 in Zupter notebook(updated) for the analysis.

4.1 INTRODUCTION

Machine learning can be applied to Fuzzy Time Series (FTS) forecasting to create more accurate predictions. Machine learning algorithms can be used to identify patterns in FTS data that may not be evident to human analysts, making them more effective at predicting future values. For example, a machine learning algorithm could identify which FTS variables are most influential in forecasting future values, allowing for more accurate predictions. Additionally, ML algorithms can be used to identify nonlinear relationships between FTS variables, which can improve the accuracy of predictions.

4.2 MACHINE LEARNING

ML enables the system the capability to automatically explore, enhance and improve from the different experiences without having to be explicitly programmed. It centers on the development of computer programs that can access data and use it to learn for themselves.

ML is powered by the diamond of statistics, calculus, linear algebra, and probability statistics are at the core of everything. Calculus tells us how to measure the rate of change of one variable with respect to another. Linear algebra helps us understand how to capture the relationship between various variables. Finally, probability allows us to measure the likelihood of different outcomes.

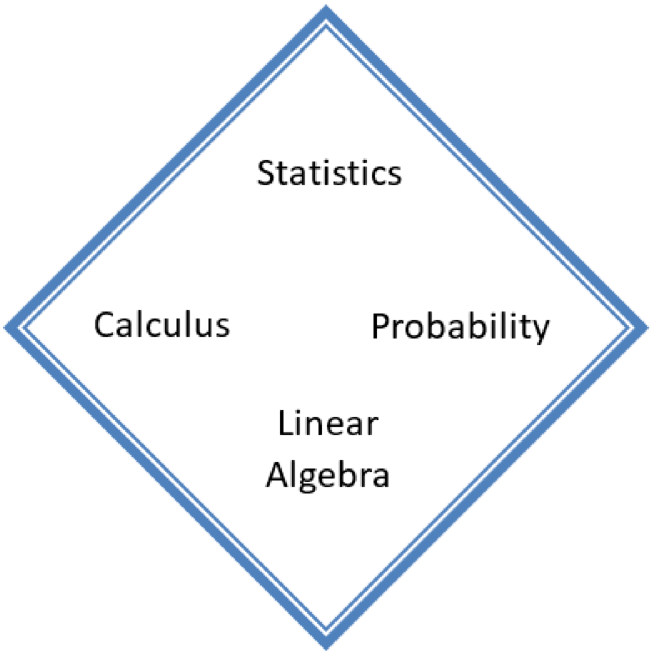


FIGURE 4.1. Component of Machine Learning

We make a set of attributes defined by an individual example. Sets of the attribute are known as features and variables also. Binary, numeric and ordinal can also represent these features. The performance metric is used for calculating the performance of machine learning.

In today's world, we are surrounded by lots of technologies. So, it is better to stay up to date with new emerging technologies.

Application of Machine learning in different sectors,

1. Health care

2. Sentiment analysis

3. Fraud detection

4. E-commerce

5. Oil and gas

6. Transportation

7. Marketing and Sales

8. Agricultural forecasting

4.3. DEVELOPMENT OF MACHINE LEARNING

Machine learning technology growing day by day in different sectors for analysis and prediction with the help of training data. So, training data is a key factor for machine learning. It tells us about the use of AI with a self-assured so also used in agriculture with a self-assured.

Before applying any data for prediction through machine learning need for some basic factors arises. To understand the factors for machine learning.

4.2.1 First Factor

Machine learning needs knowledge about around the world activities and tasks for training computers. It is helpful to explore themselves to educate them.

4.2.2 Second Factor

This factor is digital data or information collected and made accessible for the analytics process.

4.2.3 Third Factor

A recent one is the third factor, where digital changes were available for all technology-based environments and devices.

See one example of the latest technology. Technologies and deep learning algorithms are used on a drone to collect the data of crops and soil to monitor by software. And control the fertility of the soil by using the software. Some companies are developing robots and automation tools for agricultures fields to form effective ways to save a crop and also protect them from weeds. Agricultural spray machines were designed for spray accurate weeds on the plant and amount of pesticides for crops to save the crops from harmful diseases. There are many other technologies like cloud computing, IoT, Machine Learning, AI and Big data analytics are used in agriculture for better yield.

4.4. DEVELOPMENT OF DIFFERENT AREAS THROUGH MACHINE LEARNING

Machine learning is developing with technologies of big data and another fastest computer device. In the field of agriculture, machine learning is creating some new opportunities to understand the different types of data processes related to environmental function. It can be converted as the scientific formulation which will give the capability to learn without programming of the device for machines. It is used in different areas such as Biochemistry, Robotics, Medicines, Meteorology, Economic Sciences, Climatology, and Food Security.

4.5. MACHINE LEARNING METHODS

In ML agriculture, it gets to learn through agricultural processes to derive methods. In machine learning have those type of data set which depends on examples. An individual example is also used in examples of data sets. Characteristics of these sets are known as variables or helpers. These features can also describe as numerical, binary and features. The process is being calculated for the performance of ML from performance metrics.

The machine learning model obtains experience to time then improves performance. Some statistical and mathematical models are used to determine the performance of the machine learning model and machine learning algorithms. When the learning process is completed then the model may be used to classify and make the assumption, and to test data. It can be achieved after the training process will be completed.

Methods and applications of machine learning in different sectors are given below,

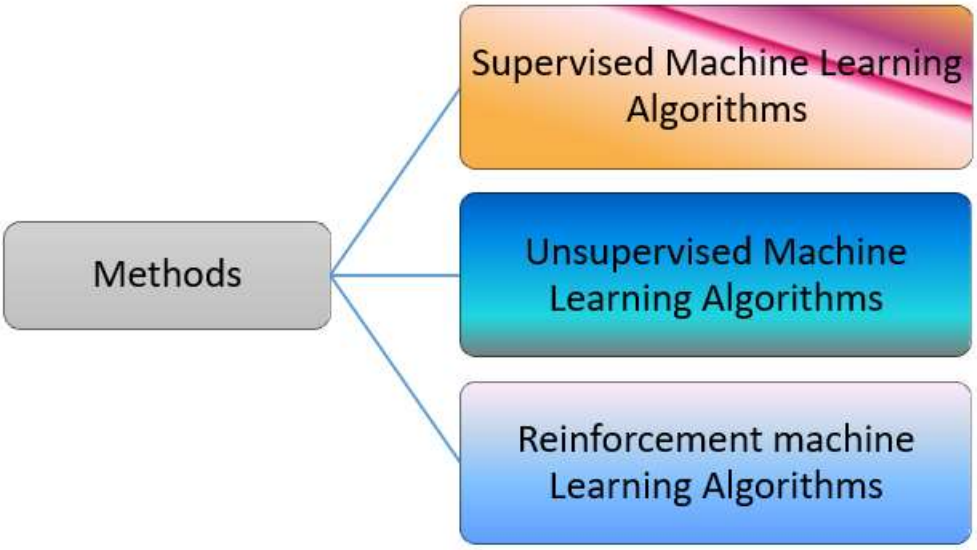


FIGURE 4.2. Methods of Machine Learning

4.5.1 Supervised machine learning Algorithm

Supervised learning is a method used to enable machines to classify/ predict object problems or situations based on labeled data fed to the machine.

Example

Suppose we take a jumble of data suppose circle, triangle, square labels of are in the labeled data we have a model training we know the answer. Very important what are you doing supervised learning, if already know the answer to the lots of the given information coming out. We have a huge couple of data coming and then you have new data coming out so then we trained the model. The model now knows the difference between circles, triangles, squares, and another we trained it. We can send square, circles go in to predict a top on the square and second on the circle.

It is used in agriculture. In the field of agriculture, there is huge data for prediction with some assumptions.

4.5.2 Unsupervised machine learning Algorithm

Un-supervised learning a machine learning model finds the hint haven the unlabeled data.

So, in the above example case what the circle is, what a triangle is, what a square is, a goes in looks as dimensions for a region it looks together preferred by the number of corners. Several models have three corners, numbers of models have two corners, numbers of models have one corner, and number of models have no corners and labels have a filter to true in together.

4.5.3 Reinforcement machine learning Algorithm

Here agent learns the property of behaviors to the environment by the performance of the act and checks the results of the action.

4.6. CASE STUDY

Our product machine learning algorithms as we know that Agriculture is an important part of GDP (Gross Domestic Product)**.** This study cussed in the advantage of insurance companies so they have efficient insurance coverage. We have taken two test data sets, one is 2csv files and another is image data set.

CSV file has lots of features like; temperature, humanity, pressure, and precipitation type like snow rains all that weather conditions, and also wind speed visibility, etc. Here we are predicting a crop field prediction. The second data set is the image data set which consists of 10 images of crop fields. Which were drone images we have manually counted the number of crops in each image. Here we using CNN regression to predict the contradict crop.

CountyNar	State	Latitude	Longitude	Date	apparentT	apparentT	cloudCove	dewPoint	humidity	precipInte	precipInte	precipProb	precipAccu	precipType	precipType	precipType	pressure	temperatu	temperatu	visibility	windBearin	windSpeed	NDVI	Day
Adams	Washingt	46.92984	-118.352	#######	18.61	-3.01	0	6.77	0.69	0	0	0	0	0	0	0	1027.95	23.93	6.96	10	9	3.8	136.1797	
Adams	Washingt	47.15033	-118.959	#######	19.67	-0.74	0	6.66	0.65	0	0	0	0	0	0	0	1028.26	25.88	8.71	10	352	6.03	135.6975	
Adams	Washingt	46.81169	-118.695	#######	20.66	-0.14	0	6.55	0.67	0	0	0	0	0	0	0	1028.29	24.67	8.26	10	25	3.59	135.677	
Adams	Washingt	47.16234	-118.7	#######	19.69	-2.66	0.03	7.32	0.69	0	0	0	0	0	0	0	1027.74	25.48	8.1	10	1	5.18	135.0058	
Adams	Washingt	47.15751	-118.434	#######	18.82	-3.04	0.04	7.62	0.7	0	0	0	0	0	0	0	1027.16	24.83	8.32	9.99	5	4.69	134.8039	
Adams	Washingt	47.00689	-118.51	#######	19.24	-2.58	0.04	7.45	0.7	0	0	0	0	0	0	0	1027.68	25.35	7.98	10	3	4.86	134.6575	
Alfalfa	Oklahoma	36.50822	-98.4543	#######	56.86	10.06	0.04	21.74	0.51	0	0	0	0	0	0	0	1015.05	56.86	24.21	9.95	352	9.71	150.4279	
Alfalfa	Oklahoma	36.6507	-98.4447	#######	56.32	9.57	0.04	20.59	0.49	0	0	0	0	0	0	0	1015.14	56.32	23.91	9.93	354	10.05	150.2179	
Alfalfa	Oklahoma	36.70245	-98.3199	#######	56.89	9.68	0.05	20.8	0.5	0	0	0	0	0	0	0	1015.13	56.89	24.06	9.93	352	10.01	150.1431	
Alfalfa	Oklahoma	36.62878	-98.5398	#######	56.15	9.32	0.04	20.1	0.49	0	0	0	0	0	0	0	1015.16	56.15	23.79	9.92	354	10.13	149.3406	
Alfalfa	Oklahoma	36.98881	-98.4046	#######	52.51	8.05	0.04	18.12	0.5	0	0	0	0	0	0	0	1015.43	52.51	22.64	9.93	0	11.89	148.1695	
Alfalfa	Oklahoma	36.99583	-98.3309	#######	52.53	8.14	0.04	18.33	0.5	0	0	0	0	0	0	0	1015.41	52.53	22.69	9.94	0	11.86	148.0441	
Allen	Kansas	37.82964	-95.5179	#######	58.43	12.48	0.08	32.08	0.7	0	0	0	0	0	0	0	1014.64	58.43	25.28	9.23	347	5.7	147.5343	
Allen	Kansas	37.93584	-95.297	#######	59.09	13.32	0.1	33.31	0.71	0	0	0	0	0	0	0	1014.73	59.09	25.75	9.24	344	5.21	146.8112	
Allen	Kansas	37.7755	-95.5177	#######	59.72	13.36	0.08	33.35	0.71	0	0	0	0	0	0	0	1014.96	59.72	25.9	9.22	344	5.04	146.53	
Allen	Kansas	37.8864	-95.403	#######	58.49	12.72	0.08	32.27	0.7	0	0	0	0	0	0	0	1014.68	58.49	25.5	9.26	346	5.53	145.9264	
Allen	Kansas	37.9449	-95.1587	#######	59.34	13.76	0.09	33.89	0.71	0	0	0	0	0	0	0	1014.74	59.34	26.09	9.27	341	4.91	145.4953	
Allen	Kansas	37.82396	-95.4433	#######	59.89	13.79	0.08	33.79	0.7	0	0	0	0	0	0	0	1014.34	59.89	26.28	9.26	342	4.96	144.9338	
Anderson	Kansas	38.07577	-95.1568	#######	59.59	13.67	0.09	34.23	0.72	0	0	0	0	0	0	0	1014.71	59.59	26.04	9.23	338	5.83	146.8112	
Anderson	Kansas	38.09794	-95.4483	#######	57.21	11.22	0.09	30.4	0.7	0	0	0	0	0	0	0	1015.02	57.21	24.28	9.19	347	6.74	146.4638	
Anderson	Kansas	38.15425	-95.0905	#######	58.79	12.39	0.08	32.29	0.72	0	0	0	0	0	0	0	1014.99	58.79	25	9.1	339	7.51	145.7837	
Anderson	Kansas	38.20991	-95.2243	#######	53.78	7.99	0.06	25.25	0.69	0	0	0	0	0	0	0	1015.5	53.78	21.91	8.9	351	10.46	143.9019	
Anderson	Kansas	38.34689	-95.1191	#######	54.04	8.31	0.05	24.62	0.68	0	0	0	0	0	0	0	1015.67	54.04	22.1	8.88	349	11.04	143.5907	
Anderson	Kansas	38.23728	-95.3387	#######	54.67	8.05	0.06	24.91	0.67	0	0	0	0	0	0	0	1015.5	54.67	21.87	8.99	352	10.25	143.0336	
Archer	Texas	33.40353	-98.5972	#######	78.75	32.37	0	50.37	0.63	0	0	0	0	0	0	0	1012.43	78.75	40.68	9.35	203	7.91	150.9683	
Archer	Texas	33.41	-98.9153	#######	77.99	28.38	0	47.08	0.61	0	0	0	0	0	0	0	1012.06	77.99	38.04	9.16	207	6.86	146.0862	
Archer	Texas	33.74978	-98.841	#######	77.55	23.83	0	42.95	0.58	0	0	0	0	0	0	0	1012.08	77.55	34.78	10	215	4.64	145.2708	
Archer	Texas	33.59957	-98.4543	#######	77.89	28.13	0	46.98	0.62	0	0	0	0	0	0	0	1012.11	77.89	37.64	10	207	6.14	144.9732	
Archer	Texas	33.48708	-98.9502	#######	77.86	27.78	0	46.56	0.6	0	0	0	0	0	0	0	1012.06	77.86	37.62	9.21	208	6.69	144.6725	
Archer	Texas	33.79193	-98.4531	#######	76.68	22.86	0	42.3	0.6	0	0	0	0	0	0	0	1012.35	76.68	34.15	10	212	4.25	142.6651	
Armstrong	Texas	35.05299	-101.295	#######	70.07	7.04	0	16.43	0.3	0	0	0	0	0	0	0	1011.37	70.07	22.14	10	305	4.45	148.5028	
Armstrong	Texas	35.17797	-101.491	#######	70.74	6.91	0	16.31	0.3	0	0	0	0	0	0	0	1011.31	70.74	22.13	10	303	4.55	148.4128	

FIGURE 4.3. CSV file having test data

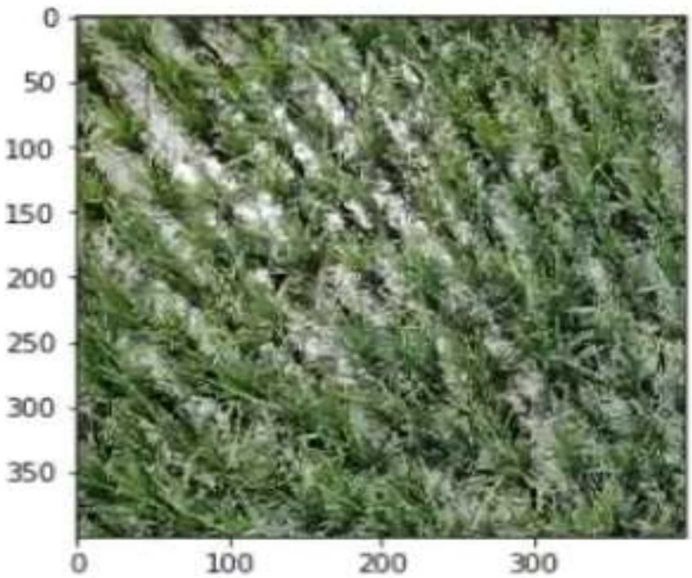


FIGURE 4.4. Crop development obtained by testing data

Tell about the pre-processing about the data set and different machine learning algorithms, we have used in our project. So, firstly in the pre-processing part, we have to input all the cetavariacle data using one odd in coding then we have to normalize the data using max scalar then again we have to again fill nine values by the mean values of that respective columns. Then we have split whole data set in the two parts, the first part is the training set, the second is testing set. Training set consisting the 70% of the data and the second is consists of 30% part of the data. And now coming to the different algorithms use. Firstly, we have used random forest then we have use support vector regression then we have used deep neural networks for regression.

To explain in detail the case study we have done we study the weather crops in the field that can be counted using the image of the set crop. We have to take 10 images using a drone. And have counted them manually several crops per image. It took about 2h to conclude this. We then took 8 of the images for training a model and 2 images for testing a model to see error rates. We have used to CNN Model and modify it for a predictive purpose by including a dancing layer with a linear activation function in the output layer. We normalized the images to 400×400 pixels with the length. So, feeding images into the model is easier. A model has about 14 hidden layers with a 25% drop out function. Since we do not have a large data set. We cannot aspect very large accuracy but we can predict a modulate level of accuracy based on a data set.

```
import keras
from keras.models import Sequential
from keras.layers import Dense, Dropout, Flatten
from keras.layers import Conv2D, MaxPooling2D
from keras.utils import to_categorical
from keras.preprocessing import image
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from tqdm import tqdm
%matplotlib inline
```

```
train = pd.read_csv('/content/drive/My Drive/Bennet/number_final.csv')    # reading the csv file
train.head()      # printing first five rows of the file
```

FIGURE 4.5. Model programming graph

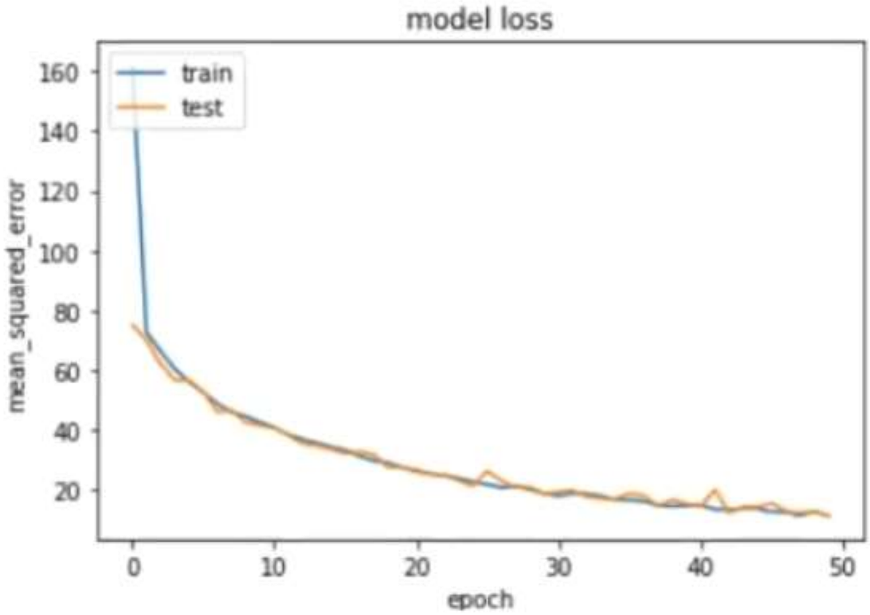


FIGURE 4.6. Model Test and Train graph

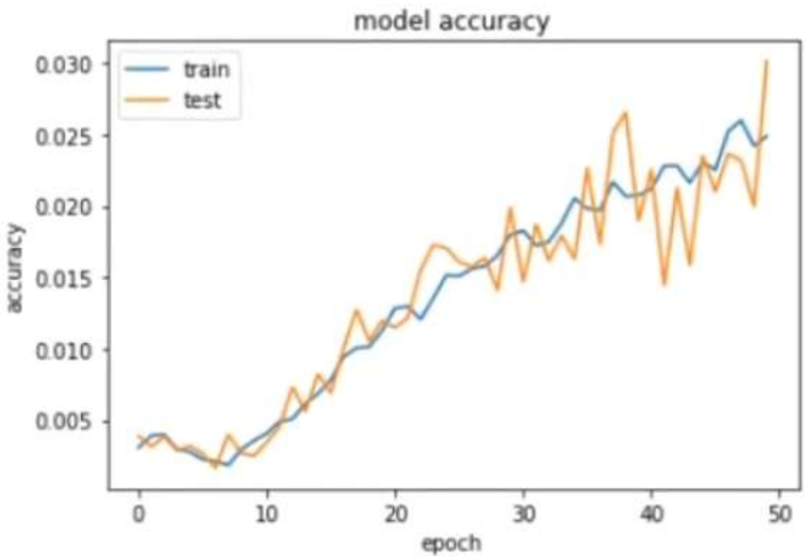


FIGURE 4.7. Model accuracy graph

Here we are using three methods,

1. Random forecast

2. Support vector regression and

3. Deep NN

We have used the parameter MAE, MSE, and are to comparing the different models, we have applied. So, as you can see the given war graph that is a good result is giving by deep NN is a mean absolute error is lowest and also the MSE is lowest.

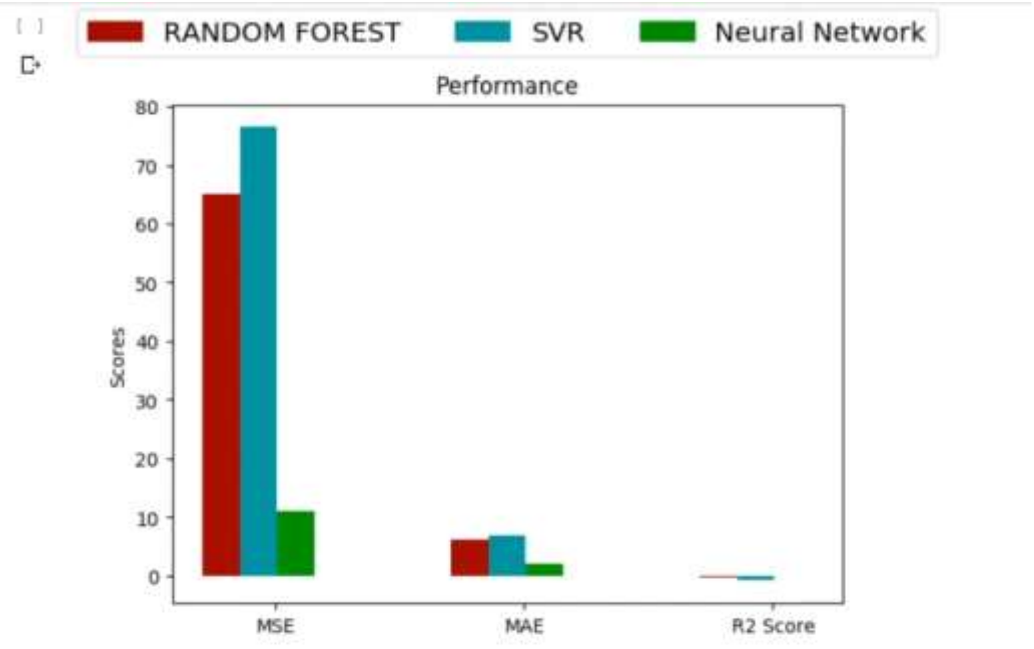


FIGURE 4.8. Performance of Methods

Now we will see the graph of the NN. As you can see with the including no. of approx the MSE and MAE are decreasing and the accuracy increasing with the increasing number of crops. Now we will come to the CNN case which we are used for counting the images.

```
plt.plot(history.history['mean_absolute_error'])
plt.plot(history.history['val_mean_absolute_error'])
plt.title('model mean_absolute_error')
plt.ylabel('mean_absolute_error')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
# summarize history for loss
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('model mean squared error')
plt.ylabel('mean_squared_error')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
```

FIGURE 4.9. Snap of Commands

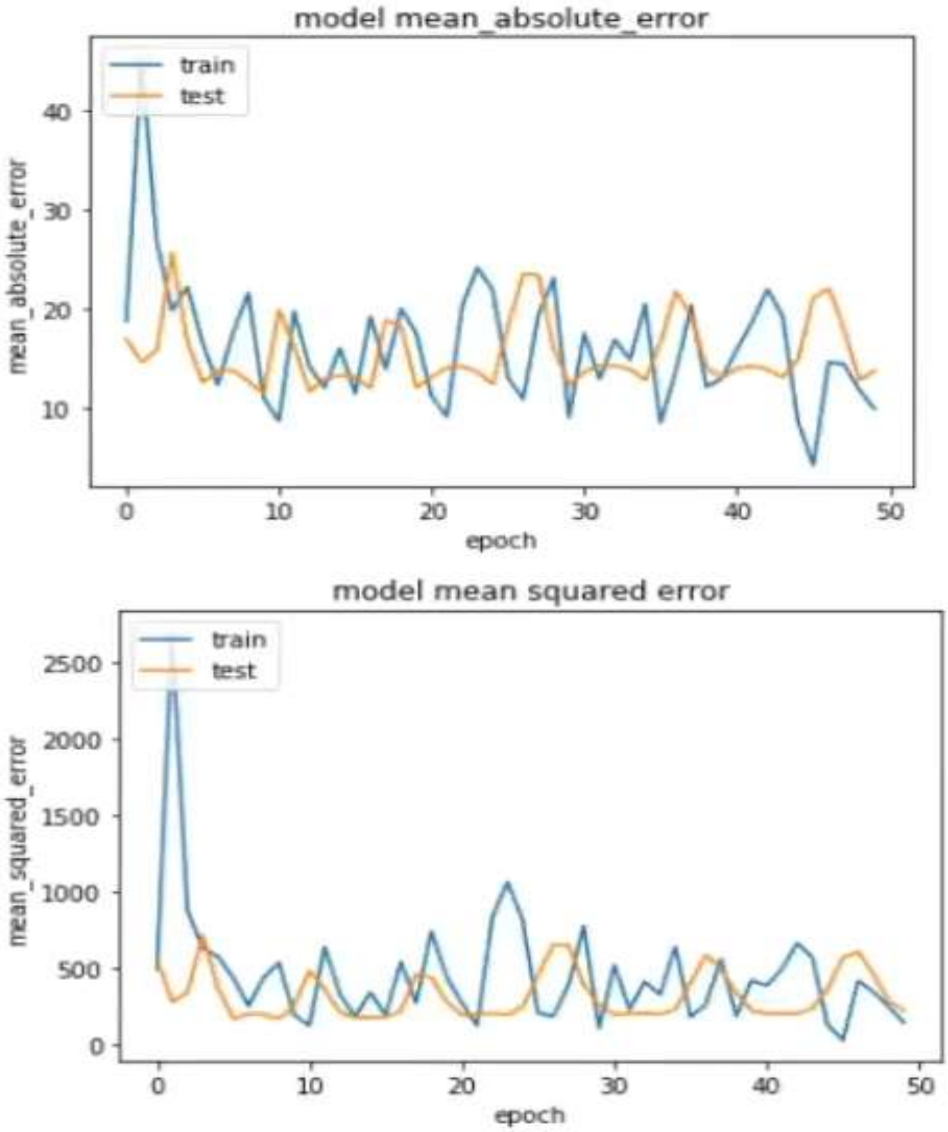


FIGURE 4.10. Error Performance

4.7. HYBRID MODEL FOR FORECASTING/ HYBRID MODEL FOR HEALTHCARE FORECASTING

4.7.1. Random Forest Algorithm

It is an ensemble learning algorithm that is used for both classification and regression tasks. It is a type of decision tree algorithm that creates multiple decision trees and then combines their results to make a more accurate prediction. Random Forest uses a technique called bagging (bootstrap aggregation) where multiple trees are created using different subsets of the data. The results of each tree are then averaged to make a more accurate prediction. Random Forest is a powerful and versatile algorithm that is used in many areas of Machine Learning. It is particularly useful for dealing with large datasets and can be used to identify the most important features in a dataset.

Also, the Random Forest algorithm applied on the analysis of forecasting of data, firstly calculate the entropy and create a decision tree. We can do this process by (i) Calculating the entropy of the data being analyzed, (ii) Creating a decision tree from the entropy, (iii) Splitting the data into training and test sets, and (iv) Evaluating the accuracy of the model using the test set.

Thus, to construct the decision tree algorithm following steps to follow:

Random Forest algorithm follows as;

Step I. Choose samples from the dataset.

Step II. For each sample, create a decision tree. Then get a forecasting result from it.

Creation of Decision Tree.

Step II. 1. Calculate the training set's Entropy (S) as follows:

$$Entropy\ (S) = -\sum_{i=1}^{k}\left\{\left[\frac{freq\left[\frac{freq(t_i,S)}{|S|}\right]}{|S|}\right] log_2\left[\frac{freq(t_i,S)}{|S|}\right]\right\}$$

Where $freq(t_i, S)$ is the number of samples included in class t_i and $|s|$ is the number of samples in t_i training set is a dependent variable, i = 1, 2....k, k is the number of classes of the dependent variable.

Step II. 2. Determine the Information Gain for test attribute X to partition:

$$Information\ Gain\ X(S) = Entropy\ (S) - \sum_{i=1}^{L}\left[\left(\frac{|s_i|}{|S|}\right) Entropy\ (s_i)\right]$$

Where $|s_i|$ is the number of dependent variables of subset s_i, number of test outputs is L, X, and s_i is a subset of S corresponding to i^{th} output.

Step II. 3. Calculate the partition information value X (Split info) acquiring for S partitioned into L subsets.

$$Split\ info(X) = \sum_{i=1}^{L}\left[\left(\frac{|s_i|}{|S|}\right)\log_2\left(\frac{|s_i|}{|S|}\right) + \left(1-\left(\frac{|s_i|}{|S|}\right)\right)\log_2\left(1-\left(\frac{|s_i|}{|S|}\right)\right)\right]$$

Step II. 4. Calculate the Gain ratio (X);

$$Gain\ ratio\ (X) = \frac{Information\ Gain_X(S)}{split\ Info\ (X)}$$

Step II. 5. The attribute with the highest gain ratio will be designated as the root node and the same calculations from step II.1- step II.4 are performed for every intermediate node until all instances are exhausted and this step II.2 According to reaches the leaf node.

Step III. Voting of the forecasted result.

Step IV. Final forecasting from the most voted forecasted result.

In the last of the year 2019, in the world Epidemic/Endemic/Pandemic disease like Covid-19 was effective to everyone, then everyone hopes on the healthcare as researcher who can make a treatment for control it. Then healthcare sector is very famous and most of the work started on it. It effects come on the healthcare stock also. Similarly, we also work on healthcare for forecasting of patient recovery. Forecasting of the patient recovery of covid-19, describe an (RF-DT) mathematical algorithm which is combination of Random Forest and Decision tress model.

For the RF-DT model for the forecasting of the patient recovery considered the data set/source file, which was collected from the Kaggle website [14]. In the data set, there are six types of parameters, i.e., observation date, province/state, country/region, confirmed, deaths, recovered cases. Python 3 used in Jupyter notebook and describe the actual data of COVID-19 and a summary of data with respect to Country/Region, and observation date as shown in Fig. 4.11, to Fig. 4.13, respectively.

	ObservationDate	Province/State	Country/Region	Confirmed	Deaths	Recovered
0	01/22/2020	Anhui	Mainland China	1	0	0
1	01/22/2020	Beijing	Mainland China	14	0	0
2	01/22/2020	Chongqing	Mainland China	6	0	0
3	01/22/2020	Fujian	Mainland China	1	0	0
4	01/22/2020	Gansu	Mainland China	0	0	0

FIGURE 4.11. Actual data of COVID-19

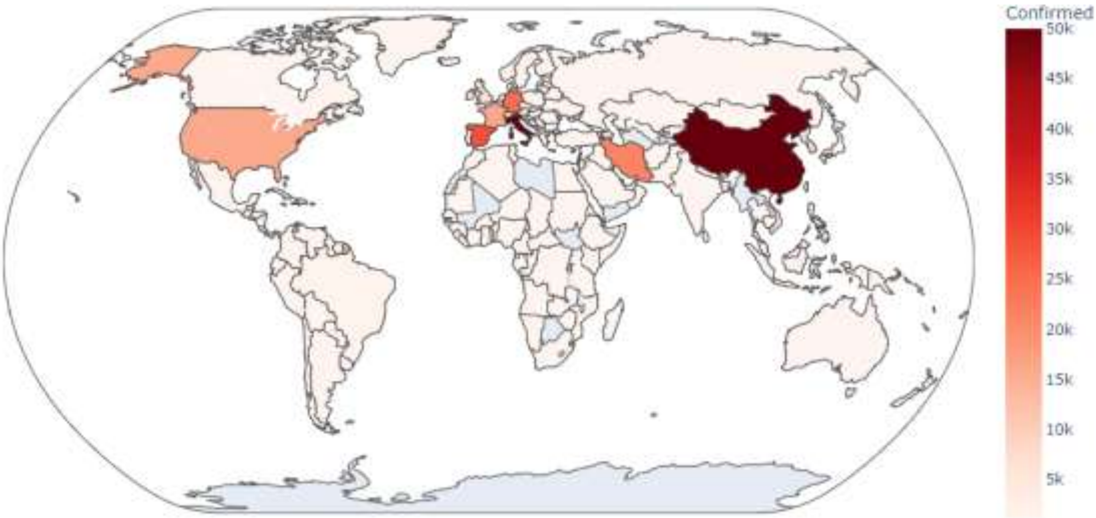


FIGURE 4.12. World map depicting confirmed cases

	Country/Region	Confirmed	Deaths	Recovered	Active
0	Mainland China	81060	3261	72252	5547
1	Italy	59138	5476	7024	46638
2	US	33276	417	178	32681
3	Spain	28768	1772	2575	24421
4	Germany	24873	94	266	24513
5	Iran	21638	1685	7931	12022
6	France	16044	674	2200	13170
7	South Korea	8897	104	2909	5884
8	Switzerland	7245	98	131	7016
9	UK	5741	282	67	5392
10	Netherlands	4216	180	2	4034

FIGURE 4.13. Summary of the data Country/Region wise with various parameters

4.7.2. Classical forecast parameter for preparation of data set

The forecasting process was completed through the python programming language. Python (Jupyter Notebook 6.4.8) along with requisite libraries are used to implement the proposed model. The RF-DT model was used for the patient recovery cases data. In python, Fig. 4.11 shows the actual data of COVID-19 with all attributes. Fig. 4.12 shows the world map of coronavirus depicting confirmed cases of various countries. Fig. 4.13 shows the summary of the dataset country-wise with various measures and for the statistical idea from the data of COVID-19, we used coding in python which shows the statistical summary of COVID-19 data as shown in Fig. 4.14.

	Confirmed	Deaths	Recovered
count	7926.000000	7926.000000	7926.000000
mean	655.640172	22.993187	237.621625
std	4978.076030	229.440512	2704.176057
min	0.000000	0.000000	0.000000
25%	2.000000	0.000000	0.000000
50%	16.000000	0.000000	0.000000
75%	130.000000	1.000000	10.000000
max	67800.000000	5476.000000	59433.000000

FIGURE 4.14. Statistical summary of covid-19 data

Statistical summary of data gives the value of standard deviation, mean, min, and maximum of the data for confirmed, deaths, and recovered cases. Counts values of data Country/Region-wise and null values of data calculated by python is shown in Fig. 4.15 and it is observed that the most frequently occurring country is Mainland China. The various parameters have been shown by region, by province, etc. Fig. 4.16 shows the summary of the data based on observation data with confirmed, deaths, recovered, and active cases parameters. Along with this, the graphical representation of these parameters has been shown in Fig. 4.18. Counts of the parameters are shown in Fig. 4.17 and the value lies from 0 to 3,00,000.

```
1  table['Country/Region'].value_counts().head()
```

```
Mainland China    1889
US                1617
Australia          323
Canada             254
France             127
Name: Country/Region, dtype: int64
```

```
1  table.isnull().sum()
```

```
ObservationDate       0
Province/State     3433
Country/Region        0
Confirmed             0
Deaths                0
Recovered             0
dtype: int64
```

FIGURE 4.15. Values counts in the form of Country/region and null values of data

	ObservationDate	Confirmed	Deaths	Recovered	Active
60	2020-03-22 00:00:00	335957	14634	97882	223441
59	2020-03-21 00:00:00	304528	12973	91676	199879
58	2020-03-20 00:00:00	272167	11299	87403	173465
57	2020-03-19 00:00:00	242713	9867	84962	147884
56	2020-03-18 00:00:00	214915	8733	83313	122869
55	2020-03-17 00:00:00	197168	7905	80840	108423
54	2020-03-16 00:00:00	181546	7126	78088	96332
53	2020-03-15 00:00:00	167447	6440	76034	84973
52	2020-03-14 00:00:00	156099	5819	72624	77656
51	2020-03-13 00:00:00	145193	5404	70251	69538
50	2020-03-12 00:00:00	128343	4720	68324	55299

FIGURE 4.16. Summary of data Date wise with parameters

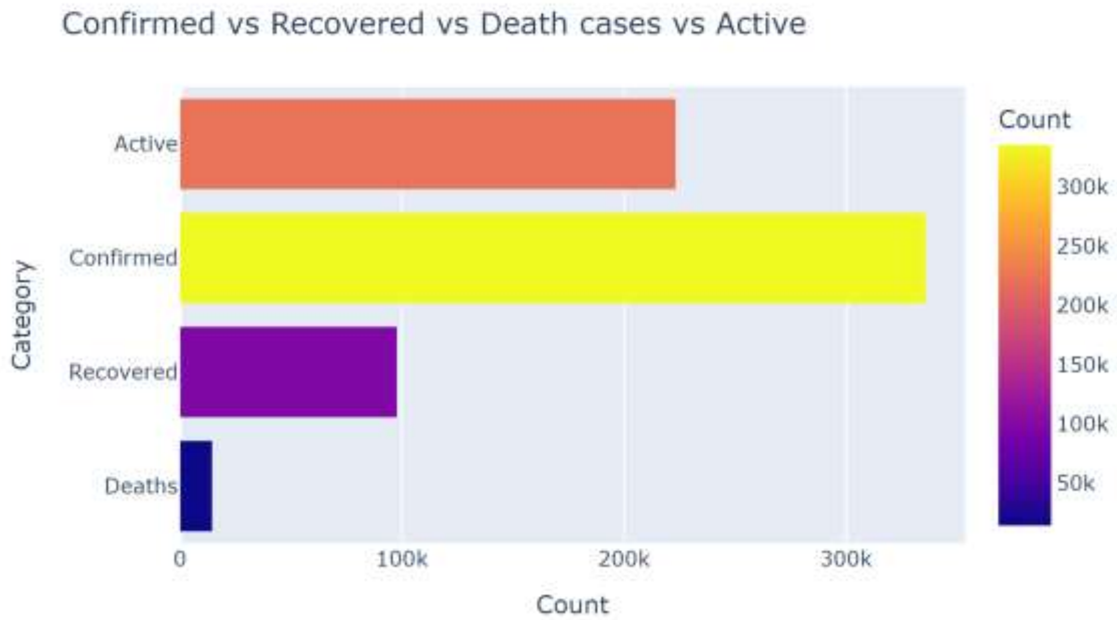


FIGURE 4.17. Counts the numbers parameters wise

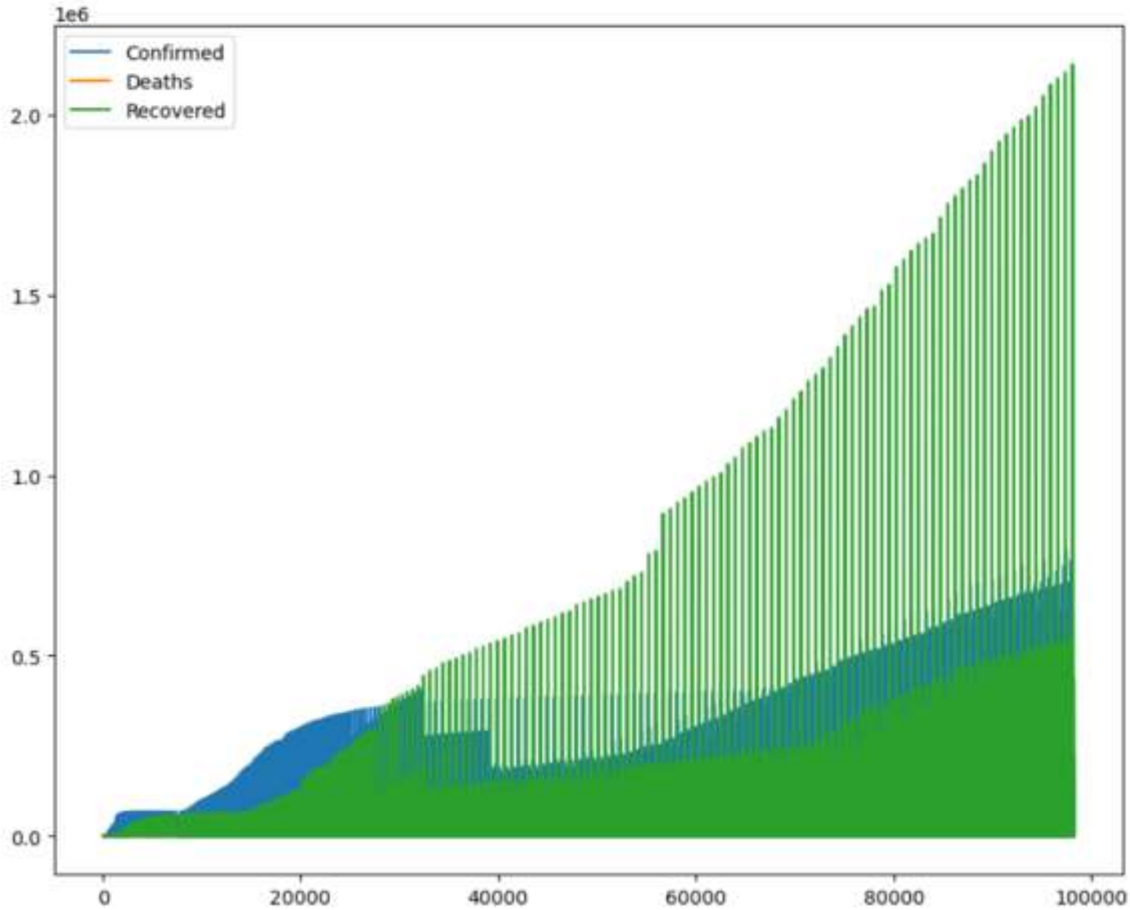


FIGURE 4.18. Graphical representation of COVID-19 dataset with different measures

Symptoms of COVID-19

According to the World Health Organisation and other health Agencies coronavirus is defined as the collection of various viruses whose symptoms can be from a mild cold to severe diseases. It is seen that it is the respiratory disease that increases the cases of COVID-19 rapidly. It has many symptoms but the most common symptoms of COVID-19 are sneezing, coughing, fever, respiratory problems, loss of sense, smell chest pain. Moreover, the given chart shows the details of symptoms in Fig. 4.19 and it is observed that fever, dry cough, and fatigue are the most common symptom.

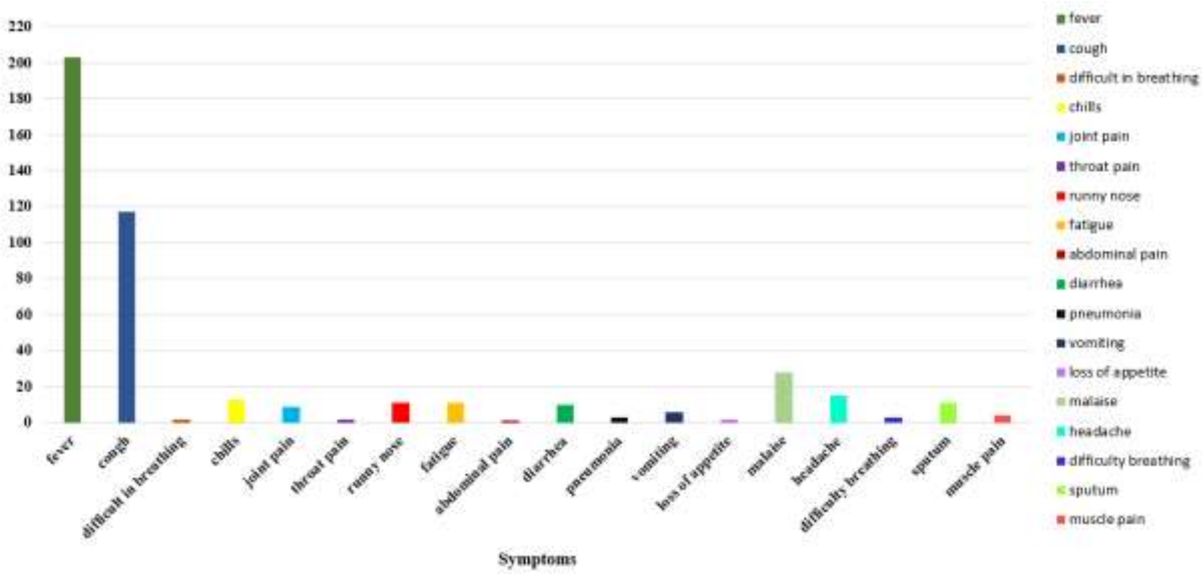


FIGURE 4.19. Graphical representation of the symptoms

4.7.3 Result and discussion

Used data mining algorithms such as the Random Forest model for Analysis and forecasting of the COVID-19 data. Used random forest model result various evaluation parameters were measured in terms of precision, accuracy, F-measure, recall, time taken to build a model, correctly classified instances percent and incorrectly classified instances percent.

However, accuracy is usually measured by the percentage of correct predictions made by the algorithm, and is calculated by

$$Accuracy = \frac{True\ positive\ +\ True\ negative}{True\ positive + True\ Negative + False\ Positive + False\ Negative}$$

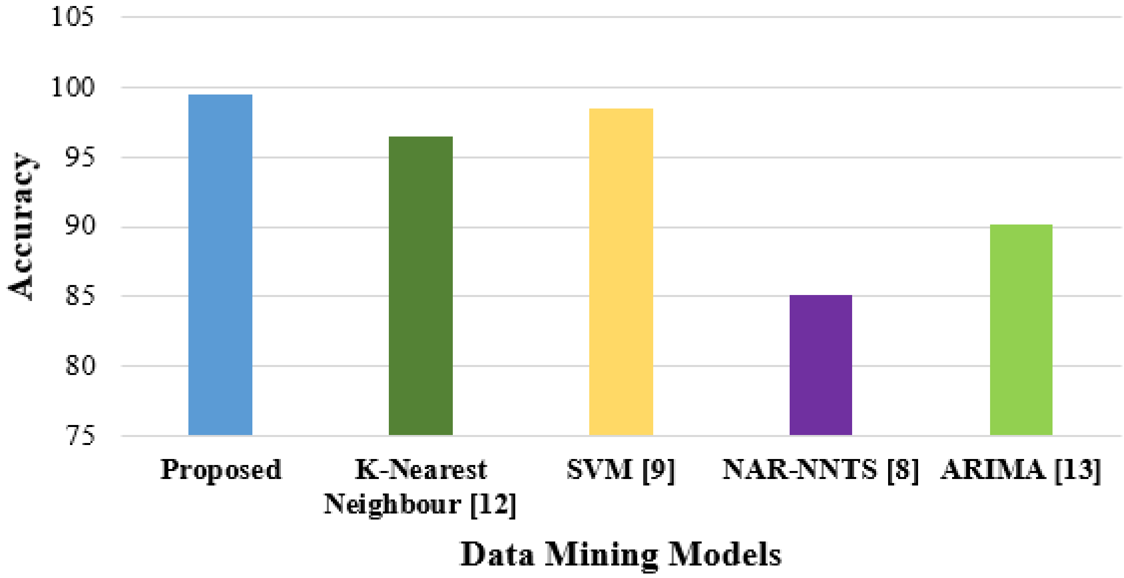


FIGURE 4.20. Proposed model performance with comparison to others model

In the proposed model, the percentage of accuracy will be 99.5%, which is the highest compared to other models such as K-Nearest Neighbour, SVM, NAR-NNTS, ARIMA models as 98.5%, 98.45%, 85.10%, 90.1%, respectively. Then we can conclude that the proposed model would be very useful and helpful in the healthcare sector such as COVID-19.

CHAPTER

V

Classical Methods/Model for Time Series Forecasting

5.1. INTRODUCTION

Time series analysis is a method of examining data over time to uncover patterns and trends. It can be used to study a variety of phenomena, such as economic trends, stock market fluctuations, population growth, and more. Time series analysis is often used to forecast future events and make decisions about investments, policy, and other decisions.

Time series can be divided into two types: stationary and non-stationary. Stationary series are those whose long-term statistical properties remain unchanged, while non-stationary series are those whose statistical properties change over time. Stationary series are usually easier to analyze and make predictions with, since the patterns in the data remain consistent. Non-stationary series are more difficult to analyze, since the patterns may change unexpectedly over time.

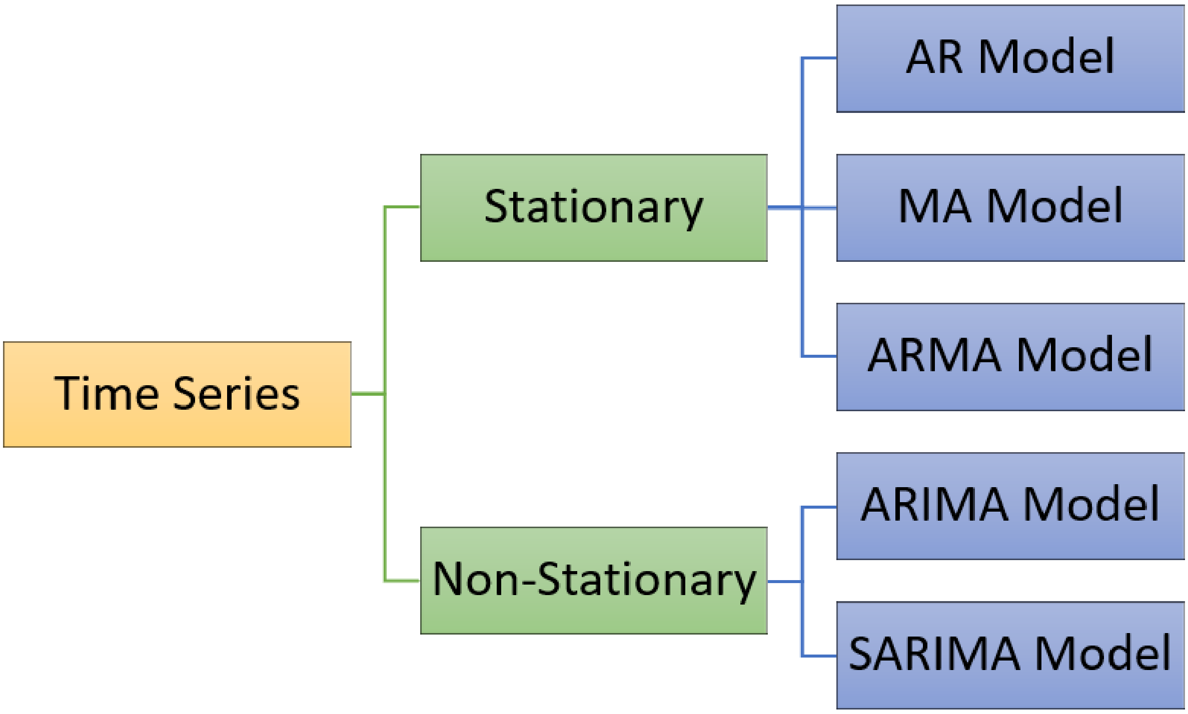


FIGURE 5.1. Types of Time Series

Classical methods for time series forecasting are methods that have been used for years to predict future values based on past values. These methods include ARMA model, ARIMA models, Holt-Winters seasonal decomposition, Vector Autoregressive (VAR) models, and Exponential Smoothing (ETS). Each of these methods uses different techniques to analyze and predict future values, and each has its own strengths and weaknesses. The choice of which model to use depends on the characteristics of the time series data being analyzed.

Classical methods for time series forecasting include Box-Jenkins ARIMA models, exponential smoothing, and regression models. ARIMA models are used to model the autocorrelation structure of the time series and extrapolate future values based on past values. Exponential smoothing methods use weighted averages of past values to forecast future values. Regression models capture the relationship between the time series and other relevant variables and use these relationships to make predictions. All three methods have been used extensively in time series forecasting and can be used together to obtain more accurate forecasts.

Classical methods for time series forecasting are a set of well-established techniques used to make forecasts about future values based on past data. These methods include ARIMA models, exponential smoothing, and Holt-Winters seasonal decomposition. ARIMA models are used to identify and quantify the relationships between past values and future values, while exponential smoothing and Holt-Winters decomposition are used to identify and quantify seasonal patterns. These methods are widely used by organizations and companies to gain insights into future trends and anticipate changes in demand, inventory, and prices.

Some classical methods for time series forecasting, where some are using and discussed by researchers in the past include:

- ❖ Autoregression (AR)
- ❖ MA (Moving Average)
- ❖ ARIMA: Autoregressive Integrated Moving Average
- ❖ Seasonal ARIMA
- ❖ Exponential Smoothing
- ❖ Holt-Winters Method
- ❖ Vector Autoregression (VAR)
- ❖ Neural Networks
- ❖ Support Vector Machines (SVMs)
- ❖ Long Short-Term Memory (LSTM)
- ❖ Prophet
- ❖ Structural Time Series (STS)

and etc, are also used for time series forecasting. Considering some classical methods here time series forecasting.

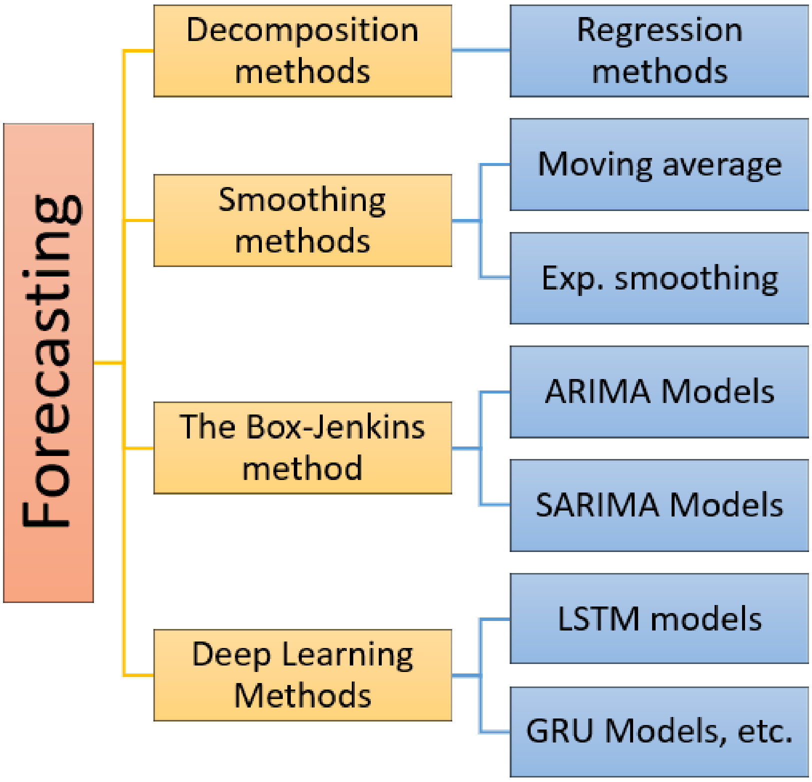


FIGURE 5.2. Classification of forecasting methods

5.2. WHEN TIME SERIES IS STATIONARY

Stationarity of a time series means that the mean, variance, and autocorrelation structure of the series does not change over time. A stationary time series is one whose statistical properties such as variance, mean, autocorrelation, etc. are constant over time. This implies that the trend component has been removed and the series is in equilibrium. Stationarity can be achieved by removing the trend component using techniques such as differencing, log-transformations, seasonal adjustment, and smoothing.

Tests such as the Augmented Dickey-Fuller, the Phillips-Perron, and KPPSS (Kwiatkowski-Phillips-Schmidt-Shin), are used to check the stationarity of time series. If the test statistic is less than the critical value, and say that time series is stationary. The autocorrelation function and partial autocorrelation function (PACF) are also used to identify stationarity. If the ACF and PACF have a sharp cut-off at lag 1 and the ACF decays exponentially to 0, then the time series is stationary.

5.3. MOVING AVERAGE

In a form of mathematics notation of moving average have two parts; moving + average.

Ordinary or Simple Moving average is data points of a series of numbers that are used to compute the average of a subset of numbers.

The moving average formula is as follows:

MA = (Sum of data points in the series) / (Number of data points in the series)

For example, if the series of numbers is [1, 2, 3, 4, 5], the simple moving average would be

(1 + 2 + 3 + 4 + 5) / 5 = 3.

For example, Also can calculate a 10-day or 20-day simple moving average, which is the average of the past 10- or 20- days' closing prices.

The moving average method is used for long-term trends of forecasting, which is a method of statistics for forecasting. Trends estimation of a trend at time t is found by an average of time series value for k period of t. Formulation of moving average of m order is,

$$T_t = \frac{1}{m} \sum_{i=-k}^{k} y_{t+i} \ , \ \ m = 2k + 1$$

It is also called the moving average of order m (m-MA).

If value of k is 2 then order is m = 2k+1 = 5 (five), similarly, k=3, m=7, and k=4, m=9, etc.

This method is also used for analyzing the stock market or other time-series data. This method is suitable for long-term trends and is used to smooth out short-term fluctuations. It is often used to identify trends and to make predictions about future values. It is a powerful tool used to smooth out data and reduce the noise from the data. It is also used to identify any seasonal patterns in the data. This method is used to eliminate any outliers and make the data more consistent and reliable.

The main advantages of moving average are; 1. Easy to construct and interpret. 2. It is easy to identify trends and seasonality. 3. It is simple to compute. 4. It is less affected by outliers. 5. Moving average smoothing is an efficient way to remove noise.

The disadvantages of moving average are; 1. It is based on a finite number of past values and so is not suitable for forecasting long-term trends. 2. It lags the current data and so may not be suitable for real-time forecasting. 3. It is sensitive to outliers and may not be able to capture sudden changes in the data.

5.4. DOUBLE MOVING AVERAGES

A moving average is a statistical measure of the average of a series of data points taken over a period of time. It is used to smooth out short-term fluctuations and make it easier to identify longer-term trends or cycles. In technical analysis, a double moving average (DMA) is a type of moving average indicator that uses the average of two separate moving averages.

Definition: A double moving average is a technical analysis tool used to identify near-term trends in a stock or commodity's price. It uses two different moving averages to generate trading signals. The first is a shorter-term moving average and the second is a longer-term moving average.

A double moving average is calculated by taking the average of two separate moving averages. The first is a short-term average, typically calculated over a period of days. The second is a longer-term average, typically calculated over a period of weeks.

The basic formula for a double moving average is:

Double Moving Average = (SMA1 + SMA2) / 2

Where SMA1 and SMA2 are the two simple moving averages. The double moving average is an average of the two simple moving averages and is used to smooth out short-term fluctuations in price and volume. It can be used to identify trends and also measure the strength of a trend.

Moving average of first order is

$$M_T = \frac{u_k + u_{k-1} + \cdots + u_1}{k}$$

Double moving average of first 'N' simple moving averages is

$$M_T^{(2)} = \frac{M_1^{(1)} + M_2^{(1)} + \cdots + M_N^{(1)}}{N}$$

Double moving average is obtained by adding (N+1)th simple moving average term and deleting $M_1^{(1)}$ term to the above equation then we get double moving average of order 'T+1'.

$$M_{T+1}^{(2)} = \frac{M_2^{(1)} + M_3^{(1)} + \cdots + M_{N+1}^{(1)}}{N}$$

Add and subtract $M_1^{(1)}$ term to numerator

$$M_{T+1}^{(2)} = \frac{M_2^{(1)} + M_3^{(1)} + \cdots + M_{N+1}^{(1)} + M_1^{(1)} - M_1^{(1)}}{N}$$

$$= \frac{M_1^{(1)} + M_2^{(1)} + M_3^{(1)} + \cdots + M_{N+1}^{(1)} - M_1^{(1)}}{N}$$

$$= M_T^{(2)} + \frac{M_1^{(1)} - M_1^{(1)}}{N}$$

This is formulae of recurrence for moving average as second order.

5.5. HIGHER ORDER MOVING AVERAGE (HOMA)

A higher order moving average (HOMA) is calculated by taking the average of n previous data points. This is done by adding up the n previous data points, then dividing the sum by n. The result is an average of the n previous data points. Higher order moving average are called 3rd order MA, 4th order MA, 5th order MA and so on. These types of moving averages use more than two data points to calculate the average. For example, if were to calculate a 5-day HOMA, the formula would be:

HOMA/ MA(5) = (x1 + x2 + x3 + x4 + x5) / 5

The resulting value is the average of the five previous data points.

Order of moving average	Notation	Formulae
3rd	$M_T^{(3)}$	$\frac{M_1^{(2)} + M_2^{(2)} + \cdots + M_N^{(2)}}{N}$
4th	$M_T^{(4)}$	$\frac{M_1^{(3)} + M_2^{(3)} + \cdots + M_N^{(3)}}{N}$
:	:	:
k-th	$M_T^{(k)}$	$\frac{M_1^{(k-1)} + M_2^{(k-1)} + \cdots + M_N^{(k-1)}}{N}$

5.6. MODEL OF MOVING AVERAGE

Given table 5.1, represented a time series stock market data on monthly bases of BSE healthcare, monthly close data values from Jan 2018 to Jul 2021 [143], where applied the moving average on data, calculate the moving average on the order (odd order) 3, 5, 7, 9. See figure 5.3, the order-changing effects of the moving average for BSE healthcare market index data.

TABLE 5.1. Actual data of BSE Healthcare Stock market index

Month	Close	Month	Close	Month	Close
Jan-18	14559.39	Apr-19	14367.02	Jun-20	16262.97
Feb-18	14113.01	May-19	13305.06	Jul-20	18284.76
Mar-18	13157.62	Jun-19	12889.34	Aug-20	18387.62
Apr-18	14153.59	Jul-19	12704.38	Sep-20	19799.24
May-18	13002.72	Aug-19	12875.4	Oct-20	19257.76
Jun-18	14003.64	Sep-19	12493.53	Nov-20	20318.54
Jul-18	14205.73	Oct-19	13229.05	Dec-20	21681.24
Aug-18	15945.17	Nov-19	13603.33	Jan-21	20628.71
Sep-18	15025.34	Dec-19	13429.11	Feb-21	20855.65

Oct-18	14726.58	Jan-20	13957.01	Mar-21	21328.21
Nov-18	14332.65	Feb-20	13480.1	Apr-21	23530.7
Dec-18	13923.37	Mar-20	12148.57	May-21	24534.98
Jan-19	13881.35	Apr-20	15332.39	Jun-21	25589.02
Feb-19	13760.65	May-20	15646.4	Jul-21	26156.18
Mar-19	14407.89				

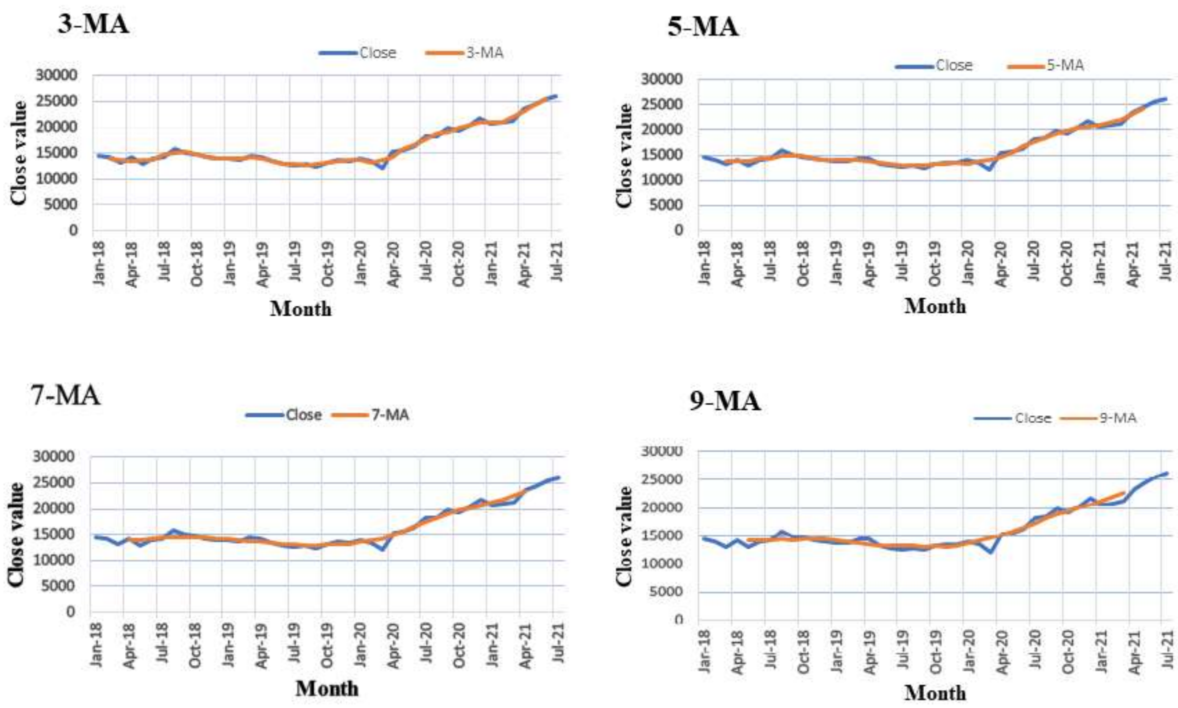


FIGURE 5.3. Different types of order moving average applied to the BSE healthcare market index data.

5.7. EXPONENTIAL SMOOTHING

This is a method of forecasting time-series data, such as sales figures or stock market prices, using weighted averages. In exponential smoothing, the most recent data points are given more weight than older data points. This technique is useful for predicting future trends based on past data and is commonly used in financial and economic forecasting.

The formula for exponential smoothing is as follows:

$$Forecasted\ Value = \alpha * Current\ Value + (1 - \alpha) * Previous\ Forecasted\ Value$$

where α is the smoothing factor. This is typically a number between 0 and 1, with higher values giving more weight to recent values.

Another form

$$F_t = \alpha * X_t + (1 - \alpha) * F_{t-1}$$

5.8. SINGLE EXPONENTIAL SMOOTHING

It is an approach used to forecast short-term demand in a time-series. It is a simple method that takes into account the most recent demand data and assigns exponentially decreasing weights to older data. This means that more recent demand data will have a higher weight when calculating the forecast than data from the past. It is also called a simple exponential smoothing. The formula used to calculate the forecast for single exponential smoothing is:

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$

Where, F_{t+1}= forecast for time point 't+1'

Y_t = actual time series observation at time 't'

F_t = forecast for time point 't'

α = smoothing constant (between 0 to 1).

5.9. AUTOREGRESSIVE MOVING AVERAGE FORECASTING MODEL

ARMA forecasting model is a combination of autoregressive and moving average models. It is model is used to represent the time series data that shows a combination of autocorrelated and non-autocorrelated noise. For forecast future values of a time series based on its past values by ARMA models.

The basic structure of an ARMA model is a linear regression model with lagged values of the variable itself as the independent variables. The lags are usually of two types: Autoregressive and Moving Average (MA) lags. AR lags are lags of the variable itself, while MA lags are lags of the errors of the model. The ARMA model can be written as follows:

$$Y_t = b_0 + b_1\ Y_{\{t-1\}} + b_2\ Y_{\{t-2\}} + ... + b_p\ Y_{\{t-p\}} + e_{\{t-1\}} + e_{\{t-2\}} + ... + e_{\{t-q\}},$$

where Y_t is the value of the time series at time t, b_0 is a constant, and b_1, b_2, ..., b_p are the autoregressive coefficients and $e_{\{t-1\}}$, $e_{\{t-2\}}$,..., $e_{\{t-q\}}$ are the moving average coefficients. The ARMA model can be used to analyze the effects of past observations on the current value of the time series, as well as the effect of random disturbances on the current value. The model can also be used to forecast future values of the time series.

The general AR(p) model was represented as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots . + \phi_p X_{t-p} + e_t \qquad (5.1)$$

Multiplying both sides by X_{t-k} in the equation (5.1)

$$X_{t-k}X_t = \phi_1 X_{t-k}X_{t-1} + \phi_2 X_{t-k}X_{t-2} + \phi_3 X_{t-k}X_{t-3} + \cdots . + \phi_p X_{t-k}X_{t-p} + X_{t-k}e_t \qquad (5.2)$$

Taking the expected value of both sides of equation (5.2) and assuming stationarity gives

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \phi_3\gamma_{k-3} + \cdots + \phi_p\gamma_{k-p} \qquad (5.3)$$

where γ_k is the covariance between X_t and X_{t-k}

The *MA(q)* model is written as

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} - \cdots - \theta_q e_{t-q} \qquad (5.4)$$

Multiplying both sides by X_{t-k} in equation (5.4)

$$X_{t-k}X_t = (e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} - \cdots - \theta_q e_{t-q})(e_{t-k} - \theta_1 e_{t-k-1} - \theta_2 e_{t-k-2} - \theta_3 e_{t-k-3} - \cdots - \theta_q e_{t-k-q}) \qquad (5.5)$$

The expected value of equation *(5.5)* will depend upon the value of k. if k=0, and all other terms of equation *(5.5)* drop out because by definition $E(e_t, e_{t+i})=0$ for $i\neq 0$ and $E(e_t, e_{t+i})=\sigma_e^2$ for $i=0$.

Thus *(5.5)* becomes

$$\gamma_k = \phi_1 E(X_t X_{t-k}) + \cdots + \phi_p E\big(X_{t-p}X_{t-k}\big) + E(e_t X_{t-k}) \qquad (5.6)$$

To obtain the initial estimates for ARMA models, combine AR md MA models:

$$\gamma_k = \phi_1 E(X_t X_{t-k}) + \cdots + \phi_p E\big(X_{t-p}X_{t-k}\big) + E(e_t X_{t-k}) - \theta_1 E(e_t X_{t-k}) - \cdots - \theta_q E\big(e_{t-q}X_{t-k}\big) \qquad (5.7)$$

If q<k. the terms $E(e_t X_{t-k}) = 0$ which leaves

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \phi_3\gamma_{k-3} + \cdots + \phi_p\gamma_{k-p}$$

If q>k, the past errors and the X_{t-k} will be correlated and the autocovariences will be affected by the moving average part of the process, requiring that it will be included.

The variance and auto-covariences of an ARMA(1,1) process are therefore obtained as follows:

$$X_t = \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1} \qquad (5.8)$$

Multiplying both sides of (5.8) by X_{t-k} gives

$$X_{t-k}X_t = \phi_1 X_{t-k}X_{t-1} + X_{t-k}e_t - \theta_1 X_{t-k}e_{t-1} \qquad (5.9)$$

Taking the expected values of (5.9) results in

$$E(X_{t-k}X_t) = \phi_1 E(X_{t-k}X_{t-1}) + \cdots + E(X_{t-k}e_t) - \theta_1 E(X_{t-k}e_{t-1})$$

If k = 0, this is

$$\gamma_0 = \phi_1\gamma_1 + E[(\phi_1 X_{t-1} + e_t - \theta_1 e_{t-1})e_t] - \theta_1 E\big[\big((\phi_1 X_{t-1} + e_t - \theta_1 e_{t-1})e_t\big)\big]e_{t-1} \qquad (5.10)$$

Since $X_t = \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1}$

$$\gamma_0 = \phi_1\gamma_1 + \sigma_e^2 - \theta_1(\phi_1 - \theta_1)\sigma_e^2$$

Similarly, if *k* = 1

$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_e^2 \qquad (5.11)$$

Solving the equations (5.10) and (5.11) for γ_0 and γ_1, get

$$\gamma_0 = \frac{1+\theta_1^2 - 2\phi_1\theta_1}{1-\theta_1^2} \qquad (5.12)$$

$$\gamma_1 = \frac{(1+\phi_1\theta_1)(\phi_1-\theta_1)}{1-\theta_1^2} \qquad (5.13)$$

Dividing (5.13) by (5.12) gives

$$\rho_1 = \frac{(1+\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} \qquad (5.14)$$

5.10. WHEN TIME SERIES IS NON-STATIONARY

Non-stationary time series are those which have time-dependent trends and/or variance. This means that the mean, variance and/or autocorrelation of the series changes over time. These time series can be caused by seasonality, changing demand, or a lack of appropriate data normalization. It can be made stationary by applying a transformation such as differencing, logging, or seasonal adjustment. Examples of non-stationary time series include stock prices, weather data, and economic data.

Non-stationary time series can be mathematically represented as follows:

$$X_t = X_{t-1} + \varepsilon_t$$

Where X_{t-1} tells the value of the time series at t-1 time, X_t tells the value of the time series at t time, and random error term is ε_t. This equation shows that the value of the time series at time t is determined by the value of the time series at time t-1, plus some random noise. This equation implies that the time series is non-stationary, since the value of the time series is dependent on the time period.

Non-stationary time series can be difficult to model since they are often highly dependent on time and can be unpredictable. To accurately model a non-stationary time series, it is often necessary to use sophisticated techniques such as ARIMA (Auto Regressive Integrated Moving Average) or GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models.

5.11. ARIMA (Auto Regressive Integrated Moving Average)

Such as everyone know that integration of two model as AR and MA is ARIMA model. Here, see the AR basic concept and MA basic concept then comes on the main topic ARIMA.

"ARIMA= AR (Integrate) MA"

For error reduction and data smoothing, moving average approaches are utilised. This smoothing has the effect of removing unpredictability, allowing this pattern to be forecasted and projected into the future. Moving average methods often employ arithmetic methods. Moving average methods are specifically used to measure trends by reducing data volatility. Extreme values have an impact on the arithmetic mean. Extreme values in this time series of data are impacted by random oscillations. Due to extreme observations, geometric mean is not impacted in the same way as arithmetic mean.

History


George E. P. Box Gwilym M. Jenkins

ARIMA was first developed in the 1950s by **George Box and Gwilym Jenkins**, two statisticians from the University of Wisconsin. The model was originally intended to be used for analyzing and forecasting the behavior of economic time series, such as unemployment, stock prices, and inflation.

Over time, the model has become increasingly popular, and is now widely used in many differ fields, like as economics, finance, engineering, and marketing.

ARIMA is a class of statistical models that incorporate both AR and moving average (MA) components. ARIMA models are used to analyze and forecast time series data. The parameters of an ARIMA model include the order of the autoregressive component (p), the order of the moving average component (q), and the degree of differencing (d). ARIMA models are used for forecasting or predicting future values of a time series. They can also be used to identify seasonality in the data and to test for stationarity. ARIMA models are particularly useful for financial time series data, such as stock prices, as well as for economic data, such as inflation and GDP.

If nonstationary is added to a combined ARMA process, then the general ARMA (p,d,q), model is implied. For example; ARIMA (1,1,1)

$$(1-B)(1-\phi_1 B)X_t = \mu' + (1-\theta_1 B)e_t$$

(i) the first difference (ii) the AR(1) part and (iii) the MA(1) aspect of the model to describe the backward shift operator. Multiplied and rearranged term as;

$$[1-B(1+\phi_1)+\phi_1 B^2]X_t = \mu' + e_t - \theta_1 e_{t-1}$$

$$X_t = (1+\phi_1)X_{t-1} - \phi_2 X_{t-2} + \mu' + e_t - \theta_1 e_{t-1} \quad (5.15)$$

As eqn. (5.15), model is similar to equation of conventional regression, and error terms are more than one.

ARIMA model: autoregressive model for p^{th} order

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \quad (5.16)$$

ARIMA model: moving average model for q^{th} order

$$Y_t = \mu + \varepsilon_t - W_1\varepsilon_{t-1} - W_2\varepsilon_{t-2} - \cdots - W_q\varepsilon_{t-q} \quad (5.17)$$

ARIMA model: ARMA(p,q) model

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - W_1 \varepsilon_{t-1} - W_2 \varepsilon_{t-2} - \cdots - W_q \varepsilon_{t-q} \quad (5.18)$$

ARIMA (0,1,1) model

$$Y_t - Y_{t-1} = \varepsilon_t - W_1 \varepsilon_{t-1} \quad (5.19)$$

Forecasting equation in exponential smoothing for ARIMA (0,1,1) model

$$\hat{Y}_{t+1} = Y_t - W_1(Y_t - \hat{Y}_{t-1}) = (1 - W_1)Y_t + W_1 \hat{Y}_t \quad (5.20)$$

ARIMA modelling is controlled by four steps; model evaluation, parameter estimation, diagnostics, and forecasting. Identifying if the time series dataset is seasonal and stationary is the first stage in this time series model. If a time series' statistical characteristics remain constant, it is stationary. The stationarity of the dataset is an important observation to make to obtain accurate forecasts. The unit root test is used to determine the stationarity of a time series. If the series is not stationary, differences are used to make the data stationary. ACF graphs and partial autocorrelation function (PACF) correlograms can be used to estimate ARIMA model parameters. The graph of the auto correlation function determines the relationship between previous and subsequent values in a time series. The partial auto correlation function graph computes the degree of correlation between lag and variable. We can estimate the best ARIMA model using maximum likelihood estimation (MLE). Once the best model for the time series data set has been chosen, the ARIMA model can be used as a forecasting model to predict future values using those parameters.

Auto Correlation Function (ACF): The auto-correlation function (ACF) is a mathematical tool used in statistics and econometrics to measure the linear relationship between two variables. It measures the degree of similarity that lies in a time series & a lagged version of itself.

In mathematics, the auto-correlation function (ACF) is defined as:

$$ACF(h) = \frac{\text{Cov}(X_t, X_{(t-h)})}{\text{Var}(X_t)}$$

Where:

$\text{Cov}(X_t, X_{(t-h)})$ is the covariance between two variables X_t and $X_{(t-h)}$,

$\text{Var}(X_t)$ is the variance of the variable X_t

The ACF can be plotted as a graph, which shows the correlation between an observation and a number of lagged observations. This can be helpful in identifying patterns in a time series and also in identifying any seasonality in the data.

Partial Auto Correlation Function (PACF): A summary of the relationship between an observation in a time series and observations at past time steps is called a partial autocorrelation. This means that the relationships between the observations that occurred in the intervening time steps have been removed. After removing the effect of any correlations that may have been caused by the terms at shorter lags, the correlation that remains is referred to as the partial autocorrelation at lag k.

5.12. BOX -JENKINS FORECASTING MODEL

The Box-Jenkins forecasting model is a statistical model used for time series analysis and forecasting. The Box-Jenkins model is widely used in economics, finance, and other fields, and has been successfully applied to many different types of data.

The Box-Jenkins model is a type of autoregressive integrated moving average model. It is used to forecast data with a high degree of accuracy and is based on the assumption that future values of a given variable can be estimated from prior values. The model uses a combination of autoregression, moving average, & differencing to identify patterns within the data and make predictions.

The Box-Jenkins model is a three-step process:

1. Identification: The first step is to identify the appropriate model for the data. This involves examining the data and examining the ACF and Partial ACF functions to determine the type of model that best fits the data.

2. Estimation: It is second step to estimate the model parameters. This is done by using the maximum likelihood method.

3. Diagnostics: The third and last step is to check the model for any errors or misspecifications. This is done by examining the residuals of the model. Once the model has been identified, estimated, and checked, it can be used to make forecasts.

5.13. ACCURACY MEASURES OF FORECASTING TECHNIQUES

Accuracy measures are used to measure the effectiveness of forecasting techniques. They help measure the accuracy of predictions made by a forecasting technique and can be used to compare different forecasting methods. Accuracy measures are used to quantify the differences between the forecasts made by a forecasting technique and the actual results. The most commonly used accuracy measures are the mean absolute error, mean squared error, root mean squared error, etc,. These measures are used in literature to compare the forecasted values with the actual values and quantify the differences between them.

Suppose, X_i be the actual data for time period i and F_i be the forecast or fitted value for the same period, then be forecast error or residual for the time period. *i* is defined as

$$e_i = X_i - F_i$$

Some important absolute and relative measures of accuracy are given by,

1. **Mean absolute error (MAE)** measures the average difference between the actual and forecasted values. This measure is expressed in the same units as the data being forecasted. A lower MAE indicates a more accurate prediction.

   Mean Absolute Error (MAE) $= \frac{\sum_{i=1}^{n}|e_i|}{n}$

2. **Mean squared error (MSE)** is like MAE, but it takes into account the magnitude of the differences between the actual and predicted values. This measure is expressed in the same units as the data being forecasted, but it is also squared. A lower MSE indicates a more accurate prediction.

   Mean Square Error (MSE) $= \frac{\sum_{i=1}^{n}|e_i^2|}{n}$

3. **Root mean squared error** is like MSE, but it is expressed in the same units as the data being forecasted. A lower RMSE indicates a more accurate prediction.

   $$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(e_i)^2}{n}}$$

   n represent the total number of data points.

4. Percentage Error $PE_i = \left[\frac{X_i - F_i}{X_i}\right]100$

5. Mean Percentage Error (MPE) = $\frac{\sum_{i=1}^{n} PE_i}{n} = \left[\sum_{i=1}^{n}\left\{\frac{X_i - F_i}{X_i}\right\}/n\right] 100$

6. Mean Absolute Percent Error (MAPE) is a measure of the average difference between the forecasted values and the actual values, expressed as a percentage. It is calculated by taking the absolute value of the difference between the forecasted value and the actual value and then dividing it by the actual value and multiplying it by 100. Mathematically, MAPE is expressed as:

   Mean Absolute Percentage Error (MAPE) = $\left[\frac{\sum_{i=1}^{n}\left|\frac{X_i - F_i}{X_i}\right|}{n}\right] 100$

7. If a forecast is being prepared for a rime horizon of one period, the most recent actual value would be used as the forecast for the next period. Under Naive Forecast 1 or NF1 the MAPE is expressed as

$$[MAPE]_{NF1} = \left[\frac{\sum_{i=2}^{n}\left|\frac{X_i - X_{i-1}}{X_i}\right|}{n-1}\right]100$$

8. If one considers the seasonality in the time series, then under Naïve Forecast 2 or NF, the MAPE is expressed as

$$[MAPE]_{NF2} = \left[\frac{\sum_{i=2}^{n}\left|\frac{X_i' - X_{i-1}'}{X_{i-1}'}\right|}{n-1}\right]100$$

where X_i' is the seasonality adjusted value of X_i.

5.14. MODEL: ARIMA MODEL FOR INDIAN CROP (RICE) PRICE FORECASTING

The Consideration of quantity-based time series of historical data from JAN 2003 to MAY 2022. i.e., 233 months data of Indian rice price which is in Indian rupee per metric ton as shown in figure 5.4. Historical data for the Study of QBTS (quantity-based time series) is taken from the Index Mundi website [144]. Process of the forecasting on time series historical data done by SPSS software.

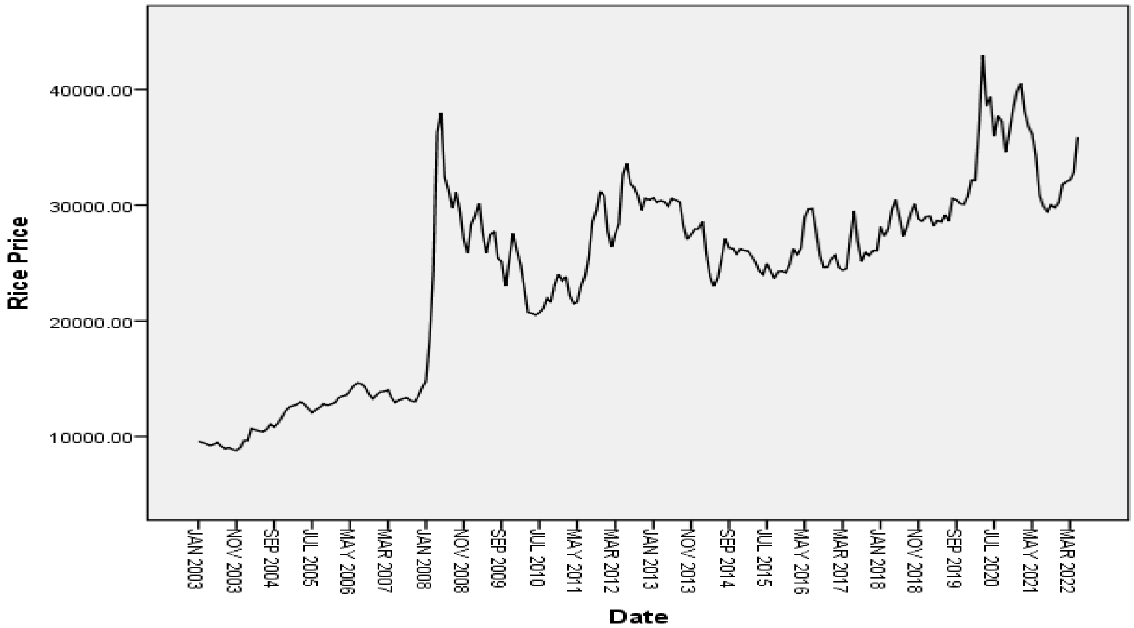


FIGURE. 5.4. Quantity-based time series data of the Indian rice price from JAN 2003 – MAY 2022.

For the application of the ARIMA (p, d, q) model first, check the stationary or non-stationary condition of historical data. Quantity-based time series of historical data of Indian rice price which is in Indian rupee per metric ton. From figure 5.4 it can be said that the series is not stationary, and figure 5.5 shows the 16 lags of ACF for the considered data series.

FIGURE. 5.5. ACF of quantity-based time series of Indian rice price

The next one is to make a series stationary with the order of integration (difference), calculate the mean order of d, and take the first-order difference (d=1), with the different series which is plotted in figure 5.6, where can see that the series is stationary at d=1.

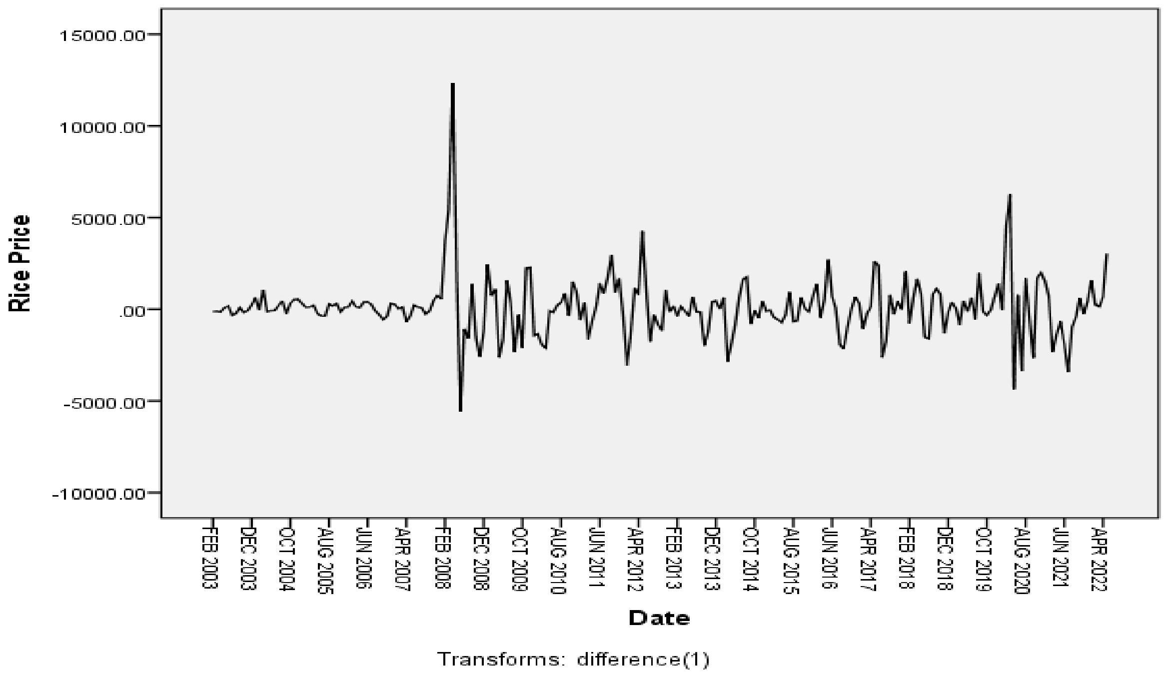
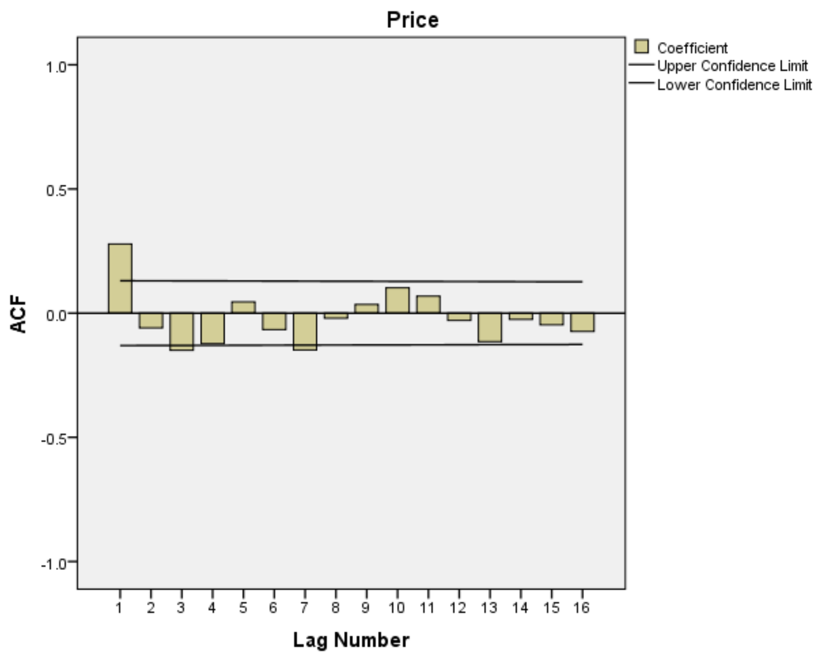


FIGURE 5.6. Quantity-based time series at the first difference (d=1)

An order of difference is 1 but for the ARIMA model, the required order of p & q can be observed by ACF and Partial ACF plot. Figure 5.7 shows the ACF & PACF at seasonal difference 1 with 16 lags.



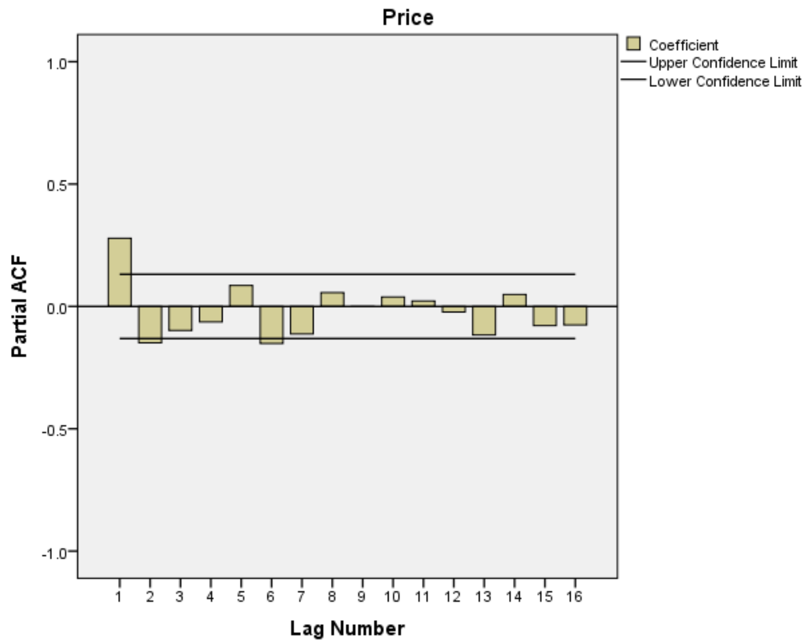


FIGURE 5.7. ACF and Partial ACF functions at the difference (d=1) with 16 lags.

The estimated order p and q from ACF and PACF for the ARIMA (p, d, q) model are shown. In this estimation, the order of p is 1 or 2, similarly, the order of q is 1 or 3. According to these, models of the ARIMA(1,1,1), ARIMA(1,1,3), & ARIMA(2,1,1), and ARIMA(2,1,3) are possible with respect to significance levels as ACF and Partial ACF have shown in figure 5.8, 5.9, 5.10, and 5.11. Evaluation of the possible models of the ARIMA for the forecasting of Indian price data performance is observed by MAPE. It is described as,

$$\text{Mean Absolute Percentage Error (MAPE)} = \left[\frac{\sum_{i=1}^{n}\left|\frac{X_i - F_i}{X_i}\right|}{n}\right] \times 100$$

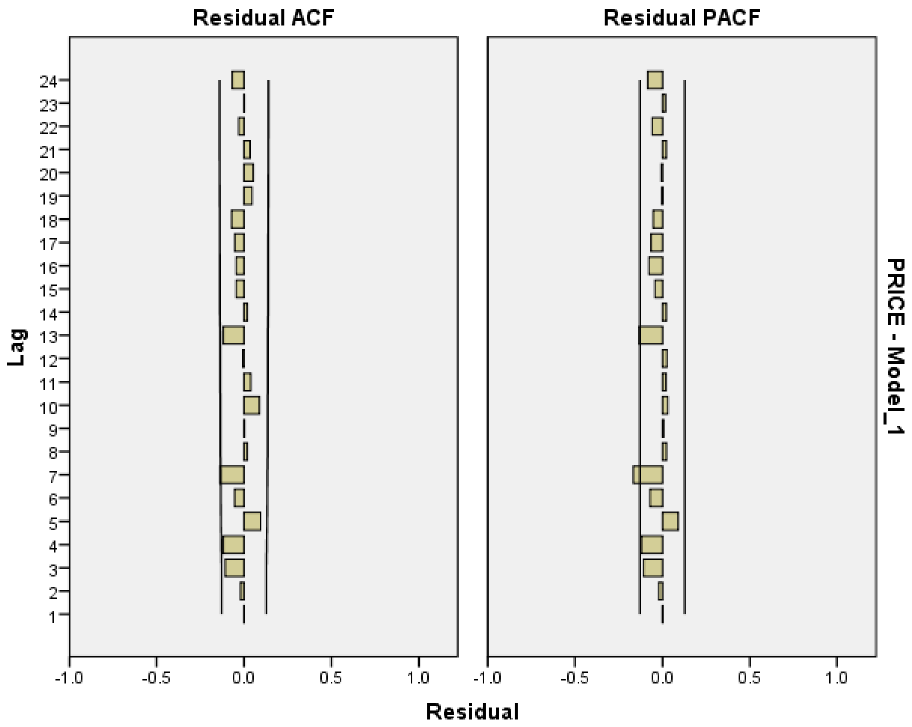


FIGURE. 5.8. ACF and PACF for model ARIMA (1,1,1)

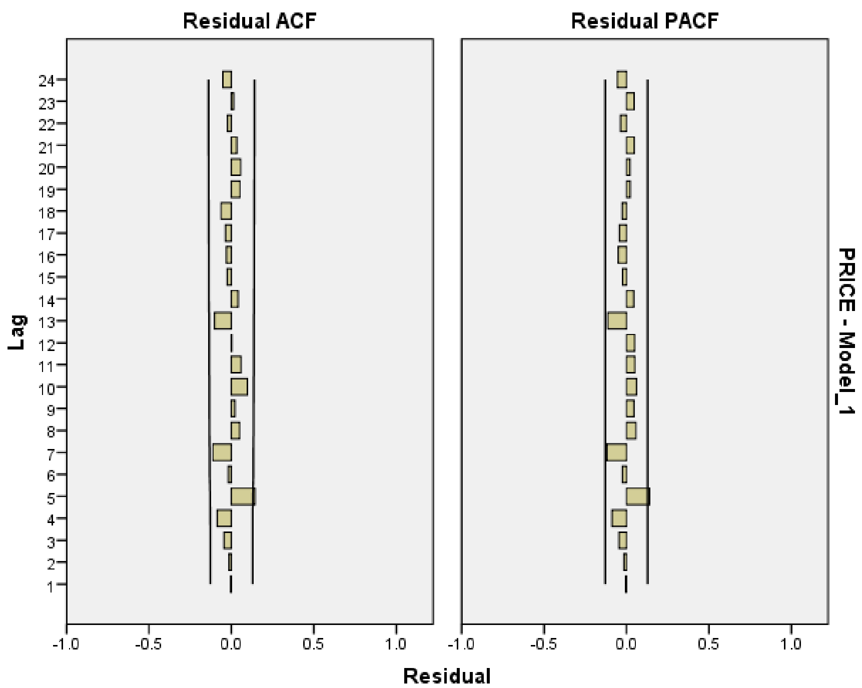


FIGURE 5.9. ACF and PACF for model ARIMA (1,1,3)

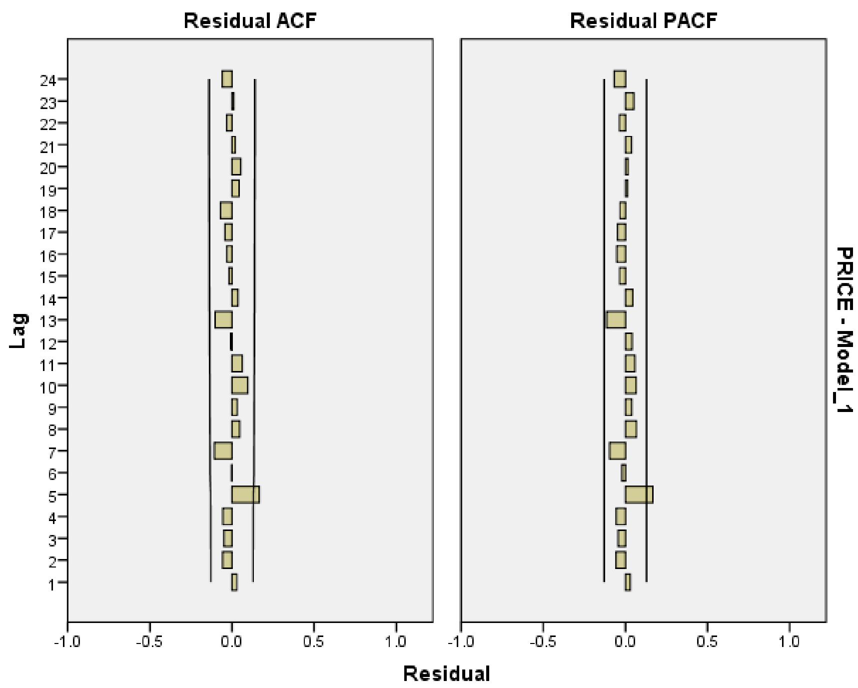


FIGURE 5.10. ACF and PACF for model ARIMA (2,1,1)

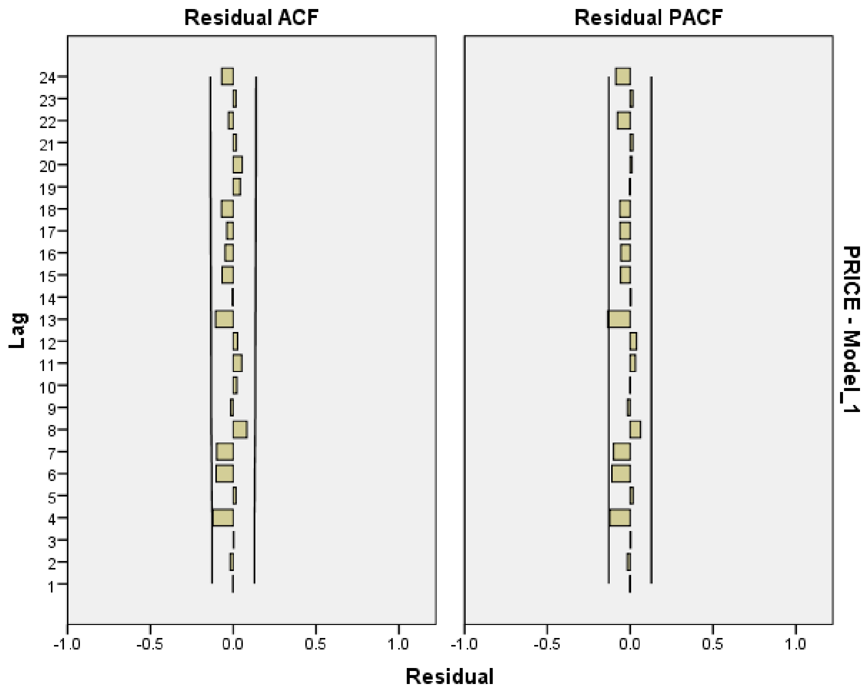


FIGURE 5.11. ACF and PACF for model ARIMA(2,1,3)

In table 5.2 to table 5.12, shows the forecasted rice price of India with respect to possible models of ARIMA with the lower and upper limits. And respective models of ARIMA (1,1,1), ARIMA (1,1,3) & ARIMA (2,1,1) & ARIMA (2,1,3) shows the forecasted graph of the future till Dec 2038 as shown in figure 5.12 to figure 5.15.

TABLE 5.2. Forecasted Price of Indian Rice form Jun 2022 to Oct 2023

Model	ARIMA(2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast
Jun 2022	33979.55	40088.42	37033.99	33874.10	39946.43	36910.27	33953.68	39995.73	36974.71	33953.68	39995.73	36974.71
Jul 2022	32284.32	42378.76	37331.54	32322.69	42074.99	37198.84	32264.16	42080.80	37172.48	32264.16	42080.80	37172.48
Aug 2022	30594.81	43553.98	37074.39	31056.50	43390.40	37223.45	31060.50	43339.12	37199.81	31060.50	43339.12	37199.81
Sep 2022	29431.42	44374.25	36902.83	30093.79	44247.48	37170.63	30249.66	44220.26	37234.96	30249.66	44220.26	37234.96
Oct 2022	28925.34	45468.13	37196.74	29378.45	44840.49	37109.47	29670.05	44884.21	37277.13	29670.05	44884.21	37277.13
Nov 2022	28526.91	46689.19	37608.05	28850.26	45276.61	37063.44	29245.43	45405.82	37325.62	29245.43	45405.82	37325.62
Dec 2022	27796.39	47610.79	37703.59	28460.83	45616.46	37038.64	28931.86	45827.73	37379.80	28931.86	45827.73	37379.80
Jan 2023	26939.81	48226.08	37582.94	28174.27	45895.52	37034.90	28701.17	46176.98	37439.07	28701.17	46176.98	37439.07
Feb 2023	26356.03	48889.05	37622.54	28533.99	46471.89	37502.94	28533.99	46471.89	37502.94	25944.81	49603.85	37774.33
Mar 2023	26042.98	49744.42	37893.70	28416.41	46725.46	37570.93	28416.41	46725.46	37570.93	25392.09	50389.68	37890.88
Apr 2023	25681.41	50582.02	38131.72	28338.10	46947.17	37642.63	28338.10	46947.17	37642.63	24873.43	51141.45	38007.44
May 2023	25132.75	51220.15	38176.45	28291.23	47144.11	37717.67	28291.23	47144.11	37717.67	24384.10	51863.89	38123.99
Jun 2023	24582.53	51758.59	38170.56	28269.74	47321.66	37795.70	28269.74	47321.66	37795.70	23920.38	52560.71	38240.54
Jul 2023	24207.35	52382.76	38295.05	28268.91	47483.95	37876.43	28268.91	47483.95	37876.43	23479.28	53234.92	38357.10
Aug 2023	23932.87	53088.11	38510.49	28284.96	47634.19	37959.57	28284.96	47634.19	37959.57	23058.34	53888.97	38473.65
Sep 2023	23588.59	53732.23	38660.41	28314.90	47774.89	38044.90	28314.90	47774.89	38044.90	22655.52	54524.90	38590.21
Oct 2023	27580.48	47569.45	37574.97	27580.48	47569.45	37574.97	28356.33	47908.03	38132.18	22269.10	55144.42	38706.76

TABLE 5.3. Forecasted Price of Indian Rice from Nov 2023 to Mar 2025

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**
Nov 2023	27606.55	47729.88	37668.22	27606.55	47729.88	37668.22	28407.29	48035.15	38221.22	21897.62	55749.01	38823.31
Dec 2023	27639.83	47889.21	37764.52	27639.83	47889.21	37764.52	28466.20	48157.49	38311.84	21539.82	56339.91	38939.87
Jan 2024	27679.01	48047.70	37863.36	27679.01	48047.70	37863.36	28531.75	48276.03	38403.89	21194.62	56918.22	39056.42
Feb 2024	27723.03	48205.55	37964.29	27723.03	48205.55	37964.29	28602.88	48391.54	38497.21	20861.05	57484.89	39172.97
Mar 2024	27771.04	48362.89	38066.96	27771.04	48362.89	38066.96	28678.72	48504.65	38591.68	20538.29	58040.76	39289.53
Apr 2024	27822.35	48519.80	38171.08	27822.35	48519.80	38171.08	28758.53	48615.85	38687.19	20225.59	58586.57	39406.08
May 2024	27876.41	48676.36	38276.38	27876.41	48676.36	38276.38	28841.71	48725.53	38783.62	19922.29	59122.98	39522.63
Jun 2024	20789.22	58547.73	39668.47	27932.74	48832.61	38382.68	28927.76	48834.01	38880.88	19627.79	59650.58	39639.19
Jul 2024	20514.17	59032.61	39773.39	27990.99	48988.60	38489.79	29016.26	48941.54	38978.90	19341.57	60169.91	39755.74
Aug 2024	20239.79	59498.08	39868.93	28050.84	49144.33	38597.59	29106.86	49048.32	39077.59	19063.14	60681.45	39872.29
Sep 2024	19997.41	59975.99	39986.70	28112.04	49299.85	38705.94	29199.26	49154.51	39176.89	18792.07	61185.62	39988.85
Oct 2024	19776.62	60464.46	40120.54	28174.38	49455.15	38814.77	29293.22	49260.24	39276.73	18527.97	61682.83	40105.40
Nov 2024	19548.43	60938.47	40243.45	28237.70	49610.25	38923.98	29388.51	49365.61	39377.06	18270.46	62173.45	40221.95
Dec 2024	19308.92	61390.13	40349.52	28301.85	49765.16	39033.51	29484.97	49470.70	39477.84	18019.22	62657.79	40338.51
Jan 2025	19078.64	61836.27	40457.45	28366.72	49919.88	39143.30	29582.43	49575.57	39579.00	17773.95	63136.17	40455.06
Feb 2025	18868.40	62290.02	40579.21	28432.21	50074.42	39253.32	29680.78	49680.28	39680.53	17534.35	63608.87	40571.61
Mar 2025	18666.22	62743.78	40705.00	28498.24	50228.79	39363.51	29779.89	49784.86	39782.37	17300.18	64076.15	40688.17

TABLE 5.4. Forecasted Price of Indian Rice from Apr 2025 to Oct 2026

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast
Apr 2025	18458.57	63184.65	40821.61	28564.74	50382.98	39473.86	29879.66	49889.35	39884.50	17071.20	64538.24	40804.72
May 2025	18249.12	63613.82	40931.47	28631.66	50537.00	39584.33	29980.03	49993.76	39986.89	16847.17	64995.38	40921.27
Jun 2025	18049.51	64042.07	41045.79	28698.96	50690.85	39694.91	30080.91	50098.12	40089.52	16627.89	65447.76	41037.83
Jul 2025	17861.29	64472.98	41167.13	28766.60	50844.54	39805.57	30182.24	50202.45	40192.35	16413.17	65895.59	41154.38
Aug 2025	17675.76	64899.80	41287.78	28834.54	50998.07	39916.30	30283.98	50306.76	40295.37	16202.83	66339.04	41270.93
Sep 2025	17488.05	65317.32	41402.69	28902.76	51151.43	40027.10	30386.07	50411.04	40398.56	15996.70	66778.28	41387.49
Oct 2025	17302.76	65728.97	41515.87	28971.23	51304.64	40137.94	30488.47	50515.32	40501.90	15794.61	67213.47	41504.04
Nov 2025	17125.31	66140.35	41632.83	29039.95	51457.69	40248.82	30591.15	50619.60	40605.37	15596.43	67644.76	41620.59
Dec 2025	16954.08	66551.19	41752.63	29108.89	51610.58	40359.73	30694.07	50723.87	40708.97	15402.01	68072.28	41737.15
Jan 2026	16784.03	66957.20	41870.61	29178.03	51763.32	40470.68	30797.22	50828.14	40812.68	15211.22	68496.18	41853.70
Feb 2026	16614.28	67357.06	41985.67	29247.38	51915.90	40581.64	30900.55	50932.42	40916.49	15023.93	68916.58	41970.25
Mar 2026	16448.20	67753.76	42100.98	29316.91	52068.34	40692.63	31004.06	51036.71	41020.39	14840.04	69333.58	42086.81
Apr 2026	16287.69	68149.65	42218.67	29386.63	52220.63	40803.63	31107.73	51141.00	41124.36	14659.41	69747.31	42203.36
May 2026	16130.74	68543.46	42337.10	29456.51	52372.76	40914.64	31211.53	51245.29	41228.41	14481.97	70157.86	42319.91
Jun 2026	15974.93	68933.02	42453.98	29526.57	52524.76	41025.66	31315.46	51349.59	41332.53	14307.59	70565.35	42436.47
Jul 2026	15820.75	69318.56	42569.65	29596.79	52676.61	41136.70	31419.50	51453.90	41436.70	14136.19	70969.85	42553.02
Aug 2026	15670.15	69702.00	42686.08	29667.17	52828.31	41247.74	31523.63	51558.21	41540.92	13967.68	71371.47	42669.58
Sep 2026	15523.41	70084.03	42803.72	29737.70	52979.87	41358.79	31627.86	51662.53	41645.19	13801.98	71770.28	42786.13
Oct 2026	15379.01	70463.53	42921.27	29808.38	53131.30	41469.84	31732.16	51766.86	41749.51	13638.99	72166.37	42902.68

TABLE 5.5. Forecasted Price of Indian Rice from Nov 2026 to Mar 2028

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**
Nov 2026	15236.04	70839.62	43037.83	29879.21	53282.58	41580.89	31836.53	51871.19	41853.86	13478.65	72559.82	43019.24
Dec 2026	15095.20	71212.95	43154.07	29950.18	53433.73	41691.95	31940.97	51975.52	41958.25	13320.89	72950.69	43135.79
Jan 2027	14957.40	71584.52	43270.96	30021.29	53584.74	41803.02	32045.47	52079.86	42062.66	13165.62	73339.06	43252.34
Feb 2027	14822.36	71954.34	43388.35	30092.55	53735.62	41914.08	32150.02	52184.20	42167.11	13012.79	73725.00	43368.90
Mar 2027	14689.18	72321.70	43505.44	30163.94	53886.36	42025.15	32254.61	52288.54	42271.58	12862.33	74108.57	43485.45
Apr 2027	14557.63	72686.40	43622.01	30235.47	54036.97	42136.22	32359.25	52392.88	42376.07	12714.17	74489.83	43602.00
May 2027	14428.25	73049.00	43738.62	30307.13	54187.45	42247.29	32463.93	52497.23	42480.58	12568.27	74868.84	43718.56
Jun 2027	14301.35	73409.94	43855.64	30378.92	54337.80	42358.36	32568.64	52601.58	42585.11	12424.56	75245.66	43835.11
Jul 2027	14176.56	73769.04	43972.80	30450.84	54488.02	42469.43	32673.38	52705.93	42689.65	12282.98	75620.34	43951.66
Aug 2027	14053.44	74125.95	44089.69	30522.89	54638.12	42580.51	32778.15	52810.28	42794.21	12143.50	75992.93	44068.22
Sep 2027	13932.01	74480.74	44206.37	30595.07	54788.09	42691.58	32882.94	52914.63	42898.79	12006.05	76363.49	44184.77
Oct 2027	13812.58	74833.77	44323.18	30667.37	54937.93	42802.65	32987.76	53018.98	43003.37	11870.59	76732.06	44301.32
Nov 2027	13695.18	75185.20	44440.19	30739.80	55087.65	42913.73	33092.60	53123.34	43107.97	11737.07	77098.68	44417.88
Dec 2027	13579.53	75534.86	44557.20	30812.35	55237.25	43024.80	33197.46	53227.69	43212.57	11605.46	77463.40	44534.43
Jan 2028	13465.44	75882.62	44674.03	30885.03	55386.73	43135.88	33302.33	53332.04	43317.19	11475.70	77826.27	44650.98
Feb 2028	13352.99	76228.62	44790.81	30957.82	55536.09	43246.95	33407.22	53436.39	43421.81	11347.75	78187.32	44767.54
Mar 2028	13242.33	76573.06	44907.69	31030.73	55685.33	43358.03	33512.13	53540.74	43526.43	11221.58	78546.60	44884.09

TABLE 5.6. Forecasted Price of Indian Rice from Apr 2028 to Aug 2029

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast
Apr 2028	13133.37	76915.96	45024.66	31103.76	55834.45	43469.11	33617.04	53645.09	43631.07	11097.15	78904.14	45000.64
May 2028	13025.94	77257.24	45141.59	31176.91	55983.46	43580.18	33721.97	53749.44	43735.71	10974.42	79259.98	45117.20
Jun 2028	12919.98	77596.87	45258.42	31250.17	56132.35	43691.26	33826.92	53853.79	43840.35	10853.35	79614.15	45233.75
Jul 2028	12815.55	77934.98	45375.26	31323.55	56281.12	43802.34	33931.87	53958.14	43945.00	10733.92	79966.69	45350.30
Aug 2028	12712.69	78271.65	45492.17	31397.04	56429.78	43913.41	34036.83	54062.49	44049.66	10616.08	80317.63	45466.86
Sep 2028	12611.32	78606.90	45609.11	31470.65	56578.33	44024.49	34141.80	54166.83	44154.31	10499.81	80667.01	45583.41
Oct 2028	12511.33	78940.66	45726.00	31544.36	56726.77	44135.57	34246.77	54271.18	44258.97	10385.08	81014.85	45699.96
Nov 2028	12412.71	79272.99	45842.85	31618.19	56875.09	44246.64	34351.76	54375.52	44363.64	10271.85	81361.19	45816.52
Dec 2028	12315.50	79603.95	45959.72	31692.13	57023.31	44357.72	34456.75	54479.86	44468.30	10160.09	81706.05	45933.07
Jan 2029	12219.67	79933.59	46076.63	31766.17	57171.42	44468.80	34561.74	54584.20	44572.97	10049.78	82049.47	46049.62
Feb 2029	12125.17	80261.90	46193.54	31840.32	57319.42	44579.87	34666.74	54688.54	44677.64	9940.90	82391.46	46166.18
Mar 2029	12031.95	80588.89	46310.42	31914.58	57467.31	44690.95	34771.75	54792.88	44782.32	9833.41	82732.06	46282.73
Apr 2029	11939.99	80914.58	46427.29	31988.95	57615.10	44802.03	34876.76	54897.22	44886.99	9727.28	83071.29	46399.28
May 2029	11849.32	81239.03	46544.17	32063.42	57762.79	44913.10	34981.78	55001.55	44991.67	9622.50	83409.17	46515.84
Jun 2029	11759.89	81562.26	46661.08	32137.99	57910.37	45024.18	35086.80	55105.89	45096.34	9519.04	83745.74	46632.39
Jul 2029	11671.66	81884.27	46777.97	32212.67	58057.84	45135.26	35191.82	55210.22	45201.02	9416.88	84081.01	46748.95
Aug 2029	11584.62	82205.08	46894.85	32287.45	58205.22	45246.33	35296.85	55314.55	45305.70	9315.99	84415.01	46865.50

TABLE 5.7. Forecasted Price of Indian Rice from Sep 2029 to Mar 2031

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**	**LCL**	**UCL**	**Forecast**
Sep 2029	11498.74	82524.72	47011.73	32362.33	58352.49	45357.41	35401.88	55418.88	45410.38	9216.36	84747.75	46982.05
Oct 2029	11414.03	82843.21	47128.62	32437.31	58499.66	45468.49	35506.92	55523.21	45515.06	9117.95	85079.26	47098.61
Nov 2029	11330.45	83160.58	47245.51	32512.39	58646.73	45579.56	35611.96	55627.54	45619.75	9020.75	85409.57	47215.16
Dec 2029	11247.98	83476.83	47362.40	32587.57	58793.71	45690.64	35717.00	55731.86	45724.43	8924.75	85738.68	47331.71
Jan 2030	11166.59	83791.98	47479.28	32662.85	58940.58	45801.72	35822.04	55836.19	45829.11	8829.91	86066.62	47448.27
Feb 2030	11086.29	84106.05	47596.17	32738.23	59087.36	45912.79	35927.08	55940.51	45933.80	8736.22	86393.42	47564.82
Mar 2030	11007.05	84419.07	47713.06	32813.70	59234.04	46023.87	36032.13	56044.83	46038.48	8643.67	86719.08	47681.37
Apr 2030	10928.85	84731.05	47829.95	32889.27	59380.63	46134.95	36137.18	56149.15	46143.17	8552.23	87043.62	47797.93
May 2030	10851.68	85042.00	47946.84	32964.93	59527.12	46246.03	36242.23	56253.47	46247.85	8461.89	87367.07	47914.48
Jun 2030	10775.51	85351.94	48063.72	33040.69	59673.52	46357.10	36347.29	56357.79	46352.54	8372.62	87689.44	48031.03
Jul 2030	10700.34	85660.88	48180.61	33116.54	59819.82	46468.18	36452.34	56462.11	46457.23	8284.42	88010.75	48147.59
Aug 2030	10626.15	85968.85	48297.50	33192.48	59966.03	46579.26	36557.40	56566.42	46561.91	8197.27	88331.01	48264.14
Sep 2030	10552.93	86275.85	48414.39	33268.52	60112.15	46690.33	36662.46	56670.74	46666.60	8111.14	88650.24	48380.69
Oct 2030	10480.65	86581.90	48531.28	33344.64	60258.18	46801.41	36767.52	56775.05	46771.29	8026.03	88968.46	48497.25
Nov 2030	10409.31	86887.02	48648.16	33420.86	60404.12	46912.49	36872.58	56879.37	46875.97	7941.92	89285.68	48613.80
Dec 2030	10338.89	87191.21	48765.05	33497.16	60549.96	47023.56	36977.65	56983.68	46980.66	7858.80	89601.91	48730.35
Jan 2031	10269.38	87494.50	48881.94	33573.56	60695.72	47134.64	37082.71	57087.99	47085.35	7776.65	89917.17	48846.91
Feb 2031	10200.77	87796.89	48998.83	33650.04	60841.39	47245.72	37187.78	57192.30	47190.04	7695.45	90231.47	48963.46
Mar 2031	10133.04	88098.39	49115.72	33726.62	60986.97	47356.79	37292.85	57296.61	47294.73	7615.19	90544.83	49080.01

TABLE 5.8. Price Forecast of Indian Rice from Apr 2031 to Oct 2032

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast
Apr 2031	10066.18	88399.03	49232.60	33803.28	61132.47	47467.87	37397.91	57400.92	47399.41	7535.87	90857.26	49196.57
May 2031	10000.17	88698.81	49349.49	33880.02	61277.88	47578.95	37502.98	57505.22	47504.10	7457.46	91168.78	49313.12
Jun 2031	9935.02	88997.74	49466.38	33956.85	61423.20	47690.03	37608.05	57609.53	47608.79	7379.96	91479.39	49429.67
Jul 2031	9870.69	89295.84	49583.27	34033.77	61568.44	47801.10	37713.13	57713.84	47713.48	7303.35	91789.11	49546.23
Aug 2031	9807.19	89593.11	49700.15	34110.77	61713.59	47912.18	37818.20	57818.14	47818.17	7227.62	92097.95	49662.78
Sep 2031	9744.51	89889.58	49817.04	34187.85	61858.66	48023.26	37923.27	57922.44	47922.86	7152.75	92405.92	49779.33
Oct 2031	9682.62	90185.24	49933.93	34265.02	62003.64	48134.33	38028.35	58026.75	48027.55	7078.74	92713.03	49895.89
Nov 2031	9621.52	90480.11	50050.82	34342.27	62148.55	48245.41	38133.42	58131.05	48132.24	7005.58	93019.31	50012.44
Dec 2031	9561.21	90774.20	50167.71	34419.61	62293.37	48356.49	38238.50	58235.35	48236.93	6933.24	93324.75	50128.99
Jan 2032	9501.66	91067.52	50284.59	34497.02	62438.11	48467.56	38343.58	58339.65	48341.61	6861.73	93629.36	50245.55
Feb 2032	9442.88	91360.09	50401.48	34574.52	62582.76	48578.64	38448.65	58443.95	48446.30	6791.04	93933.17	50362.10
Mar 2032	9384.84	91651.90	50518.37	34652.09	62727.34	48689.72	38553.73	58548.25	48550.99	6721.14	94236.17	50478.65
Apr 2032	9327.55	91942.97	50635.26	34729.75	62871.84	48800.79	38658.81	58652.55	48655.68	6652.03	94538.38	50595.21
May 2032	9270.98	92233.31	50752.14	34807.48	63016.26	48911.87	38763.89	58756.85	48760.37	6583.71	94839.81	50711.76
Jun 2032	9215.14	92522.92	50869.03	34885.30	63160.60	49022.95	38868.97	58861.15	48865.06	6516.16	95140.47	50828.32
Jul 2032	9160.01	92811.83	50985.92	34963.19	63304.86	49134.03	38974.05	58965.45	48969.75	6449.37	95440.37	50944.87
Aug 2032	9105.59	93100.03	51102.81	35041.16	63449.05	49245.10	39079.13	59069.75	49074.44	6383.33	95739.52	51061.42
Sep 2032	9051.86	93387.53	51219.70	35119.20	63593.15	49356.18	39184.21	59174.05	49179.13	6318.03	96037.92	51177.98
Oct 2032	8998.82	93674.34	51336.58	35197.33	63737.19	49467.26	39289.30	59278.34	49283.82	6253.47	96335.59	51294.53

TABLE 5.9. Forecasted Price of Indian Rice from Nov 2032 to May 2034

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast
Nov 2032	8946.46	93960.48	51453.47	35275.53	63881.14	49578.33	39394.38	59382.64	49388.51	6189.63	96632.53	51411.08
Dec 2032	8894.77	94245.94	51570.36	35353.80	64025.02	49689.41	39499.46	59486.93	49493.20	6126.51	96928.76	51527.64
Jan 2033	8843.75	94530.75	51687.25	35432.15	64168.83	49800.49	39604.54	59591.23	49597.89	6064.10	97224.28	51644.19
Feb 2033	8793.38	94814.89	51804.13	35510.57	64312.56	49911.56	39709.63	59695.53	49702.58	6002.39	97519.09	51760.74
Mar 2033	8743.66	95098.39	51921.02	35589.07	64456.21	50022.64	39814.71	59799.82	49807.27	5941.37	97813.22	51877.30
Apr 2033	8694.57	95381.25	52037.91	35667.64	64599.80	50133.72	39919.80	59904.11	49911.96	5881.04	98106.66	51993.85
May 2033	8646.12	95663.47	52154.80	35746.28	64743.31	50244.79	40024.88	60008.41	50016.65	5821.38	98399.43	52110.40
Jun 2033	8598.30	95945.07	52271.69	35825.00	64886.75	50355.87	40129.97	60112.70	50121.33	5762.39	98691.53	52226.96
Jul 2033	8551.10	96226.05	52388.57	35903.79	65030.11	50466.95	40235.05	60217.00	50226.02	5704.06	98982.96	52343.51
Aug 2033	8504.51	96506.42	52505.46	35982.64	65173.41	50578.03	40340.14	60321.29	50330.71	5646.38	99273.75	52460.06
Sep 2033	8458.52	96786.18	52622.35	36061.57	65316.63	50689.10	40445.22	60425.58	50435.40	5589.35	99563.89	52576.62
Oct 2033	8413.13	97065.34	52739.24	36140.57	65459.78	50800.18	40550.31	60529.88	50540.09	5532.95	99853.39	52693.17
Nov 2033	8368.33	97343.92	52856.13	36219.65	65602.87	50911.26	40655.39	60634.17	50644.78	5477.19	100142.25	52809.72
Dec 2033	8324.12	97621.90	52973.01	36298.79	65745.88	51022.33	40760.48	60738.46	50749.47	5422.06	100430.50	52926.28
Jan 2034	8280.49	97899.31	53089.90	36377.99	65888.83	51133.41	40865.57	60842.76	50854.16	5367.54	100718.12	53042.83
Feb 2034	8237.43	98176.15	53206.79	36457.27	66031.70	51244.49	40970.65	60947.05	50958.85	5313.64	101005.13	53159.38
Mar 2034	8194.94	98452.41	53323.68	36536.62	66174.51	51355.56	41075.74	61051.34	51063.54	5260.34	101291.54	53275.94
Apr 2034	8153.01	98728.12	53440.56	36616.03	66317.25	51466.64	41180.83	61155.63	51168.23	5207.63	101577.35	53392.49
May 2034	8111.63	99003.27	53557.45	36695.51	66459.92	51577.72	41285.91	61259.93	51272.92	5155.53	101862.56	53509.04

TABLE 5.10. Forecasted Price of Indian Rice from Jun 2034 to Dec 2035

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast
Jun 2034	8070.80	99277.88	53674.34	36775.06	66602.53	51688.79	41391.00	61364.22	51377.61	5104.00	102147.19	53625.60
Jul 2034	8030.52	99551.93	53791.23	36854.68	66745.07	51799.87	41496.09	61468.51	51482.30	5053.06	102431.24	53742.15
Aug 2034	7990.78	99825.46	53908.12	36934.36	66887.54	51910.95	41601.17	61572.80	51586.99	5002.69	102714.72	53858.70
Sep 2034	7951.56	100098.44	54025.00	37014.10	67029.95	52022.03	41706.26	61677.10	51691.68	4952.89	102997.62	53975.26
Oct 2034	7912.88	100370.90	54141.89	37093.92	67172.29	52133.10	41811.35	61781.39	51796.37	4903.65	103279.97	54091.81
Nov 2034	7874.71	100642.84	54258.78	37173.79	67314.57	52244.18	41916.43	61885.68	51901.06	4854.97	103561.76	54208.36
Dec 2034	7837.07	100914.27	54375.67	37253.73	67456.78	52355.26	42021.52	61989.97	52005.75	4806.84	103842.99	54324.92
Jan 2035	7799.93	101185.18	54492.55	37333.74	67598.93	52466.33	42126.61	62094.27	52110.44	4759.26	104123.68	54441.47
Feb 2035	7763.30	101455.58	54609.44	37413.81	67741.01	52577.41	42231.69	62198.56	52215.13	4712.22	104403.83	54558.03
Mar 2035	7727.17	101725.49	54726.33	37493.94	67883.03	52688.49	42336.78	62302.85	52319.82	4665.71	104683.45	54674.58
Apr 2035	7691.54	101994.90	54843.22	37574.13	68024.99	52799.56	42441.87	62407.15	52424.51	4619.73	104962.53	54791.13
May 2035	7656.40	102263.81	54960.11	37654.39	68166.89	52910.64	42546.95	62511.44	52529.20	4574.28	105241.09	54907.69
Jun 2035	7621.75	102532.24	55076.99	37734.71	68308.72	53021.72	42652.04	62615.73	52633.89	4529.35	105519.13	55024.24
Jul 2035	7587.58	102800.19	55193.88	37815.09	68450.50	53132.79	42757.12	62720.03	52738.57	4484.93	105796.66	55140.79
Aug 2035	7553.88	103067.65	55310.77	37895.54	68592.21	53243.87	42862.21	62824.32	52843.26	4441.02	106073.67	55257.35
Sep 2035	7520.67	103334.65	55427.66	37976.04	68733.86	53354.95	42967.30	62928.61	52947.95	4397.62	106350.18	55373.90
Oct 2035	7487.92	103601.17	55544.54	38056.60	68875.45	53466.03	43072.38	63032.91	53052.64	4354.71	106626.19	55490.45
Nov 2035	7455.63	103867.24	55661.43	38137.23	69016.98	53577.10	43177.47	63137.20	53157.33	4312.31	106901.70	55607.01
Dec 2035	7423.80	104132.84	55778.32	38217.91	69158.44	53688.18	43282.55	63241.49	53262.02	4270.39	107176.73	55723.56

TABLE 5.11. Forecasted Price of Indian Rice from Jan 2036 to Jul 2037

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast
Jan 2036	7392.43	104397.98	55895.21	38298.66	69299.85	53799.26	43387.64	63345.79	53366.71	4228.96	107451.26	55840.11
Feb 2036	7361.52	104662.68	56012.10	38379.46	69441.20	53910.33	43492.72	63450.08	53471.40	4188.01	107725.32	55956.67
Mar 2036	7331.05	104926.92	56128.98	38460.32	69582.50	54021.41	43597.80	63554.38	53576.09	4147.54	107998.90	56073.22
Apr 2036	7301.02	105190.72	56245.87	38541.25	69723.73	54132.49	43702.89	63658.67	53680.78	4107.55	108272.00	56189.77
May 2036	7271.43	105454.09	56362.76	38622.22	69864.90	54243.56	43807.97	63762.97	53785.47	4068.02	108544.63	56306.33
Jun 2036	7242.28	105717.01	56479.65	38703.26	70006.02	54354.64	43913.06	63867.26	53890.16	4028.96	108816.80	56422.88
Jul 2036	7213.56	105979.51	56596.53	38784.36	70147.08	54465.72	44018.14	63971.56	53994.85	3990.36	109088.51	56539.43
Aug 2036	7185.27	106241.58	56713.42	38865.51	70288.08	54576.79	44123.22	64075.86	54099.54	3952.21	109359.76	56655.99
Sep 2036	7157.40	106503.22	56830.31	38946.72	70429.03	54687.87	44228.31	64180.15	54204.23	3914.52	109630.56	56772.54
Oct 2036	7129.96	106764.44	56947.20	39027.98	70569.91	54798.95	44333.39	64284.45	54308.92	3877.28	109900.91	56889.09
Nov 2036	7102.93	107025.25	57064.09	39109.30	70710.75	54910.03	44438.47	64388.75	54413.61	3840.48	110170.81	57005.65
Dec 2036	7076.31	107285.64	57180.97	39190.68	70851.52	55021.10	44543.55	64493.04	54518.30	3804.12	110440.28	57122.20
Jan 2037	7050.11	107545.62	57297.86	39272.11	70992.24	55132.18	44648.63	64597.34	54622.99	3768.20	110709.31	57238.75
Feb 2037	7024.31	107805.19	57414.75	39353.60	71132.91	55243.26	44753.72	64701.64	54727.68	3732.72	110977.90	57355.31
Mar 2037	6998.91	108064.37	57531.64	39435.15	71273.52	55354.33	44858.80	64805.94	54832.37	3697.66	111246.06	57471.86
Apr 2037	6973.91	108323.14	57648.53	39516.74	71414.08	55465.41	44963.88	64910.24	54937.06	3663.03	111513.80	57588.41
May 2037	6949.31	108581.52	57765.41	39598.40	71554.58	55576.49	45068.96	65014.54	55041.75	3628.82	111781.11	57704.97
Jun 2037	6925.10	108839.50	57882.30	39680.10	71695.02	55687.56	45174.04	65118.84	55146.44	3595.03	112048.01	57821.52
Jul 2037	6901.28	109097.09	57999.19	39761.86	71835.42	55798.64	45279.12	65223.13	55251.13	3561.66	112314.49	57938.07

TABLE 5.12. Forecasted Price of Indian Rice from Aug 2037 to Dec 2038

Model	ARIMA (2,1,3)			ARIMA (2,1,1)			ARIMA (1,1,3)			ARIMA (1,1,1)		
	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast	LCL	UCL	Forecast
Aug 2037	6877.85	109354.30	58116.08	39843.68	71975.76	55909.72	45384.19	65327.43	55355.81	3528.70	112580.56	58054.63
Sep 2037	6854.80	109611.13	58232.96	39925.54	72116.04	56020.79	45489.27	65431.74	55460.50	3496.15	112846.21	58171.18
Oct 2037	6832.13	109867.57	58349.85	40007.46	72256.28	56131.87	45594.35	65536.04	55565.19	3464.00	113111.47	58287.73
Nov 2037	6809.84	110123.64	58466.74	40089.44	72396.46	56242.95	45699.43	65640.34	55669.88	3432.26	113376.32	58404.29
Dec 2037	6787.92	110379.33	58583.63	40171.46	72536.59	56354.03	45804.51	65744.64	55774.57	3400.91	113640.77	58520.84
Jan 2038	6766.37	110634.66	58700.52	40253.54	72676.67	56465.10	45909.58	65848.94	55879.26	3369.97	113904.82	58637.40
Feb 2038	6745.20	110889.61	58817.40	40335.67	72816.69	56576.18	46014.66	65953.24	55983.95	3339.41	114168.49	58753.95
Mar 2038	6724.38	111144.20	58934.29	40417.85	72956.66	56687.26	46119.74	66057.55	56088.64	3309.24	114431.76	58870.50
Apr 2038	6703.93	111398.43	59051.18	40500.08	73096.59	56798.33	46224.81	66161.85	56193.33	3279.46	114694.65	58987.06
May 2038	6683.84	111652.30	59168.07	40582.36	73236.46	56909.41	46329.89	66266.15	56298.02	3250.06	114957.15	59103.61
Jun 2038	6664.10	111905.81	59284.95	40664.70	73376.28	57020.49	46434.96	66370.46	56402.71	3221.05	115219.28	59220.16
Jul 2038	6644.72	112158.96	59401.84	40747.08	73516.05	57131.56	46540.04	66474.76	56507.40	3192.41	115481.02	59336.72
Aug 2038	6625.69	112411.77	59518.73	40829.51	73655.77	57242.64	46645.11	66579.07	56612.09	3164.14	115742.39	59453.27
Sep 2038	6607.01	112664.23	59635.62	40912.00	73795.44	57353.72	46750.19	66683.37	56716.78	3136.25	116003.39	59569.82
Oct 2038	6588.67	112916.34	59752.51	40994.53	73935.06	57464.79	46855.26	66787.68	56821.47	3108.73	116264.03	59686.38
Nov 2038	6570.68	113168.11	59869.39	41077.11	74074.63	57575.87	46960.33	66891.99	56926.16	3081.57	116524.29	59802.93
Dec 2038	6553.02	113419.54	59986.28	41159.74	74214.15	57686.95	47065.40	66996.29	57030.85	3054.77	116784.19	59919.48

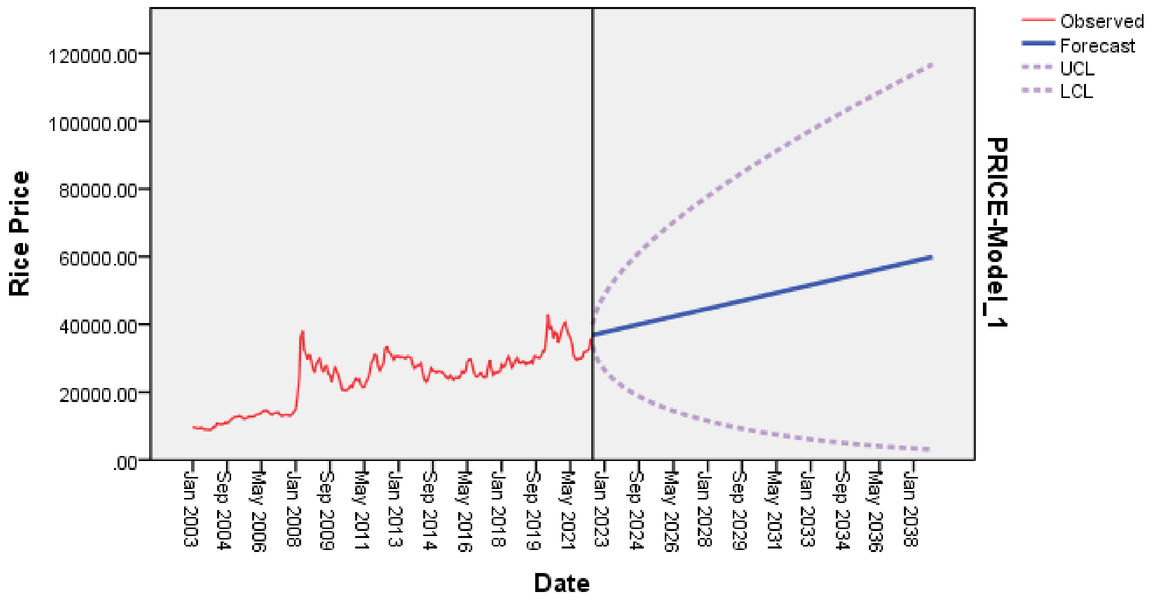


FIGURE 5.12. ARIMA(1, 1, 1) model along with the data set for forecasting from Jun 2022 to Dec 2038

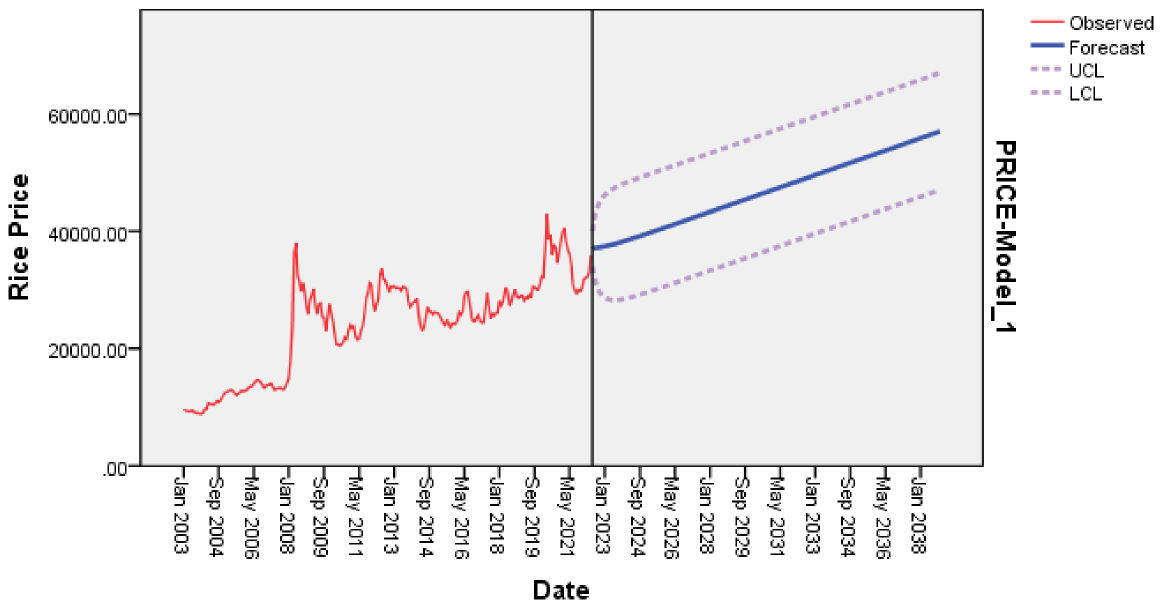


FIGURE 5.13. ARIMA(1, 1, 3) model along with the data set for forecasting from Jun 2022 to Dec 2038

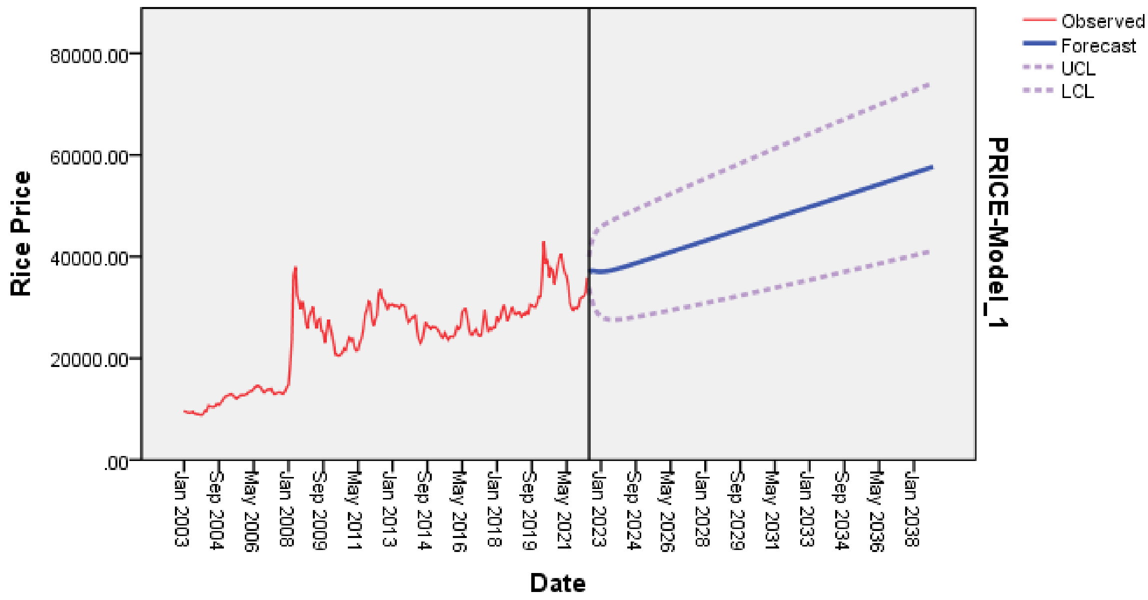


FIGURE 5.14. ARIMA (2, 1,1) model along with the data set for forecasting from Jun 2022 to Dec 2038

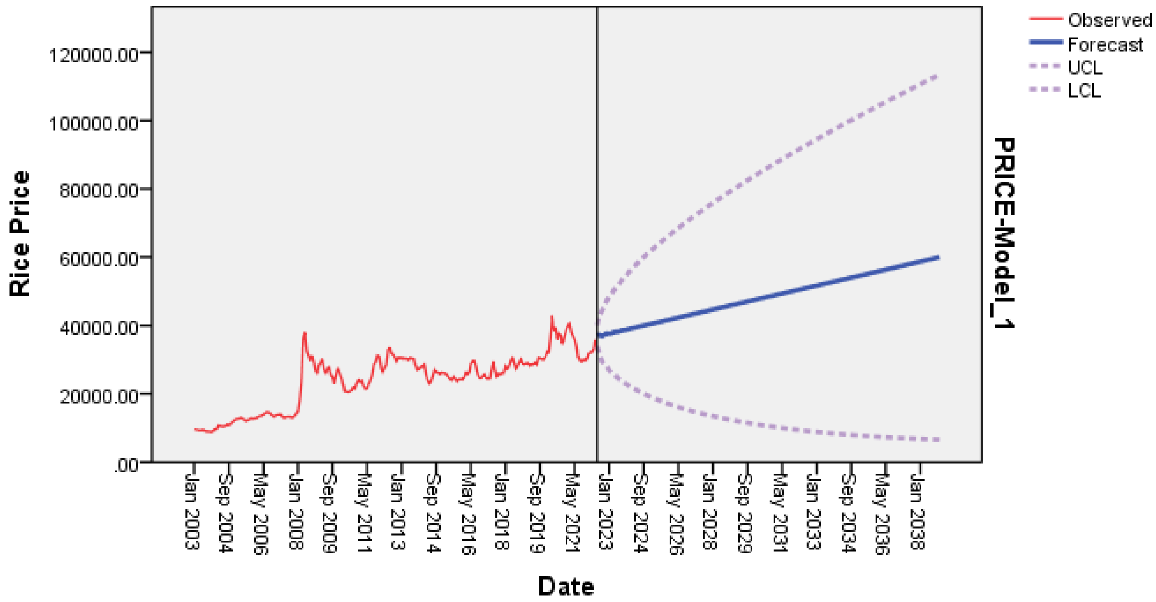


FIGURE 5.15. ARIMA (2, 1, 3) model along with the data set forecasting from Jun 2022 to Dec 2038

TABLE 5.13. Forecasted evaluation of the ARIMA models

Models	Model fit Statistics			Ljung-- Box Q(18)		
	MAPE	R_squared	Normalized BIC	Statistics	DF	Sig.
ARIMA(1,1,1)	3.630	0.964	14.778	22.758	16	0.120
ARIMA(1,1,3)	**3.530**	0.965	14.792	18.632	14	0.180
ARIMA(2,1,1)	3.552	0.965	14.774	20.681	15	0.147
ARIMA(2,1,3)	3.591	0.965	14.833	17.507	13	0.177

5.14.1. Result and Discussion

used the model of ARIMA to forecasting of the rice price data/historical data. Table 5.13, an evaluation of the ARIMA model in different order i.e., ARIMA(1,1,1),ARIMA(1,1,3), ARIMA(2,1,1), & ARIMA(2,1,3) shows MAPE value 3.630, 3.530, 3.552, 3.591 respectively. MAPE value is one of the crucial factors for the accuracy of model. MAPE value is inversely proportional to the accuracy. Lower value Higher will be the accuracy. In this, the MAPE value for ARIMA(1, 1, 3) model is **3.530**, which is a good level of forecasting. Considered time series data set of rice price reached 36974.71 in Jun 2022 after the applied model of forecasting ARIMA (1,1,3) and forecasting of long-term for Dec 2038 is 57030.85. Similarly, ARIMA (1,1,1) shows the forecasted value of price rice for Jun 2020 is 36864.59 and reached to 59919.48 in Dec 2038. ARIMA (2,1,1) and ARIMA (2,1,3) show the forecasted rice price values reached 57686.95 and 59986.28, respectively in the year 2038. Estimated rice price values by the forecasted models of ARIMA are graphically presented in figure 5.12-5.15 and clearly, it can be observed that the trends of the rice price are in increasing order in the future. For forecasting of rice price analysis, ARIMA(1,1,3) was found as the best-fitted possible model.

5.15. MODEL: FORECASTING OF RICE CROP PRICE

This model is based on the big data analysis concept of price time series data. Where used a machine learning technique for time series data analysis. For the time series data analysis purpose use a time series data of rice crop, price data (in USD) is taken for the period Feb 1992 to Jan 2022 from Index Mundi website [145]. In the dataset have recorded 360 months values of rice price. In the machine learning technique used python 3 (Jupyter notebook 6.4.8). Figure 5.16 shows the time series of the rice price data in USD. In this series, it makes more sense to assume in the sum of three terms, i.e., trends, seasonality, residual. Figure 5.17 show the decomposition of the rice price time series into three component such as trends, seasonality, residuals.

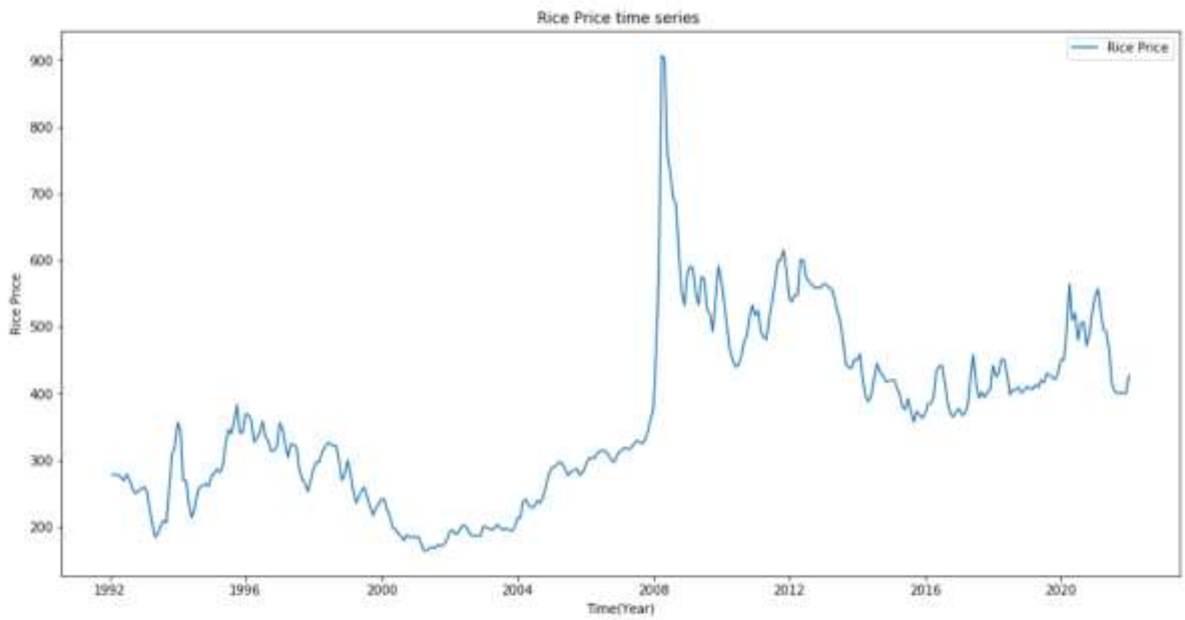


FIGURE 5.16. Rice Price Data from Feb 1992 to Jan 2022

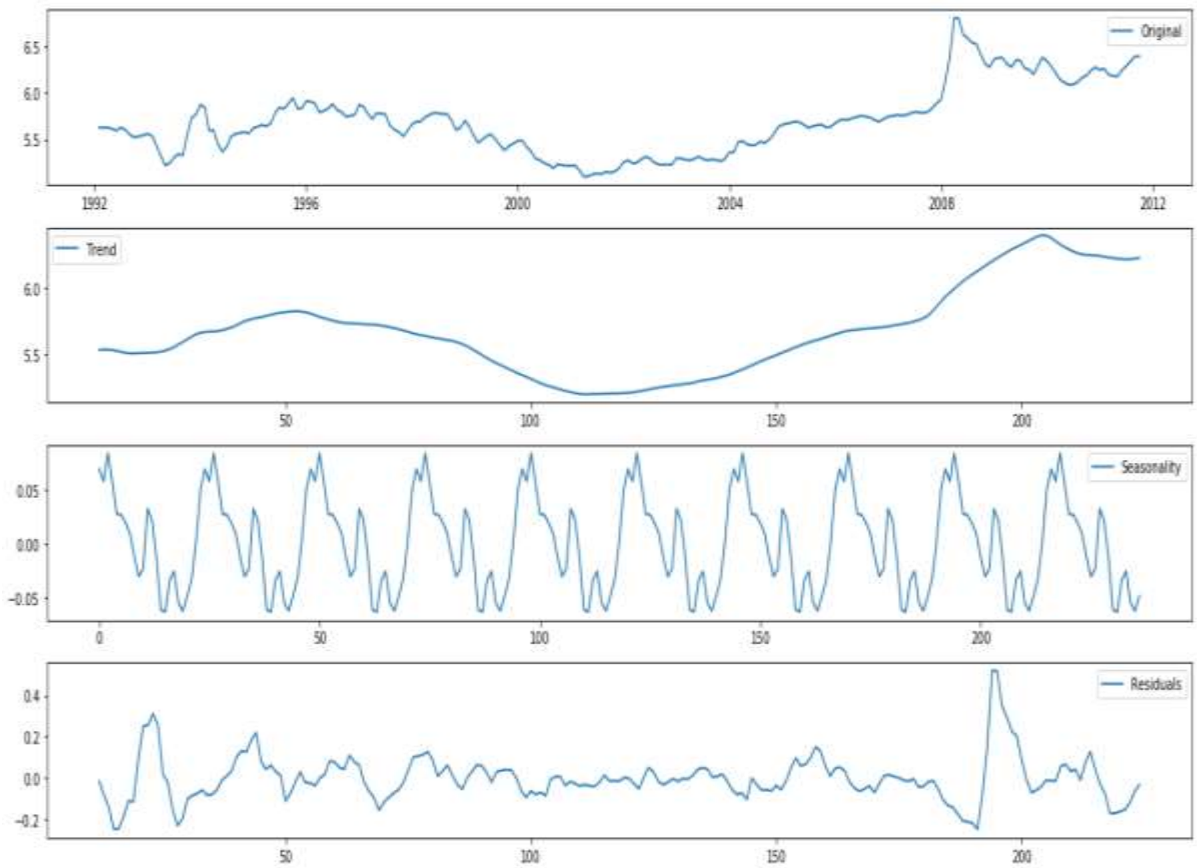


FIGURE 5.17. Decomposition of rice price time series into three components, i.e., Trend, Seasonality, Residuals

This big data time series analysis can the ARIMA (p,q,d) model with the different order values of p, q, & d. For the analysis used ARIMA model needed values of p and q, calculation of these values required a plotting of ACF and PACF functions, which helps to get the order values of AR and MA. Figure 5.18. Shows ACF and PACF of time series at 40 lags.

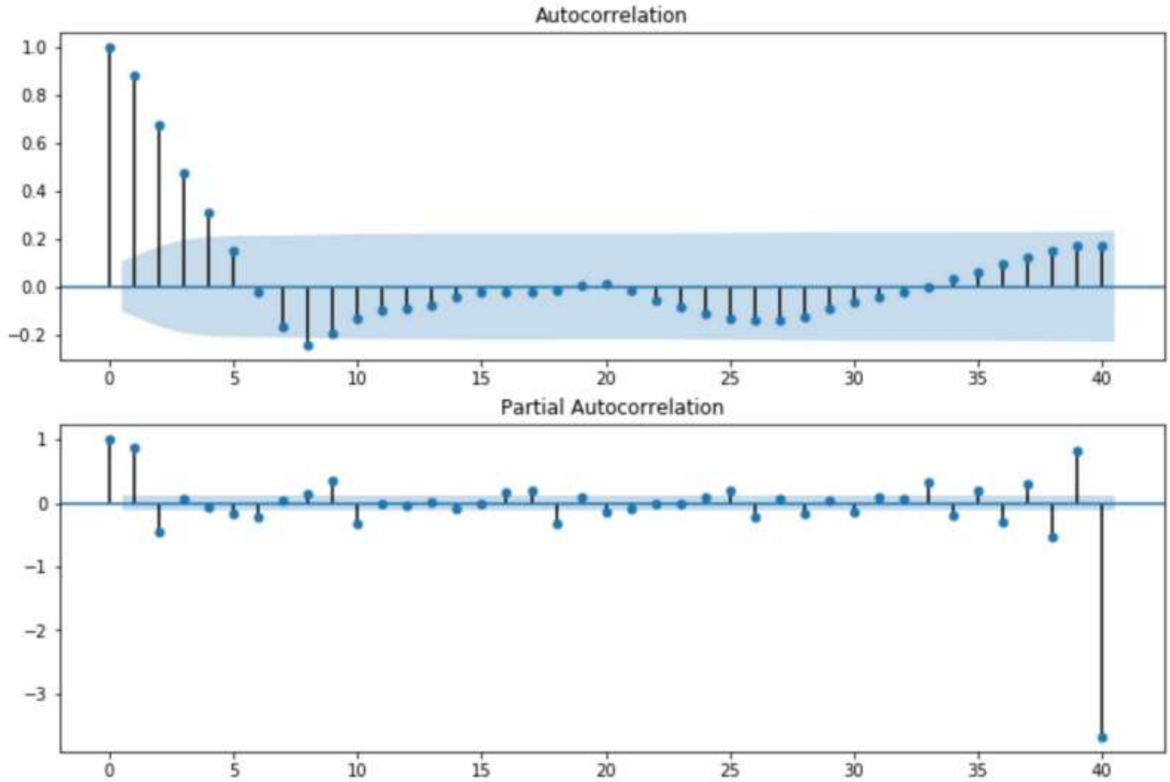


FIGURE 5.18. ACF and Partial ACF function of rice price time series at 40 lags

5.15.1. Result and Discussion

In the price time series big data analysis of rice crop using models;

1. ARIMA(1,1,1)
2. ARIMA(1,1,0)
3. Naïve Forecast

ARIMA model applied for the analysis for the order of (1,1,1) and (1,1,0) in python, are shown in figure 5.19 & figure 5.20, Respectively.

ARIMA Model Results

Dep. Variable:	D.Price	**No. Observations:**	359
Model:	ARIMA(1, 1, 1)	**Log Likelihood**	-1672.190
Method:	css-mle	**S.D. of innovations**	25.501
Date:	Thu, 17 Mar 2022	**AIC**	3352.380
Time:	00:37:16	**BIC**	3367.913
Sample:	03-01-1992	**HQIC**	3358.557
	- 01-01-2022		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.4410	1.869	0.236	0.814	-3.223	4.105
ar.L1.D.Price	-0.0260	0.137	-0.190	0.850	-0.294	0.242
ma.L1.D.Price	0.4260	0.125	3.407	0.001	0.181	0.671

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-38.5095	+0.0000j	38.5095	0.5000
MA.1	-2.3473	+0.0000j	2.3473	0.5000

FIGURE 5.19. ARIMA (1,1,1) model analysis of rice price time series data

ARIMA Model Results

Dep. Variable:	D.Price	**No. Observations:**	359
Model:	ARIMA(1, 1, 0)	**Log Likelihood**	-1677.277
Method:	css-mle	**S.D. of innovations**	25.867
Date:	Thu, 17 Mar 2022	**AIC**	3360.553
Time:	00:37:22	**BIC**	3372.203
Sample:	03-01-1992	**HQIC**	3365.186
	- 01-01-2022		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.4497	2.054	0.219	0.827	-3.576	4.476
ar.L1.D.Price	0.3363	0.050	6.771	0.000	0.239	0.434

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	2.9734	+0.0000j	2.9734	0.0000

FIGURE 5.20. ARIMA (1,1,0) model analysis of rice price time series data

Analysis of time series data by the model ARIMA(1,1,1,), ARIMA(1,1,0) of taken data values from Fab 1992 to Jan 2022, and for future calculated the forecast for the time series data, which is see in the figure 5.21 and figure 5.22 for ARIMA(1,1,1) & ARIMA(1,1,0), respectively.

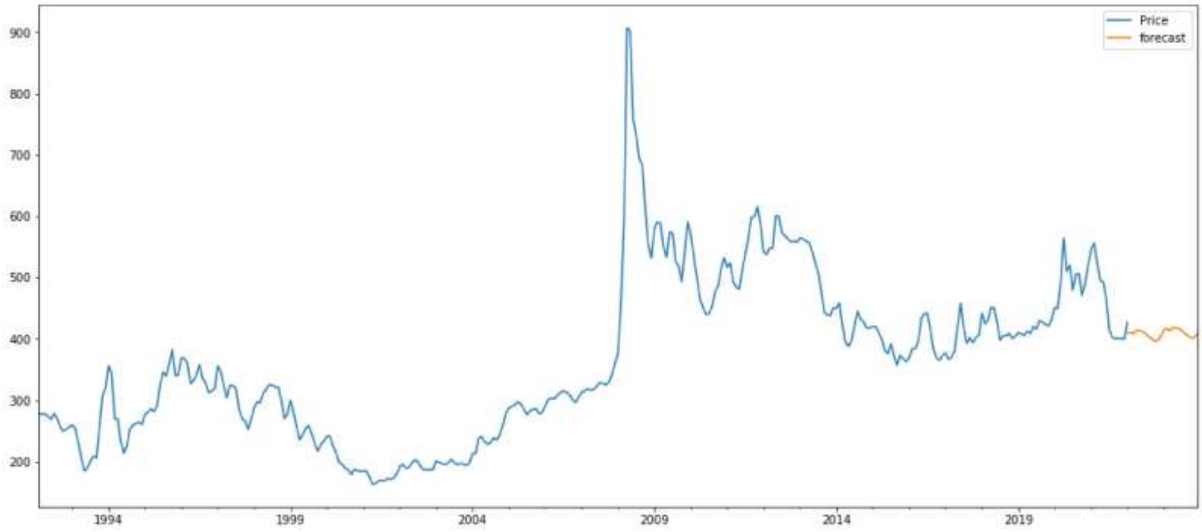


FIGURE 5.21. Rice price time series forecasted result by model ARIMA(1,1,1)

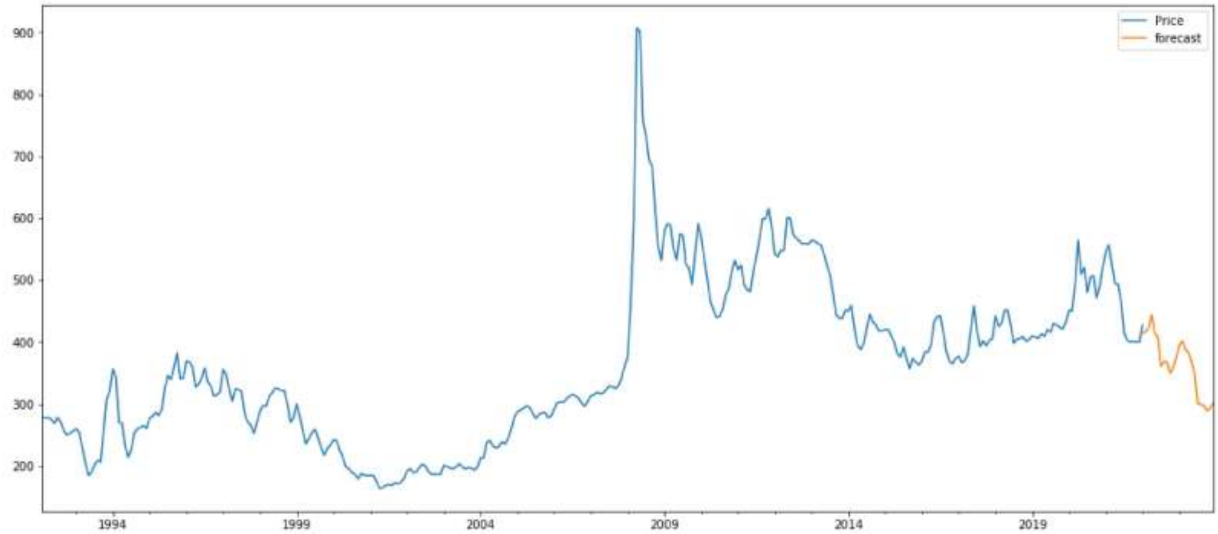


FIGURE 5.22. Rice price forecasted result by model ARIMA(1,1,0)

In naïve forecast analysis of rice price time series data used programming processes in the jupyter notebook are shown in the figure 5.23. Analysis by the Naive forecast model of price time series data is shown in figure 5.24.

```python
# predictions using naive approach for the validation set.
dd= np.asarray(train['Price'])
y_hat = valid.copy()
y_hat['Naive'] = dd[len(dd)-1]
plt.figure(figsize=(12,8))
plt.plot(train.index, train['Price'], label='Train')
plt.plot(valid.index,valid['Price'], label='Valid')
plt.plot(y_hat.index,y_hat['Naive'], label='Naive Forecast')
plt.legend(loc='best')
plt.title("Naive Forecast")
plt.show()
```

FIGURE 5.23. Programming processes of Naïve forecast model in Jupyter notebook

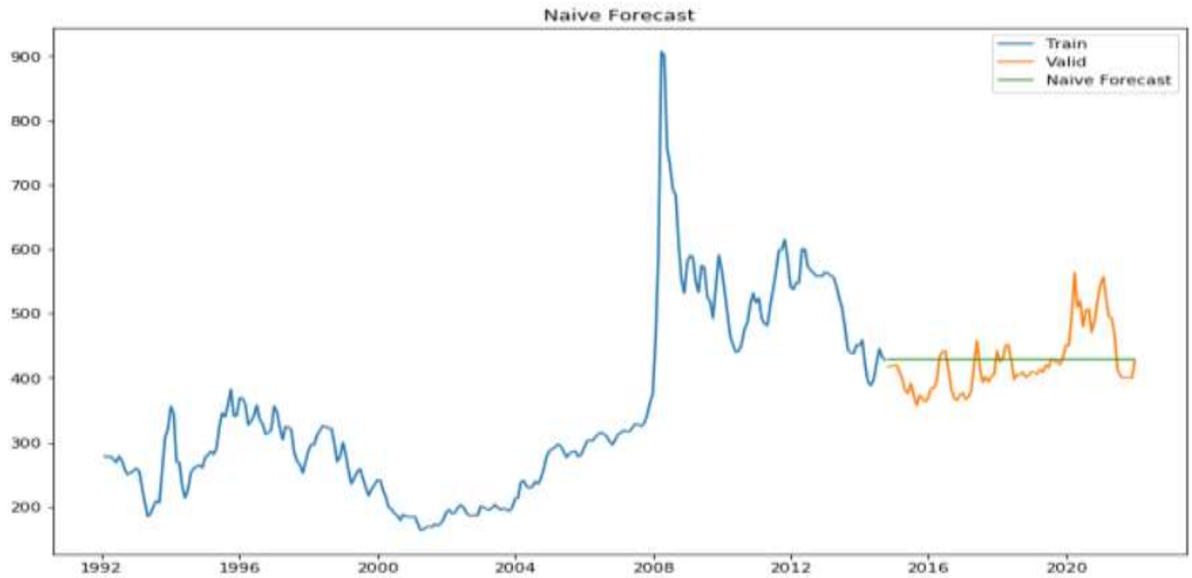


FIGURE 5.24. Analysis by the Naive forecast model of price data

This model is based on forecasting analysis of rice price time series data and for that, used models are ARIMA(1,1,1), ARIMA(1,1,0), and Naïve forecast. Which helps to others to understand, how can apply models on time series data. For the clear own concepts or knowledge, and ideas then do practice themself.

5.16. MODEL: FORECASTING OF NON-STATIONARY TIME SERIES

Analysis of the non-stationary time series data of the World stock index. The data contained in dataset of NASDAQ Composite, TSEC weighted, Shenzhen, SSE Composite world stock index [146,147,148,149]. Approaches of the forecasting analysis used the auto-regression integrated moving average model, which is a combination or integration of the auto-regression and moving average, at a several orders of p, d, and q. Representation of the ARIMA (p,d,q) as,

$$X_t = \theta_0 + \varphi_0 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_u X_{t-u} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_v e_{t-v}$$

Considered time series data is non-stationary or stationary, this is first purity to make a stationary to considered series, otherwise cannot apply ARIMA model, then can apply ARMA model for the analysis if time series data is stationary. Steps for forecasting analysis are;

Step 1. Check series is non-stationary or stationary.

Step 2. If series is non-stationary, then find out the series' first difference.

$$\Delta X_t = X_t - X_{t+1}$$

Step 3. Apply step 1 on the series' first difference. If stationary then follow step 4.

Step 4. Find out the order of AR and MA from ACF and PACF graphs.

Step 5. Make possible order model of ARIMA.

Step 6. Apply model of ARIMA on considered time series data for forecast.

For stationary and non-stationary confirmation from the p-value if p-value is less than 0.05 then series is stationary or series is non-stationary, if p-value is greater than 0.05, this processes for the p-value. If not checking p-value then drawn series of data help to identify for this.

Root mean square error defined by,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(forecasted\ value_i - actual\ value_i)^2}{n}}$$

Where, n denote the total number of data points. Minimum value of RMSE describe the accuracy of model.

Mean absolute percentage error (MAPE), used for accuracy as a percentage, defined by;

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{forecasted\ value_i - actual\ value_i}{actual\ value_i}\right|}{n} \times 100$$

Mean absolute error/ deviation (MAE or MAD) defined as;

$$MAE = \frac{\sum_{i=1}^{n}|forecasted\ value_i - actual\ value_i|}{n}$$

Total number of data points denoted by n.

Evaluate the accuracy of forecasting by R-square, which is defined as,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(forecasted\ value_i - actual\ value_i)^2}{\sum_{i=1}^{n}\left(\overline{actual\ value}_i - actual\ value_i\right)^2}$$

Bayesian information criterion (BIC) defined as;

$$BIC = n\log(MSE) + k\log n$$

Sample size represented by n, total number of observations by k, MSE is mean square error.

5.16.1. Computational Forecasting

Considered time series data of stock index as NASDAQ Composite, TSEC weighted, Shenzhen, SSE Composite history shown in the figure 5.25.

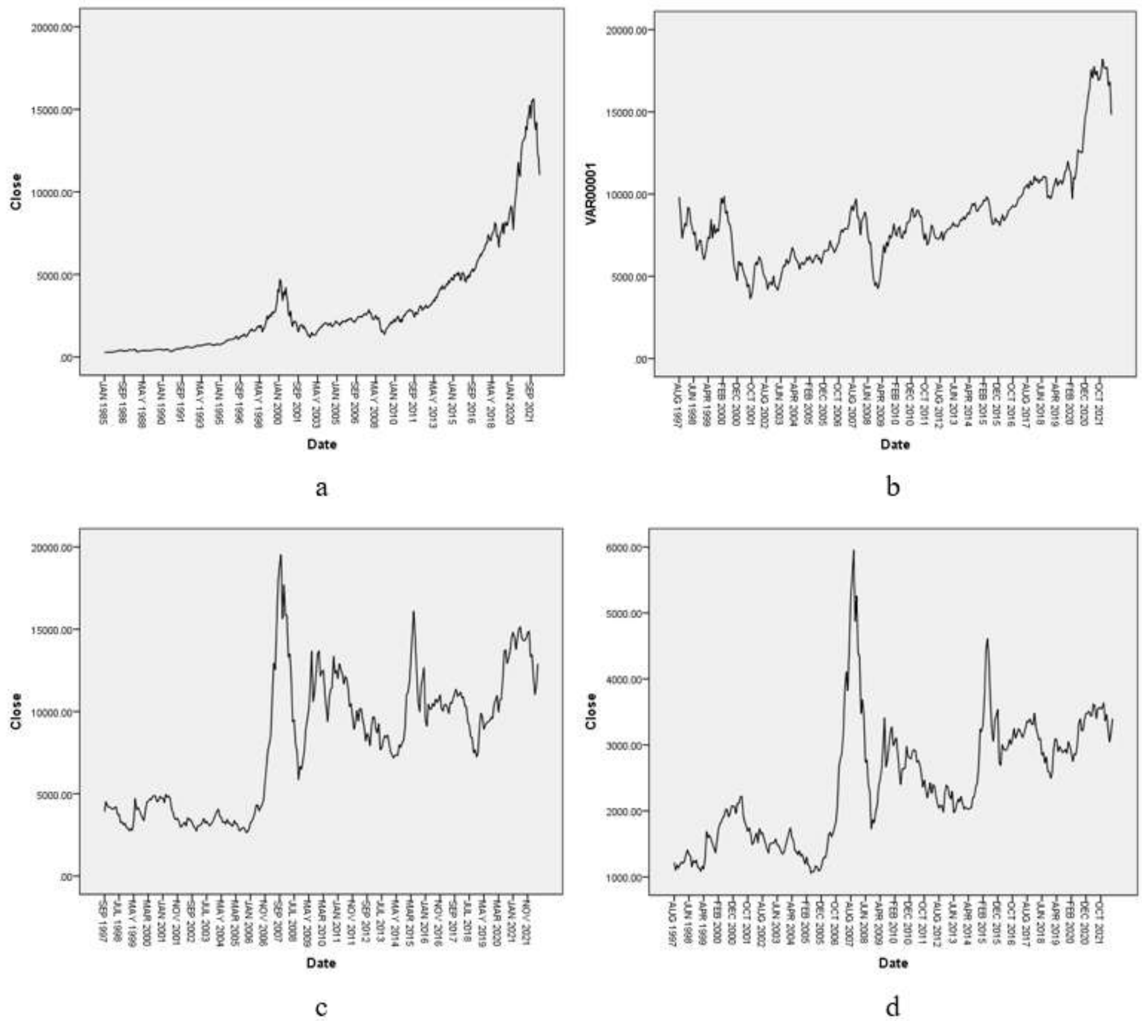


FIGURE 5.25. History of Stock Index time series data (a). NASDAQ Composite, (b). TSEC weighted, (c). Shenzhen, and (d). SSE Composite

Time series data of stock index are non-stationary, then comes to step 2, series' first difference apply on these series, where order of d is 1, (d=1). And see figure 5.26, these series are stationary at first difference.

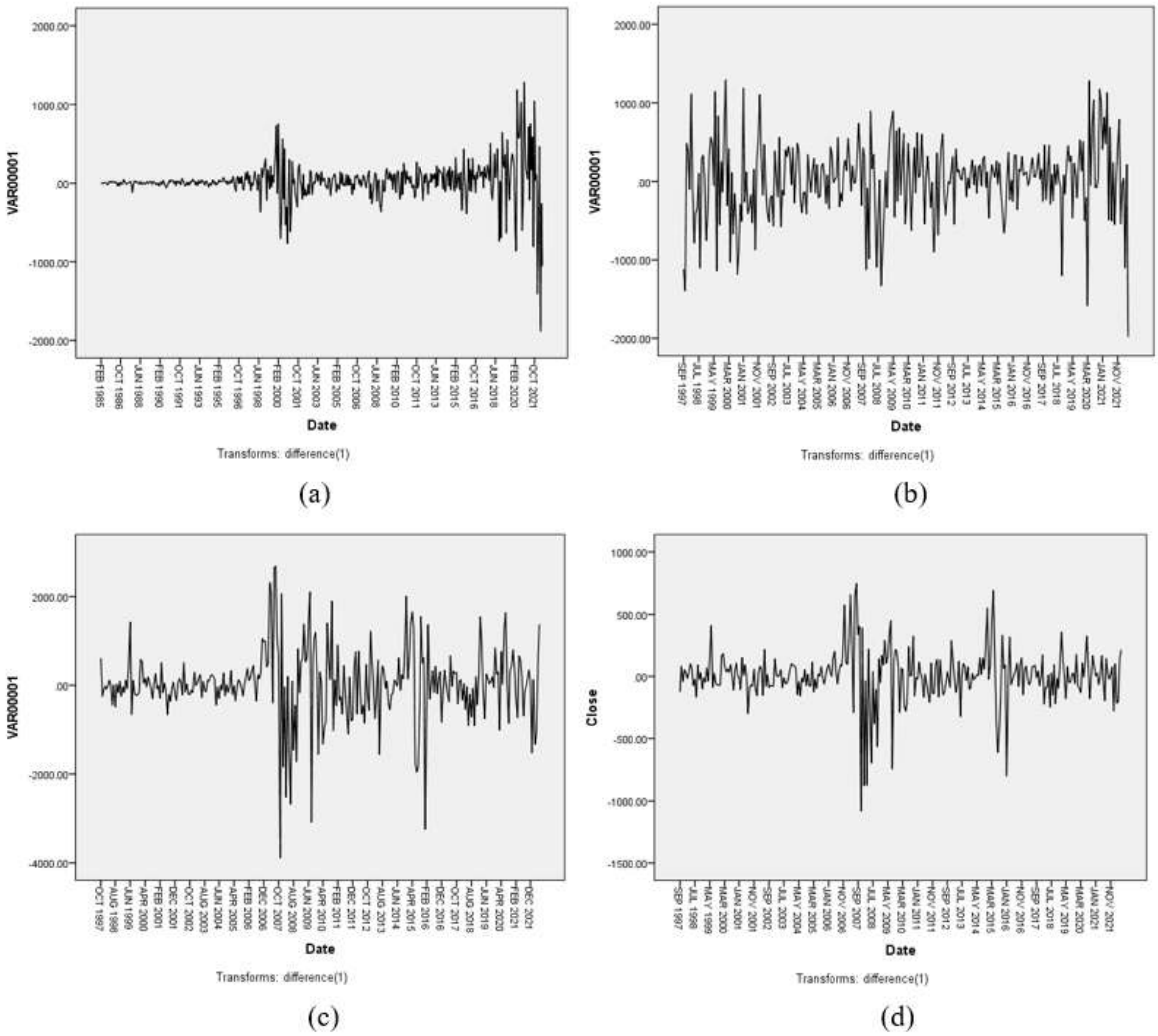


FIGURE 5.26. First series difference (d=1) of stock index (a). NASDAQ Composite, (b). TSEC weighted, (c). Shenzhen, and (d). SSE Composite.

Next step 4, graph of the ACF & Partial ACF of first series difference (d=1) as shown in the figure 5.26, where find out the order of AR & MA. Form the series NASDAQ Composite gives an order of AR & MA are 4 & 4 at first difference, respectively. First difference series of the stock TSEC weighted gives AR order is 2 and MA order also 2. Third Shenzhen stock series at the first difference order of AR is 4 and MA is 4. Similarly, Fourth SSE Composite stock series at the first difference order of AR is 2, 4 and MA also 2, 4. Possible order combination for the stock index which are considered at the first difference are shown in Table 5.14.

TABLE 5.14. Possible order combinations for the considered world stock index.

World Stock Index	Possible order combinations	Index Model Representation
NASDAQ Composite	(4,1,4)	NASDAQ_(4,1,4)
TSEC weighted	(2,1,2)	TSEC_(2,1,2)
Shenzhen	(4,1,4)	Shenzhen_(4,1,4)
SSE Composite	(2,1,2), (2,1,4), (4,1,2), (4,1,4)	SSE_(2,1,2), SSE_(2,1,4), SSE_(4,1,2), SSE_(4,1,4)

Clearly define in the above table, possible order combinations for the stock index forecasting till year 2035 by ARIMA model.

5.16.2. Result and Discussion

In the above session discussed the considered non-stationary time series, which is converted into stationary series with the first difference of series (d=1), and define possible order conjunction of the models for the forecasting analysis of the considered world stock index by use of ACF and PACF functions of the series.

Forecasting of NASDAQ Composite time series data possible order (4,1,4) of the ARIMA. By ARIMA (4,1,4) model forecasting of the series decreasing firstly than increasing. Forecasting of the series to till last of the year 2035 by ARIMA (4,1,4) model value reached to 10842.53. Possible order of forecasting as ARIMA (2,1,2) for TSEC weighted stock time series gave an increasing flow of the series and value reached 16734.53 in last of year 2035. For the Shenzhen stock index time series possible order of ARIMA is (4,1,4) as ARIMA (4,1,4), also shows the increasing flow with some zig-zag motion in the starting of forecasting, but the forecasting value of the year 2035 is 18474.78. Similarly, the SSE Composite stock market time series have four possible order combinations as (2,1,2), (2,1,4), (4,1,4), and (4,1,4). These four models ARIMA (2,1,2) reached at 4689.77, ARIMA (2,1,4) value reached 4667.41, ARIMA (4,1,4) value is 4711.49 and ARIMA (4,1,4) value reached 4578.81 in last of the year 2035. Forecasting results graphically shown in figure 5.27 & figure 5.28. Table 5.15(a) to 5.15(j), shows the value of the forecasting each considered world stock indices with respect to possible models.

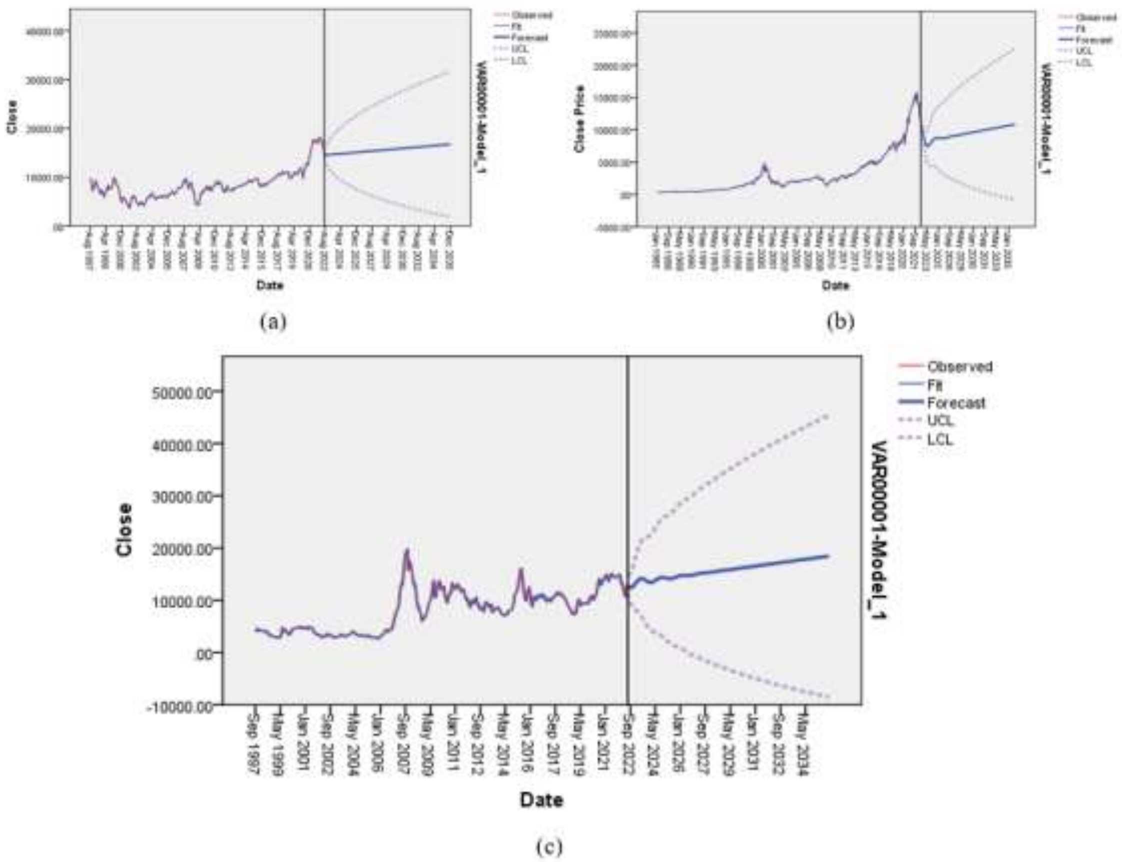


FIGURE 5.27. Forecasting of World stock indices (a). NASDAQ Composite, (b). TSEC weighted, and (c). Shenzhen to till last of year 2035 by possible order.

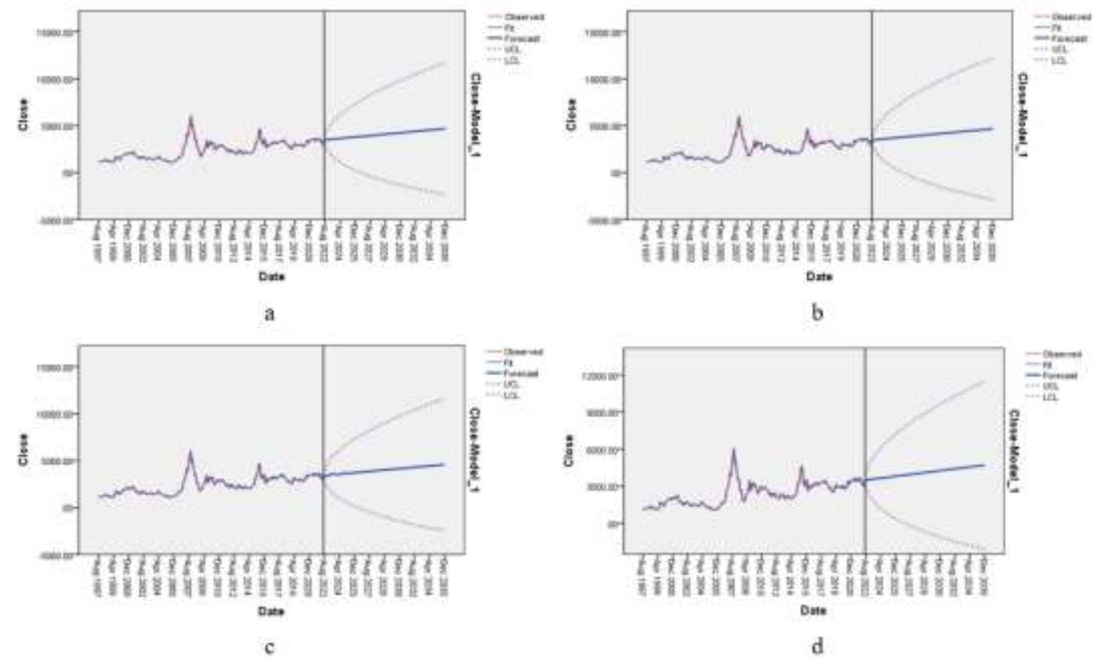


FIGURE 5.28. Forecasting of the SSE Composite stock market time series to last of year 2035 by possible order as (a). ARIMA (2,1,2), (b). ARIMA (2,1,4), (c). ARIMA (4,1,4) (d). ARIMA (4,1,2).

TABLE 5.15.a. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE_(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
3414.97	3474.43	3452.06	3427.61	12940.29	14881.73	10896.66	**Jul 2022**
3323.56	3465.35	3393.45	3462.93	12523.89	14476.92	10447.13	**Aug 2022**
3331.02	3473.44	3408.86	3477.92	12331.82	14581.33	9996.74	**Sep 2022**
3333.56	3529.48	3462.95	3496.11	12525.79	14538.80	9649.01	**Oct 2022**
3330.12	3536.49	3487.18	3506.18	12569.56	14572.43	9129.44	**Nov 2022**
3377.13	3518.68	3502.80	3517.79	12818.13	14577.08	8723.88	**Dec 2022**
3406.92	3534.00	3513.94	3526.15	13195.18	14594.57	8415.05	**Jan 2023**
3448.28	3560.60	3523.07	3535.23	13459.37	14606.80	8038.82	**Feb 2023**
3494.89	3558.95	3531.27	3543.01	13765.27	14621.27	7816.70	**Mar 2023**
3524.31	3555.37	3539.05	3551.11	14015.21	14634.80	7665.81	**Apr 2023**
3553.06	3571.92	3546.64	3558.69	14126.07	14648.73	7501.59	**May 2023**
3568.55	3585.75	3554.14	3566.42	14182.42	14662.50	7476.48	**Jun 2023**
3571.54	3585.51	3561.59	3573.94	14144.07	14676.33	7484.30	**Jul 2023**
3568.92	3590.22	3569.03	3581.52	14018.14	14690.14	7495.00	**Aug 2023**
3557.15	3603.96	3576.46	3589.01	13870.97	14703.95	7602.51	**Sep 2023**

TABLE 5.15.b. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE_(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
3542.98	3612.27	3583.88	3596.54	13705.64	14717.77	7705.40	**Oct 2023**
3529.38	3615.20	3591.31	3604.03	13554.12	14731.58	7808.57	**Nov 2023**
3517.67	3623.17	3598.73	3611.53	13455.29	14745.39	7962.97	**Dec 2023**
3511.98	3633.95	3606.15	3619.02	13410.44	14759.21	8086.84	**Jan 2024**
3512.42	3640.51	3613.57	3626.51	13429.75	14773.02	8204.85	**Feb 2024**
3519.36	3645.95	3620.99	3634.00	13513.62	14786.83	8339.59	**Mar 2024**
3532.58	3654.64	3628.42	3641.49	13642.62	14800.65	8433.60	**Apr 2024**
3550.00	3663.47	3635.84	3648.97	13801.66	14814.46	8519.79	**May 2024**
3570.09	3669.92	3643.26	3656.46	13970.58	14828.27	8603.60	**Jun 2024**
3590.74	3676.78	3650.68	3663.95	14126.08	14842.09	8649.89	**Jul 2024**
3609.88	3685.26	3658.10	3671.44	14253.33	14855.90	8691.63	**Aug 2024**
3626.20	3693.11	3665.52	3678.93	14341.06	14869.71	8724.76	**Sep 2024**
3638.66	3699.92	3672.94	3686.41	14383.87	14883.53	8731.22	**Oct 2024**
3646.97	3707.41	3680.37	3693.90	14385.29	14897.34	8739.04	**Nov 2024**
3651.47	3715.49	3687.79	3701.39	14353.34	14911.15	8739.45	**Dec 2024**
3652.91	3722.97	3695.21	3708.88	14300.11	14924.97	8725.76	**Jan 2025**
3652.45	3730.14	3702.63	3716.37	14240.44	14938.78	8718.72	**Feb 2025**

TABLE 5.15.c. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE_(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
3651.34	3737.82	3710.05	3723.85	14188.12	14952.59	8707.90	**Mar 2025**
3650.81	3745.61	3717.47	3731.34	14154.73	14966.41	8693.09	**Apr 2025**
3651.87	3753.01	3724.89	3738.83	14148.34	14980.22	8687.60	**May 2025**
3655.18	3760.42	3732.32	3746.32	14172.05	14994.04	8681.23	**Jun 2025**
3661.03	3768.10	3739.74	3753.80	14224.42	15007.85	8676.82	**Jul 2025**
3669.35	3775.72	3747.16	3761.29	14300.06	15021.66	8681.47	**Aug 2025**
3679.69	3783.16	3754.58	3768.78	14390.55	15035.48	8686.34	**Sep 2025**
3691.41	3790.67	3762.00	3776.27	14486.14	15049.29	8695.29	**Oct 2025**
3703.73	3798.31	3769.42	3783.75	14577.27	15063.10	8710.96	**Nov 2025**
3715.88	3805.86	3776.84	3791.24	14655.79	15076.92	8726.50	**Dec 2025**
3727.20	3813.34	3784.26	3798.73	14716.21	15090.73	8745.75	**Jan 2026**
3737.21	3820.90	3791.69	3806.22	14756.13	15104.54	8768.65	**Feb 2026**
3745.67	3828.49	3799.11	3813.71	14776.35	15118.36	8790.51	**Mar 2026**
3752.57	3836.02	3806.53	3821.19	14780.45	15132.17	8814.69	**Apr 2026**
3758.13	3843.54	3813.95	3828.68	14774.04	15145.98	8839.84	**May 2026**
3762.73	3851.10	3821.37	3836.17	14763.74	15159.80	8863.23	**Jun 2026**
3766.82	3858.67	3828.79	3843.66	14756.14	15173.61	8887.60	**Jul 2026**

TABLE 5.15.d. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE (2,1,4)	SSE_(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
3770.90	3866.20	3836.21	3851.14	14756.85	15187.42	8911.23	**Aug 2026**
3775.40	3873.74	3843.64	3858.63	14769.73	15201.24	8932.94	**Sep 2026**
3780.65	3881.30	3851.06	3866.12	14796.56	15215.05	8954.83	**Oct 2026**
3786.83	3888.85	3858.48	3873.61	14836.91	15228.86	8975.30	**Nov 2026**
3793.99	3896.39	3865.90	3881.09	14888.43	15242.68	8994.23	**Dec 2026**
3802.02	3903.93	3873.32	3888.58	14947.38	15256.49	9013.08	**Jan 2027**
3810.71	3911.49	3880.74	3896.07	15009.24	15270.30	9030.59	**Feb 2027**
3819.80	3919.03	3888.16	3903.56	15069.52	15284.12	9047.20	**Mar 2027**
3828.96	3926.58	3895.59	3911.05	15124.33	15297.93	9063.84	**Apr 2027**
3837.94	3934.13	3903.01	3918.53	15170.96	15311.74	9079.56	**May 2027**
3846.49	3941.68	3910.43	3926.02	15208.10	15325.56	9095.05	**Jun 2027**
3854.49	3949.22	3917.85	3933.51	15235.97	15339.37	9110.73	**Jul 2027**
3861.89	3956.77	3925.27	3941.00	15256.09	15353.18	9125.98	**Aug 2027**
3868.72	3964.32	3932.69	3948.48	15270.99	15367.00	9141.45	**Sep 2027**
3875.10	3971.86	3940.11	3955.97	15283.71	15380.81	9157.23	**Oct 2027**
3881.19	3979.41	3947.54	3963.46	15297.34	15394.62	9172.92	**Nov 2027**
3887.19	3986.96	3954.96	3970.95	15314.56	15408.44	9189.01	**Dec 2027**

TABLE 5.15.e. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE_(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
3893.26	3994.50	3962.38	3978.43	15337.28	15422.25	9205.42	**Jan 2028**
3899.55	4002.05	3969.80	3985.92	15366.44	15436.06	9221.89	**Feb 2028**
3906.18	4009.60	3977.22	3993.41	15401.96	15449.88	9238.77	**Mar 2028**
3913.18	4017.15	3984.64	4000.90	15442.85	15463.69	9255.85	**Apr 2028**
3920.53	4024.69	3992.06	4008.39	15487.41	15477.51	9273.00	**May 2028**
3928.19	4032.24	3999.48	4015.87	15533.58	15491.32	9290.46	**Jun 2028**
3936.05	4039.79	4006.91	4023.36	15579.25	15505.13	9307.97	**Jul 2028**
3944.01	4047.34	4014.33	4030.85	15622.56	15518.95	9325.52	**Aug 2028**
3951.94	4054.88	4021.75	4038.34	15662.19	15532.76	9343.21	**Sep 2028**
3959.75	4062.43	4029.17	4045.82	15697.44	15546.57	9360.87	**Oct 2028**
3967.38	4069.98	4036.59	4053.31	15728.34	15560.39	9378.49	**Nov 2028**
3974.78	4077.52	4044.01	4060.80	15755.52	15574.20	9396.16	**Dec 2028**
3981.95	4085.07	4051.43	4068.29	15780.12	15588.01	9413.72	**Jan 2029**
3988.93	4092.62	4058.86	4075.77	15803.53	15601.83	9431.23	**Feb 2029**
3995.76	4100.17	4066.28	4083.26	15827.21	15615.64	9448.72	**Mar 2029**
4002.51	4107.71	4073.70	4090.75	15852.43	15629.45	9466.09	**Apr 2029**
4009.26	4115.26	4081.12	4098.24	15880.11	15643.27	9483.42	**May 2029**

TABLE 5.15.f. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE_(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
4016.07	4122.81	4088.54	4105.73	15910.77	15657.08	9500.72	**Jun 2029**
4022.99	4130.35	4095.96	4113.21	15944.41	15670.89	9517.93	**Jul 2029**
4030.04	4137.90	4103.38	4120.70	15980.62	15684.71	9535.11	**Aug 2029**
4037.24	4145.45	4110.81	4128.19	16018.64	15698.52	9552.27	**Sep 2029**
4044.56	4153.00	4118.23	4135.68	16057.53	15712.33	9569.38	**Oct 2029**
4051.99	4160.54	4125.65	4143.16	16096.30	15726.15	9586.50	**Nov 2029**
4059.47	4168.09	4133.07	4150.65	16134.06	15739.96	9603.61	**Dec 2029**
4066.96	4175.64	4140.49	4158.14	16170.16	15753.77	9620.70	**Jan 2030**
4074.43	4183.19	4147.91	4165.63	16204.25	15767.59	9637.81	**Feb 2030**
4081.84	4190.73	4155.33	4173.11	16236.30	15781.40	9654.93	**Mar 2030**
4089.17	4198.28	4162.76	4180.60	16266.58	15795.21	9672.06	**Apr 2030**
4096.40	4205.83	4170.18	4188.09	16295.58	15809.03	9689.21	**May 2030**
4103.56	4213.37	4177.60	4195.58	16323.95	15822.84	9706.37	**Jun 2030**
4110.65	4220.92	4185.02	4203.07	16352.37	15836.65	9723.55	**Jul 2030**
4117.70	4228.47	4192.44	4210.55	16381.44	15850.47	9740.75	**Aug 2030**
4124.73	4236.02	4199.86	4218.04	16411.62	15864.28	9757.95	**Sep 2030**

TABLE 5.15.g. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE_(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
4131.77	4243.56	4207.28	4225.53	16443.17	15878.09	9775.17	**Oct 2030**
4138.84	4251.11	4214.70	4233.02	16476.11	15891.91	9792.40	**Nov 2030**
4145.96	4258.66	4222.13	4240.50	16510.28	15905.72	9809.63	**Dec 2030**
4153.14	4266.21	4229.55	4247.99	16545.33	15919.54	9826.87	**Jan 2031**
4160.37	4273.75	4236.97	4255.48	16580.85	15933.35	9844.11	**Feb 2031**
4167.64	4281.30	4244.39	4262.97	16616.36	15947.16	9861.35	**Mar 2031**
4174.94	4288.85	4251.81	4270.45	16651.45	15960.98	9878.59	**Apr 2031**
4182.26	4296.39	4259.23	4277.94	16685.79	15974.79	9895.83	**May 2031**
4189.58	4303.94	4266.65	4285.43	16719.21	15988.60	9913.07	**Jun 2031**
4196.87	4311.49	4274.08	4292.92	16751.68	16002.42	9930.30	**Jul 2031**
4204.14	4319.04	4281.50	4300.41	16783.30	16016.23	9947.53	**Aug 2031**
4211.38	4326.58	4288.92	4307.89	16814.29	16030.04	9964.75	**Sep 2031**
4218.59	4334.13	4296.34	4315.38	16844.96	16043.86	9981.97	**Oct 2031**
4225.76	4341.68	4303.76	4322.87	16875.61	16057.67	9999.19	**Nov 2031**
4232.92	4349.23	4311.18	4330.36	16906.54	16071.48	10016.40	**Dec 2031**
4240.06	4356.77	4318.60	4337.84	16937.96	16085.30	10033.62	**Jan 2032**

TABLE 5.15.h. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE_(2,1,2)	Shenzhen _(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
4247.20	4364.32	4326.03	4345.33	16970.01	16099.11	10050.83	**Feb 2032**
4254.35	4371.87	4333.45	4352.82	17002.70	16112.92	10068.03	**Mar 2032**
4261.51	4379.41	4340.87	4360.31	17035.99	16126.74	10085.24	**Apr 2032**
4268.70	4386.96	4348.29	4367.79	17069.71	16140.55	10102.45	**May 2032**
4275.91	4394.51	4355.71	4375.28	17103.66	16154.36	10119.65	**Jun 2032**
4283.13	4402.06	4363.13	4382.77	17137.64	16168.18	10136.86	**Jul 2032**
4290.37	4409.60	4370.55	4390.26	17171.44	16181.99	10154.07	**Aug 2032**
4297.62	4417.15	4377.98	4397.75	17204.91	16195.80	10171.28	**Sep 2032**
4304.87	4424.70	4385.40	4405.23	17237.96	16209.62	10188.49	**Oct 2032**
4312.11	4432.25	4392.82	4412.72	17270.56	16223.43	10205.70	**Nov 2032**
4319.35	4439.79	4400.24	4420.21	17302.76	16237.24	10222.91	**Dec 2032**
4326.58	4447.34	4407.66	4427.70	17334.66	16251.06	10240.13	**Jan 2033**
4333.79	4454.89	4415.08	4435.18	17366.39	16264.87	10257.34	**Feb 2033**
4340.99	4462.43	4422.50	4442.67	17398.09	16278.68	10274.55	**Mar 2033**
4348.18	4469.98	4429.92	4450.16	17429.90	16292.50	10291.77	**Apr 2033**
4355.37	4477.53	4437.35	4457.65	17461.94	16306.31	10308.98	**May 2033**

TABLE 5.15.i. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE _(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
4362.55	4485.08	4444.77	4465.13	17494.26	16320.12	10326.20	**Jun 2033**
4369.74	4492.62	4452.19	4472.62	17526.88	16333.94	10343.42	**Jul 2033**
4376.93	4500.17	4459.61	4480.11	17559.79	16347.75	10360.63	**Aug 2033**
4384.12	4507.72	4467.03	4487.60	17592.90	16361.57	10377.85	**Sep 2033**
4391.32	4515.27	4474.45	4495.09	17626.14	16375.38	10395.06	**Oct 2033**
4398.53	4522.81	4481.87	4502.57	17659.40	16389.19	10412.28	**Nov 2033**
4405.75	4530.36	4489.30	4510.06	17692.58	16403.01	10429.49	**Dec 2033**
4412.97	4537.91	4496.72	4517.55	17725.62	16416.82	10446.71	**Jan 2034**
4420.20	4545.45	4504.14	4525.04	17758.47	16430.63	10463.92	**Feb 2034**
4427.42	4553.00	4511.56	4532.52	17791.11	16444.45	10481.14	**Mar 2034**
4434.64	4560.55	4518.98	4540.01	17823.56	16458.26	10498.35	**Apr 2034**
4441.86	4568.10	4526.40	4547.50	17855.86	16472.07	10515.57	**May 2034**
4449.08	4575.64	4533.82	4554.99	17888.07	16485.89	10532.78	**Jun 2034**
4456.28	4583.19	4541.25	4562.47	17920.27	16499.70	10549.99	**Jul 2034**
4463.49	4590.74	4548.67	4569.96	17952.51	16513.51	10567.21	**Aug 2034**

TABLE 5.15.j. Forecasted result of the proposed model of considered index

SSE_(4,1,4)	SSE_(4,1,2)	SSE_(2,1,4)	SSE_(2,1,2)	Shenzhen_(4,1,4)	TSEC_(2,1,2)	NASDAQ_(4,1,4)	Forecast
4470.69	4598.28	4556.09	4577.45	17984.84	16527.33	10584.42	**Sep 2034**
4477.89	4605.83	4563.51	4584.94	18017.31	16541.14	10601.63	**Oct 2034**
4485.09	4613.38	4570.93	4592.43	18049.92	16554.95	10618.85	**Nov 2034**
4492.29	4620.93	4578.35	4599.91	18082.66	16568.77	10636.06	**Dec 2034**
4499.49	4628.47	4585.77	4607.40	18115.51	16582.58	10653.27	**Jan 2035**
4506.69	4636.02	4593.20	4614.89	18148.41	16596.39	10670.49	**Feb 2035**
4513.90	4643.57	4600.62	4622.38	18181.33	16610.21	10687.70	**Mar 2035**
4521.11	4651.12	4608.04	4629.86	18214.23	16624.02	10704.92	**Apr 2035**
4528.32	4658.66	4615.46	4637.35	18247.06	16637.83	10722.13	**May 2035**
4542.75	4673.76	4630.30	4652.33	18312.44	16665.46	10756.56	**Jul 2035**
4549.97	4681.30	4637.72	4659.81	18344.99	16679.27	10773.77	**Aug 2035**
4557.18	4688.85	4645.14	4667.30	18377.47	16693.09	10790.99	**Sep 2035**
4564.39	4696.40	4652.57	4674.79	18409.91	16706.90	10808.20	**Oct 2035**
4571.60	4703.95	4659.99	4682.28	18442.34	16720.71	10825.41	**Nov 2035**
4578.81	4711.49	4667.41	4689.77	18474.78	16734.53	10842.63	**Dec 2035**

Forecasting of the world indices which are considered and get the future forecast result which is discussed above at the possible models according to non-stationary time series. Proposed result of the forecasting analysis models, evaluations analysis of model shown in the table 5.16.

TABLE 5.16. Performance evaluation of ARIMA (p,d,q)

World Stock Index	ARIMA (p,d,q)	RMSE	MAPE	MAE/ MAD	R-square	BIC	Ljung-Box Q (18)		
							Statistic	DF	Sig.
NASDAQ Composite	(4,1,4)	250.193	4.818	137.952	0.994	4.818	20.264	10	0.027
TSEC weighted	(2,1,2)	485.879	4.696	361.073	0.972	12.468	16.845	14	0.265
Shenzhen	(4,1,4)	785.382	6.038	507.760	0.963	13.505	19.759	10	0.032
SSE Composite	(2,1,2)	209.992	5.313	132.813	0.948	10.790	34.540	14	0.002
	(2,1,4)	206.194	5.355	133.837	0.950	10.791	21.240	12	0.047
	(4,1,2)	207.236	5.282	132.124	0.950	10.802	25.868	12	0.011
	(4,1,4)	203.531	5.313	132.464	0.952	10.804	12.390	10	0.260

Forecasting analysis of the stock index by differ or possible models according to non-stationary time series, the model ARIMA (4,1,4) at NASDAQ COMPOSITE stock index close value reached to 10842.63 and ARIMA (2,1,2), TSEC weighted stock index close value reached in the future as 16734.53, Shenzhen stock index value increase in future and reached to 18474.7 at ARIMA (2,1,2), and SSE Composite index analysis by four model, close value reached 4689.77 by ARIMA(2,1,2), 4667.41 by ARIMA(2,1,4), 4711.49 by ARIMA (4,1,2), and 4578.81 by ARIMA (4,1,4) to till last of the year 2035. Above table 5.16, shows the performance evaluations of the model, 203.531 is lowest RMSE value at ARIMA (4,1,4), but ARIMA (2,1,2) gives 10.790 of BIC, 0.948 of R-square values and 132.124 MAE value in SSE composite stock index forecasting models. 4.818 is lowest BIC value, which is near to zero at ARIMA (4,1,4) on the NASDAQ Composite stock market index.

5.16.3. Conclusion

Forecasting analysis of stock indexes by possible models, NASDAQ COMPOSITE stock index close value will be reached to 10842.6, TSEC weighted stock index reached up to 16734.53 and 18474.7 of Shenzhen stock index value in future. SSE Composite index forecasting by ARIMA(4,1,4) has a lowest RMSE value compare to other possible models.

CHAPTER

VI

Forecasting with Regression Model

6.1. REGRESSION MODEL

Forecasting with regression models is a statistical method used to predict future outcomes based on past performance. It uses a linear regression model to make predictions based on the relationships between (v) independent and (u) dependent variables. Regression models are used to estimate the values of one variable based on the values of one or more other variables.

The basic idea behind regression modeling is that the dependent variable can be predicted from the independent variables. The model assumes that there is a linear relationship between the independent and the dependent variables. This means that the dependent variable is a function of the independent variables and that the value of the dependent variable can be determined from the values of the independent variables.

Mathematically, regression models are expressed as equations of the form

$$u = f(v)$$

where u and v are dependent and independent variables, respectively. The exact form of the equation depends on the type of regression model used. Example, a linear regression model is expressed as

$$u = av + b,$$

Coefficients a and b, represent the strength of the relationship between the independent and dependent variables. In addition, other types of regression models, such as polynomial regression models, may be used to better fit the data.

Once the regression model is developed, it can be used to forecast future values of the dependent variable. This is done by entering values of u into the equation and computing the corresponding value of v. The resulting values can then be used to estimate future trends in the dependent variable.

Forecasting with regression models can be a powerful tool for predicting future trends. However, the accuracy of the forecasts depends on the quality of the data and the accuracy of the model. If the data is unreliable or the model is poorly constructed, the forecasts will not be accurate.

Regression models use linear equations to predict a response variable given a set of predictor variables. The most common type of regression model is linear regression, which uses a linear equation to model the relationship between the response variable and one or more predictor variables. Other types of regression models include logistic regression, polynomial regression, and multivariate regression. This model can be used to develop forecasts, identify relationships between variables, and evaluate the impact of changes in one or more variables on an outcome.

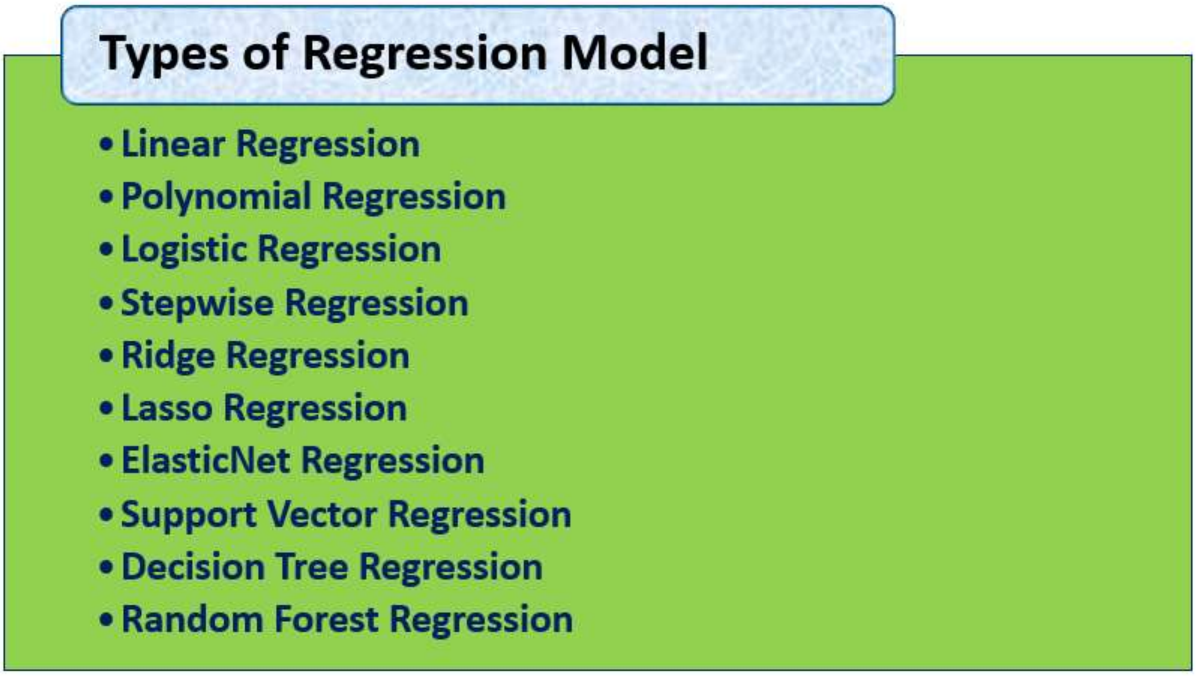


FIGURE 6.1. Types of Regression Model

6.2. LINEAR REGRESSION

Linear regression is a statistical technique used to predict numerical values by establishing a linear relationship between a dependent and one or more than one independent variables. It is one of the oldest and most widely used predictive techniques in both academic and business settings.

The basic mathematics behind linear regression is that a straight line can be used to describe the relationship between two variables, one independent ($\mathcal{X}$) and one dependent ($\mathcal{Y}$). The equation of the line is typically written as:

$$\mathcal{Y} = \beta_0 + \beta_1\mathcal{X}$$

where β_0 is the intercept, β_1 is the slope.

The formula for linear regression is a generalization of the two-variable linear equation. It is written as:

$$\mathcal{Y} = \beta_0 + \beta_1\mathcal{X}_1 + \beta_2\mathcal{X}_2 + \cdots + \beta_n\mathcal{X}_n$$

where intercept is b_0, and slopes are $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$, and $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ are the independent variables.

6.3. VARIABLES: DEPENDENT AND INDEPENDENT

Regression models are a type of statistical model used to assess the relationship between one or more independent (predictors) and one or more dependent variables (outcomes). The independent variables are the variables that are used to predict the dependent variable, and the dependent variables are that being predicted.

Independent variables are the predictors or inputs of the regression model and are usually denoted by $\mathcal{V}$. They are the causes that are used to explain the variation in the dependent variable. Independent variables can be continuous, discrete, or categorical.

Dependent variables are the outcomes or responses of the regression model and are usually denoted by $\mathcal{U}$. They are the variables that are being predicted by the independent variables. Dependent variables are usually continuous but can also be discrete or categorical.

Simple and multiple linear regression are the two types of linear regression.

FIGURE 6.2. Types of Linear Regression

6.3.1. Simple Linear Regression Model

In linear regression, the goal is to determine a linear equation that best describes the relationship between the input (or independent) and the output (or dependent) variable. The equation is used to make predictions about the output variable based on the input variables.

The mathematical formula for a simple linear regression model is,

$$\mathcal{U} = \beta_0 + \beta_1 \mathcal{V}$$

$\mathcal{U}$ is the response variable (dependent) and $\mathcal{V}$ is the predictor (independent) variable, and β_0 and β_1 are the regression coefficients.

6.3.2. Multiple Linear Regression Model

This model is an extension of a simple linear regression model that is used to predict a dependent variable using more than one independent variable. It is used to explain the relationship between one dependent and two or more independent variables, allowing for a more accurate prediction of the outcome of the dependent variable. The multiple linear regression model is useful for analyzing complex relationships between variables, and can help to identify the impact of different independent variables on the dependent variable.

Mathematically, model takes the following form:

$$\mathcal{U} = \beta_0 + \beta_1\mathcal{V}_1 + \beta_1\mathcal{V}_2 + \cdots + \beta_n\mathcal{V}_n + \varepsilon$$

Where, the intercept is denoted by β_0, $\mathcal{U}$ denote dependent variable, independent variables denote by $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_n$, and β_1, β_2, …, β_n are the coefficients corresponding to the independent variables, and ε tell the error.

It can be used to make forecasts based on past data, to identify relationships between independent and dependent variables, and to test hypotheses about the relationships between the variables. It can also be used to identify non-linear relationships between variables.

6.4. MODEL: FORECASTING OF THE CROP PRODUCTION (RICE, WHEAT, POTATO) IN INDIA

The forecasting of crop production (rice, wheat, and potato) of India can be done using regression models. [150] Past production data of the three crops taken from 2000 to 2019 for rice, wheat, and 2000 to 2018 for potato. Rice data shown in table 6.1 and figure 6.3 represent a graphical form.

TABLE 6.1. Indian Rice Production data (Yearly)

Year	Production (MT)	Year	Production (MT)	Year	Production (MT)
2000	84977	2007	96682	2014	105482
2001	93334	2008	99172	2015	104408

2002	71814	2009	89083	2016	109698
2003	88552	2010	95970	2017	112760
2004	83127	2011	105301	2018	116420
2005	91785	2012	105241	2019	115000
2006	93345	2013	106646		

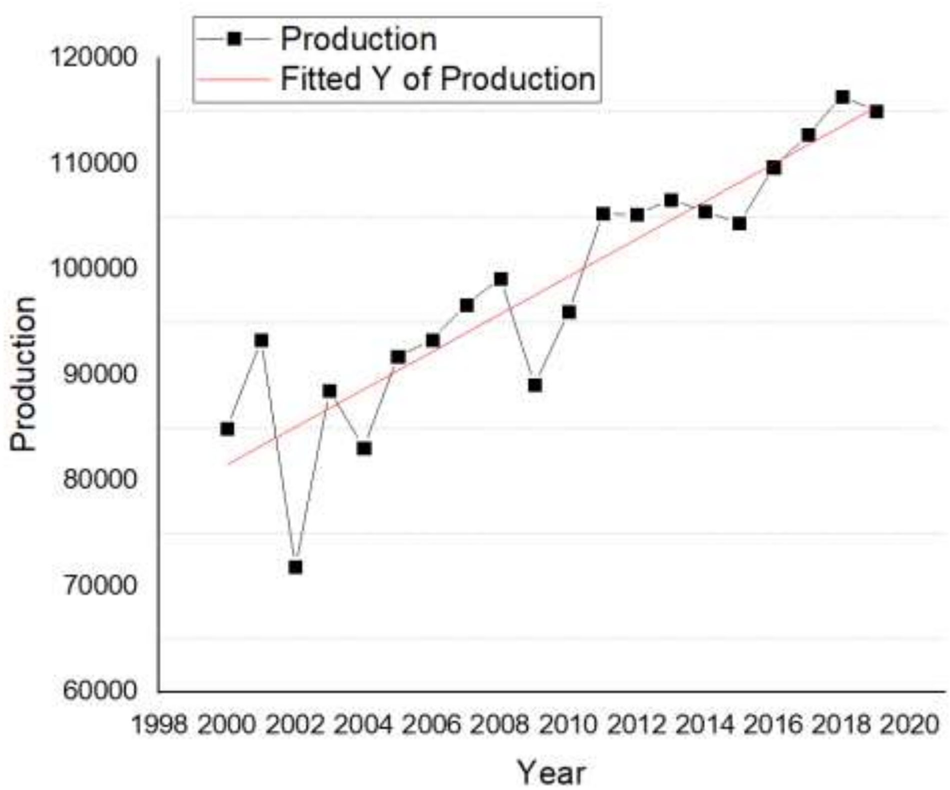


FIGURE 6.3. Indian Rice Production data

The regression model applied on production data of rice of India from 2000 to 2019, then get the output summary in Figure 6.4.

Regression Statistics	
Multiple R	0.9029207
R Square	0.8152658
Adjusted R Square	0.8050028
Standard Error	5156.9379
Observations	20

ANOVA

	df	SS	MS	F	Significance F
Regression	1	2112555736	2.11E+09	79.43728	5.09038E-08
Residual	18	478692154.6	26594009		
Total	19	2591247891			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-3483195	401856.3897	-8.66776	7.68E-08	-4327463.686	-2638925.8	-4327463.686	-2638925.8
Year	1782.3511	199.9774746	8.912759	5.09E-08	1362.214045	2202.48821	1362.214045	2202.488211

FIGURE 6.4. Output summary of the Rice Production

From 2000-2019 India's wheat production data is shown in Table 6.2. Figure 6.5 is a graphical representation of wheat data with a linear equation.

TABLE 6.2. Production data of Wheat

Year	Area Harvested (HA)	Production (MT)	Year	Area Harvested (1000/HA)	Production (MT)	Year	Area Harvested (HA)	Production (MT)
2000	27486	76369	2007	27995	75807	2014	30473	95850
2001	25731	69681	2008	28039	78570	2015	31466	86527
2002	26345	72766	2009	27752	80679	2016	30220	87000
2003	25196	65761	2010	28457	80804	2017	30785	98510
2004	26595	72156	2011	29069	86874	2018	29651	99870
2005	26383	68637	2012	29865	94882	2019	29850	102190
2006	26484	69355	2013	30003	93506			

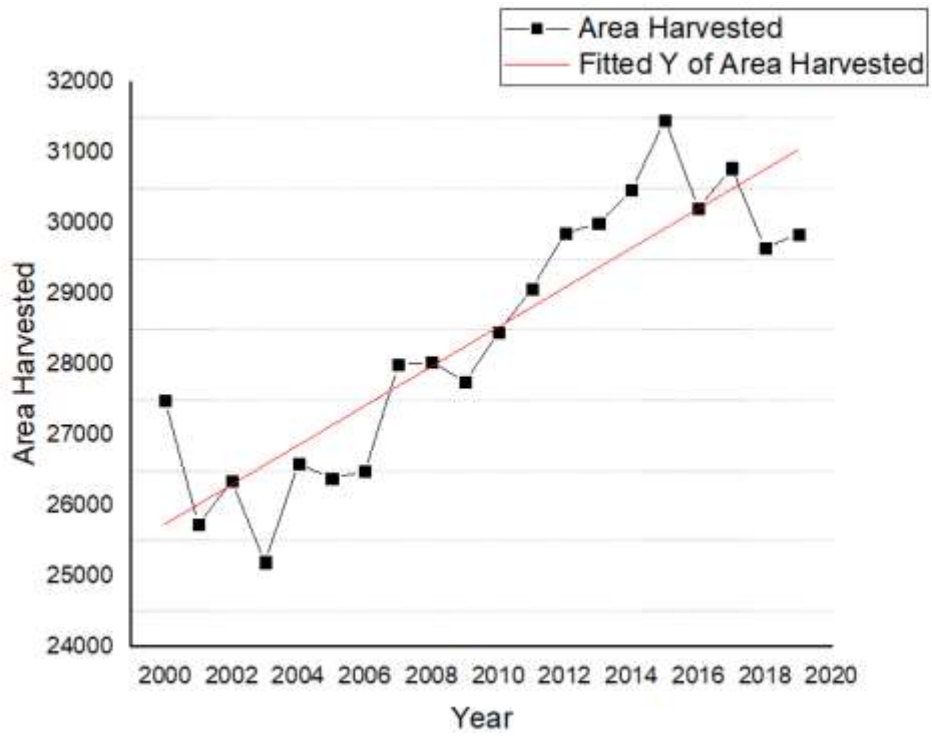


FIGURE 6.5. Area Harvested (Wheat)

Linear equation of wheat production from 2000-2019 is shown in figure 6.6 and production (Million Tons) over the area harvested is shown in a graphical form, figure 6.7.

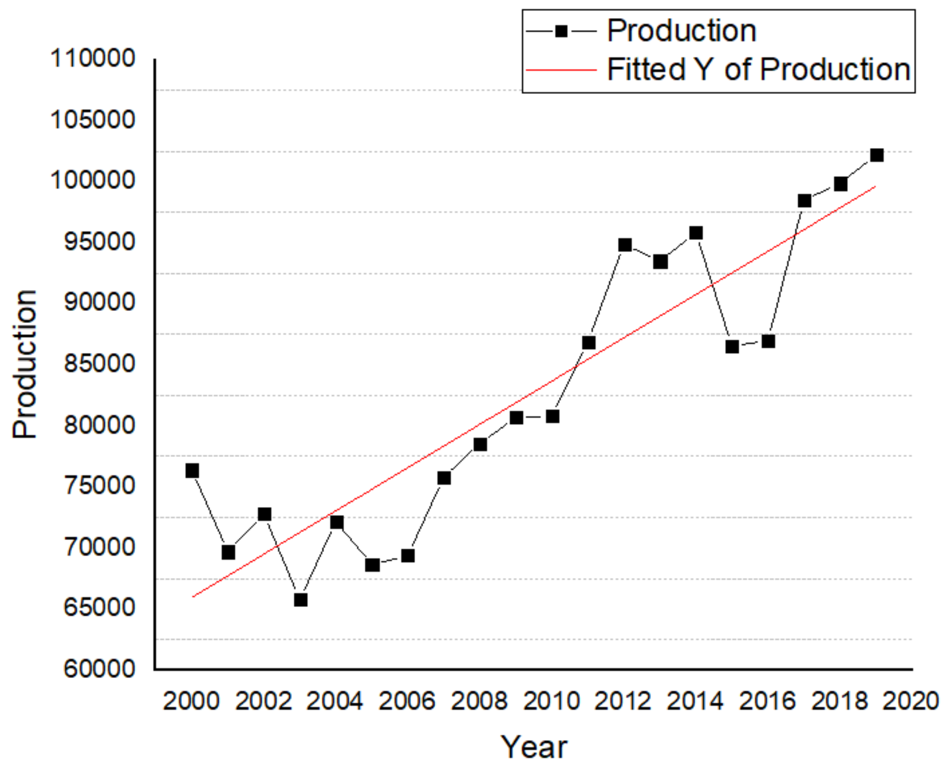


FIGURE 6.6. Production of Wheat

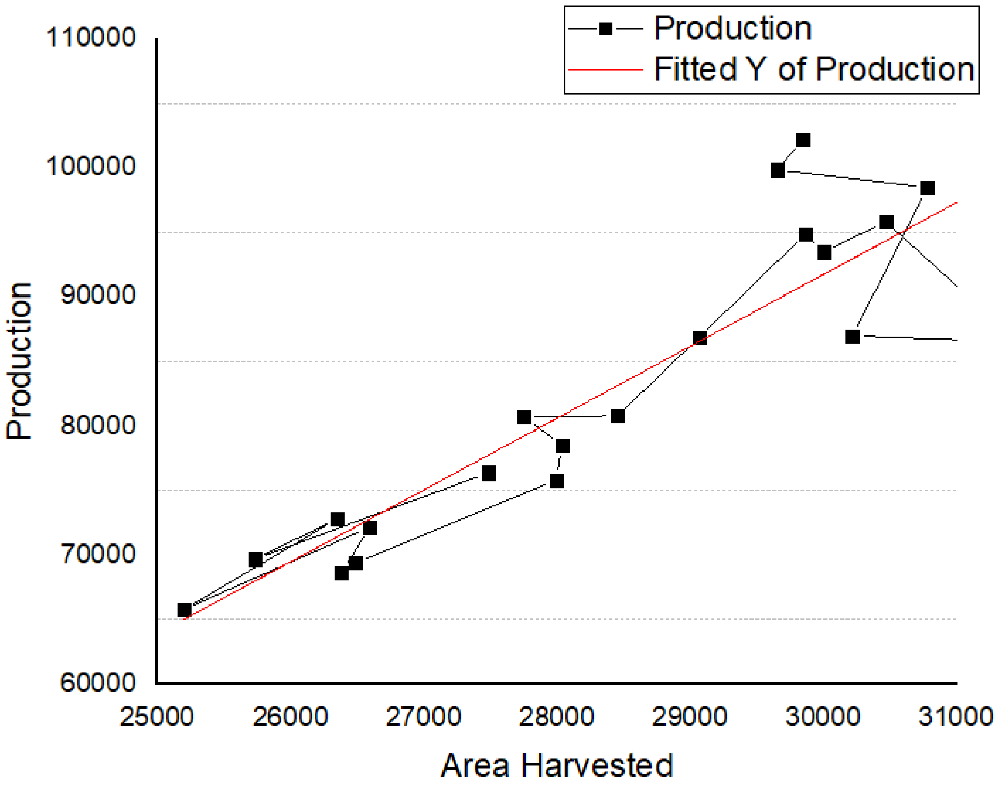


FIGURE 6.7. Production over area harvested

Regression Statistics	
Multiple R	0.8909714
R Square	0.79383
Adjusted R Square	0.7823761
Standard Error	5422.9945
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2038228530	2.04E+09	69.306593	1.38343E-07
Residual	18	529359648.2	29408869		
Total	19	2567588178			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-75362.43	19035.78319	-3.95899	0.0009202	-115355.1311	-35369.7383	-115355.1311	-35369.7383
AREA HARVESTED	5.5702572	0.669095291	8.325058	1.383E-07	4.164540146	6.97597423	4.164540146	6.975974231

FIGURE 6.8. Output summary of the Wheat Production

Similarly, for the calculation of the production of potato in India from 2000 to 2018 by a Regression model. Table 6.3. Show a year-wise production of potato in India from 2000 to 2018. Figure 6.3 is a graphical representation of rice data with a linear equation.

TABLE 6.3. Potato production in India (year-wise)

Year	Yield (kg/HA)	Area (Million HA)	Production (MT)	Year	Yield (kg/HA)	Area (Million HA)	Production (MT)
2000	17886	1.32	23.61	2010	18810	1.83	34.39
2001	18443	1.34	24.71	2011	19951	1.84	36.58
2002	18404	1.22	22.49	2012	22724	1.86	42.34
2003	19806	1.21	23.92	2013	21753	1.91	41.48
2004	17300	1.35	23.27	2014	22760	1.99	45.34
2005	17887	1.29	23.06	2015	21060	1.97	41.56
2006	17923	1.32	23.63	2016	23126	2.08	48.01
2007	17508	1.4	23.91	2017	20509	2.12	43.42
2008	14943	1.48	22.18	2018	22303	2.18	48.6
2009	18331	1.55	28.47				

Regression Statistics	
Multiple R	0.9991367
R Square	0.9982741
Adjusted R Square	0.9980583
Standard Error	0.4438231
Observations	19

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1822.912305	911.4562	4627.175	7.87391E-23
Residual	16	3.151663245	0.196979		
Total	18	1826.063968			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-33.34131	0.957014425	-34.8389	1.62E-16	-35.3700904	-31.312531	-35.37009	-31.312531
AREA	19.891549	0.479727872	41.46423	1.03E-17	18.87457091	20.908526	18.8745709	20.9085262
YIELD	0.0017033	7.23497E-05	23.54237	7.64E-14	0.001549908	0.0018567	0.00154991	0.00185666

FIGURE 6.9. Output summary of the Potato Production

6.4.1. Result and Discussion

Forecasting of crop production for the year 2020 with a regression model with the hypothesis. In the production of rice, a hypothesis is rejected at a significance level should be less than 0.5. and R square value is 0.8153 calculated with standard error 5156.938. The standard error in intercept and Year are 401856.4 and 199.9775, respectively. The Forecast of rice production is 117154.02 (1000 MT) for the year 2020. The crop of Wheat production output expression shows the R square value is closer to 1. Suppose the area is 30395 (1000/HA) then the forecast for the production of wheat is 93947.2685 (1000 MT) for the year 2020. Production of Potato depends on the area and yield of crops where to apply the multiple linear regression model for calculate the Potato production of data from 2000-2018. Calculated values of R, R square, and estimate the standard error are 0.999137, 0.998274, and 0.443823, respectively. And higher significant level ($p<0.05$). If Potato production such as Area = 2.5 (Million HA) and Yield = 206000 (kg/HA) then Predicted production is 51.47 (Million Tons). Such as selected crops of India and its Regression line equation for each crop production by changing area, yield, and year in variables:

1. Wheat, $\mathcal{Y} = -3483195 + 1782.351\mathcal{X}$
2. Rice, $\mathcal{Y} = -75362.4 + 5.5703\mathcal{X}$
3. Potato, $\mathcal{Y} = -33.3413 + 19.89155\mathcal{X}_1 + 0.001703\mathcal{X}_2$

Where dependent variables $\mathcal{Y}$ and $\mathcal{X}, \mathcal{X}_1, \mathcal{X}_2$ are independent variables.

6.4.2. Conclusion

From this analysis, we conclude that the forecasting for production analysis of rice in India 2020 is described as an increasing factor with respect to the previous years, which is highest in recorded data. In present, the study of wheat production analysis significance is short according to the last three years of the data. The most demanded crop in the market is Potato. So, forecasting of potato production in India are high progress compare to the last twenty years.

REFERENCES

1. Joshi, B. P., & Kumar, S. (2012). A computational method of forecasting based on intuitionistic fuzzy sets and fuzzy time series. In Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011, 993-1000. Springer, New Delhi.
2. Yabuuchi, Y., Watada, J., & Toyoura, Y. (2004). Fuzzy AR model of stock price, *Scientiae Mathematicae Japonicae*, **60**(2), 303-310.
3. Chen, S. M., & Chen, C. D. (2010). TAIEX forecasting based on fuzzy time series and fuzzy variation groups, *IEEE Transactions on Fuzzy Systems*, **19**(1), 1-12.
4. Huarng, K., & Yu, T. H. K. (2006). Ratio-based lengths of intervals to improve fuzzy time series forecasting, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **36**(2), 328-340.
5. Huarng, K. H., Yu, T. H. K., & Hsu, Y. W. (2007). A multivariate heuristic model for fuzzy time-series forecasting. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **37**(4), 836-846.
6. Kumar, D. A., & Murugan, S. (2013, February). Performance analysis of Indian stock market index using neural network time series model. *In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, 72-78. IEEE.
7. Chen, S. M., Chu, H. P., & Sheu, T. W. (2012). TAIEX forecasting using fuzzy time series and automatically generated weights of multiple factors. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, **42**(6), 1485-1495.
8. Abdel-Basset, M., Chang, V., Mohamed, M., & Smarandache, F. (2019). A refined approach for forecasting based on neutrosophic time series. *Symmetry*, **11**(4), 457.
9. Das, D., Mahanta, S., Chutia, R., & Baruah, H. K. (2013). Construction of normal fuzzy numbers: case studies with stock exchange data. *Annals of Fuzzy Mathematics and Informatics*, **5**(3), 505-514.
10. Singh, P. (2017). A brief review of modeling approaches based on fuzzy time series. *International Journal of Machine Learning and Cybernetics*, **8**(2), 397-420.
11. Chou, M. T. (2018). Fuzzy Forecast Based on Fuzzy Time Series, *Time Series Analysis-Data, Methods, and Applications*. IntechOpen.
12. Garg, B., & Garg, R. (2016). Enhanced accuracy of fuzzy time series model using ordered weighted aggregation, *Applied Soft Computing*, **48**, 265-280.

13. Liu, H. T., Wei, N. C., & Yang, C. G. (2009). Improved time-variant fuzzy time series forecast, *Fuzzy Optimization and Decision Making*, **8**(1), 45-65.
14. Krollner, B., Vanstone, B. J., & Finnie, G. R. (2010, April). Financial time series forecasting with machine learning techniques: A survey. In European Symposium on Artificial Neural Networks: Computational Intelligence and Machine Learning (*ESANN)*, 25-30.
15. Kajitani, Y., Hipel, K. W., & McLeod, A. I. (2005). Forecasting nonlinear time series with feed-forward neural networks: a case study of Canadian lynx data. *Journal of Forecasting*, ***24***(2), 105-117.
16. Tsaur, R. C. (2012). A fuzzy time series-Markov chain model with an application to forecast the exchange rate between the Taiwan and US dollar. *International journal of innovative computing, information and control*, ***8***(7), 4931-4942.
17. Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. *Energy*, ***34***(9), 1413-1421.
18. Smarandache, F. (2003, September). Definiton of neutrosophic logic-a generalization of the intuitionistic fuzzy logic. Multispace & multistructure. Neutrosophic transdisciplinarity (100 collected papers of science), **4**, 396.
19. Susruth, M. (2017). Financial Forecasting: An Empirical Study on Box–Jenkins Methodology with reference to the Indian Stock Market. *Pacific Business Review International*, **10**(2), 115-123.
20. Yu, T. H. K., & Huarng, K. H. (2008). A bivariate fuzzy time series model to forecast the TAIEX. *Expert Systems with Applications*, ***34***(4), 2945-2952.
21. Yu, H. K. (2005). Weighted fuzzy time series models for TAIEX forecasting. *Physica A: Statistical Mechanics and its Applications*, ***349***(3-4), 609-624.
22. Huarng, Kun-Huang, and Tiffany Hui-Kuang Yu. (2017). Using qualitative approach to forecasting regime switches. *Quality & Quantity*, **51**(5), 2035-2048.
23. Zhang, R., Ashuri, B., & Deng, Y. (2017). A novel method for forecasting time series based on fuzzy logic and visibility graph. *Advances in Data Analysis and Classification*, **11**(4), 759-783.
24. Eyoh, I., John, R., & De Maere, G. (2017, July). Time series forecasting with interval type-2 intuitionistic fuzzy logic systems. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1-6. IEEE.
25. Song, Q., & Chissom, B. S. (1993). Fuzzy time series and its models. *Fuzzy sets and systems*, **54**(3), 269-277.

26. Huarng, K., & Yu, T. H. K. (2006). The application of neural networks to forecast fuzzy time series. *Physica A: Statistical Mechanics and its Applications*, **363**(2), 481-491.
27. Bhuriya, D., Kaushal, G., Sharma, A., & Singh, U. (2017, April). Stock market predication using a linear regression. In *2017 international conference of electronics, communication and aerospace technology (ICECA)*, **2**, 510-513, IEEE.
28. Lv, P., Wu, Q., Xu, J., & Shu, Y. (2022). Stock Index Prediction Based on Time Series Decomposition and Hybrid Model. *Entropy*, **24**(2), 146.
29. Herawati, N., Saidi, S., Nisa, K., & Setiawan, E. (2022). Performance of High-Order Chen Fuzzy Time Series Forecasting Method and Feedforward Backpropagation Neural Network Method in Forecasting Composite Stock Price Index. *American Journal of Computational and Applied Mathematics*, **12**(1), 1-7.
30. Abbasimehr, H., & Paki, R. (2022). Improving time series forecasting using LSTM and attention models. *Journal of Ambient Intelligence and Humanized Computing*, ***13***(1), 673-691.
31. Yolca, O. C., & Alpaslan, F. (2018). Prediction of TAIEX based on hybrid fuzzy time series model with single optimization process. *Applied Soft Computing*, **66**, 18-33.
32. Pal, S. S., & Kar, S. (2022). Fuzzy transfer learning in time series forecasting for stock market prices. *Soft Computing*, **26**(14), 6941-6952.
33. Bisht, K., & Kumar, S. (2016). Fuzzy time series forecasting method based on hesitant fuzzy sets. *Expert Systems with Applications*, **64**, 557-568.
34. Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data*, **4**(1), 15.
35. Sousa, R., Lima, T., Abelha, A., & Machado, J. (2021). Hierarchical Temporal Memory Theory Approach to Stock Market Time Series Forecasting. *Electronics*, **10**(14), 1630.
36. Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep Learning for Time Series Forecasting: A Survey. *Big Data*, ***9***(1), 3-21.
37. Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, **379**(2194), p. 20200209.
38. Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. (2021). Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change*, ***163***, p. 120480.
39. Anufriev, M., Hommes, C., & Makarewicz, T. (2019). Simple forecasting heuristics that make us smart: Evidence from different market experiments. *Journal of the European Economic Association*, **17**(5), 1538-1584.

40. Miciuła, I. (2014). The concept of FTS analysis in forecasting trends of exchange rate changes. *Economics & Sociology*, **7**(2), 172-182.
41. Tozan, H., & Vayvay, O. (2008). Fuzzy forecasting applications on supply chains. *WSEAS Transactions on Systems*, **7**(5), 600-609.
42. Javedani Sadaei, H., & Lee, M. H. (2014). Multilayer stock forecasting model using fuzzy time series. *The Scientific World Journal, **2014**, Article ID 610594, 10 pages*. https://doi.org/10.1155/2014/610594
43. Dai, Y., Han, D., & Dai, W. (2014). Modeling and computing of stock index forecasting based on neural network and Markov chain. *The Scientific World Journal*, ***2014***, 124523, 9 pages. https://doi.org/10.1155/2014/124523
44. Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on systems, Man, and Cybernetics*, **1**, 28-44.
45. Gonchar, N. S. (2000). Mathematical model of a stock market. *Condensed Matter Physics*. **3**(23), 461-496.
46. Couts, D., Grether, D., & Nerlove, M. (1966). Forecasting non-stationary economic time series. *Management science*, ***13***(1), 1-21.
47. Young, P. C. (1999). Nonstationary time series analysis and forecasting. *Progress in Environmental Science*, **1**, 3-48.
48. Coulibaly, P., & Baldwin, C. K. (2005). Nonstationary hydrological time series forecasting using nonlinear dynamic methods. Journal of Hydrology, **307**(1-4), pp. 164-174.
49. Liu, C., Hoi, S. C., Zhao, P., & Sun, J. (2016, February). Online arima algorithms for time series prediction. In *Thirtieth AAAI conference on artificial intelligence*, **30**(1).
50. Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014, March). Stock price prediction using the ARIMA model. In *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*, pp. 106-112. IEEE.
51. Abhishekh, Gautam, S. S., & Singh, S. R. (2020). A new method of time series forecasting using intuitionistic fuzzy set based on average-length. *Journal of industrial and production engineering*, **37**(4), pp. 175-185.
52. Wilinski, A. (2019). Time series modeling and forecasting based on a Markov chain with changing transition matrices. *Expert Systems with Applications*, **133**, 163-172.
53. Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *omega*, ***29***(4), 309-317.

54. Sah, M., & Konstantin Y, D. (2005). Forecasting enrollment model based on first-order fuzzy time series. *In World Academy of Science, Engineering and Technology*, ***1***, 375-378.
55. Chen, T. L., Cheng, C. H., & Teoh, H. J. (2007). Fuzzy time-series based on Fibonacci sequence for stock price forecasting. *Physica A: Statistical Mechanics and its Applications*, ***380***, 377-390.
56. Wang, J. Z., Wang, J. J., Zhang, Z. G., & Guo, S. P. (2011). Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, ***38***(11), 14346-14355.
57. Wong, H. L., Tu, Y. H., & Wang, C. C. (2010). Application of fuzzy time series models for forecasting the amount of Taiwan export. *Expert Systems with Applications*, ***37***(2), 1465-1470.
58. Chu, H. H., Chen, T. L., Cheng, C. H., & Huang, C. C. (2009). Fuzzy dual-factor time-series for stock index forecasting. Expert systems with applications, **36**(1), 165-171.
59. Cheng, C. H., Teoh, H. J., & Chen, T. L. (2007, August). High-order fuzzy time series based on rough set for forecasting TAIEX. In *2007 International Conference on Machine Learning and Cybernetics*, **3**, 1354-1358, IEEE.
60. Efendi, R., Ismail, Z., & Deris, M. M. (2015). A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand. Applied Soft Computing, **28**, pp. 422-430.
61. Weron, R. (2008). Forecasting wholesale electricity prices: A review of time series models. Financial Market: Principles of Modelling, Forecasting and Decision-Making, W. Milo, P. Wdowinski, eds., FindEcon Monograph Series, WUL, Lodz, 71-82
62. Gordillo-Orquera, R., Lopez-Ramos, L. M., Muñoz-Romero, S., Iglesias-Casarrubios, P., Arcos-Avilés, D., Marques, A. G., & Rojo-Álvarez, J. L. (2018). Analyzing and forecasting electrical load consumption in healthcare buildings. *Energies*, ***11***(3), 493.
63. Delima, A. J. P. (2019). Application of Time Series Analysis in Projecting Philippines' Electric Consumption. *International Journal of Machine Learning and Computing*, ***9***(5), 694-699.
64. Mazengia, D. H. (2008, November). Forecasting spot electricity market prices using time series models. In *2008 IEEE International Conference on Sustainable Energy Technologies*, pp. 1256-1261, IEEE.
65. Weron, R., & Misiorek, A. (2005). Forecasting spot electricity prices with time series models. *The European Electricity Market EEM*-05, proceeding vol. 133, pp. 133-141.

66. Cai, Q., Zhang, D., Zheng, W., & Leung, S. C. (2015). A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. *Knowledge-Based Systems*, *74*, 61-68.
67. Junior, P. R., Salomon, F. L. R., & de Oliveira Pamplona, E. (2014). ARIMA: An applied time series forecasting model for the Bovespa stock index. *Applied Mathematics*, *5*(21), 3383.
68. Zhiqiang, G., Huaiqing, W., & Quan, L. (2013). Financial time series forecasting using LPP and SVM optimized by PSO. *Soft Computing*, ***17***, 805-818.
69. Tanuwijaya, K., & Chen, S. M. (2009, July). A new method to forecast enrollments using fuzzy time series and clustering techniques. In *2009 International Conference on Machine Learning and Cybernetics*, **5**, 3026-3029. IEEE.
70. Vovan, T. (2019). An improved fuzzy time series forecasting model using variations of data. *Fuzzy Optimization and Decision Making*, ***18***, 151-173.
71. Atsalakis, G. S., & Valavanis, K. P. (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert systems with Applications*, ***36***(7), 10696-10707.
72. Bisht, K., & Kumar, S. (2018). Intuitionistic fuzzy set-based computational method for financial time series forecasting. *Fuzzy Information and Engineering*, ***10***(3), 307-323.
73. Tsai, M. C., Cheng, C. H., & Tsai, M. I. (2019). A multifactor fuzzy time-series fitting model for forecasting the stock index. *Symmetry*, ***11***(12), 1474.
74. Hassan, S., Jaafar, J., Samir, B. B., & Jilani, T. A. (2012). A Hybrid Fuzzy Time Series Model for Forecasting. *Engineering Letters*, ***20***(1).
75. Bollapragada, R., Savin, I., & Kerbache, L. (2013). Price forecasting and analysis of exchange traded fund. *Journal of Mathematical Finance*, **3**, 181-191.
76. Mehtab, S., & Sen, J. (2020). Stock price prediction using convolutional neural networks on a multivariate timeseries. *arXiv preprint arXiv:2001.09769*.
77. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
78. Mathur, P. (2018). *Machine learning applications using python: Cases studies from healthcare, retail, and finance*. Apress.
79. Zimmermann, H. J. (2011). *Fuzzy set theory—and its applications*. Springer Science & Business Media.
80. Anderson, H. (1960). Problems peculiar to export sales forecasting. *Journal of Marketing*, ***24***(4), 39-42.

81. Janerich, D. T. (1984). Forecasting cancer trends to optimize control strategies. *Journal of the National Cancer Institute*, ***72***(6), 1317-1321.
82. Grabowski, H., & Vernon, J. (1990). A new look at the returns and risks to pharmaceutical R&D. *Management Science*, ***36***(7), 804-821.
83. Weiss, T. W., Ashton, C. M., & Wray, N. P. (1993). Forecasting areawide hospital utilization: a comparison of five univariate time series techniques. *Health services management research*, ***6***(3), 178-190.
84. Kosirog, J. L., Rospond, R. M., Destache, C., & Hall, P. (1993). Aminoglycoside forecasting in neutropenic patients with cancer. *Clinical pharmacokinetics*, ***24***(1), 79-87.
85. Goldman, R. L. (1994). Using Wild Cards for Facilities, Program and Trend Forecasting. *Health marketing quarterly*, ***12***(1), 9-22.
86. Abidi, S. S., & Goh, A. (1998). Neural network based forecasting of bacteria-antibiotic interactions for infectious disease control. *STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS*, **1**, 571-571.
87. Grover, S. A., Coupal, L., Paquet, S., & Zowall, H. (1999). Cost-effectiveness of 3-hydroxy-3-methylglutaryl–coenzyme A reductase inhibitors in the secondary prevention of cardiovascular disease: forecasting the incremental benefits of preventing coronary and cerebrovascular events. *Archives of Internal Medicine*, ***159***(6), 593-600.
88. Hulme, D., & Xu, S. (2001, December). Application of genetic algorithms to the optimisation of neural network configuration for stock market forecasting. In *Australian Joint Conference on Artificial Intelligence* (pp. 285-296). Springer, Berlin, Heidelberg.
89. Xue, J. L., Ma, J. Z., Louis, T. A., & Collins, A. J. (2001). Forecast of the number of patients with end-stage renal disease in the United States to the year 2010. *Journal of the American Society of Nephrology*, ***12***(12), 2753-2758.
90. Bae, J. M., Jung, K. W., & Won, Y. J. (2002). Estimation of cancer deaths in Korea for the upcoming years. *Journal of Korean medical science*, ***17***(5), 611-615.
91. Hao, J., Var, T., & Chon, J. (2003). A forecasting model of tourist arrivals from major markets to Thailand. *Tourism Analysis*, ***8***(1), 33-45.
92. Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with applications*, ***29***(4), 927-940.

93. Lakshminarayanan, S., Weckman, G. R., Snow, A., & Marvel, J. (2006). Stock market hybrid forecasting model using neural networks. In *IIE Annual Conference. Proceedings* (p. 1). Institute of Industrial and Systems Engineers (IISE).
94. Rodriguez, J. (2007). A critical look at the forecasting ability of real estate mutual fund managers. *Journal of Real Estate Portfolio Management*, ***13***(2), 99-106.
95. Masursky, D., Dexter, F., O'Leary, C. E., Applegeet, C., & Nussmeier, N. A. (2008). Long-term forecasting of anesthesia workload in operating rooms from changes in a hospital's local population can be inaccurate. *Anesthesia & Analgesia*, ***106***(4), 1223-1231.
96. Sun, Y., Heng, B. H., Seow, Y. T., & Seow, E. (2009). Forecasting daily attendances at an emergency department to aid resource planning. *BMC emergency medicine*, ***9***(1), 1-9.
97. Kam, H. J., Sung, J. O., & Park, R. W. (2010). Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthcare informatics research*, ***16***(3), 158-165.
98. Froelich, W., Papageorgiou, E. I., Samarinas, M., & Skriapas, K. (2012). Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer. *Applied Soft Computing*, ***12***(12), 3810-3817.
99. Garg, B., Beg, M. S., & Ansari, A. Q. (2012, August). A new computational fuzzy time series model to forecast number of outpatient visits. In *2012 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)* (pp. 1-6). IEEE.
100. Claudio, D., Miller, A., & Huggins, A. (2014). Time series forecasting in an outpatient cancer clinic using common-day clustering. *IIE Transactions on Healthcare Systems Engineering*, ***4***(1), 16-26.
101. Bro, R., Kamstrup-Nielsen, M. H., Engelsen, S. B., Savorani, F., Rasmussen, M. A., Hansen, L., & Dragsted, L. O. (2015). Forecasting individual breast cancer risk using plasma metabolomics and biocontours. *Metabolomics*, ***11***(5), 1376-1380.
102. Golmohammadi, K., & Zaiane, O. R. (2015, October). Time series contextual anomaly detection for detecting market manipulation in stock market. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1-10. IEEE.
103. Dang, H. S., Huang, Y. F., Wang, C. N., & Nguyen, T. M. T. (2016). An application of the short-term forecasting with limited data in the healthcare traveling industry. *Sustainability*, ***8***(10), 1037.

104. Ganguly, A., & Nandi, S. (2016). Using statistical forecasting to optimize staff scheduling in healthcare organizations. *Journal of Health Management*, ***18***(1), 172-181.

105. Aripin, A., Suryono, S., & Bayu, S. (2016, June). Web based prediction of pollutant PM10 concentration using Ruey Chyn Tsaur fuzzy time series model. In *AIP Conference Proceedings*, **1746**(1), p. 020046, AIP Publishing LLC.

106. Calegari, R., Fogliatto, F. S., Lucini, F. R., Neyeloff, J., Kuchenbecker, R. S., & Schaan, B. D. (2016). Forecasting daily volume and acuity of patients in the emergency department. *Computational and mathematical methods in medicine*, ***2016***, p. 8. https://doi.org/10.1155/2016/3863268

107. Iqelan, B. M. (2017). Forecasts of female breast cancer referrals using grey prediction model GM (1, 1). *Applied Mathematical Sciences, 11*(54).

108. Kaushik, S., Choudhury, A., Sheron, P. K., Dasgupta, N., Natarajan, S., Pickett, L. A., & Dutt, V. (2020). AI in healthcare: time-series forecasting using statistical, neural, and ensemble architectures. *Frontiers in big data*, **3**, p. 4.

109. Harrou, F., Dairi, A., Kadri, F., & Sun, Y. (2020). Forecasting emergency department overcrowding: A deep learning framework. *Chaos, Solitons & Fractals*, **139**, 110247. https://doi.org/10.1016/j.chaos.2020.110247

110. Firmino, P. R. A., de Sales, J. P., Júnior, J. G., & da Silva, T. A. (2020). A non-central beta model to forecast and evaluate pandemics time series. *Chaos, Solitons & Fractals*, ***140***, 110211. https://doi.org/10.1016/j.chaos.2020.110211

111. Budiharto, W. (2021). Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM). *Journal of big data*, ***8***(1), 1-9.

112. Zhang, H., Luo, Y., Song, J., Zhang, H., Xia, J., Zhao, Q. (2011). Predicting As, Cd and Pb uptake by rice and vegetables using field data from China. Journal of Environmental Sciences, **23**(1), 70-78.

113. Hare, D. J. (1980). Impact of defoliation by the Colorado potato beetle on potato yields. Journal of Economic Entomology, **73**(3), 369-373.

114. El-Nazer, T., & McCarl, B. A. (1986). The choice of crop rotation: a modeling approach and case study. American Journal of Agricultural Economics, **68**(1), 127-136.

115. Johnson, K. B., (1992). Evaluation of a mechanistic model that describes potato crop losses caused by multiple pests. Crops, **9**, 20.

116. Gandhi, N., Petkar, O., & Armstrong, L. J. (2016). Rice crop yield prediction using artificial neural networks. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 105-110. IEEE.

117. Jadhav, V., Chinnappa, R. B., and Gaddi, G. M. (2017). Application of ARIMA model for forecasting agricultural prices, J. Agr. Sci. Tech., **19**, 981-992.

118. Biswas, R. and Bhattacharyya, B. (2013). ARIMA modeling to forecast area and production of rice in West Bengal. Journal of crop and weed, **9**(2), 26-31.

119. Elsamie, M. A., Ali, T. and Zhou, D. (2021). Using a dynamic time series model (Arima) for forecasting of Egyptian cotton crop variables. Journal of Animal and Plant Sciences, **31**(3), 810-23.

120. Ho, S. L., Xie, M. and Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. Computers & Industrial Engineering, **42**(2-4), 371-375.

121. Devi, M., Kumar, J., Malik, D. P., and Mishra, P. (2021). Forecasting of wheat production in Haryana using hybrid time series model. Journal of Agriculture and Food Research, **5**, 100175. https://doi.org/10.1016/j.jafr.2021.100175

122. Purohit, S. K., Panigrahi, S., Sethy, P. K., and Behera, S. K. (2021). Time series forecasting of price of agricultural products using hybrid methods. Applied Artificial Intelligence, **35**(15), 1388-1406.

123. Awal, M. A., and Siddique, M. A. B. (2011). Rice production in Bangladesh employing by ARIMA model. Bangladesh Journal of Agricultural Research, **36**(1), 51-62.

124. Singh, N. O., Singh, N. G., Singh, L. N., Kumar, S. and Paul, A. K. (2020). Forecasting model of potato yield from farmers' fields in Manipur, Int. J. Agricult. Stat. Sci. **16**(1), 401-405.

125. Wang, J., Li, H., & Lu, H. (2018). Application of a novel early warning system based on fuzzy time series in urban air quality forecasting in China. Applied Soft Computing, **71**, 783-799.

126. Rahman, M., Shakeri, M., Khatun, F., Tiong, S. K., Alkahtani, A. A., Samsudin, N. A., & Hasan, M. K. (2022). A comprehensive study and performance analysis of deep neural network-based approaches in wind time-series forecasting. *Journal of Reliable Intelligent Environments*, **9**(2), 183-200.

127. Gupta, A., Kumar, A., & Boopathi, K. (2021). Day-ahead and intra-day wind power forecasting based on feedback error correction. *IET Renewable Power Generation*, **15**(13), 2840-2848.

128. Alyousifi, Y., Othman, M., Faye, I., Sokkalingam, R., & Silva, P. C. (2020). Markov weighted fuzzy time-series model based on an optimum partition method for forecasting air pollution. *International Journal of Fuzzy Systems*, ***22***, 1468-1486.

129. Huarng, K. H., Yu, T. H. K., & Solé Parellada, F. (2011). An innovative regime switching model to forecast Taiwan tourism demand. *The Service Industries Journal*, **31**(10), 1603-1612.

130. Xiao, Y., Liu, J. J., Hu, Y., Wang, Y., Lai, K. K., & Wang, S. (2014). A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting. *Journal of Air Transport Management*, *39*, 1-11.

131. Sagaert, Y. R., Aghezzaf, E. H., Kourentzes, N., & Desmet, B. (2018). Temporal big data for tactical sales forecasting in the tire industry. *Interfaces*, ***48***(2), 121-129.

132. Haque, M. E., Hossain, M. I., & Rahman, K. M. M. (2005). Forecasting fish production in Bangladesh using AR1MA model. *Journal of the Bangladesh Agricultural University*, *3*(452-2018-3870), 381-392.

133. Tan, G., Yan, J., Gao, C., & Yang, S. (2012). Prediction of water quality time series data based on least squares support vector machine. *Procedia Engineering*, ***31***, 1194-1199.

134. Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental modelling & software*, ***15***(1), 101-124.

135. Ni, J., Ma, H., & Ren, L. (2012, May). A time-series forecasting approach based on KPCA-LSSVM for lake water pollution. In *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, 1044-1048, IEEE.

136. Paraschiv, D., Tudor, C., & Petrariu, R. (2015). The textile industry and sustainable development: a Holt–Winters forecasting investigation for the Eastern European area. *Sustainability*, ***7***(2), 1280-1291.

137. Deng, W., Wang, G., & Zhang, X. (2015). A novel hybrid water quality time series prediction method based on cloud model and fuzzy forecasting. *Chemometrics and Intelligent Laboratory Systems*, ***149***, 39-49.

138. Kogekar, A. P., Nayak, R., & Pati, U. C. (2021, July). Forecasting of water quality for the river ganga using univariate time-series models. In *2021 8th International Conference on Smart Computing and Communications (ICSCC)*, 52-57, IEEE.

139. Dutta, B., Barman, M. P. and Patowary, A. N. (2020). Application of ARIMA Model for forecasting road accident deaths in India. Int. J. Agricult. Stat. Sci., **16**(2), 607-615.

140. Data source: https://finance.yahoo.com/quote/%5ETWII/history?period1=1514764800&period2=1577750400&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true.

141. Data source: https://finance.yahoo.com/quote/%5EIXIC/history/.

142. Data source: https://finance.yahoo.com/quote/%5EDJI/history/.

143. Data source: https://www.bseindia.com/Indices/IndexArchiveData.html.

144. Data source: https://www.indexmundi.com/commodities/?commodity=rice&months=240&currency=inr.

145. Data source: https://www.indexmundi.com/commodities/?commodity=rice&months=360.

146. Data source: https://finance.yahoo.com/quote/%5EIXIC/history?p=%5EIXIC.

147. Data source: https://finance.yahoo.com/quote/%5ETWII/history?p=%5ETWII.

148. Data source: https://finance.yahoo.com/quote/399001.SZ/history?p=399001.SZ

149. Data source: https://finance.yahoo.com/quote/000001.SS/history?p=000001.SS

150. Data source: https://www.indexmundi.com/agriculture/?country=in&commodity=milled-rice&graph=production.

List of Publication

Paper Published

1. Garov, A. K. (2022, May). Quantity Based weights forecasting for TAIEX. In *Journal of Physics: Conference Series* (Vol. 2267, No. 1, p. 012151). IOP Publishing.
2. Koul, S., Awasthi, A., & Garov, A. K. (2020). Decision Making Model for Stock Index. *Journal of Xidian University*, 14(3), 1261-1265.
3. Garov, A. K., & Awasthi, A.K. (2021). A computational mathematical model for forecasting of Indian crop, The Pharma Innovation Journal, 10 (7S), 05-08.
4. Garov, A. K., & Awasthi, A.K. (2023). Time Series ARIMA Based Forecasting of Rice Price, INTL. JL. OF AGRICULTURAL & STATISTICAL SCIENCES, 19(1), pp. 285-303.
5. Awasthi, A. K., Garov, A. K., Sharma, M. and Sinha, M. (2023). GNN Model Based On Node Classification Forecasting in Social Network, 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), Greater Noida, India, pp. 1039-1043, doi: 10.1109/AISC56616.2023.10085118.
6. Awasthi, A. K., Sharma, M., Garov, A. K., & Chaudhary, P. (2023). Presentation of futuristic Malarial Disease through a Hybrid Model of A.I. and Big data, *2023 International Conference on Artificial Intelligence and Smart Communication (AISC), Greater Noida, India*, pp. 1044-1050, doi: 10.1109/AISC56616.2023.10085407.
7. Garov, A. K. & Awasthi, A.K. Quantity Based Time Series Fuzzified Approach for Forecasting Stock Index, AIP Conference Proceeding, 23 June 2023; 2768 (1): 020006.
8. A.K. Awasthi, Minakshi Sharma and Arun Kumar Garov, Forecasting Analysis of COVID-19 Patient Recovery Using RF–DT Model, AIP Conference Proceedings 23 June 2023; 2768 (1): 020009.

Book Chapter Published

1. Awasthi, A. K., & Garov, A. K. (2020). Agricultural modernization with forecasting stages and machine learning. In *Smart Agriculture: Emerging Pedagogies of Deep Learning, Machine Learning and Internet of Things* (pp. 61-80). Chapman and Hall/CRC.

2. Awasthi, A. K., Kumar, S., & Garov, A. K. (2022). IoT in the Healthcare Sector. In *Machine Learning, Deep Learning, Big Data, and Internet of Things for Healthcare* (pp. 107-123). Chapman and Hall/CRC.
3. Garov, A. K., & Awasthi, A. K. (2022). Case Study-Based Big Data and IoT in Healthcare. In *Machine Learning, Deep Learning, Big Data, and Internet of Things for Healthcare* (pp. 13-35). Chapman and Hall/CRC.
4. Garov, A. K., Awasthi, A. K., Kumar, R., & Sankat, M. (2023). Application and Some Fundamental Study of GNN In Forecasting. In V. Kumar & D. Rajput (Eds.), Concepts and Techniques of Graph Neural Networks (pp. 132-154). IGI Global. https://doi.org/10.4018/978-1-6684-6903-3.ch009.

List of Conferences

1. Presented a paper title **"Presentation of Futuristic Malarial Disease through a Hybrid Model of A.I. and Big Data"** in the AICTE-sponsored International Conference on Computational Methods in Science & Technology, 19th-20th January 2023 at Chandigarh Engineering College, CGC Landran.

2. Presented a paper title **"GNN Model Based on Node Classification Forecasting in Social Network"** in the AICTE-Sponsored International Conference on Computational Methods in Science & Technology, 19th-20th January 2023 at Chandigarh Engineering College, CGC Landran.

3. Presented a paper title **"Quantity based Time Series Fuzzified Approach for Forecasting Stock Index"** in the International Conference on Computational Applied Sciences and its Applications" (ICCASA-2022), 28-29 April, 2022, University of Engineering and Management, Jaipur, Rajasthan.

4. Presented a paper title **"Quantity based Weighted Forecasting for TAIEX"** in the International Conference, RAFAS 2021, Lovely Professional University, Punjab

5. Presented a paper title **"Quantity Based Mathematical Approach for Forecasting Stock Index"** in the International Conference, RAFAS 2019, Lovely Professional University, Punjab

List of Workshops

1. Participated in 15 Days Training-CUM-Certificate Programme on “training novel technologies in Agriculture, Animal Husbandry, Fisheries Science and Their Allied Systems” on January 15 to 30, 2023 in online mode by Society of Agriculture Research and Social Development, College of Agriculture, Tripura, & Malla Reddy University, Hyderabad.
2. Participated in 5-day virtual Faculty Development Program (vFDP) on “Machine Leaning Using Python: Hands-on Approach” from 20-06-2022 to 25-06-2022 organized by J. C. Bose University of Science & Technology, YMCA, Faridabad, Haryana, India.

3. Participated in 5-day virtual Faculty Development Program (vFDP) on “Machine Leaning & Deep Learning Approach towards Computer Vision” from 28-03-2022 to 02-04-2022 organized by Dr. B. C. Roy Engineering College, Durgapur.

4. Participated in 5-day virtual Faculty Development Program (vFDP) on “Probability Theory and Reliability Theory (IWPRT-2021)” on 20th November 2021 Organized by the Department of Applied Mathematics, Shri Shankaracharya Technical Campus, Bhilai, Chhattisgarh.

5. Participated in 5-day AICTE Training and Learning (ATAL) Academy, FDP on “Machine learning and Optimization Technique: Applications to Financial Markets” from 12-07-2021 to 16-07-2021 at Indian Institute of Technology Mandi.

6. Participated in 5-day AICTE Training and Learning (ATAL) Academy, FDP on “Data Science & Machine Leaning with Python” from 05-07-2021 to 09-07-2021 at Manipal University Jaipur.

7. Participated in AICTE Training and Learning (ATAL) Academy, FDP on “Artificial Intelligence in Contemporary Biomedical & Healthcare Application: Fundamentals & Hands on MATLAB” from 21-06-2021 to 25-06-2021 at Gautam Buddha University, Greater Noida, U. P.

8. Participated in AICTE Training and Learning (ATAL) Academy, FDP on “Role of Augmented Reality (AR) & Virtual Reality (VR) in Real-Time Applications” from 07-06-2021 to 11-06-2021 at Karunya Institute of Technology and Science.

9. Participated in 5-day AICTE Training and Learning (ATAL) Academy, FDP on “Artificial Intelligence” from 22-02-2021 to 26-02-2021 at Model Engineering College.

10. Participated in 5-day AICTE Training and Learning (ATAL) Academy, FDP on “Data Sciences” from 07-01-2021 to 11-01-2021 at Cochin University of Science and Technology.

11. Participated in AICTE Training and Learning (ATAL) Academy, FDP on "Mathematics for Machine learning" from 18-01-2021 to 22-01-2021 at Sri Jayachamarajendra College of Engineering, JSS Science and Technology University.

Nomenclature

QBFTS- Quantity Based Fuzzy Time Series

QBMF- Quantity Based Main Factor

QBSF- Quantity Based Secondary Factor

QBFLRs- Quantity Based Fuzzy Logical Relationships

QBFLRGs- Quantity Based Fuzzy logical relationship groups

TAIEX- Taiwan Stock Exchange Capitalization Weights Stock Index

BSE - Bombay Stock Market

ACF- Auto Correlation Function

PACF- Partial Auto Correlation Function

ARIMA- Autoregressive Integrated Moving Average

ML- Machine Learning

RMSE- Root mean square error

MSE- Mean square error

MAE - mean absolute error

Multiple R- Coefficient of correlation

R^2- Coefficient of determination

SS- Sum of Squares

MS- Mean squared error

F- Overall test for the null hypothesis

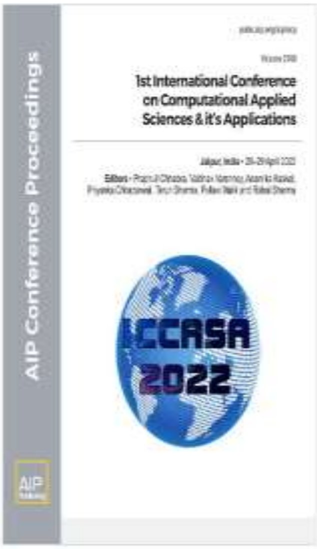
T-State - Computed t statistic for intercept and slope

Lower 95% and Upper 95% - Confidence interval estimate of intercept and slope

Published Matter

Volume 2768, Issue 1

23 June 2023



1ST INTERNATIONAL CONFERENCE ON COMPUTATIONAL APPLIED SCIENCES & IT'S APPLICATIONS

28–29 April 2020

Jaipur, India

‹ Previous Article Next Article ›

RESEARCH ARTICLE | JUNE 23 2023

Quantity based time series fuzzified approach for forecasting stock index

Arun Kumar Garov ; A. K. Awasthi

+ Author & Article Information

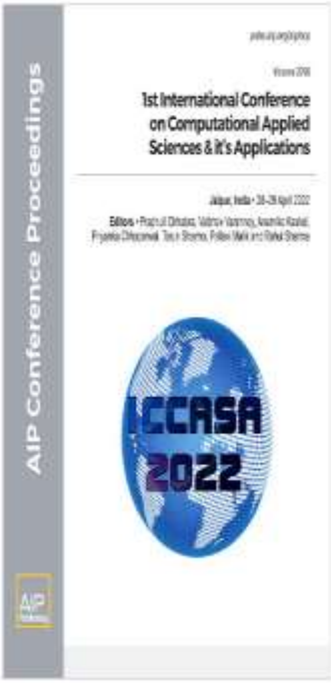
AIP Conference Proceedings 2768, 020006 (2023)

https://doi.org/10.1063/5.0148310

In the present paper, we have considered Taiwan Stock Exchange Capitalization Weights Stock Index (TAIEX) index, the NASDAQ and Dow Jones indexes for study and finding predicted trends for the year 2018 using Quantity based fuzzy time series (QBFTS). In each quantity-based index which has many fields, on that basis we have taken two types of quantity-based factors, first is the main quantity-based factor and second is the secondary quantity-based factor are applied for forecasting of stock index. For forecasting, the methods are QBFTS and QBFVGs. In the succession of that, a fuzzifying variation of the main quantity-based factor, that make some relation of the fuzzy sets this type of relation is called QBFLR. Also, arrange the QBFLRGs with the help of linguistic term, which is represented by fuzzy sets. In this fuzzified approach, these terms and groups are used for calculation of different variation of quantity-based factors and weights of the quantity-based factors. Finally, based on the weights of quantity based fuzzy variation appearing in the QBFVGs and QBFLRGs, opted for forecasting of Stock Index with help of software.

Volume 2768, Issue 1

23 June 2023



1ST INTERNATIONAL CONFERENCE ON COMPUTATIONAL APPLIED SCIENCES & IT'S APPLICATIONS

28–29 April 2020

Jaipur, India

‹ Previous Article Next Article ›

RESEARCH ARTICLE | JUNE 23 2023

Forecasting analysis of COVID-19 patient recovery using RF-DT model

A. K. Awasthi; Minakshi Sharma; Arun Kumar Garov

Check for updates

+ Author & Article Information

AIP Conference Proceedings 2768, 020009 (2023)

https://doi.org/10.1063/5.0148356

Share ∨ Tools ∨

This paper is to analyse and to foretell the effect of COVID-19. This virus constantly changes due to mutation and hence there are many variants like Delta, Omicron, Alpha, Beta, and Gamma. In this paper, we are using the RF-DT model by machine learning technique for the forecasting analysis of COVID-19 patient recovery data. The proposed result of forecasting analysis of COVID-19 patient recovery data compares with other data mining techniques like Naïve Bayes, linear regression, SVM. The proposed RF-DT model shows 99.5% accuracy. Moreover, a python programming language is used for the analysis and graphical presentation along with statistical summarization.

Quantity Based weights forecasting for TAIEX

A K Garov

Department of Mathematics, Lovely Professional University, Punjab, 144411, India

email: arunkumar170@yahoo.com

Abstract. Many models of the forecasting have been proposed on stock index. In this paper, the consideration of Taiwan Stock Exchange Capitalization Weights Stock Index (TAIEX) has been taken and founded forecasted index value. Here forecasting is done by using Quantity based fuzzy time series (QBFTS), Quantity based factor TAIEX and forecasting methods are QBFLR and QBFLRGs. The approach of Statistical weights assigned for different weights to various quantity based fuzzy logical relationships for considered quantity-based dataset of TAIEX. Hence got the comparative result better than the other models.

1. Introduction

Stock market index forecasting approached by researchers with different adopted models. Sadaei et al. [1] developed a method of multi-layer model for stock market index forecasting with five logical significant layers use by fuzzy time series. Smarandache [2] generalized the intuitionistic fuzzy logic (IFL) and other neutrosophic logic (NL). Chen at el. [3] applied fuzzy time series and generated weights of multiple factors for the forecasting of the stock market index. Malik and Bhatt [4] forecasted the exchange rate of the financial market by various models i.e. artificial neural network, particle swarm optimization, genetic programming, and fuzzy logic. Dai et al. [5] described a forecasting method by the combination of back-propagation (BP) neural network and Markov chain with the modelling and computational techniques. Tozan et al. [6] evaluated the performance effects of fuzzy linear regression, fuzzy time series and fuzzy GM (1, 1) forecasted model on the supply chain. Qiu et al. [7] discussed the concept of generalized fuzzy logical relationship with clustering techniques, and computational methods for forecasting of stock index. Yabuuchi et al. [8] described the comparison between the fuzzy autoregressive (AR) model and fuzzy autocorrelation model on the stock price. Thenmozhi [9] applied neural network model to predict the daily returns of the Bombay stock index (BSE).

The Autoregressive integrated moving average (ARIMA) model was applied for the prediction on a short-term basis by Ariyo et al. [10]. Lin et al. [11] used a back propagation network method with statistics to prediction analysis of Taiwan stock index. Study of Moving average (MA) with buy and holding strategy on the stock index by Wong et al. [12]. Zhang et al. [13] proposed a novel fuzzy time series forecasting model with multiple linear regression and time series clustering for forecasting market prices. Zadeh [14] discussed the conventional quantitative techniques for system analysis with linguistic variable. Cheng & Li [15] developed a novel fuzzy smoothing method by Hidden Markov Model (HMM) for forecasting.

Song and Chissom [16] disused the fuzzy time series and its model to forecast the enrollments of a university. Kumar and Murugan [17] disused the basic ideas of times series and ANN with forecasting by neural network time series model. Saxena et al. [18] proposed a method based on fuzzy time series

Int. J. Agricult. Stat. Sci. Vol. 19, No. 1, pp. 285-303, 2023
DOI: https://doi.org/10.59467/IJASS.2023.19.285
www.connectjournals.com/ijass
ISSN : 0973-1903, e-ISSN : 0976-3392
ORIGINAL ARTICLE

TIME SERIES ARIMA BASED FORECASTING OF RICE PRICE

Arun Kumar Garov* and A.K. Awasthi
Department of Mathematics, Lovely Professional University, Jallandhar - 144 402, India.
E-mail: arunkumar170@yahoo.com

Abstract: Rice crop is one of the most produced crops in India due to this the price of rice is also an important factor for buyers and suppliers (Farmers). In this article, the discussion, and methodology are used for time series forecasting of the Indian rice price with Autoregressive Integrated Moving Average Models. The model is based on historical data of Time series data for the Indian rice price in terms of Indian rupee per metric ton and the model applied for the computational forecasting analysis from June 2022 to December 2038 of the Indian rice price is ARIMA model with different possible orders which are suitable for the quantity-based data set. Evaluation of the model is done by MAPE and the process is carried out by SPSS.

Key words: Agriculture, Rice, Quantity based time series, Forecasting, ARIMA model.

Cite this article

Arun Kumar Garov and A.K. Awasthi (2023). Time Series ARIMA based Forecasting of Rice Price. *International Journal of Agricultural and Statistical Sciences.* DOI: https://doi.org/10.59467/IJASS.2023.19.285

1. Introduction

The preeminent crop of India is rice, which can see in southern and eastern areas and staple food of these areas. One effective factor for people is the Rice price. The natural effect is harmful to crops as heavy rainfall causes farmers to lose the crop but the demand for rice remains the same, which means demand is more and the production rate is low due to which the price increases. For the forecasting of the price of crops, various researchers applied various techniques and software for analysis. Jadhav *et al*. (2017) forecasted the price of Paddy, Ragi and Maize in the Indian state of Karnataka for the year 2016 on time series data from 2002 to 2016 by the ARIMA model for the forecasting up to the year 2020. Garov and Awasthi (2021) applied a regression model for forecasting rice, wheat and potato production. Biswas and Bhattacharyya (2013) applied ARIMA (p, d, q) model for the forecasting of the area and production of rice in West Bengal. Awasthi and Garov (2020) discussed the forecasting stages of agricultural modernization with the use of mathematical, statistical, and machine learning techniques and discussed in detail by easy examples to understand and case study to define agriculture by machine learning techniques. Elsamie *et al*. (2021) used the ARIMA models on the time series of the data set for the forecasting of productivity, cultivated area, and production of cotton crops. Koul *et al*. (2019) applied a completely randomized design (CRD) model to different methods to forecast stock index data sets. Ho *et al*. (2002) used neural networks and Box-Jenkins ARIMA on time series forecasting for failures of repairable systems. Box-Jenkins ARIMA and ANN were used to forecast the behavior of data for production and used a hybrid approach on the time series data of wheat production in Haryana by Devi *et al*. (2021). Garov and Awasthi (2023) discussed the time series data, plotted the ACF and PACF function, moved the average on different orders for the time series data set and discussed forecasting for the time series data with the fuzzy set and statistical weight methods. Purohit *et al*. (2021) used hybrid methods for the forecasting of the price of agricultural products such as tomatoes, Onions and Potatoes. Awal and Siddique(2011) used the ARIMA model to forecast rice production in Bangladesh. Dutta *et al*. (2020) applied

**Author for correspondence   Received May 14, 2022   Revised December 20, 2022   Accepted January 18, 2023*

2023 International Conference on Artificial Intelligence and Smart Communication (AISC)

GNN Model Based On Node Classification Forecasting in Social Network

2023 International Conference on Artificial Intelligence and Smart Communication (AISC) | 979-8-3503-2230-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/AISC56616.2023.10085118

1st A.K. Awasthi
dept. of Mathematics,
Lovely Professional University,
Phagwara, Punjab, India
dramitawasthi@gmail.com

2nd Arun Kumar Garov
dept. of Mathematics,
Lovely Professional University,
Phagwara, Punjab, India
arunkumar170@yahoo.com

3rd Minakshi Sharma
dept. of Mathematics,
Lovely Professional University,
Phagwara, Punjab, India
minakshisohareya@gmail.com

4th Mrigank Sinha
dept. of Mathematics,
Lovely Professional University,
Phagwara, Punjab, India
mrigank.sinhal1311@gmail.com

***Abstract*—In the time of ever-growing technology, engineering, and deep learning methods, one thing that has caught the attention of people is the invention of Neural Networks, also known as Artificial Neural Networks [1]. These are the subset of machine learning and are at the core of deep learning. Their structure and nomenclature are modeled after the human brain, mimicking the communication between biological neurons [2].**

This work is presented and explained by ANN, Graphical Neural Network [3], which is a type of NN that works on graphs. In today's world, one can see various real-life applications of GNNs like those in various social networks, prediction of molecules, and drug preparation in medical sciences, road traffic, etc.

The article deals with the application of the GNN showing how can a GNN helps in forecasting information about a person in a social network based on various given datasets. In the end, one can easily forecast the information of a person using various tools like Pytorch, etc.

Keywords— ANN, GNNs, Forecasting, Social Networks, Machine Learning.

I. Introduction

A social network is a collection of people and their mutual interactions with each other in a community. Collection of people and their mutual interactions with each other in a community is known as a social network. Such a kind of community can commonly be seen these days on any social media like Facebook, Instagram, Twitter, etc.

Actually, a social network is very much analogous to a Graphical Neural network in real life. For example; In a graph there are various entities called nodes with pre-defined personal features and various mutual relationships between these nodes which connect them as edges [4]. Similarly in a social network, the people in a community can be seen as nodes where every person has its own personal traits and the various relationships between the people such as various social interactions, number of mutual friends, number of followers, etc. can be seen as edges.

So, a social network becomes an ideal subject to implement and base a GNN on. By doing this one can perform various tasks on a social network using a GNN. Such tasks include prediction, regression, etc. There are various levels of such tasks [5]. [6] CNN regression for crop field prediction, and [10] forecast of the stock index by QBFTS and statistical weights on the relationship of QBFLRGs.

II. Graph Convolution

A. General Graph Convolution formula

Some insight will be given to the particular graph-based NN model f(X, A) that will be used in this research. Multiple-layer Graph Convolutional [9] with the propagation rule (layer-wise) is described below:

$$H^{(l+1)} = \sigma(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}) \quad (1)$$

The adjacency matrix of G (undirected graph) with additional self-connections $\tilde{A}$ is shown here as

$$A + I_N = \tilde{A}$$

For graph G, A is an adjacency matrix, (I_N) an identity matrix, and a layer-specific trainable weight matrix is $W^{(l)}$, σ (.) denote the activation function. The hidden feature matrix for l^{th} layer is $H^{(l)}$. Here, proposed how the uses of this propagation rule may be motivated by a 1st-order approximation of localised spectrum filters on graphs [17, [13].

B. Spectral Graph Convolution

Used spectral convolution as x, a signal as belongs to $\mathbb{R}^n$ and multiplied with filter, $g_\theta = \text{diag}(\Theta)$, parameterized by $\theta \in \mathbb{R}^N$ for Fourier domain, the result is a spectral convolution on the graph [13].

$$g_\theta(L) * x = U g_\theta(\Lambda) U^T x \quad (2)$$

in the above equation (2), the unitary matrix U of the eigenvectors of the symmetrically normalized Laplacian matrix,

$$L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

Here, D is a diagonal degree matrix, into eigen decomposed $U\Lambda U^T$. Diagonal matrix Λ consists of eigenvalues of L and $U^T x$ represents the graph's Fourier transform of x, and known as a function of L's eigenvalues, g_θ

Presentation of futuristic Malarial Disease through a Hybrid Model of A.I. and Big data

A.K. Awasthi
Department of Mathematics
Lovely Professional University
Phagwara, India
dramitawasthi@gmail.com

Minakshi Sharma
Department of Mathematics
Lovely Professional University
Phagwara, India
minakshisohareva@gmail.com

Arun Kumar Garov
Department of Mathematics
Lovely Professional University
Phagwara, India
arunkumar170@yahoo.com

Prayanshu Chaudhary
Department of Mathematics
Lovely Professional University
Phagwara, India
prayanshuchd512@gmail.com

***Abstract*—Transmission of parasites of female Anopheles mosquito is a major cause for the fatal disease like Malaria and most common symptoms are high fever, headache, abdominal pain, muscle pain, vomiting, diarrhea, Anemia, etc.**

Healthcare is a field where Forecasting is most beneficial and helpful for the cure of diseases and it can be possible only by using models of Neural Networks, Regression, and LSTM for predicting that how many confirmed cases will occur, how many people will get recovered and among them how many get dead cause of any disease based on past and present data for forecasting the future trend of these cases. Even though forecasting can be done by traditional methods but traditional methods are time consuming. The machine learning techniques are fast and accurate then also for adding efficiency and accuracy to these methods and techniques, Artificial Intelligence takes place in the field of Data Sciences. Artificial Intelligence came into the picture 100 years ago and shows remarkable growth in every field in the past few years. There are a lot of models proposed for different types of diseases in the field of healthcare. The visible patterns of symptoms and cases related to the disease can help in the medical field to prevent and cure it. This work is going to deal with the applications of Artificial Intelligence in Data Sciences which is to make an early reliable prediction of Malarial disease using Neural Networks, Regression, and LSTM model. It Compares the trends of different models and comparative results will demonstrate that the machine learning techniques practiced to forecast the malarial disease along with visible patterns gives information related to the patient's data.

Keywords— Artificial intelligence, Neural Networks, LSTM, ARIMA-SARIMA

I.INTRODUCTION

Vector-borne diseases are illnesses caused by the transmission of viruses, bacteria, and parasites through vectors. Every year around 1 billion people get infected, and more than 7 lakhs people die from diseases like Japanese encephalitis, yellow fever, Chagas disease, schistosomiasis, dengue, and malaria. In 2014, World Health Organization (WHO) created a theme related to vector-borne diseases along with the slogan "Small bite, big threat" to spread awareness of these fatal diseases among people.

Among all these vector-borne diseases, around 40% of the cases are of malaria, and WHO considers Malaria to pose the biggest threat to the world. Malaria is a parasitic infection and one of the vector-borne diseases which is caused by the transmission of parasites from female mosquitoes known as Anopheline mosquitoes. It causes more than 215 million cases and results in more than 4 lakhs deaths every year and the most vulnerable to this disease are children who are below 5 years.

Malaria also causes the deflation of economic growth by 1% point per year in endemic countries. Malaria and poverty are closely connected. The global allocation of per-capita gross domestic product shows a distinctive correlation between malaria disease and poverty.

Although, Malaria is a curable disease, and it can be prevented if it is done timely cured but deaths are very common among children and old age people, which is unavoidable. Every year the number of deaths increases with an increase in cases. There are a lot of countries that are still not able to control the inflation of cases and due to which several people and children die. Even with the advancement of technologies, diagnostics, and treatment modalities, this disease will remain a public health problem in countries that are still developing. There were 228 million estimated cases of malaria that occurred worldwide in 2018. World Health Organization (WHO) African Region was the region with the most cases, followed by 3.4 % of the cases in the South-East Asia Region and 2.1% cases of in WHO Eastern Mediterranean.

II.RELATED STUDY

Awasthi et al. [1] discussed the internet of things in healthcare. The various machine learning techniques and deep learning techniques help the common people in healthcare. Tai & Dhaliwal [2] used machine learning models and deep learning model also proposed to forecast individual malaria risk. If machine learning models doesn't works only then it is the alternative method. As this study works on Malaria Gen dataset hence this study provides genetic knowledge of malaria which can improve various forecasting methods of individual risk. Sharma [3] discusses various data mining techniques in healthcare and shows the various fields along with the various techniques applied to them. The author discusses all the techniques along with various examples of forecasting. Kumar et al. [4] explained the case study of malaria of Delhi (India) and the climatic effect on malaria Patients. In this the author forecasted the malaria cases by using ARIMA (0,1, 1), ARIMA (0,1,0) in which it is observed that ARIMA (0,1,0) shows exceptional result. Among the

The Pharma Innovation Journal 2021; SP-10(7): 05-08

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; SP-10(7): 05-08

www.thepharmajournal.com
Received: 04-05-2021
Accepted: 10-06-2021

Arun Kumar Garov
Department of Mathematics,
Lovely Professional University,
Phagwara, Punjab, India

AK Awasthi
Department of Mathematics,
Lovely Professional University,
Phagwara, Punjab, India

Corresponding Author:
Arun Kumar Garov
Department of Mathematics,
Lovely Professional University,
Phagwara, Punjab, India

A computational mathematical model for forecasting of Indian crop

Arun Kumar Garov and AK Awasthi

Abstract
In this paper, we have considering a forecasting of crops, such as Rice, Wheat, and Potatoes of Indian data from 2000 to 2019 year. Mathematical model with regression equation used for forecasting of Indian crop data. Where output experimental result use for the forecasting of next year production of the data.

Keywords: Forecasting, regression equation, crops, wheat, rice, potatoes

Introduction

For the forecasting of the production of the crop have several methods, which based on theoretical calculation and software calculation. Related work of crop production and agricultural field are completed by many agricultural and other field experts. For this paper, we also search some previous work and discuss it. Zhang *et al.* [1] applied single- variable regression in rice grain/vegetables versus natural log-transformed concentration in soil on cropland of China. Hare [2] discussed on the impact of defoliation on Potato yield and applied experiment was repeated seven times in intervals, where one interval is equal to two weeks. Risk- neutral and risk-averse formulations are applied on northeastern Oregon farms data by Nazer *et al.* [3]. Regression based model applied to estimate the impact of multiple pets on crop productivity by Johnson [4]. Gandhi *et al.* [5] applied a neural network for predict the rice production of districts of Maharashtra.

Regression Model

A good measure of the relationship between two variables is given by the coefficient of correlation which tells about strength and direction of a relationship. After correlation between two variables, we determine a mathematical relationship between them to achieve the following:

1. The value of a variable based on the other variable could be predicted.
2. How does the change in the value of one variable could impact the other explained.

For geometrical convenience and ease, we fit a linear relationship and analyze a regression model. The term regression was coin by Sir Francis Galton in 1885. To figure out the predictive power of an independent variables on the dependent variables there is a term used which is known as "Regression". It deals with the way changes in one variable based on how one or more other variables changes. Regression provides more information rather than correlation.

Dependent and independent variables:

Out of the two variables is considered dependent variable and other one is considered independent variable. For example, out of rainfall and yield of crop rainfall is independent and yield is dependent.

1. Simple linear Regression Model: It is a relation between one explanatory variable (X) and one response variable (Y). The simplest relationship is given by X and Y.

$$Y = a + b\,X + e$$

Where, the dependent variable is Y, the independent variable is X, a constant, b is coefficient and e represents error.

2. Multiple Linear regression Model: In some cases, the response variable Y may depend on

THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22-Issue-37-December -2019

Experimental model approach for decision making in Stock Index

Sumit Koul[1], A.K. Awasthi[2], Arun Kumar Garov[3]

Department of mathematics, Lovely Professional University, Punjab, India[1,2,3]
amit.25155@lpu.co.in[2], dramitawasthi@gmail.com[2], Sumit.24693@lpu.co.in[1], arunkumar170@yahoo.com[3]

Abstract

Purpose of study in this paper is to find the variance analysis experimental model approach between stock exchange trend obtained by the comparison of the RMSEs by different methods of forecasting used by Chen[1], Yu & Huarng[2], Huaing[3], Yu[5], Yu & Huarng[6] and Yu & Huarng[7] .

Key words: *Forecasting, fuzzy relation equation, mathematical model, CRD, RMSEs.*

Nomenclature

z_{ij}	dependent variable,
γ	general mean effect
β_i	treatment due to i^{th} effect,
ε_{ij}	error effect,
N	Number of observation
G	Grand total
C.F.	Correction Factor
RSS	Raw Sum of Square
TSS	Total sum of square
SST_R	Sum of square due to treatment
SSE	Sum of square due to Error
ANOVA	Analysis of variance
CRD	completely randomized design
S.O.V._1	Source of variation_1
D.F._1	Degree of Freedom_1
S.S.	Sum of squares_1
M.S.S._1	mean sum of square_1
l.o.s.	level of significance

Introduction

In Chen[1] Presenting a new method to forecast the Taiwan Stock Exchange Capitalization Weighted Stock Index(TAIEX) derived from the fuzzy time series and fuzzy variation groups. Huarng et al.[2] working on a handle non- linear problem used by fuzzy time series model and generate the non- linear arrangement of the neural network for the forecasting. Huang [3] analyzed the TAIEX historical data for forecasting based on fuzzy time series and some multivariate heuristic function. These heuristic functions are extended and integrate, which is

