

FORECASTING RICE PRICE IN INDIA USING STATISTICAL APPROACH

Submitted in partial fulfillment of the requirements for the award of

Master of Science Degree in Mathematics

By

HEMAVARTHINI. S (41630002)



**DEPARTMENT OF MATHEMATICS
SCHOOL OF SCIENCE AND HUMANITIES**

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

**Accredited with Grade "A" by
NAAC I 12B Status by UGC I Approved by AICTE
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119**

APRIL – 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC I 12B Status by UGC I Approved by AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

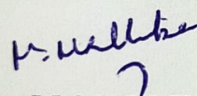
www.sathyabama.ac.in

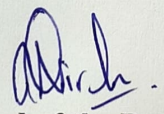
DEPARTMENT OF MATHEMATICS

BONAFIDE CERTIFICATE

This is to certify that this project report is the bonafide work of **HEMAVARTHINIS (41630002)** who carried out the project entitled "FORECASTING RICE PRICE IN INDIA USING STATISTICAL APPROACH" under our supervision from **JANUARY 2023 to APRIL 2023**.

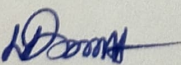
Internal Guide

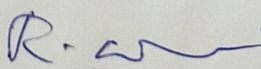

(Dr. M. MALLIKA)


Head of the Department
Dr. M. NIRMALA, M.Sc., M.Phil., Ph.D.,
Head of the Department
Dept. of Mathematics
Sathyabama Institute of Science and Technology
(Deemed to be University) Chennai-600 119.
(Dr. M. NIRMALA)

Submitted for Viva voce Examination held on

11/05/2023


Internal Examiner


External Examiner

DECLARATION

I HEMAVARTHINI.S (41630002) hereby declare that the project report titled "FORECASTING RICE PRICE IN INDIA USING STATISTICAL APPROACH" done by me under the guidance of Dr. M. MALLIKA at SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY submitted in partial fulfillment of the requirements for the award of Master of Science Degree in Mathematics.

DATE: 26-04-2023

PLACE: Chennai

S. Hemavarthini

SIGNATURE OF THE CANDIDATE

ABSTRACT:

Rice is the staple food crop of majority of India. India is the world's second largest producer of rice and the largest exporter of rice in the world. Prediction of the prices may help the farmers, traders and policy makers in making necessary decisions in minimizing and managing the risk of price fluctuations. The study compares the Moving Average model and Holt-Winters Exponential Smoothing model in terms of forecasting the monthly wholesale rice price per kg in India from January 2012 to December 2021. The models were used to analyze data and evaluated based on the RMSE and MAPE values. By comparing the error measures of the models, Holt Winters exponential smoothing model is found to be best model that fits the data. Therefore the Holt-Winters Exponential smoothing model is used to predict the rice price per Kg in India from January 2022 to August 2023.

CONTENT

CHAPTER 1	PAGE NO.
1.INTRODUCTION	1
1.1 IMPORTANCE OF RICE	1
1.2 TIME SERIES	2
1.3 TIME SERIES ANALYSIS	2
1.4 COMPONENTS OF TIME SERIES	3
1.5 DATA DESCRIPTION	3
CHAPTER 2	
2.1 LITERATURE SURVEY	4
CHAPTER 3	
3.1 AIM AND SCOPE	6
CHAPTER 4	
4. METHODOLOGY	7
4.1 MOVING AVERAGE	7
4.2 HOLT-WINTERS (TRIPLE EXPONENTIAL SMOOTHING)	7
4.3 ERROR MEASURE	9
4.4 MEAN ABSOLUTE PERCENTAGE ERROR(MAPE)	9
4.5 ROOT-MEAN-SQUARE ERROR (RMSE)	9
CHAPTER 5	
5. RESULTS AND DISCUSSION	11
5.1 MOVING AVERAGE MODEL	12
5.2 HOLT-WINTERS MODEL	13

5.3 MOVING AVERAGE VS HOLT-WINTERS EXPONENTIAL SMOOTHING	15
--	----

5.4 FUTURE PREDICTION	16
-----------------------	----

CHAPTER 6

6.1 CONCLUSION	18
----------------	----

REFERENCE	19
-----------	----

CHAPTER 1

1.INTRODUCTION

1.1 IMPORTANCE OF RICE

Rice is a staple food in India and is an integral part of the country's culture and cuisine. It is consumed by people from all walks of life, across various regions and communities in the country. It is the primary food grain for more than half of the country's population. It is a source of daily sustenance for millions of people, especially those who live in rural areas. Rice is a good source of carbohydrates, which are essential for providing energy to the body. It also contains important nutrients such as vitamins, minerals, and dietary fibers, which are necessary for maintaining good health. Rice is an important crop for the Indian economy and is grown in large quantities across various regions of country. It is also an important export commodity, generating significant foreign exchange earnings for the country. Rice cultivation and processing provide employment to millions of people in India, especially in rural areas where job opportunities are limited.

Rice (*Oryza Sativa*) is one of the most important agricultural commodities in the agricultural supply chain. India is one of the largest producers and exporters of rice in the world, and its contribution to the global rice trade is significant. India is the second-largest producer of rice in the world, after China. In the 2020-21 crop year, India produced around 120 million tones of rice. India is also one of the largest exporters of rice in the world. In the 2020-21 financial year, India exported around 14 million tones of rice, worth about \$7 billion. India exports different types of rice, including Basmati rice, non-Basmati rice, and parboiled rice. Basmati rice is a premium variety of rice, known for its fragrance, and is highly sought after in international markets. India's major markets for rice exports include countries in the Middle East, Africa, and Southeast Asia. Some of the top importers of Indian rice are Bangladesh, Saudi Arabia, and Iraq. The Indian government provides support to rice farmers and exporters through various policies and schemes, such as minimum support prices for rice, export incentives, and subsidies for rice processing and transportation. Overall, India plays a significant role in the global rice trade, and its contribution to rice exports is likely to remain important in the coming years.

1.2 TIME SERIES

A time series is a set of observations taken at specified times, usually at equal intervals. Mathematically, a time series is defined by the functional relation $y_t = f(t)$. In time series, the independent variable is time and the dependent variable is observation correspond to time t , where y_t is the value of the variable under consideration over time t . Thus, if the values of a variable at time points $t_1, t_2, \dots, t_n$ are $y_1, y_2, \dots, y_n$ respectively, then the series

$$\begin{array}{lcl} t & : & t_1 \quad t_2 \quad t_3 \quad \dots, \quad t_n \\ y_t & : & y_1 \quad y_2 \quad y_3 \quad \dots, \quad y_n \end{array}$$

constitute a time series.

Example:

- The selling amount (y_t) of a particular grocer's shop in every day (t).
- The number of new born baby in India (y_t) in each year (t).

1.3 TIME SERIES ANALYSIS

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. Time series are analyzed to discover past patterns of variability that can be used to forecast future values. Time series analysis comprises methods for analysing time series data in order to extract meaningful statistics and other characteristics of the data.

Examples of time series analysis in action include:

- Weather data
- Rainfall measurements
- Temperature readings
- Automated stock trading
- Industry forecasts

1.4 COMPONENTS OF TIME SERIES

The components of a time series can be divided into four main categories:

Trend: This refers to the long-term direction or pattern of the data over time. It represents the overall movement of the series, and can be upward, downward, or stable.

Seasonality: This refers to the pattern of periodic fluctuations in the data that occur at regular intervals, such as daily, weekly, or monthly. These fluctuations can be caused by factors such as weather, holidays, or other seasonal events.

Cyclical variations: This refers to longer-term fluctuations in the data that occur over periods of several years, and can be caused by economic or political factors, among others.

Random or irregular variations: This refers to the short-term fluctuations or noise in the data that cannot be explained by the other three components. These fluctuations can be caused by random events, measurement errors, or other factors.

1.5 DATA DESCRIPTION

Monthly rice price of India from 2012 January to 2021 December was taken as the data set. This research used the secondary data taken from the Ministry of Consumer Affairs, Food and Public Distribution website. It is the Ministry of Government of India responsible for the formulation and implementation of policies related to consumer affairs, food and public distribution in India. There are 120 observations in the dataset with some Statistical measures such as mean=36.12, median=36.375, Standard deviation=4.087, minimum value= 27.21, maximum value=42.33.

CHAPTER 2

2.1 LITERATURE SURVEY

Nurhamidah, Nusyirwan, Ahmad Faisol are the authors in the paper [6]“Forecasting Seasonal Time Series Data using Holt Winters Exponential Smoothing Method of Additive Models” researched to predict seasonal time series data using Holt-winters exponential smoothing additive model. The data used is the number of passengers departing at Hasanudin Airport in 2009 to 2019. They discussed the result that the Holt Winters exponential smoothing method on the passengers number contained trend patterns and seasonal patterns. They forecasted with initial values and smoothing parameters minimize forecast errors.

Kurumatani .K in the paper[4], “Time series forecasting of Agricultural product prices based on Recurrent Neural Network” and its Evaluation method implemented two methods for time series forecasting based on Recurrent Neural network one of which is called time-alignment of time point forecast (TATP), and another one is called direct future time series forecast (DFTS). TATP of LSTM shows superior performance when compared to other method.

Jalikatti, V.N.; Raghavendra, C.; Ahmad, D.G.; Shreya, A.; Sheikh, S. in the paper[3] “Forecasting the prices of onion in Belgaum market of northern Karnataka using ARIMA technique” has made an attempt to forecast the prices of onion at Belgaum district of Northern Karnataka. The forecasted values revealed an increasing trend for the future years. Hence, farmers need to plan the production process in such a way that a good price for the produce would be expected.

Rathod, S.;Saha, A.; Patil, R.; Ondrasek, G.; Gireesh, C.; Anantha, M.S.; Rao, D.V.K.N.; Bandumula, N.; Senguttuvel, P.; Swarnaraj, A.K.; Et al. in the paper[7], “ Two-stage spatiotemporal time series modelling approach for rice yield prediction & advanced agroecosystem management” proposed two stage Space Time Auto Regressive Moving Average(STARMA-II) for prediction of Rice yield in Andhra Pradesh in India. This method outperformed both ARIMA and STARMA models.

The authors I Djakaria and S E Saleh in the paper[2] “Covid-19 forecast using Holt-Winters exponential smoothing” researched the covid-19 pandemic data of Gorontalo, Indonesia from April 10 to October 13, 2020 and they found that the smoothing parameter $\alpha = 0.1$, $\gamma = \delta = 0.5$ for trend and seasonality respectively, which gives the smallest MAPE value of 6.14.

The authors Md. Habibur Rahman, Umma Salma, Md. Moyazzen Hossain, Md. Tareq Ferdous Khan in the paper[5] “Revenue Forecasting using Holt Winters exponential smoothing” researched on monthly revenue of Bangabandhu multipurpose bridge. They focus on the analysis of seasonal time series data using additive and multiplicative seasonal model of Holt-winters exponential smoothing and forecasted monthly revenue with best model additive Holt-winters exponential smoothing up to January 2021.

CHAPTER 3

3.1 AIM AND SCOPE

Average Rice Price of India monthly wise from Jan 2012 to Dec 2021 is taken as the data set.

AIM:

- To obtain a model to do analysis and forecast of India's average rice price with time series analysis.
- The model which gives less error measure is used to predict the future values in this paper.

SCOPE:

- Predicting the future price of Rice in India can be very useful for various stakeholders in the rice industry, including farmers, traders, and policymakers.
- Farmers are the primary producers of rice. They are directly affected by rice prices, and accurate price forecasting can help them make informed decisions about when to sell their crops.
- The government of India is the major stakeholder in the rice industry, and it plays an important role in regulating and supporting the industry. Accurate price forecasts can help the government make informed decisions about various policies and programs related to rice production, procurement, and distribution.

CHAPTER 4

4. METHODOLOGY

4.1 MOVING AVERAGE

The moving average is a statistical technique that is used to forecast long-term trends. The moving average is a simple and flexible process of trend measurement that is quite accurate under certain conditions. This method establishes a trend by using a series of averages covering overlapping periods of the data. The technique represents taking an average of a set of numbers in a given range while moving the range.

Moving Average can be calculated for the Time period 2, 3, 4....n.

FORMULA

The forecast value using Simple Moving Average is given by,

$$F_t = \frac{A_{t-1} + A_{t-2} + A_{t-3} + \dots + A_{t-n}}{n}$$

Where,

- F_t = Forecast for the period 't'.
- n = Number of terms to be averaged.
- $A_{t-1}, A_{t-2}, A_{t-3} \dots$ are the actual occurrences in the past period.
- two period ago, three period ago and so on respectively.

4.2 HOLT-WINTERS (TRIPLE EXPONENTIAL SMOOTHING)

The Holt-Winters model is a forecasting method for time series data that takes trends, seasonality, and the level of the time series. It is used for both additive and multiplicative time series models. The multiplicative Holt-Winters model is appropriate when the seasonal variations in the time series are proportional to the level of the time series. The Holt-Winters model for a multiplicative time series involves three components.

Level : The level of the time series is the average value of the series over time. The level is estimated using a weighted average of past observations, where more recent observations are given higher weights.

Trend : The trend component represents the rate of change of the time series over time. The trend is estimated using a weighted average of the differences between successive observations, where more recent differences are given higher weights.

Seasonality : The seasonal component represents the periodic variations in the time series that occur at fixed intervals, such as daily, weekly, monthly, or yearly. The seasonal component is estimated by averaging the deviations of the observed values from the level, and dividing by the average deviation for that season. The model requires several parameters one for each smoothing (α , β , γ), the length of a season, and the number of periods in a season.

FORMULA

$$S_t = \alpha \frac{y_t}{I_{t-m}} + (1-\alpha)(S_{t-1} + b_{t-1})$$

$$b_t = \beta(S_t - S_{t-1}) + (1 - \beta)b_{t-1}$$

$$I_t = \gamma \frac{y_t}{S_t} + (1 - \gamma)I_{t-1}$$

$$F_{t+m} = (S_t + mb_t)I_{t-m}$$

Where,

- y_t is the observation
- S_t is the level smoothing
- b_t is the trend smoothing
- I_t is the seasonal smoothing
- F_{t+m} is the forecast at m periods ahead
- t is an index denoting a time
- α, β and γ are the constants that must be estimated in such a way that the RMSE error is minimized.

4.3 ERROR MEASURE

Forecast error is a metric that defines the actual quantity and how it deviates from the forecasted quantity. Forecast error focuses on the magnitude of the error rather than whether it is positive or negative.

4.4 MEAN ABSOLUTE PERCENTAGE ERROR(MAPE)

The Mean Absolute Percentage Error (MAPE) is also called the Mean Absolute Percentage Deviation (MAPD). It measures accuracy of a forecast system. It represents the average of the absolute percentage errors of each entry in the dataset to calculate how accurate the forecasted quantities were in comparison with the actual quantities.

FORMULA

$$MAPE = \frac{1}{n} \sum \left| \frac{A_t - F_t}{A_t} \right| \times 100$$

Where,

- n = Sample size
- A_t = Actual value
- F_t = Forecast value

4.5 ROOT-MEAN-SQUARE ERROR (RMSE)

The Root-Mean-Square Error (RMSE) or Root-Mean-Square Deviation(RMSD) is a frequently used to measure the differences between values (sample or population values) predicted by a model or an estimator and the values observed. The RMSE represents the square root of mean of squared error.

FORMULA

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}$$

where,

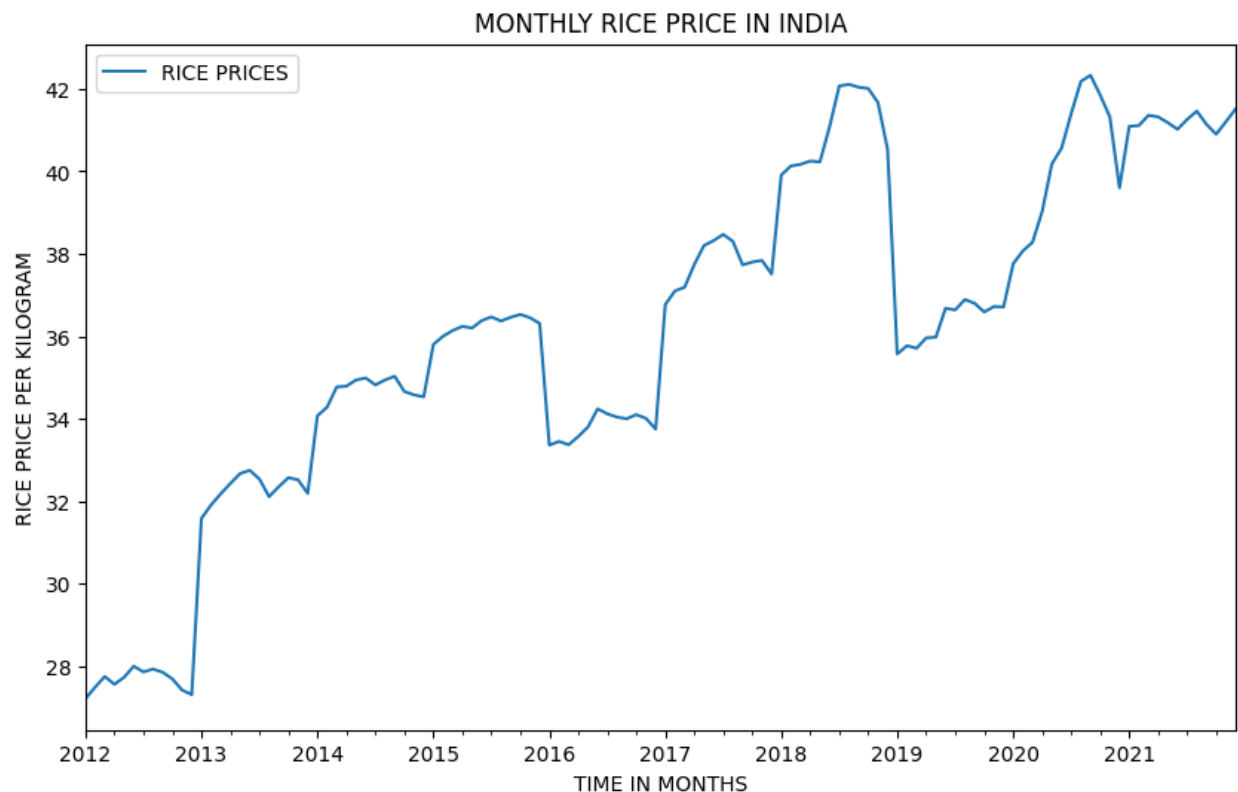
- x_i =Actual data
- y_i =Estimated data
- N =Number of data points

CHAPTER 5

5.RESULTS AND DISCUSSION

The data set from Jan 2012 to Dec 2021 of Rice price of India per kg undergoes analysis. The length of dataset is 120. To visualise this time series data for the observation of price pattern, the data set is plotted and the graph is shown below. The blue line indicates the actual price of rice in India. The price pattern shows the upward trend and seasonality.

GRAPH



The dataset is split into two sets which are training data and test data for the calculation of error measure. The first 100 observations are considered as training data and remaining 20 are considered as test data. The splitting of data may helps the time series models to analyse train data to forecast the test data.

The models Moving average and Holt-Winters exponential smoothing were applied in the training set to forecast test data.

5.1 Moving Average Model:

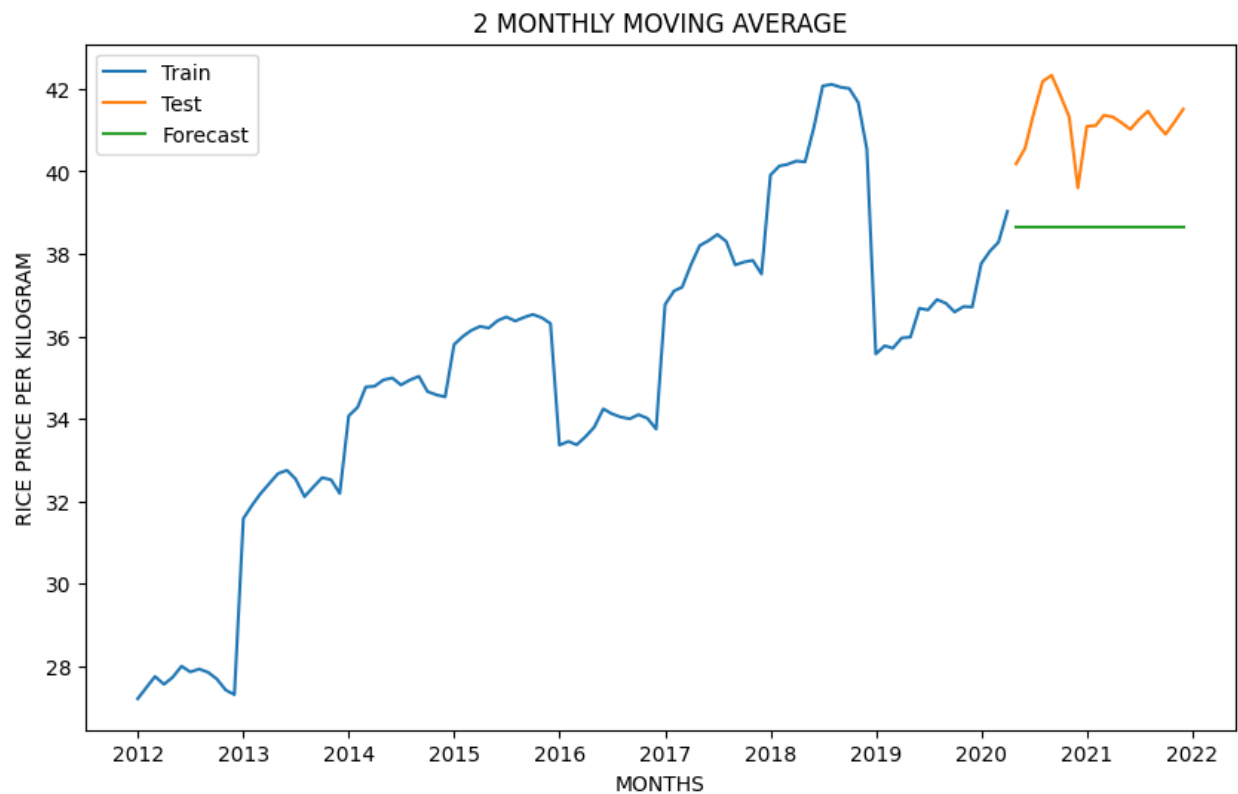
Moving Average model is performed for different time period on training data and compared with actual data to observe the error measure.

MOVING AVERAGE	RMSE	MAPE
2 Monthly Moving Average	2.612	6.154
3 Monthly Moving Average	2.802	6.62
4 Monthly Moving Average	2.973	7.052
5 Monthly Moving Average	3.282	7.817
6 Monthly Moving Average	3.487	8.322
7 Monthly Moving Average	3.652	8.729
8 Monthly Moving Average	3.75	8.970
9 Monthly Moving Average	3.817	9.133
10 Monthly Moving Average	3.89	9.324

Here, the table shows that 2 monthly moving average error measure is lower than all other. Lower the error higher the accuracy. Therefore among different time period of moving average, 2 monthly moving average has better forecast.

The Graph fitted for 2 monthly moving average consists of training, testing and forecasting plots. Forecasting is done with Python code in Jupyter notebook. The graph below shows the forecast of test data and error between actual and forecast data.

GRAPH



Here, the blue, orange and green line indicates train data, test data and forecast for test data.

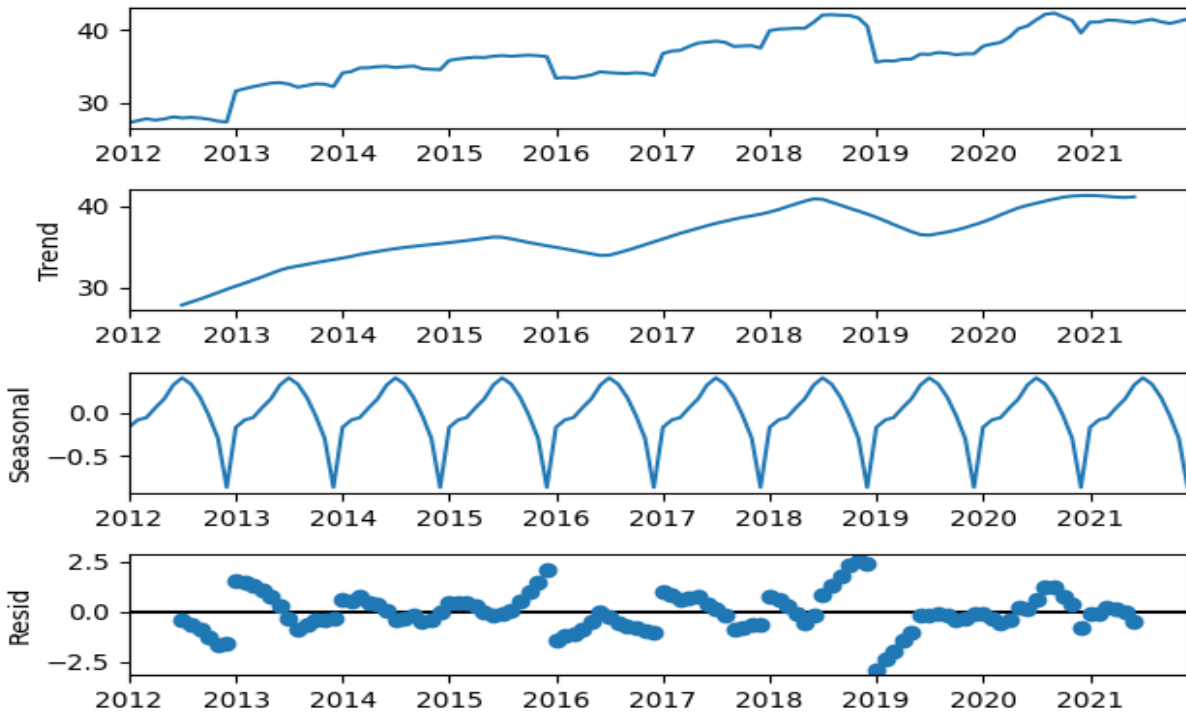
5.2 HOLT-WINTERS MODEL

Before performing Holt Winters model, the data is decomposed to notice trend and seasonal parts to perform this model more efficiently.

DECOMPOSITION

Decomposition of time series is a method used to separate a time series into its individual components, such as trend, seasonality, error. These three time series components were decomposed in graph. The original time series (top), the trend component (second from top), the seasonal component (third from top), and the irregular component (fourth from top) are all shown in the graph below.

GRAPH OF DECOMPOSED DATA



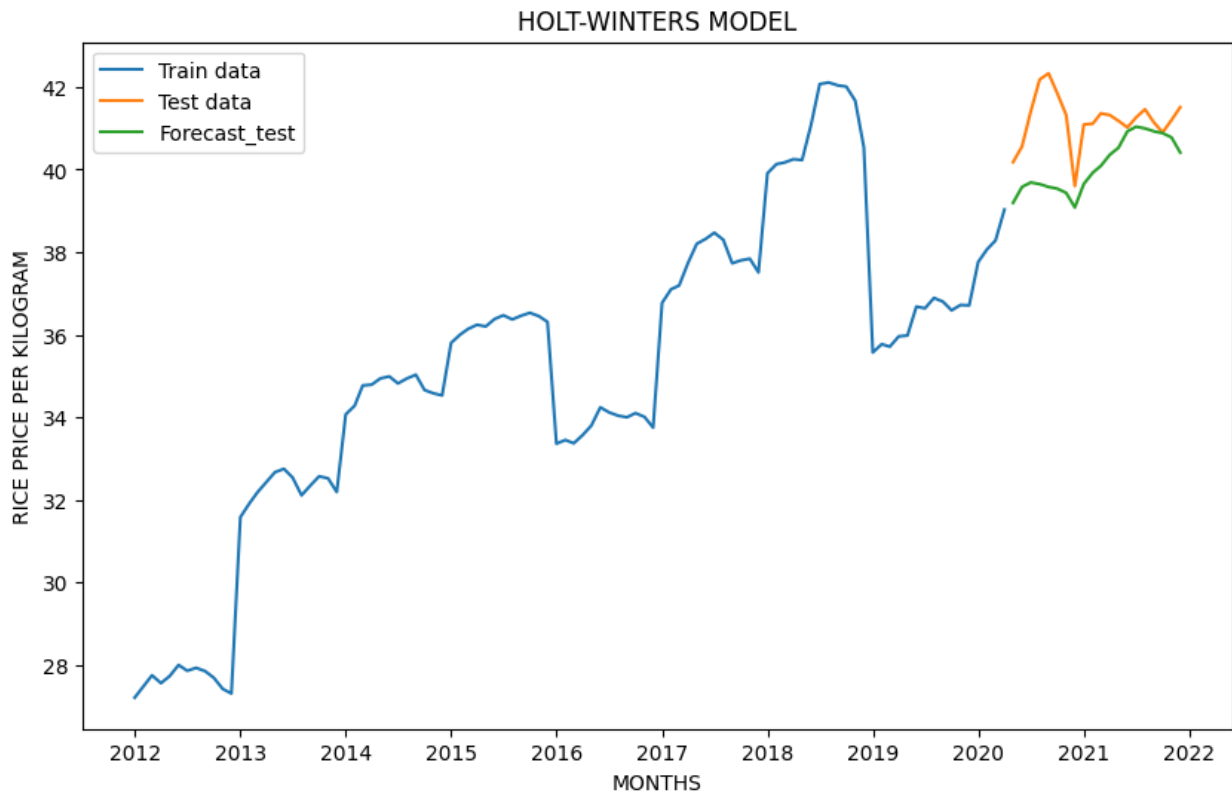
We observe that there is the trend and seasonal component is present in the data. There is gradual upward trend and with the seasonal period of 12. In 120 data, first 100 are trained to fit the Holt-Winters model for forecasting. The values of the constants α, β, γ takes from 0 to 1. Different combinations of these constants may use for forecasting but only one combination gives the forecast value which are closer to actual value. The combination $(1.0, 0.00, 3.2822e^{-12})$ is found to be the best fit for forecasting test data. The initial values for each smoothing were obtained. The initial level smoothing $S_0=27.096$, initial trend smoothing $b_0=1.002$, initial season $I_1=1.001, I_2=1.005, I_3=1.006, I_4=1.01, I_5=1.012, I_6=1.019, I_7=1.019, I_8=1.015, I_9=1.010, I_{10}=1.006, I_{11}=1.0015, I_{12}=0.989$. The forecasted values are compared with the actual value for capturing the error for this method.

The error found in this method are

MAPE=2.631

RMSE=1.361

GRAPH



Here the blue line, orange line and green line indicates training data, test data and forecast for test data respectively.

5.3. MOVING AVERAGE VS HOLT WINTERS EXPONENTIAL SMOOTHING

By the result of Moving Average, 2 monthly moving average is better among them and now it is compared with Holt Winters model. The table shows the comparison of model with their error.

ERROR MEASURE	2 MONTHLY MOVING AVERAGE MODEL	HOLT-WINTERS MODEL
RMSE	2.61	1.361
MAPE	6.15	2.63

The table shows the error measures. The RMSE and MAPE values of Holt Winters is lower than the other model. Therefore Holt Winters model is better than Moving average model.

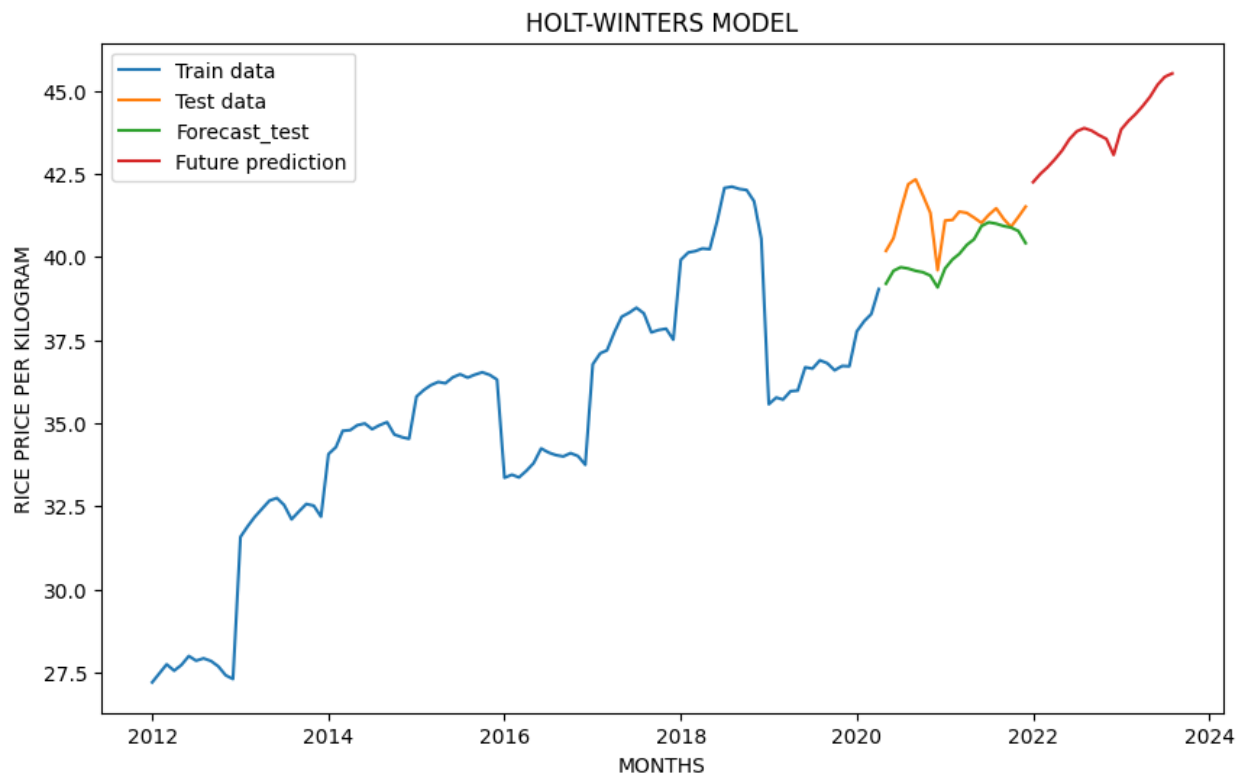
5.4. FUTURE PREDICTION

To predict the average rice price of India from 2022 January to 2023 August, Holt Winters model is applied on whole data. The values of the smoothing parameters α, β, γ are 1.0, $2.559e^{-11}$, 0.00 found to be the best fit. The initial level smoothing $S_0=25.26$, initial trend smoothing $b_0=1.003$, initial season $I_1=1.007, I_2=1.077, I_3=1.078, I_4=1.081, I_5=1.084, I_6=1.090, I_7=1.092, I_8=1.091, I_9=1.086, I_{10}=1.079, I_{11}=1.073, I_{12}=1.058$.

Here is the table shows the prediction of rice price from January 2022 to August 2023.

MONTHS	AVERAGE RICE PRICE IN INDIA PER KILOGRAM
2022 JAN	42.24
2022 FEB	42.49
2022 MAR	42.68
2022 APR	42.92
2022 MAY	43.19
2022 JUN	43.54
2022 JUL	43.77
2022 AUG	43.87
2022 SEP	43.79
2022 OCT	43.65
2022 NOV	43.54
2022 DEC	43.06
2023 JAN	43.82
2023 FEB	44.09
2023 MAR	44.28
2023 APR	44.53
2023 MAY	44.81
2023 JUN	45.17
2023 JUL	45.41
2023 AUG	45.51

GRAPH



The graph shows training, testing, forecast value for testing data and future prediction which are blue, orange, green and red in color respectively.

CHAPTER 6

CONCLUSION

Among many forecasting models for time series prediction, here the Holt-winters and moving average models were performed to show the results. According to the analysis and discussion of Average Rice Price of India from 2012 January to 2021 December monthly wise, it can be concluded that the forecasting model Holt-Winters exponential smoothing is better than that of 2 monthly moving average. The models are compared with their error measures such as RMSE and MAPE. RMSE of Holt-Winters is less than moving average model ($1.36 < 2.61$) and also MAPE of Holt-Winters is less than moving average model ($2.63 < 6.15$). Therefore the Holt winters Exponential smoothing model is used for future prediction from 2022 January to 2023 August.

REFERENCES:

- [1] Headey, D.D. Rethinking the Global Food Crisis: The Role of Trade Shocks. Food Policy, **2011**.
- [2] I Djakaria and S E Saleh. Covid-19 forecast using Holt-Winters exponential smoothing, SEA-STEM **2020**.
- [3] Jalikatti, V.N.; Raghavendra, C.; Ahmad, D.G.; Shreya, A.; Sheikh, S. Forecasting the prices of onion in Belgaum market of northern Karnataka using ARIMA technique. Int. J. Agric. Econ. **2014**.
- [4] Kurumatani, K. Time series forecasting of Agricultural product prices based on Recurrent Neural Network and its Evaluation method. SN appl. Sci. **2020**.
- [5] Md. Habibur Rahman, Umma Salma, Md, Moyazzen Hossain, Md, Tareq Ferdous Khan. Revenue Forecasting using Holt Winters exponential smoothing, **2017**.
- [6] Nurhamidah, Nusyirwan, Ahmad Faisol. Forecasting Seasonal Time Series Data using Holt Winters Exponential Smoothing Method of Additive Models. **2020**.
- [7] Rathod, S.; Saha, A.; Patil, R.; Ondrasek, G.; Gireesh, C.; Anantha, M.S.; Rao, D.V.K.N.; Bandumula, N.; Senguttuvel, P.; Swarnaraj, A.K.; Et al. Two-stage spatiotemporal time series modelling approach for rice yield prediction & advanced agroecosystem management. Agronomy **2021**.
- [8] Yohana James Mgale, Yunxian Yan, Shauri Timothy. A Comparative study of ARIMA and Holt Winters exponential smoothing models for Rice price Forecasting in Tanzania, **2021**