Indian Trade Data Analysis and Forecasting

*By*

Akarsh Somani (162) & Gaurav Misra (172)



*Bachelor Thesis submitted to*

Indian Institute of Information Technology Kalyani

*For the partial fulfillment of the degree of*

**Bachelor of Technology**

**in**

**Computer Science and Engineering**

**May, 2020**

Certificate

This is to certify that the thesis entitled “Indian Trade Data Analysis and Forecasting” being submitted by Akarsh Somani and Gaurav Misra, an undergraduate student, Reg. No 162 and 172, Roll No. 39/CSE/16005 and 39/CSE/16015, in the Department of Computer Science and Engineering, Indian Institute of Information Technology Kalyani, West Bengal 741235, India, for the award of Bachelors of Technology in Computer Science and Engineering is an original research work carried by him under my supervision and guidance. The thesis has fulfilled all the requirements as per the regulations of Indian Institute of Information Technology Kalyani and in my opinion, has reached the standards needed for submission. The work, techniques and the results presented have not been submitted to any other University or Institute for the award of any other degree or diploma.

Dalia Nandi

Assistant Professor (IIIT Kalyani)

Electronics and Communication Engineering Department

Declaration

I hereby declare that the work being presented in this thesis entitled, “Indian Trade Data Analysis and Forecasting”, submitted to Indian Institute of Information Technology Kalyani in partial fulfillment for the award of the degree of **Bachelor of Technology** in Computer Science and Engineering during the period from January, 2020 to May, 2020 under the supervision of Dr. Dalia Nandi, Department of Electronics and Communication Engineering, Indian Institute of Information Technology Kalyani, West Bengal 741235, India, does not contain any classified information.

Akarsh Somani and Gaurav Misra

Reg No/Roll No: 39/CSE/16005/162 and 39/CSE/16015/172

Department: Computer Science and Engineering

Institute Name: Indian Institute of Information Technology Kalyani

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

..................................

Dr. Dalia Nandi

Position: Assistant Professor

Departmental address: Electronics and Communication Engineering

Place: Kalyani

Date: May 1, 2020

Acknowledgments

Firstly, we would like to thank our supervisor Dr. Dalia Nandi for her support and guidance to complete this project work without whom we would not be able to make out this far. We would also like to thank our friends who supported us greatly and were always willing to help us. We are very grateful to Department of Computer Science and Engineering, Indian Institute of Information Technology Kalyani, West Bengal 741235, India, for providing us this wonderful opportunity. Lastly, we would like to thank our Parents a God for their never-ending grace.

Akarsh Somani and Gaurav Misra

Reg No/Roll No: 39/CSE/16005/162 and 39/CSE/16015/172

Department: Computer Science and Engineering

Institute Name: Indian Institute of Information Technology Kalyani

Place: Kalyani

Date: May 1, 2020

Abstract

This thesis investigates about the Indian Trade. There is a lot of things which is very important of the economy of our country and trade is one of them. We have done Data Analysis on the Indian trade and based on the results we will be drawing some conclusions.

Trade is the economic concept which invokes on BUY and Sell of the commodities, or exchanging goods and services. Trade increases competition and decreases overall world wise cost for a product. Here we will be dealing with the Indian trade with the other countries and the impact of the trade. Trade is one of the important factors for the economy of our country and that is why this trade data analysis is very important. All the commodities are divided into 99 Classes called HS Code (Harmonized System Code). Each class represent some commodity of trade and here our analysis starts. We have all the countries present from which India trade and for each country we have the HS Code which India trade from that particular country.

Looking at the above importance we tried to find some answers like

1. What HS Code we import/export the most?
2. From which country we import export the most?
3. Rate of change of trades in years.
4. From the country we import/export the most, what are the products HS Code traded.
5. Any future predictions on the Indian trade.
6. What category we trade the most year wise.
7. Highest trading country year wise.

Table of Contents

[**Objective –** 1](#_Toc41824334)

[**Overview about the data –** 1](#_Toc41824335)

[**Methods –** 2](#_Toc41824336)

[**Lets first discuss the types of timeseries –** 2](#_Toc41824337)

[**Exponential Smoothing(Exponential Averaging) –** 3](#_Toc41824338)

[**Auto Regressive Model –** 4](#_Toc41824339)

[**Moving Average Model –** 5](#_Toc41824340)

[**Holt-Winters Method –** 6](#_Toc41824341)

[**ARIMA Multiplicative –** 7](#_Toc41824342)

[**ARIMA Additive –** 8](#_Toc41824343)

[**Seasonal ARIMA –** 9](#_Toc41824344)

[**Conclusion –** 10](#_Toc41824345)

# **Objective –**

Trade is the economic concept which invokes on BUY and Sell of the commodities, or

exchanging goods and services. Trade increases competition and decreases overall world wise

cost for a product. Here we will be dealing with the Indian trade with the other countries and the

impact of the trade. Trade is one of the important factors for the economy of our country and that

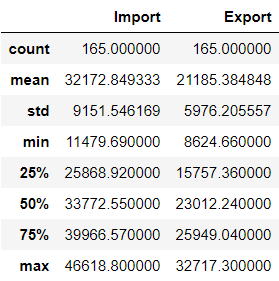
is why this Indian trade data analysis and forecasting is very important.

Here we will be forecasting using different models and check the prediction with the available data.

# **Overview about the data –**

The data is taken from the [MINISTRY OF COMMERCE AND INDUSTRY DEPARTMENT, GOVT. OF INDIA](https://commerce.gov.in/).

The data is monthly data from Jan 2006 to Sept 2019.The value is in million US Dollar $ and the stats are as follows–



# **Methods –**

We built seven models to forecast –

* Exponential Smoothing
* Auto Regressive
* Moving Average
* Holt-Winters model
* ARIMA Multiplicative
* ARIMA Additive
* Seasonal ARIMA

Now we will go one by one with each model and see there methodologies and prediction –

## **Lets first discuss the types of timeseries –**

There are basic two kind of timeseries – Multiplicative and Additive

If the amplitude does not vary then that is Additive and if amplitude varies then that is Multiplicative.

Multiplicative Time Series = Trend \* Seasonality \* Randomness



Additive Time Series = Trend + Seasonality + Randomness



## **Exponential Smoothing(Exponential Averaging) –**

Exponential Smoothing is the technique for smoothing(Averaging) the timeseries using exponential window function. In simple moving average the past observation are equally weighted where as here it exponentially decreases over the time.

In this method we use



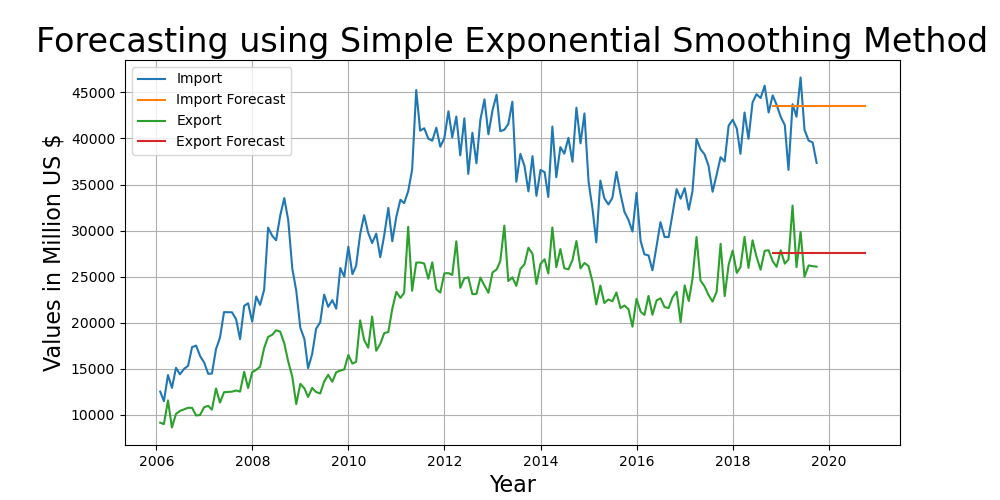
Where S0 is the value at time t = 0 and the forecast at time t is given as St

The term smoothing factor applied to α here is something of a misnomer, as larger values of α actually reduce the level of smoothing, and in the limiting case with α = 1 the output series is just the current observation.

As mentioned, it uses the exponential window function, we substitute the value of the above equation back to itself.



In other words, it forms a GP which is a discrete version of an exponential function.



RMSE Import/Export - 3454.44, 2073.22

## **Auto Regressive Model –**

It is used when a value from the time series has dependency on previous values, like Yt on Yt-1 .

Ex - Yt = f(Yt-1)

The order of an autoregression is the number of immediately preceding values in the series that are used to predict the value at the present time.

An order of 2 denotes that the time t is predicted based on t-1 and t-2.

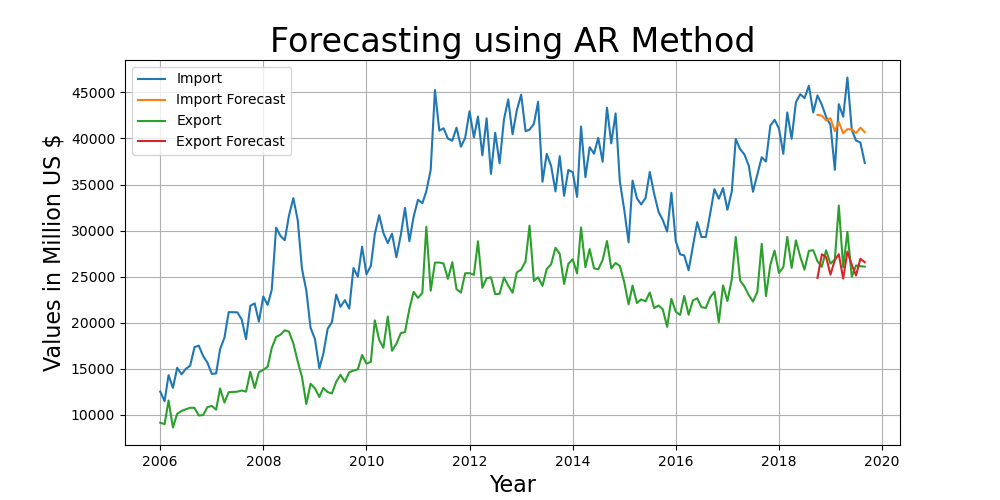
Terminologies - **Autocorrelation Function** (**ACF**) and **Partial Autocorrelation Function** (**PACF**).

ACF = CORR (Yt , Yt-k) ,

PACF can be calculated as

* Remove the linear dependency from the timeseries
* Calculate the correlation.

PACF is used to find the order of the autoregressive model.



RMSE Import/Export - 2530.75, 1932.5

## **Moving Average Model –**

Moving Average is a technique that calculates the overall trend in the dataset. As the name suggest that we go by taking the Average over a fixed rolling size window. The MAt is calculated by taking the unweighted mean of the previous window\_size (here 4) data.

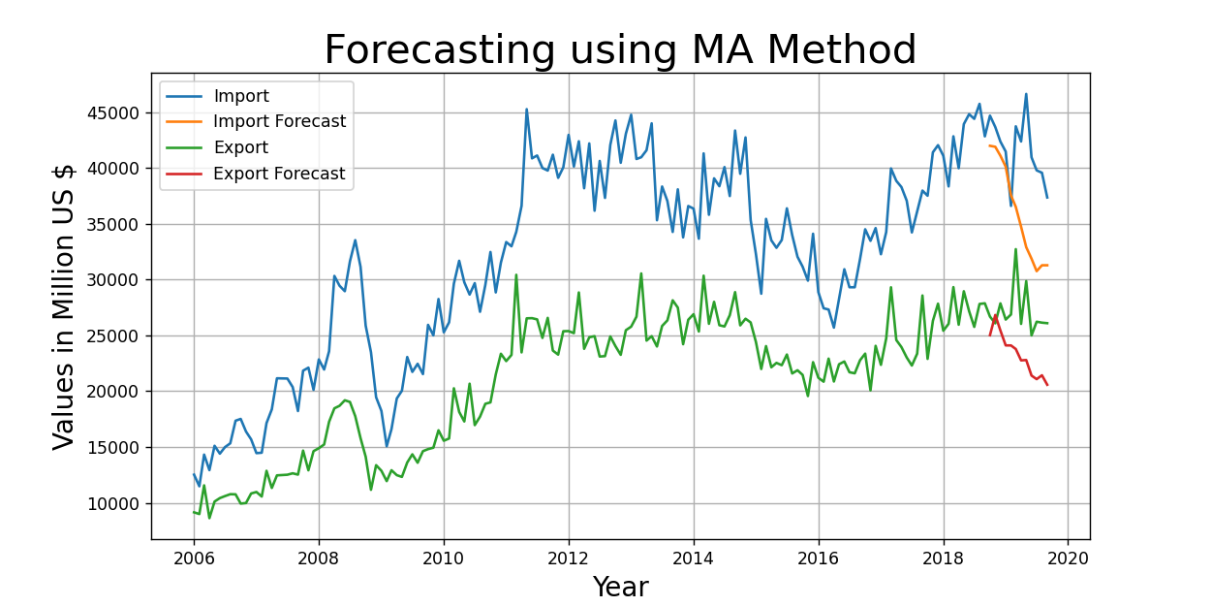
The moving average is thus calculated as –



MA4 = X4 + X3 + X2 + X1

And the successive value can be calculated as –

MAn = MAn-1 + (Xn – Xn-window\_size)

In our model we have chosen the window of 12

RMSE Import/Export - 6969.83, 4609.61

## **Holt-Winters Method –**

It is the extension over the simple exponential smoothing method. Here we use triple smoothing with the factor - seasonal period, trend type and seasonal type. Here seasonal and trend type means *Multiplicative* or *Additive*.

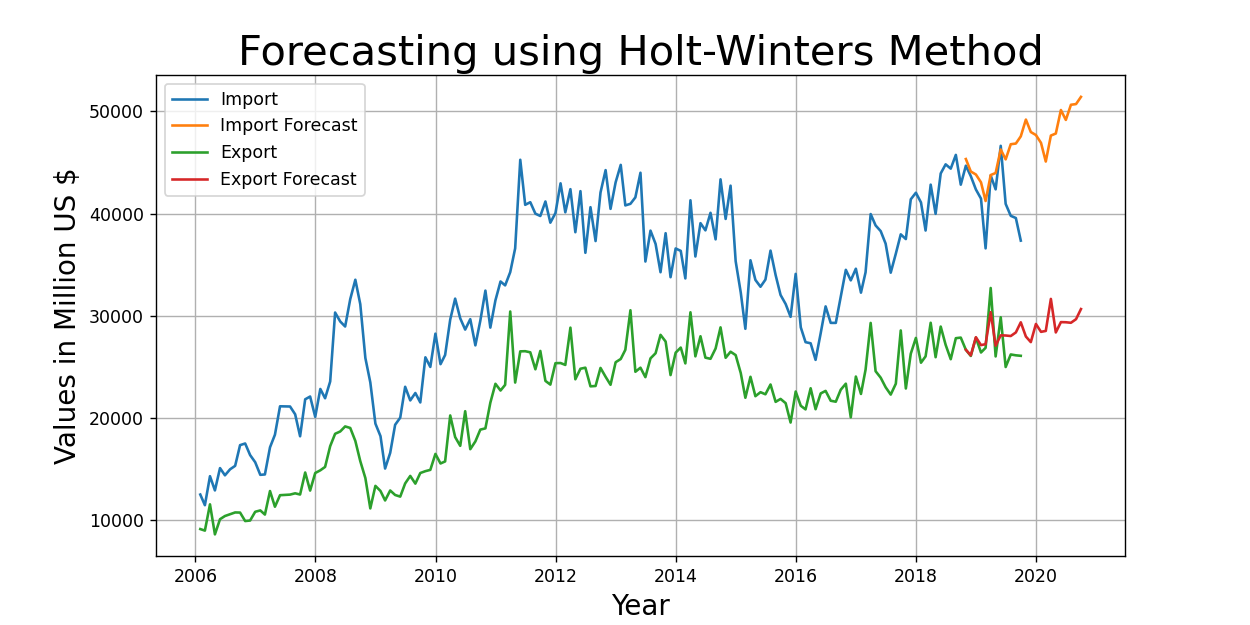
st represents the smoothed value of the constant part for time t. bt represents the sequence of best estimates of the linear trend that are superimposed on the seasonal changes. ct is the sequence of seasonal correction factors. ct is the expected proportion of the predicted trend at any time t mod L in the cycle that the observations take on. As a rule of thumb, a minimum of two full seasons (or 2L periods) of historical data is needed to initialize a set of seasonal factors.

The output of the algorithm is again written as Ft+m, an estimate of the value of x at time t+m, m>0 based on the raw data up to time t. Triple exponential smoothing with multiplicative seasonality is given by the formulas{\displaystyle {\begin{aligned}s\_{0}&=x\_{0}\\s\_{t}&=\alpha {\frac {x\_{t}}{c\_{t-L}}}+(1-\alpha )(s\_{t-1}+b\_{t-1})\\b\_{t}&=\beta (s\_{t}-s\_{t-1})+(1-\beta )b\_{t-1}\\c\_{t}&=\gamma {\frac {x\_{t}}{s\_{t}}}+(1-\gamma )c\_{t-L}\\F\_{t+m}&=(s\_{t}+mb\_{t})c\_{t-L+1+(m-1)\mod L},\end{aligned}}}



where α is the data smoothing factor, 0 < α < 1, β is the trend smoothing factor, 0 < β < 1, and γ is the seasonal change smoothing factor, 0 < γ < 1. [Reference](https://en.wikipedia.org/wiki/Exponential_smoothing)

In our model we have used seasonal period = 12 and trend type and seasonal type as multiplicative.



RMSE Import/Export - 3768.38, 1642.84

## **ARIMA Multiplicative –**

ARIMA stands for *Autoregressive Integrated Moving Average* which is a combination of three terms –

The AR part of ARIMA indicates the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

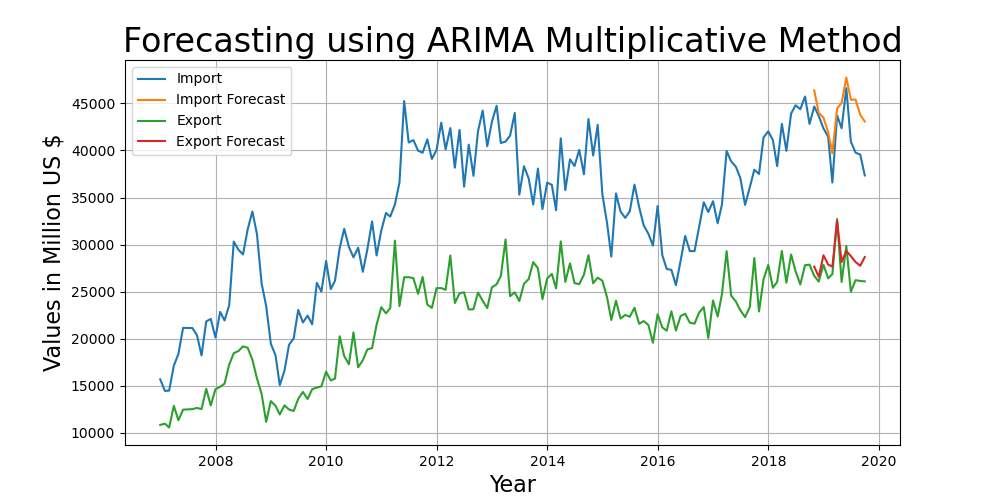
Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model.

We have already seen the Autoregressive and Moving average individually.

Multiplicative Time Series = Trend \* Seasonality \* Randomness, As explained previously.

For predicting the p and q we use partial autocorrelation and autocorrelation function. We have already discussed about them earlier. We used plots to find the optimal value of p and q graphically.

Here we are taking the (p,d,q) = (1,0,5). d as 0 because we do not require the differencing as our randomness is stationary.



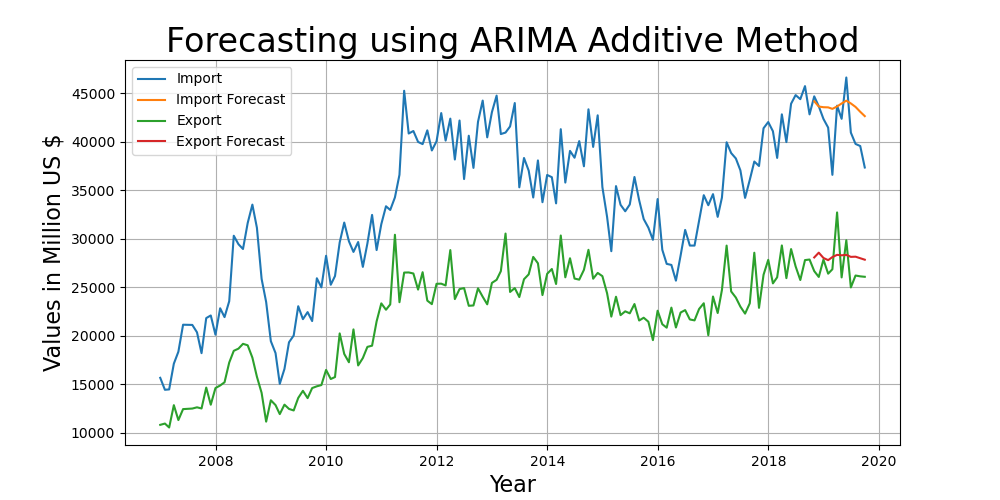
RMSE Import/Export - 3238.24, 1762.79

## **ARIMA Additive –**

This is similar to the above method except for the fact that we have considered our time series as Additive (just for comparison).

Additive Time Series = Trend + Seasonality + Randomness

Here we are taking the (p,d,q) = (1,0,5). d as 0 because we do not require the differencing as our randomness is stationary.

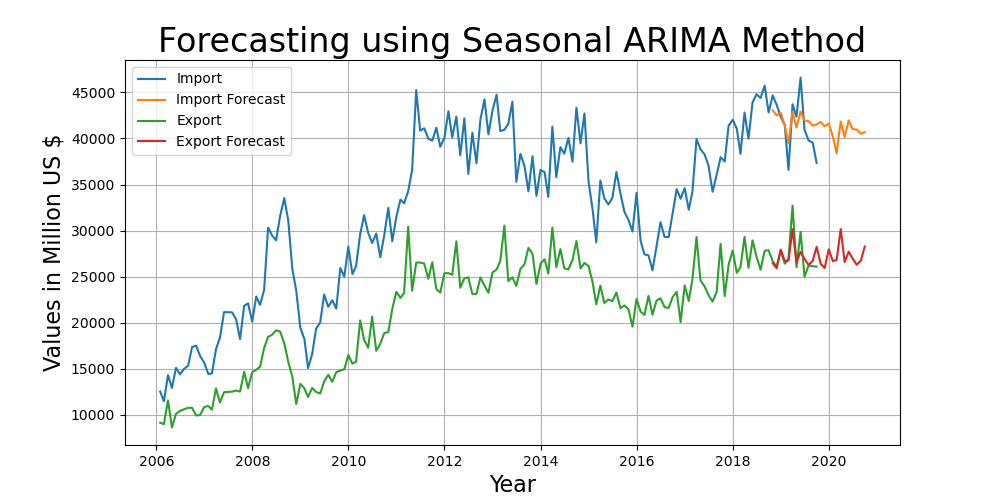


RMSE Import/Export - 3221.36, 2210.54, we can see that the forecast does not fits right.

## **Seasonal ARIMA –**

Seasonal ARIMA here is ARIMA Multiplicative method with a seasonality factor m. Here we have chosen the multiplicative factor of 12 as ours is a monthly data.

Also we used brute force to find the best p,d,q values which reduces the RMSE. Import (1,0,0) and Export (1,0,2). This model has performed quite well and thus below is the plot.

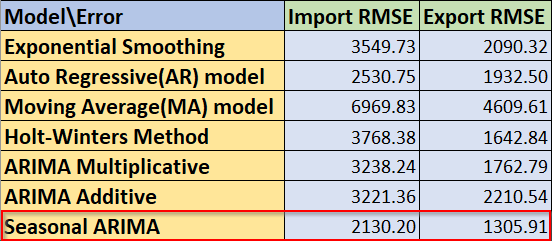


RMSE Import/Export - 2130.2, 1305.91

# **Conclusion –**

All the models perform quite well but the best observation which we have observed is of Seasonal ARIMA.

Below is the RMSE Comparisons –



The comparisons are so close because one model is the extension of the other with some advancement.

Holt-winters is the advancement of the Simple exponential smoothing. ARIMA is the combination of AR and MA model, Seasonal ARIMA is the advancement of the ARIMA model.

The results are quite satisfactory and we see that the there is a downward trend in Import forecast and upward trend in the Export forecast. The more we export and the less we import is beneficial for the Indian economy. This shows that India will be playing pretty good in terms of trade.