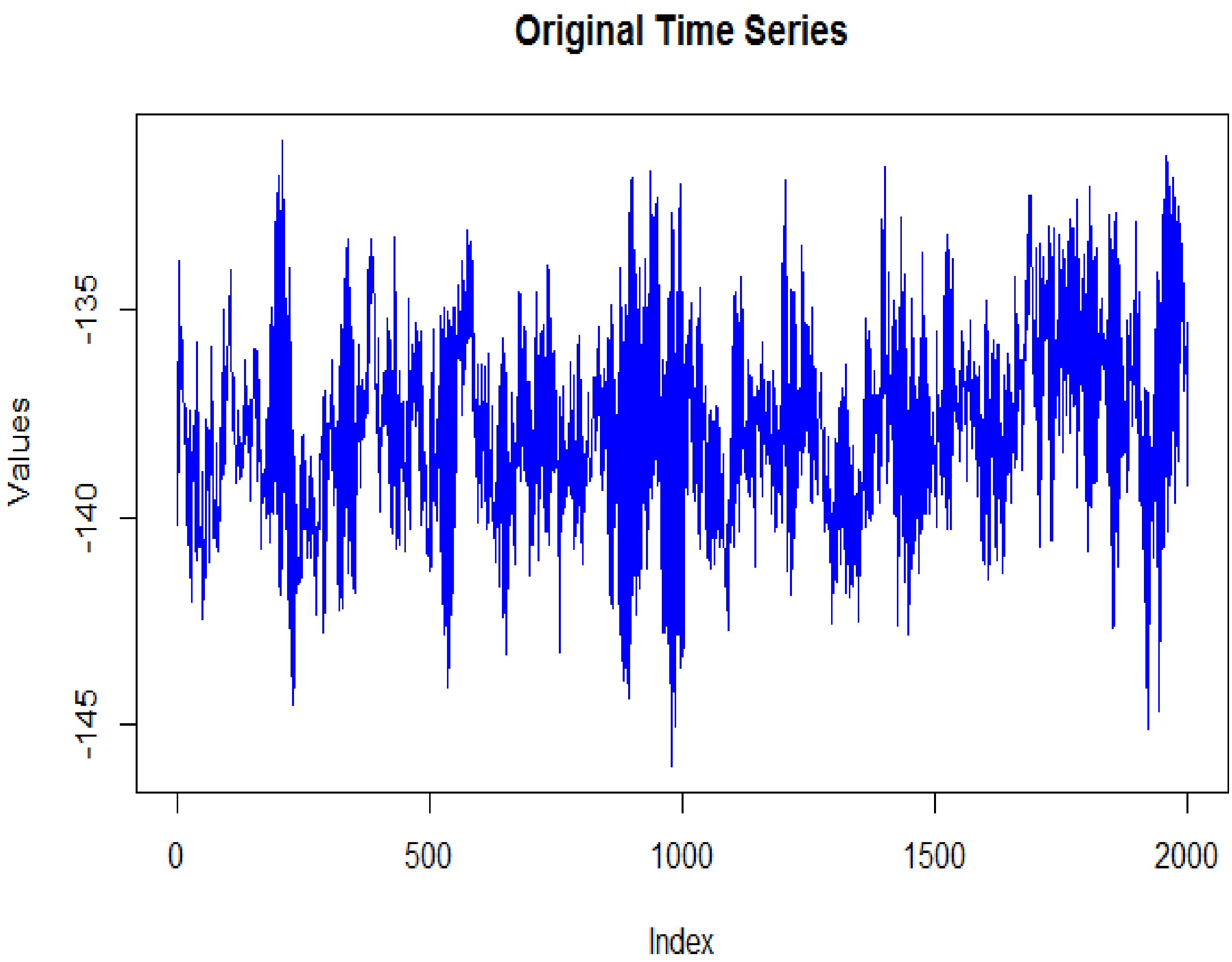
CSC – 591 Internet of Things Analytics (Project 3: Forecasting)

Name :- Harsh Jatinbhai Patel
Student Id :- 200258486
Unity Id :- hpatel8

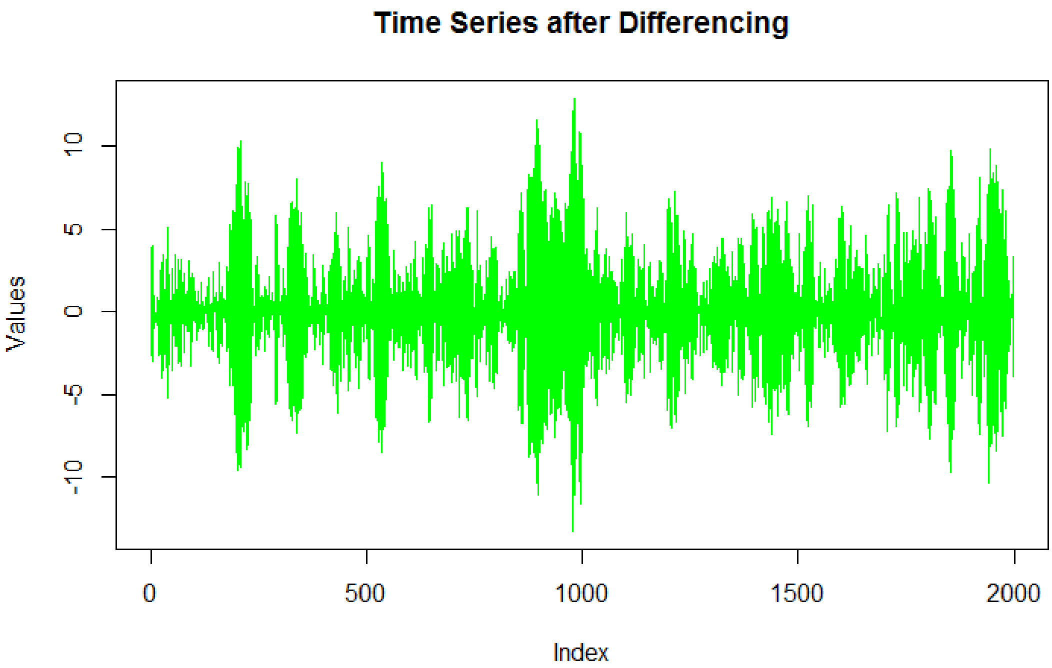
<u>Task: 1 Check for stationarity.</u>

- A stationary time series is one whose properties do not depend on the time at which the series is observed. A stationary series has the property that the mean, variance and autocorrelation structure do not change over time.

- **<u>Original Time Series Plot from data given</u>**



- <u>**Time Series Plot after performing differencing operation**</u>



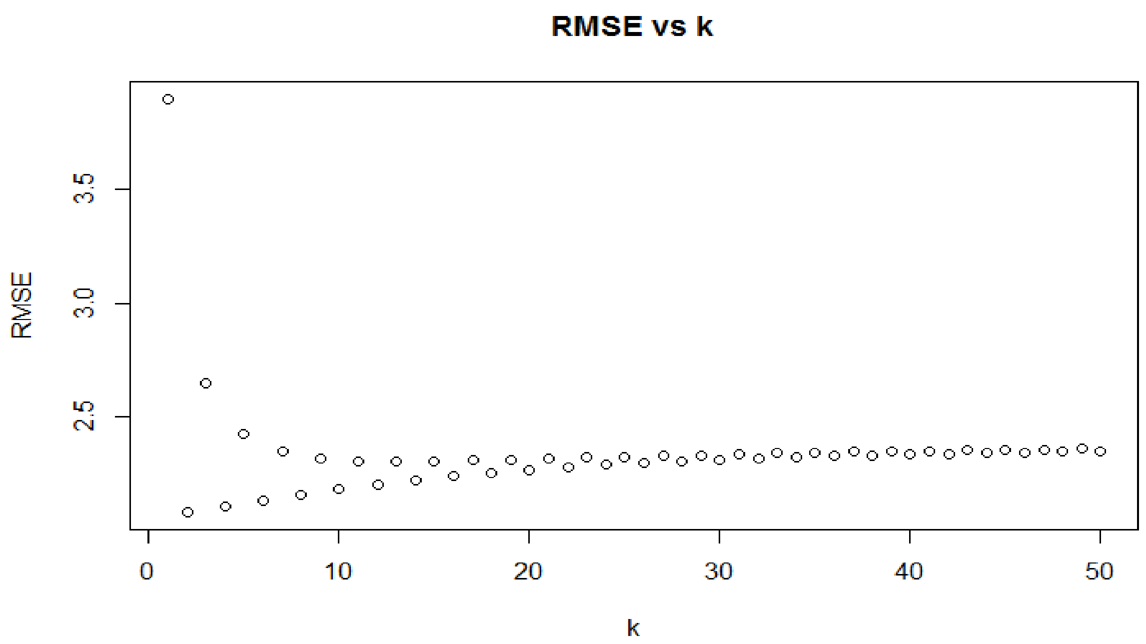
- **Here we can see that after performing the differencing operation our time series visually appears to be stationary and we can confirm that by performing the following test.**

```
	Augmented Dickey-Fuller Test

data:  x
Dickey-Fuller = -15.175, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

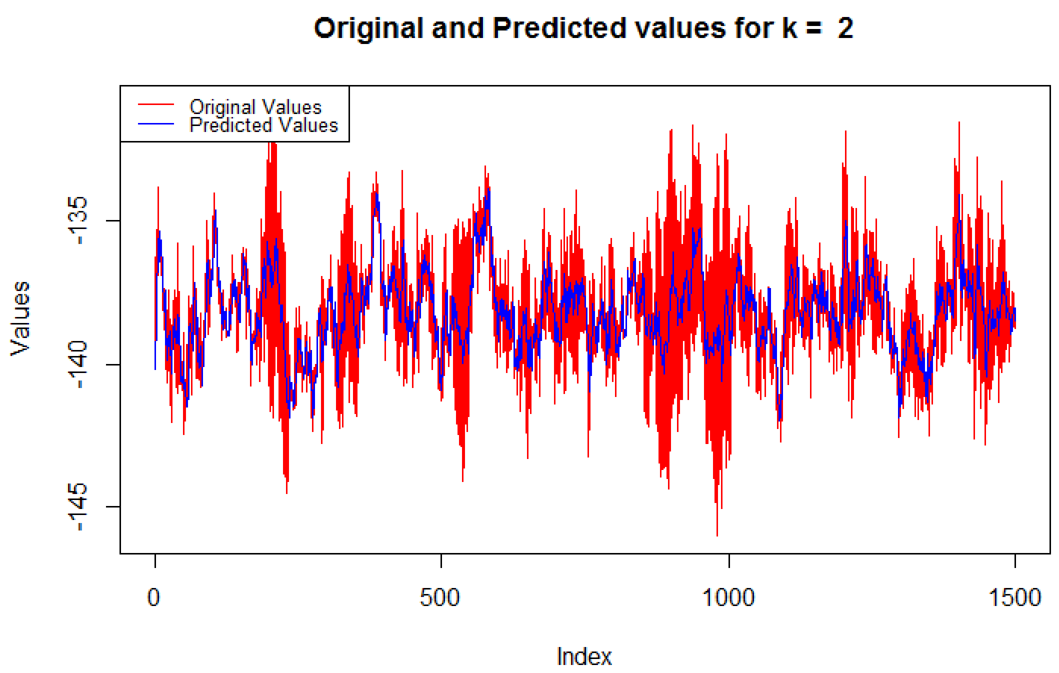
- **Here the value of P is very small so we reject the null hypothesis and accept our alternative hypothesis which proves that our Time Series is stationary.**

Task 2. Fit a simple moving average model (use the training set)

- Here we vary the value of k from 1 to 50 and plot the values of k vs RMSE

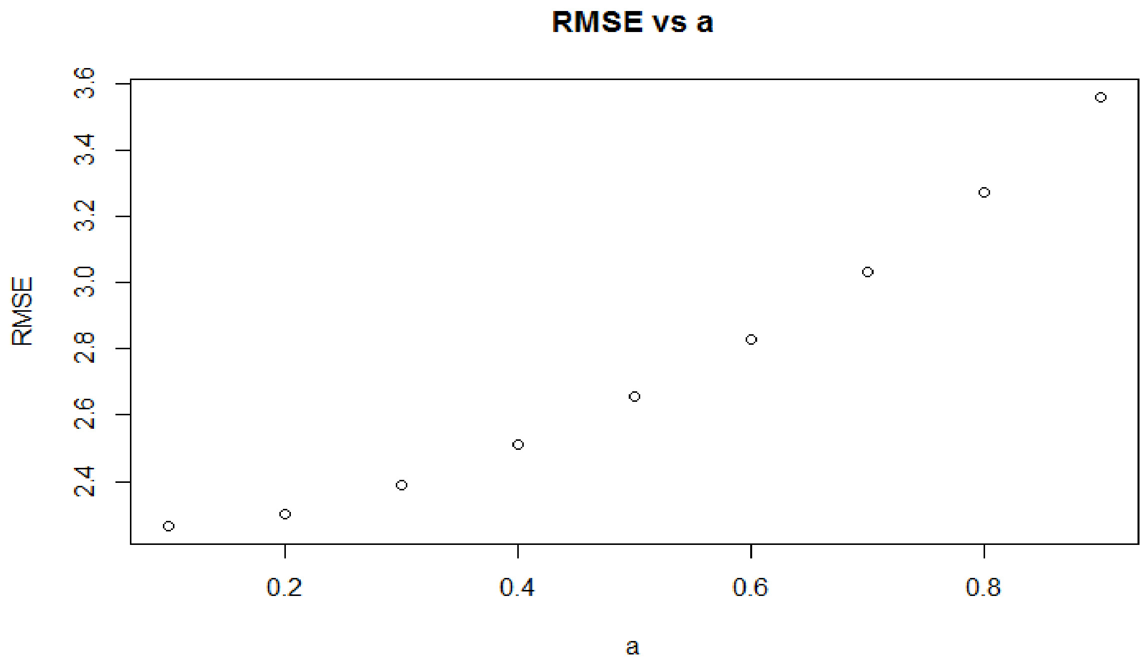


- **Here from the above plot we see that at k = 2 we get the RMSE as the lowest. So we select k = 2 and plot the graph of predicted vs original values.**

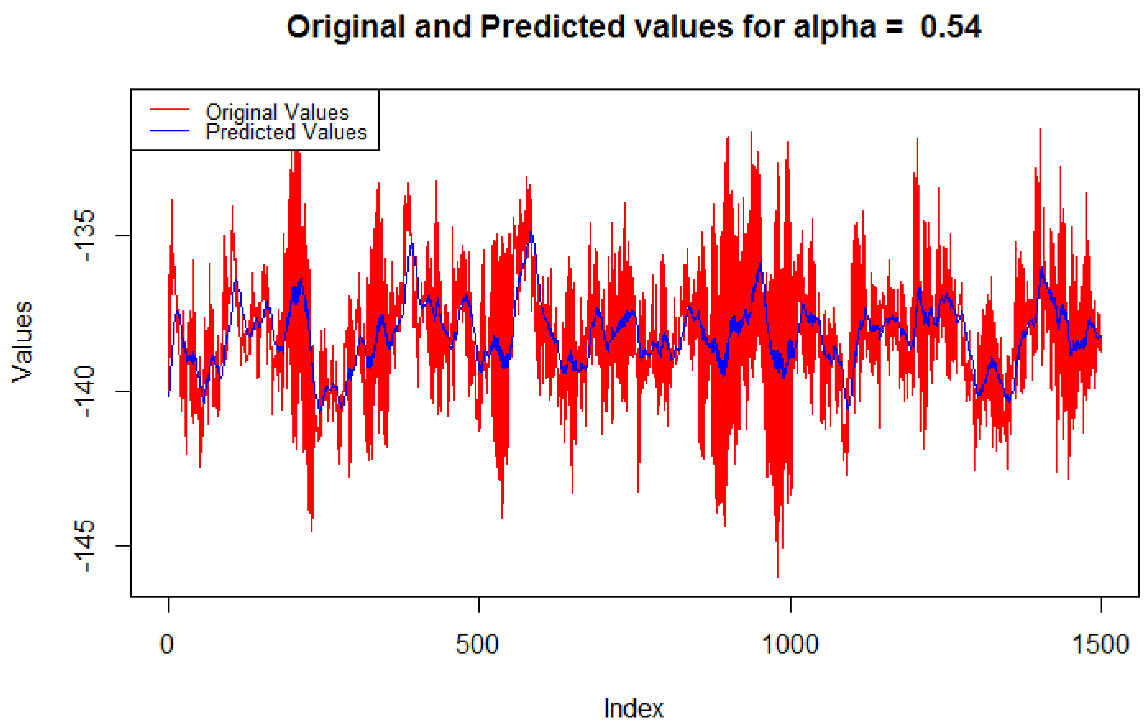


- **From the above plot we can see a decent fit of our values and now we shall fit other models on our data and analyse the performance**

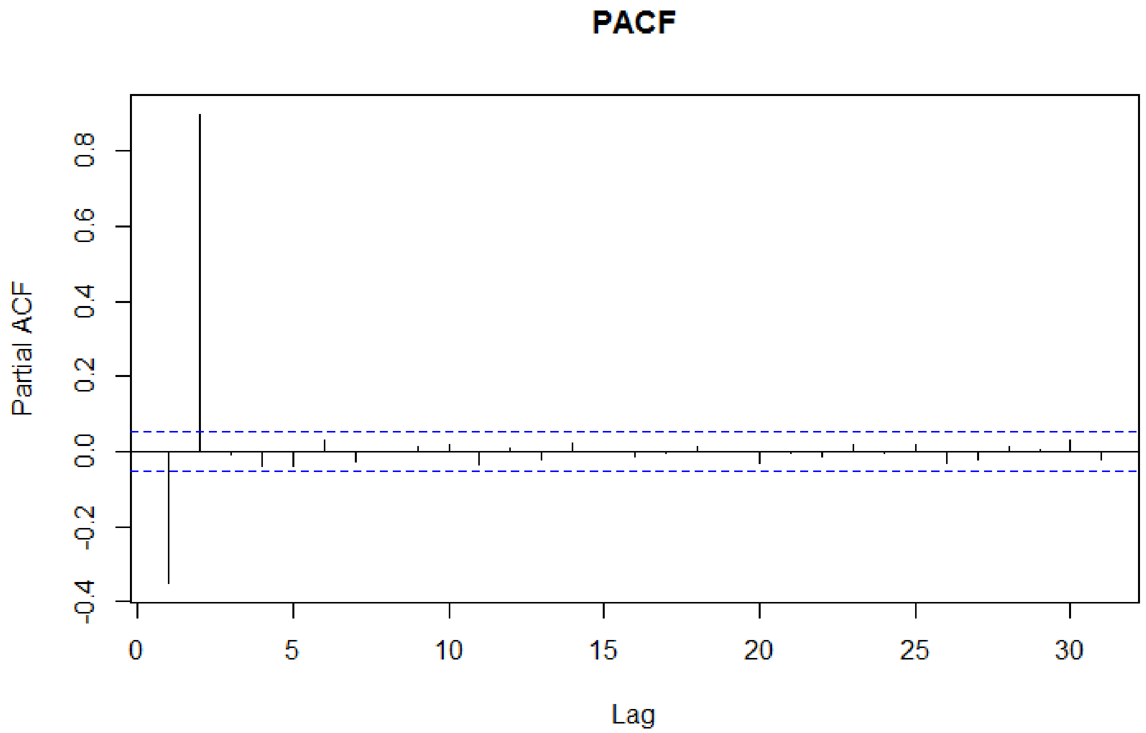
Task 3. Fit an exponential smoothing model (use the training set)



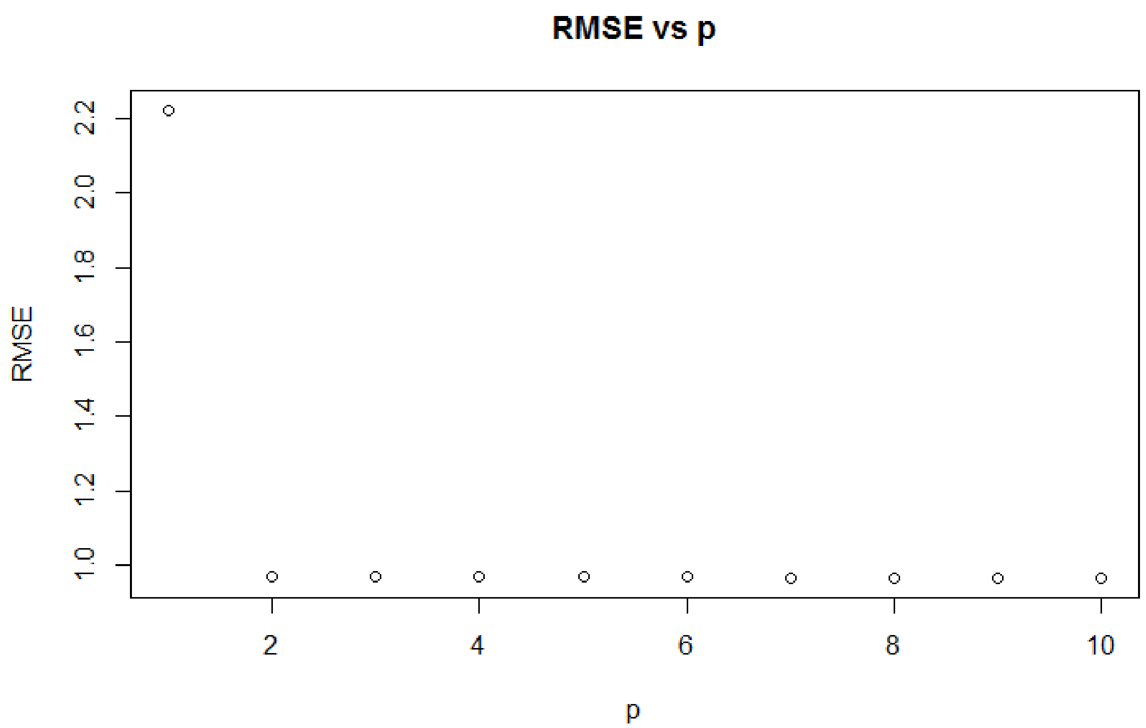
- From the above plot we can that for the value of a = 0.1 we get the value of RMSE as the lowest. So we shall use a = 0.1 and plot the graph of original vs predicted values.



<u>Task 4. Fit an AR(p) model (use the training set)</u>



- ❖ **From the above plot we can see that the lag gets close to zero at lag = 3 so we consider the value just before it which is 2 as our order.**

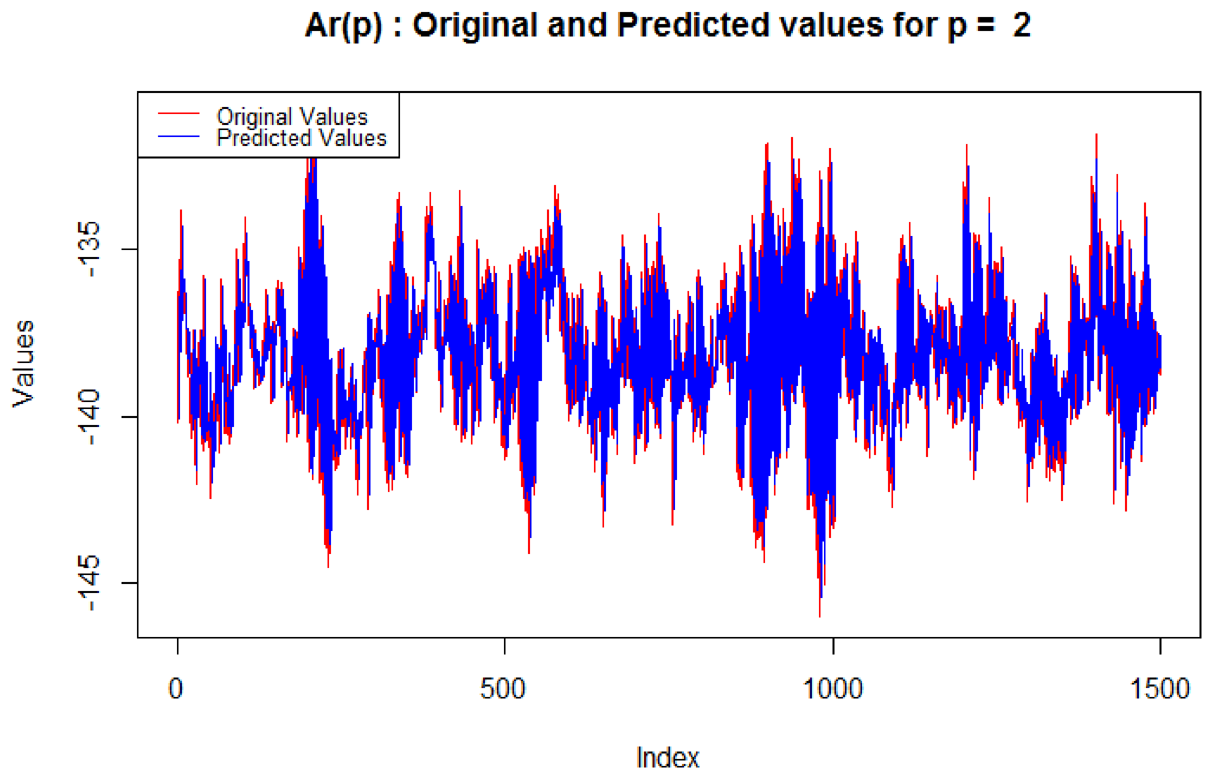


- **From the above plot it is clearly visible that the RMSE is the least for p = 2 , hence we select p = 2 and plot the graph of original vs predicted values for the AR(2) model.**

  Parameters for our model:

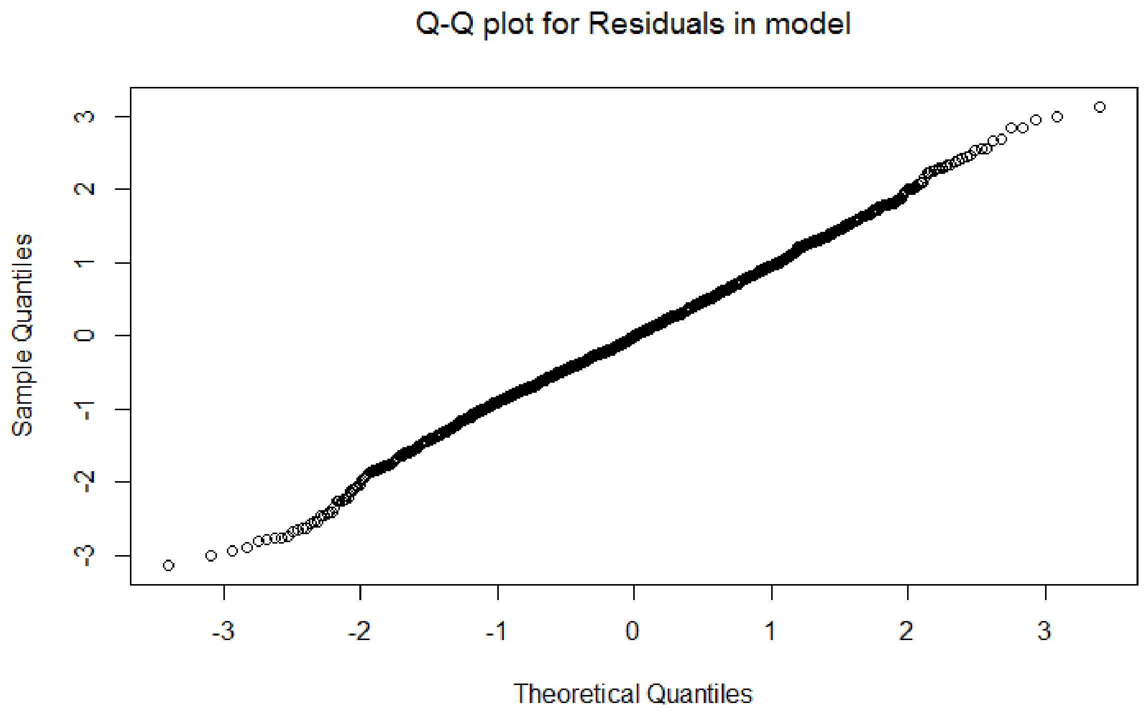
```
Call:
ar(x = train, aic = FALSE, order.max = 2)

Coefficients:
      1        2
-0.0355   0.8987
```

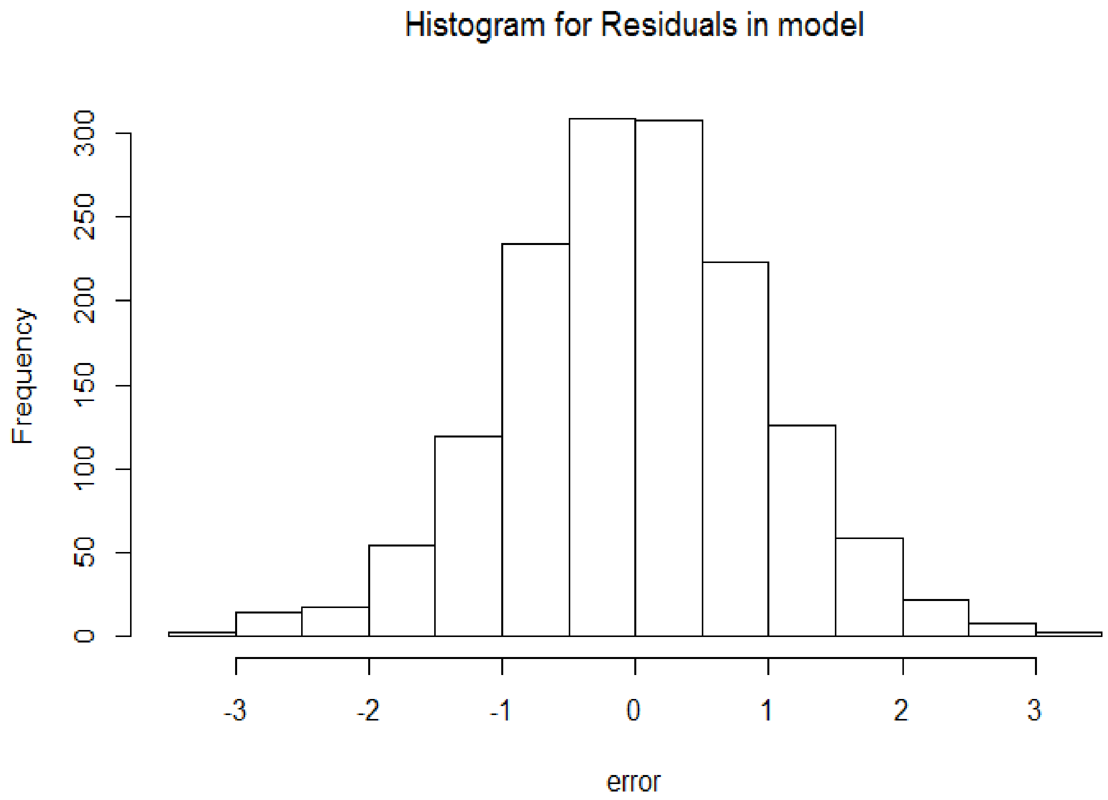
- **Plot of Original vs Predicted values for p = 2**



- **From the above plot we can visually determine that using the AR(p) model we get the best fit for our original vs predicted values in the training data set among the above three tasks.**

- **Q-Q Plot of the pdf of residuals.**

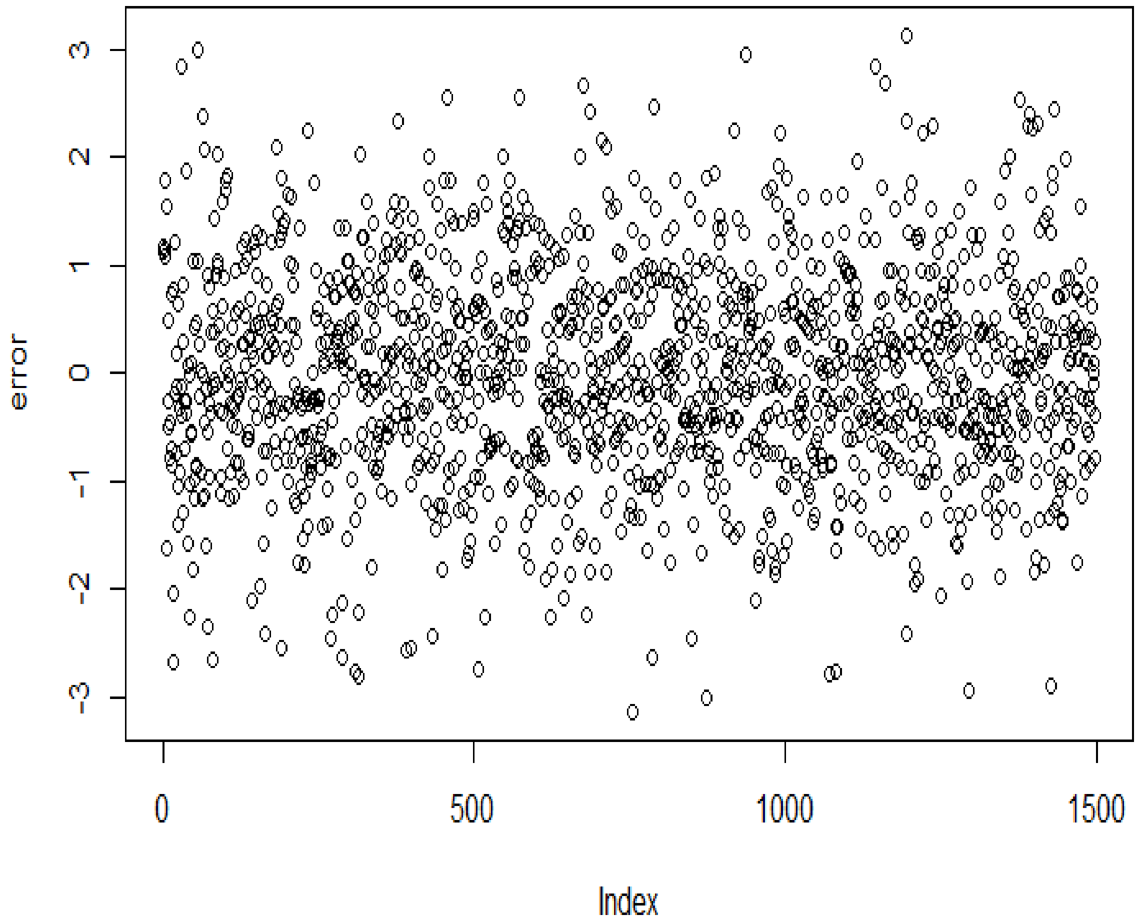




- **Performing the chi-square test and here we see that the p-value is larger than 0.1 and so we can say that we reject the null hypothesis and can say that the distribution is not normal.**

```
Pearson chi-square normality test
p-value = 0.4331
```

- **Scatter Plot of the Residuals.**

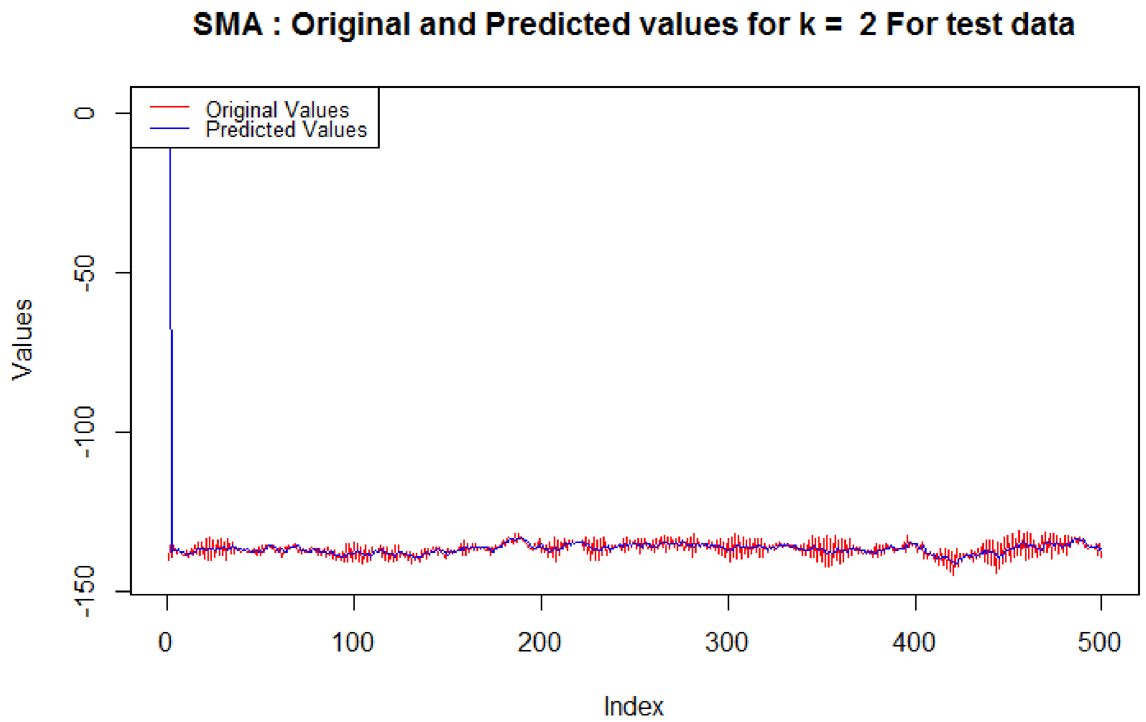


- **From the above scatter plot it is clearly evident that there are no correlation trends among the residuals**.

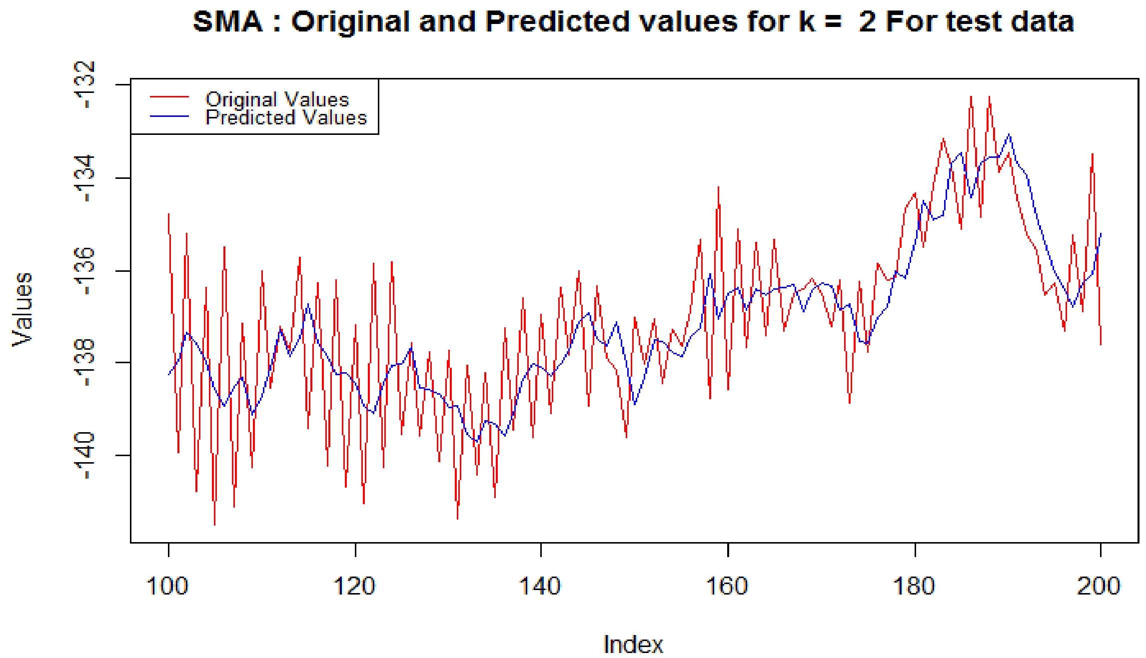
<u>**Task 5. Comparison of all the models (use the testing set)**</u>

1.) Simple Moving Average

<u>**Plot for the Test Data**</u>

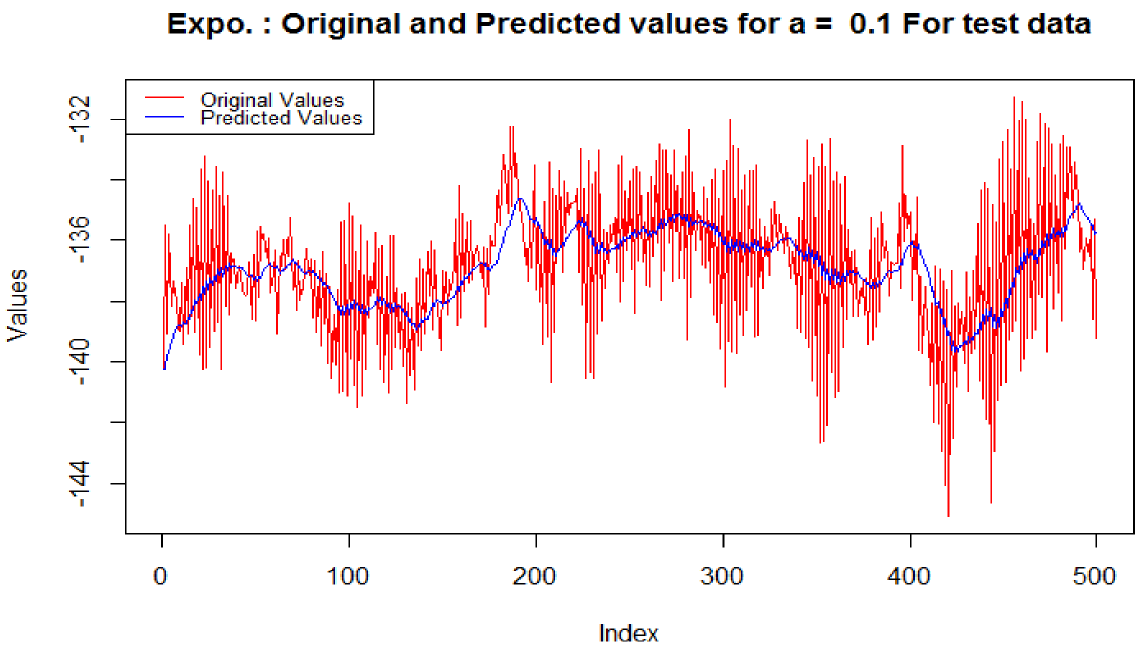


<u>**Plot for (100-200) rows of the test data to view plot more clearly**</u>

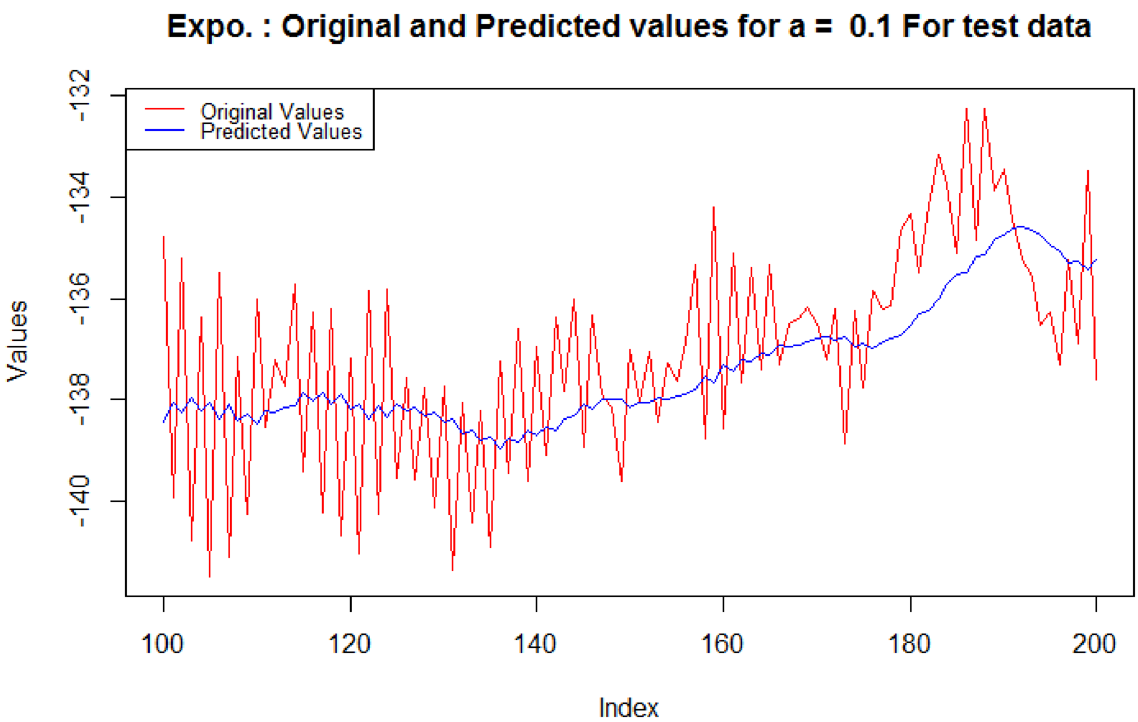


2.) Exponential Smoothing Model

❖ **Plot for the Test Data**

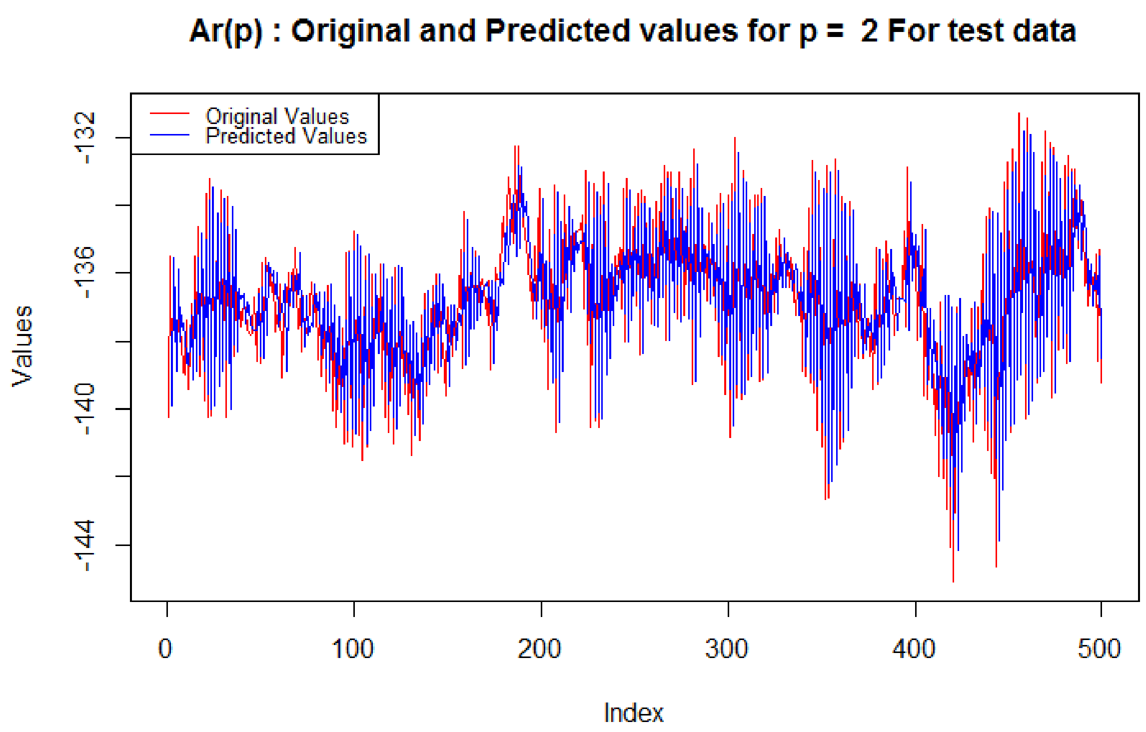


❖ **Plot for (100-200) rows of the test data to view plot more clearly**

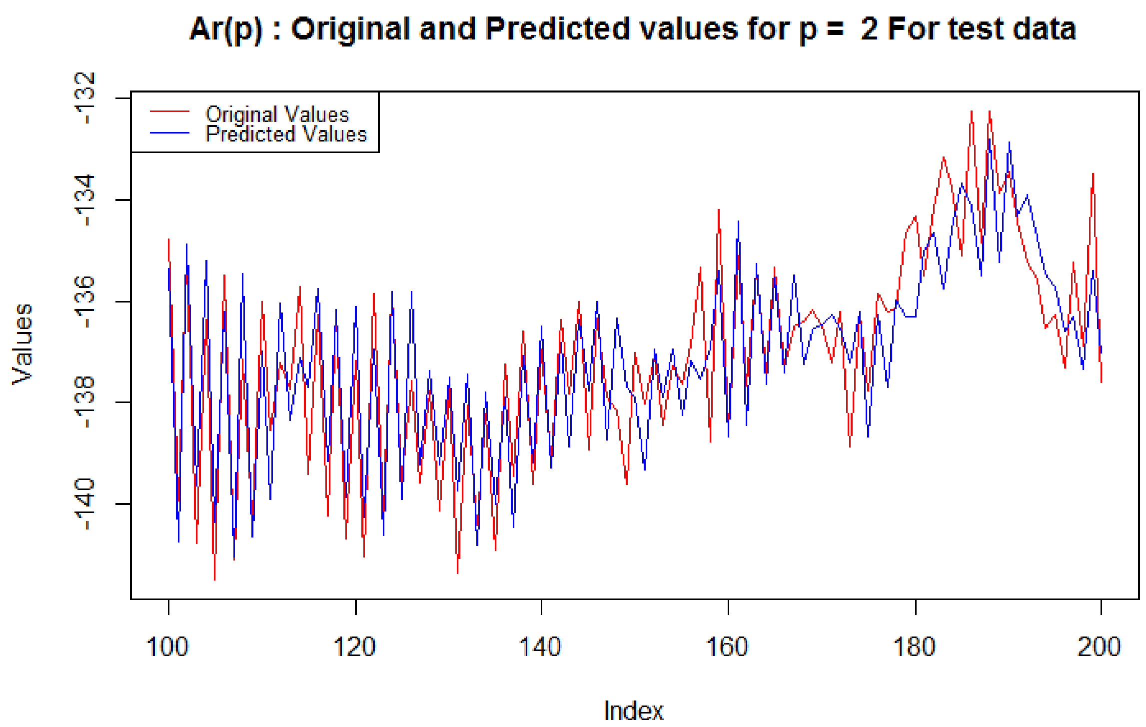


3.) AR(p) Model with P = 2 for test data

❖ **Plot for the Test Data**



❖ **Plot for (100-200) rows of the test data to view plot more clearly**



- **From all the above plots on test data we can say the following.**

- **RMSE for all the three models are as follows:**

  <u>Simple Moving Average Model</u>: 2.073583

  <u>Exponential Smoothing Model</u>: 2.266261

  <u>AR(P) Model</u> : 0.9723

- **The AR(p) model has the least RMSE and along with that we can visually see that the AR(p) model fits the best amongst the other models and so the AR(p) model is the best fit for our data set.**