IoT Analytics: Project 03

TASK 1

1.1

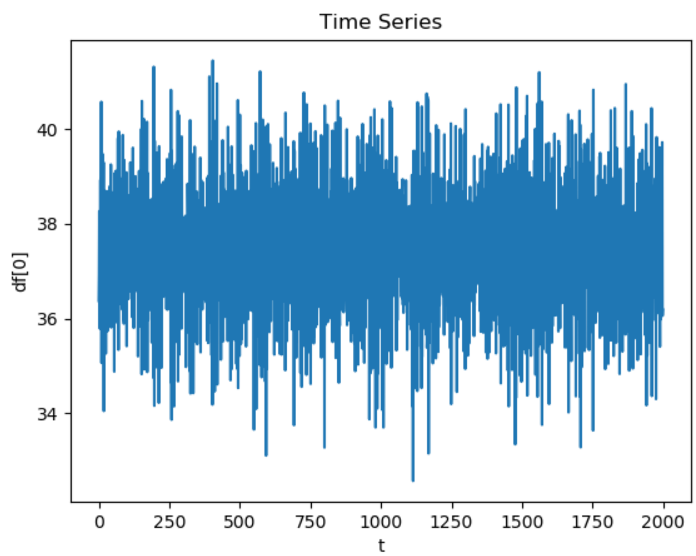

Figure 1: Time Series

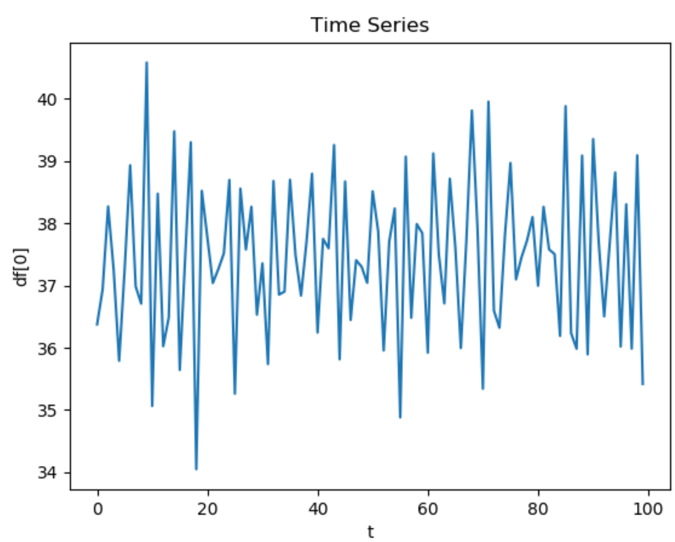

Figure 2: Time Series of first 100 values of the data frame

This (fig.1) is the plot of the entire time series for the given dataset ‘shdeshpa.csv’. To get a clearer view of the series, the first 100 values of the series are plotted to check for stationarity(fig.2). From both figures, it is apparent that the series appears visually stationary.

The transformations applied are logarithmic transformation, and first order differencing.

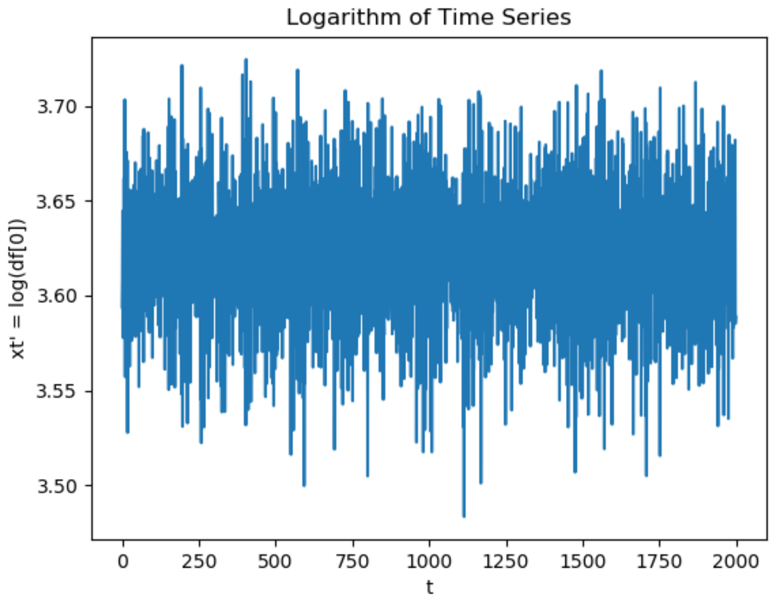

Figure 3: Logarithmic Transformation of Time Series

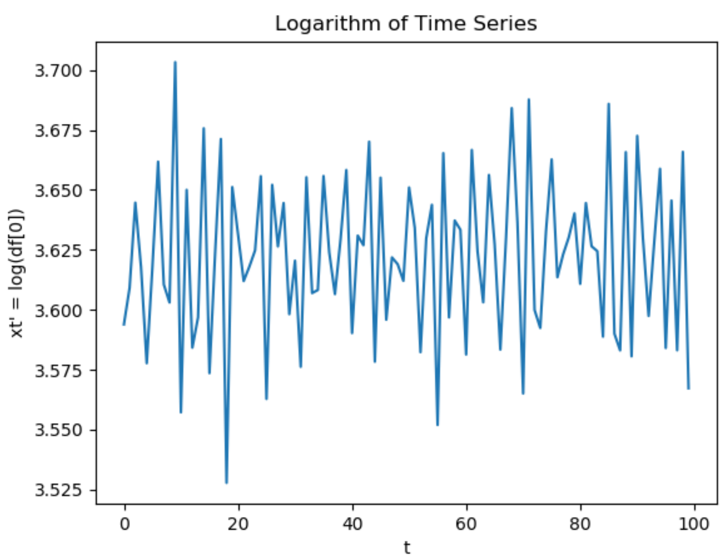

Figure 4: Logarithmic Transformation of Time Series of first 100 values

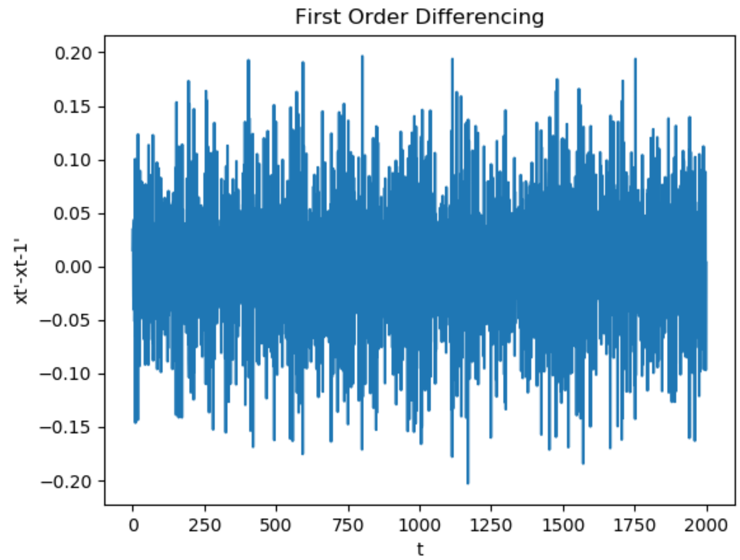

Figure 5: First Order Differenced Time Series

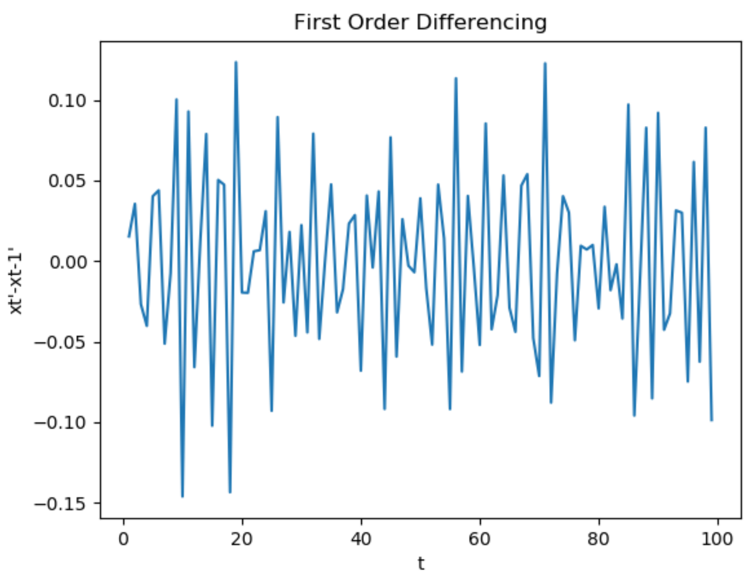

Figure 6: First Order Differenced Time Series of first 100 values

Logarithmic transformations are theoretically applied to remove variability. But this **dataset**, is already **stationary**. This means the dataset,

- Has no trends
- Has no variable variance
- Has no seasonality effect

Hence, applying logarithmic transformations and first order differencing doesn't make a difference in the plot.

TASK 2

2.1 The model is applied to df_train (training data of dataframe 'df').

2.2

```
error = mean_squared_error(df_train, predictions)
print('MSE: %.3f' % error)
rmse = np.sqrt(error)
print('RMSE:',rmse)
```

Figure 7: Code for calculation of MSE and RMSE

```
MSE: 1.887
RMSE: 1.3738476090886447
```

Figure 8: MSE and RMSE for k=1250

As shown in fig. 8, the RMSE and MSE values are the lowest obtained values for a given **'k',** i.e. k=1250. The lowest **RMSE Value= 1.3738.**

2.3

Given below, is a table of varying window sizes and it's corresponding MSE and RMSE values.

k	MSE	RMSE
1	6.289	2.5078
2	3.727	1.9304
3	2.541	1.5941
100	2.072	1.4393
1000	1.955	1.3983
1250	1.887	1.3738

Table 1: MSE, RMSE values for varying k

As seen from the table, the **lowest RMSE** value is for **k=1250, i.e., RMSE Value= 1.3738**.

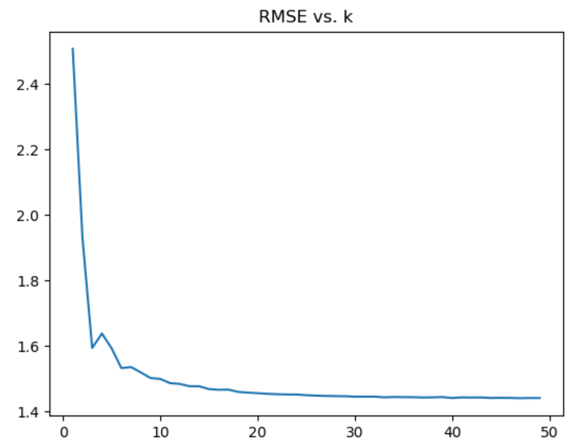

Figure 9: RMSE vs k

As seen from the figure above(fig.9), the value of RMSE decreases with increase in window size. The minimum value of k obtained from subsequent range iterations was at k=1250. Therefore k=1250 is used for obtaining further results.

2.4

For **k=1250,** the predicted values are plotted against the original values of df_train.

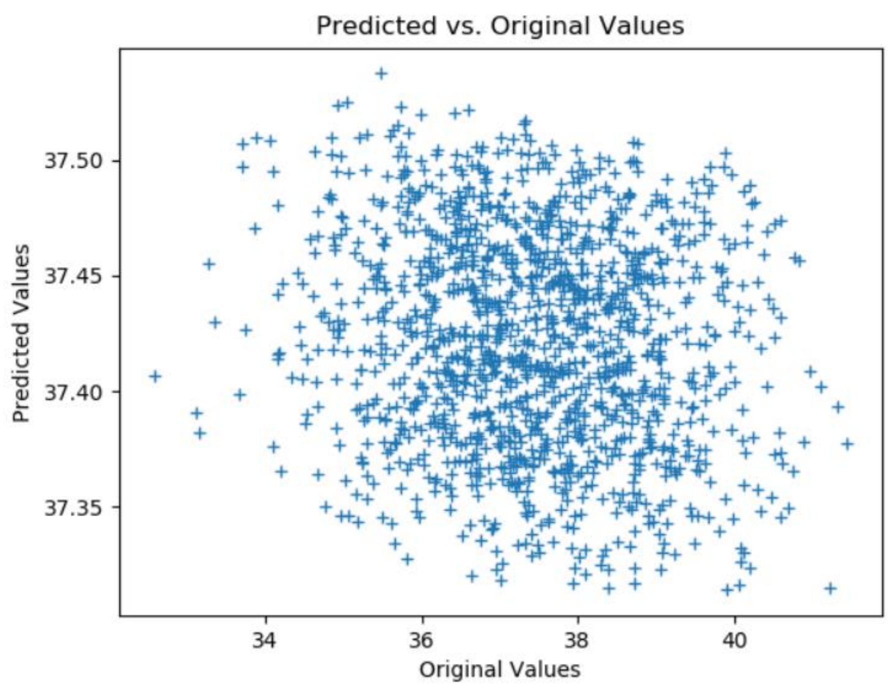

Figure 10: Predicted Values vs. Original Values for k=1250

RESULT ANALYSIS

From the Fig.10 it can be said that the values are correlated. Hence, model accuracy is good.

TASK 3

3.1

The exponential smoothing model is applied to the training dataset for **'a'** value ranging from **a=0.1** to **a=0.9**.

3.2

The RMSE for **'a'** value ranging from **a=0.1** to **a=0.9** is calculated along with MSE for the same **'a'** values.

```
a: 0.1
RMSE: 1.4409946456186549
MSE: 2.0764655687016327
```

Figure 11: RMSE and MSE Values for a=0.1

3.3

The step 3.2 of this task is repeated for **'a'** value ranging from **a=0.1** to **a=0.9** with step size of 0.1**.** The RMSE and MSE are calculated for this range of **'a'** values.

```
a: 0.1
RMSE: 1.4409946456186549
MSE: 2.0764655687016327
a: 0.2
RMSE: 1.5152688316217082
MSE: 2.2960396320842165
a: 0.30000000000000004
RMSE: 1.5959362633329768
MSE: 2.547012556621224З
a: 0.4
RMSE: 1.683769219418641
MSE: 2.8350787842616603
a: 0.5
RMSE: 1.779701597038563
MSE: 3.1673377745016116
a: 0.6
RMSE: 1.8846838354664317
MSE: 3.5520331596684596
a: 0.7000000000000001
RMSE: 1.9996904460022038
MSE: 3.998761879832493
a: 0.8
RMSE: 2.1257161540612275
MSE: 4.518669167636856
a: 0.9
RMSE: 2.2637794828149227
MSE: 5.124697546813798
```

Figure 12: RMSE and MSE values for range of ‘a’ values

As seen from Fig.12, the RMSE value is lowest for **a=0.1.** The value of RMSE for **a=0.1 is RMSE = 1.440994.**

3.4

The plot of RMSE vs. a is plotted and is seen in Fig. 13. As seen from Fig.13 it is observed that, RMSE is the lowest for a=0.1 and increases as value of ‘a’ increases.

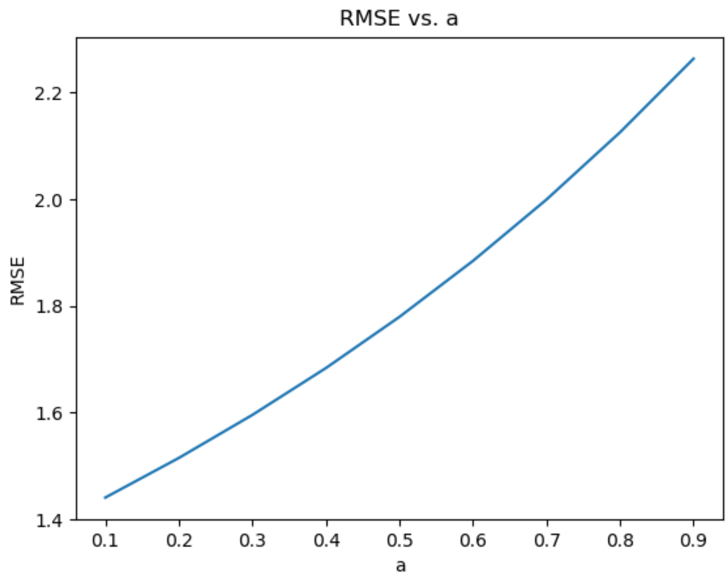

Figure 13: RMSE vs. a for Exponential Smoothing Model

3.5

The lowest value of RMSE is obtained for **a=0.1.**

Therefore, the plot of Predicted Values vs. Original Values is obtained by keeping value of a=0.1. This plot can be seen in Fig.14.

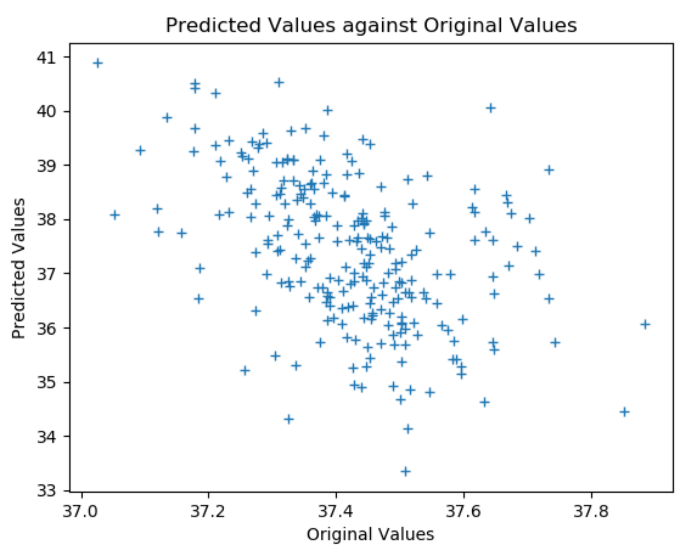

Figure 14: Predicted Values vs. Original Values for a=0.1

<u>RESULT ANALYSIS:</u>

The predicted values and original values (as seen from Fig.14) are correlated. Hence, the model accuracy is good. Stronger the correlation, better will be the accuracy of the model.

TASK 4

4.1

After PACF analysis, the lag value where the **PACF** chart crosses the upper confidence interval for the first time. If noticed closely, in this case, **p=1** of AR(p) model.

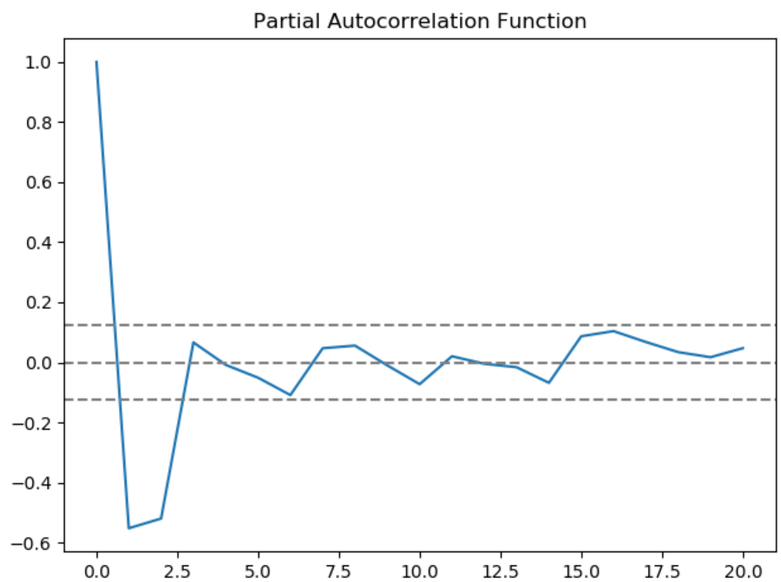

Figure 15: PACF

4.2

```
MSE: 1.399
RMSE: 1.1829900228061843
```

Figure 16: MSE and RMSE values for AR model

The RMSE value is 1.18299.

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      0   No. Observations:                 1500
Model:                     ARMA(1, 0)   Log Likelihood               -2380.539
Method:                       css-mle   S.D. of innovations              1.183
Date:                Sun, 11 Nov 2018   AIC                           4767.079
Time:                        21:28:31   BIC                           4783.018
Sample:                             0   HQIC                          4773.017

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         37.4203      0.020   1904.733      0.000      37.382      37.459
ar.L1.0       -0.5550      0.021    -25.841      0.000      -0.597      -0.513
                                    Roots
=============================================================================
                 Real           Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1          -1.8018           +0.0000j            1.8018            0.5000
-----------------------------------------------------------------------------
```

Figure 17: AR Model Summary

The parameters estimated are:

$$\delta = 37.4203$$

a1= -0.5550

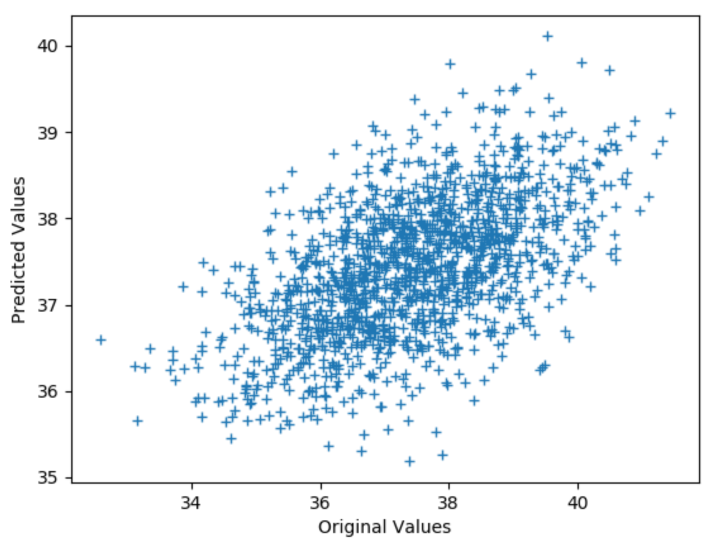

Figure 18: Predicted Value Against Original Value

<u>RESULT ANALYSIS:</u>

The predicted and original values are strongly correlated. Hence, the model accuracy is good as predicted values closely follow the original values.

4.3

a.

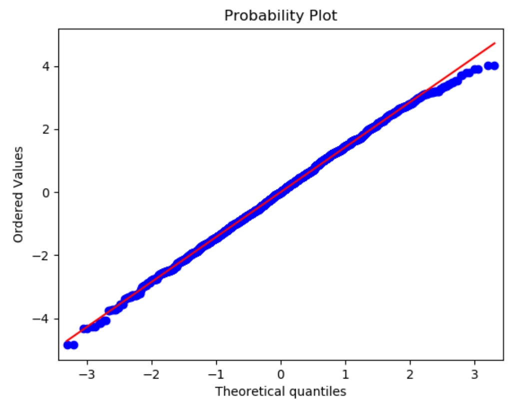

Figure 19: Q-Q Plot

As seen in fig.19, it is observed that in the Q-Q plot the distribution is following the standard normal distribution.

```
Chi Squared Test: NormaltestResult(statistic=1.8664568573512055, pvalue=0.3932819758157069)
```

Figure 20: Chi-Squared Test

From the **Chi-Squared Test** it's observed that **p-value=0.3932.** This signifies the following,

P-value > α: Fail to reject H_0
(α = 0.05 (assumption-for a one-tailed test))

The p-value observed in Fig.20 is larger than the significance level. Hence, this will cause failure to reject the null hypothesis. This means, Ho is accepted. Therefore, it can be said that residual closely follows the normal distribution curve.

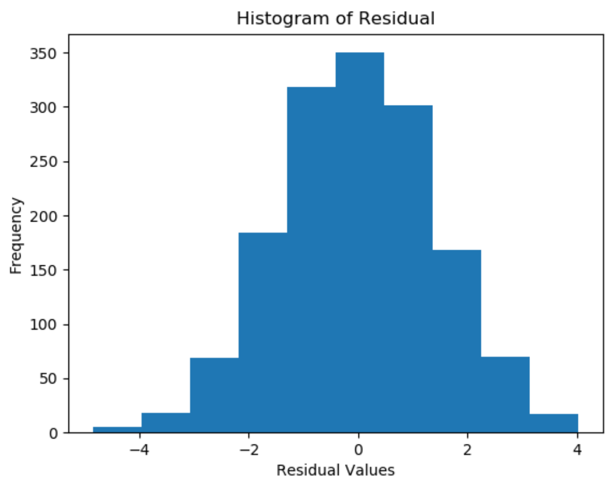

Figure 21: Histogram of Residuals

As seen in Fig.21, it can be said that the histogram has a standard normal distribution.

b.

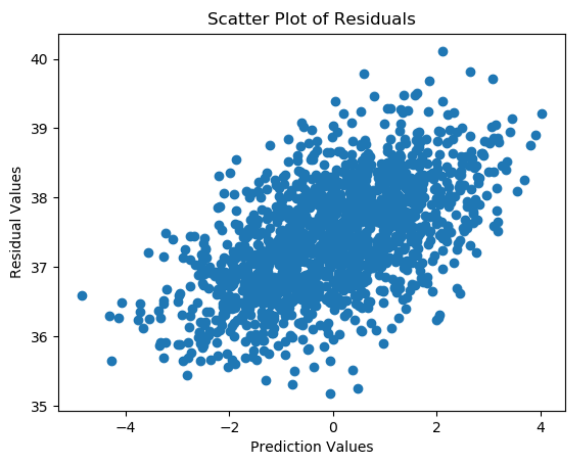

Figure 22: Scatter Plot of Residuals

As seen from fig.22, the residuals are not strongly correlated.

TASK 5

5.1 Simple Moving Average Model

The window size for minimum RMSE value of trained dataset is k=1250.

But for testing dataset, since there are only 500 observations, value of k=100 is found to have the lowest RMSE value.

MSE: 1.850

RMSE: 1.3601709558273363

Figure 23: RMSE and MSE value for k=100

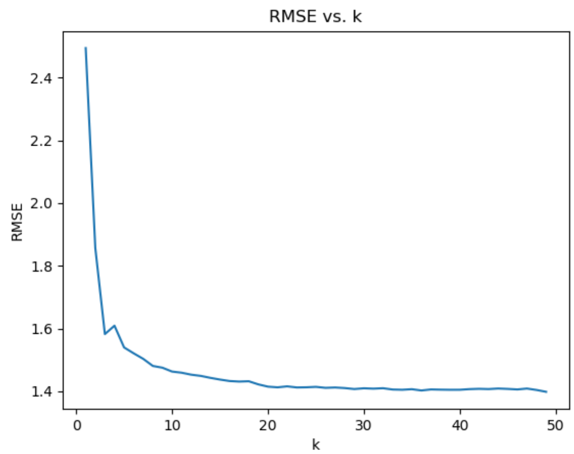

Figure 24: RMSE vs. k

As seen from the figure above(fig.24), the value of RMSE decreases with increase in window size. The minimum value of k obtained from subsequent range iterations was at k=100. Therefore k=100 is used for obtaining further results.

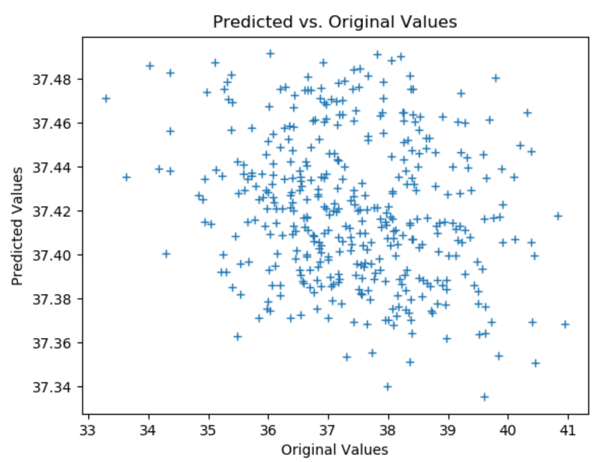

Figure 25: Predicted Values vs. Original Values for k=100

The predicted and original values are sparsely correlated. Hence, the model accuracy can be improved.

5.2 Exponential Smoothing Model

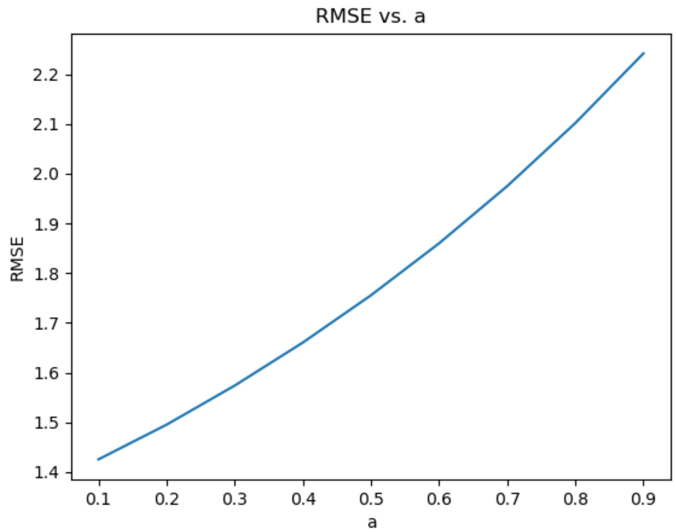

Figure 26: RMSE vs. a for test dataset

As seen from the RMSE vs. a curve(Fig.26), RMSE is least for a = 0.1. Shown in the figure below(Fig.27), it is seen that RMSE for a=0.1 is **RMSE=1.42566**.

```
a: 0.1
RMSE: 1.4256688353259352
MSE: 2.0325316280196084
a: 0.2
RMSE: 1.4954376872625694
MSE: 2.2363338764852223
a: 0.30000000000000004
RMSE: 1.5739878821580031
MSE: 2.477437853180236
a: 0.4
RMSE: 1.660472684986775
MSE: 2.75716953758719
a: 0.5
RMSE: 1.7554868668057815
MSE: 3.0817341395275792
a: 0.6
RMSE: 1.8599388932077614
MSE: 3.4593726864669123
a: 0.7000000000000001
RMSE: 1.9749065535589132
MSE: 3.9002558952899444
a: 0.8
RMSE: 2.1016457019762904
MSE: 4.416914656635414
a: 0.9
RMSE: 2.2416493412398357
MSE: 5.02499176908099
```

Figure 27: RMSE and MSE Values for ‘a’ values

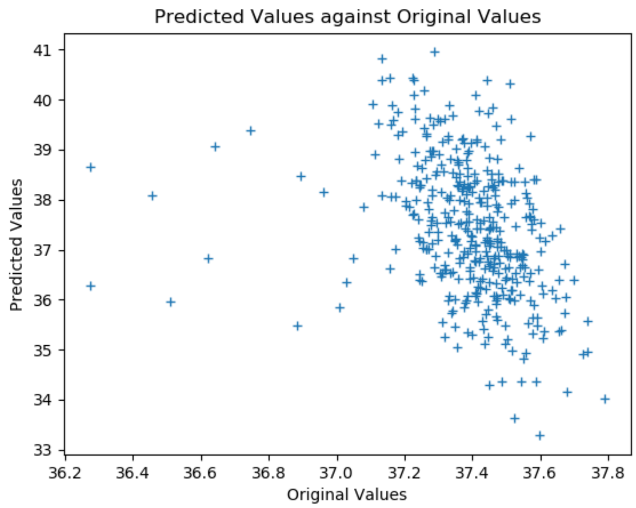

Figure 28: Predicted Values vs. Original Values for a=0.1

The values are strongly correlated. Hence, accuracy of the model is good.

5.3 AR Model

```
MSE: 1.269
RMSE: 1.126511070087771
```

Figure 29: MSE and RMSE Values for df_test of AR Model

The RMSE value of this model for test dataset is,

RMSE = 1.1265

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      0   No. Observations:                  500
Model:                     ARMA(1, 0)   Log Likelihood                -769.193
Method:                       css-mle   S.D. of innovations              1.126
Date:                Sun, 11 Nov 2018   AIC                           1544.386
Time:                        22:17:51   BIC                           1557.030
Sample:                             0   HQIC                          1549.348

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         37.4188      0.032   1180.983      0.000      37.357      37.481
ar.L1.0       -0.5910      0.036    -16.411      0.000      -0.662      -0.520
                                    Roots
=============================================================================
                 Real           Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1          -1.6919           +0.0000j            1.6919            0.5000
-----------------------------------------------------------------------------
```

Figure 30: Model Summary

The estimated parameters are (as seen in Fig.30),

$$\boldsymbol{\delta = 37.4188}$$

a1= -0.5910

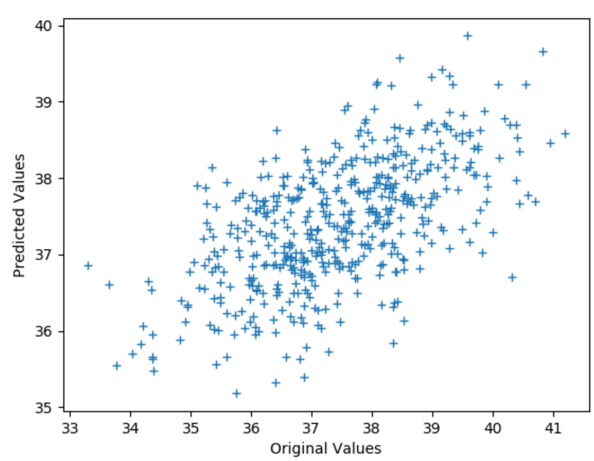

Figure 31: Predicted Values Against Original Values for AR Model Test Dataset

As seen in Fig.31, the predicted and original values are strongly correlated, hence, the model's accuracy is good.

CONCLUSION:

Out of all the models run on the “Test Dataset”, the **RMSE value** is the **least** for **AR Model.** Hence, **AR model** is the **best model** out of the three models for the test dataset.