Indian Institute of Technology Patna

CS-575
Applied Time Series Analysis
Mini Project
Report

Faculties In-charge:

Dr. Jimson Mathew Dr. Mayank Agarwal

Department of Computer Science and Engineering IIT Patna

Submitted by:
Durgesh Vikram Yadav
2011MT10

OBJECTIVE

Analysis of financial time series

To consider a time series stock values of various companies for a given duration and find an appropriate model for the stock prediction.

To estimate the parameters of the chosen time series model and compare at least with three different models.

METHODOLOGY AND MATHEMATICAL BACKGROUND

1. ARIMA Model

ARIMA, short for 'Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

Any 'non-seasonal' time series that exhibits patterns and is not a random white noise can be modelled with ARIMA models.

An ARIMA model is characterized by 3 terms: p, d, q

where,

p is the order of the AR term

q is the order of the MA term

d is the number of differencing required to make the time series stationary

A pure Auto Regressive (AR only) model is one where Yt depends only on its own lags. That is, Yt is a function of the 'lags of Yt'.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1$$

A pure Moving Average (MA only) model is one where Yt depends only on the lagged forecast errors.

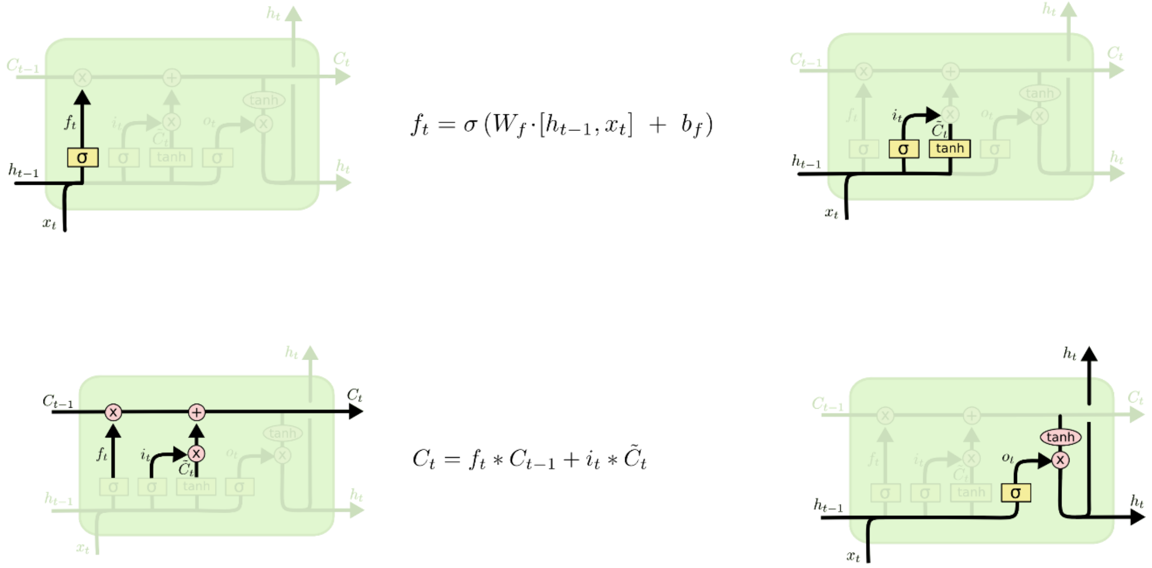
$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

An ARIMA model is one where the time series was differenced at least once to make it stationary and you combine the AR and the MA terms. So, the equation becomes:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

2. LSTM (Long short-term memory)

It is special kind of recurrent neural network that is capable of learning long term dependencies in data. This is achieved because the recurring module of the model has a combination of four layers interacting with each other.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Fig: LSTM Steps

3. <u>Triple Exponential Smoothening</u>

A simple method that assumes no systematic structure, an extension that explicitly handles trends, and the most advanced approach that add support for seasonality.
Triple Exponential Smoothing is an extension of Exponential Smoothing that explicitly adds support for seasonality to the univariate time series.
This method is sometimes called Holt-Winters Exponential Smoothing, named for two contributors to the method: Charles Holt and Peter Winters.

The basic equations for their method are given by:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad \text{OVERALL SMOOTHING}$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \qquad \text{TREND SMOOTHING}$$

$$I_t = \beta \frac{y_t}{S_t} + (1-\beta)I_{t-L} \qquad \text{SEASONAL SMOOTHING}$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \qquad \text{FORECAST ,}$$

where,

- y is the observation
- S is the smoothed observation
- b is the trend factor
- I is the seasonal index
- F is the forecast at *m* periods ahead
- t is an index denoting a time period

and α, β, and γ are constants that must be estimated in such a way that the MSE of the error is minimized.

4. <u>Prophet Model</u>

The Prophet library is an open-source library designed for making forecasts for univariate time series datasets. It is easy to use and designed to automatically find a good set of hyperparameters for the model in an effort to make skillful forecasts for data with trends and seasonal structure by default.

RESULTS AND ANALYSIS

1. <u>IBM Dataset</u>

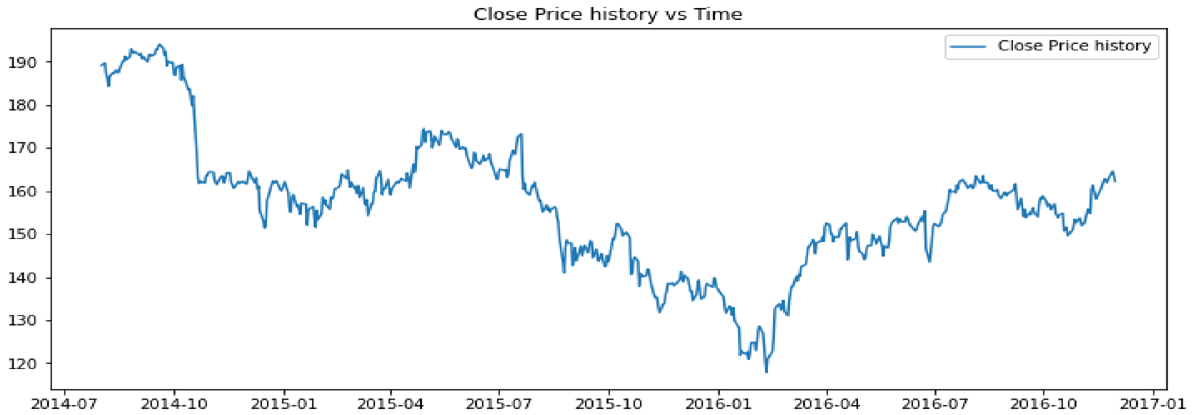


Fig: Close Price History vs Time plot

a) ADF and KPSS Test

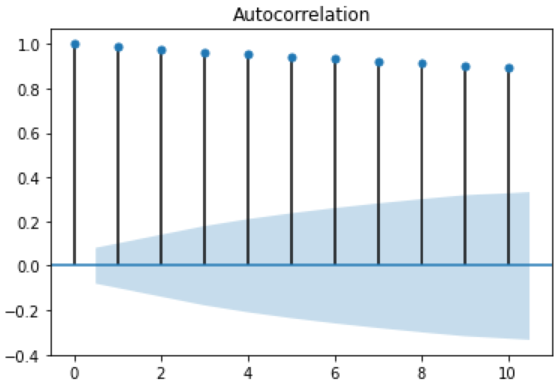
a. ADF Test: Since p value is greater than 0.05, therefore, we fail to reject the null hypothesis and the data has a unit root and is non-stationary.

b. KPSS Test: From KPSS Test, the test statistic is greater than the critical value, we reject the null hypothesis (series is not stationary).

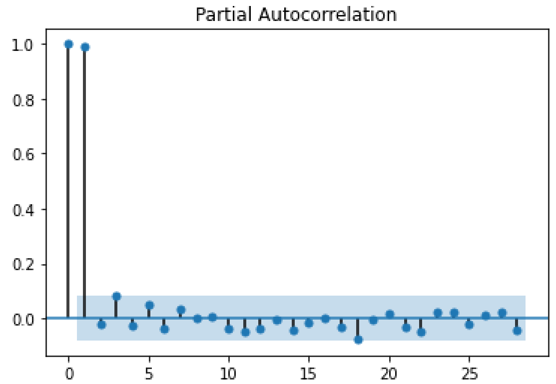
Both tests conclude that the series is not stationary -> series is not stationary

b) ACF and PACF Test

ACF Test



PACF Test



From ACF and PACF plots, since more than 5% of the plot is outside the shaded region, the data is non stationary.

c) *Apply ARIMA Model*

For p = 1, q = 1 and d = 0

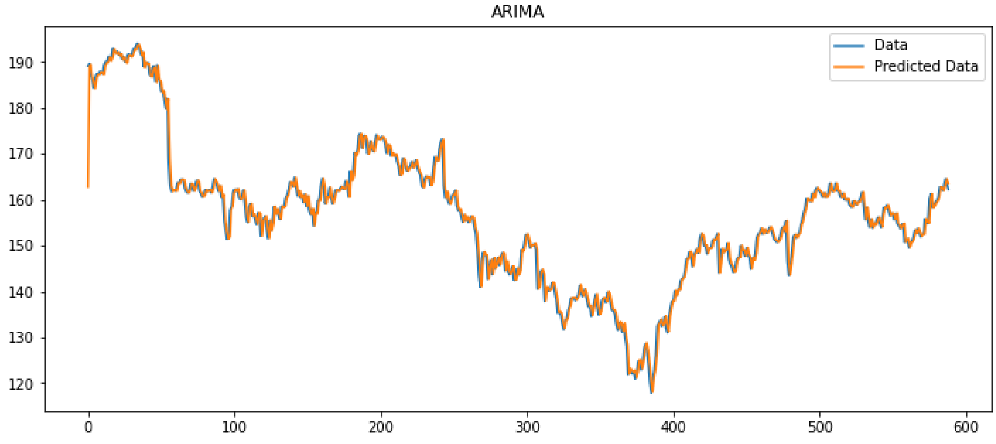


Fig: ARIMA Model plot

RMSE Value = 2.2488489672255034

d) *Apply LSTM Model*

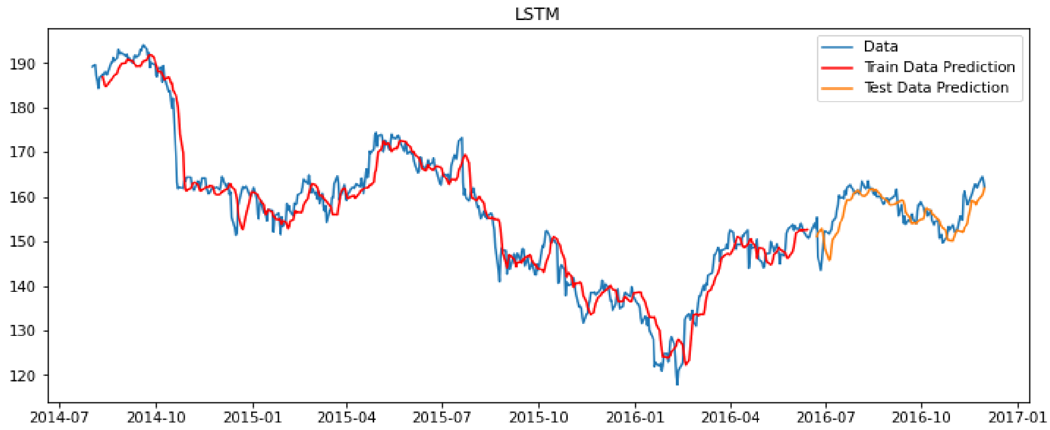


Fig: LSTM Model plot

RMSE Value of train data with LSTM model = 3.459834091805127
RMSE Value of test data with LSTM model = 2.8068576537297374

e) *Apply Triple Exponential Smoothening Model*

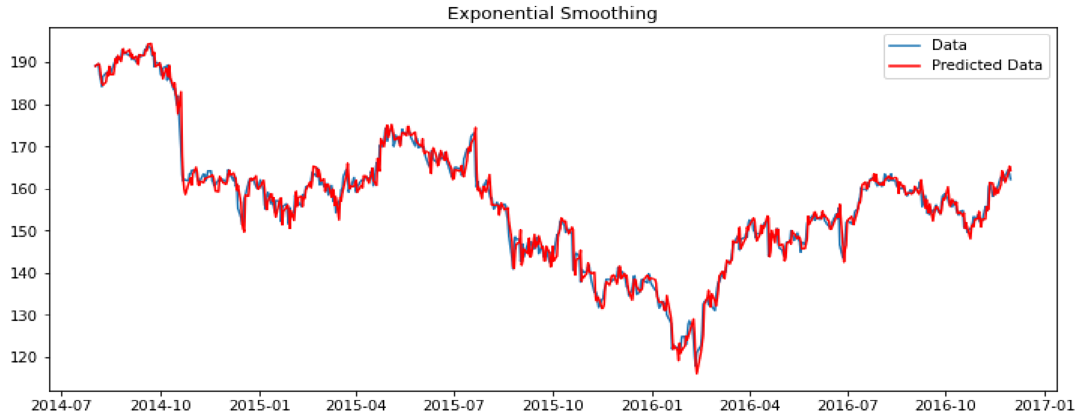


Fig: Triple Exponential Smoothening Model plot

RMSE value with Exponential Smoothing = 2.2503866549923073 *f) Apply Prophet Model*

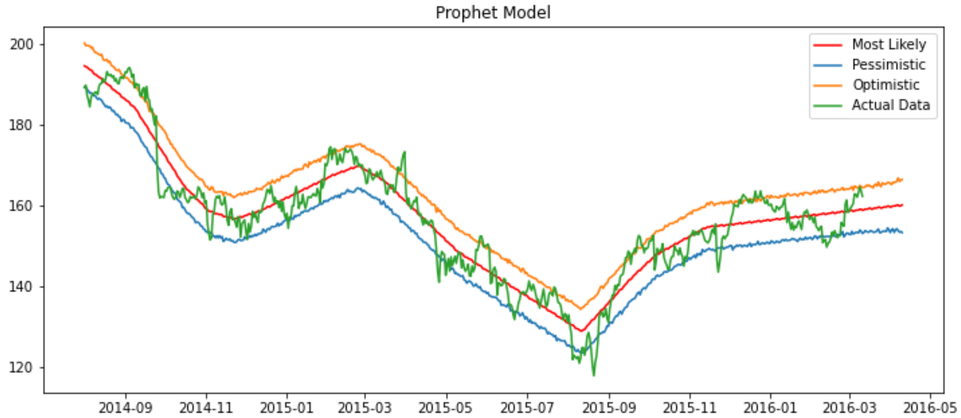


Fig: Prophet Model plot

Among the 4 models applied on IBM Dataset, LSTM Model gives higher RMSE Value. Also, it is evident from the plot that prediction for other models than LSTM are better, thus, we can use any of the ARIMA, Exponential Smoothening or Prophet Models.

Implementation file: Python (Google Colab)

2. <u>Apple Dataset</u>

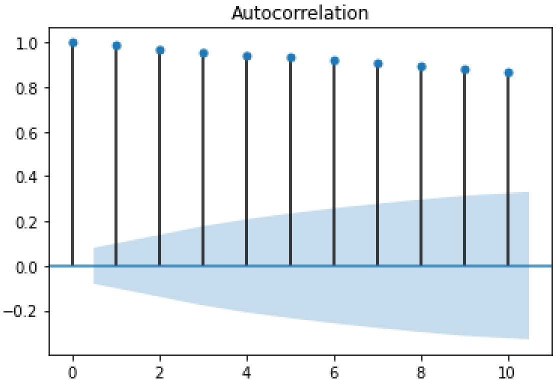
Fig: Close Price History vs Time plot

a) ADF and KPSS Test

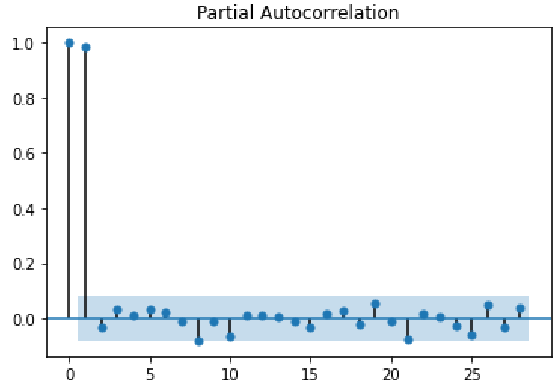
a. ADF Test: Since p value is greater than 0.05, therefore, we fail to reject the null hypothesis and the data has a unit root and is non-stationary.
b. KPSS Test: From KPSS Test, the test statistic is greater than the critical value, we reject the null hypothesis (series is not stationary).

Both tests conclude that the series is not stationary -> series is not stationary *b) ACF and PACF Test*

ACF Test



PACF Test



From ACF and PACF plots, since more than 5% of the plot is outside the shaded region, the data is non stationary.

c) Apply ARIMA Model

For p = 1, q = 1 and d = 0

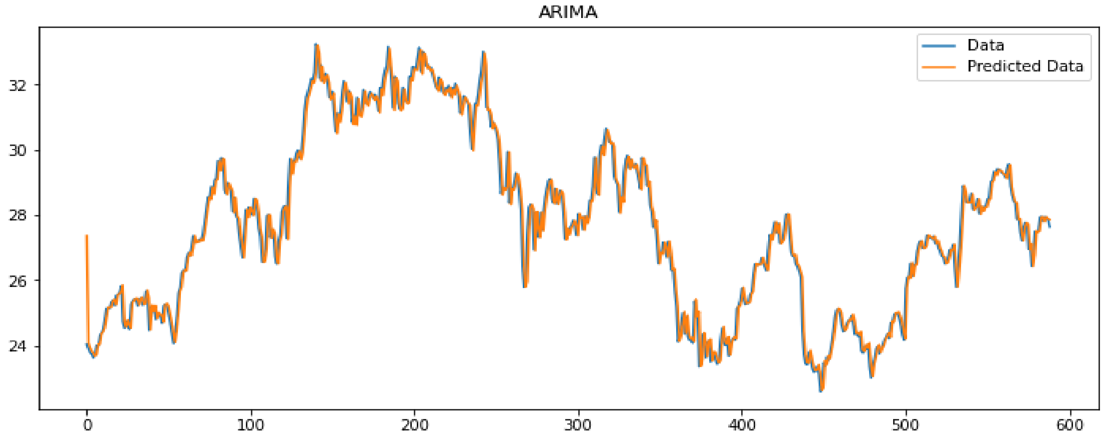


Fig: ARIMA Model plot

RMSE Value = 0.44705122244112044

d) Apply LSTM Model

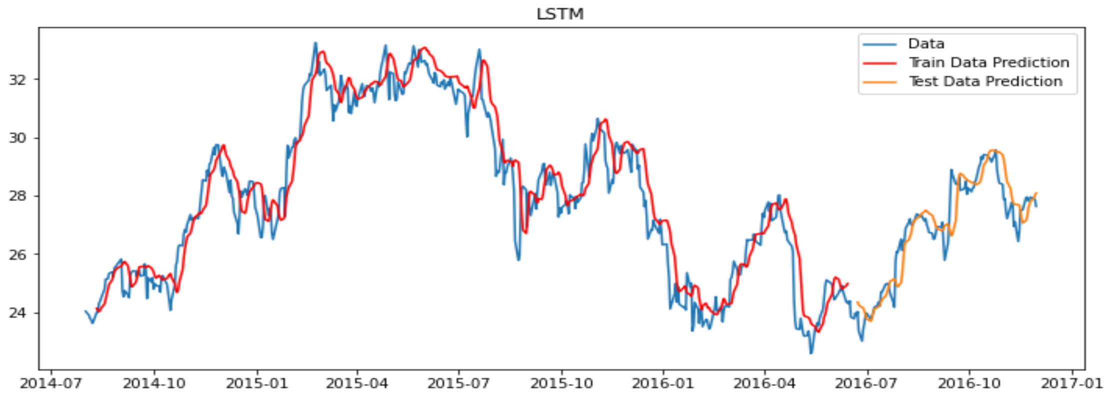


Fig: LSTM Model plot

RMSE Value of train data with LSTM model = 0.7202594350488373

RMSE Value of test data with LSTM model = 0.5758232838708677 *e) Apply Triple Exponential Smoothening Model*

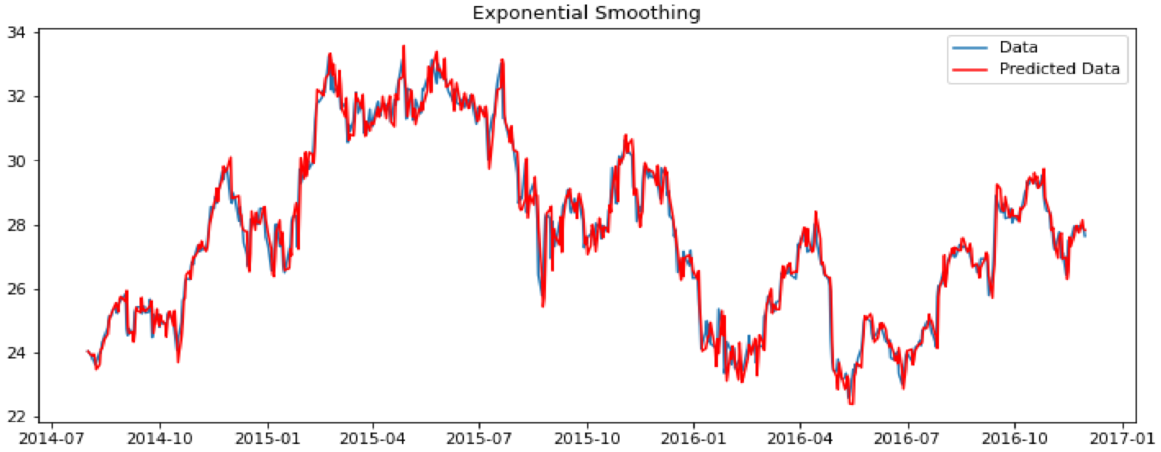


Fig: Triple Exponential Smoothening Model plot

RMSE value with Exponential Smoothing = 0.47501061625775814

f) Apply Prophet Model

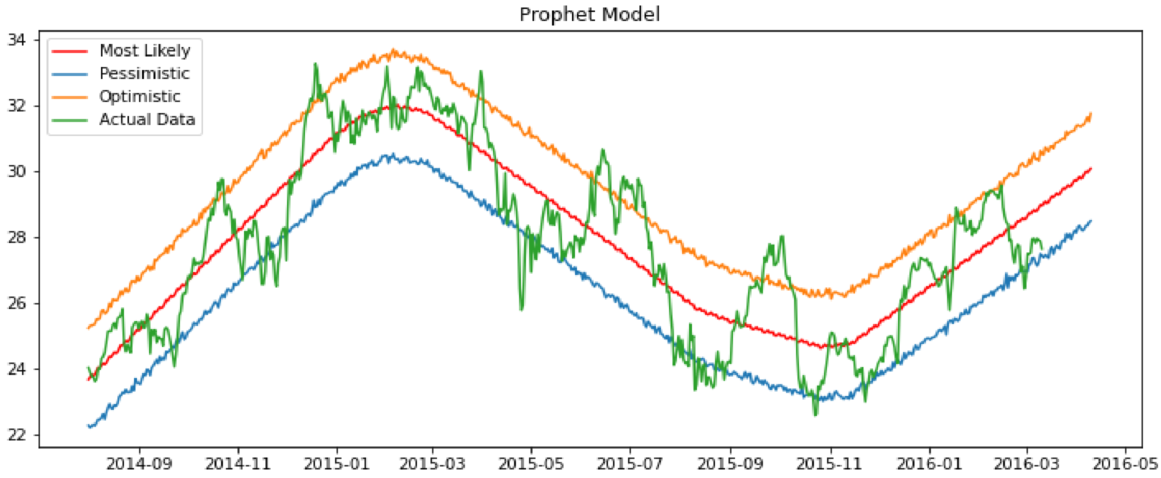


Fig: Prophet Model plot

Among the 4 models applied on IBM Dataset, LSTM Model gives higher RMSE Value. Also, it is evident from the plot that prediction for other models than LSTM are better, thus, we can use any of the ARIMA, Exponential Smoothening or Prophet Models.

Implementation file: Python (Google Colab)

3. <u>Facebook Dataset</u>

Fig: Close Price History vs Time plot

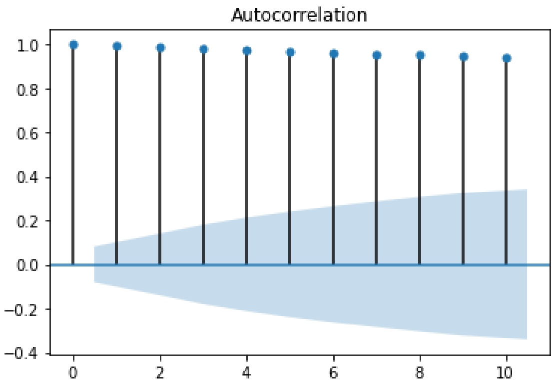
a) ADF and KPSS Test

a. ADF Test: Since p value is greater than 0.05, therefore, we fail to reject the null hypothesis and the data has a unit root and is non-stationary.
b. KPSS Test: From KPSS Test, the test statistic is greater than the critical value, we reject the null hypothesis (series is not stationary).

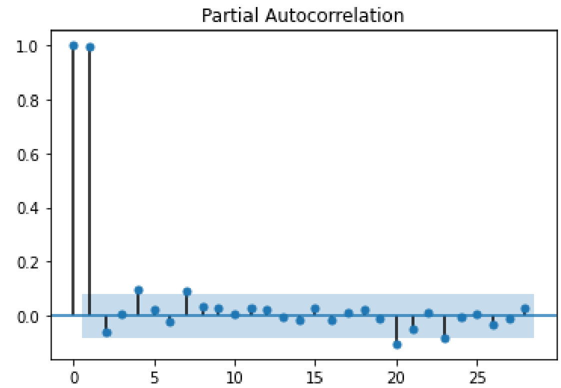
Both tests conclude that the series is not stationary -> series is not stationary

b) ACF and PACF Test

ACF Test



PACF Test



From ACF and PACF plots, since more than 5% of the plot is outside the shaded region, the data is non stationary.

c) Apply ARIMA Model

For p = 1, q = 1 and d = 0

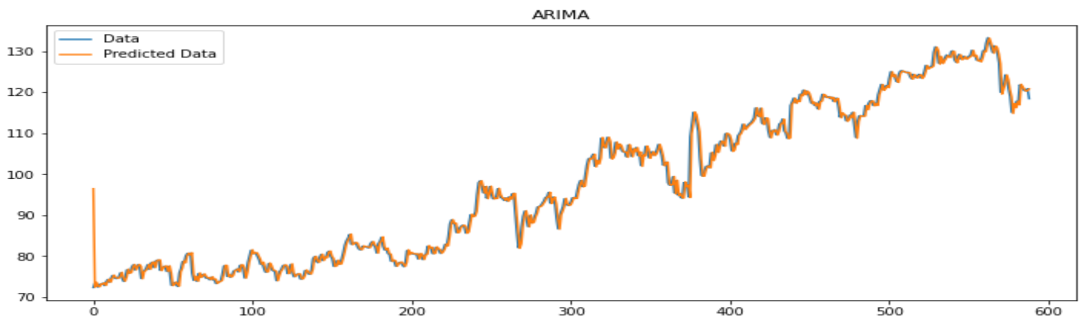


Fig: ARIMA Model plot

RMSE Value = 1.914996479972129

d) Apply LSTM Model

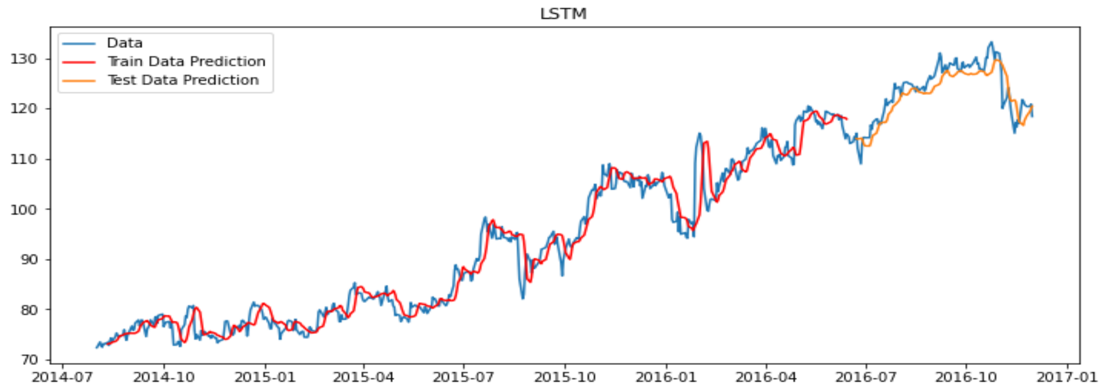


Fig: LSTM Model plot

RMSE Value of train data with LSTM model = 2.62398208238286
RMSE Value of test data with LSTM model = 2.4286688619873114

e) Apply Triple Exponential Smoothening Model

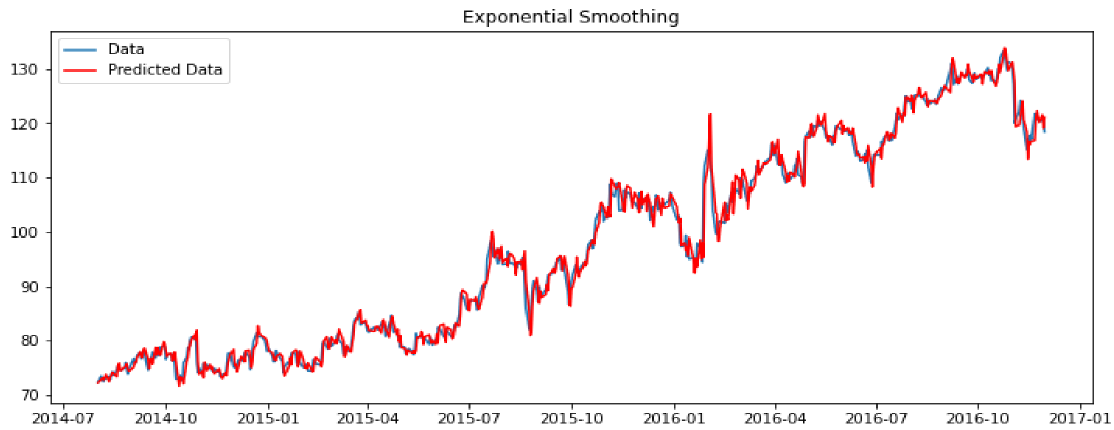


Fig: Triple Exponential Smoothening Model plot

RMSE value with Exponential Smoothing = 1.888335218694675 *f) Apply Prophet Model*

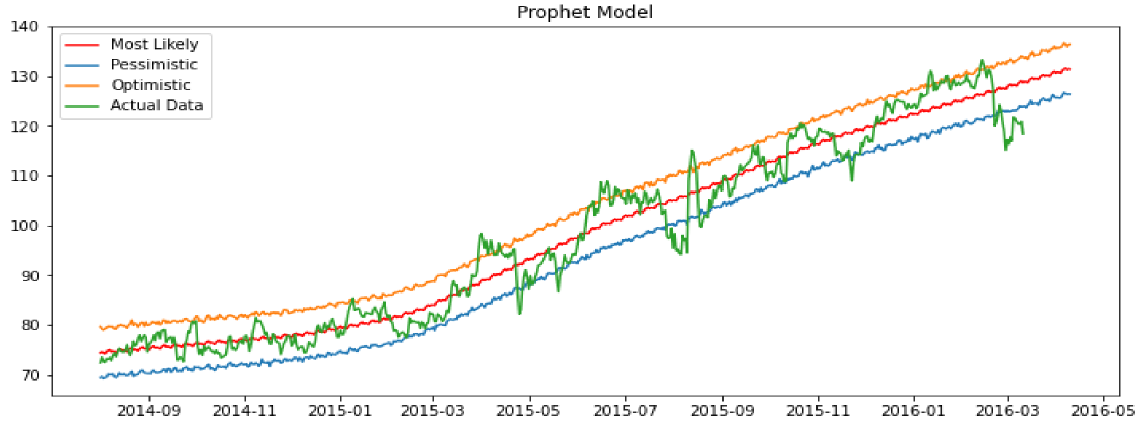


Fig: Prophet Model plot

Among the 4 models applied on IBM Dataset, LSTM Model gives higher RMSE Value. Also, it is evident from the plot that prediction for other models than LSTM are better, thus, we can use any of the ARIMA, Exponential Smoothening or Prophet Models.

Implementation file: Python (Google Colab)

4. Google Dataset

Fig: Close Price History vs Time plot

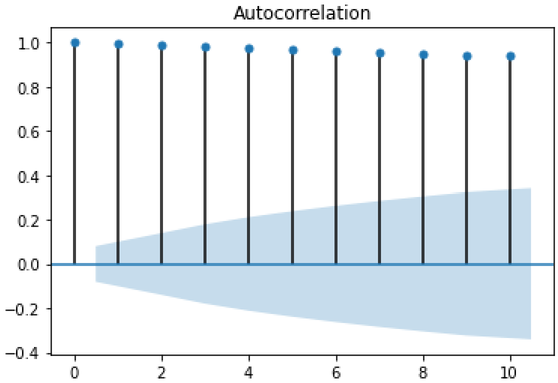
a) ADF and KPSS Test

a. ADF Test: Since p value is greater than 0.05, therefore, we fail to reject the null hypothesis and the data has a unit root and is non-stationary.
b. KPSS Test: From KPSS Test, the test statistic is greater than the critical value, we reject the null hypothesis (series is not stationary).

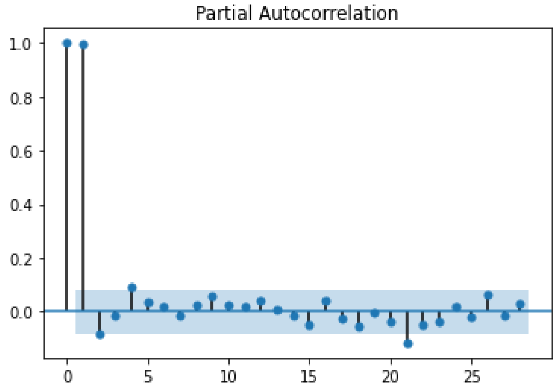
Both tests conclude that the series is not stationary -> series is not stationary

b) ACF and PACF Test

ACF Test



PACF Test



From ACF and PACF plots, since more than 5% of the plot is outside the shaded region, the data is non stationary.

c) Apply ARIMA Model

For p = 1, q = 1 and d = 0

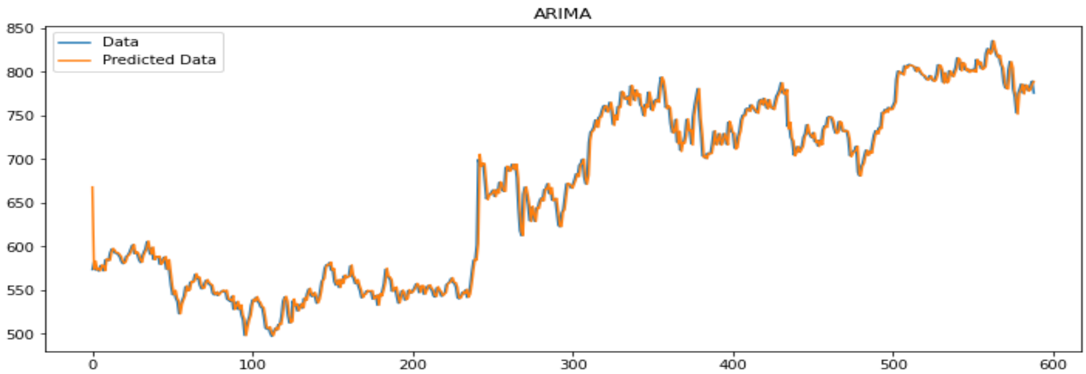


Fig: ARIMA Model plot

RMSE Value = 10.665839090335492

d) Apply LSTM Model

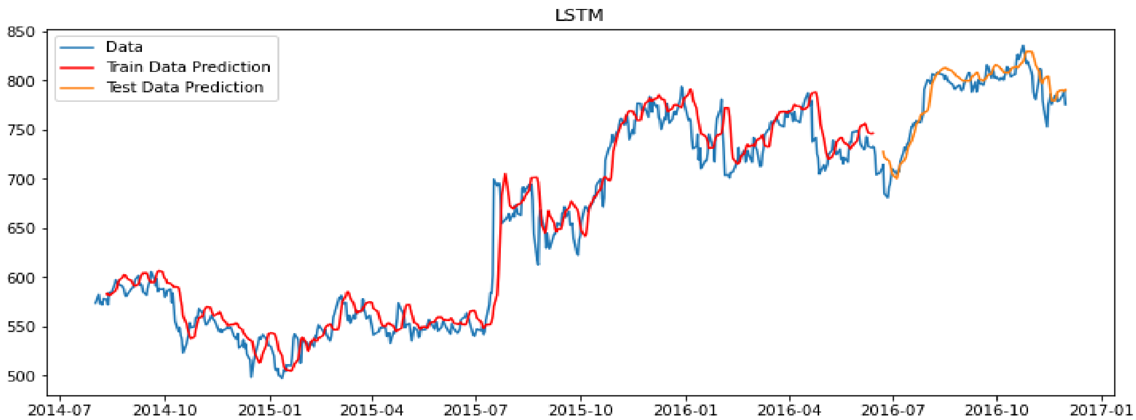


Fig: LSTM Model plot

RMSE Value of train data with LSTM model = 18.925341191607686
RMSE Value of test data with LSTM model = 13.286848770607543

e) Apply Triple Exponential Smoothening Model

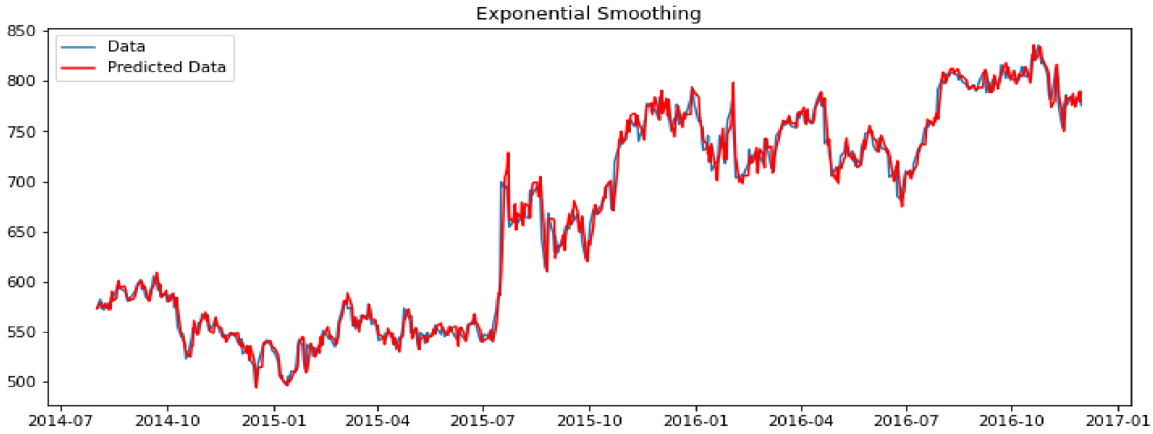


Fig: Triple Exponential Smoothening Model plot

RMSE value with Exponential Smoothing = 11.238657636531293

f) Apply Prophet Model

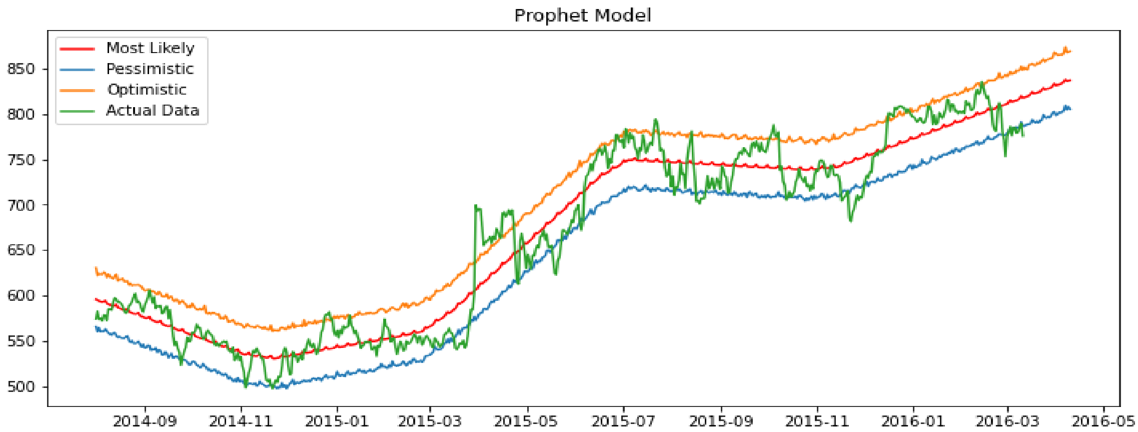


Fig: Prophet Model plot

Among the 4 models applied on IBM Dataset, LSTM Model gives higher RMSE Value. Also, it is evident from the plot that prediction for other models than LSTM are better, thus, we can use any of the ARIMA, Exponential Smoothening or Prophet Models.

Implementation file: Python (Google Colab)

CONCLUSION

Among the 4 models applied on IBM Dataset, LSTM Model gives higher RMSE Value. Also, it is evident from the plot that prediction for other models than LSTM are better, thus, we can use any of the ARIMA, Exponential Smoothening or Prophet Models.

All these datasets were non stationary. Thus, they can be made stationary with different methods like differencing to get better results.