FORECASTING USD-INR EXCHANGE RATE

Kiran.B

Kavya Ramesh

Tejaswini Halmare

Harishankar Giri

Pratik Trivedi

Sudhin Sudhakaran

BUSINESS OBJECTIVE

- **Data provided is related to USD/INR Exchange rates. The objective is to understand the underlying structure in the given dataset and come up with a suitable forecasting model which can effectively forecast USD/INR exchange rate for the next 30 days.**
- **This forecasting model will be used by exporting and importing companies to understand the currency movements and accordingly set their revenue expectations.**

DATA SET

Given dataset is in .csv format and it contains two columns having “Date” and “Rate” (Indian Rupee values). There are 12649 rows present in the dataset

[3]:

	Date	Rate
0	1973-01-02	8.02
1	1973-01-03	8.02
2	1973-01-04	8.00
3	1973-01-05	8.01
4	1973-01-08	8.00
...	...	...
12644	2021-06-21	74.18
12645	2021-06-22	74.37
12646	2021-06-23	74.13
12647	2021-06-24	74.14
12648	2021-06-25	74.14

12649 rows × 2 columns

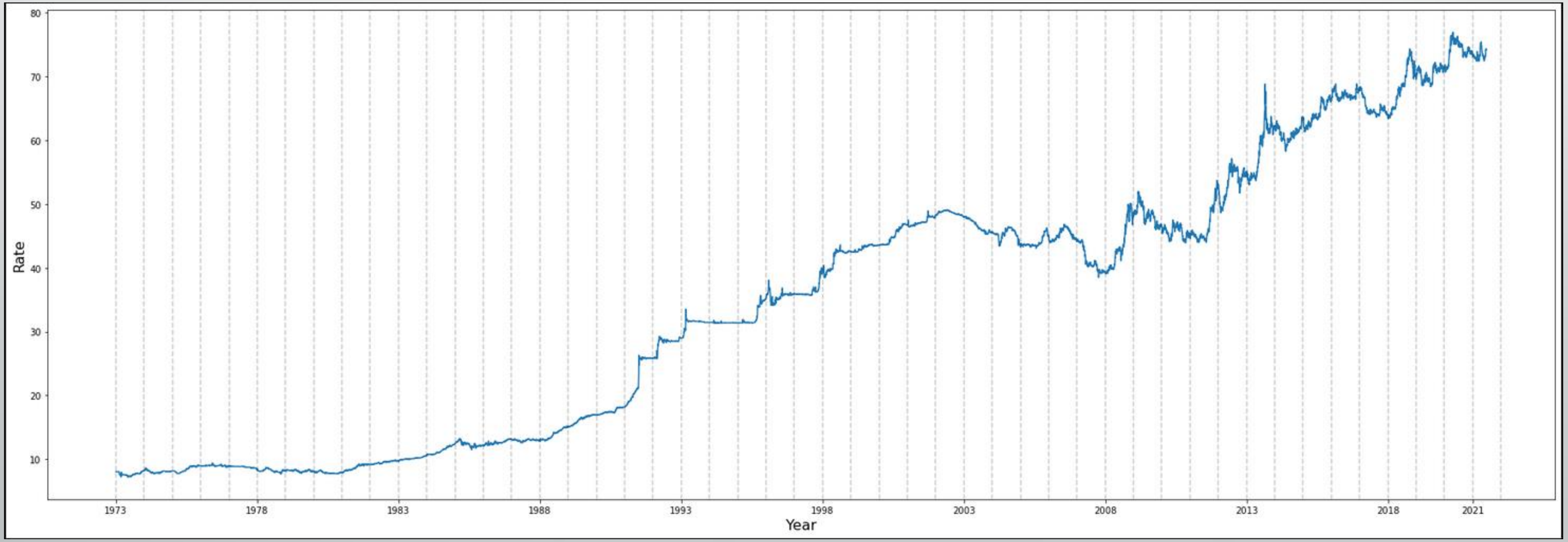
DATA PRE-PROCESSING

- Data imputation was done as there were multiple Null-values in the data set. The other option was to drop the rows with Null-values. As there were around 494 rows containing Null-values, it was not a viable option to the drop the rows, considering that around 4% of data will be lost.
- Hence data was imputed using the below technique.

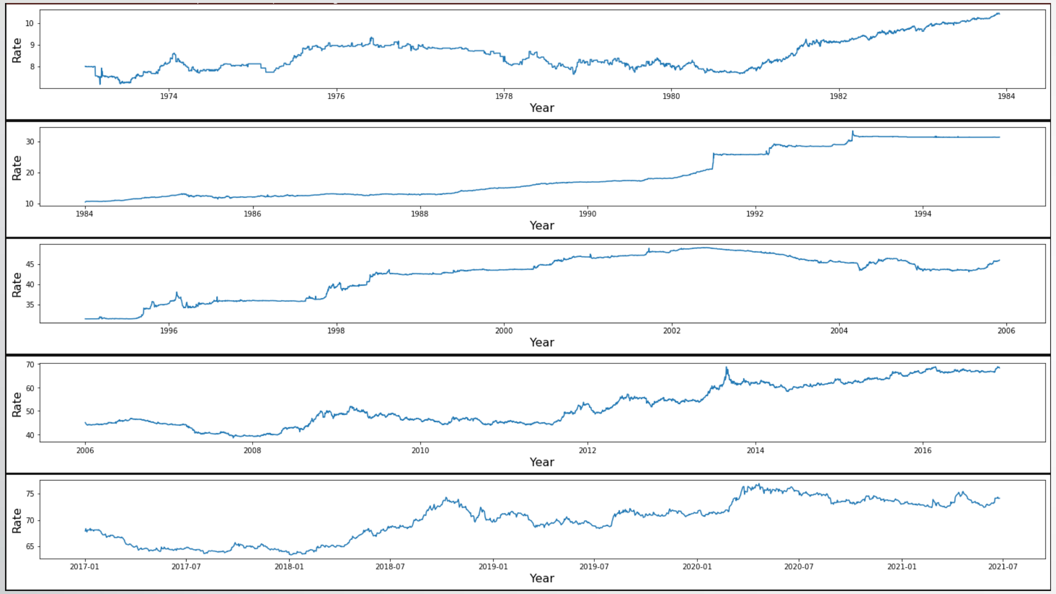
```
[5]: data.isnull().sum()

[5]: Date      0
     Rate    494
     dtype: int64

[6]: # Imputing the null values of the Rate column

     data['Rate'] = data['Rate'].fillna(value= 0.0)

     for i in range(0, len(data['Rate'])):
         if data.Rate[i] == 0.0:
             data.Rate[i] = (data.Rate[i-1] + data.Rate[i-2])/2.0
     data
```

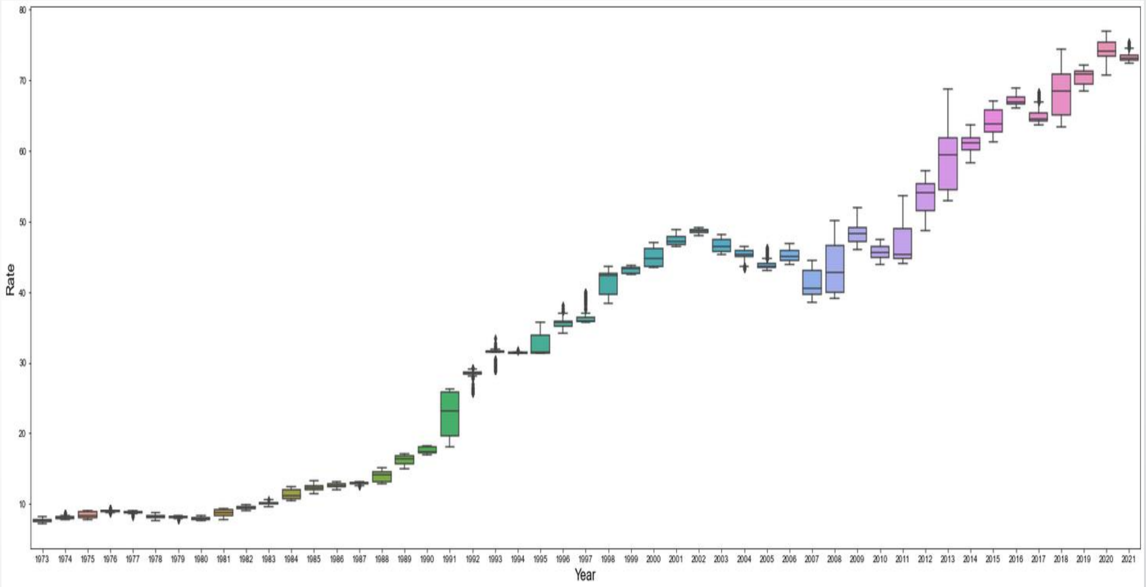
DATA VISUALIZATION

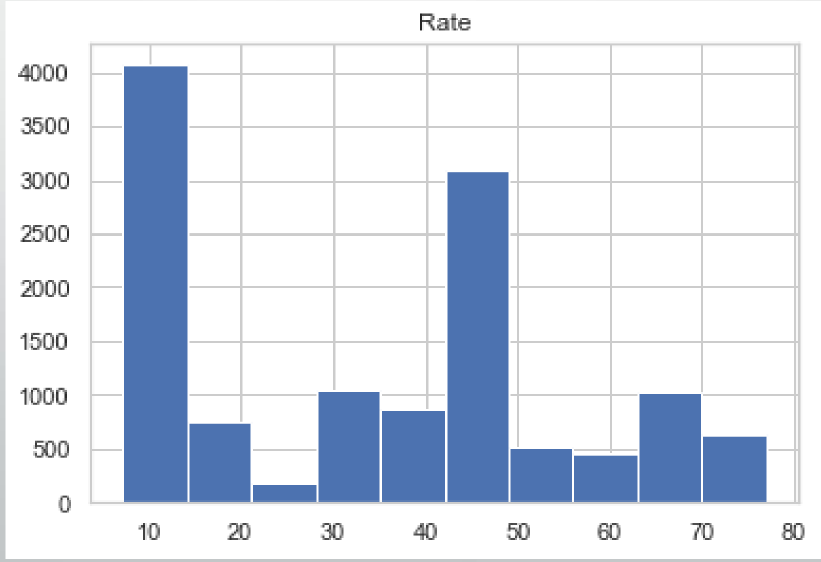
We visualized the given data set, using several inbuilt python plots, so that we can figure out some insights and also to find out the economic reasoning behind sudden changes in the currency exchange rates, which will eventually help us to build better forecasting models



MORE VISUALIZATIONS…

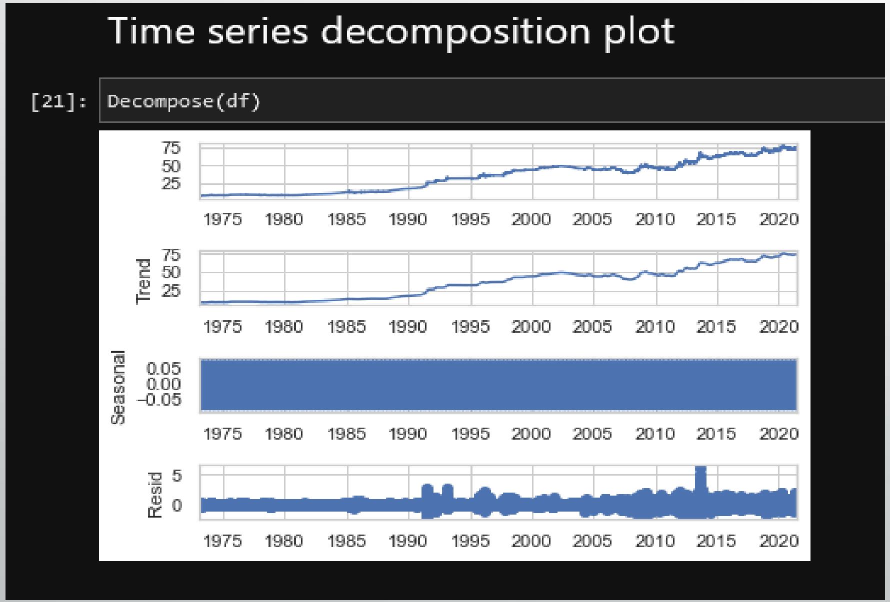


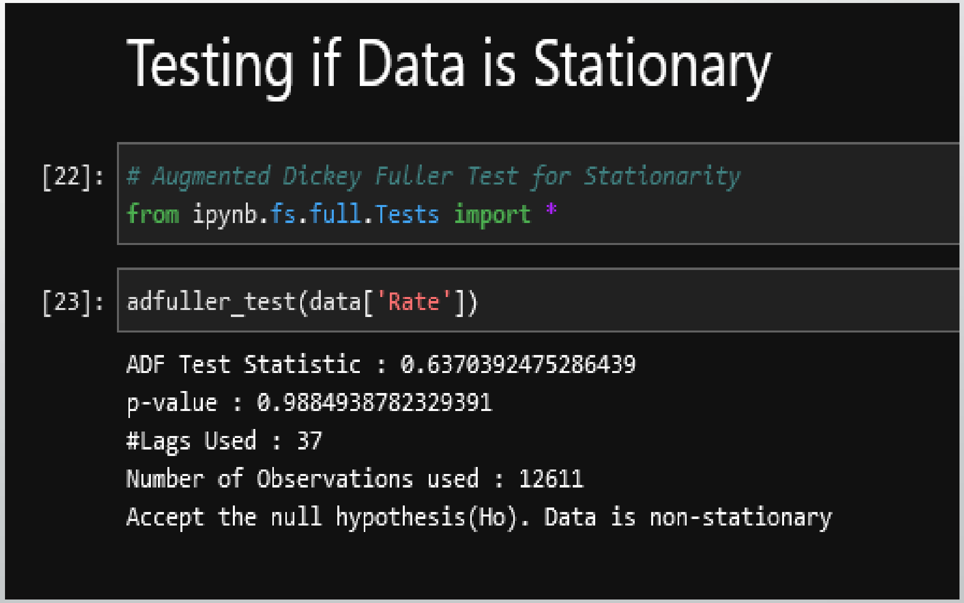




DATA STATIONARITY CHECK

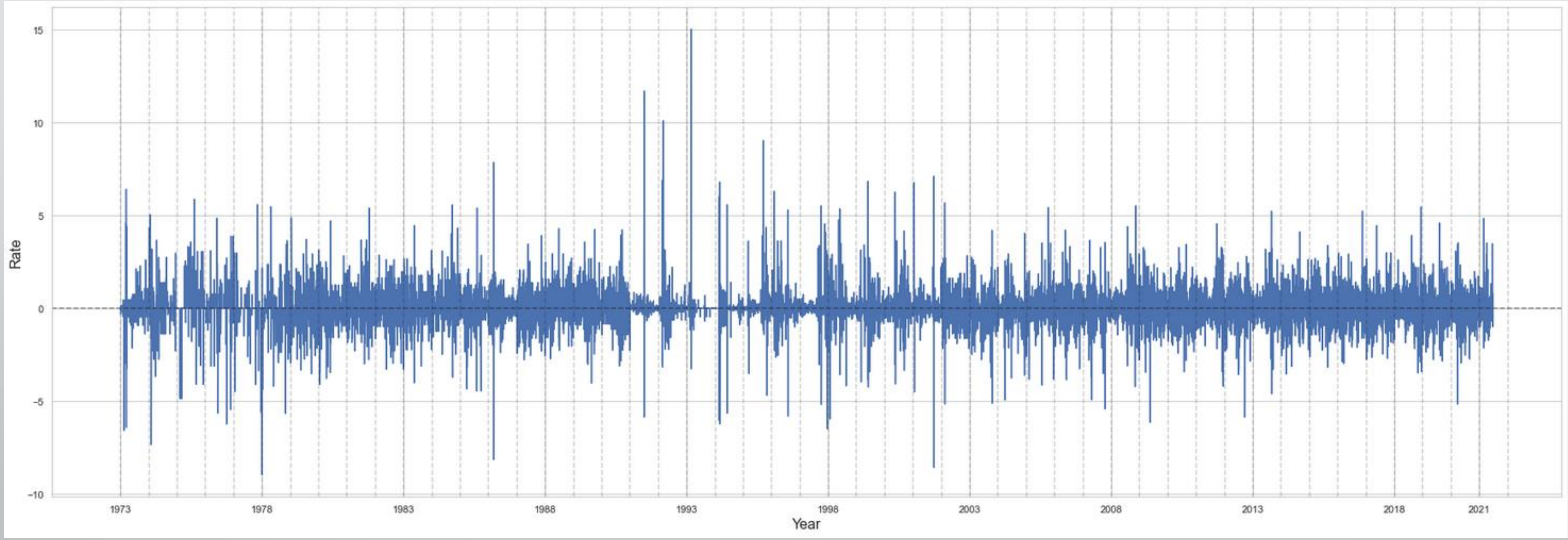
- We check if our data set has any ‘Trend’, ‘Seasonality’, ‘Cyclicity’ or ‘Irregularity’ and we remove them. Some of our forecasting models are susceptible to these components and the efficiency of the model decreases, due to these being present.
- We can figure out the presence of these components by checking the “Decomposition plot” and also we can confirm it by using the “Augmented Dickey-Fuller test”. We check if the p-value from the test is less than 0.05, if not the data is not stationary.





REMOVING TREND AND VOLATILITY

- Before removing the seasonality we normalize our data and then we remove the Trend and Volatility
- Trend component is removed by calculating the first difference and the Volatility component is removed by calculating the standard deviation over time



MODEL BUILDING AND EVALUATION

- Once the previous processes are applied and the data is set, we build the forecasting models. We have used both the Model Based and Data Driven approaches here to get an efficient forecasting result.
- Totally we have built 8 models and evaluated those model results using the RMSE score. Also we have compared their R2 scores. Some of the prominent models we built are AR, ARIMA, Simple Exponential, Holt's Winters Linear Trend, FB Prophet, and three Deep learning models(ANN,LSTM and GRU)
- Out of all the 8 models, ARIMA, GRU and LSTM models gave the best RMSE score(least value). And ARIMA was our best model with an RMSE value of 0.208. Hence we have used this model to forecast the future 30 days' data.
- Some of the forecasting models with efficient results are discussed here…

AR MODEL

RMSE: 0.581, MAPE: 0.0069

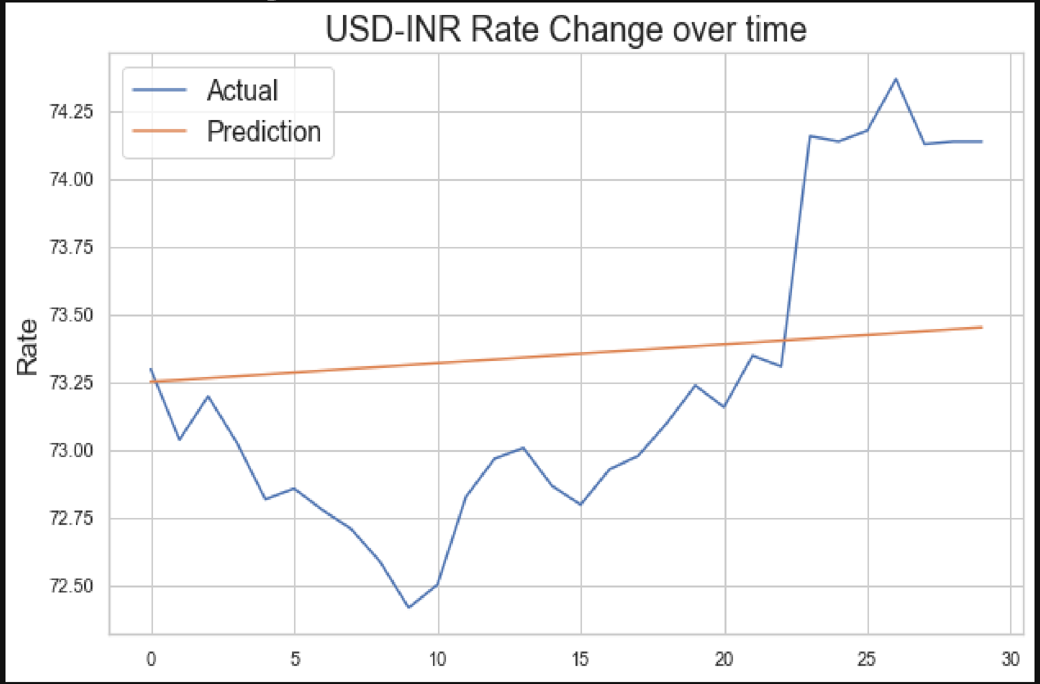
Auto Regressive Model

1. AR Model ¶

yt = c + ϕ1.yt−1 + ϕ2.yt−2 + εt

```
[36]: from ipynb.fs.full.Models import *

[37]: AR_Model(df)

Model fitting time: 0.006324768066406250
Coefficients: [0.00387901 0.9333923  0.06665569]
                            AutoReg Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                12618
Model:                     AutoReg(2)   Log Likelihood                3970.996
Method:               Conditional MLE   S.D. of innovations              0.177
Date:                Wed, 18 Aug 2021   AIC                             -3.467
Time:                        17:13:03   BIC                             -3.464
Sample:                             2   HQIC                            -3.466
                                12618
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0039      0.003      1.295      0.195      -0.002       0.010
y.L1           0.9334      0.009    105.064      0.000       0.916       0.951
y.L2           0.0667      0.009      7.502      0.000       0.049       0.084
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0000           +0.0000j            1.0000            0.0000
AR.2          -15.0031           +0.0000j           15.0031            0.5000
-----------------------------------------------------------------------------
```

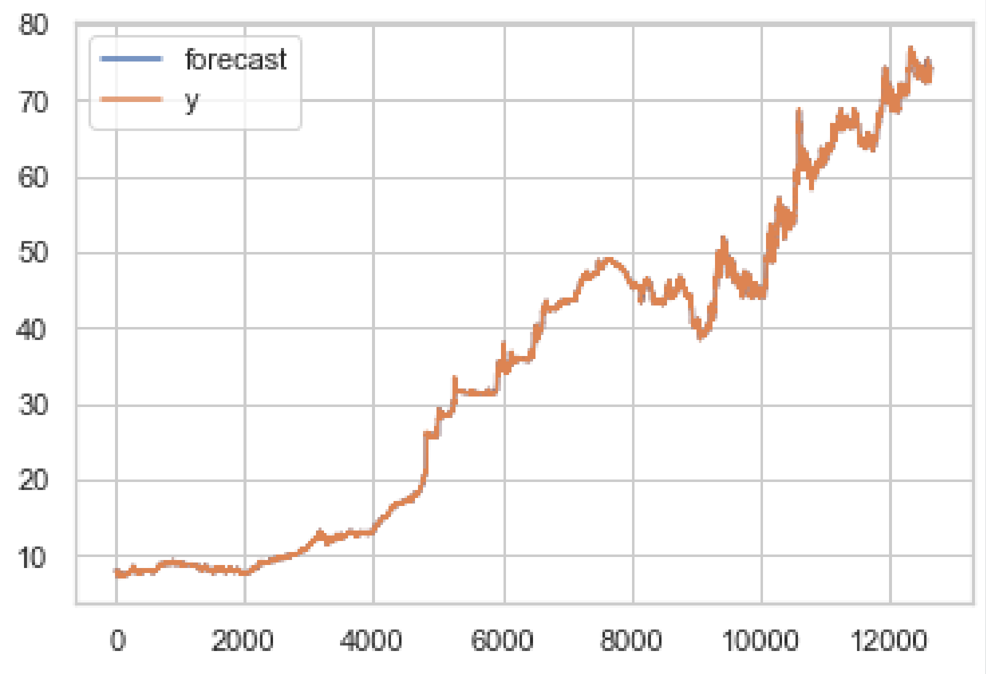
```
Root Mean Squared Error: 0.581
Mean Absolute Percentage Error: 0.0069
```



ARIMA MODEL

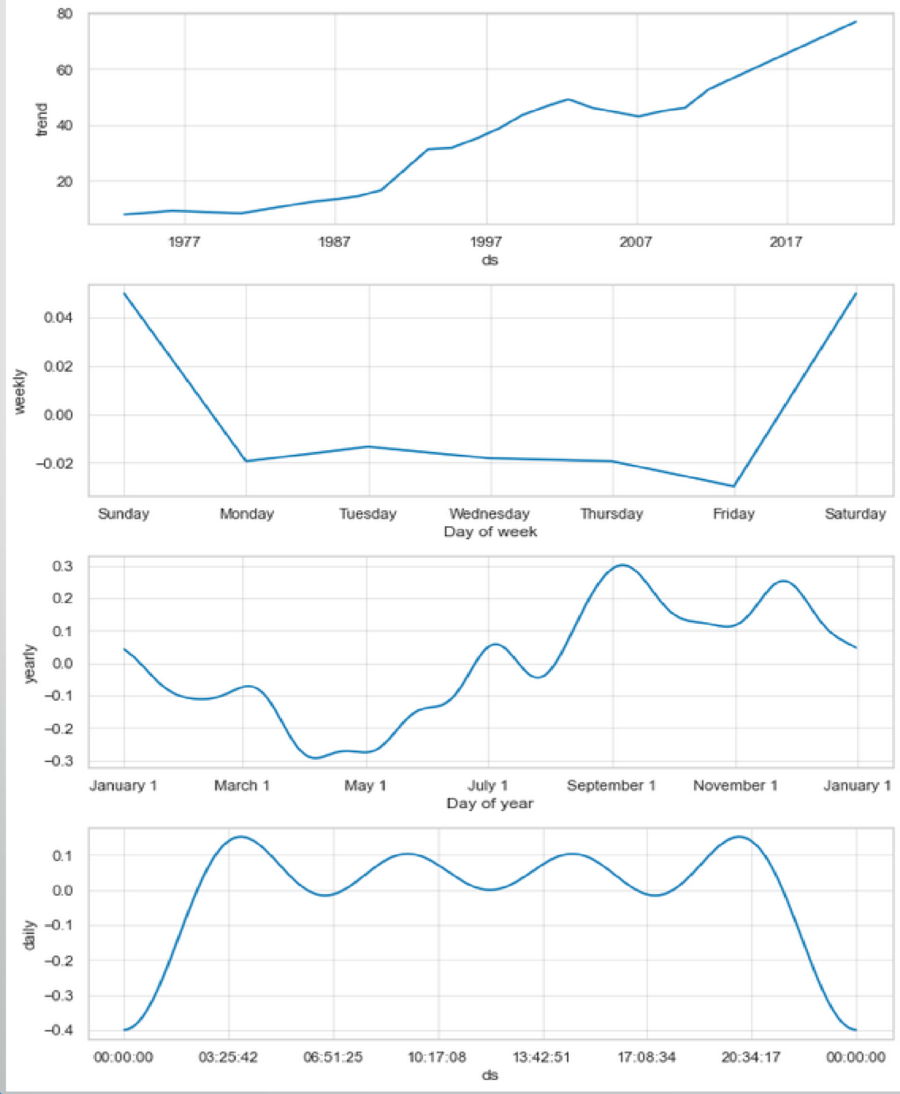
RMSE: 0.208, MAPE: 0.002

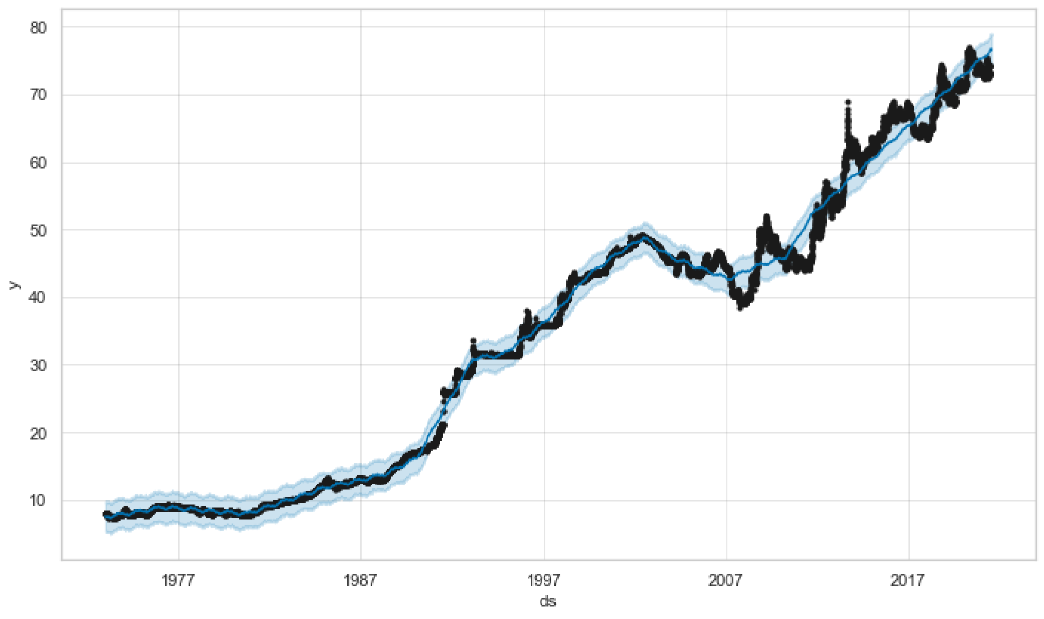
```
Root Mean Squared Error: 0.208
Mean Absolute Percentage Error: 0.002
                              ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                12646
Model:                 ARIMA(1, 1, 1)   Log Likelihood                3978.309
Method:                       css-mle   S.D. of innovations              0.177
Date:                Wed, 18 Aug 2021   AIC                          -7948.619
Time:                        17:13:44   BIC                          -7918.838
Sample:                             1   HQIC                         -7938.655

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0052      0.001      3.648      0.000       0.002       0.008
ar.L1.D.y      0.2073      0.084      2.458      0.014       0.042       0.373
ma.L1.D.y     -0.2768      0.083     -3.350      0.001      -0.439      -0.115
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            4.8236           +0.0000j            4.8236            0.0000
MA.1            3.6130           +0.0000j            3.6130            0.0000
-----------------------------------------------------------------------------
```

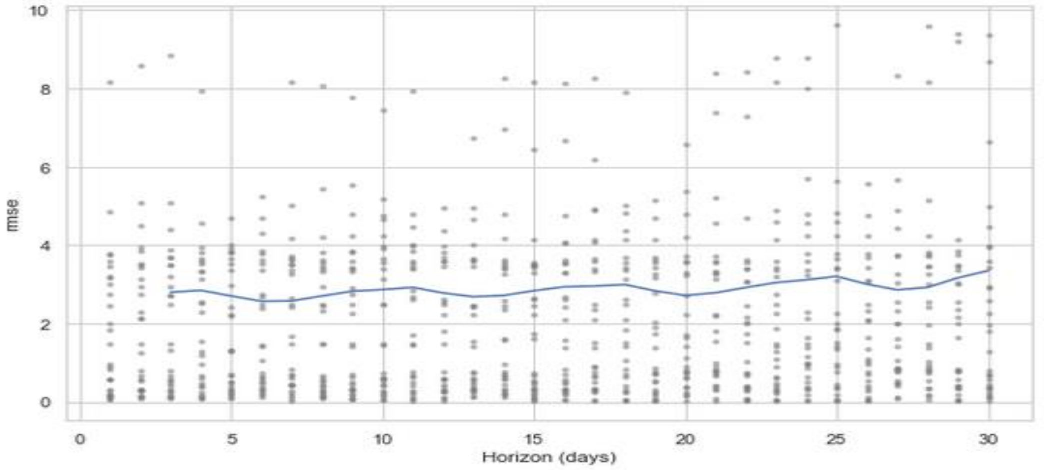


FB PROPHET MODEL

RMSE: 2.821







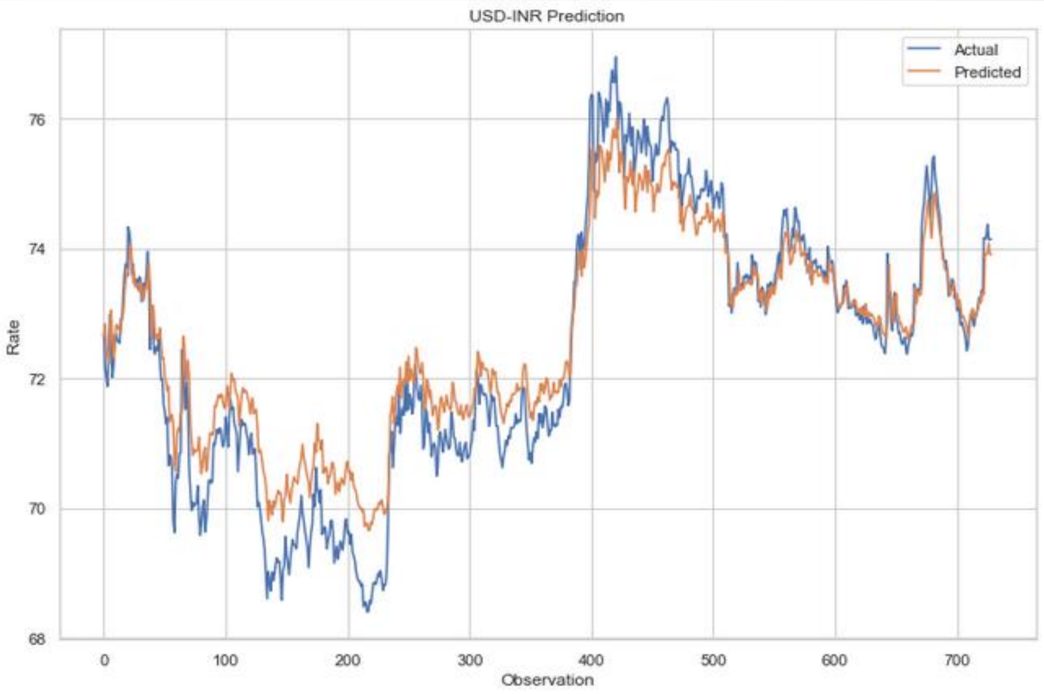
FB PROPHET MODEL

ARTIFICIAL NEURAL NETWORK(ANN)

RMSE: 0.617, R2 Score: 0.908

1. Artificial Neural Network (ANN)

```
[46]: y_test1,y_pred1 = ANN_Model(df)

Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 12)                24
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 13
=================================================================
Total params: 37
Trainable params: 37
Non-trainable params: 0
_________________________________________________________________
Epoch 1/200
11918/11918 [==============================] - 4s 309us/step - loss: 1.1665e-05 - rmse: 0.0034
Epoch 2/200
11918/11918 [==============================] - 4s 306us/step - loss: 2.5748e-04 - rmse: 0.0160
Epoch 3/200
11918/11918 [==============================] - 4s 315us/step - loss: 9.2683e-05 - rmse: 0.0096
Epoch 00003: early stopping
The R2 score on the Test set is:        0.908
The Adjusted R2 score on the Test set is:       0.908
Root Mean Squared Error: 0.617
```

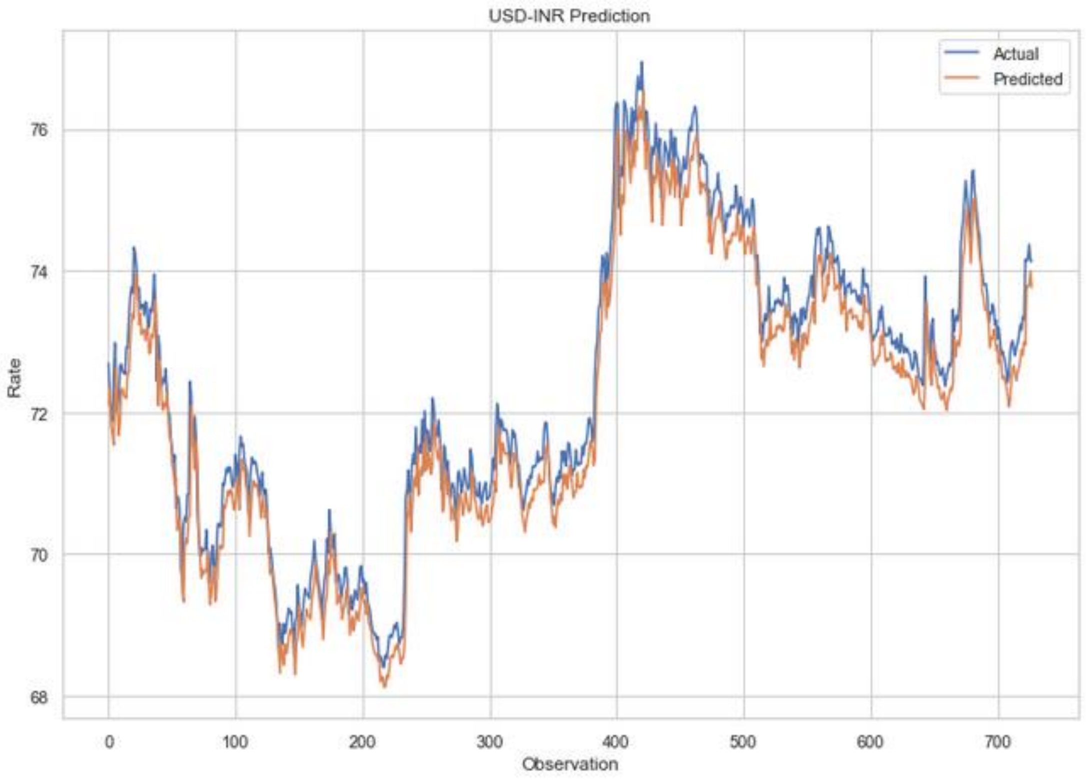


LONG SHORT TERM MEMORY(LSTM)

RMSE: 0.446, R2 Score: 0.952

2. Long Short Term Memory (LSTM Recurrent Neural Network)

```
[48]: y_test2,y_pred2 = LSTM_Model(df)

Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 100)               40800
_________________________________________________________________
dense (Dense)                (None, 1)                 101
=================================================================
Total params: 40,901
Trainable params: 40,901
Non-trainable params: 0
_________________________________________________________________
Epoch 1/5
11917/11917 [==============================] - 7s 530us/step - loss: 2.0440e-04 - rmse: 0.0143
Epoch 2/5
11917/11917 [==============================] - 6s 530us/step - loss: 1.7607e-05 - rmse: 0.0042
Epoch 3/5
11917/11917 [==============================] - 6s 542us/step - loss: 1.7308e-05 - rmse: 0.0042
Epoch 4/5
11917/11917 [==============================] - 7s 559us/step - loss: 1.5695e-05 - rmse: 0.0040
Epoch 5/5
11917/11917 [==============================] - 7s 569us/step - loss: 1.5510e-05 - rmse: 0.0039
The R2 score on the Test set is:        0.952
The Adjusted R2 score on the Test set is:       0.952
Root Mean Squared Error: 0.446
```

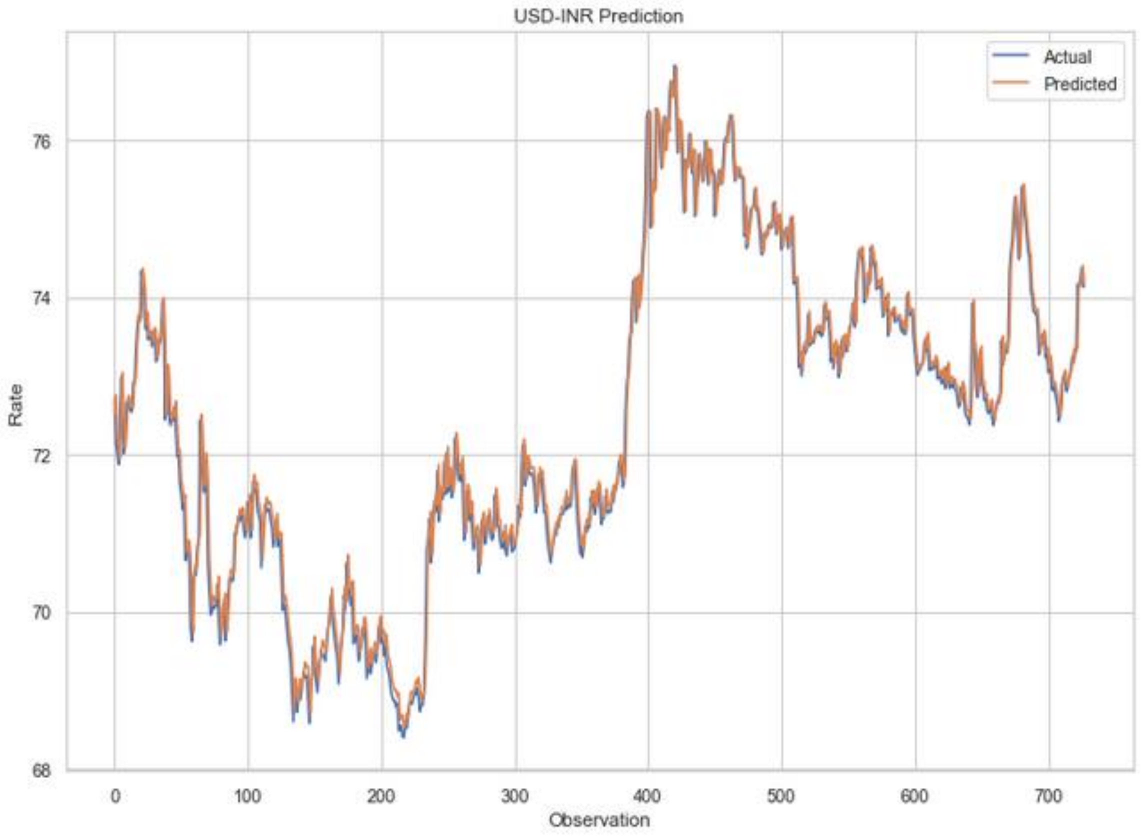


GATED RECURRENT UNIT(GRU)

RMSE: 0.289, R2 Score: 0.980

3. Gated Recurrent Unit(GRU)

```
[50]: y_test3,y_pred3 = GRU_Model(df)

Epoch 1/100
596/596 [==============================] - 1s 548us/step - loss: 1.5833e-05 - rmse: 0.0040
Epoch 2/100
596/596 [==============================] - 0s 584us/step - loss: 6.2941e-05 - rmse: 0.0079
Epoch 3/100
596/596 [==============================] - 0s 596us/step - loss: 1.9298e-05 - rmse: 0.0044
Epoch 4/100
596/596 [==============================] - 0s 582us/step - loss: 9.7537e-06 - rmse: 0.0031
Epoch 5/100
596/596 [==============================] - 0s 621us/step - loss: 8.0950e-06 - rmse: 0.0028
Epoch 6/100
596/596 [==============================] - 0s 595us/step - loss: 8.1079e-06 - rmse: 0.0028
Epoch 7/100
596/596 [==============================] - 0s 610us/step - loss: 8.2660e-06 - rmse: 0.0029
Epoch 00015: early stopping
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
gru (GRU)                    (None, 12)                540
_________________________________________________________________
dense (Dense)                (None, 1)                 13
=================================================================
Total params: 553
Trainable params: 553
Non-trainable params: 0
_________________________________________________________________
The R2 score on the Test set is:        0.980
The Adjusted R2 score on the Test set is:        0.980
Root Mean Squared Error: 0.289
```

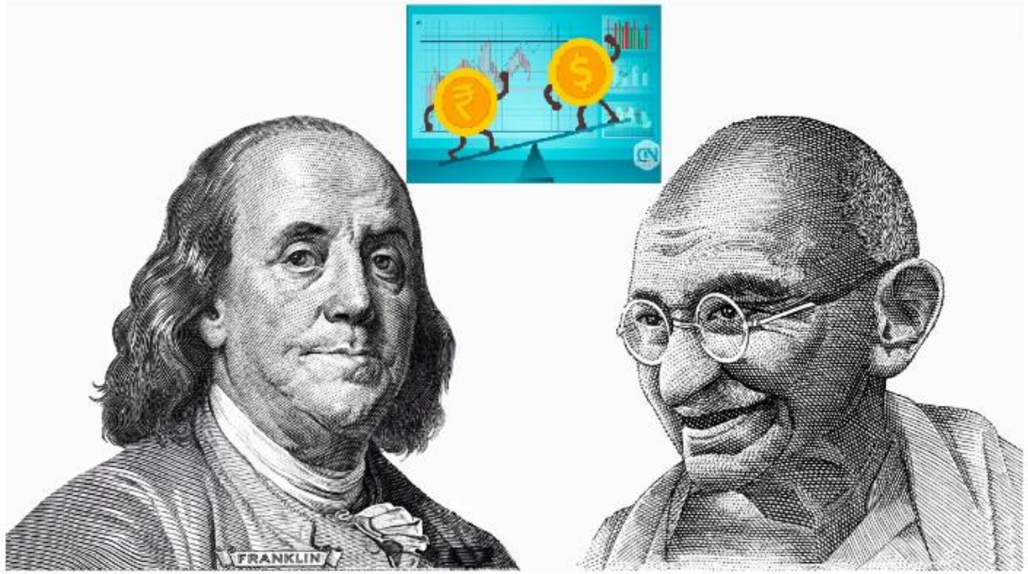


DEPLOYMENT

- After the Model building and Evaluation process, we have deployed the code using “Streamlit”
- We have selected the best 3 models(ARIMA,GRU and LSTM) and used in our deployment phase. Our web-app loads the data on its own, displays the data visualizations and then gives the user an option to select the Model to be used to forecast the future data

Forecasting USD-INR Timeseries Data

Data Exploration and Forecasting the data using ARIMA, LSTM and GRU Models



USD vs. INR

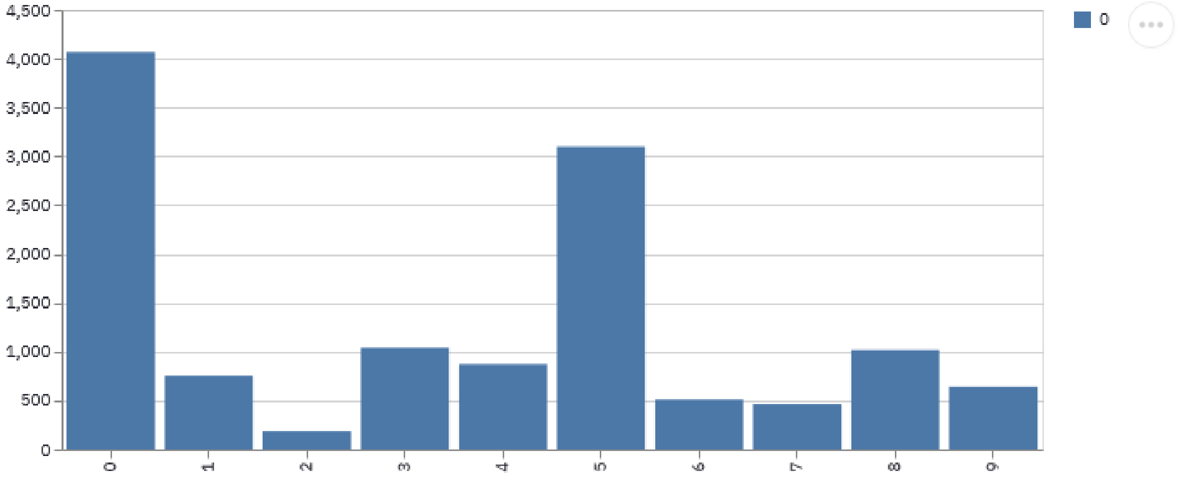
```
Loading data...done!
```

Processed data

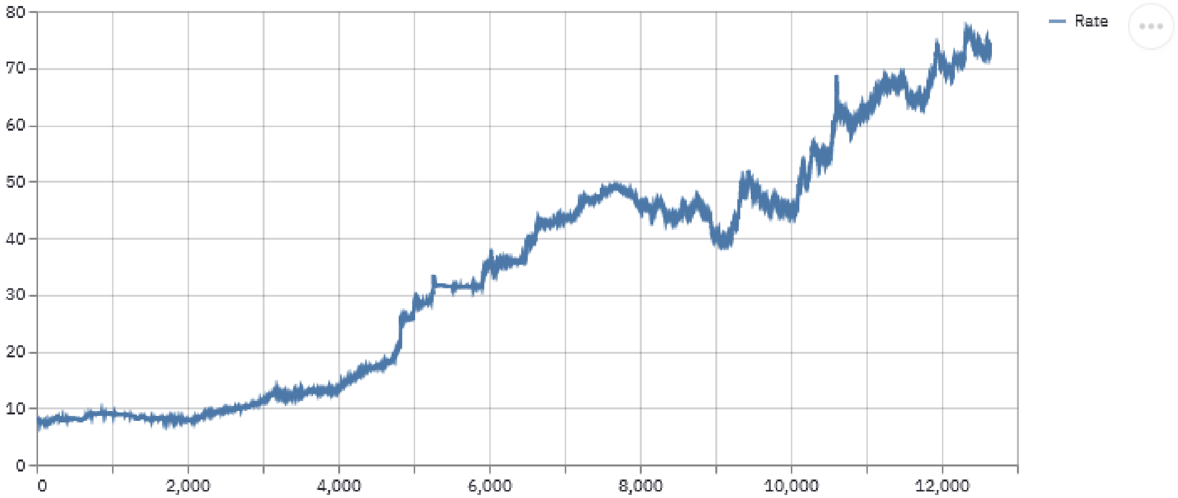
	Date	Rate
0	1973-01-02T00:00:00	8.0200
1	1973-01-03T00:00:00	8.0200
2	1973-01-04T00:00:00	8.0000
3	1973-01-05T00:00:00	8.0100
4	1973-01-08T00:00:00	8.0000
5	1973-01-09T00:00:00	8.0000
6	1973-01-10T00:00:00	8.0000
7	1973-01-11T00:00:00	8.0000
8	1973-01-12T00:00:00	8.0000
9	1973-01-15T00:00:00	8.0000
10	1973-01-16T00:00:00	8.0100

Visualizing our data

Histogram



Line Plot



Model Selection

Select a Model you would like to select, to forecast the future 30 days' data

ARIMA

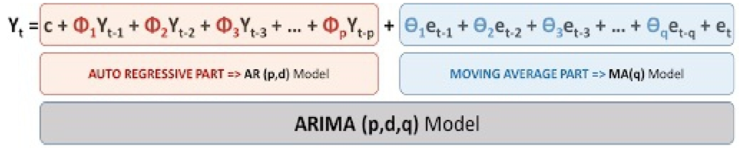
- ARIMA
- LSTM
- GRU

Model Selection

Select a Model you would like to select, to forecast the future 30 days' data

ARIMA

You selected: ARIMA



Structure of ARIMA

Root Mean Squared Error: 0.20713

Mean Absolute Percentage Error: 0.00806

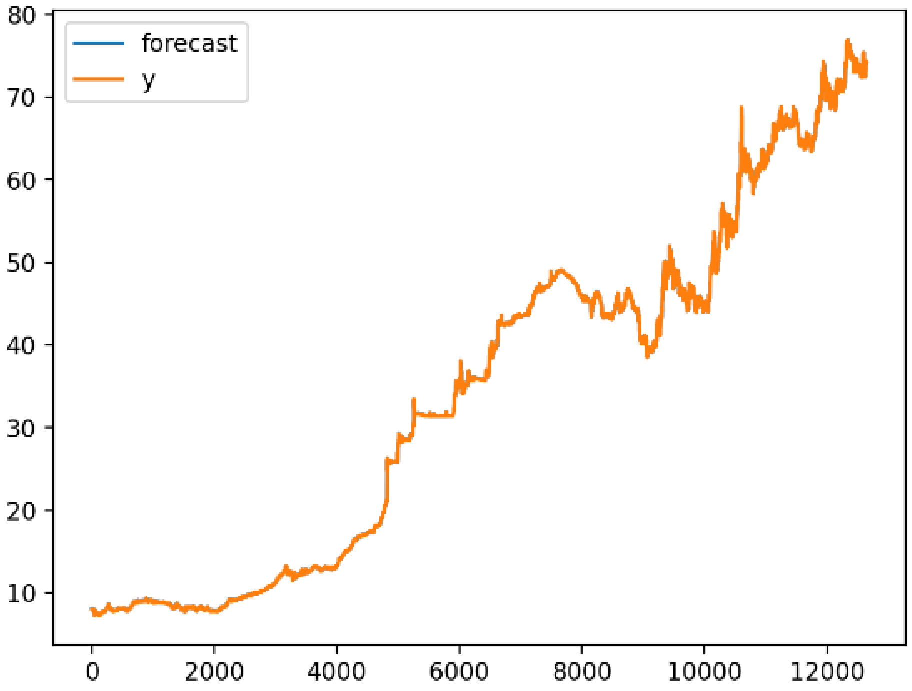
The R2 score on the Test set is: 0.86589

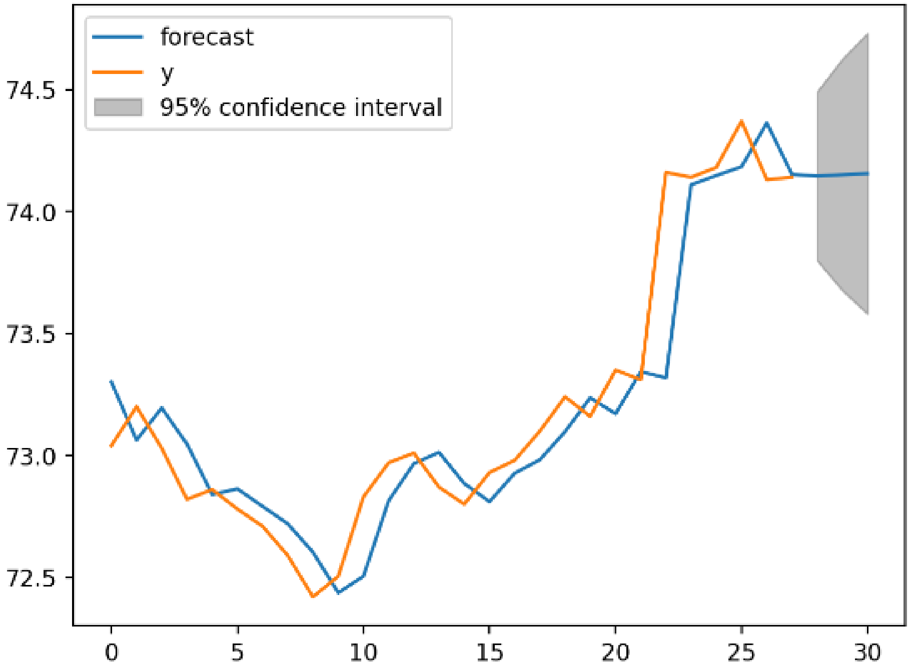
Dep. Variable:	D.y	**No. Observations:**	12646
Model:	ARIMA(1, 1, 0)	**Log Likelihood**	3974.597
Method:	css-mle	**S.D. of innovations**	0.177
Date:	Fri, 20 Aug 2021	**AIC**	-7943.194
Time:	03:03:36	**BIC**	-7920.859
Sample:	1	**HQIC**	-7935.721

ARIMA Model Results

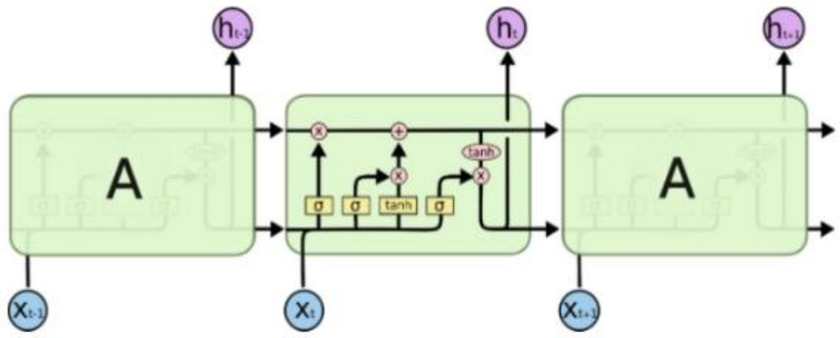
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0052	0.001	3.549	0.000	0.002	0.008
ar.L1.D.y	-0.0666	0.009	-7.504	0.000	-0.084	-0.049

	Real	Imaginary	Modulus	Frequency
AR.1	-15.0206	+0.0000j	15.0206	0.5000





You selected: LSTM



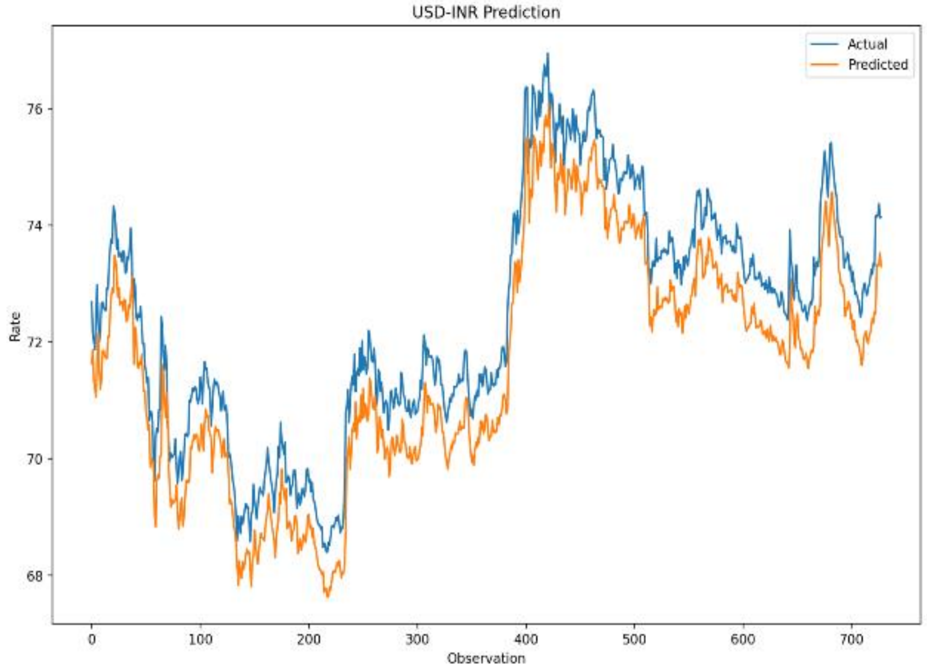
Structure of LSTM

	0
0	0.0143

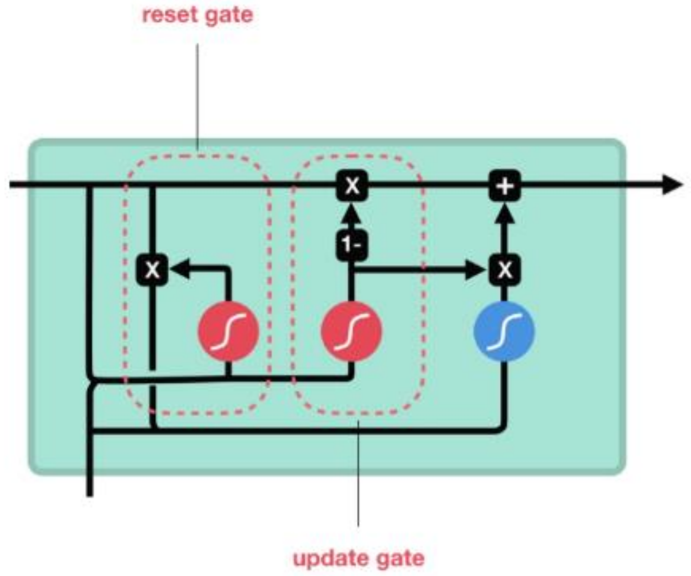
Root Mean Squared Error: 0.86793

The R2 score on the Test set is: 0.81778

The Adjusted R2 score on the Test set is: 0.81753



You selected: GRU



Structure of GRU

None

Root Mean Squared Error: 0.28865

The R2 score on the Test set is: 0.97985

The Adjusted R2 score on the Test set is: 0.97982