Statistical Machine Learning

Foreign Exchange Rate Prediction

I. PROBLEM STATEMENT AND MOTIVATION

Foreign exchange rate forecasting helps the business man and brokers make better decisions to minimize risks and maximize returns. The relative economic strength of a country relates to the exchange rates with the currency of other countries.A high economic strength and strong economic environments are more likely to attract investments from foreign investors. To make an investment in the desired country, an investor has to first buy that country's currency. If the amount of currency flowing around the country is higher the exchange rate for that country will be low. Currently the world's strongest currency is Kuwaiti Dinar. It has the highest valued currency against the United States Dollar. The Iranian Rial is the weakest currency in the world. Crashes in the foreign exchange markets are quite different from that in the stock markets. Stock markets affect many countries but changes in the foreign exchange rate affects that country only.

An increase in the exchange rate is in favour of the businessman as it will result in more value or worth of goods for their money. Decrease in exchange rates will require the businessman to spend more money for the same value or worth of goods. Multinational Corporation is more affected by changes in exchange rate than a national company. The fluctuation of exchange rate will affect the cash flow through transactions and lead to economic risk.

Thus knowing the exchange rate beforehand can save the people as well as the country from an economic risk. This was the main motivation behind choosing this topic.

II. INTRODUCTION

Foreign exchange rate market also known as Forex market is a globally decentralized market for trading currencies. It is a huge market in the world processing billions worth of Dollars everyday. Exchange rate is how much it costs to exchange one currency for another. These rates can differ from weeks,daily,monthly or yearly.

If it is given that USD/CAD is 1.55, that means it costs 1.55 Canadian Dollars for 1 U.S. Dollar. The first currency always stands for one unit of that currency and the exchange rate shows how much of the second currency is needed to purchase the one unit of first currency. This is how the currency

exchange rates are read. To know how much it costs to buy one Canadian Dollar using U.S. dollars we can use the following formula :1/exchange rate. The conversion rates used by the banks are not the raw rate, The bank or currency exchange house will always markup the price so that they might get their profit. The credit card companies and payment services also increase the currency rates to increase gains.

III. LITERATURE REVIEW

Many machine learning techniques are being used to predict stock markets in the economic world. But there is much less work in predicting Foreign exchange rates.

Research made by Rojas and Herman from Stanford University [1] on the Mexican Peso (MXN) and USD exchange rate using two datasets namely market variables dataset and fundamental variable dataset. They had used the factors which are responsible for change of exchange rates in their dataset. Factors like Industrial production, Trade Balance (Export- Import), Unemployment rate, Consumer Price Index, Producer Price Index, National Debt and Money Supply were used as the independent variable to predict the foreign exchange rates. The models used were Linear,regularized linear, Support Vector Machine Regression, Gradient Boosting Regression and artificial neural networks. Ridge regression gave the best accuracy.

Another researcher from Egypt named Areej Baasher and Fakhr [2] used time series data on exchange rate for Egyptian currency and made it into a classification problem. They classified it as uptrend and downtrend. They made different duration and then detected uptrend and downtrend. Feature selection methods like PCA and LDA has been used for improving the accuracy of RBF kernel, Multi layer Perceptron, Support vector machine.

Dinesh k. Sharma [3] from University of Maryland used regression techniques to predict exchange rates. They mainly focus on the comparison and implementation of regression techniques and ensemble methods. Data-set used was a time series value. They have used cross validation techniques for improving the model and found that bagging and boosting techniques work better with their data than normal regression techniques.

The [4] paper by Farrell and Correa mainly focuses on the application of Gaussian Processes in the Stock market analysis. Data is taken from 8 stocks over a span of 10 years. They have tried their model on various kernels. study is mainly focused on the rise and fall of prices rather than the prediction of prices.

IV. DATASETS

A. Details

We have used 2 datasets one of them contains the exchange rate of 22 countries' currency with the U.S.Dollar. The data has a date column where date varies from 01 January 2000 to 31 December 2019. so the data consists of 20 years. so it is a time series data with 23 columns. The number of rows are 5217. The other represents USD to INR (United States Dollar to Indian Rupee) conversion rates over a period of time. It has 7 columns and 4051 rows.

B. Source

The data-set has been taken from Kaggle [5] The datasets have been generated on the Federal Reserve's download data program.

1. Dataset-1
2. Dataset-2

C. Preparation of data-sets

In the data-set some of the values are missing. Even some of the dates are also missing. Some rows have NA values. If the rows have missing values then the values are filled using the top and bottom values of the missing rows or are not considered.

V. PROPOSED ARCHITECTURE

We are using time series data on India and the U.S. Dollar exchange with dates. so basically only the data of two columns namely INR/USD exchange rate and date is used from the second Dataset while the dataset 1 is completely used.

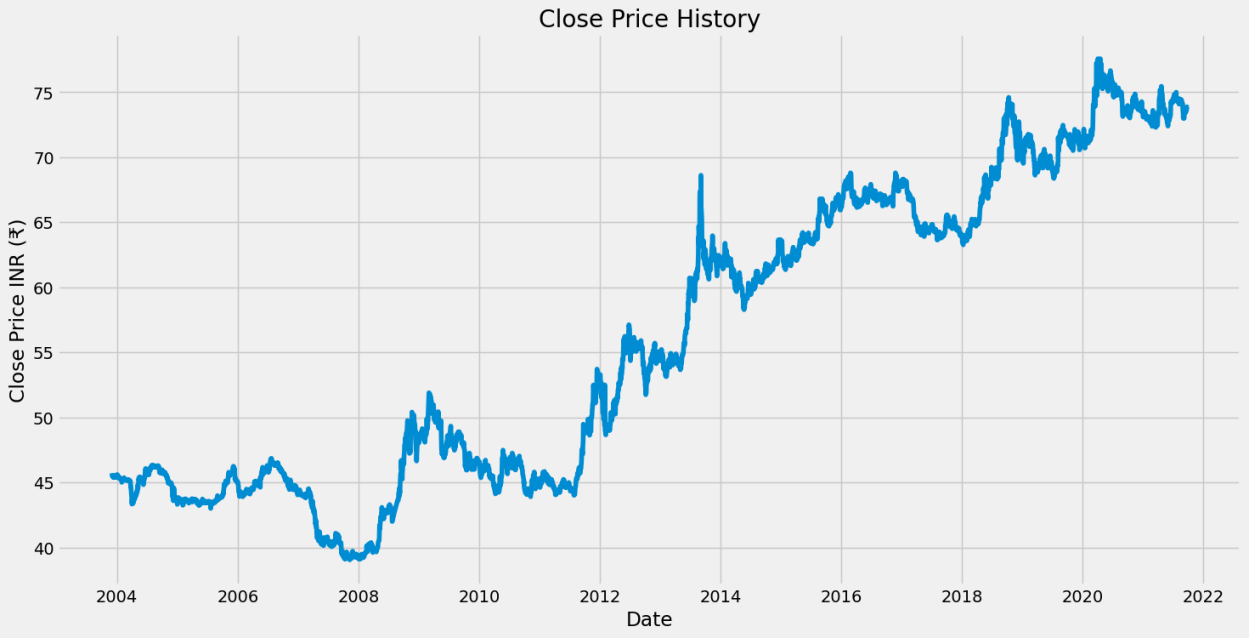
A. Linear , Ridge Regression

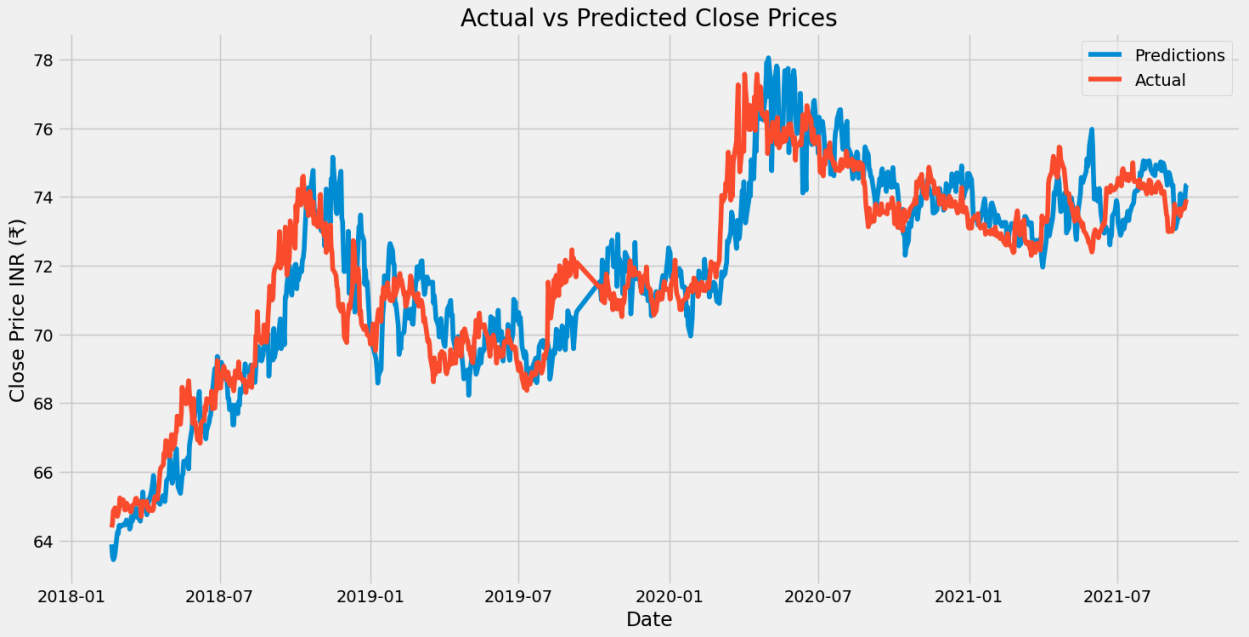
We have applied Linear, Ridge regression. Regression techniques analyze the use of lagged dependent variable. it estimates the independent variable to forecast the values of the dependent variable. Linear regression can have large coefficients. Ridge uses L2 normalization to get the regression curve. Ridge penalizes the sum of squares coefficients.

B. ARIMA

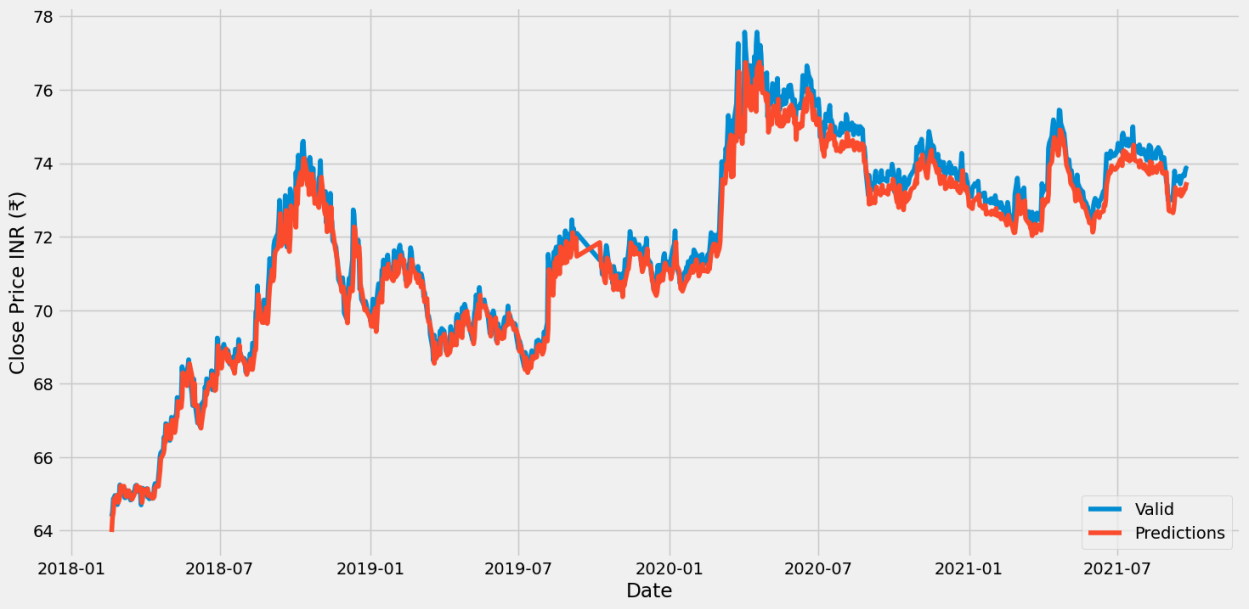
ARIMA stands for Auto Regressive Integrated Moving Average. It is made up of auto regression and moving average integrated together. It is considered as the best model for forecasting time series data. ARIMA works on stationary data only. Stationary data is tested with dickey fuller test and rolling mean and variance test. We found that our data was not stationary as the critical point did not reject the null hypothesis. we have tried different variations to make the data stationary like centralizing the data, taking log etc. Auto correlation function and correlation function have been used to get the values of p and q respectively. The d value for ARIMA is found using number of operations done to get stationary data. We have tried with many different values and best Mean Square Error is achieved at p=25,q=0 and d=1.

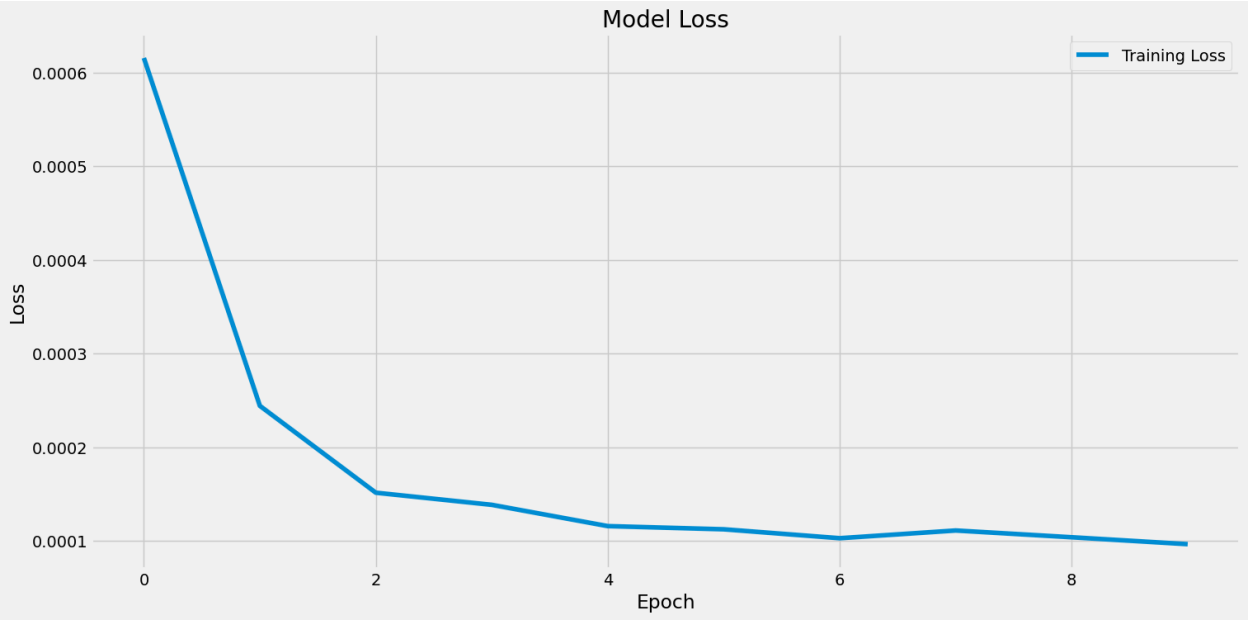
VI. RESULTS



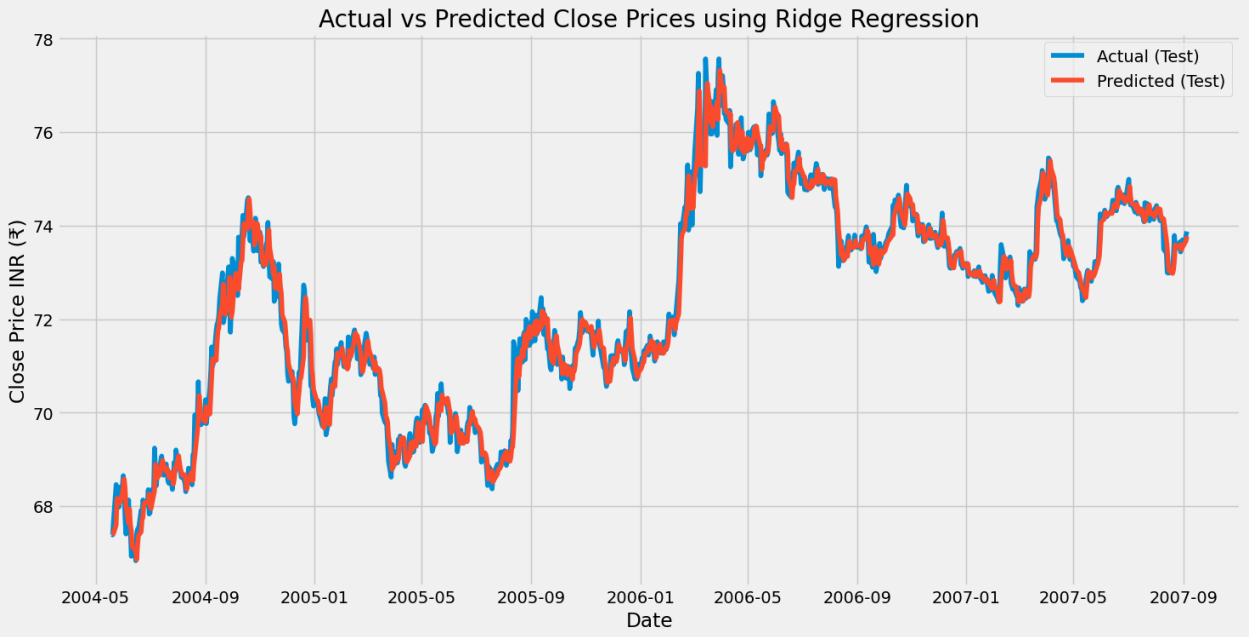

Actual vs Predicted Close Prices
Predictions
Actual
Close Price INR (₹)
78
76
74
72
70
68
66
64
2018-01
2018-07
2019-01
2019-07
2020-01
2020-07
2021-01
2021-07
Date

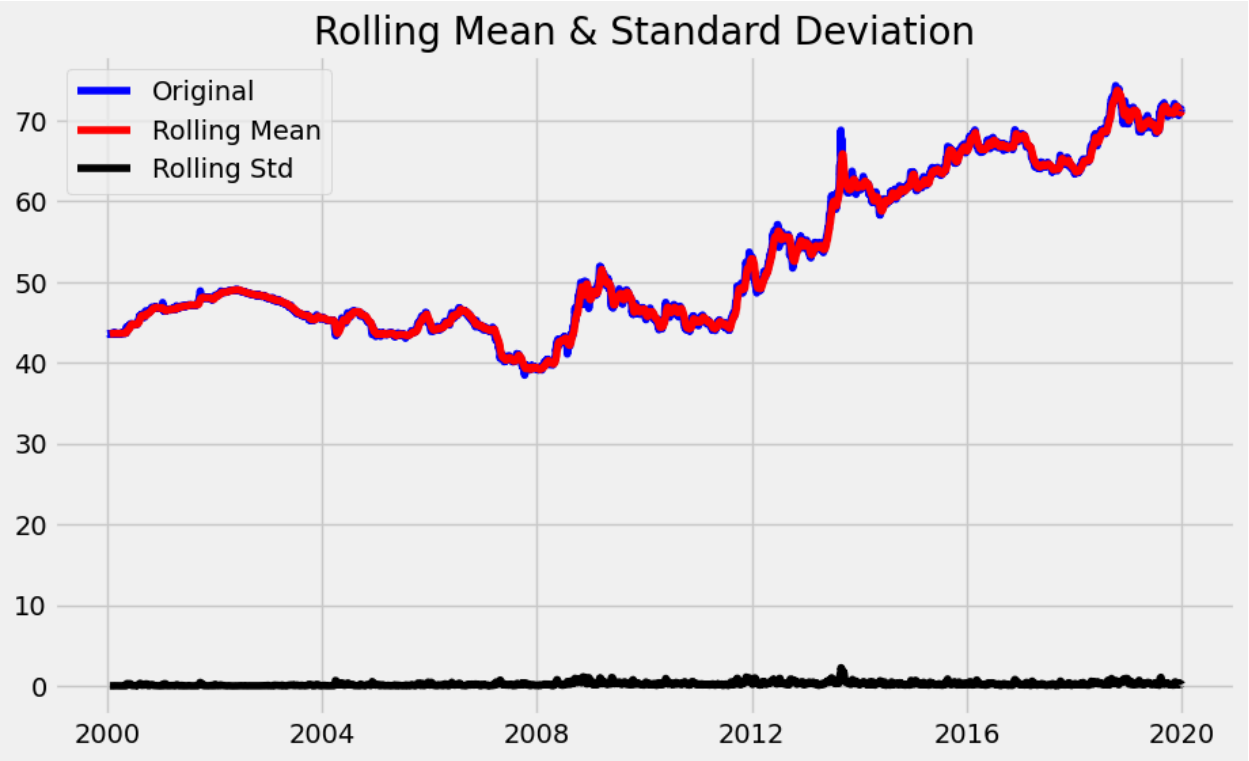

Close Price INR (₹)
75
70
65
60
55
50
45
40
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Date
Train
Valid
Predictions

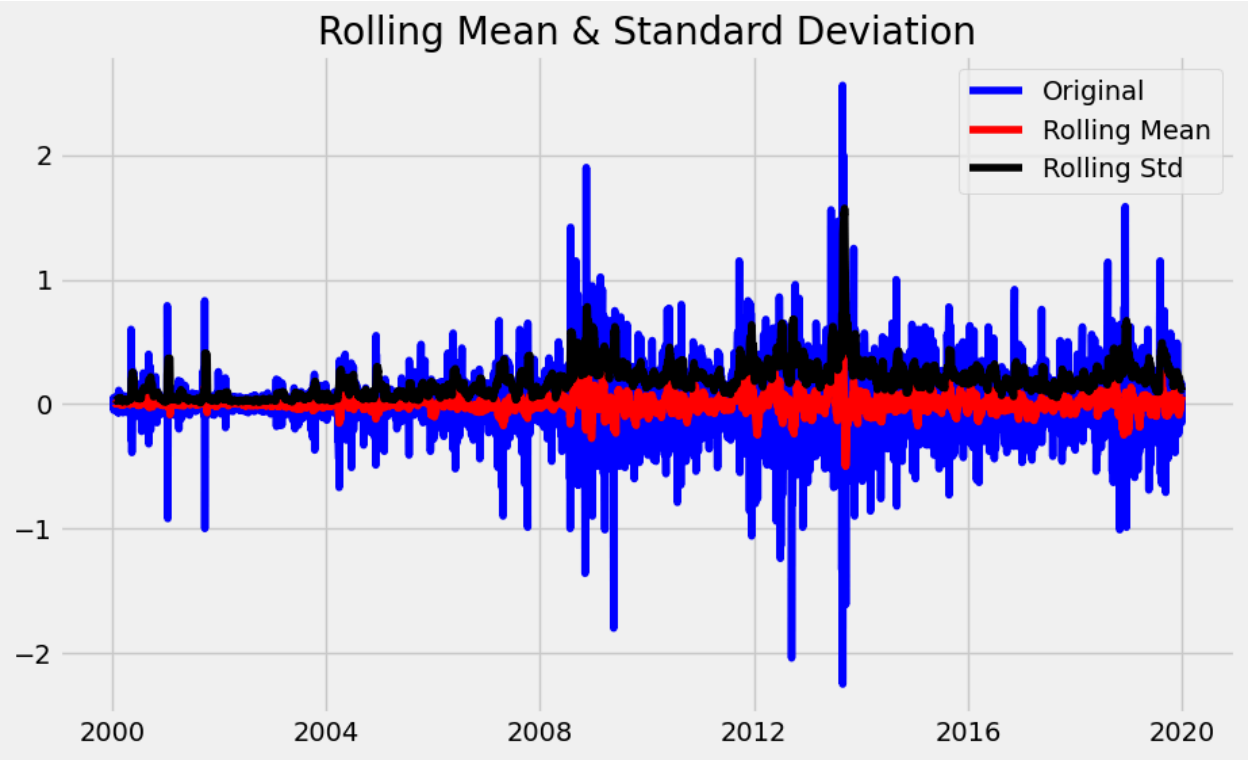

78
76
74
72
70
68
66
64
Close Price INR (₹)
2018-01
2018-07
2019-01
2019-07
2020-01
2020-07
2021-01
2021-07
Date
Valid
Predictions

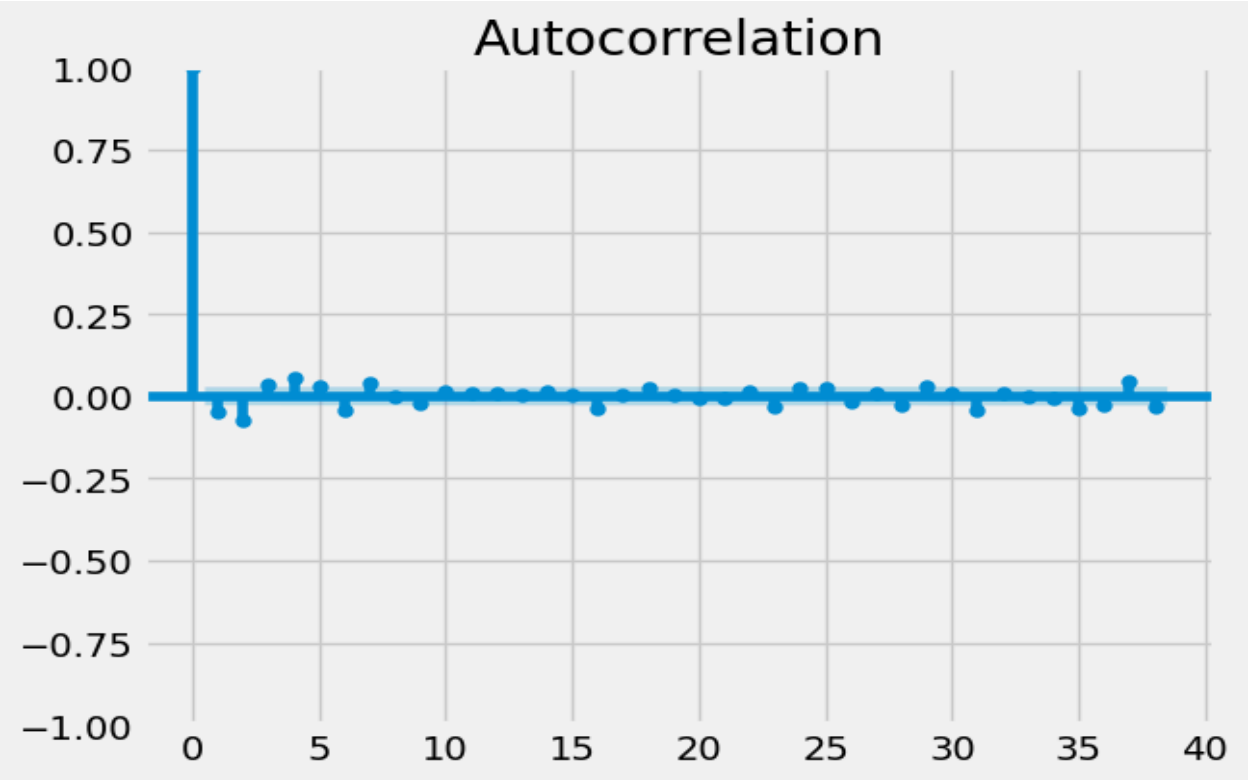

Model Loss
Training Loss
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
Loss
0
2
4
6
8
Epoch

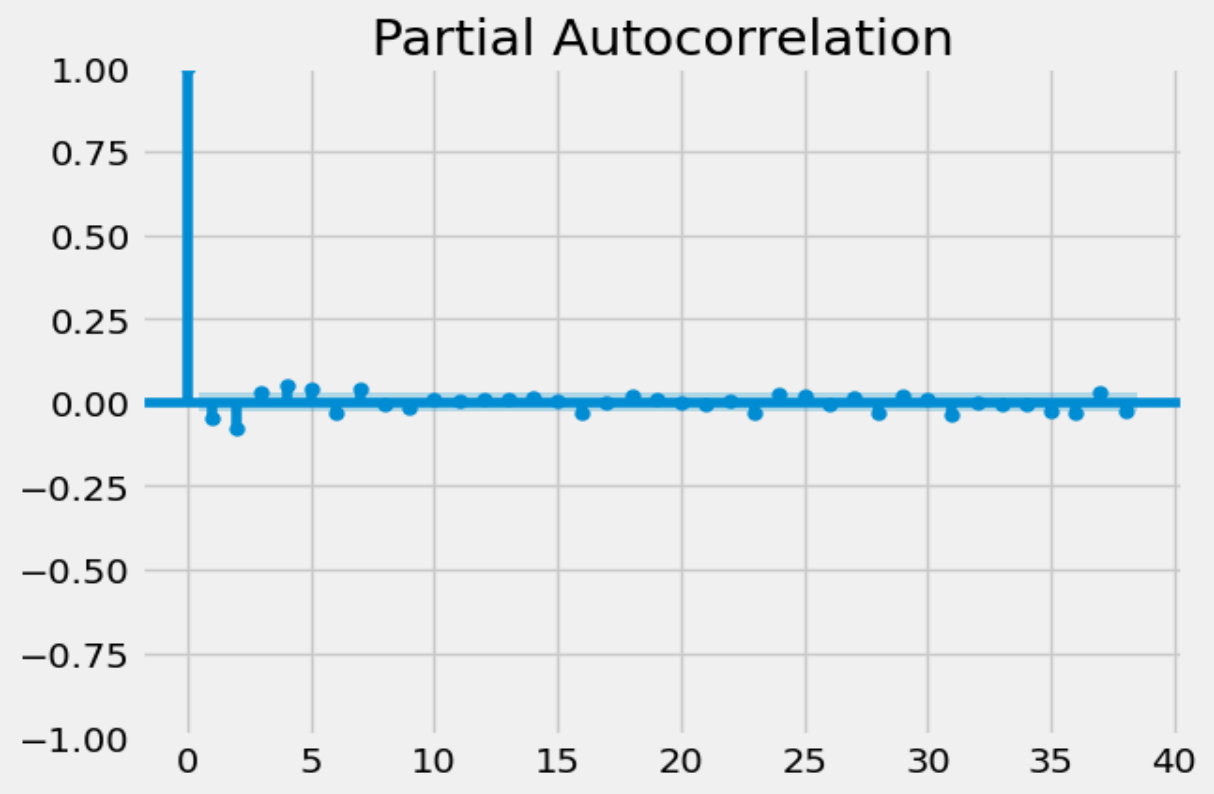

Close Price History
Close Price INR (₹)
Date
75
70
65
60
55
50
45
40
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022

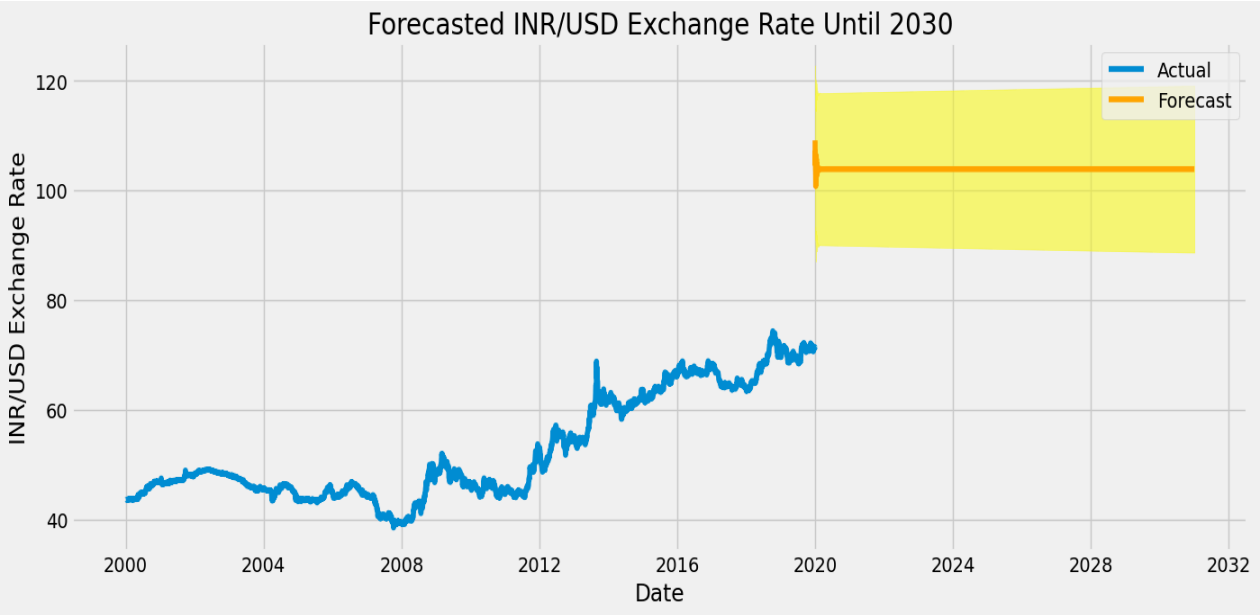

Actual vs Predicted Close Prices using Ridge Regression
Actual (Test)
Predicted (Test)
Close Price INR (₹)
Date
78
76
74
72
70
68
2004-05
2004-09
2005-01
2005-05
2005-09
2006-01
2006-05
2006-09
2007-01
2007-05
2007-09


Rolling Mean & Standard Deviation
Original
Rolling Mean
Rolling Std
70
60
50
40
30
20
10
0
2000
2004
2008
2012
2016
2020


Rolling Mean & Standard Deviation
Original
Rolling Mean
Rolling Std
2
1
0
−1
−2
2000
2004
2008
2012
2016
2020


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20
25
30
35
40


Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20
25
30
35
40


Forecasted INR/USD Exchange Rate Until 2030
Actual
Forecast
INR/USD Exchange Rate
120
100
80
60
40
2000
2004
2008
2012
2016
2020
2024
2028
2032
Date