A Comprehensive ML study in US Economic Trends: Inflation, Unemployment & the Real Interest Rate

ROHIT KUMAR
I200327
1st YEAR MA APPLIED QUANTITATIVE FINANCE

ABSTRACT

Real Interest Rate has been a valuable variable to analyse the growth of a country over the years. This time series project focuses on United States data over a period of 1970-2020 exploring a dynamic relationship between the Real Interest Rate (RIR) & key economic indicators like Consumer Price Index (CPI) inflation, Gross Domestic Product (GDP) deflator inflation & Unemployment. The aim is to show how these macroeconomic variables influence the Real Interest Rate & assess the prediction capability of various ML models. (Mishkin, Frederic S.,1982.)

The data rings a comprehensive timeline, capturing the economy by period of growth, recession & financial crisis. The main focus is to understand the intricate relation between the Real Interest Rate & (CPI & GDP deflator) Inflationary pressures as well as the impact of Unemployment. (Pennacchi, George G.,1991)

I have used MS excel to bring all the data at one place from all the sources then analysed it in python in Google Collaboratory.
In the first half of the project exploratory data analysis is in work to visualize trends & identify potential patterns in data. Showing it in the graphs & maps that how the variables had there move across the time period in U.S. Later half there are ML models including Long Short-Term Memory (LSTM RNN), Random Forest (RF), Support Vector Machine (SVM), K- nearest neighbour (KNN). They are trained & evaluated to predict the Real Interest Rate based on the selected macroeconomic indicators. Results of models are punctiliously analysed to ascertain the factors that contribute to variation in the Real Interest Rate. Also, this project investigates the relative importance of each feature in influencing the target variable. Information gained from the analysis contributes to a refined understanding of the tangled relationship within the US economy.

The project will demonstrate the versatility of ML models in predicting & intricate relation between my target variable & other macroeconomic variables. The findings of the project will implicate financial analysts, economists & policymakers. It is blend of hoe machine learning models can be helpful in a study related to macroeconomics of nation.

1. INTRODUCTION

The economic landscape is heavily woven with threads of interest rate, unemployment & inflation with each one influencing the other. To understand this is very important for economists & policy makers aiming to navigate monetary & economic stability. This ML project focuses on the dynamic relationship between Real Interest rate (RIR), Consumer price (CPI) inflation, Gross domestic product (GDP) deflator inflation & Unemployment. (Lucas Jr, Robert E., and Leonard A. Rapping.,1969)
Overall, these macroeconomic variables are very important from a point of view of nations' development. Time series data refers to data that is collected over time at fixed intervals, such as daily, weekly, monthly or yearly. This analysis is a statistical method that is used to analyse & predict trends in time series.

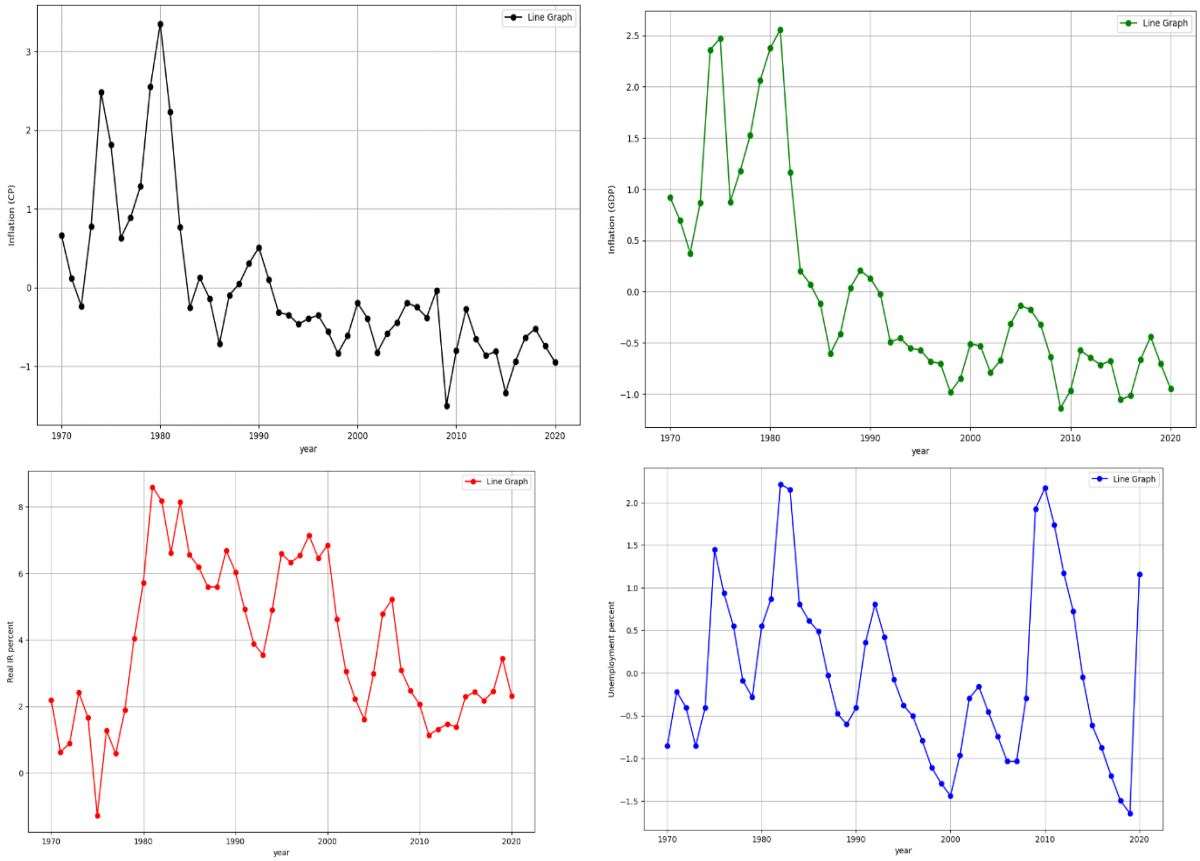

Fig 1: There is a graphical indication of all the measures separately over time.

Then there is a graphical representation of all the macroeconomic variables over the years. The real IR is the nominal interest rate adjusted for inflation. Inflation (CPI) is the average price paid by consumers for goods & services. Inflation (GDP) is a broad measure of prices of services & goods. Unemployment is the not employed labour force but are available or seeking for work.

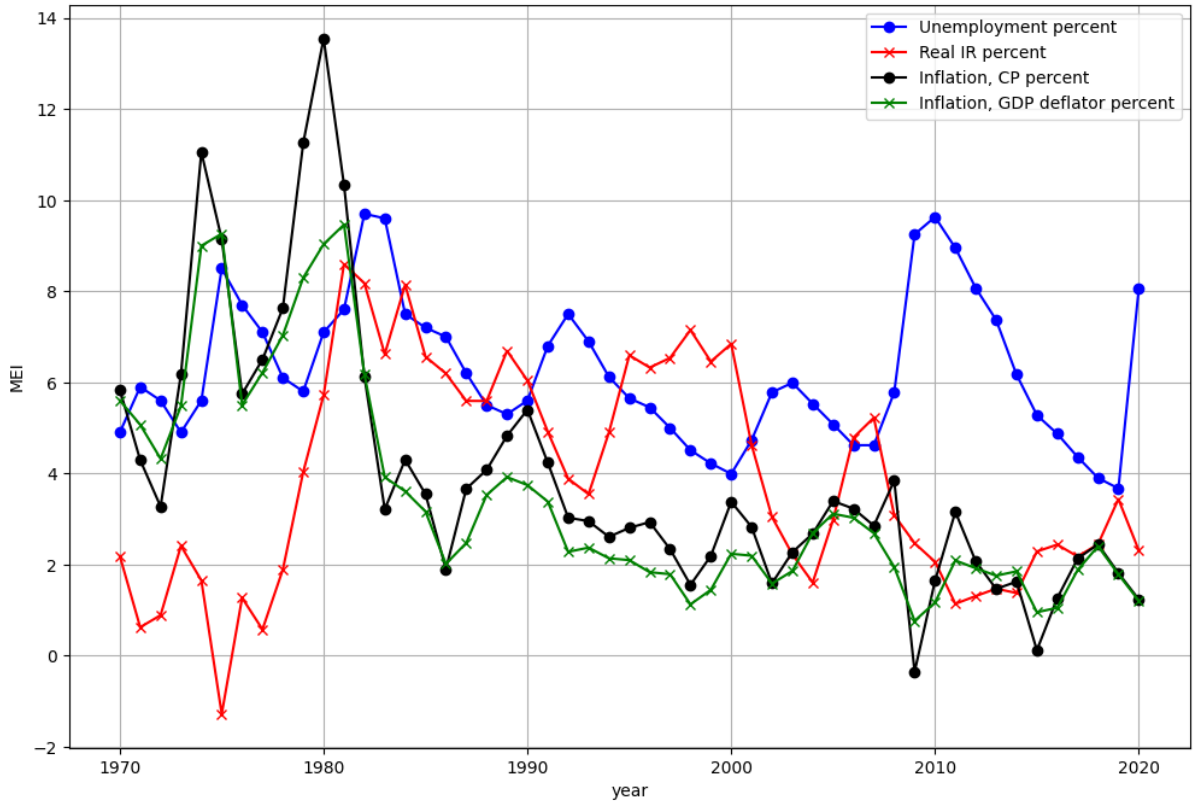

Fig 2: representation of all the variables over the years on the same plane.

The Heat Map provides a visual way to identify patterns, correlation. With each cell's color intensity patterns, variations within data. These are depicting the relationship between two categories of variables or visualization. Two dimensions distribution of values are there which make the data more exploratory & conveying patterns in a concise & accessible form. (Zhao, Shilin, et al.,2014)

- Strong positive linear relationship when correlation coefficient close to +1.
  We can also say that when one value increases other value also increase. +1 means a perfect positive correlation which means both variables shift in perfect harmony.

In the above two cases when the value of correlation is far from 1, be it close to 0 or be it close to -1, shows weaker relationship between the variables.

- Strong negative linear relationship when correlation coefficient close to -1.
  -1 means a relationship in which one variable tends to reduce/decrease & the other tends to rise/increase
- Little to no linear relationship when correlation coefficient close to 0.
  0 indicated that there is no relation between the variables.

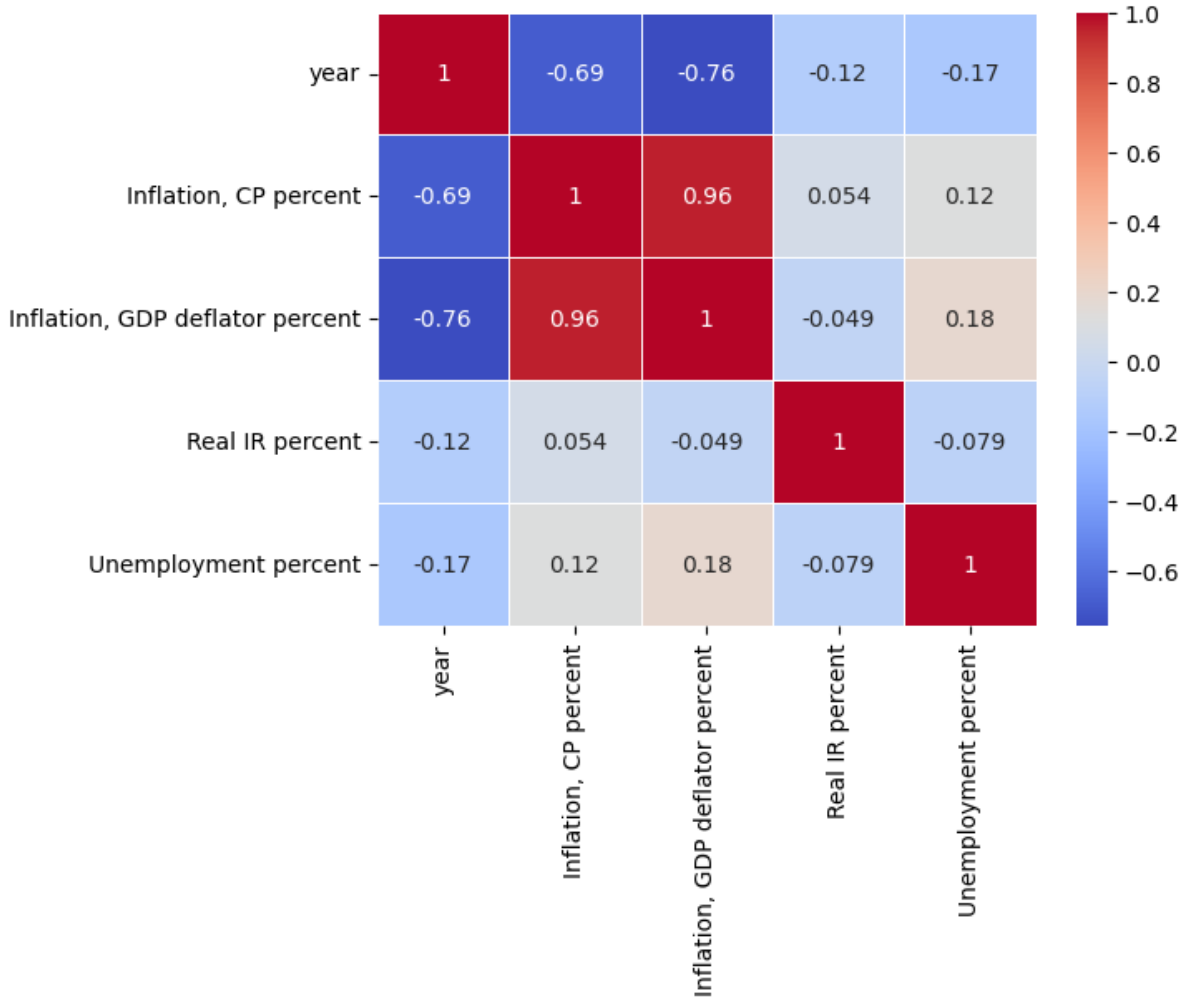

Fig 3: Heatmap representation of the variables.

Several ML models are used for the analysis, each having its own strength & weakness. Here are some of the models used:

1. Long Short-Term Model (LSTM RNN)
2. Random Forest (RF)
3. Support Vector Machines (SVM)
4. K- nearest neighbour (KNN)

Each model has its own potential, so we will figure out which model suits best for our analysis & will help in our findings. By including a data driven approach the aim is to unveil patterns, correlations & offer an understanding of the forces shaping the US economy over time. (Ayodele, Taiwo Oladipupo.,2010)

2. LITERATURE REVIEW

Being a very raw topic of discussion, this project does not have any exact previous work and reviews. Considering my macroeconomics variables the previous works on them has been examined by me. How the macroeconomic variables have been affecting each other and ultimately the growth of nation.

In the 1990 Canada was experiencing a economic underperformance period, in the first half. The economy had lower per capita real GDP and unemployment was high compared to that of 1930s & other developed countries. The Bank of Canada's monetary policy was identified as factor that contributed to the outcomes. With IR high which lead to lower spreading, output & employment. The inflation rate was low and unemployment was high compared to U.S. Author argued that the monetary policy of Canada was restrictive excessively & the primary cause of the economic difficulties was not fiscal policy. (Fortin, Pierre. CSLS, 2001.)

The dynamic interactions between unemployment, inflation & federal funds rate in the U.S. economy using a VECM is examined. In the short run, aggregate demand shocks & monetary policy shocks move unemployment & inflation in opposite directions, as expected in traditional fluctuations in economics. In the long run permanent supply shocks are the primary drivers of inflation & unemployment. The findings differed from the predictions of natural rate models predictions & suggest a propagation mechanism linking productivity shocks to inflation & unemployment at medium & low frequencies. There is also an interpretation of low frequency between unemployment & inflation over the last decades. The empirical analysis includes dynamic Phillips curve & uses data per month, with smoothed inflation data & an effective measure of expected inflation based on average value of inflation over past four periods. This study emphasized on unemployment in differences instead of levels which can be explained referring to micro founded models. The discussion is on identification of long run shocks & the impulse response functions, as well as the trade-offs between unemployment & inflation in short run. (Ribba, Antonio, (2006))

3. COMPARISON OF ABOVE MODELS

A number of things have impacted our decision that these models have been selected. And our aim is to find that which one is the best for our study. These models have been selected on the site of study of data, problem understanding, the constraints of the data, the accuracy. Now generally seeing which model is best when: (rjunaidraza.medium.com, 2020)

3.1 Speed

On the data's nature, dimension & size. For training Random Forest is slow, KNN is slower than LSTM. SVM is the slowest.

3.2 Memory

KNN is more memory intensive. It is also costly to train. It keeps the track of all the trained data. Many models are good with small datasets.

3.3 Flexibility

LSTM is less capable & flexible to capture more complex data relations. Both linear & non-linear solutions are supported by SVM. KNN is better when data has high SNR. Random forest is accurate & robust.

4. DATA METHODOLOGY

4.1 Data Collection:

The foundation of this study rests on a dataset taken from reputable databases to ensure accuracy & reliability. The dataset amalgamates information from various esteemed sources, harmonizing multiple economic indicators for a comprehensive analysis. CPI inflation data has been taken from Federal Reserve Economic Data (FRED). It is a trusted repository known for providing high quality data. GDP inflation, Unemployment Rate & Real Interest Rate (RIR) the main point of our study, was sourced from the world bank. (Fund, Nordic Trust.,2013)

To create a good dataset, data obtained from sources were integrated. There is a consistency in the units of measurement across variables. The dataset synthesis of diverse yet interrelated economic indicators, serves as the bed for our machine learning analysis.

4.2 Methodology

4.2.1. LSTM RNN

A Long Short-Term Memory (LSTM) is a type of artificial neural network that is capable of processing and predicting sequential data, such as time-series or text data. They are good for time series data forecasting due to its ability to retain information over extended sequence. LSTM models have been widely used in various applications, including stock price prediction. (Abdel-Nasser, Mohamed, and Karar Mahmoud.,2019)

4.2.2. Random Forest

It is technique for model's predictions & behaviour & is built on decision trees. The multiple decision trees helps improve the performance. This technique considers taking the one with the majority of selections as a selected prediction. It is useful for dealing with high-dimensional & complex data. (Rigatti, Steven J.,2017)

4.2.3. Support Vector Machines (SVM)

Support Vector Machines is used to solve classification & tasks related to regression. It is amazing solving binary classification problems, which classifies data sets in two sections. It is also used for non-linear relationships through kernels. (Joachims, Thorsten.,1998).

4.2.4. K-Nearest Neighbours (KNN)

K-Nearest Neighbours is the most basic algo in ML. It is non-parametric as it does not make any assumptions about distribution of data. It handles both numerical & categorical data, making it a flexible choice for classification & regression related tasks. It is less sensitive to outliers compared to the rest of the algorithms. (Xiong, Lei, and Ye Yao. (2021).

4.3 Modelling & Analyses

The collected data is pre-processed for the models scaling them correctly. The inputs are defined like CPI inflation, GDP inflation & Unemployment & the target variable as Real IR rate. Then the data is done in training & testing datasets. Then that data is processed & fed in the machine learning model i.e. Long Short-Term Memory (LSTM) RNN, Random Forest (RF), Support Vector Machines (SVM) & K-Nearest Neighbours (KNN) respectively. The

Mean Squared Error & Root Mean Squared Error metrics helps evaluate the performance of these models. This predicts the difference between actual & predicted values. (Deb, Sougata.,2016).
The observed data is shown below.

4.3.1 Long Short-Term Memory (LSTM) network

This model is trained for 50 epochs (iterations), with a batch size of 32. The Mean Squared Error (3.65). This means the difference between actual value & predicted value is 3.65 units. This shows that the degree of accuracy of the model is quite low, as evidenced in Fig 4. in which it shows actual vs predicted with Target variable on Y-axis & Data points at X-axis. The graphs are not together with both the actual and predicted values. This suggests that our model is not accurate & has low predicting power.

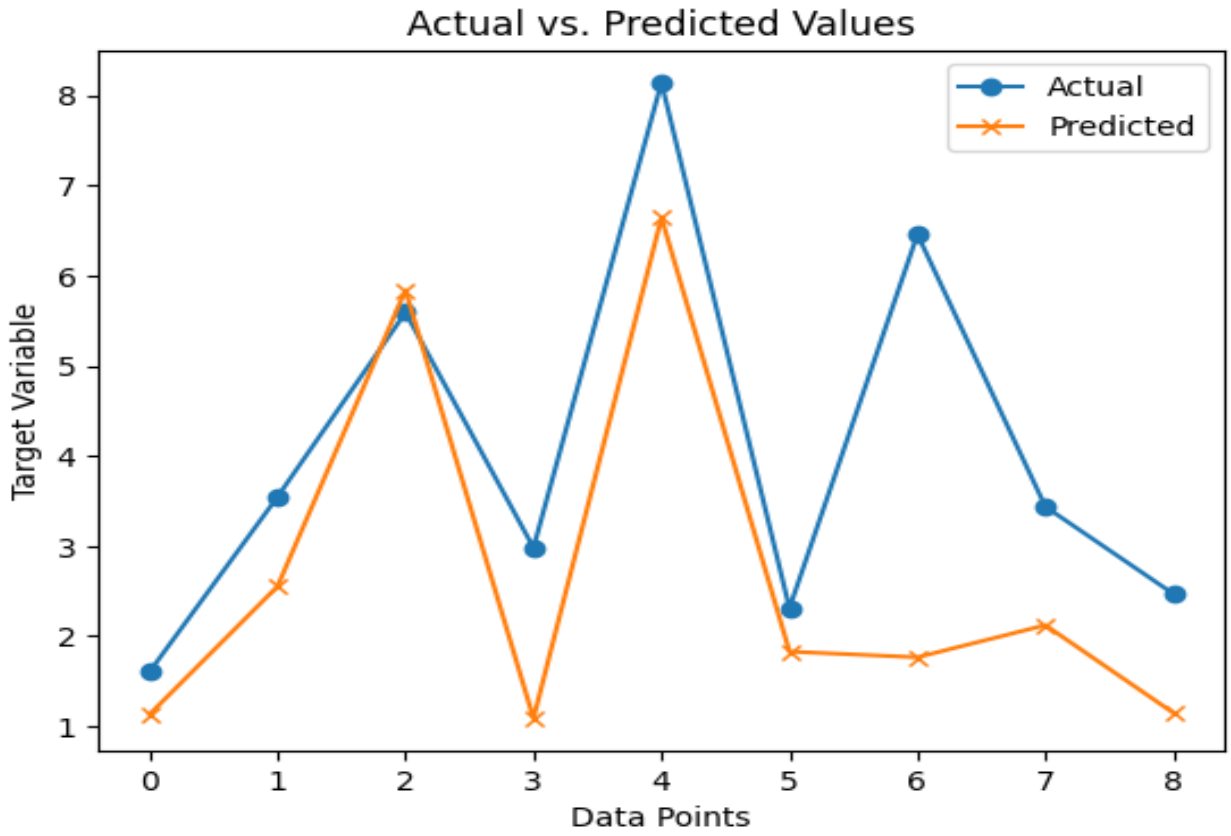

Fig 4: Predicted vs. Actual of the LSTM model

After observing the predictability of LSTM model, to search for the best Model we will shift to check this with Random Forest.

4.3.2 Random Forest (RF)

This model is trained for 100 epochs (iterations), with random state of 42. The Mean Squared Error (0.73). This means difference between actual value & predicted value is 0.73 units. The R-Squared value is (0.98). That shows the difference of 0.98 units. This shows that the degree of accuracy of the model is quite high, as evidenced the Fig 5. in which it shows actual vs predicted with Target variable on Y-axis & Data points at X-axis. The graphs are together of both the actual and predicted values. This suggests that our model is accurate & has high predicting power.

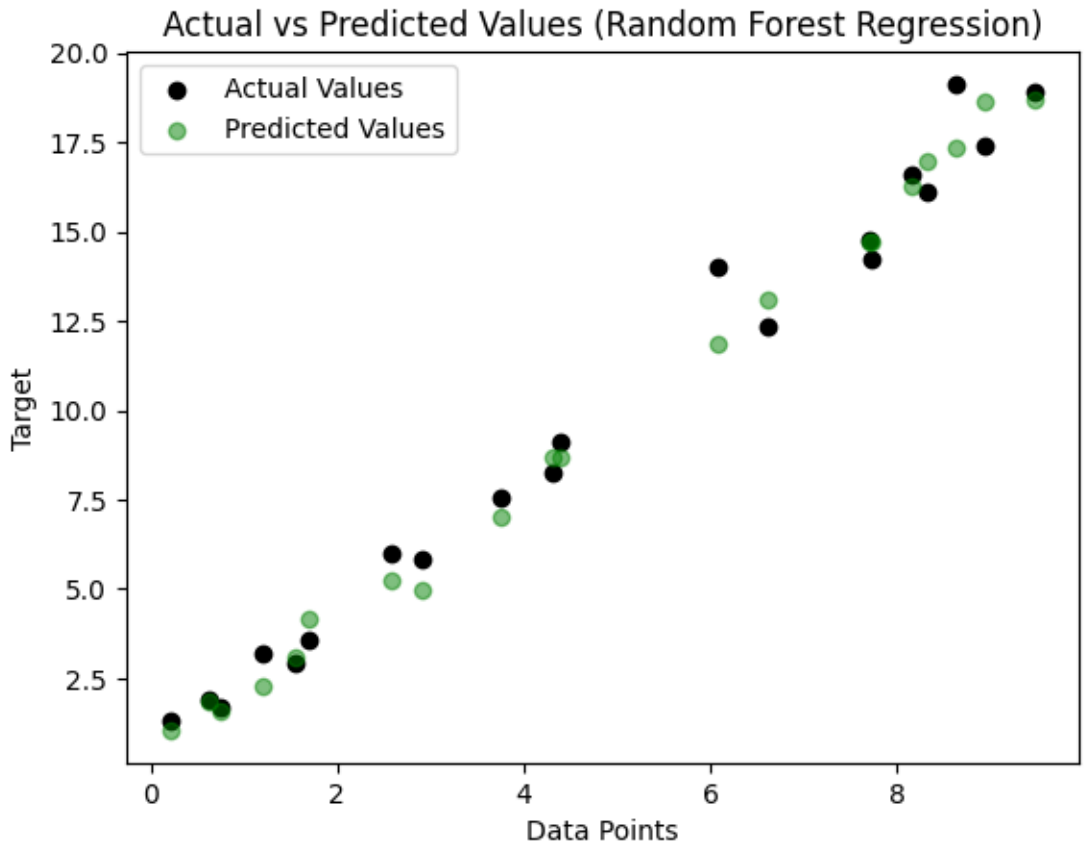

Fig 5: Predicted vs. Actual of the RF model

After seeing our prediction in RF model, we will check few more machine learning model & see how this model works in them.

4.3.3 Support Vector Machines (SVM)

This model is trained for 100 epochs (iterations), with random state of 42 & with linear kernel. The accuracy shown by the model is 0.95. This means our model is 95% accurate. The R-squared value is 0.4447. This shows that the degree of accuracy of the model is quite high, as evidence the Fig 6. in which it shows actual vs predicted with Target variable on Y-axis & Data points at X-axis. The graphs are together of both the actual and predicted values. This suggests that our model is accurate & has high predicting power.

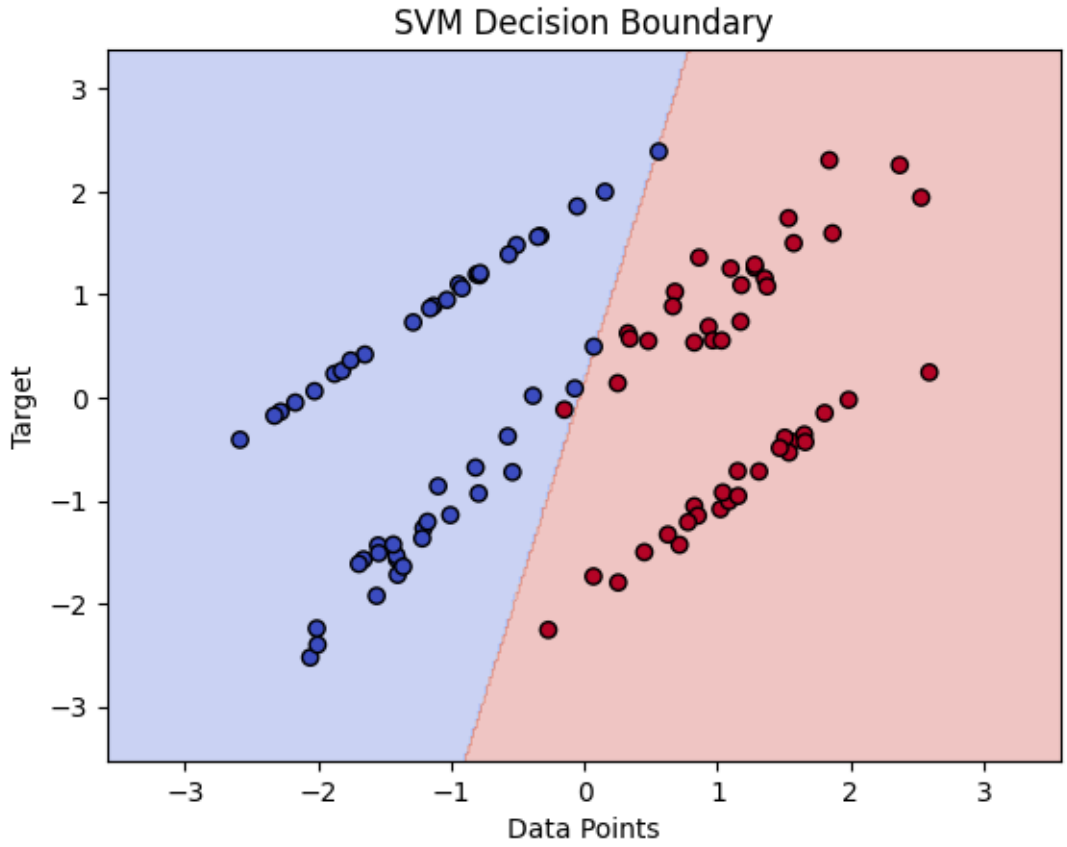

Fig 6: Predicted vs. Actual of the SVM model

Now we will shift our focus on checking the KNN model and see how our model turns out to be considering all the variables.

4.3.4 K-Nearest Neighbours (KNN)

This model is trained for 100 epochs (iterations), with random state of 42 & with linear kernel. The accuracy shown by the model is 0.95. This means our model is 95% accurate. The R-squared value is 0.4447. This shows that the degree of accuracy of the model is quite high, as evidence the Fig 7. in which it shows actual vs predicted with Target variable on Y-axis & Data points at X-axis. The graphs are together of both the actual and predicted values. This suggests that our model is accurate & has high predicting power.

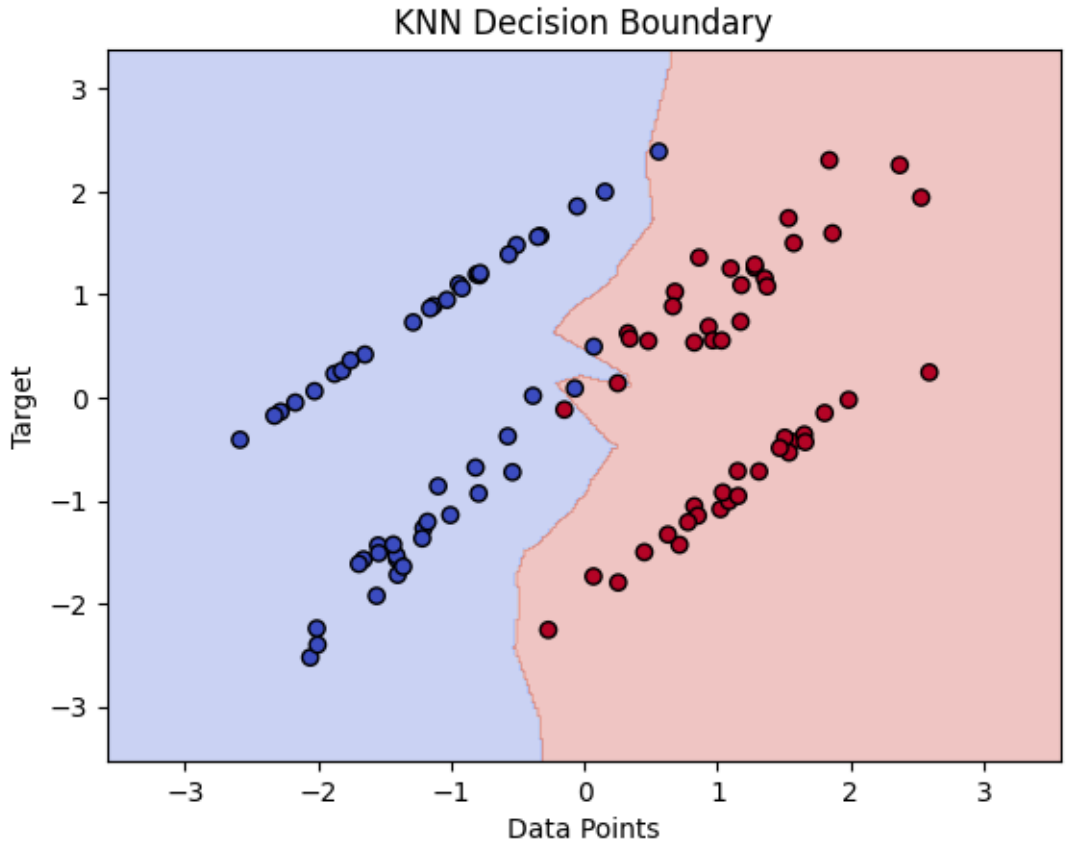

Fig 7: Predicted vs. Actual of the KNN model

Checking Best Fit Model

After analysing to check with model is the best for our study, all the models were again trained & the mean squared error value got calculated. Some findings showed that those models had the value at a greater distance from 0.

MODELS	MSE Value
LSTM RNN	5.3063755612520670
Random Forest	0.4356164047458618
SVM	5.8521724704085045
KNN	0.6522750233447270

5. Conclusion

In conclusion, we have demonstrated the model for our target variable Real IR rate using four machine learning models: Long Short-Term Memory (LSTM) RNN network, Random Forest (RF), Support Vector Machines (SVM) & K-Nearest Neighbours (KNN) respectively. The models have shown potential of the accuracy & predicting power. All the models take a unique route for modelling & explore the relationship between the variable. We can also say that there is no over fitting as the data does not show any unusual performance on testing.

So, in our models we observed how Support Vector Machines (SVM) & K-Nearest Neighbours (KNN) has same accuracy & R-squared values. The LSTM model was not the best fit model. When we compared the mean squared values under all the models we found that Random Forest is the best model compared to other models & has the least mean squared values. Machine Learning is very important specially when it comes to Economics. Firstly it helps us to find patterns & predict trends. Secondly there are many things that humans cannot reach & there ML helps us to think through. Thirdly it can give inferences in crucial areas in economic events & help economics understand the link between the variables more preciously. Fourthly we know how economics rotate around huge data so it can help us handle the data more easily. Lastly, in this project of mine the study was to find how the other macroeconomic variable influence the Real IR rate, and with our RF model we can predict as the difference between actual & predicted is very less. Also, the mean squared error value was the minimum for RF indicating that it is the best model. This model can help in real world as it uses time series data, so countries can predict there Real IR rate considering other variables. (Dittman, David, et al.,2011.)

Reference

- *(Ayodele, Taiwo Oladipupo. "Types of machine learning algorithms." New advances in machine learning 3 (2010): 19-48.*
- *(Abdel-Nasser, Mohamed, and Karar Mahmoud. "Accurate photovoltaic power forecasting models using deep LSTM-RNN." Neural computing and applications 31 (2019): 2727-2740.)*
- *(Deb, Sougata. "A novel robust r-squared measure and its applications in linear regression." International Conference on Computational Intelligence in Information System. Cham: Springer International Publishing, 2016).*
- *(Dittman, David, et al. "Random Forest: A reliable tool for patient response prediction." 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW). IEEE, 2011.)*
- *(Fund, Nordic Trust. "World Bank." Human rights impact assessments: a review of the literature, differences with other forms of assessments and relevance for development. Washington DC: The World Bank (2013).)*
- *(Fortin, Pierre. Interest rates, unemployment and inflation: the Canadian experience in the 1990s. CSLS, 2001.)*
- *(Joachims, Thorsten. Making large-scale SVM learning practical. No. 1998, 28. Technical report, 1998.)*
- *(Lucas Jr, Robert E., and Leonard A. Rapping. "Real wages, employment, and inflation." Journal of political economy 77.5 (1969): 721-754.)*
- *(Mishkin, Frederic S. The real interest rate: a multi-country empirical study. No. w1047. National Bureau of Economic Research, 1982.)*

- *(Pennacchi, George G. "Identifying the dynamics of real interest rates and inflation: Evidence using survey data." The review of financial studies 4.1 (1991): 53-86.)*
- *(Ribba, Antonio. "The joint dynamics of inflation, unemployment and interest rate in the United States since 1980." Empirical Economics 31 (2006): 497-511.)*
- *(Rigatti, Steven J. "Random Forest." Journal of Insurance Medicine 47.1 (2017): 31-39.)*
- *(Xiong, Lei, and Ye Yao. "Study on an adaptive thermal comfort model with K-nearest-neighbors (KNN) algorithm." Building and Environment 202 (2021): 108026.)*
- *(Zhao, Shilin, et al. "Advanced heat map and clustering analysis using heatmap3." BioMed research international 2014 (2014).)*