**Data Science Practicum 2**

**Lakshmi Deepika Pinnamaraju**

**[lpinnamaraju@regis.edu](mailto:lpinnamaraju@regis.edu)**

**Project Title:**

Modeling and Forecasting U.S. Inflation in R

**Project Description:**

Inflation refers to the sustained increase in the general price level of goods and services in an economy over a period of time. It can be measured by calculating the percentage change in the Consumer Price Index(CPI) or other price indices. Overall, inflation can have a significant impact on the economy, affecting the purchasing power of individuals and businesses and influencing investment and spending decisions. Due to various factors such as war in Ukraine, supply chain disruptions, changes in money supply, interest rates and few more have caused increase in inflation. This means there is rise in prices of various economic indicators such as oil, commodities, housing, education and many more.

The main purpose of this project is to model inflation in the United States using various economic indicators like oil prices, commodities prices, unemployment, federal funds rate, change in money supply and consumer sentiment index. After modeling the data, then it will be focused on forecasting inflation in the coming few months.

**Data:**

I want to collect the data from two sources : first source is The United States Bureau of Labor Statistics,  (<https://bls.gov/>) and next source of data is the St. Louis Federal Reserve’s economic data (<https://fred.stlouisfed.org/>). The data I have taken is from 1950’s to 2023 and it is a time series data. I want to model inflation using this data and to forecast inflation in the coming few months like to August 2023. Data will be of a moderate size not too big or small. It may take a week or more to gather the relevant data, perform data cleaning and merge it into a single dataset as per my project requirement.

**Data Analysis:**

I am using different economic variables to discuss about inflation and each variable in the final dataset will be on a monthly interval. After getting the relevant data, it will be cleaned and then merged to create a final dataset. After merging the data, data analysis will be done to explore the characteristics of the data and their relationship with inflation. After data exploration, it is time for modeling the data using linear regression and then the assumptions made by this linear model will be verified. As we are working with time series data, two other models, ARIMA(Auto regressive Integrated Moving Average) and NNAR(Neural Network Auto regressive) will also be used in data analysis as these both are time-series forecasting models. In both these models, different combination of variables will be tested and then they will be compared.The best models will be ensembled and then be used to forecast inflation in the next few months.

1. **Data Cleaning and Data merging into single dataset**
2. **Exploratory Data Analysis**
3. **Correlation Matrix**
4. **Data Visualizations**
5. **Linear Regression Model**
6. **Verifying Assumptions by Linear Model**
7. **ARIMA modeling:** ARIMA (Auto regressive Integrated Moving Average) is a statistical modelthat is used to analyze and forecast time-series data. ARIMA models use past values of a time series data to predict future values based on the relationship between the values and their historical trends. These models are often used for univariate time series data.
8. **NNAR modeling:** NNAR (Neural Network Auto regressive) is a type of artificial neural network that is specifically designed for time-series forecasting. NNAR models are based on the idea that time-series data can be modeled as a nonlinear function of its past values. These models can be used for both univariate and multivariate time-series data.
9. **Final model selection and Ensembling**
10. **Forecasting Results**

**Anticipated Difficulties:**

* Data Collection of various variables in a monthly interval from two different sources and merging them into single dataset might be difficult. I will overcome this by carefully collecting the data the way I want for my project and properly cleaning the data without missing anything and finally merge into a single dataset by year and month.
* Performing ARIMA and NNAR modeling to forecast U.S. Inflation for the next few months will be challenging for me as I have no idea or knowledge about them at all. But I will overcome this by researching and knowing more about them.

**Timeline:**

Week 1: Project Proposal

Week 2: Data Collection, cleaning and merging into final dataset

Week 3: Exploratory Data Analysis and Data Visualizations

Week 4: Linear Regression Model and Verifying Assumptions

Week 5: ARIMA modeling

Week 6: NNAR modeling

Week 7: Final model selection and Forecasting Results

Week 8: Adding entire project to Github Repository and Presentation

**Github Repository Link:**