Project Proposal: Foreign Exchange Rates Predictor

by
Sagar Padekar (20318463) & Pragya Madaan(17316889)
Masters in Computer Science (Future Networked Systems)
Machine Learning Module- Group 60

1. Motivation

Our project aims to assist international students like us to decide the best time for sending or receiving money from international transactions based on the predictions for foreign exchange rates. Further, the project will also be able to provide businesses with an expected timeframe for huge international business investments.

2. Dataset

We have gathered the raw data from the official website of International Monetary Fund. The same can be viewed at: https://www.imf.org/external/np/fin/data/rms_mth.aspx?SelectDate=2020-09-30&reportType=REP

Based on the dataset format, a conversion from .tsv to .csv will be performed by us. Further, as there are multiple datasets for each month exchange rates, we will be combining all those to obtain the daily exchange rates for the last 10 years. Pre-processing techniques will be then applied to the prepared data. We are planning to handle missing values using 'sklearn.preprocessing.impute', checking for outliers and scale the data to improve the prediction efficiency for our models. We will also visualize our data and check for its distribution, skewness etc. for further processing.

3. Method

As we will be predicting a continuous dependent variable, our data is suitable for performing regression techniques. Hence, we will be using the following Machine Learning models learned in our module for predicting the exchange rates and compare their predictions:

- Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Lasso Regression (L1 Regularization)
- Ridge Regression (L2 Regularization)

As our project is mainly focussed over time-series data analysis at a macro-level, we might also consider using ARIMA Model if time permits.

4. Intended Experiments

For evaluating the performance of our models, we will use metrics such as Mean Square Error, Mean Absolute Error, R2 Score through which we will find our model's accuracy and will compare them against a baseline model. We'll also perform Sliding Window and Forward Chaining schemas for Cross-Validation for the time-series data. We will optimize our hyperparameters using Grid Search and Random Search. Further, we plan to perform experiments for obtaining a best-fitted ensembled model from our trained models.