Nibedita Bal

Fall 2021_ Time series forecasting

Milestone 5: Update on individual progress

In this Traffic time series final project, I'm working on the 'Traffic volume' feature with 48,204 data points. Here I'm analysing the monthly traffic volume from the year 2012 to 2018. In the part of exploratory data analysis, I visualized the traffic pattern monthly basis where it is almost found that the traffic is high sometimes between June to September and around November to december. This is the time where most of the accidents happen. So in this project, I am trying to see whether traffic volume is going up actually.

Analysis: I converted the date time column to date and time and processed the data and fit it to different models.

Preprocessing: converted date time column to date and time. Also, splitted date into year, month and day, grouped traffic volume by month. Plotted histogram, scatter plot to observe overall data trend.

Model building: I tried the Linear model but found that it's not a good model because of the weak correlation. Then I plotted ACF, PACF and eacf to observe which model to select, however I ended up choosing AR(2) by looking at the acf and pacf plot. I also took the difference(first order) of the time series. I applied box-cox transformation to make the series stationary. To confirm with the AR2 model, I ran auto.arima and found that ARIMA(2,0,3) is the best model. However, the model building process is not stopped here. I'll do residual analysis to see if there is any auto correlation among residuals and if white noise is present. Next, I'll proceed with backtesting and Garch modelling. In addition, I'll also do forecasting for the next 1 or 2 years to observe the traffic volume pattern.