**Electricity Load forecast**

Nowadays electricity playing a major role. We are completely revolving around electricity. In this generation all the equipments are electrified and even now vehicles are also come under this. Forecasting the electricity for future would be a task. Here we have previous data to forecast the electricity load. By analysing the data we are going to predict the load .

Dataset containing columns like temperature, apparent temperature, dew point, wind speed, cloud cover, humidity date, and time, these are the information provided to predict the electricity load. Initially dataset has been imported along with all necessary libraries. Checking the dataset for null values for all columns, null values has been dropped. Features in the dataset was plotted to see the distribution of values using histogram plots. Distribution was checked whether it follows normal distribution, left skewed or right skewed. Feature transformation are applied to wind speed which requires to transform. To find the outliers in numerical columns, box plot was plotted and find out the outliers present in data and IQR method is followed to find outliers and replace the outliers with the lower and upper bounds values. Same has been plotted for crosschecking the numerical. Correlation matrix will help to see the correlation between features. Even though all features having good correlation, humidity only having negative correlation.

Month, weekday, time are fetch from date time using the pandas date time extraction method and descriptive statistics like mean, median, standard deviation, min ,max values are checked for numerical. Various plots are plotted using line and scatter plot for data exploration. Load vs. year and load vs. temperature shows the same scenario where the load is increasing the month of April i.e. the summer and it constantly start decreasing after the month of November. The peak of load will be in may and June. Temperature playing the major role in electricity prediction, as temperature increases the load also simultaneously increased. Load vs. humidity graphs clearly says that humidity in range of 0.1 to 0.3 will consume more energy. Dew point with the range of 0 -5 requires more load after that load is constant.

Time which has 96 values each value denotes 15 minutes interval. For one day we have 96 values, those values are label encoded with sklearn learn library label encoder function, though its categorical variable we opt for label encoding. To convert Weekday and month values, here we are doing one h0t encoding to get more clarity on data one hot encoding is helpful. Splitting the data for modelling, we have three years of data so 33 months of data for training and 3 months of data for testing purpose. Once splitted the data, scaling of data has been done using standard scalar. Standard scalar follows (X - xmin)/(xmax - xmin) Once data is scaled it will come in range of 0-1 . though models have high values and low values to normalise the data scaling is required.

Models like Linear regression, Decision tree, Random forest, XGboost are trained and Rscore values for each models are checked to pick the best model out of four model. All models Mean absolute error percentage are seen. Metrics to evaluate model has been tried and compared the RMSE values . All the test score , train score and MAPE values are appended in the dataframe for easy visualisation. All models are over fitted except Linear regression, because models which is performed in train data is failed to perform in test data. After checking the all the model scores the Linear regression performance is good and its not over fitted both test score and train score is almost 0.7. Finally concluded that linear regression model for further forecast of NaN values data.

Conclusion of this would be a best feature which helps to predict load is definitely would be temperature, apparent temperature. Based on the climatic condition more load of electricity consumption is happening in summer season in the month of May to September. Load consumption is quite less in weekends because holidays and high in weekdays. In other months except summer also we need electricity but load consumption of electricity is constant. We will buy electricity based on the above factors.