# Capstone Project Submission

**Name: Priyanka Shinde**

**Email:** shindepriya7709@gmail.com

**Contributor: Individual Project**

**Github Link:** https://github.com/priyankashinde-DS/Appliances_Energy_Prediction.git

**In this project we predict Appliance energy consumption for a house based on factors like temperature , humidity & pressure . In order to achieve this , we need to develop a supervised learning model using regression algorithms . Regression algorithims are used as data consist of continuous features and there are no identification of appliances in dataset.**

**After loading csv file into the pandas dataframe the data cleaning is the most important step. But, when I check the Nan values and duplicates their is no dupliactes and NaN values present in it.**

**In second step we check the distribution of dependent variable i.e Consumption.and the how the other independent variables like temperature,pressure etc. Impact on the target variable in visualization. Using IQR method we remove the outliers and using log transform normalize the data.**

**In data preprocessing we encoded the column and after converting all textual column into numeric we scale the datapoints using standared scalar.after scaling we split the dataset for train and test and bulid the baseline model.from the baseline model we choose few models for hyper parameter tuning.after that we get our final model which predict the energy consumption of appliances.**

**Following are the some important colusion we get from this project:-**

**1) Nine baseline models are created to give a general understanding of performance.**
**2) The Linear Regression model has performed extremely poorly. This is because there is no significant linear relationship between the dependable variable and the independent variable.**
**3) As compared to other models, the Random Forest Regression model has performed best with the highest R*2 scores in both Testing and Training.**
**4) It is also evident from the result that the Random Forest Regression model has more variance as the difference between testing and training r*2 is greater.**

**5) K-Nearest Neighbors Regression model, Support Vector Regression model, and Light Gradient Boosting Machine model also performed well and have low bias and variance compared to the Random Forest Regressor model.**
**6) A Gradient Boosting Regression model (GBM) or the Extreme Gradient Boosting Regression model (XGB) also poorly performed.**
**7) For hyperparameter tuning i used only three models based on the low RMSE SCORE.**
**8) We observe that the model Extreme Gradient Boosting Regression model (XGB) is the best with low RMSE and High R2 SCORE as compared to other models.**
**9) Monday has the highest total energy consumption as well as the highest mean energy consumption.**
**10) Tuesday has the lowest total energy consumption as well as the lowest mean energy consumption.**
**11) The consumption of energy from late night to early morning is very low. This is because appliances are used less during the night.**
**12) In the evening, energy consumption is highest.This is because appliances are used more during the evening.Windspeed:-On Average windspeed is higher during the day than the night and early morning.**
**13) All the temperature variables from T1-T9 and T_out have positive correlation with the target.**

**DriveLink:** https://drive.google.com/drive/u/0/folders/1b-fLdLRv0GFwhjs8p7QybVmnJCcvsnj2