**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Rohit Kumar Sharma [rohitraw6@gmail.com](mailto:rohitraw6@gmail.com) |
| **Please paste the GitHub Repo link.** |
| GitHub Link: - https://github.com/Rohitkumar1011/Appliance-Energy-Prediction |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Summary**  Energy consumption matters both to our environment and our economy. World needs energy to support economic and social progress and build a better quality of life. So, to reduce consumption of energy first we have to manage its usage. This generated a need to develop a supervised machine learning model that can predict Appliance energy Prediction. In this project we will be analysing the appliance usage in the house gathered via home sensors.  All readings are taken at 10 min intervals for 4.5 months. The goal is to predict energy consumption by appliances. In this first step we have pre-processed the data by removing non relevant features and then renaming the column. Renaming of the columns is done according to there measurement. We have checked null values but, fortunately there was no null value in this dataset and we haven’t done any operations.  After that I have done EDA by comparing various variables like Energy usage with pressure, temperature, Visibility, Windspeed, humidity. Exploratory data analysis is the crucial part of project because it helps in visualizing relation between all the independent and dependent variables.  Correlation heatmap is used to do feature selection all the features which were highly correlated are dropped. Dropping the features was challenging task because after dropping some features model testing is done again and again. After some trials I came to the conclusion that dropping 12 features can increase the accuracy.  Before fitting the data to the model standardization and normalization is done, it restricts the range of variable between 0 and 1. Standardization makes mean approximate to zero and standard deviation equal to one.  Dataset is divided into two parts with help of train and split module. Major part of the data is used for training. After that I have compared various models like Random Forest regressor, Gradient Boosting, Xgboost, LightBGm, SVR, Linear regression on the basis of Adjusted R^2 score.  By comparing them we came to the conclusion that Random Forest regressor is giving best score.  And after that I have done hyperparameter tuning this was bit longer process because it took lots of computational power and time. After some time, it gave optimal parameters that can increase accuracy, but accuracy was increased slightly. Random Forest regressor and Gradient Boosting regressor model is fitted. They both gave good results. Gradient Boosting score was 0.87 and Random Forest score was 0.9.  Finally, we come to the conclusion that Random Forest regressor model is best to use for prediction. |