Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Name	Email	Contribution
Sanket Chouriya (Solo Project)	Sanket.db72332@gmail.com	Research, Data preparation and cleaning, EDA, Visualization, Model training and implementation TechnicalDocumentation and Presentation.

Please paste the GitHub Repo link.

Github Link: https://github.com/Sanket7994/Appliance_Energy_Prediction_Project

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

We hereby present the summary of the project with the title "Appliance Energy Prediction Project" – Capstone Project 2

Problem Description:

The data set is at 10 min for about 4.5 months. The house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. Each wireless node transmitted the temperature and humidity conditions around 3.3 min. Then, the wireless data was averaged for 10 minutes' periods. The energy data was logged every 10 minutes with m-bus energy meters. Two random variables have been included in the data set for testing the regression models and to filter out non-predictive attributes.

Project Context:

The goal of our project is to predict the energy consumption of appliances in households based on the sensor data that we have from a random apartment and corresponding weather reports with help of Machine Learning Algorithms.

Project Preparation:

- As first step Performed Data Preprocessing and clean the data, handled the missing values.
- Performed EDA on different features in dataset, to observe trend and relation between features.

- After the EDA part, for model prediction split the data in train and test.
- Subsequently, in order to proceed with the classification, I used the StandardScaler, to standardize the data values into a standard format.

- Some of the most popular regression algorithms that I have used in this project:

  - Polynomial regression
  - Lasso Regression
  - Random Forest regressor
  - XGBoost regressor
  - ExtraTree Regressor

- And for Hyper parameter tuning I have used RandomizedSearchCV:

The RandomizedSearchCV class in Sk-learn serves a dual purpose in tuning this model. The class allows you to,

- Apply a random search to an array of hyper-parameters
- Cross-validate your model using k-fold cross validation

- After model implementation, ExtraTree Regressor had performed better than the other regressors, with R2 score of around 0.657 while overall model accuracy is 90%.

-