Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Yogesh. K, Yogeshiaf399@gmail.com Individual Project.
Please paste the GitHub Repo link.
Github Link:- https://github.com/Yogeshkrishn/Appliances-Energy-Prediction-.git
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

PROBLEM STATEMENT

Using different data sources and environmental parameters (indoor and outdoor conditions),specifically, data from a nearby airport weather station, temperature and humidity in different rooms in the house from a wireless sensor network and one sub-metered electrical energy consumption (lights) have been calculated. Our goal is to predict the energy use by appliances.

SUMMARY

This project aims to predict the energy consumption by home appliances. With the advent of smart homes and rising need for energy management, existing smart home systems can benefit from accurate prediction. If the energy usage can be predicted for every possible state of appliances, then device control can be optimized for energy savings as well. This is a case of Regression analysis which is part of the Supervised Learning problem. Appliance energy usage is the target variable while sensor data and weather data are the features.

APPROACHES INVOLVED

- Data Cleaning
  i) Checking null values
  ii) Checking data types
  iii) Detecting outliers
- Exploratory Data Analysis
  i) Univariate
  ii) Bivariate
  iii) Multivariate
- Feature Selection
- Preprocessing
  i) Transformation
  ii) Scaling
- Model fitting
  i) Simple Modeling
     a) Evaluation
     b) Feature Importance
  ii) Modeling with Feature selection
     a) Evaluation
  iii) Modeling with Ensembling and CV
     a) Evaluation
     b) Model Explainability

CONCLUSION

- As the first step, we understand the data & checked for null values, and outliers and performed EDA to get better understanding of variables .
- As part of data pre-processing, we performed feature scaling and outlier removal
- As so we have a Timestamp in our data, we needed to see the periodicity and trend of our dependent and independent variables.
- We tried multiple simple models and multiple advanced models with performed hyper parameter tuning and cross validation.
- Models Built: Linear Regression, SVR, RandomForest, Gradient Boosting XGBoost
- Advanced Models: Stacking Regressor, Voting Regressor, Average Ensemble
- Based on our targeted evaluation metric - RMSE and R2 scorel, we chose Stacking Regressor as the suggested model.