Capstone Project Submission

Team Member's Name, Email and Contribution:

S. No	Team Member	Email Id
1	Amrutha B.S	Amrutha123bs@gmail.com
2	Mahima Shree	mahimashree21@gmail.com
3	Purnima Rai	0204purnimarai@gmail.com
4	Tanmay Bohra	tanmaybohra@gmail.com
5	VikramPratap	vikramlearn@gmail.com
6	Pooja Rana	rana.pooja800@gmail.com

Please paste the GitHub Repo link.

Github Link:- https://github.com/Vikrampr/Appliances-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)

INTRODUCTION:

In this time of global unreliability world needs energy to support economic and social progress and build a better quality of life. In today's time there are many places particularly developing countries where there is interruption of services. These fallouts are primarily because of excess load consumed by appliances at home. Heating and cooling appliances takes most power in house. In this project we will be analyzing the appliance usage in the house gathered via sensor network. All readings are taken at 10 minutes' intervals for 4.5 months. The goal is to predict energy consumption by appliances. It is important to study the energy consumption in the residential sector and predict the energy consumption by home appliances as it consume maximum amount of energy in the residence.

PROBLEM STATEMENTS:

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. Weather from the nearest airport weather station (Chievres Airport, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru) and merged together with the experimental data sets using the date and time column. Two random variables have been included in the data set for testing the regression models and to filter out non-predictive attributes (parameters).

APPROACH:
- Problem Statement
- Data Description
- Data Wrangling
- Exploratory View – All Features, Target Variable
- EDA
- Correlation Matrix
- Feature Engineering
- Data Splitting & Standardizing
- Model Training & Hyper-Parameter Tuning
- Model Performance Comparison
- Conclusion
- Scope of Improvement

CONCLUSION:
- Dataset doesn't have any null values.
- We have observed very less co-relation between the target and feature variables.
- Dropped features like rv1 & rv2 as it has infinity VIF.
- Top 2 models were Extra Tree Regressor & Random Forest.
- Worked on Multi-Collinearity, but not much significant effect on the dataset.
- Tree based models are the best ones while dealing with features which has very less correlation with the target variable. Thus, the linear models, Ridge & Lasso performed the worst.

CONTRIBUTOR ROLE:

1. Amrutha B.S :

   ➢ Understanding the Data
   ➢ Worked on Temperature Analysis
   ➢ Worked on Regression Plots
   ➢ Worked on Data Pre-processing
   ➢ Worked on Linear Regression
   ➢ Story for Team Colab
   ➢ Worked on Cat-Boost (Just to check how it functions with the given dataset)

2. Mahima Shree:

   ➢ Understanding the Data
   ➢ Worked on EDA – All features
   ➢ Worked on Data Pre-processing
   ➢ Worked on KNN & SVR
   ➢ Worked on Hyper-parameter tuning – Random Forest
   ➢ Story for Team Colab
   ➢ Worked on Cat-Boost (Just to check how it functions with the given dataset)

3. Purnima Rai:

   ➢ Understanding the Data
   ➢ Worked on EDA – Irregular plots
   ➢ Worked on hourly usage of appliances
   ➢ Worked on Humidity Analysis
   ➢ Worked on Column Management
   ➢ Worked on Random Forest
   ➢ Worked on Feature Engineering – Reducing Multi-Collinearity
   ➢ Worked on PPT Presentation

- ➢ Worked on Technical Documentation
- ➢ Worked on Cat-Boost (Just to check how it functions with the given dataset)

4. Tanmay Bohra:

- ➢ Understanding the Data
- ➢ Data Wrangling
- ➢ Worked on week with highest energy consumption
- ➢ Worked on Data Pre-processing
- ➢ Worked on Extra Tree Regressor
- ➢ Worked on Hyper-parameter tuning – Extra Tree Regressor
- ➢ Worked on Feature Engineering – Reducing Multi-Collinearity
- ➢ Worked on Boruta Feature Selection (Could not implement due to time – constraints)
- ➢ Worked on PPT Presentation
- ➢ Worked on Technical Documentation
- ➢ Worked on Summary Document
- ➢ Worked on Cat-Boost (Just to check how it functions with the given dataset)

5. VikramPratap:

- ➢ Understanding the Data
- ➢ Data Wrangling
- ➢ Worked on overall energy consumption for the entire data
- ➢ Worked on Regression Plots
- ➢ Worked on Column Management
- ➢ Worked on Gradient Boost
- ➢ Worked on Technical Documentation
- ➢ Worked on Summary Document
- ➢ Worked on Cat-Boost (Just to check how it functions with the given dataset)

6. Pooja Rana:

- ➢ Understanding the Data
- ➢ Worked on Weekday v/s Weekend
- ➢ Worked on understanding the target feature and its outliers
- ➢ Worked on Co-relation matrix
- ➢ Worked on Column Management
- ➢ Worked on XGBoost
- ➢ Worked on PPT Presentation
- ➢ Worked on Technical Documentation
- ➢ Worked on Cat-Boost (Just to check how it functions with the given dataset)

Drive Link:

https://drive.google.com/drive/folders/1s4lUMz9D6rrSwCmdIIBp4a20DyvVhB6o?usp=sharing

Thank You
From- The team- Shivalik 01.