

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Name: Dipanshu Kumar

Email: dipanshukumar449@gmail.com

Contribution: Individual Project

Please paste the GitHub Repo link.

Github Link: <https://github.com/DipanshuKumar449/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)

Energy plays an essential role in economic and social development. It provides opportunities for nations to reach their goals of better living standards. So, planning and operating the energy production and energy consumption is a necessity these days. For better management, its being a requirement to understand the usage of energy. This provided an opportunity to develop a supervised machine learning algorithm to predict the Appliance Energy usage.

Based on our dataset, Data prepressing was the first step followed. I have understood the data found that the dataset fortunately is not having any null values, got a clear description about the features involved.

After that Exploratory Data Analysis and Data Visualization has provided a brief understanding about the relationship present between features and label i.e., the dependent variable also it has given an idea about the features to be selected for the further process.

Heatmap was used to understand the correlation between independent variables, based on which important features were selected. Selecting the correct features was challenging in order to get a better accuracy.

Before fitting the model, Standardization was an important step, it makes the feature values in the data have zero mean and unit variance. While we implement any Machine Learning algorithm it could be a possibility that objective function will not work properly without normalization.

After training and testing was done. I have made the use of Linear, Lasso, Ridge, Elasticnet, Gradient Boosting, Random Forest, XGBoosting techniques. Checked and compared various matrices and came to conclusion that Random Forest Regressor is giving the best score.

Chosen The Random Forest Regressor and did Hyperparameter Tuning for better accuracy and to reduce overfitting, I have made the use of Gridsearch cross validation to achieve the best parameter. These parameters enhanced the predicting capability of our model.

Finally Model Explainability using ELI5 was done which helped us to understand the involvement of features and their impact on the target variables.

In conclusion I would say that Random Forest Regressor have proven to be the best model for our dataset.

Please paste the drive link to your deliverables folder. Ensure that this folder consists of the project Colab notebook, project presentation and video.

<https://drive.google.com/drive/folders/1aBDdgnRgUp4PuXPPbDr9WCJWCH7Q2pVV>