**This file consists of Python source code for Residential Energy Appliance Classification.**

Authors(Group No. 23):

Srinithi Punniamurthy, Alonso Gomez Zeballos, and Chaitanya Srinevas

The directory contents of this distribution are as follows:

* group23\_ass2\_impl.ipynb : Python Jupyter Notebook to run the classification model
* pred\_labels.csv : File containing predicted labels for various appliances after running the model present in “group23\_ass2\_impl.ipynb”.

**Environment Details:**

· Python Version: Python 3.0 or higher

· Anaconda Distribution Version: 4.9.2 or higher

· Jupyter Notebook Version: 6.0.1 or higher

Packages required in the JuPyter Notebook:

* TSfresh: For Time series processing
* Shap: Feature Importance
* Seaborn- for visualization
* Pandas - data manipulation
* Numpy - mathematical operations
* Sklearn: developing machine learning models.
* Matplotlib- for visualization
* os : Accessing to directory
* Mpl\_toolkits
* Scipy- scientific operations, statistic and correlations
* Xgboost: XG Boost Machine Learning Model

**Instructions for setup and running the code:**

1. Import the required libraries and packages as mentioned above.

2. The training data file is named train\_data\_withlabels.csv and the test data is named as test\_data\_nolabels.csv .

3. Run the code present in group23\_ass2\_impl.ipynb file which generates a csv file of predicted labels.

4. Tsfresh, and the models take time for execution around 15-20 minutes.

5. The output file is named as pred\_label.csv

**Input Data:**

1. The data used for training the model is “train\_data\_withlabels.csv”.
2. The data in “train\_data\_withlabels.csv” consists of following three columns:

* Load
* Ac- Appliance 1
* Ev - Appliance 2
* Oven- Appliance 3
* Wash- Appliance 4
* Dryer- Appliance 5
* Hourofday
* Dayofweek
* Dif-Difference between two sequential load data points
* Absdiff- Absolute Difference between two sequential load data points
* Max: it depicts the largest mean shift between two consecutive windows
* Var: Variance of load over a neighbourhood time window of 30 minutes around each load data point.
* Entropy: Measures forecastability and uncertainty in time series
* Nonlinear
* Hurst: measure of the long-term memory of a time series,