DJS-Compute

Machine Learning Assignment-
6

Topic:
DECISION TREE, BAGGING & BOOSTING

General Instructions:

- The deadline for completing the tasks is Feb 17, 2024. A discussion session, addressing any doubts and task-related topics, will be conducted in the following days.
- You can refer to additional online resources if needed.

DECISION TREE & RANDOM FOREST

Task:

Dataset Link:
Dataset

Colab Link:
https://colab.research.google.com/drive/15yYyqatntfFs9SEOOBKMh0vTGhYj17an?usp=sharing

BOOSTING (XGBoost)

Task-1:

Dataset Link: https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption?r esource=download&select=AEP_hourly.csv
Download the AEP_hourly dataset and perform the following task, you can also use other datasets for your own practice

Colab Link:
https://colab.research.google.com/drive/1PLEJw6WHRwlqdXJgwr8XDwcPc7YVcLKg#scrollTo=ntiwtBTZgr4P

Create a copy of the colab link and follow the instructions to fit the model.

BAGGING

Task-1:

Dataset Link: https://drive.google.com/file/d/1zGgO1P7eCw2K9HjMxFb-wtbVy7ChA1EN/view? usp=sharing

The dataset is already cleaned and processed and also scaled. But the catch is the dataset is imbalanced so treat it before feeding it to the ML model by appropriate method. Apply bagging (random forest) to the following dataset. Mention the metric you would focus on to give the results in this type of imbalanced environment and give reasons.