ML Predicting Task

General Instruction :

- Please go through the instructions of questions carefully and submit your answers through the submission link.
- Do not trim or shuffle the test data or else your submission will be rejected.
- Carefully rename your answers before submitting, as mentioned in the instructions.
- You can approach questions in any order.

Submission Link:
https://docs.google.com/forms/d/e/1FAIpQLSfdYrwqp8LKRanGXk-jyOX-uvmmiRvl-cIgvsiuGoXmSRENsw/viewform?usp=sf_link

1. Classification Problem

A dataset with 10 attributes and 8.3k instances was given. All data points are classified into 12 classes. Build a machine learning model that can classify the test data more accurately.

Participants are free to use any machine learning algorithm that is suited for the provided data.

Data Provided :

- Dataset classified into 12 classes.
- **Train Data :**
    - Training dataset contains 8.3k data points and 10 attributes.
    - Download Link : https://drive.google.com/file/d/14WonmlfGc_lfFMEy7hiWK_QgWUADV_ap/view?usp=sharing
- **Test Data :**

- Test data contains 2k data points.
- Participants have to predict the most precise class for the given input test data.
- Download Link : https://drive.google.com/file/d/1MjiU0ugxc2YmSNae7NLuqYL1zAGSfAho/view?usp=sharing

Instructions :

- Participants would build a model using training data provided. And submit ipynb file(ParticipantID_classification .ipynb)
- Predicted precise class for test data and submit it as (ParticipantID_classification.csv)

Upload Link : Add upload link here

2. Regression Problem

Participants are expected to build a machine learning model for the given regression dataset. Dataset might be, Some amount of exploratory data analysis (EDA) would need to be performed.
Participants are free to use any machine learning technique to solve the problem.

Data Provided :

- **Training Data** :
  - The Training Dataset contains 5.8k instances and 150 dimensions.
  - Download Link : https://drive.google.com/file/d/1-CWo3fRm8yG1wsFMagBf8Yipntrflp9w/view?usp=sharing
- **Testing Data :**
  - Participants have to predict the most precise output value(y) for the given input test data.
  - Download Link : https://drive.google.com/file/d/1byvSq6FoX8WqBuAszvzCKsuG4l2c4R3V/view?usp=sharing

Instructions :

- Sufficient data cleaning, pre pre-processing must be done to achieve good results.
- Participants must add required comments in the code.
- Participants should build a model and upload the model (ParticipantID_regression .ipynb jupyter notebook).
- Predicted output values(y) for test data and submit it as (IParticipantID_regression.csv)
- Root-mean-square error(RMSE) would be used for evaluating model performance.

3. Forecasting given time series problem

Participants are expected to build a model for a given multivariate time series data. Dataset might be containing missing values and outliers. Some amount of exploratory data analysis (EDA) would need to be performed.
Participants are free to use any technique provided it is suited for the variety and volume of data provided.

Data Provided :

- Hourly data of 16 features were extracted and stored.
- **Training Data** :
    - Hourly data from 01 April 2004 to 16 Feb 2005 is publicly available as a training dataset.
    - Download Link : https://drive.google.com/file/d/1L2dlDAVTBxjnjTQHhw3ojf-OFE0k1OU-/view?usp=sharing

- **Participants Test Data** :
    - Hourly data from 17 Feb 2005 to 26 Feb 2005 is publicly available as a test dataset. Use it to evaluate the model from the participants' side.
    - Download Link : https://drive.google.com/file/d/153SjlM0P9jQm3i1AEuTfCIiuI2mfDyEV/view?usp=sharing

- **Final Evaluation Test Data** :

- Hourly data from 27 Feb 2005 to 08 March 2005 will be used for final evaluation. Participants would need to forecast it and have to submit it as a .csv file.
- Link : https://drive.google.com/file/d/1opP6bUnd4P0VpWKJ7AgbOWvFZcovmHWG/view?usp=sharing

Instructions :

- Sufficient data cleaning, pre pre-processing must be done to achieve good results.
- Participants must add required comments in the code.
- Participants would build a model and upload the model (ParticipantID_forecasting.ipynb jupyter notebook) .
- Evaluate the model with given participants test data and submit your Mean absolute percentage error(MAPE).
- And train the same model with both Training Data + Participants Test Data and forecast Final test data given above. The final forecasting should be submitted in .csv format(ParticipantID_forecasting.csv).