WEEK 1

Resources

Please refer to these links below for tutorials needed to approach this week's Problem Statement:-

Numpy: https://www.youtube.com/watch?v=pJoYVxgl840

Pandas: https://www.kaggle.com/learn/pandas

Data visualization: https://www.kaggle.com/learn/data-visualization

Intro to ML: https://www.kaggle.com/learn/intro-to-machine-learning

Intermediate Machine Learning:

https://www.kaggle.com/learn/intermediate-machine-learning

<u> Assignment - 1</u>

Deadline - 20th December 2022

PROBLEM STATEMENT

The problem is a classification problem. Try to solve this problem using algorithms like XGBOOST/ KNN CLASSIFIER/DECISION TREE / Naive Bayes or Logistic Regression. However, we also think that this problem could be efficiently solved using Neural Networks.

Your submissions would be evaluated based on:

- the evaluation metrics used (area under the ROC curve between the predicted probability and the observed target.)
- the data pre-processing techniques to be done before moving onto the prediction stage (basic EDA, feature selection/elimination/transformation)

Submission Guidelines: Upload your COLAB/jupyter Notebook file(.ipynb file), which clearly separates each of the steps involved in your prediction pipeline and provides explanations for each of the steps taken in Markdown in this part

Submission Link: https://classroom.github.com/a/mLiXX631

NOTE:- For each id in the test set, you must predict a PROBABILITY of failure. The file should contain a header and have the following format:

id	failure
26570	0.2
26571	0.1
26572	0.9

LINK TO THE DATASET

- https://drive.google.com/drive/folders/1xl_8KHIJrywdc6sUEgmuhgv-4D5uS6eB?usp=share_link
- This data represents a product quality assessment. For each "product_code" you are given a number of product "attributes" (fixed for the code) as well as a number of "measurement" values for each individual product, representing various lab testing methods. Each product absorbs a certain amount of fluid (denoted by the column -"loading") to see whether or not it fails.
- Your task is to use the data to predict individual product failures of new codes with their individual lab test results.
- FEATURES TYPES : both numerical and categorical
- TARGET COLUMN: failure