Final Project

Classification Project

For the final project, you are allowed to pick your own classification dataset from various resources and create your own project. The required aspects are:

Data

You can find any dataset is the available repositories, such as:

- data.world
- Kaggle
- Papers with code

or any other source available. The dataset should include:

1. At least 500 samples
2. Each item should have at leaset 100 features/attributes
3. The classification problem can be either binary or multiclass.

Goal

With this project it is expected to have the highest possible correct classification percentage. In order to achieve that you are expected to perform attribute selection (note: cross-validation in attribute selection is also required), and then go for classification with the selected attributes.

Classification Algorithms

Project Groups (should be between 1 to 3 people) are free to use **ANY** classification algorithm/technology that can be found in the literature. **ANY** programming language and platforms including machine learning packages **except WEKA, KNIME, and MATLAB** can be used. If a group programs the project, the executable build is requested.

Performance Measures

Sensitivity, *specificity*, *Precision*, *F1* score and *AUC* is requested as the output of the program performance. If a group programs the project from scratch, AUC is not necessary.
AUC = Area under the ROC curve.
DEADLINE IS JANUARY 22ND, 2024. Return your reports to: nalbantoglu@odev.erciyes.edu.tr.
Your reports should include **Your problem, your dataset spects (number of samples, number of features, types of features, target classes)**, **Your solution: the algorithms you used and the experimental setup**, and **Your results: the test results based on the given performance metrics.**
Good Luck!