[2025] Machine Learning Projects – Milestone 2

The objective of the projects is to prepare you to apply different machine learning algorithms to real-world tasks. This will help you to increase your knowledge about the workflow of the machine learning tasks. You will learn how to apply pre-processing, feature engineering, regression, and classification methods.

- **Delivering Milestone 2: Practical exam.**

- You must deliver a detailed report for milestone 2 contains all your work in this phase. Combine both reports and deliver a complete report for the project (Hardcopy).

- Each team should work on their project's updated dataset for milestone 2.

- **In the practical exam**:

  - We will give you two unseen test sets, **one for regression and one for classification.**

  - Make sure you **save your trained model** and create a test script that takes the new csv file, **loads the saved models,** and outputs predictions. This is to allow us to test your model without re-training.

    Hint 1: You can use libraries such as 'pickle' to save and load your models.
    Hint 2: Any model that you need to 'fit' or 'learn' during training means you need to save it and reload it for the test to work correctly.

  - You should be able to handle missing values for features in a test sample. (You can't drop an entire test sample row).

- You must Show the MSE and R2 score of the regression models and the classification accuracy of each classifier on the test set.

- Each team member will be graded individually according to their response to the oral questions related to their project.

➢ In the second milestone, you will apply the following: -

Classification:

- Split your dataset into 80% training and 20% testing.

- Train at least 3 different models to classify each sample into distinct classes.

- Choose at least two hyperparameters to vary. Study **at least three different choices** for each hyperparameter. When varying one hyperparameter, all the other hyperparameters should be fixed.

- [**Extra Requirement Mandatory for Teams of 6 Only**]: Apply (heteregenous) ensemble learning using different machine learning models to get the output. You should try both voting and stacking approaches.

(Note: Ensemble methods based on the same base model e.g. random forest will not be counted as doing the extra task)

Milestone 2:

➢ Classification and Hyperparameter tuning.

Milestone 2 Report Must Include:

❖ Summarize the **classification accuracy**, **total training time**, and **total test time** using three bar graphs.

- Note that your **Feature Selection** process may differ in this phase (classification) than the previous (regression), If so, explain your feature selection process and how it was proved or disproved.

- Explain in details how **hyperparameter tuning** affected your models' performance.

- Finally, write a **conclusion** about this phase of the project and what intuition you had about your problem and how it was proved/disproved.

Project(3): Parkinson's Disease Prediction

An **updated dataset** will be provided for each project in the second milestone.

Updated Dataset Snapshot:

Cholester	Cholester	UPDRS	MoCA	Functional	DoctorInC	WeeklyPh	MedicalHisto	Symptoms	Diagnosis
25.542044	237.29080	4.1616200	28.626479	5.3550554	DrXXXCon	4:14	{'FamilyHisto	{'Tremor': 'No', 'Ri	0
23.0981	150.13	176.22040	20.310768	9.9279976	DrXXXCon	0:59	{'FamilyHisto	{'Tremor': 'No', 'Ri	0
66.076196	66.871416	133.281	20.614059	5.7043076	DrXXXCon	5:38	{'FamilyHisto	{'Tremor': 'Yes', 'F	1
41.725854	248.16348	155.95202	4.2376960	7.2504347	DrXXXCon	5:02	{'FamilyHisto	{'Tremor': 'No', 'Ri	1
23.251948	127.74769	49.523001	21.475758	6.1191300	DrXXXCon	0:08	{'FamilyHisto	{'Tremor': 'No', 'Ri	0
71.763342	88.026496	82.731035	25.411267	9.7364274	DrXXXCon	1:53	{'FamilyHisto	{'Tremor': 'Yes', 'F	1
99.203744	314.08643	36.223156	28.442607	8.7604145	DrXXXCon	1:53	{'FamilyHisto	{'Tremor': 'No', 'Ri	0
42.278671	322.18642	120.20836	6.8956536	0.8634021	DrXXXCon	2:10	{'FamilyHisto	{'Tremor': 'Yes', 'F	1
92.713900	127.16783	122.03778	23.416523	9.33717	DrXXXCon	0:18	{'FamilyHisto	{'Tremor': 'No', 'Ri	1
90.350296	359.08665	121.55887	26.580741	5.9778049	DrXXXCon	8:48	{'FamilyHisto	{'Tremor': 'No', 'Ri	0
90.496007	88.871618	149.53111	8.6615604	3.8288200	DrXXXCon	6:29	{'FamilyHisto	{'Tremor': 'Yes', 'F	1

Updated Dataset Description:

- A New column is added **"Diagnosis"**. **Diagnosis** can be one of two values {0 or 1} referring to Yes or No.

Milestone 2 Classification task:

Classify diagnosis into one of two categories: {0 or 1} using **the updated dataset.**