ML 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 provided updated dataset for milestone 2 (not the original dataset).

- **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' 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 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 can be fixed.

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 detail 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.

Car Price Prediction Dataset Update

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

Updated Dataset Snapshot:

symb	CarName	fueltype	aspiration	doorn	carbody	drive	engin	wheelbase	carlength	carwidt	carheight	curbweigh	enginetyp	cylinc	engines	fuelsysten	borerati	stroke	compressi	horsepow	peakrpm	citymp	highwaym	category
3	alfa-romei	gas	std	two	convertibl	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	High
1	alfa-romei	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823	ohcv	six	152	mpfi	2.68	3.47	9	154	5000	19	26	High
3	alfa-romei	gas	std	two	convertibl	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	High
2	audi 100 l	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337	ohc	four	109	mpfi	3.19	3.4	10	102	5500	24	30	High
1	audi 100ls	gas	std	four	sedan	fwd	front	105.8	192.7	71.4	55.7	2844	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	High
2	audi 100ls	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824	ohc	five	136	mpfi	3.19	3.4	8	115	5500	18	22	High
1	audi 4000	gas	turbo	four	sedan	fwd	front	105.8	192.7	71.4	55.9	3086	ohc	five	131	mpfi	3.13	3.4	8.3	140	5500	17	20	High
1	audi 5000	gas	std	four	wagon	fwd	front	105.8	192.7	71.4	55.7	2954	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	High
0	audi 5000	gas	turbo	two	hatchback	4wd	front	99.5	178.2	67.9	52	3053	ohc	five	131	mpfi	3.13	3.4	7	160	5500	16	22	High
2	audi fox	gas	std	two	sedan	fwd	front	99.8	177.3	66.3	53.1	2507	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	High
0	bmw 320i	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2395	ohc	four	108	mpfi	3.5	2.8	8.8	101	5800	23	29	High
2	bmw 320i	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2395	ohc	four	108	mpfi	3.5	2.8	8.8	101	5800	23	29	High

Updated Dataset Description:

- The **"price"** column used in the previous milestone as the actual output has been removed.
- A New column is added **"category"**. A car can belong to a price category which can be {Low or High}.