

# ML Projects (CS) – 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. The link for the **updated dataset for each project** will be announced on MS Teams.
  - **In the practical exam:**
    - We will give you two unseen test sets, **one for regression and one for classification.**
    - In case of the movies dataset you will receive three csv files for regression and three csv files 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 should 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 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(1): **Player Value Prediction**

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

### Updated Dataset Snapshot:

CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN
RM	LWB	LDM	CDM	RDM	RWB	LB	LCB	CB	RCB	RB	PlayerLevel
63+2	57+2	51+2	51+2	51+2	57+2	56+2	50+2	50+2	50+2	56+2	C
43+2	54+2	61+2	61+2	61+2	54+2	58+2	71+2	71+2	71+2	58+2	A
68+2	53+2	48+2	48+2	48+2	53+2	51+2	47+2	47+2	47+2	51+2	A
65+2	70+2	68+2	68+2	68+2	70+2	70+2	70+2	70+2	70+2	70+2	A
59+2	62+2	58+2	58+2	58+2	62+2	61+2	58+2	58+2	58+2	61+2	C
53+2	39+2	35+2	35+2	35+2	39+2	37+2	34+2	34+2	34+2	37+2	C
66+2	50+2	46+2	46+2	46+2	50+2	47+2	40+2	40+2	40+2	47+2	A
69+2	72+2	70+2	70+2	70+2	72+2	72+2	70+2	70+2	70+2	72+2	A
63+2	64+2	67+2	67+2	67+2	64+2	64+2	66+2	66+2	66+2	64+2	C
63+2	66+2	63+2	63+2	63+2	66+2	65+2	62+2	62+2	62+2	65+2	B
66+2	67+2	68+2	68+2	68+2	67+2	65+2	63+2	63+2	63+2	65+2	B
72+2	72+2	73+2	73+2	73+2	72+2	71+2	71+2	71+2	71+2	71+2	S
57+2	62+2	65+2	65+2	65+2	62+2	63+2	67+2	67+2	67+2	63+2	C
72+2	60+2	58+2	58+2	58+2	60+2	57+2	48+2	48+2	48+2	57+2	S

### Updated Dataset Description:

- The “**value**” column used in the previous milestone as the actual output has been removed.
- A New “**PlayerLevel**” column has been added instead. Each player can have a level that is either {S, A, B, C or D}.

### Milestone 2 Task:

Classify a player into one of five levels: S, A, B, C or D based on the provided features in **the updated dataset**.

## Project(2): **Movie Revenue Prediction**

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

### Updated Dataset Snapshots:

movie_title	release_date	genre	MPAA_rating	MovieSuccessLevel
Recess: School S	16-Feb-01	Comedy	G	C
D2: The Mighty	25-Mar-94	Comedy	PG	B
Home on the Range	2-Apr-04	Comedy	PG	C
Young Black J	25-Dec-03	Adventure	G	D
What's Love Got	9-Jun-93	Drama	R	B
Lady and the Tr	22-Jun-55	Drama	G	S
Corky Romano	12-Oct-01	Comedy	PG-13	C
The Waterboy	6-Nov-98	Comedy	PG-13	A
Captain America	4-Apr-14	Action	PG-13	A
Pirates of the	7-Jul-06	Adventure	PG-13	S

### Updated Dataset Description:

- The “**revenue**” column used in the previous milestone as the actual output has been removed.
- A New “**MovieSuccessLevel**” column has been added instead. Each movie can have a success level that is either {S, A, B, C or D}.

### Milestone 2 Classification task:

Classify each movie (row) into one of five categories {S, A, B, C or D} based on the provided features **in the updated dataset**.