

[2024] Pattern 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.
- **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 (heterogenous) 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 (1): Apartment Rent Prediction

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

Updated Dataset Snapshots:

amenities	bathrooms	bedrooms	currency	fee	has_photo	pets_allow	RentCategory	price_type	square_feet	address	cityname
Clubhouse	2.5	2	USD	No	Thumbnail	Cats,Dogs	Medium-Priced Rent	Monthly	800		Cary
Cable or Satellite	1	1	USD	No	Thumbnail	Cats,Dogs	Medium-Priced Rent	Monthly	795	640 West \	Madison
Basketball	1	1	USD	No	Thumbnail	Cats,Dogs	Medium-Priced Rent	Monthly	560	2777 SW A	Gainesvill
located at	1	1	USD	No	Thumbnail	Cats,Dogs	Low Rent	Monthly	600	534-542 P	Omaha
located at	3	3	USD	No	Thumbnail		High Rent	Monthly	1600	2216 S Pal	Los Angel
located at	2	3	USD	No	Thumbnail		Medium-Priced Rent	Monthly	2300	680 Broml	Baton Rou
Gym,Pool	1	2	USD	No	Yes		High Rent	Monthly	1079		Lynnwood
located at	2	3	USD	No	Thumbnail	Cats,Dogs	Medium-Priced Rent	Monthly	1177	1712 Dona	Shawnee
located at	1	1	USD	No	Thumbnail	Cats,Dogs	Medium-Priced Rent	Monthly	678	2975 Blac	Dallas
Cable or Satellite	1	2	USD	No	Thumbnail		Medium-Priced Rent	Monthly	995	290 9th Av	Forest Lak
Wood Floor	1	2	USD	No	Thumbnail	Cats,Dogs	Low Rent	Monthly	414		Norfolk

Updated Dataset Description:

- The “price_display” column used in the previous milestone as the actual output has been removed.
- A New “**RentCategory**” column has been added instead. Each apartment can have a category of {Low Rent, Medium-Priced Rent or High Rent}.

Milestone 2 Classification task:

Classify each apartment rent into one of three categories: (Low Rent, Medium-Priced Rent or High Rent) based on the provided features **in the updated dataset**.

Project(2): **Electronic Device Rating Prediction**

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

Updated Dataset Snapshot:

brand	processor	processor	processor	ram_gb	ram_type	ssd	hdd	os	graphic_card	weight	warranty	Touchscreen	msoffice	Price	rating	Number of Reviews	Number of Ratings
ASUS	Intel	Pentium Q	Not Availa	4 GB	LPDDR4X	0 GB	512 GB	64-bit Win	0 GB	Casual	No warran	No	No	45990	Bad Rating	0	0
ASUS	Intel	Core i5	10th	8 GB	LPDDR4	0 GB	512 GB	32-bit DOS	4 GB	Casual	No warran	No	No	106167	Good Rating	1881	241
ASUS	Intel	Celeron D	Not Availa	4 GB	DDR4	0 GB	512 GB	64-bit Win	0 GB	Casual	No warran	No	No	22990	Bad Rating	0	0
Lenovo	Intel	Core i3	11th	4 GB	DDR4	256 GB	0 GB	64-bit Win	0 GB	Casual	1 year	No	Yes	44990	Good Rating	42	4
HP	AMD	Ryzen 7	Not Availa	16 GB	DDR4	512 GB	0 GB	64-bit Win	2 GB	Casual	1 year	No	Yes	79990	Bad Rating	23	6
APPLE	M1	M1	10th	8 GB	DDR4	512 GB	0 GB	64-bit Mac	0 GB	Casual	1 year	No	No	101990	Good Rating	809	91
HP	Intel	Core i5	11th	8 GB	DDR4	512 GB	0 GB	64-bit Win	0 GB	ThinNlight	1 year	No	Yes	63990	Good Rating	3	0
ASUS	Intel	Pentium Q	Not Availa	4 GB	DDR4	0 GB	1024 GB	64-bit Win	0 GB	Casual	No warran	No	No	30780	Bad Rating	0	0
HP	Intel	Core i3	11th	8 GB	DDR4	512 GB	0 GB	64-bit Win	0 GB	ThinNlight	1 year	No	Yes	43970	Good Rating	571	56
Lenovo	Intel	Celeron D	Not Availa	4 GB	DDR4	0 GB	1024 GB	32-bit Win	0 GB	Casual	No warran	No	No	29999	Bad Rating	19	1
DELL	Intel	Core i3	11th	4 GB	DDR4	256 GB	0 GB	32-bit Win	0 GB	Casual	No warran	No	No	43900	Bad Rating	0	0
ASUS	Intel	Core i3	10th	4 GB	DDR4	512 GB	0 GB	64-bit Win	0 GB	ThinNlight	1 year	No	Yes	42990	Good Rating	1869	234

Updated Dataset Description:

- The “**rating**” column used in the previous milestone as the actual output has been removed.
- A New column is added “**RatingCategory**”. A device can have a rating category of {Bad Rating or Good Rating}.

Milestone 2 Classification task:

Classify a device into one of two categories: Good or Bad based on the provided features in **the updated dataset**.