

[AI&Rob 2024] 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 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: Doctor Fee Prediction

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

Updated Dataset Snapshots:

Doctor Name	City	Specialization	Doctor Qualification	Experience (Years)	Total Reviews	Patient Satisfaction	Avg Time to Wait (min)	Time to Wait (min)	Hospital Affiliation	Doctors Link	Fee Category
Dr. Umair Iqbal	GUJRANWALA	Ent Specialist	MBBS, FCF	6	11	100	19	6	Central Hospital	https://www.marham.pk/doctors/gujranwala/umair-iqbal	Expensive
Dr. Haris Saeed	RAJAN-PUR	General Physician	MBBS	1	0	94	14	11	No Address	No Link Available	Cheap
Dr. Iqra Rehman	MIRPUR-KI-KHANOSE	General Physician	MBBS, RMI	6	9	100	10	0	Rehman Clinic	https://www.marham.pk/doctors/mirpur-ki-khanose/iqra-rehman	Medium-Priced
Dr. Erum Mehrooz	MHYDERABAD	Gynecologist	MBBS, FCF	11	71	96	18	10	Mehmood Hospital	https://www.marham.pk/doctors/hyderabad/erum-mehrooz	Cheap
Dr. Aisha Ali	LAHORE	Dermatologist	MBBS, FCF	12	199	100	16	2	Skinnovation	https://www.marham.pk/doctors/lahore/aisha-ali	Medium-Priced
Dr. Saadat Islam	ISLAMABAD	Urologist	MBBS, FCF	15	22	96	15	9	Maryam Hospital	https://www.marham.pk/doctors/islamabad/saadat-islam	Expensive
Dr. M. Irfan	GUJRANWALA	Urologist	MBBS, FCP	10	0	94	14	11	Gondal Medical	No Link Available	Medium-Priced
Prof. Dr. Atif Khairpur	KHAIRPUR	Urologist	MBBS, MCh	33	4	100	14	11	No Address	https://www.marham.pk/online-consultation/atif-khairpur	Medium-Priced
Dr. Awais Ahmad	NOWSHEER	Urologist	MBBS, MCh	3	0	94	14	11	No Address	No Link Available	Medium-Priced
Dr. Aftab Ali	JHELUM	General Physician	MBBS, FCP	10	0	94	14	11	Cmh Hospital	No Link Available	Medium-Priced
Assoc. Prof. Faisal	ISLAMABAD	Urologist	MBBS, FCF	8	94	100	14	11	Saad Medical	https://www.marham.pk/doctors/faisal	Expensive

Updated Dataset Description:

- The “Fee” column used in the previous milestone as the actual output has been removed.
- A New “**Fee Category**” column has been added instead. Each apartment can have a category of {Cheap, Medium-Priced or Expensive}.

Milestone 2 Classification task:

Classify each doctor fee into one of three categories: {Cheap, Medium-Priced or Expensive} based on the provided features **in the updated dataset**.