Deadline: December 23, 2025, at 17:00

Project Grade Percentage: 15%

PROJECT ASSIGNMENT

Project groups must consist of **4 students**. Students who wish to form a smaller group must provide a valid justification and obtain approval from the instructor.

Select a dataset that is appropriate for the domain you want to work on. Datasets may be chosen from Kaggle, the UCI Repository, or any open-source platform.

Your project report must include the following steps, and the results must be clearly presented.

---

1. Work with Matlab/Python

a) Data Analysis

You must extract fundamental information such as the content of the dataset, the meaning of the features, the number of classes, and the number of samples per class.
Features are expected to be ranked according to their discriminative power, and missing values (if any) must be handled using appropriate methods.

b) Classification / Regression / Clustering

Identify suitable classification, regression, or clustering methods for your dataset, and select **at least two** to examine in detail.
Model performances will be compared.

During data analysis, using the methods covered in the course is **mandatory**:
K-Nearest Neighbors, Naïve Bayes, Decision Trees, Support Vector Machines, ensemble methods (Random Forest, AdaBoost, XGBoost, Gradient Boosting, etc.).

For image-processing datasets, CNNs may be used for feature extraction; however, **deep learning–based classifiers are not allowed**.

---

2. Project Report

The report must follow an academic article format and include **introduction, methodology (development), and conclusion** sections.

- **Introduction:** Present the objective and introduce your dataset.
- **Methodology / Development:** Explain the preprocessing techniques and classification algorithms you used. Do not describe the theory; instead, explain how you applied these methods in your model.
- **Conclusion:** Compare the performance of your models.
    - Which model performs better?
    - Why?
    - If a model performed poorly, what caused the issue?

- What could be improved?

You may access the article template via:
https://template-selector.ieee.org/secure/templateSelector/downloadTemplate?publicationTypeId=1&titleId=27&articleId=3&fileId=145

3. Project Presentation

Prepare PowerPoint slides including the problem definition and your analysis results (max. 10 minutes).
All group members must be present on the presentation day and answer questions.
Anyone who does not attend will receive **40% less** than the assigned presentation grade.

IMPORTANT DATES

Group Formation and Dataset Selection

Groups must be formed and the dataset must be selected **by November 28, 2025**.

The group spokesperson must send the following information via email (baslan@yildiz.edu.tr):

- A short description of the dataset (a few sentences)
- The dataset link
- The problem to be solved (what are you trying to achieve?)
- A brief explanation of how the problem will be solved

Once the dataset is approved, the group number will be assigned by email.

The group spokesperson is responsible for uploading all project files to the system.

Group / Dataset Changes

Requests for member changes or dataset changes must be submitted with a valid justification **no later than December 5, 2025**.
No changes will be allowed after this date.
Therefore, complete your preliminary analysis on the dataset before the deadline.

Submission

By **Tuesday, December 23, 2025, at 17:00**:

- The report (Word format, article style) must be submitted **in printed form**. If I am not in my office, you may slide it under the door.
- PowerPoint presentation + Project Report + Program Code for Python or Session file for Matlab must be uploaded as a single compressed file named: **Group_No_DatasetName**

via **online.yildiz.edu.tr**.
(Data files should NOT be uploaded.)

Only the group spokesperson needs to upload the project.

Important: The system will close on December 23, 2025, at 17:00.
Do not attempt last-minute uploads.
Submissions via email will **not** be accepted.
Late submissions will receive a **–30 points penalty per day**.

DELIVERABLES

- **Word document:** Article format, double-column, **maximum 4 pages**.
- **PowerPoint slides:** Problem definition and analysis results (**max. 10 minutes**).
- **Program Code Files:** Matlab codes/ Session file or Python source files

The Word document must also be submitted **in printed form**. (Printing the PowerPoint slides is not required.)

PRESENTATION DATES

December 25, 2025 – January 8, 2026 (during class hours)

EVALUATION CRITERIA

- Clear and accurate presentation of the problem
- Correctness of the data-cleaning process
- Quality of data analysis and feature evaluation
- Originality of the approach
- The selection of the dataset is highly important. Choosing a dataset that is sufficiently challenging and suitable for conducting meaningful analysis will significantly affect your grade performance
- Clear presentation of the dataset in the report
- Appropriate comparison of results (tables, graphs, etc.)
- Proper interpretation of results and insightful suggestions
- Correct citation of all used references