

Lab Final

Deadline: 7 April, 2025

Task1:

- Find a dataset
- A short description of the dataset (You can use pandas library for descriptive statistics)
- Provide the source of the dataset.
- Identify your objective, for example: classification or regression.
- Write details of the objective.
- Identify suitable machine learning algorithms to achieve your objective.
- Discuss the algorithms.
- Discuss why they are suitable for your dataset.
- Identify at least 3 research papers from google scholar on the same dataset. Discuss what they have done with this dataset.
- Print a hard copy of the assignment and submit on **7 April**, during submission.
- Use a formal writing style.
- Use a good cover page with your Full name and registration number or ID.
- It's an individual task.

Remember Task 1 should be completed individually. If I get any kind of copy paste type work, he/she will get penalties for these.

Then form a group with maximum 2 members. If anyone wants to do the tasks individually, it wouldn't be a problem. But I won't allow any group with 3 or more members.

Now discuss with your project partner and fix a dataset between you and your partner that completed in Task 1.

If your dataset is Tabular.

Task 2: Exploratory Data Analysis (EDA).

- Perform EDA on the dataset you have used in Task 1.
- Plot Scatter Plot, Histogram, Bar chart, pie chart, heat map wherever necessary

Task 3: Data Preprocessing

- Replace the Null values. Print the total count after and before removing the null values.
- Plot boxplot and identify outliers.
- Scale your data, with Min-Max Scaling or Normalization or Standardization or Robust Scaling (Whichever is appropriate).
- Check class imbalance issues.

If your dataset is Image.

Task 2. Exploratory Data Analysis (EDA):

- Examine a few images to understand the dataset's structure, dimensions, color channels (grayscale or RGB), and label distribution (if classification) on the dataset you have used in Task 1.
- Plot a few sample images with their respective labels (if available) to visualize class variety.
- Create a bar chart showing the distribution of images across classes (if it's a classification task).
- Analyze the count of images per class to check for class imbalances.

Task 3: Data Preprocessing

- Resize images to a consistent shape if they vary in dimensions.
- Normalize the pixel values using the mean and standard deviation calculated in EDA.
- Apply augmentations like rotation, flipping, cropping, and brightness adjustments to increase dataset variability and improve model generalization.

Notes:

I'll take the submission in person. Every person in a group must be presented during submission. I'll see the colab notebook. Later I'll give you google form to take group information and share a sheet with you to track submission time.