

Machine Learning LAB MSE - November, 2025

Class: BSAI-F23	Course Instructor: Dr. Abid Ali
Semester: 5th	Lab Engineer: Ms. Sana Saleem
Course: Machine Learning	Date: 4 th - November-2025
Week 08: Mid Semester Exam	School Of Computing-PAFIAST

Duration: 2 Hours

Total Marks: 35

Instructions:

- Use Python and the OpenCV or PIL libraries to complete the tasks.
- Write clear comments in your code to explain each step.
- Ensure output images are displayed with appropriate titles and saved as separate files in a designated folder.
- Use your name and registration number to save a file.(i.e- Sanasaleem-m22f0019ds002)
- Must submit a proper **PDF** report along with code file in teams folder within given time
 - Anyone caught cheating will be marked zero on spot.
 - No Extra time will be given.

Section A (5 Marks)

- Q1. Select any real-world dataset (minimum 300 rows) and perform data loading, display the first 10 records, and show dataset shape. **(2 Marks)**
- Q2. Identify missing values in the dataset and apply two different techniques to handle missing data. Provide before and after screenshots. **(2 Marks)**
- Q3. Explain the importance of data preprocessing in Machine Learning. Why is it considered a crucial step? **(1 Marks)**

Section B: (5 Marks)

- Q1. Write a Python program to plot any two visualizations (Histogram, Scatter Plot, Bar Plot, Heatmap) from the selected dataset. Add a summary of insights gained from each plot. **(2 Marks)**
- Q2. Write a Python program to detect and remove outliers using the Z-score or IQR method. Show before and after dataset shapes. **(2 Marks)**
- Q3. Explain what correlation is. How does correlation help in feature selection? **(1 Marks)**

Section C: (5 Marks)

- Q1. Train a Simple Linear Regression model on your dataset to predict any continuous target variable. Show model equation, MSE and R² score. **(2 Marks)**

"May your results reflect your hard work — and your journey remind you how strong you truly are."

- Q2. Train a Logistic Regression model on the same dataset after converting the problem into classification. Show confusion matrix and accuracy score. **(2 Marks)**
Q3. Explain the difference between Linear Regression and Logistic Regression with examples. **(1 Marks)**

Section D: (10 Marks)

- Q1. Implement Naïve Bayes or KNN classification on your dataset. Compare results with Logistic Regression. **(5 Marks)**
Q2. Evaluate model performance using Precision, Recall, and F1-Score. Provide interpretation. **(5 Marks)**
-

Section E: (10 Marks)

- Q1. Select ONE Machine Learning model used in this exam (Linear Regression, Logistic Regression, Naïve Bayes, or KNN). Conduct a detailed analysis including: **(5 Marks)**

- Model Assumptions
- Reasons for Selecting the Model
- Model Training Steps
- Performance Evaluation
- Strengths & Limitations
- Final Conclusion

- Q2. You are required to evaluate the performance of a binary classification model using the **ROC Curve and AUC Score**. **(5 Marks)**

Using any binary classification dataset of your choice (e.g., Titanic Survival, Heart Disease, Breast Cancer, Diabetes dataset):

1. Import and preprocess the dataset (handle missing values if any, and split into training and testing sets).
2. Train a **Logistic Regression** classification model.
3. Use the model to predict probabilities for the test set.
4. Plot the **ROC Curve** based on model predictions.
5. Calculate and display the **AUC (Area Under the Curve) Score**.
6. Write **3–4 lines interpreting** the ROC Curve and AUC value. Explain what the results indicate about your model's performance.

"Every challenge is a step toward your strength."

*** GOOD LUCK***

"May your results reflect your hard work — and your journey remind you how strong you truly are."