

# Assignment: Classification of Real-World Datasets

## Applied Machine Learning

### Objective

You are required to perform a detailed classification analysis on the following two datasets from the UCI Machine Learning Repository:

1. **Banknote Authentication:** <https://archive.ics.uci.edu/dataset/267/banknote+authentication>
2. **Haberman's Survival:** <https://archive.ics.uci.edu/dataset/43/haberman+s+survival>

You must build classifiers to predict the target variable in each dataset using at least **three** of the following algorithms:

- Naive Bayes
- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest

### Report Structure (to be submitted as a single PDF)

Your report must be self-contained and include the following sections:

1. **Title Page:** Assignment title, your name, roll number, course title, and date.
2. **Problem Definition:** Define the classification task clearly for each dataset.
3. **Mathematical Details (Methodology):** Provide clear mathematical explanations of the algorithms used. Handwritten derivations or formulas may be included as scanned images or photos.
4. **Dataset Description:** Summarize each dataset—number of features, instances, classes, and preprocessing steps.
5. **Experiments:**
  - Describe your experimental setup and train-test split strategy.

- **Hyperparameter Tuning:** Report the tuning method and parameter ranges explored.

**6. Results on Test Data:** For each model, report the following metrics:

- Accuracy
- Macro F1-Score
- Macro Precision
- Macro Recall

**7. Conclusion:** Summarize findings and explain which models performed better and why.

**8. Future Work:** Suggest possible improvements or extensions.

## Submission Guidelines

- Submit a single **PDF file** that includes:
  - Your written report
  - Exported Jupyter notebooks (.ipynb) as PDF (either appended to the report or attached separately)
- If the total size exceeds 10 MB, zip the report and notebook PDFs and upload to the **Taxila portal**.

## Evaluation Criteria (Total: 10 Marks)

Component	Marks
Problem Understanding	1
Mathematical Explanation	2
Data Preprocessing and Description	1
Code Implementation & Hyperparameter Tuning	2
Evaluation Metrics & Result Analysis	3
Report Clarity and Presentation	1
<b>Total</b>	<b>10</b>

## Deadline:

[05 May 2025, 11:59 PM]

Late submissions are not allowed.