Machine Learning Assignment — Guidelines

Objective

The objective of this assignment is to enable students to collect a local dataset, apply a new machine learning algorithm, evaluate and explain the model, and optionally integrate it into a front-end system.

Task Description

Students must identify a real-world problem, collect or compile a dataset, apply a machine learning algorithm not taught in lectures, train and evaluate it, and explain the results using XAI techniques. **Avoid developing an image processing application.**

Guidelines

1. Problem Definition & Dataset Collection **(15 marks)**
    - Clearly describe the problem and its relevance.
    - Explain:
        - Data source (how and from where it was collected),
        - Features and target variable,
        - Size of the dataset,
    - Any preprocessing done (cleaning, encoding, normalization).
        - Ensure ethical data use (no personal or sensitive data without consent).

2. Selection of a New Machine Learning Algorithm **(15 marks)**
    - Choose an algorithm **not covered during lectures and** avoid using deep learning models.
    - Justify:
        - Why was this algorithm selected,
        - How it differs from standard models (e.g., decision trees, logistic regression, k-NN, etc.).
3. Model Training and Evaluation (20 marks)
    - Explain:
        - Train/validation/test split,
        - Hyperparameter choices,
        - Performance metrics used (accuracy, F1, RMSE, AUC, etc. depending on task),
        - Results obtained and what they indicate.
    - Include tables, graphs, or plots where appropriate.

4. Explainability & Interpretation (20 marks)
   - Apply at least **one explainability method**, such as:
     - SHAP
     - LIME
     - Feature importance analysis
     - Partial Dependence Plots (PDP)
   - Explain:
     - What the model has learned,
     - Which features are most influential,
     - Whether the model's behavior aligns with domain knowledge.

5. Critical Discussion (10 marks)
   - Limitations of the model,
   - Data quality issues,
   - Risks of bias or unfairness,
   - Potential real-world impact and ethical considerations.

6. Report Quality & Technical Clarity (10 marks)
7. Bonus: Front-End Integration (10 marks)
   Bonus marks will be awarded for:
   - Integrating the trained model into a front-end system (web app, dashboard, mobile app, etc.).
   - Allowing users to input data and view predictions/explanations.
   - Examples: Streamlit app, Flask + HTML, React frontend, etc.

Submission Requirements

- **Written Report (PDF)**
Including problem description, methodology, results, interpretation, and discussion.

- **Source Code (ZIP / GitHub link)**
Including data preprocessing, training, evaluation, and explainability scripts.

- **Dataset (if publicly shareable)** or a description of how it was obtained.

- **Demo video (3–5 minutes)** showing the front-end system.

Upload your submission to the Moodle course page.