



Hashemite University
Prince Al-Hussein bin Abdullah II
Faculty for Information Technology
Department of Computer Information
System

Data Mining Mini Project (WEKA-Based)

Course: Data Mining (151002351)

Mini Project – WEKA Based

Weight: 10 Marks

Type: Team Assignment (3–4 students)

Tool: WEKA (Explorer Mode)

Submission Mode: PDF via LMS (Microsoft Teams (Assignments)).

1. Assignment Overview

This mini project aims to help students apply data mining concepts and techniques using the WEKA tool on a real academic dataset. The assignment integrates data preprocessing, classification, Model evaluation, Result interpretation, and teamwork into one coherent task.

Students will work in small teams to analyze the dataset, build classification models, evaluate their performance, and explain results clearly

2. Learning Outcomes Addressed

This assignment addresses the following Course Learning Outcomes (CLOs):

- CLO 2: Understand different data preprocessing methods
 - CLO 3: Demonstrate knowledge in using WEKA
 - CLO 4: Apply preprocessing, transformation, and interpretation methods
 - CLO 5: Apply classification algorithms
 - CLO 6: Work effectively as a member or leader of a team
-

3. Team Structure

- Team size: 3–4 students
 - Required roles:
 - Data Preprocessing Lead
 - Modeling Lead
 - Evaluation & Interpretation Lead
 - Team Coordinator (optional)
-

4. Mini Project Topic (Fixed for All Teams)

Student Performance Prediction (Classification)

Objective: Predict whether a student will PASS or FAIL based on academic and behavioral data.

Example attributes include: Attendance, Study hours, GPA, Assignment or quiz scores.

Suggested algorithms: J48, NaiveBayes, RandomForest.

5. Dataset Used (OFFICIAL DATASET)

Source website: UCI Machine Learning Repository

dataset name: “Student Performance Data Set”

Authors: Paulo Cortez and Alice Silva

Institution: University of Minho, Portugal

This dataset is:

- Publicly available
 - Widely used in academic research
 - Free for educational purposes
-

6. Dataset Description

The dataset contains academic and behavioral information about secondary school students.

The original dataset includes a final grade attribute **G3** (0–20).

Adaptation for This Mini Project

For this project, the final grade is converted into a binary class:

- **Pass** → if $G3 \geq 10$
- **Fail** → if $G3 < 10$

Target Attribute

FinalResult ∈ {Pass, Fail}

7. Selected Attributes

Students will work with a selected subset of attributes, such as:

- **studytime** – weekly study time
 - **failures** – number of past class failures
 - **absences** – number of school absences
 - **G1** – first period grade
 - **G2** – second period grade
 - **FinalResult** – Pass / Fail (target attribute)
-

8. Project Tasks and Steps

Step 1: Dataset Understanding

- Load dataset in WEKA → Explorer → Preprocess
- Identify number of instances, attributes, and target attribute (Pass/Fail)
 - ❖ Screenshot required

Step 2: Data Preprocessing

Apply at least TWO preprocessing techniques:

- Replace missing values
- Normalize or standardize numeric attributes
- Remove irrelevant attributes

Explain why each step was used.

- ❖ Screenshots required

Step 3: Classification Algorithms

- Apply at least TWO algorithms (J48 (Decision Tree), NaiveBayes, RandomForest)
- Use 10-fold cross-validation
- Screenshots required

Step 4: Evaluation and Comparison

- Report accuracy and confusion matrix
 - Compare models and select best one
- Select the best model with justification

Step 5: Interpretation

- Explain results in simple language
 - Identify important attributes
 - Provide 1–2 real-world insights
-

9. Submission Requirements

Each team must submit:

1. Mini project report (3–4 pages PDF), or more
 2. WEKA screenshots
 3. Dataset used
 4. Team roles and contributions
-

10. Evaluation Rubric (10 Marks)

Dataset and problem understanding: 2 marks

Data preprocessing: 2 marks

Algorithms applied: 2 marks

Evaluation & comparison: 2 marks

Interpretation & insights: 1 mark

Teamwork & organization: 1 mark

Total : 10 Marks

11. Penalty Rules

- Only one algorithm used: -2 marks
 - Missing WEKA screenshots: -2 marks
 - No interpretation or explanation: -1 mark
-

12. Submission Deadline and Grading Policy

- Submission deadline will be announced on LMS.
 - Late submissions may be penalized according to university regulations.
 - Plagiarism or copied work will receive zero marks.
-

13. Submission Checklist (Before Uploading on LMS):

- Report is 3–4 pages and saved as PDF.
- Team information included (names, IDs, roles).
- Dataset description included (instances, attributes, target).
- At least TWO preprocessing steps described (with filter names).
- Screenshots included: BEFORE preprocessing and AFTER preprocessing.
- At least TWO classification algorithms applied (e.g., J48 + NaiveBayes / RandomForest).
- 10-fold cross-validation used (for classification).
- Evaluation included: Accuracy + Confusion Matrix + short comparison.
- Best model selected with clear justification.
- Interpretation included: important attributes + 1–2 real-world insights.
- File name follows required format (e.g., TeamXX_MiniProject.pdf).

Note: Missing screenshots or using only one algorithm will reduce marks according to the penalty rules.

Instructor: Ms. Marwa Al-Lataifeh

Good Luck all 😊