Data Mining Project

Intelligent Student Performance Analytics and Learning Strategy Optimization

Project Overview

This project focuses on applying **data mining techniques** to discover patterns, predict outcomes, and generate actionable insights for improving student learning.
You will analyze one or more educational datasets to understand factors that influence performance and engagement, predict at-risk students, and propose data-driven strategies to support learning success.

You are encouraged — though not required — to **combine or enrich multiple datasets** to build a more comprehensive view of student behavior and outcomes.

Project Goals

By the end of this project, you should be able to:

1. Perform **data preparation, feature engineering, and exploratory analysis** on educational datasets.

2. Build and interpret **supervised (classification/regression)** and **unsupervised (clustering/association)** models.

3. Evaluate model performance and interpretability.

4. Translate findings into **data-driven recommendations** that can inform academic decisions.

5. Communicate results effectively through visuals and concise reporting.

What You Need to Do

- Select **at least one** educational dataset (see suggestions below). You may **combine or augment** multiple datasets if you wish.

- Conduct **exploratory data analysis (EDA)** to understand distributions, relationships, and anomalies.

- Prepare data (handle missing values, encode categories, engineer features).

- Build and evaluate at least **three predictive models** (e.g., to predict pass/fail, dropout, or final grade).

- Apply at least **one unsupervised method** (e.g., clustering to group students or association rule mining to uncover patterns).

- Summarize **key findings, visualizations, and implications** for improving student performance.

- Document your workflow, insights, and recommendations in a short report using the expected format ([Springer Nature Latex Template](#)).

🏫 Suggested Datasets

You can use any of the following individually or in combination:

Dataset	Description	Link
OULAD (Open University Learning Analytics Dataset)	32K+ students, demographics, VLE engagement, and results	https://archive.ics.uci.edu/dataset/349/open+university+learning+analytics+dataset
xAPI-Edu-Data (Kaggle)	Behavioral data (raised hands, discussion participation, etc.)	https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data
UCI Student Performance Dataset	Grades, demographic, and social data	https://archive.ics.uci.edu/ml/datasets/Student+Performance
Educational Process Mining Dataset	Sequential event logs of student interactions	https://archive.ics.uci.edu/dataset/346/educational+process+mining+epm+a+learning+analytics+data+set

| MOOC or Feedback Datasets | Optional — to enrich behavioral or sentiment features | Search on https://www.kaggle.com/datasets |

Expected Deliverables

1. **Data Mining Notebook / Code**

   ○ Cleaned dataset and analysis workflow (Jupyter or Colab notebook).

   ○ Predictive and unsupervised modeling results with visualizations.

2. **Research Paper**

   ○ Problem overview and objectives.

   ○ Data description and preparation summary.

   ○ Methods and results (with concise visuals).

   ○ Discussion of insights, ethical aspects, and recommendations.
   ○ **The best performing group will be invited to submit the paper to a journal co-authored with the lecturer for bonus points.**

3. **Presentation (10 minutes)**

   ○ Highlights of methodology, insights, and recommendations.
   ○ Slides + Recorded Video

🧠 Assessment Criteria

Component	Weight	Description
Data Understanding & Preparation	25%	Quality of preprocessing and feature engineering
Modeling & Analysis	35%	Rigor, correctness, and interpretation of models
Insight & Recommendations	25%	Depth, novelty, and usefulness of findings
Report & Presentation	15%	Clarity, structure, and professionalism

👬 Group Work

- Use GitHub or shared workspace for version control.

- Submit:

  - Notebook/script files

  - Research Paper (PDF)

  - Presentation slides + video (10 minutes)
  - Statement of individual contribution
  - Project Journal - make an entry every time you do something for the project. We will use it to determine individual contribution

💡 Tips for Success

- Try **combining datasets** for richer insights

- Focus on **interpretability** — educators should understand your findings.

- Manage time: allocate roughly

  - Week 1–2 → Data understanding & cleaning

  - Week 3 → Modeling

  - Week 4 → Insights & recommendations

  - Week 5 → Report & presentation