COMP47670 Assignment 2

Spring 2025

Overview:

Learning analytics is the process of collecting and analysing data about students and their activities to improve learning outcomes. This assignment explores the idea of predicting student performance for a hypothetical university module based on data describing student activity on a Virtual Learning Environment (VLE).

The VLE dataset is stored as two CSV files in the archive **`vle.zip`** available on Brightspace.

The first file, **`vle_log.csv`**, contains activity log data, where each row represents a single activity and includes the following fields: 1) the date on which the activity took place; 2) the ID number of the student involved; and 3) the type of activity (e.g. module page visit, video lecture view, quiz attempt).

The second data file, **`vle_grades.csv`**, lists the final grade achieved by each student (distinction, merit, pass, fail), with students identified by their ID numbers.

The three tasks listed below should be implemented in a single Jupyter Notebook (not a script file). Your notebook should be clearly documented, using comments and Markdown cells to explain the code and interpret the results of your analysis.

Tasks:

1. **Data Characterisation**
    a. Perform an initial characterisation of the activity log data to explore its structure and overall trends.
    b. Produce a visualisation showing the overall distribution of student grades.

2. **ABT Creation**
    a. From the VLE activity log data, create an appropriate Analytics Base Table (ABT) such that each row in the resulting table corresponds to a unique student. Columns in this table should represent aggregated features derived from the activity log.

    Features could include (but are not limited to) aspects such as the number of activities of a given type for each student or when a student typically engaged in VLE activities.

    Finally, the table should include a column representing the final grade achieved by each student.

    b. Produce visualisations showing how the values of the final grade relate to each of the other features in the ABT.

3. **Classification and Evaluation**
    b. Using the ABT created in Task 2, apply two different classification algorithms of your choice to predict final student grades based on aggregated VLE activity. Compare the performance of the two algorithms using an appropriate evaluation strategy.
    c. Experiment with applying the two classification algorithms from Task 2a in combination with different subsets of features. Which features appear to be most useful for predicting student grades?

Guidelines:

- The assignment should be completed individually. All submissions will be subject to plagiarism checking. Any evidence of plagiarism can result in a 0 grade.
- The grade awarded will depend on the complexity of the analysis and level of detail, i.e., data characterisation, preparation, analysis, classification, evaluation, interpretation of results etc.
- See the associated Grading Rubric for a detailed breakdown of marks for each task.
- Submit your assignment via Brightspace. Your submission should comprise two files, your notebook in ipymb and in html format.
- Penalties will apply for late submissions after the specified deadline:
    - 1-5 calendar days late: 1 grade point deduction, e.g. B to B-
    - 6-10 calendar days late: 2 grade point deduction, e.g. B to C+
    - Assignments will not be accepted any later than 10 calendar days without Extenuating Circumstances formally approved by UCD.