Assignment 2 - Autumn 2025/26

Flight Delays

Summary:

Flight operations analytics uses data about flights, routes, and conditions to understand and improve airline punctuality. This assignment explores predicting departure delays for commercial flights in the UK and Ireland using a synthetic dataset. For classification purposes, a flight is considered to be **delayed** if it departs at least 15 minutes after its scheduled time.

The flight data is stored as two CSV files in the archive at the link below:

http://mlg.ucd.ie/modules/python/delays.zip

The first file, *flights_schedule.csv*, contains flight schedule and route data, where each row represents a single flight with a unique identifier.

The second file, *flight_conditions_outcomes.csv*, records the origin-airport weather conditions and the departure delay for each flight.

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 Preparation and Characterisation**

   - Combine the two flight delay data sources and organise them in a suitable format for analysis. Apply preprocessing to the data, including creating derived features that will enhance your analysis.

   - Perform a characterisation of the combined data using appropriate analyses and visualisation techniques. Discuss the trends you observed in the data.

2. **Classification and Evaluation**

   - Using the combined representation created in Task 1, apply three different classification algorithms of your choice to predict whether flights will be delayed.

   - Compare the performance of the three selected algorithms using appropriate evaluation strategies and measures. Discuss the comparative performance of the algorithms and justify your evaluation approach.

3. Predictive Feature Analysis

- Perform analyses to examine the predictive power of different features in the data for flight delay classification purposes.

- Discuss which features appear to be most useful for predicting delays and provide evidence to support your conclusions.

Guidelines:

- The assignment should be completed <u>individually</u>. All submissions will be subject to plagiarism checking. Any evidence of plagiarism can result in a 0 grade.

- The grade awarded will be based on the complexity and appropriateness of the analysis and the level of detail, including data cleaning and preparation, analysis, interpretation, and related aspects.

- Generative AI tools (such as ChatGPT, code assistants, or chatbots) **can be used** for this assignment. If used, clearly state which tool/model you used and how you used it.

- See the associated Grading Rubric document for a detailed breakdown of marks for each task.

- Submit your assignment via Brightspace. Your submission should be in the form of a single ZIP file containing your notebook file (.IPYNB).

- 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.