CMPU 4011 Machine Learning for Predictive Analytics
ASSESSMENT – BUILD A CLASSIFIER

TU DUBLIN COMPUTER SCIENCE

DEADLINE: 2nd Dec 2024

TASK DESCRIPTION

In this assignment you will develop a classifier that uses data to predict if a patient has diabetes. The classifier model has to be one of those studied in this course.

Please include at least **two different** classifiers in your code and documentation.

The data is available on Kaggle, and also uploaded on Brightspace (diabetes.csv):

https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

SUBMISSION DETAILS

Deadline: 2nd Dec 2024

What you need to submit:

Submission is through Brightspace. You need to submit 2 separate files:

1. the Python code for you classifier,
2. documentation – text document describing how you solved the problem and any decisions you had to make – including why did you chose the specific classifier, any issues with the data and how it was handled, how testing was performed, etc.

SAMPLE MARKING SCHEME

This assignment is 20% of your overall mark for this module.

The overall mark for the module is 70% exam and 30% CA, with the 30% CA mark split as 10% for weekly labs and 20% for this submission. Below are some general guidelines for marking the assessment:

Data Exploration and pre-processing	20%	Data analysis/cleaning, handling data quality issues (if any)
Classifier development and testing	40%	Model selection, training, and evaluation (including results)
Documentation	40%	*For example:* Discussion of the data, possible data issues, Justify the choice of classifier Discussion on developing/training/testing classifier, Evaluation Discussion of results