Dublin Business School
Assessment Brief

Assessment Details

Module Title:	Machine Learning & Pattern Recognition
Module Code:	B9AI104
Module Leader:	Dr Shahram Azizi
Stage (if relevant):	
Assessment Title:	CA two
Assessment Number (if relevant):	
Assessment Type:	Individual
Restrictions on Time/Length :	Submission before deadline
Individual/Group:	
Assessment Weighting:	
Issue Date:	
Hand In Date:	
Planned Feedback Date:	
Mode of Submission:	Online

Guideline:
- All questions are mandatory.
- Use Python to solve questions and perform analytics.
- Submit, a report, code and dataset in the form of colab notebook.

Question 1:   [50 marks]

Use a real-world dataset (i.e. relational, text, image, video, voice files), prepare the dataset for modelling, consider one categorical variable in the dataset, and apply three classification task. To do so:

a) Provide the functional form of the predictive model for each algorithm.

[10]

b) Train each model using different ratios of the trainset and visualize the performance of models using accuracy (y -axis) in terms of different ratio of trainsets (x-axis). Elaborate on the insights.   [15]

c) Apply ensemble methods (bagging, boosting, stacking) on the base models, evaluate the performance of each ensemble technique in 100 Monte Carlo runs and visualize the performance of models using Boxplot. [20]

d) Select the best classifier and elaborate on its advantages and limitations.

[5]

Question 2: [20 marks]

Consider a continuous attribute in your dataset as the target variable, perform regression analysis using different ensemble methods, visualize and interpret the results.

Question 3 [30 marks]

Use a dataset,

    a. apply a feature extraction (LDA or PCA) to reduce the data dimensionality so that at least 90% of information of dataset is explained through extracted features. How many features do you choose? Explain the reason. [15]

    b. Apply a classifier or regression on the extracted features, evaluate and validate the model performance. Compare the result versus the performance of the classifier without feature extraction. [15]