ROBERT GORDON UNIVERSITY ABERDEEN

CM2604 Machine Learning

Academic Year	**2022/2023**
Semester	**Semester 2**
Module Number	**CM2604**
Module Title	**Machine Learning (ML)**
Assessment Method	**Coursework (Individual) – 40% Overall Weightage**
Due Date	**26th March 2023 23:59**
Submission	**Soft copy of the report in pdf format should be submitted to the Campus Moodle**
Word Limit (See Assessment Word Limit Statement)	**N/A**
Module Co-ordinator	**Prasan Yapa**

What knowledge and/or skills will I develop by undertaking the assessment?

By the successful completion of this coursework students will be able to satisfy the requirements which have been set off under the learning outcome 1, creating a dataset for ML using data and feature engineering methods applied to a real-world data collection and learning outcome 3, using ML evaluation methodologies to compare and contrast supervised and un-supervised ML algorithms using an established machine learning framework.

On successful completion of the assessment students will be able to achieve the following Learning Outcomes:

1. To create a dataset for ML using data and feature engineering methods applied to a real-world data collection.
2. To use ML evaluation methodologies to compare and contrast supervised and un-supervised ML algorithms using an established machine learning framework.

Please also refer to the Module Descriptor, available from the module Moodle study area.

What is expected of me in this assessment?

Task(s) – content

You are expected to perform a simple *classification problem* - that of labelling emails as spam or non-spam, based on their content in terms of words. The dataset has been taken from UCI Machine learning repository (https://archive.ics.uci.edu/ml/datasets/Spambase). This must be achieved using two machine learning models based on **K Nearest Neighbors (KNN)** and **Decision Trees**.

The meta information, class distribution, attributes, attribute statistics etc. of the corpus can be found in the provided link. Optimal strategies should be followed for preparing the dataset for the proposing models. Respective libraries, frameworks, tools etc. must be used for model implementation purposes. The implemented models should be compared based on the optimal evaluation metrices. Experimental results should be showcased for both model experimental settings.

Task(s) – format

The implementation must be governed through GIT. A report should be prepared including corpus preparation, solution methodology, evaluation criteria, model evaluation, experimental results, any limitations and possible further enhancements. The project GIT URL should be publicly accessible and should be mentioned in the report. The report should have an appendix that contains all the source code (added as text, not as screenshots). If the source code is not added as text in the appendix, it won't be accepted as a valid submission.

A viva will be conducted so that you are required to present your models during your viva slot.

How will I be graded?	
A grade will be provided for each criterion on the feedback grid which is specific to the assessment. The overall grade for the assessment will be calculated using the algorithm below.	
A	At least 50% of the feedback grid to be at Grade A, at least 75% of the feedback grid to be at Grade B or better, and normally 100% of the feedback grid to be at Grade C or better.
B	At least 50% of the feedback grid to be at Grade B or better, at least 75% of the feedback grid to be at Grade C or better, and normally 100% of the feedback grid to be at Grade D or better.
C	At least 50% of the feedback grid to be at Grade C or better, and at least 75% of the feedback grid to be at Grade D or better.
D	At least 50% of the feedback grid to be at Grade D or better, and at least 75% of the feedback grid to be at Grade E or better.
E	At least 50% of the feedback grid to be at Grade E or better.
F	Failing to achieve at least 50% of the feedback grid to be at Grade E or better.
NS	Non-submission.

ROBERT GORDON UNIVERSITY ABERDEEN

Feedback grid

GRADE	A	B	C	D	E	F
DEFINITION / CRITERIA (WEIGHTING)	**EXCELLENT** Outstanding Performance	**COMMENDABLE/VERY GOOD** Meritorious Performance	**GOOD** Highly Competent Performance	**SATISFACTORY** Competent Performance	**BORDERLINE FAIL**	**UNSATISFACTORY** Fail
CORPUS PREPERATION (15 %) **Grade:** ☐	70% - 100%	60% - 69%	50% - 59%	40% - 49%	30% - 39%	0% - 29%
IMPLEMENTATION (40 %) **Grade:** ☐	70% - 100%	60% - 69%	50% - 59%	40% - 49%	30% - 39%	0% - 29%
EXPERIMENTS (35 %) **Grade:** ☐	70% - 100%	60% - 69%	50% - 59%	40% - 49%	30% - 39%	0% - 29%
LIMITATIONS (10 %) **Grade:** ☐	70% - 100%	60% - 69%	50% - 59%	40% - 49%	30% - 39%	0% - 29%

Coursework received late, without valid reason, will be regarded as a non-submission (NS) and one of your assessment opportunities will be lost.

ROBERT GORDON UNIVERSITY ABERDEEN

What else is important to my assessment?

What is plagiarism?

"Plagiarism is the practice of presenting the thoughts, writings or other output of another or others as original, without acknowledgement of their source(s) at the point of their use in the student's work. All materials including text, data, diagrams or other illustrations used to support a piece of work, whether from a printed publication or from electronic media, should be appropriately identified and referenced and should not normally be copied directly unless as an acknowledged quotation. Text, opinions or ideas translated into the words of the individual student should in all cases acknowledge the original source" (RGU 2022).

What is collusion?

"Collusion is defined as two or more people working together with the intention of deceiving another. Within the academic environment this can occur when students work with others on an assignment, or part of an assignment, that is intended to be completed separately" (RGU 2022).

For further information please see Academic Integrity.

What is the Assessment Word Limit Statement?

It is important that you adhere to the Word Limit specified above. The Assessment Word Limit Statement lists what is included and excluded from the word count, along with the penalty for exceeding the upper limit.

What if I'm unable to submit?

- The University operates a Fit to Sit Policy which means that if you undertake an assessment then you are declaring yourself well enough to do so.
- If you require an extension, you should complete and submit a Coursework Extension Form. This form is available on the RGU Student and Applicant Forms page.
- Further support is available from your Course Leader.

What else is important to my assessment?

What additional support is available?

- RGU Study Skills provide advice and guidance on academic writing, study skills, maths and statistics and basic IT.
- RGU Library guidance on referencing and citing.
- The Inclusion Centre: Disability & Dyslexia.
- Your Module Coordinator, Course Leader and designated Personal Tutor can also provide support.

What are the University rules on assessment?

The University Regulation 'A4: Assessment and Recommendations of Assessment Boards' sets out important information about assessment and how it is conducted across the University.