SIT720 Machine Learning
Assessment Task 5: Machine Learning Project.

This document supplies detailed information on Assessment Task 5 for this unit.

Key information

- Due: **Sunday 10 October 2021 by 8.00 pm (AEST),**
- Weighting: 35%

Learning Outcomes

This assessment assesses the following Unit Learning Outcomes (ULO) and related Graduate Learning Outcomes (GLO):

Unit Learning Outcome (ULO)	Graduate Learning Outcome (GLO)
ULO6 - Perform model selection and compute relevant evaluation measure for a given problem. **ULO7 -** Use concepts of machine learning algorithms to design solution and compare multiple solutions.	**GLO1 -** Discipline-specific knowledge and capabilities **GLO2** - Communication **GLO4** - Critical thinking **GLO5** - Problem solving **GLO6** - Self-management

Purpose

This assessment is an extensive machine learning project. The task is open in nature, where students should make all design decisions to solve a problem and justify their decisions. In addition, they have to design and develop solutions that are better than any existing solutions.

Assessment 5 **Total marks = 35**

Submission Instructions

a) Submit your solution codes into a **notebook file with ".ipynb"** extension. Write discussions and explanations including outputs and figures into a separate file and **submit as a PDF file**.
b) Submission other than the above-mentioned file formats will not be assessed and given **zero** for the entire submission.
c) Insert your Python code responses into the cell of your submitted ".ipynb" file **followed by the question** i.e., copy the question by adding a cell before the solution cell. If you need multiple cells for better presentation of the code, add question only before the first solution cell.
d) Your submitted code should be executable. If your **code does not generate** the submitted solution, then you will **get zero** for that part of the marks.
e) Answers must be **relevant and precise**.
f) No **hard coding** is allowed. Avoid using specific value that can be calculated from the data provided.
g) Use all the **topics covered in the unit** for answering this assignment.
h) Submit your assignment **after running each cell individually**.
i) The submitted notebook **file name** should be of this form "SIT720_A5_studentID.ipynb". For example, if your student ID is 1234, then the submitted file name should be "SIT720_A5_1234.ipynb".

Questions

Background

In this project you are given a dataset and an article that uses this dataset. The authors have developed ten ML models for predicting survival of patients with heart failure and compared their performance. You must read the article to understand the problem, the dataset, and the methodology to complete the following tasks.

Dataset

The dataset contains the medical records of patients who had heart failure, collected during their follow-up period. Each patient profile has 13 clinical features. A detailed description of the dataset can be found in the Dataset section of the provided article (*patient_survival_prediction.pdf*).

Tasks:

1. Read the article and reproduce the results presented in Table-4 using Python modules and packages (including your own script or customised codes). Write a report summarising the dataset, used ML methods, experiment protocol and results including variations, if any. During reproducing the results: **(10 marks)**

i) you should use the same set of features used by the authors.
ii) you should use the same classifier with exact parameter values.
iii) you should use the same training/test splitting approach as used by the authors.
iv) you should use the same pre/post processing, if any, used by the authors.
v) you should report the same performance metrics as shown in Table-4.

N.B.
(i) Some of the ML methods are not covered in the current unit. Consider them as HD tasks i.e., based on the knowledge gained in the unit you should be able to find necessary packages and modules to reproduce the results.
(ii) If you find any issue in reproducing results or some subtle variations are found due to implementation differences of packages and modules in Python then appropriate explanation of them will be considered during evaluation of your submission.
(iii) Similarly, variation in results due to randomness of data splitting will also be considered during evaluation based on your explanation.
(iii) Obtained marks will be proportional to the number of ML methods that you will report in your submission with correctly reproduced results.
(iv) Make sure your Python code segment generates the reported results, otherwise you will receive zero marks for this task.

Marking criteria:

- *i) Unsatisfactory (x<5): tried to implement the methods but unable to follow the approach presented in the article. Variation of marks in this group will depend on the quality of report.*
- *ii) Fair (5<=x<6): appropriately implemented 50% of the methods presented in the article. Variation of marks in this group will depend on the quality of report.*
- *iii) Good (7<=x<8): appropriately implemented 70% of the methods presented in the article. Variation of marks in this group will depend on the quality of report.*
- *iv) Excellent(x>=8): appropriately implemented >=90% of the methods presented in the article. Variation of marks in this group will depend on the quality of report.*

2. Design and develop your own ML solution for this problem. The proposed solution should be different from all approaches mentioned in the provided article. This does not mean that you must have to choose a new ML algorithm. You can develop a novel solution by changing the feature selection approach or parameter optimisations process of used ML methods or using different ML methods or different combinations of them. This means, the proposed system should be substantially different from the methods presented in the article but not limited to only change of ML methods. Compare the result with reported methods in the article. Write a technical report summarising your solution design and outcomes. The report should include: **(20 marks)**

i) Motivation behind the proposed solution.
ii) How the proposed solution is different from existing ones.
iii) Detail description of the model including all parameters so that any reader can implement your model.
iv) Description of experimental protocol.
v) Evaluation metrics.
vi) Present results using tables and graphs.
vii) Compare and discuss results with respect to existing literatures.
viii) Appropriate references (IEEE numbered).

N.B. This is a HD (High Distinction) level question. Those students who target HD grade should answer this question (including answering all the above questions). For others, this question is an option. This question aims to demonstrate your expertise in the subject area and the ability to do your own research in the related area.

Marking criteria:
(i) Unsatisfactory (<10): an appropriate solution presented whose performance is lower than the reported performances in the article (Table 11). The variation in the marking in this group will depend on the quality of the report.
(i) Fair (10 - <14): an appropriate solution presented whose performance is at least equal with the lowest performance reported in the article (Table 11). The variation in the marking in this group will depend on the quality of the report.
(ii) Good (>=14): an appropriate solution presented whose performance is better than the best reported performances in the article (Table 11). The variation in the marking in this group will depend on the quality of the report.

3. Present your result in a 3 minutes video using PowerPoint slides/animation. **(5 marks)**
Marking criteria:
(i) Quality of audio presentation
(ii) Quality of slides/animation.
(iii) *Completeness of the information.*

Submission details
Deakin University has a strict standard on plagiarism as a part of Academic Integrity. To avoid any issues with plagiarism, students are strongly encouraged to run the similarity check with the Turnitin system, which is available through Unistart. A Similarity score MUST NOT exceed 39% in any case. Late submission penalty is 5% per each 24 hours from- Sunday 10 October 2021 by 8.00 pm (AEST), No marking on any submission after 5 days (24 hours X 5 days from- Sunday 10 October 2021 by 8.00 pm (AEST),).

Extension requests
Requests for extensions should be made to Unit/Campus Chairs well in advance of the assessment due date. If you wish to seek an extension for an assignment, you will need to submit a request using the “Extension

Request" link of the "Assessment" menu in the unit site, as soon as you become aware that you will have difficulty in meeting the scheduled deadline, but at least 3 days before the due date. When you make your request, you must include appropriate documentation (medical certificate, death notice) and a copy of your draft assignment. Conditions under which an extension will normally be approved include:

Medical To cover medical conditions of a serious nature, e.g. hospitalisation, serious injury or chronic illness. Note: Temporary minor ailments such as headaches, colds and minor gastric upsets are not serious medical conditions and are unlikely to be accepted. However, serious cases of these may be considered.

Compassionate e.g. death of close family member, significant family and relationship problems.

Hardship/Trauma e.g. sudden loss or gain of employment, severe disruption to domestic arrangements, victim of crime. Note: Misreading the timetable, exam anxiety or returning home will not be accepted as grounds for consideration.

Special consideration
You may be eligible for special consideration if circumstances beyond your control prevent you from undertaking or completing an assessment task at the scheduled time. See the following link for advice on the application process: http://www.deakin.edu.au/students/studying/assessment-and-results/special-consideration.

Assessment feedback
This project is alternative to final exam. The result of this project will be published with the unit's final result. We will not publish it separately like other assessments.

Referencing
You must correctly use the IEEE numbered method in this assessment as suggested in the task section.

Academic integrity, plagiarism and collusion
Plagiarism and collusion constitute extremely serious breaches of academic integrity. They are forms of cheating, and severe penalties are associated with them, including cancellation of marks for a specific assignment, for a specific unit or even exclusion from the course. If you are ever in doubt about how to properly use and cite a source of information refer to the referencing site above.
Plagiarism occurs when a student passes off as the student's own work, or copies without acknowledgement as to its authorship, the work of any other person or resubmits their own work from a previous assessment task.
Collusion occurs when a student obtains the agreement of another person for a fraudulent purpose, with the intent of obtaining an advantage in submitting an assignment or other work.
Work submitted may be reproduced and/or communicated by the university for the purpose of assuring academic integrity of submissions: https://www.deakin.edu.au/students/study-support/referencing/academic-integrity.