

This document supplies detailed information on Assessment Task 4 for this unit.

### Key information

- Due: **Monday 27 September 2021 by 8.00 pm (AEST)**,
- Weighting: 25%

### 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)   |
|--|---|
| <b>ULO3</b> - Perform linear regression, classification using logistic regression and linear Support Vector Machines.<br><b>ULO4</b> - Perform non-linear classification using KNN and SVM with different kernels.<br><b>ULO5</b> - Perform non-linear classification using Decision trees and Random forests.<br><b>ULO6</b> - Perform model selection and compute relevant evaluation measure for a given problem.<br><b>ULO7</b> - Use concepts of machine learning algorithms to design solution and compare multiple solutions. | <b>GLO1</b> - through the assessment of student ability to apply advanced data processing techniques through programming for prediction.<br><b>GLO5</b> - through assessment of student ability to deal with defined data set and solve problems. |

### Purpose

Students will be given a specific data set for analysis and will be required to develop and compare various classification techniques. Each student must demonstrate skills acquired in data representation, classification, and evaluation.

### Assessment 4

**Total marks = 30**

### Submission Instructions

- 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**.
- Submission other than the above-mentioned file formats will not be assessed and given **zero** for the entire submission.
- 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.
- 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.
- Answers must be **relevant and precise**.
- No **hard coding** is allowed. Avoid using specific value that can be calculated from the data provided.
- Use **topics covered till week 10** for answering this assignment.
- Submit your assignment **after running each cell individually**.
- The submitted notebook **file name** should be of this form “SIT720\_A4\_studentID.ipynb”. For example, if your student ID is 1234, then the submitted file name should be “SIT720\_A4\_1234.ipynb”.

---

Questions

---

1. What is an ensemble classifier? Name some of the popular ensemble methods (at least three) and which one you prefer and why? **(2 marks)**
  2. Let's assume we have a noisy dataset. You want to build a classifier model. Which classifier is appropriate for your dataset and why? **(2 marks)**
- 

Background

In the modern world, customer details are very important to suggest any product for buying. Gender, age and education have impact on level of consumption of different products. So, it is essential for businesses to analyse their customer details to better understand consumer behaviour and their impact on various products.

**Dataset filename:** Customer relationship marketing (CRM).csv

**Dataset description:** This dataset includes data on customer details and their response to buy any products. The data contains 20 attributes and 9134 records.

**Features and labels:** The attribute names are listed below.

- I. State
- II. Customer Lifetime Value
- III. Response
- IV. Coverage
- V. Education
- VI. Effective To Date
- VII. EmploymentStatus
- VIII. Gender
- IX. Income
- X. Location Code
- XI. Marital Status
- XII. Monthly Premium Auto
- XIII. Months Since Last Claim
- XIV. Number of Open Complaints
- XV. Number of Policies \* Policy
- XVI. Renew Offer Type
- XVII. Sales Channel
- XVIII. Total Claim Amount
- XIX. Vehicle Class

---

Questions

---

3. Load and pre-process the dataset if necessary. Explain steps that you have taken. Are there any alternative ways for doing that? Explain. **(5 marks)**

4. Analyse the importance of the features for predicting customer response using two different approaches. Explain the similarity/difference between outcomes. **(5 marks)**
5. Create three supervised machine learning (ML) models except any ensemble approach for predicting customer response. **(10 Marks)**
  - a. Report performance score using a suitable metric. Is it possible that the presented result is an overfitted one? Justify.
  - b. Justify different design decisions for each ML model used to answer this question.
  - c. Have you optimised any hyper-parameters for each ML model? What are they? Why have you done that? Explain.
  - d. Finally, make a recommendation based on the reported results and justify it.
6. Build three ensemble models for predicting customer response. **(6 Marks)**
  - a. When do you want to use ensemble models over other ML models?
  - b. What are the similarities or differences between these models?
  - c. Is there any preferable scenario for using any specific model among set of ensemble models?
  - d. Write a report comparing performances of models built in question 5 and 6. Report the best method based on model complexity and performance.
  - e. Is it possible to build ensemble model using ML classifiers other than decision tree? If yes, then explain with an example.

*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.*

### **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- Monday 27 September 2021 by 8.00 pm (AEST), No marking on any submission after 5 days (24 hours X 5 days from- Monday 27 September 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**

The results with comments will be released within 15 business days from the due date.

#### **Referencing**

You must correctly use the Harvard method in this assessment. See the Deakin referencing guide.

#### **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>.