THE UNIVERSITY OF NEW SOUTH WALES

SCHOOL OF INFORMATION SYSTEMS AND TECHNOLOGY MANAGEMENT

TERM 2 2024

COMM1190: DATA, INSIGHTS, & DECISIONS

SAMPLE EXAMINATION

NOTE: THE INSTRUCTIONS ON PAGES 1–2 ARE COPIED VERBATIM FROM THE T2 FINAL EXAM AND DO NOT NECESSARILY REFLECT THE CONTENTS OF THIS SAMPLE EXAM.

1. Writing Time: 3 Hours.
2. Reading and Submission Allowance Time: 1 Hour
3. This is an Online Open-book Exam, your responses must be your original work. You must attempt this Final Exam by yourself without any help from others. Thus, you have **NOT** worked, collaborated or colluded with any other persons in formulating your responses. The work that you are submitting for your Final Exam is your **OWN** work.
4. Release date/time (via Moodle): **Monday, August 12, 8:00am** (Australian Eastern Time Zone)
5. Submission date/time (Via Turnitin): **Monday, August 12, 12:00pm** (Australian Eastern Time Zone)
6. Failure to upload the exam by the submission time will result in a penalty of **15%** of the available marks **per hour** of lateness.
7. This Examination Paper has 17 pages, including the cover pages.
8. Total number of Questions: 4 Questions.
9. Answer all 4 Questions.
10. Total marks available: 100 marks. This examination is worth **50%** of the total marks for the course.
11. Questions are NOT of equal value. Marks available for question sub-parts are shown on this examination paper.
12. All questions have word limits as indicated on the question. These word limits must be adhered to. Text in excess of the specified word limit(s) may not be considered in the marking process.
13. Answers to questions are to be written in the exam answer sheet template provided. Please ensure that you provide all details required on the cover sheet of your Final Exam answer sheet.
14. Failure to submit exam answers with the correct exam answer sheet will result in **10%**

penalty of your overall exam marks.

15. Students are reminded of UNSW's rules regarding Academic Integrity and Plagiarism. Plagiarism is a serious breach of ethics at UNSW and is not taken lightly. For details, see Examples of plagiarism.
16. The use of any generative AI tools or services to search for or generate information or answers is prohibited. If its use is detected, it will be regarded as serious academic misconduct and subject to the standard penalties, which may include 00FL, suspension and exclusion.
17. This Final Exam is an open book/open web. Further information is available "Here".
    - You are permitted to refer to your course notes, any materials provided by the course convenor or lecturer, books, journal articles, or tutorial materials.
    - It is sufficient to use in-text citations that include the following information: the name of the author or authors; the year of publication; the page number (where the information/idea can be located on a particular page when directly quoted), For example, (McConville, 2011, p. 188).
    - You are required to cite your sources and attribute direct quotes appropriately when using external sources (other than your course materials).
    - When citing Internet sources, please use the following format: website/page title and date.
    - If you provide in-text citations, you MUST provide a Reference List. The Reference list will **NOT BE** counted towards your word limit.
18. Students are advised to read the Final Exam paper thoroughly before commencing.
19. The Lecturer-in-Charge (LiC) / Exam Referee will be available online (via Moodle) after the Open-book Exam paper is released for a period of one hour.

Note: This questions on this sample exam are compiled from prior COMM1190 final exams. The T2 exam will be similar in overall scope and organization. But, while these represent some of the kinds of questions and topics that might be asked, it is important to note that your final exam may cover different topics or ask you to engage with similar topics in different ways.

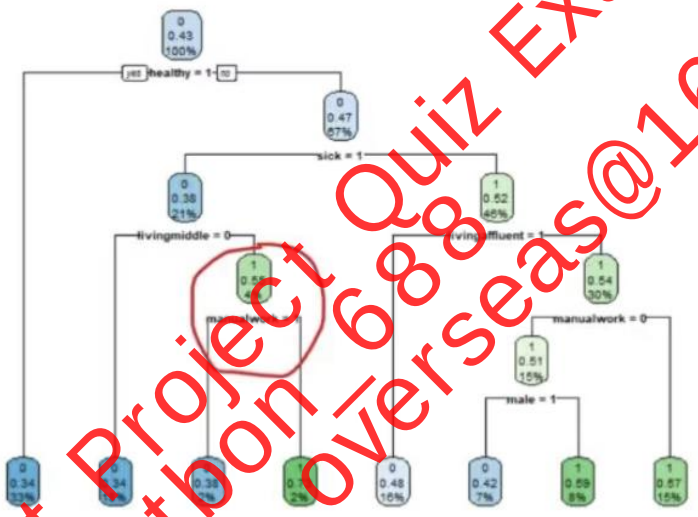
The answer sheet provided is based on the answer sheet being used this term. You are STRONGLY encouraged to practice entering your answers there, including checking that your word count for each question does not exceed the specified maximum.

QUESTION 1 **10 MARKS**

Suppose you work as an actuary and want to model the 5-year mortality rates (i.e., the probability of death within five years) for a group of 1,000 life insurance policyholders aged 50. You have the following information for each:

- Gender
- Type of work (manual or non-manual)
- Living area (High affluence, Middle class, Deprived)
- Health condition (Terminally ill, Sick, healthy)

The following classification tree is obtained for the binary variable denoting the Status of the policyholder as observed in the data. The Status variable takes a value equal to 0 if the policyholder is alive at the end of the study (5 years, where the policyholders are observed starting at the age of 50 until age 55) and 1 if dead.



Required:

With reference to the above case, please answer all of the following questions:

Part A [max 100 words] *(5 marks)*

What does each number in a node represent? Please describe the results in the node circled in red.

Part B [max 100 words] *(5 marks)*

What are the most influential variables characterising policyholders' mortality in order of importance?

Part A [max 100 words] *(5 marks)*

What does each number in a node represent? Please describe the results in the node circled in red.

Part B [max 100 words] *(5 marks)*

What are the most influential variables characterising policyholders' mortality in order of importance?

QUESTION 2 **30 MARKS**

A financial institution has decided to implement a Machine Learning (ML) algorithm to automate its loan approval process. The algorithm, developed by a third-party AI solutions provider, is designed to analyse a range of consumer data to determine loan eligibility and terms. To maintain a transparent process, the firm offers applicants the option to discuss the loan outcome decision with a representative. The representatives are well-trained in explaining how the applicant's data influenced the loan decision.

The third-party developer has provided the firm with two versions of the algorithm. The first version is "gender-blind," designed to exclude any variables that could directly or indirectly reveal an applicant's gender. The second version incorporates gender-specific features, which can enhance the accuracy of loan predictions based on statistical differences in financial behaviour across genders. The third-party developer is also open to feedback on variables to include in the model and how to determine who is loan-worthy.

Required:

With reference to the above case, please answer all of the following questions:

Part A [max 80 words] *(4 marks)*

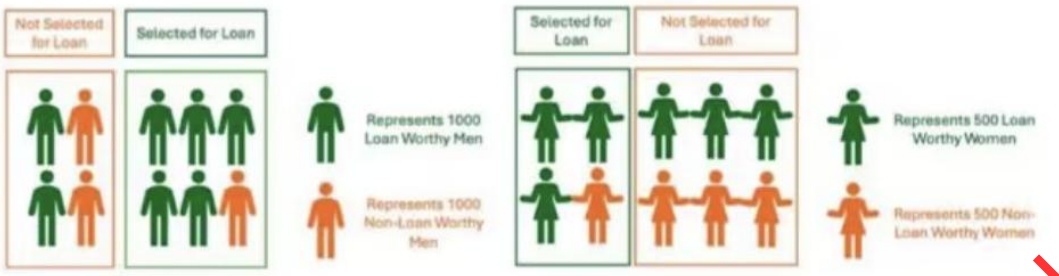
Given a choice between the gender-blind version and the gender-specific version of the loan decision algorithm, identify which option the firm should choose to use and any general adjustments that could be made to make the algorithm more responsible.

Part B [max 120 words] *(6 marks)*

Considering the scenario, identify and discuss the issues that raise concerns under the Australian Responsible AI Principles, focusing specifically on fairness, accountability, transparency, and contestability.

Part C [max 240 words] *(12 marks)*

The firm has a dataset of 15,000 people. The results of the model are graphically presented as follows:



Create confusion matrices to determine the following:

I. The overall accuracy rate
II. The overall false positive rate
III. The accuracy rate for men
IV. The false positive rate for men
V. The accuracy rate for women
VI. The false positive rate for women

Part D [max 160 words] *(8 marks)*

I. Based on the confusion matrix, explain why you believe or do not believe the predictive model is fair.
II. 1500 new clients (1000 men and 500 females) have applied for loans and need to be assessed. The third-party developer can adjust the model so that it selects 500 men and 250 females to be selected for loans. What are the advantages and disadvantages of this approach? Should the firm update the model to this approach?

QUESTION 3 **30 MARKS**

Cookie Cats is a popular puzzle game app. As players progress through the levels of the game, they occasionally encounter gates that force them to wait a non-trivial amount of time or make an in-app purchase to progress. In addition to driving in-app purchases, these gates serve the important purpose of giving players an enforced break from playing the game, hopefully resulting in the player's enjoyment of the game being increased and prolonged. The question to be explored is whether the placement of the entry-level gate matters. The hypothesis being investigated is whether moving the entry-level gate from 30 to 40 will increase player activity (and hopefully in-game purchases).

Data was collected from an A/B test where a total of 90,198 players were randomly assigned to a treatment group (entered at gate level 40, *g_40*=1) or a control group (entered at gate level 30, *g_40*=0). The output variables of interest included in the data are:

ret_1 = 1 if the player plays within one day of installing the app; and = 0 otherwise;
games = total number of games played in the first 14 days of installing the app.