MIS710 – Machine Learning in Business - Trimester 1 2025
Assessment Task 2 – Analytical Report and Business Report – Individual

DUE DATE:	**Wednesday 28 May 2024, 8:00 pm AEST**
PERCENTAGE OF FINAL GRADE:	**60% including Part A (40%) and Part B (20%)**
WORD COUNT:	**Part A Analytical Report: Maximum number of words: 2000 words** **Part B Business Report: Maximum number of words: 1000 words**

Description

Purpose

There are two parts in this assignment.

- Part A provides you with opportunities to learn a range of machine learning methods and Python skills (GLO1 & ULO1) and apply your digital literacy to research and develop a machine learning solution (GLO3, GLO5, and ULO2). By completing this task, you will gain knowledge and skills in selecting and applying one or more appropriate machine learning algorithm(s) to develop and evaluate a machine learning solution and interpret the outcomes.
- In Part B, you will report your application of machine learning and make recommendations to the business and management audience. By completing this task, you will gain ability to explain and justify machine learning options and discuss their pros and cons to the business audience.

Project Overview

You're engaged to work with Data2Intel, an Australian learning analytics consulting service that specialises in delivering various data analytics services within the primary education sector. The current project aims to predict primary school students who are at risk of underperforming in numeracy in Year 3. This initiative is part of a broader effort by a consortium of forty primary schools to enhance educational outcomes and provide targeted support to students in need.

Context/Scenario

Australian school education comprises K-12, a system akin to those in many other countries. Primary schools encompass scholastic years K-6, while secondary schools cover years 7-12. In Australia, most Kindergarten students must be at least 5 years old by January of the calendar year they commence schooling. The early years of schooling, typically K-2 (Kindergarten to Year 2), encompass children aged 5-7 years.

The National Assessment Program – Literacy and Numeracy (NAPLAN) tests are annual assessments for students in Years 3, 5, 7, and 9 to provide a key benchmark for assessing students' foundational literacy and numeracy skills. Research, although more limited in scope compared to secondary and tertiary education, underscores the profound importance and influence of these early years on future academic performance, employability, wellbeing, and career progression. The foundational skills acquired during these formative years set the stage for long-term educational and professional success.

You are provided with a dataset of 2,000 students across over forty schools. The dataset focuses on their reading and numeracy skills during the early years of Year 1 and Year 2. These skills were measured through localised, formative assessments, which, while validated and consistent, are not solely "pen and paper" tests.

Given the young age of the students, these assessments often include dialogue or interview-based evaluations administered by trained teachers. In the dataset, students at risk of underperforming in numeracy, Year3_Numeracy_At_Risk, were determined by their NAPLAN results in Year 3. In addition, the dataset also provides students' demographic and family backgrounds as well as disability conditions. Further details can be found in the supplied data description.

This comprehensive dataset spans five continuous years, from 2016 to 2020, and has been curated for learning support and research purposes. Importantly, there is no missing data, as records with incomplete information have been removed for this exercise. However, you are still required to check data quality and preprocess the data as needed.

Specific Requirements

You are tasked with performing two analytical tasks: 1/ uncovering data insights and 2/ exploring machine learning opportunities, and reporting the findings to Data2Intel.

Regarding the first task - uncovering data insights, you are required to respond to the following enquiries:

1. Student demographics analysis: Analyse student demographics factors to gain insights about the students included in the dataset and report findings and insights.
2. What are students' numeracy skills in Year 1, including Counting-01, Place Value-01, Addition and Subtraction-01, and Multiplication and Division-01. Is there a relationship between students' numeracy skills in Year 1 and Year3_Numeracy_At_Risk?
3. What are students' reading skills, for example TextLevel and Writing Vocabulary in Year 1? Are there relationships between these and their Year3_Numeracy_At_Risk?
4. Are students' literacy skills and numeracy skills related, in Year 1 and in Year 2? Are there relationships between these and their Year3_Numeracy_At_Risk?
5. Describe the students' disability conditions in the dataset. Are there relationships between these conditions and their Year3_Numeracy_At_Risk?
6. Are there other insights that might inform early interventions to improve students' numeracy skills?

Regarding the second task - machine learning opportunities, you are required to

1. Develop and evaluate two predictive models to identify students at risk of underperforming in numeracy in Year 3 (Year3_ Numeracy_At_Risk).
2. Develop one clustering model to explore possible clusters of students.

Based on your findings from the above both tasks, you are required to provide actionable insights and recommendations to primary schools and educators to implement targeted interventions and support mechanisms. In addition, you should advise the Client of the potential ethical and legal implications of the models.

Dataset provided: LA4Schools.csv
Data description provided: LA4Schools data description

You are required to deliver two (2) reports to the Client, Data2Intel learning analytics:

- The first **analytics** report (Part A) should present your analysis and findings to Peter Vo, Data2Intel Director of Data and Insights **– a fictitious character**. This report should detail your approach to exploring the dataset, the machine learning techniques used, and your findings. Your findings should be supported by relevant visualisations and statistical analysis. This report should also develop, test and compare two predictive machine learning models and one clustering model, and recommend a

predictive machine learning model, inform model deployment, and recommend future engagements with the client. See further details in the Specific Requirements section below.

- The second consultancy report (Part B) should be developed for Elaine Race, **Data2Intel** Director of Education and Engagement **– a fictitious character**. The report should include your response to the client's six (6) data analytics questions, the proposed machine learning models, and recommendations for use. You should also discuss the limitations of your approach and any potential areas for future improvements.

See further details in the Specific Requirements section below.

T*he dataset has been* prepared *and provided by Data2Intel, mirrors the real-life data, and has been pre-processed by Associate Professor Lemai Nguyen for assessment purposes.*

Deliverable Requirements

You are required to:

- Develop your business and data understandings using BACCM.
- Prepare and explore the provided dataset, cleanse and pre-process data as needed. Undertake an exploratory data analysis (EDA) to respond to the client's six questions.
- Undertake supervised machine learning model development, evaluation, comparison and selection. Two predictive models should be developed, tested, and compared.
- Undertake unsupervised machine learning using clustering analytics.
- Develop two reports:
  - **The first technical report (Part A)** should present your EDA (Exploratory Data Analysis) and machine learning findings to Peter Vo.
  - **The second consultancy report (Part B)**, for Elaine Race, Director of Education and Engagement, should present responses to the six (6) specific requests about data, insight from clustering analytics, and a predictive machine learning model.
- Format and present your report professionally. Two sample report templates are provided under Assessment Resources.
- Correctly use the APA7 style of referencing, and include in-text citations when quoting, referring to, summarising, or paraphrasing from any source:

  https://www.deakin.edu.au/students/studying/study-support/referencing

Deliverables:

Part A. Case Study Report

Part A.1 Machine Learning Solution

- A cover page (**not** included in the word count) that includes:
  - Report Title
  - Unit code and name
  - Student name and student ID
- A table of contents (**not** included in the word count)

- GenAI acknowledgement statements (if applicable; **not** included in the word count; see the section Use of Generative Artificial Intelligence (genAI) in this assessment below)

- An executive summary of max. 200 words is required (**included** in the word count).
- The report should include:
    1. Introduction:
        - Objective: the business problem to be addressed in its business context, and the value proposition of the project.
    2. Approach:
        - Overview of the machine learning approach, including machine learning types and problem(s), and prediction target(s).
    3. Data preparation and Exploratory Data Analysis (EDA):
        - Data sources, data size, types, quality, cleansing and pre-processing, and any observations.
        - EDA: statistical analysis and visualisation.
        - Key insights gained from EDA to inform feature selection and data splitting.
    4. Model development and evaluation:
        - Supervised Machine Learning:
            - Two predictive models and performance metrics.
            - Model comparison based on your selection criteria.
        - Unsupervised Machine Learning:
            - Clustering analytics results, cluster profiling, and justification of the number of clusters.
    5. Solution recommendation:
        - Interpretation and discussion of results obtained from the validation and comparison.
        - Solution recommendation – what model is to offer to the client.
        - Future engagements with the client.
    6. Technical recommendations:
        - Summary of the development and testing environment, such as software libraries, the programming language and computing environment used.
        - To inform model deployment, provide your machine process diagram and data pre-processing.
        - Suggestions for maintenance of accuracy and relevance over time (based on your research).

- References (**not** included in the word count)
- Optional appendices (**not** included in the word count – not subject to assessment), such as additional technical details, supplementary figures and tables.

Part A.2 Files

- A python notebook with detailed comments to guide the deployment team, AND
- A PDF version of the Python notebook.

Part B. Business report

- A cover page (**not** included in the word count) that includes:

    - Report Title
    - Unit code and name
    - Student name and student ID
- A table of contents (**not** included in the word count)
- An executive summary of max. 100 words is required (**included** in the word count).
- The report should include:
    1. Introduction:
        - Business understanding of the project using the Business Analysis Core Concept Model (BACCM) framework[1] .
    2. Insights from Exploratory Data Analysis (EDA):
        - Answers to the Client's six (6) questions.
        - Additional insights, such as comments on data quality or observations beyond the client's six (6) questions and possible insights gained from clustering analytics.
    3. Proposed machine learning solution:
        - The selected machine learning model.
        - Interpretation of its performance and discussion of pros and cons.
    4. Recommendations and conclusions:
        - Recommendations of business applications.
        - Potential benefits to stakeholders and how they relate to the value proposition.
        - Implications such as changes to business processes and decision making and possible impacts.
        - Recommendations for model improvements.
- References (**not** included in the word count)
- Optional appendices (**not** included in the word count – not subject to assessment), such as supplementary figures and tables.

Important Notes

- The final submission should be presented professionally. The reports should use clear, concise, and relevant language to communicate the content relevant to the target audiences.
- You should undertake research and use various tools to solve the business problem. In the end, you must exercise and understand the Python code yourself for your own learning purposes, develop and present your business understandings and solution to the client(s). Cite and reference any sources you use.

Submission

You must submit your assignment in the Assignment Dropbox in the unit CloudDeakin site on or before the due date. The submission must include two files:

- Two (2) report documents in **PDF for marking**. Name your documents using the following syntax: **<your surname_your first name_your Deakin student ID number_[unitcodeA1].pdf**. For example,

[1] *Business Analysis Core Concept Model (BACCM) https://www.iiba.org/business-analysis-blogs/6-steps-to-applying-the-baccm/ see also Topic 1*

'MIS710A1_Jones_Barry_123456789_MIS710A2 ReportA.pdf' and 'MIS710A1_Jones_Barry_123456789_MIS710A2 ReportB.pdf'

- Two (2) report documents **in Word as back-ups**. Name your documents using the following syntax: **<your surname_your first name_your Deakin student ID number_[unitcodeA1].doc (or '.docx')**. For example, 'MIS710A1_Jones_Barry_123456789_MIS710A2 ReportA.doc' and 'MIS710A1_Jones_Barry_123456789_MIS710A2 ReportB.doc'

- One (1) **Python notebook**

- **PDF version of the Python notebook**. From Google Colab, select File, Print and Save as PDF

Submitting a hard copy of this assignment is not required. You must keep a backup copy of every assignment you submit until the marked assignment has been returned to you. In the unlikely event that one of your assignments is misplaced you will need to submit your backup copy.

Any work you submit may be checked by electronic or other means for the purposes of detecting collusion and/or plagiarism and for authenticating work.

When you submit an assignment through your CloudDeakin unit site, you will receive an email to your Deakin email address confirming that it has been submitted. You should check that you can see your assignment in the Submissions view of the Assignment Dropbox folder after upload and check for, and keep, the email receipt for the submission.

Learning Outcomes

This task allows you to demonstrate your achievement towards the Unit Learning Outcomes (ULOs) which have been aligned to the Deakin Graduate Learning Outcomes (GLOs). Deakin GLOs describe the knowledge and capabilities graduates acquire and can demonstrate on completion of their course. This assessment task is important in determining your achievement of the ULOs. If you do not demonstrate achievement of the ULOs you will not be successful in this unit. You are advised to familiarise yourself with these ULOs and GLOs as they will inform you on what you are expected to demonstrate for the successful completion of this unit.

The learning outcomes that are aligned with this assessment task are:

Unit Learning Outcomes (ULOs)		**Graduate Learning Outcomes (GLOs)**
ULO1	Analyse and frame business challenges using machine learning concepts, techniques, and the machine learning model development lifecycle.	GLO1: Discipline-specific knowledge and capabilities
ULO2	Select and apply appropriate machine learning techniques to solve business problems and evaluate the machine learning model performance.	GLO3: Digital literacy GLO5: Problem solving
ULO3	Explain the application of machine learning and interpret the outcomes to the various stakeholders	GLO2: Communication

Use of Generative Artificial Intelligence (genAI) in this assessment

Deakin welcomes the opportunity to engage with emerging technologies in education and seeks to build your capability in the **ethical** and **responsible** use of current and emergent technology. Deakin also upholds a commitment to academic integrity and to ensuring high-quality educational outcomes that prepare you for an AI-driven future.

Using genAI as an assistant is appropriate in this assessment task.

To support your learning in this assessment task, it is recommended that you limit genAI use to assist with specific tasks such as ***identifying and editing grammatical and spelling errors, generating images for illustrative purposes, and assisting in generating Python code for specific tasks.*** You must modify any AI-generated content you use. Your final submission should be your own work and show how you have used your own critical thinking skills and what you have learnt in this unit.

It is important that you take responsibility for your final submission, including:

- **Evaluating the accuracy and quality** of any genAI generated material.
- **Acknowledging how you used genAI** tools in this assessment to ensure you are making informed decisions about your learning, demonstrating learning you have gained in the unit, and acting with integrity.

Please use the Acknowledgement statements to guide how you acknowledge the use of genAI in this assessment.

Student Toolkits

A set of toolkits was prepared by experienced Deakin students to help you learn the generic skills required in the Business & Law professions: https://d2l.deakin.edu.au/d2l/home/93063

You will find the following tool kits to be useful:

- Communication Skills - especially Writing Skills: https://d2l.deakin.edu.au/d2l/le/content/93063/viewContent/6086619/View
- Use APA7 style of referencing **and include in-text citations**: https://www.deakin.edu.au/students/studying/study-support/referencing

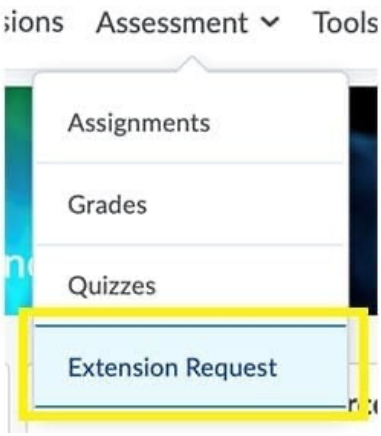
Marking and Feedback

The marking rubric indicates the assessment criteria for this task. It is available in the CloudDeakin unit site in the Assessment folder, under Assessment Resources. Criteria act as a boundary around the task and help specify what assessors are looking for in your submission. The criteria are drawn from the ULOs and align with the GLOs. You should familiarise yourself with the assessment criteria before completing and submitting this task.

Students who submit their work by the due date will receive their marks and feedback on CloudDeakin 15 working days after the submission date.

Extensions

Extensions can only be granted for exceptional and/or unavoidable circumstances outside of your control. Requests for extensions must be made by noon on the submission date using the online Extension Request form under the Assessment tab on the unit CloudDeakin site. All requests for extensions should be supported by appropriate evidence (e.g., a medical certificate in the case of ill health).



Applications for extensions after 12 noon on the submission date require University level special consideration and these applications must be submitted via StudentConnect on your DeakinSync site.

Late Submission Penalties

If you submit an assessment task after the due date without an approved extension or special consideration, 5% of the **total marks** of the task (40%) is deducted for each new calendar day up to seven calendar days*. Work submitted more than seven days after the due date will not be marked and will receive 0% for the task. The Unit Chair may refuse to accept a late submission where it is unreasonable or impracticable to assess the task after the due date.

*'Day' means calendar day for electronic submissions and working day for paper submissions.

An example of how the calculation of the late penalty based on an assignment being due on a Thursday at 8:00 pm is as follows:

- 1 day late: submitted after Thursday 11:59 pm and before Friday 11:59 pm– 5% penalty.
- 2 days late: submitted after Friday 11:59 pm and before Saturday 11:59 pm – 10% penalty.
- 3 days late: submitted after Saturday 11:59 pm and before Sunday 11:59 pm – 15% penalty.
- 4 days late: submitted after Sunday 11:59 pm and before Monday 11:59 pm – 20% penalty.
- 5 days late: submitted after Monday 11:59 pm and before Tuesday 11:59 pm – 25% penalty.
- 6 days late: submitted after Tuesday 11:59 pm and before Wednesday 11:59 pm – 30% penalty.
- 7 days late: submitted after Wednesday 11:59 pm and before Thursday 11:59 pm – 35% penalty.

The Dropbox closes the Thursday after 11:59 pm AEST/AEDT time.

Support

The Division of Student Life provides a range of **Study Support** resources and services, available throughout the academic year, including **Writing Mentor** and **Maths Mentor** online drop-ins and the SmartThinking 24-hour writing feedback service at **this link**. If you would prefer some more in-depth and tailored support, **make an appointment online with a Language and Learning Adviser**.

Referencing and Academic Integrity

Deakin takes academic integrity very seriously. It is important that you (and if a group task, your group) complete your work in every assessment task Any material used in this assignment that is not your original work must be acknowledged as such and appropriately referenced. You can find information about referencing (and avoiding breaching academic integrity) and other study support resources at the following website: http://www.deakin.edu.au/students/study-support

Your Rights and Responsibilities as a Student

As a student, you have both rights and responsibilities. Please refer to the document ***Your Rights and Responsibilities as a Student*** in the Unit Guide & Information section in the Content area on the CloudDeakin unit site.