School of Information Technology

Diploma in Applied AI & Analytic
IT3301 Applied of Machine Learning
AY2024 Semester 1

Assignment Guide

40% of Module Grade Submission by 28th July 2024 at 2359 Hrs

1. Introduction

This is an individual assignment and contributes to 40% of the ICA. The assignment is divided into two distinct parts with Part 1 accounting for 15% and Part 2 for 25% of the total assessment. The deadline for submission is Week 15, 28th July 2024, at 2359.

2. Assignment Submission

The delivery of your assignment should be in three parts:

a. Write-up (word file) for Part A.
b. Source codes files in **Google Colab** Jupyter notebook format for Part B. Please retain the output results during submission.
c. Video presentation for Part B. Presentation of your business objectives, source code and recommendation through video recording.
    i. Each student is to make a video presentation of his/her assignment of walking through his/her codes in Jupyter Notebook file with dataset.
    ii. The duration of video will be 5 minutes ONLY.
    iii. Do highlight the key insights you have found in your analysis in the video
    iv. Upload your video to your preferred online storage (dropbox, google drive etc.) and include the download link in your PoliteMall submission.

All the files in this assignment (write-up for Part A, Jupyter Notebook file, and video recording link) are to be submitted into PoliteMall by end of Week 15 - 28th July 2024 at 2359 Hrs.

Multiple submission attempts, if done before the submission deadline are allowed. Only the most recent submission will be graded.

Late submission penalties will be applied in accordance as stated in Section 3.

3. Penalty for Late Submission

a. Cap at 50% of base mark for submission within 5 calendar days from deadline
b. From 6th day within deadline onwards, award 0 mark
c. Note that “day” includes non-working days (Sat, Sun and public holidays).

Part A: Part A: AI Ethics case study (15m)
Students are required to research one AI company (Google, OpenAi, Microsoft, Amazon or IBM) and analyze its practices and policies through the lens of AI ethics. The analysis should cover various ethical dimensions, including data privacy, bias and fairness, transparency, accountability, and the societal impact of AI technologies. Write a report on the relevance of AI ethics for the organization that you have researched on.

- The analysis should be 1500-2000 words in length.
- Use at least 5 credible sources, including academic papers, industry reports, and official company documents.
- The report can be organized based on the following key points:
    - Company Analysis (3 marks)
    - Data Privacy and Security (3 marks)
    - Bias and Fairness (3 marks)
    - Transparency and Accountability (3 marks)
    - Societal Impact (1 mark)
    - Writing Quality (2 marks)

Assessment
The submission will be based on the following rubrics.

Category	**Excellent (A)**	**Good (B)**	**Average (C)**	**Poor (D)**
Company Analysis (3m)	Provides a detailed and insightful analysis of the chosen AI company's practices and policies. Covers multiple ethical dimensions with in-depth evaluation.	Offers a solid analysis of the company's practices and policies. Covers several ethical dimensions with adequate evaluation.	Basic analysis of the company's practices and policies. Covers some ethical dimensions with limited evaluation.	Insufficient analysis of the company's practices and policies. Ethical dimensions are poorly covered or not at all.
Data Privacy and Security (3 marks)	Thoroughly examines the company's approach to data privacy and security. Provides specific examples and	Examines the company's approach to data privacy and security with some examples. Evaluation is mostly clear.	Basic examination of data privacy and security practices. Few examples and limited evaluation.	Minimal or no examination of data privacy and security. Lacks examples and evaluation.

	evaluates effectiveness.			
Bias and Fairness (3 marks)	Comprehensive analysis of how the company addresses bias and fairness in its AI systems. Includes detailed examples and critical evaluation.	Good analysis of bias and fairness issues. Provides examples and adequate evaluation.	Basic analysis of bias and fairness. Few examples and limited evaluation.	Little to no analysis of bias and fairness. Lacks examples and evaluation.
Transparency and Accountability (3 marks)	In-depth discussion on the company's transparency and accountability measures. Provides specific examples and critical evaluation.	Adequate discussion on transparency and accountability. Provides examples and some evaluation.	Basic discussion on transparency and accountability. Few examples and limited evaluation. Transparency and Accountability (3 marks)	Little to no discussion on transparency and accountability. Lacks examples and evaluation.
Societal Impact (1 mark)	Thorough analysis of the societal impact of the company's AI technologies. Includes specific examples and critical evaluation.	Good analysis of societal impact with some examples and evaluation.	Basic analysis of societal impact with limited examples and evaluation.	Minimal or no analysis of societal impact. Lacks examples and evaluation.
Writing Quality (2 marks)	Well-organized, clear, and properly cited sources.	Organized and properly cited sources.	Some organization and clarity issues. Some citation issues.	Poorly organized. Improper or missing citations.

Part B: Data Science ethics application design protocol. (25 marks)

HousePricePredict Inc. is a tech startup specializing in predictive analytics for real estate. Their primary goal is to develop a robust machine learning model that accurately predicts house prices based on various features. Please refer to the housing_pricing.xls and data_description.txt from more information. By using the given dataset, perform the following.

- Data Understanding and Preparation including feature engineering (3m)
- Model Selection (2m)
- Performance Measurement (2m)
- Hyperparameter Tuning (5m)
- Extra feature and consideration (3m)
- AI Ethic consideration (5m)
- Documentation and Code Quality (3m)
- Project Presentation (Video Recording) (2m)

Marking Rubrics (Total 25m)

Category	**Excellent (A)**	**Good (B)**	**Average (C)**	**Poor (D)**
Data Understanding and Preparation including feature engineering (3m)	Conducts thorough and appropriate data understanding, preparation and applying appropriate data cleaning techniques.	Conducts appropriate data understanding and preparation but may overlook some aspects.	Conducts some data understanding and preparation but may overlook important aspects or not apply appropriate techniques.	No data understanding and preparation.
Model Selection (2m)	Simulate at least 4 models in your analysis. Chooses the most suitable machine learning model based on a deep understanding of the problem, data, and model characteristics.	Simulate at least 3 models in your analysis. Demonstrates good understanding in choosing appropriate machine learning algorithms to achieve the model requirements.	Simulate at least 2 models in your analysis. Demonstrates a basic grasp of selecting machine learning algorithms, though there may be gaps in considering all relevant factors.	Simulate at least 1 model in your analysis. Shows some understanding of machine learning algorithm selection but lacks depth in considering important factors.

Performance Measurement (2m)	Appropriate use and excellent interpretation of performance metrics for evaluating model performance.	Appropriate use of performance metrics for evaluating model performance with minor misconceptions in interpretation.	Use and interpretation of performance metrics with some misconceptions.	Inappropriate use and interpretation of performance metrics.
Hyperparameter Tuning (5m)	Carefully selects a comprehensive set of hyperparameters, considering the characteristics of the model and the dataset. Demonstrates an advanced understanding of the impact of hyperparameters on model performance.	Chooses a good set of hyperparameters, taking into account the model and dataset characteristics. Shows understanding of the importance of hyperparameter tuning.	Selects basic hyperparameters, but some choices may not be well-suited to the model or dataset.	Fails to choose appropriate hyperparameters, resulting in poor model performance.

Extra features and considerations/recommendation (3m)	Assessment of additional features or considerations /recommendation beyond basic requirements.	Implementation of some additional features or considerations/recommendation beyond basic requirements.	Basic implementation of extra features or considerations/recommendation, but lacking depth or innovation.	Limited or no implementation of extra features or considerations/recommendations.
AI Ethic consideration (5m)	Comprehensive and detailed documentation of ethical considerations, including proactive risk assessment. Transparent communication about every aspect of the AI system, including potential ethical challenges. Rigorous measures to mitigate biases, ensure fairness, and prioritize ethical considerations in decision-making.	Comprehensive documentation of ethical considerations. Transparent communication about data sources, model decisions, and potential biases. Measures in place to address and mitigate biases and ensure fairness.	Documentation of ethical considerations with an awareness of potential biases. Transparent communication about data sources and model decision-making. Consideration of fairness and potential impacts on different user groups.	Basic documentation of ethical considerations. Limited transparency on data collection and model decision-making. Some acknowledgment of potential bias without concrete actions to address it.
Documentation and Code Quality (3m)	Code is highly readable, well-structured and modular. Optimized for efficiency. Reveals a thorough and comprehensive understanding of the project, providing high-quality documentation that is clear, well-organized, and insightful.	Code is generally readable and adheres to standard conventions, with adequate modularity. Demonstrates an understanding of the project through documentation, but there are areas that require improvement or may be lacking in certain aspects	Code quality meets basic standards but may have minor issues in readability or modularity. Grasp the project adequately, and with documentation offering a fundamental overview but falling short in depth or detail in specific areas.	Code quality is significantly below standard, hindering comprehension, maintenance, and performance. The documentation is incomplete or unclear, making it challenging to grasp key aspects or severely lacks detail.

Project Presentation (Video Recording) (2m)	Provides a comprehensive code walkthrough, explaining the problem, methodology, and results clearly. The explanations cover key aspects of the code with depth and detail.	Presents the code well, with some room for improvement in organization and clarity. The explanations cover key aspects of the code but may lack some depth or detail.	The code explanation is satisfactory but may be lacking in organization and clarity. Explanations cover basic aspects of the code but lack majority of depth and detail.	The code explanation is lacking in organization and clarity. Explanations are incomplete or unclear, making it challenging to understand key aspects of the code.