
Module Title:	Machine Learning
Assessment Type:	CA3
Assessment Title:	Report
Release Date:	7 th May 2021
Submission date:	23 rd May 2021
Assignment Compiler:	Dr. Muhammad Iqbal
Method of Submission:	Upload one zip file (Report, Code and Datasets) on Moodle
Group/ Individual:	Individual assessment

Learning Outcomes Assessed: Machine Learning

List the module learning outcomes to be assessed (delete as necessary)

MLO 1 - Implement Machine Learning Algorithms to solve analytical problems.

(Linked to PLO 1, PLO 2, PLO 5)

MLO 2 - Determine whether a given data analysis problem requires the use of supervised, semi-supervised or unsupervised learning methods. Develop and implement the chosen learning method.

(Linked to PLO 2, PLO 4, PLO 5)

MLO 5 - Critically evaluate and optimise the performance of Machine Learning models.

(Linked to PLO 3)

Assessment Details

Question 1:

What are the major types of machine learning approaches? Briefly explain them with at least one example for each type. Assume you are working on a data set that deals with an understanding of customer's churn rate using a Machine Learning model. Which frame you would like to use to solve this problem? How much data will you allocate for the training, validation, and test sets? Briefly describe and justify your approach.

(400 words, 20 marks)

Question 2:

Suppose you are working on a classification problem. For validation purposes, you have randomly sampled the training data set into train and validation. The accuracy is high during the validation of modelling results. When you have tested Machine Learning (ML) model on unseen data, the testing accuracy is poor. Can you explain the reasons for this poor accuracy and explain further steps to improve this accuracy of your ML model.

(400 words, 20 marks)

Question 3:

a) Briefly describe the design issues that the data scientists faced during the development of models based on Decision Trees. Demonstrate the difference between Gini Impurity and Entropy in a Decision Tree Models. Provide an example that demonstrate the usage of Gini index and Entropy on the decision trees.

(300 words, 15 marks)

b) What is Regression? Explain the difference between L1 and L2 regularizations. Briefly discuss and highlight the difference between Linear and Logistic Regression by considering any dataset of your choice.

(300 words, 15 marks)

Question 4:

a) Explain the role of Reinforcement Learning (RL) models on a larger canvas of Machine learning and identify the suitable areas for the implementation of reinforcement learning models. Distinguish the active and passive RL approaches.

(250 words, 10 marks)

b) What is role of Natural Language Processing (NLP) in Machine Learning? Highlight the impact of NLP for social media data processing. Consider any dataset or paragraph comprised of (at least 200 words) of your own choice and show that the outcomes of the text analytics operations (for example, tokenization, normalization, stemming and etc.) for text classification. Perform sentiment analysis on the chosen dataset and calculate TF-IDF.

(400 words, 20 marks)

Requirements:

Note:

- The number of words used to answer after each question must be clearly mentioned.
- Different Datasets should be used in CA3 than used in the class.
- Citations and references style must be Harvard in the report.
- Must upload any datasets used, code examples etc, Jupyter notebook along with this submission.

Provide detailed information on the requirements of the assessment:

Requirements
Number of words should be mentioned after every part of the questions
MS word/ pdf copy should be uploaded on Moodle
Captions must be used for the figure numbers/ illustrations

Marking Scheme

Weighting	Module
40% of continuous assessment	Machine Learning

Module	Breakdown	
Machine Learning for Business	Criteria	Weighting
	Q1	0 to 20
	Q2	0 to 20
	Q3 (a)	0 to 15
	Q3 (b)	0 to 15
	Q4 (a)	0 to 10
	Q4 (b)	0 to 20
	(expand as needed)	
Shared	Enter any shared required components eg. Reflective journal	