
Module Titles:	Statistics, Data Visualization and Machine Learning HDip in Sci in AI Applications / HDIP in Sci in Data Analytics for Business (Feb 21 start)
Assessment Type:	Practical and Report
Assessment Title:	Final Integrated Assessment
Release Date:	4 th May 2021
Submission date:	23rd May 2021 (23:59 Irish standard time)
Assignment Compilers:	Aldana Louzán Grandi, David McQuaid and Muhammad Iqbal
Method of Submission:	All program files, report and dataset in one zip file submitted to Moodle (allow submission to one central module only)
Group/ Individual:	An Individual assessment

Learning Outcomes Assessed: Statistical techniques for Data Analysis

List the module learning outcomes to be assessed

MLO 1 - Explore and evaluate datasets using descriptive statistical analyses.

(Linked to PLO 1)

MLO 3- Apply regression analysis to appropriate datasets and demonstrate an awareness of the limitations of regression models.

(Linked to PLO 2)

MLO 4 - Use and understand current software tools and languages to produce result sets from existing data(e.g. Python).

Learning Outcomes Assessed: Data Visualization

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

MLO 1. Develop strategies, incorporating basic programming skills (input / output and basic data structures) for identifying and handling missing and out-of-range data.

(linked to PLO 4)

MLO 3. Engineer new features selection in data with the goal of improving the performance of machine learning models.

(linked to PLO 2, PLO 4)

MLO 4. Critically evaluate and implement suitable data-encoding techniques for a variety of machine learning algorithms.

(linked to PLO 1, PLO 5)

MLO 5. Collaboratively perform a critical analysis of a data set to optimise the data for a given problem space. Document the rationale behind the group's decisions to peers and stakeholders.

(linked to PLO 1,PLO 3, PLO 6)

Learning Outcomes Assessed: Machine Learning

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

MLO1: Implement Machine Learning Algorithms to solve analytical problems.

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

MLO2: 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)

MLO3: Implement a range of classification and regression techniques and detail / document their suitability for a variety of problem domains.

(Linked to PLO 5)

MLO5: Critically evaluate and optimise the performance of Machine Learning models.

(Linked to PLO 3)

NOTE: You may not use a dataset already covered in any module. You Must Source a suitable Dataset with at least 5000 observations and a minimum of 8 features This MUST be included in your upload zip file.

Assessment Details

Section 1 (Statistics):

With your chosen dataset, you should carry out the following tasks:

Question 1:

Perform a complete exploratory data analysis (Summary descriptive statistics, providing a short analysis of the variables in your dataset, as well as appropriate graphical representations to visualize the variables dispersion).

(10 marks)

Question 2:

Pick two variables and build a model that allows you to predict information about those variables. Interpret your results, and provide a short explanation and conclusion based on your findings.

(20 marks)

Section 2 (Data Preparation)

Question 1:

Characterisation of the data set: source URLs; size; number of attributes; has/does not have missing values and number of examples etc.

(5 marks)

Question 2:

Application of Data Preparation methods and EDA visualizations including a clear and concise explanation.

(12.5 marks)

Question 3:

Use of encoding techniques (appropriate to the Machine Model) and PCA for dimensionality reduction.

(12.5 marks)

Section 3 (Machine Learning)

Question 1:

Consider the dataset that you have used in section 1 (Statistics) for Exploratory Data Analysis (EDA) in Question 1. Implement any two appropriate machine learning algorithms (Supervised or Unsupervised) to provide the prediction or classification based on the problem domain of your dataset. Compare the outcomes of both models. Justify that Machine Learning models are not over or underfitted. Write a brief report based on the outcomes of Machine learning algorithms employed in this study that your modelling results can be used for the decision making.

(20 marks)

Question 2:

Describe the role of Reinforcement Learning (RL) models on a larger canvas of Machine learning. What areas/ fields in the real-world are considered as attractive for the implementation of reinforcement learning models? Briefly distinguish the active and passive RL learning approaches using three specific examples. Use illustrations to support your justifications.

[Examples discussed in the class lecture or tutorial notes are not allowed to use Question 2]

(500 – 600 words, 15 marks)

Section 4 (Report)

Appropriate report style and appropriate list of references and Harvard referencing style.

(5 marks)

Requirements:

Provide detailed information on the requirements of the assessment:

Module	Requirements
Statistics	Jupyter notebook using Python programming (you may use colab but the file submission must be ipynb)
Data Preparation	Jupyter notebook using Python programming (you may use colab but the file submission must be ipynb)
	Jupyter notebook using Python programming (you may use colab but

Machine Learning	the file submission must be ipynb)
------------------	------------------------------------

Marking Scheme

Weighting	Module
40% of continuous assessment	Statistics
40% of continuous assessment	Data Visualization
40% of continuous assessment	Machine Learning

Module	Breakdown	
Statistics	Criteria	Weighting
	Summary statistics	0 to 5
	Exploratory visualizations	0 to 5
	Appropriate usage of Python modules and functions	0 to 5
	Linear regression analysis	0 to 10
	Conclusion/interpretation	0 to 5
Data Visualization	Criteria	Weighting
	Characterisation of the data set:	0 to 5
	Application of Data Preparation methods	0 to 7.5
	EDA	0 to 5
	Use of encoding techniques (appropriate to the Machine Model)	0 to 7.5
	PCA	0 to 5

Machine Learning	Criteria	Weighting
	Proper implementation of ML model/ algorithm 1	5 marks
	Proper implementation of ML model/ algorithm 2	5 marks
	Comparison of algorithms	5 marks
	Brief report to justify the results for the chosen algorithms/ models. Jupyter notebook must be provided for the implementation.	5 marks
	Evaluation for the role of Reinforcement Learning. Describe the clear difference between active and passive learning with examples.	15 marks
Report		
	Appropriate report style and appropriate list of references and Harvard referencing style.	0 to 5 marks