Assignment 1 - Report

Start Assignment

- Due 17 Nov by 23:59
- Points 100
- Submitting a file upload

This assignment contributes 50% to your final module mark.

The following generic learning outcomes will be assessed, however please note the **RUBRIC** should be followed in terms of points and structure, the RUBRIC is the table below to see how the marks are awarded:

Knowledge

K1 Demonstrate a deep understanding of data science problems and hence raise awareness of the broad range of tasks/skills required to be a data scientist.

K2 Understanding the broad types of data (including temporal and spatial), variables to suitably employ in the range of techniques for data transformation, fusion and analysis and data presentation and visualisation.

K3 Deep understanding of the Data Science lifecycle (collect, clean, analyse, share, act) and a related broad understanding data science platforms (e.g. Hadoop, Spark) and data science programming languages (e.g., R, Python)

Important Information

You are required to submit your work within the bounds of the University Infringement of Assessment Regulations (see your Programme Guide). Plagiarism, paraphrasing and downloading large amounts of information from external sources, will not be tolerated and will be dealt with severely. Although you should make full use of any source material, which would normally be an occasional sentence and/or paragraph (referenced) followed by your own critical analysis/evaluation. You will receive no marks for work that is not your own. Your work may be subject to checks for originality which can include use of an electronic plagiarism detection service.

Where you are asked to submit an individual piece of work, the work must be entirely your own. The safety of your assessments is your responsibility. You must not permit another student access to your work.

Where referencing is required, unless otherwise stated, the **Harvard referencing** system must be used (see your Programme Guide).

Please ensure that you retain a duplicate of your assignment. We are required to send samples of student work to the external examiners for moderation purposes. It will also safeguard in the unlikely event of your work going astray.

Submission Date and Time   See Canvas for deadline

Submission Location   Via Canvas

Submit a Word document. Not a PDF!

The assignment will be assessed and counts for **50% of the marks** for module CETM24. You are required to produce up to **2,000 word report (max)**. Going over the word limit will result in reduced marks. The references are not included in word count. Also, note you cannot reuse material from other assignments. Please note, no programming or data analysis is required, this is a report only. **NO MORE THAN 10 PAGES MAXIMUM.**

You are free to choose a topic from any area of data science; this can be anything of importance such as economics, finance, industry, or public health applications. Other possible areas can be analysis of internet data, Internet of Things etc. For example, detecting fraudulent patterns in client activity or identifying customers likely to take-up an offer, social network mining. However, please discuss your ideas first with course tutor.

You must be specific - so if for example you choose medical data mining, you should select a particular area of it such as analysis of heart rate data for diagnosing and predicting health issues.

This tool needs to be loaded in a new browser window

CETM24 - Assignment 1 - Report

Criteria	Ratings				Pts
1) Introduction Should describe all the main elements of the report.	**20 to >13.0 Pts Excellent** Excellent introduction that covers all areas of report in great detail.	**13 to >11.0 Pts Good** Good introduction that covers most topics in good detail.	**11 to >7.0 Pts Pass** Some topics missing in introduction or little attempt to discuss them	**7 to >0 Pts Fail** No real consideration of introduction and how it provides a summary of the work	20 pts
2) Application area and data Should describe the chosen domain area. Why it is important and needs data science treatment? What are the challenges and opportunities? What are the types of data encountered in this area and the challenges they present?	**30 to >20.0 Pts Excellent** Excellent consideration of application area. Great understanding of the issues facing company with challenges & opportunities in their business	**20 to >17.0 Pts Good** Good consideration of application area. Clear definition of challenges & opportunities and how to realize them.	**17 to >11.0 Pts Pass** Little consideration of application area. No real explanation of what the area is about does and how challenges & opportunities are to be tackled.	**11 to >0 Pts Fail** No real consideration of what the application area is about.	30 pts
3) Algorithms Describes how algorithms can analyze data. What algorithms have been used in the past and why they have succeeded.	**30 to >20.0 Pts Excellent** Excellent effort made to evaluate and describe algorithms and their features	**20 to >17.0 Pts Good** good/very good effort made to evaluate and describe algorithms and their features.	**17 to >11.0 Pts Pass** Reasonable effort made to evaluate and describe algorithms and their features	**11 to >0 Pts Fail** Little or no effort made to evaluate and describe algorithms and their features.	30 pts
4) Structure Structure, look of report & references	**20 to >13.0 Pts Excellent** Excellent , well structured and professionally presented report. Full use made of	**13 to >11.0 Pts Good** Structure good/very good and reads well. Good use made of references	**11 to >7.0 Pts Pass** Reasonable readability and structuring but could be improved. Some	**7 to >0 Pts Fail** Difficult to understand. Poor structuring. No diagrams, No or few references	20 pts

Criteria	Ratings				Pts
	references, diagrams and tables.	especially journals.	references used, possibly missing details. Few journals.	with incomplete details. No journal article's used.	

CETM24 - Assignment 1 - Report (1)

Criteria	Ratings				Pts
1) Introduction Should describe all the main elements of the report.	**20 Pts** **Excellent** Excellent introduction that covers all areas of report in great detail.	**13 Pts** **Good** Good introduction that covers most topics in good detail.	**11 Pts** **Pass** Some topics missing in introduction or little attempt to discuss them	**7 Pts** **Fail** No real consideration of introduction and how it provides a summary of the work	20 pts
2) Application area and data Should describe the chosen domain area. Why it is important and needs data science treatment? What are the challenges and opportunities? What are the types of data encountered in this area and the challenges they present?	**30 Pts** **Excellent** Excellent consideration of application area. Great understanding of the issues facing company with challenges & opportunities in their business	**20 Pts** **Good** Good consideration of application area. Clear definition of challenges & opportunities and how to realize them.	**17 Pts** **Pass** Little consideration of application area. No real explanation of what the area is about does and how challenges & opportunities are to be tackled.	**11 Pts** **Fail** No real consideration of what the application area is about.	30 pts
3) Algorithms Describes how algorithms can analyze data. What algorithms have been used in the past and why they have succeeded.	**30 Pts** **Excellent** Excellent effort made to evaluate and describe algorithms and their features	**20 Pts** **Good** good/very good effort made to evaluate and describe algorithms and their features.	**17 Pts** **Pass** Reasonable effort made to evaluate and describe algorithms and their features	**11 Pts** **Fail** Little or no effort made to evaluate and describe algorithms and their features.	30 pts
4) Structure Structure, look of report & references	**20 Pts** **Excellent** Excellent , well structured and professionally presented report. Full use made of	**13 Pts** **Good** Structure good/very good and reads well. Good use made of references	**11 Pts** **Pass** Reasonable readability and structuring but could be improved. Some references	**7 Pts** **Fail** Difficult to understand. Poor structuring. No diagrams, No or few references with	20 pts

Criteria	Ratings				Pts
	references, diagrams and tables.	especially journals.	used, possibly missing details. Few journals.	incomplete details. No journal article's used.	
				Total points: 100	