Report

Start Assignment

- Due Thursday by 12:00
- Points 50
- Submitting a file upload
- Attempts 0
- Allowed attempts 2
- Available 19 Feb at 9:00 - 2 May at 12:00

Your task is three-fold. First, you must understand the given topic (Topic 1) in data mining, using either the recommended text or any other resources you can find. Second, you must apply this topic to a data set of your choice and compare and contrast your results with those obtained using the second given topic (Topic 2). Finally, you must demonstrate some level of data preprocessing. You will be able to choose a data set from a selection that will be made available to you. If you wish to use a data set not in this selection, then you need to discuss this with Dr John Evans so that the suitability of this data set can be confirmed. You
are also free to use any other areas in data mining you wish (such as an area discussed in class) if you think this will help your extraction of key features from the chosen dataset.

Your report must be a maximum of 2 sides of A4, not including any code that you write, and your font size should be size 11 with font being Times New Roman. Your margins should be no less than 1 inch/2.5cm. All figures, tables, illustrations etc. should be clearly numbered with an appropriate title. Any references used must be clearly listed at the end of the report and are not part of the page limit. You are encouraged to use LATEX but this is not required.

Please submit your final report, Python code as well as any datasets used.

Topics 1 and 2 will be assigned to you in your individual study group.

Report Assignment

Criteria	Ratings				Pts
Content A maximum of 10 marks will be awarded for content, i.e. ‘is the report correct?’, ‘is the report suitable?’, ‘is the report sufficient?’ etc.	**10 to >7.0 Pts** **Full marks** 8-10 marks for those reports with significant content that is very suitable and has little to no error. There should be an excellent comparison between topics and clear evidence of data preprocessing.	**7 to >4.0 Pts** **Middling** 5-7 marks will be awarded for a reasonable attempt that is relevant and sufficient, but perhaps lacks detail or may have non-trivial errors which are not overly significant. There should be at least some comparison between topics and some evidence of data preprocessing which is not simply superficial.	**4 to >0.0 Pts** **Poor** 1-4 marks awarded for minimal content that either barely addresses the topic or has significant errors. There will be very limited comparison between topics and limited data preprocessing.	**0 Pts** **No marks** No submission or no content on display.	10 pts
Application A maximum of 10 marks will be awarded for application, i.e. 'has the topic been properly applied?', 'is the application fleshed out?', 'is the application suitable?' etc.	**10 to >7.0 Pts** **Full marks** 8-10 marks will be for those who provide a thorough application of the chosen data set and obtain novel or deep results that demonstrate a strong understanding of the topic.	**7 to >4.0 Pts** **Pass** 5-7 marks will be awarded for some clear attempt to apply data mining techniques to the data set, but in which detail is lacking, or the results obtained are uninspired.	**4 to >0.0 Pts** **Poor** 1-4 marks will be awarded if very little effort has been shown (e.g. plugging random numbers into algorithms or Wikipedia style ‘can be used for...’ statements which lack any justification).	**0 Pts** **No marks** No submission or no application.	10 pts

Criteria	Ratings				Pts
Coding A maximum of 10 marks will be awarded for coding, i.e. 'has code of a suitable level been applied?', 'is the student attempting to find deep and interesting patterns in the data, or is it quite superficial?' etc.	**10 to >7.0 Pts** **Full marks** 8-10 marks will be awarded if excellent coding skills are demonstrated.	**7 to >4.0 Pts** **Pass** 5-7 marks will be applied if a reasonable attempt has been made to use Python (or any other suitable programming language) in a way that complements the overall goal of the report.	**4 to >0.0 Pts** **Poor** 1-4 marks will be awarded if only limited coding is applied (e.g. simply using Python to compute the mean of each column in a data matrix, or code in which several parts do not run as intended).	**0 Pts** **No marks** No submission.	10 pts
Originality This is to reward those who demonstrate more than an ability to copy from various resources. Full marks will be awarded for imaginative and interesting applications which fully address the topic.	**5 to >4.0 Pts** **Full marks** Imaginative and interesting applications used which fully address the topic.	**4 to >2.0 Pts** **Pass** This covers those who have either attempted something reasonably imaginative (but not so much that it deserves full marks), or in which an original idea has been attempted, but is not particularly well fleshed out.	**2 to >0.0 Pts** **Poor** 1-2 marks for standard applications which lack any originality.	**0 Pts** **No marks** No submission.	5 pts
Logic and Discussion A maximum of 5 marks will be awarded for logic and discussion, i.e. ‘is the report logically structured?’, ‘is there a suitable discussion of the points made in the report?’ etc.	**5 to >4.0 Pts** **Full marks** The report is logically structured and there is suitable discussion throughout.	**4 to >2.0 Pts** **Pass** Reasonable logic and/or discussion, but in which either or both of these are missing. For example, perhaps there are bits in which logic could be improved, or several parts lack suitable discussion.	**2 to >0.0 Pts** **Poor** 1-2 marks for poor logic, or where discussion is mostly absent.	**0 Pts** **No marks** No submission.	5 pts

Criteria	Ratings				Pts
Spelling, grammar and layout A maximum of 10 marks will be awarded for spelling, grammar and layout, i.e. ‘is the spelling and grammar correct?’, ‘is there an introduction and conclusion present?’, ‘are the references clearly listed alphabetically?’ etc. Roughly, consider a maximum of 3 marks for spelling and grammar, 3 marks for layout and 4 marks (2+2) for references with citation.	**10 to >7.0 Pts** **Full marks**	**7 to >4.0 Pts** **Pass**	**4 to >0.0 Pts** **Poor**	**0 Pts** **No marks**	10 pts
					Total points: 50