

Student: LAUREN COLLINSPercentage: 67%

QUESTION	EXCEPTIONAL 5	MERITORIOUS 4	COMPETENT 3	PASS 2	FAIL 1
Dataset description (15%)	Infers multiple pre-processing AND analysis techniques from descriptive statistics	Infers pre-processing OR analysis techniques from descriptive statistics	Provides descriptive statistics for the dataset	Correctly defines the data types; describes the original data source	Misinterprets data types and meanings
Data pre-processing (10%+10%)	Performs multiple complex primary pre-processing operations	Codes data sensibly and discusses any data removal or manipulation	Performs obvious coding (Yes/No) OR justifies data manipulation	Describes the data used but does not perform any significant pre-processing	Performs no primary pre-processing
	Performs multiple secondary pre-processing operations	Recodes data to allow further analyses AND adds new calculated fields	Recodes data to allow further analyses OR adds new calculated fields	Performs appropriate secondary pre-processing	Performs no secondary pre-processing
Data analysis - data exploration (15%+10%)	Uses multiple methods other than decision trees Full and detailed discussion of the selected model	Uses multiple types of algorithm AND multiple options with each of these.	Uses multiple types of algorithm appropriately and shows a logical progression within an analysis	Uses the same type of algorithm in each analysis	Incorrect, misleading or missing
	Fully tests each model in each analysis	Uses MVT methodology correctly	Includes prediction and comparison of algorithms	Only creates models on the total dataset	Incorrect, misleading or missing
Reporting of results (20%)	Four complete analyses with discussion of the implications of the results	Three complete analyses with discussion of the implications of the results	Two complete analyses with discussion of the implications of the results	One complete analysis with discussion of the implications of the results	Incorrect, misleading or missing
Reflection (10%)	Full and detailed reflection with discussion of how the analysis might be repeated and extended	Discussion of how the research might be repeated OR extended.	Discussion of satisfying analyses and problems encountered.	List of skills learned.	Incorrect, misleading or missing
Report writing (10%)	Reads like a technical review article	Good use of signposting	Clear sectioning of content	Basic use of introduction, body and conclusion; correctly referenced	No clear structure; incorrectly referenced

Good introduction and description of the dataset. Strong discussion of initial visualisation.
Very good handling of missing/incorrect data.
Excellent coverage of primary preprocessing.
Clear statement of research questions.

Showing and comparing the results of multiple algorithms would be useful, as it allows you to get a feel for their different capabilities. Similarly, although you can clearly cluster on two attributes, what about 3? ~~4~~ Or 4? Or 8? When exploring a dataset like this try lots of different things and document how effective they are.

In 5.3.6 you give a confusion matrix - you could compare this with the CM generated by other algorithms or, better still, look at the result and say 'there are so many operating companies that they overpower the dataset - if I remove them, can I distinguish effectively between the other 3 categories?'

How realistic is the \$14M forecast figure? What's the average amount for the recent years? Once you've forecast something like this you should check that it's a sensible forecast.

Data Mining – Final Poster			
Student	LAUREN COLLINS		
Dataset	Start-up Venture Funding.		
Marking			
Abstract (Clarity, readability, flow, accuracy, completeness) The abstract normally also contains a brief outline of the results.			(10 marks) 7
Dataset Description (Where did the data come from? What does the data describe? Is the sample large enough? What data formats and types are involved? What were the objectives of the project?) Nice thorough description.			(10 marks) 8
Data Processing (What primary and secondary preprocessing was needed? How was data integrity maintained?) Multiple preprocessing matters considered, with examples.			(20 marks) 16
Analysis (What analyses were carried out? Are they appropriate?) Needs to say which algorithms are to be used. It's clear that clustering and classification are used, but which algorithms were they?			(20 marks) 12
Results (Are the results clearly stated? Do they meet the objectives?) Clearly presented in the right places.			(20 marks) 16
Graphic Design (Is it balanced and readable? Do the colours work? Are the graphics appropriate?) Layout and colours are good, but the text is too small. Images are on the limit of readability.			(20 marks) 14
General comments Bullet points for some of these would allow you to use a larger text size. Good linkage of questions with results.			
Date	20/5/19	Agreed Final Grade (%)	72
Tutor Signature	Nik Whitehead		
Second Marker Signature	 		

$$14^{47} - 15^{01} + 5$$

Data Mining – Final Presentation			
Student	LAUREN COLLINS		
Dataset	Venture funding dataset.		
Marking			
Presentation/Slides (Clarity, readability, flow, accuracy, completeness, performance) Good confident presentation style. Clear slides, although sometimes a bit small. Try not to read from the slides. Put 1 image for analysis on each slide.			(10 marks) 6
Dataset Description (Where did the data come from? What does the data describe? Is the sample large enough? What data formats and types are involved? What were the objectives of the project?) Good level of detail.			(10 marks) 8
Preprocessing (What primary and secondary preprocessing was needed? How was data integrity maintained?) Clear listing of added attributes. Nice highlighting of specific issues. Good calculation coverage.			(25 marks) 20
Analysis (What analyses were carried out? Are they appropriate?) All questions defined well. Good justification of tool/technique used.			(20 marks) 15
Results (Are the results clearly stated? Do they meet the objectives?) Clear linkage to questions. Good consistent presentation style, although graphs are often a bit too small to read.			(25 marks) 18
Results (What further analyses could be carried out if you had more time?) Plenty of scope for further work. Good examples			(20 marks) 14
General comments Better to put Tableau and Weka visualisations on separate slides. Slow down!			
Date	20/5/19	Agreed Final Grade (%)	84
Tutor Signature	Nik Whitehead		
Second Marker Signature	 		

81