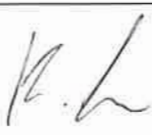


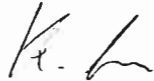
Student: ZHANG YANPercentage: 70%

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

Very good abstract, although it would probably be better to call it an introduction.
Good summary of the initial pre-processing. Good interpretation of the data visualisation.
Excellent use of InfoGain Attribute Eval :)
Good reporting of the use of other algorithms as well as J48.
Remember to state the research question that you're investigating at the start of each analysis section, e.g. in the association analysis section state that you are looking for sets of ~~extra~~ common contributors to accidents. Excellent discussion of the rule sets generated. Good conclusion.
Again, with the time series segment, be clear in stating your research question. This section requires ~~more~~ expansion. Good idea to compare with actual results.
Very good set of conclusions. Sometimes there isn't a big interesting result, but the overall results lead to a better understanding of the problem and its context. This is still all good and useful work.
Also, pre-processing is a pain, and feels as if it will never end. :)

Data Mining – Final Poster			
Student	YAN ZHANG		
Dataset	Road safety.		
Marking			
Abstract (Clarity, readability, flow, accuracy, completeness) This should include the results, not just act as an introduction.			(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 basic information but needs more detail on the significant attributes.			(10 marks) 6
Data Processing (What primary and secondary preprocessing was needed? How was data integrity maintained?) Needs much more detail. What was deleted/converted, and why?			(20 marks) 11
Analysis (What analyses were carried out? Are they appropriate?) What are the research questions? Better to give the questions and say which tools/techniques will be used to investigate each one.			(20 marks) 9
Results (Are the results clearly stated? Do they meet the objectives?) Clear set of results, but no questions to link them too.			(20 marks) 13
Graphic Design (Is it balanced and readable? Do the colours work? Are the graphics appropriate?) Good use of colour. Sensible layout, but be careful that you don't hide words in curved corners. Too much empty space.			(20 marks) 13
General comments A bit too much blank space, which could have been used to provide the extra information needed.			
Date	24/5/19	Agreed Final Grade (%)	58
Tutor Signature	Nik Whitehead		
Second Marker Signature	 		

$$15^4 - 16^0 + 5$$

Data Mining – Final Presentation			
Student	YAN ZHANG		
Dataset	ROAD SAFETY		
Marking			
Presentation/Slides (Clarity, readability, flow, accuracy, completeness, performance) Good clear slides. Good presentation style. Good summary.			(10 marks) 8
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 statement of origins and contents.			(10 marks) 7
Preprocessing (What primary and secondary preprocessing was needed? How was data integrity maintained?) Good description of and justification of changes made. Good use of filtering mechanism to support secondary preprocessing.			(25 marks) 20
Analysis (What analyses were carried out? Are they appropriate?) Good questions, clearly stated. Sensible choice of analysis techniques/tools.			(20 marks) 15
Results (Are the results clearly stated? Do they meet the objectives?) Good discussion of possible reasons/explanations for the results. Good comparison of prediction and actual figures.			(25 marks) 20
Results (What further analyses could be carried out if you had more time?) Sensible extended problems and questions. Good link to driverless cars.			(20 marks) 14
General comments Good use of slides – good interaction with audience.			
Date	20/5/19	Agreed Final Grade (%)	84
Tutor Signature	Nik Whithead		
Second Marker Signature	 		

84