

Student: CAO DONGSHENGPercentage: 59%

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 discussion
of visualisation

Always check your tables of contents and figures before submitting a report. :)

Good description of the data set. Descriptive statistics would be useful.

Clear statement of the research questions to be investigated.

Excellent explanation of the missing data problem and how it was solved by getting more data. Good description of overall preprocessing.

Well-presented discussion of initial visualisations with thoughtful analysis

Junction Control - Tables 1 and 2 are all that's needed initially together with the confusion matrix. There's no need to ~~to include~~ include the extra statistics. Even better would be to compare the results generated by J48 and ~~other~~ other algorithms. Comparing n-folds and percentage split is less important than comparing different algorithms.

~~So~~ Serious/Fatal Collision - JRip is not too good, but what about other algorithms?

Clustering - what do the clusters represent?

Potential for further research is good, but what did you learn/find challenging about this assignment?

More discussion of the results is needed. ~~How~~ How might these results be useful to different user groups?

Data Mining – Final Poster			
Student	DONGSHENG CAO		
Dataset	Road Traffic Collision Data.		
Marking			
Abstract (Clarity, readability, flow, accuracy, completeness) would be better to define the questions and give a brief summary of the results.			(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 clear description.			(10 marks) 7
Data Processing (What primary and secondary preprocessing was needed? How was data integrity maintained?) short and to the point. clearly states what was done.			(20 marks) 14
Analysis (What analyses were carried out? Are they appropriate?) This needs to say what the questions are, and what tools/techniques were used to answer each question.			(20 marks) 8
Results (Are the results clearly stated? Do they meet the objectives?) Difficult to see which questions are answered by which graphic. Good use of future work. Explain the pictures			(20 marks) 11
Graphic Design (Is it balanced and readable? Do the colours work? Are the graphics appropriate?) Too busy visually - needs to lose the background graphic and reduce the number of colours. Layout and text sizing is good. Use the same font!			(20 marks) 12
General comments			
Date	20/5/19	Agreed Final Grade (%)	58
Tutor Signature	Nik Whitehead		
Second Marker Signature	 		

$$14^{29} - 14^{43} + 3$$

Data Mining – Final Presentation			
Student	DONGSHENG CAO		
Dataset	RTINCIDENT DATA.		
Marking			
Presentation/Slides (Clarity, readability, flow, accuracy, completeness, performance) Text is too small/too much to read. Good use of slides – interacting with screen. Clear coverage of problems.			(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?) Clear description of multiple files, distinction between serious & other accidents			(10 marks) 8
Preprocessing (What primary and secondary preprocessing was needed? How was data integrity maintained?) Good description of complex preprocessing. Good use of most common value as filler. Discussed 2ndary preprocessing.			(25 marks) 20
Analysis (What analyses were carried out? Are they appropriate?) Sensible analysis, good discussion of possible causes for the results. Gives details on all analysis tools/techniques.			(20 marks) 14
Results (Are the results clearly stated? Do they meet the objectives?) Clear use of multiple techniques and algorithms			(25 marks) 19
Results (What further analyses could be carried out if you had more time?) Good ideas on both analyses and tools that could be used to further investigate the dataset.			(20 marks) 14
General comments Good answers to questions. Nice confident presentation.			
Date	20/5/19	Agreed Final Grade (%)	84
Tutor Signature	Nik Whitehead		
Second Marker Signature	 		

81