

Student**Lowri Jones****Grade****62%**

QUESTION	EXCEPTIONAL 7	EXCELLENT 6	MERITORIOUS 5	COMPETENT 4	SATISFACTORY 3	CONDONABLE 2	FAIL 1
Purpose & practice of data mining (10%)	Professional-level description	Gives examples of techniques that can be used	Identifies links between statistics, ML and DB fields	Provides suitable examples	Recognises the search for hidden patterns in data	Techniques for handling large data sets	Modern data processing method
Dataset description (15%)	Professional-level description	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	No discussion of the content of the dataset.
Data pre-processing (25%): <i>Primary</i> (15%)	Professional-level data handling	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	Formats the data appropriately to the tool being used	Describes the data used but does not perform any significant pre-processing	Performs no primary pre-processing
<i>Secondary</i> (10%)	Professional-level data handling	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 basic secondary pre-processing	Recognises that secondary processing might be done but doesn't actually do any	Performs no secondary pre-processing
Data analysis (30%): <i>Visual exploration</i> (15%)	Professional-level data visualisation	All of: demographics, relationships, trends, forecasts	Any 3 of: demographics, relationships, trends, forecasts	Any 2 of: demographics, relationships, trends, forecasts	Any 1 of: demographics, relationships, trends, forecasts	Simple plot of full dataset	Performs no visual exploration
<i>Data mining approaches</i> (15%)	Full and detailed discussion of the selected model	Compares at least 3 algorithms AND uses multiple methods other than decision trees	Compares at least 3 algorithms OR uses multiple methods other than decision trees	Provides description of the selected model; discusses the confusion matrix	Provides accuracies for algorithms used; recommends a model	Builds a single model	Incorrect or misleading
Business recommendations (10%)	Full and detailed recommendations with multiple options	Discusses and justifies the recommendation	Explains how the model can be used in practice AND includes B/B items	Explains how the model can be used in practice OR includes B/B items	Provides a business recommendation based on the A/A items	Misleading recommendation	Incorrect recommendation
Reflection (10%)	Linkage to other modules	Discussion of how the research might be repeated AND extended	Discussion of how the research might be repeated OR extended.	Discussion of satisfying analyses AND problems encountered.	Discussion of satisfying analyses OR problems encountered.	List of skills learned.	Incorrect or misleading.

Comments

If you're including figures or tables, include a table of figures and/or a table of tables as appropriate.

Section 2 is a bit long; you'd be better saving the words for the analytical part of the report.

Be a bit more precise about where the data comes from; either say it comes from the UCI Machine Learning Repository or, even better, check the paper listed at the MLR and reference that.

A lot of the Weka visualisations you've put in would be better displayed in Tableau. Investigating the data does allow you to make some inferences about relationships within the data.

Adding the extra attribute in pre-processing is a good idea.

I think you're getting 100% accuracy because you've left the original numerical 1,0 in, and this makes it very easy to predict whether something is true or false. You should use the new field rather than the old field, not as well as it.

A good Tableau viz would be to show time since last donation versus the total amount of blood donated for each of the individual donors (using disaggregated data).

Good business recommendations based upon your finding.