SIT717 2019T2 Assignment 2: Technical Report

SIT717 Enterprise Business Intelligence 2019 T2

Assignment 2 (Project 2): Data Analytic Technical Report

Individual task – 50%+(10% bonus)

Due Date: 11:59pm (AEST), Monday, September 9th, 2019

A document (pdf or word document) should be submitted via CloudDeakin. NO email or Hardcopy assignments accepted. Photos of the document or photos/scanned copy of the handwritten documents are NOT accepted.

[Description]: This project is designed to provide students a good opportunity to use data mining and machine learning method in discovering knowledge from a dataset and explore the applications for business intelligence. It is the second part of the individual project work, and you are required to implement the required analysis together with a written report. This written assessment will be a technical report with no less than 3000 words. This is a hands-on task that requires you to utilize suitable methods and models to explore the data from your chosen topic in Assignment 1. This task evaluates your technical skills on the mining of projected data in real applications. You will practise the problem solving, self-guided information discovery and written communication.

[Tasks and Requirements]:
The content of your technical report should include the following A-H aspects:
A. A meaningful title (5-20 words) followed by your name and student ID
The title describes your topic and points out your research direction, so it is very important and we list the tips below.

[Topic selection tips]:
Tip 1: **The title may be narrowed down further from the title of your Assignment 1.**
Tip 2: Make the best use of Practicals to implement your data analysis project.
First, please learn a framework of basic skills of using Weka to prepare, process and analyze data in Practicals 1-3 and focus on Practical 6 (Performance Evaluation). Practical 6 comprises the skills to compare multiple data mining methods in different metrics, which directly helps your comparison and evaluation of different techniques in your technical report.

Then, you can focus on one of the following techniques for analysing different types of data according to your topic:

· Practical 4 or Practical 5 for processing general numeric data
· Practical 7 Predicting Time Series
· Practical 8 Text Mining
· Practical 9 Image Analysis
· Others learnt by yourself

Practical 10 Recommendation provides an application example using techniques in Practicals 4-9.

Tip 3: **An example.** If the title in Assignment 1 is "*Using Classification Method to Discover Events from Twitter Data*", then the title in Assignment 2 may be "*Using*

Decision Tree to Predict A Event Trend from Twitter Data". If you choose this topic for Assignment 2, you can focus on Practical 8, which provides you details for classifying short text documents using Weka. Here, the event trend can be replaced by "a user's preference trend for a commercial product" and twitter data (taking one tweet as one short text document) also can be other short text data, such as user's comments for the product, etc. You may prefer to Decision Tree (J48 In Weka) as shown in this title, however, you also will present a counterpart method to do comparison.

*B. An abstract (**100-200 words**)*

*C. An introduction of a data analytic application background, motivation and aim (**200-300 words**)*

*D. A summary of your dataset, including data type (general numeric data, short text, time series, image etc.), data size, data quality and data pre-processing) (**300-500 words**)*

*E. The main data mining techniques you adopt to satisfy your application aim (**800-1000 words**)*

In this section, you will point out

- **whether it is a clustering problem or classification problem,**
- **what is the data mining algorithm you will adopt to analyse your data, what are the steps of the adopted algorithm and what are its advantages and disadvantages,**
- **what is the counterpart algorithm that maybe an alternative choice to analyse your data.**

*F. Evaluation and demonstration (**800-1000 words**)*

- **what is the difference between your adopted algorithm and the counterpart algorithm, you may use performance evaluation skills learnt in Practical 6 to compare them. And you must demonstrate (i.e., show result accuracies as evidence) why one is better than another.**

*G. Conclusions (**100-200 words**)*

*H. List of References (**IEEE and Harvard are preferred**).*

Please prepare your references according to the guidance at
http://www.deakin.edu.au/students/study-support/referencing

You can reuse the reference from Assignment 1 and add more publications. Most of them should be formal publications/papers.

[Submission]:

• You must submit your completed document (pdf or word doc) in the Dropbox in CloudDeakin.

• Remember that late submissions will be penalised. Further, the CloudDeakin server is the ultimate time keeper when it comes to determining whether your submission has been received on time.

• You are also reminded to keep a backup copy for record.

Marking Criteria

The technical report of data analysis will be marked using the following marking criteria:

1. (**1 mark**) The title of the report is clearly specified.
2. (**1 mark**) The abstract effectively summarises all the content.
3. (**3 marks**) The introduction is clearly specified the application purpose
4. (**5 marks**) The dataset is described clearly.
5. (**15 marks**) The main techniques (at least two data mining algorithms) are provided in detail.
6. (**20 marks**) Adequate evaluation and comparison of the experimental results for at least two data mining algorithms are provided.
7. (**1 mark**) The conclusion is made and data analysis experience is summarised.
8. (**4 marks**) The technical report is clearly structured (title, abstract, introduction, dataset, main techniques, experimental evaluation, conclusions and references), nicely presented, and well written. The length of the report is within the scope given in the guideline.
9. (**10 bonus marks**) An improved algorithm is provided in pseudocodes based on an existing algorithm. Use the skills learnt by yourself to implement the proposed algorithms by programming, and use experimental results to demonstrate the proposed algorithm is better.

Table 1. Marking Scheme.

Assignment Task 2: 50%+10%(bonus)=60 Marks				
Criteria	Excellent	Good	Marginal	Not Shown
1: Specify a title.	Clear **(1 Mark)**	Intelligible but not sharp **(0.5 Mark)**	Not specified **(0 Mark)**	
2: Specify an abstract.	Effective summary **(1 Mark)**	Limited summary (**0.5 Mark**)	Not specified **(0 Mark)**	
3: Specify an introduction.	A focused purpose **(3 Mark)**	Need further narrow down **(2 Mark)**	Infeasible purpose **(1 Mark)**	Not specified **(0 Mark)**
4: Describe the dataset used	Clear in data type, data size and data quality and preprocessing is conducted **(4-5 Marks)**	Clear in data type, data size and data quality. No preprocessing is conducted **(2-3 Marks)**	Marginal description of data type, data size and data quality. **(1 Marks)**	Not specified **(0 Mark)**
5: Detail the main techniques.	At least two data mining algorithms are provided in detail with adequate theoretical comparison **(12-15 Marks)**	At least two data mining algorithms are provided in detail with limited theoretical comparison **(8-11 Marks)**	Only one data mining algorithm is provided and present clearly **(4-7 Marks)**	Only one algorithm is presented with limited details. **(0-3 Mark)**
6. Demonstrate by experiments	Adequate evaluation and comparison of the experimental results between at least two data mining algorithms. **(15-20 Marks)**	The experimental results between at least two algorithms are provided but with limited evaluation and comparison. **(10-14 Marks)**	The experimental results of only one algorithm are provided and presented clearly. **(5-9 Marks)**	Limited experimental results are provided for any algorithm. **(0-4 Marks)**
7. Specify an conclusion.	Clear **(1 Mark)**	Intelligible but not sharp **(0.5 Mark)**	Not specified **(0 Mark)**	
8. Report with suitable structure and writing skills.	Clear structure, nice presentation and excellent written **(4 Marks)**	Intelligible structure, good presentation and written **(2-3 Marks)**	Marginal content and hard to follow **(1 Mark)**	Not specified **(0 Mark)**
***9. Bonus marks (Improved data mining algorithm is provided and implemented in programming)**	Provide an improved data mining algorithm in pseudocodes based on an existing algorithm. Demonstrate it is better by implementing in programming (screenshot is provided). **(7-10 Marks)**	Provide an improved data mining algorithm in pseudocodes based on an existing algorithm. No programming demonstration but with adequate theoretical analysis. **(4-6 Marks)**	Provide an improved data mining algorithm in pseudocodes based on an existing algorithm, with only limited theoretical demonstration. **(1-3 Marks)**	Not specified **(0 Mark)**

*Not everyone will achieve bonus marks, it is only for exceptional students.