Continuous Assessment 2 (25%) CA683

Project Specification:

Propose and execute a research project using data mining techniques as a team of 3-4 participants. Your research question should be answered using an appropriate method (i.e., via KDD, CRISP-DM) and presented via a conference-style report. You have full freedom to choose any appropriate method(s) but must use appropriate literature to defend your choice(s).

The overarching goal of the project is to leverage methods learned in the module and the course as a whole to execute a significant data mining study. You are expected to source your own data sets for this project.

Submitted papers will be assessed based on their novelty, technical quality, potential impact, insightfulness, depth, clarity, and reproducibility. Code and datasets should be put up on Github and the link should be referred to in the paper.

Project Details

- Teams of a maximum of 5 people.
- Project deadline 19th of April 2020
- **Datasets**
    - Pick a dataset from a published datasets or pick your own

Make sure it is large enough, at least 100,000 records and preferably from a mix of sources, i.e. a database system.

Stipulate the questions you want to answer.

Rules of engagement:

Project breakdown:

- Week 5 24th of February: Project Proposal

- Week 12 19th of April (25%)
    - Final Data Analysis
    - Final Data Modelling
    - Results Presentation
- Marks:
    - Paper 90%
    - Presentation 10% (5 minute video)
    - Report/Paper

Suggested Paper Structure

Abstract: 150-250 words

Introduction remainder of 1st page (+ up to 1 column). Should motivate the work, present and discuss the research question(s) / objective(s) of the paper and (optionally) provide a concise overview of the following sections (max 1-2 lines per each).

Related Work: ideally 1 page (12 or more references in total) – this should not only summarise related work, but also critically evaluate (positive and negative aspects) related work with respect to the topic question, i.e. how well/badly does the related work artefact answer this question, what aspects are useful to consider, what are the limitations, and so on. Also discuss here any foundational papers that substantiate your study design or upon which you build.

Data Mining Methodology (can be named differently): how you have approached answering your question. Additional (technical) detailscan also be discussed here. Essentially, you should recount how you applied either KDD or CRISP-DM to answer your research question(s).

Evaluation/Results – how you have used your methodology to answer the question (evaluation methodology), how do you know your approach is good, results of your evaluation, and a discussion on their implications / impact. If you have to parameterise part of your approach how have you done that, and why were these choices made, and what impacts can different parameterisations have on your results? You should also discuss the results in detail in this section: what are their implications? What do they show / not show, etc.

Conclusions and future work: summarise your findings and discuss limitations / extensions that were you to have more time, you would do next to improve / extend your study. Summarise the partial) answer to the research question(s) at a high level and note the contribution to knowledge the paper has made.

References

Paper Formating and Length Papers must follow the IEEE conference format and should be a maximum of 8 double column pages in length (this includes all figures and references). For this exercise IEEE style referencing, not Harvard referencing, should be used.
Papers over 8 pages (even if it is only 1 word) will be subjected to a 5-percentage point penalty. Word and LaTeX templates are available here:
http://www.ieee.org/conferences_events/conferences/publishing/templates.html

ASSESSMENT	*EXCELLENT >90%*	*COMMENDABLE >=80 & <90%*	*GOOD >=70 & < 80%*	*SATISFACTORY >=50 & < 70%*	*THRESHOLD >=40 & <50%*	*FAIL <40%*
Review of Literature and the rationale for the research (20%)	Excellent critical analysis of substantive and relevant literature leading to compelling rational for the proposed research	Very good critical analysis of substantive and relevant literature leading to convincing rationale for proposed research	Good analysis of the relevant literature leading to clear rational for the proposed research	Adequate analysis of mostly relevant literature leading to an adequate rational for the proposed research	Some review of some relevant literature but limited evidence of understanding and weak rational for proposed research	Little relevant literature reviewed, very limited evidence of understanding and weak rational for proposed research
Research methods, analysis and ethics. (30%)	Excellent application of research principles in terms of appropriate methodology, methods for generating and analysing the data. Ethical issue were all dealt with appropriately.	Very good application of research principles in terms of appropriate methodology, methods for generating and analysing the data. Ethical issue were all dealt with appropriately	Good application of research principles in terms of appropriate methodology, methods for generating and analysing the data. Ethical issue were all dealt with appropriately	Adequate application of research principles in terms of appropriate methodology, methods for generating and analysing the data. Ethical issue were all dealt with appropriately	Weak application of research principles in terms of appropriate methodology, methods for generating and analysing the data. Ethical issue were all dealt with appropriately	Poor application of research principles in terms of appropriate methodology, methods for generating and analysing the data. Ethical issue were not dealt with appropriately

S

ASSESSMENT	*EXCELLENT >90%*	*COMMENDABLE >=80 & <90%*	*GOOD >=70 & < 80%*	*SATISFACTORY >=50 & < 70%*	*THRESHOLD >=40 & <50%*	*FAIL <40%*
Evaluation & Conclusions (20%)	Excellent choice and implementation of the evaluation strategy. The results clearly support the conclusions outlined in the paper. All assumptions were meet.	Very Good choice and implementation of the evaluation strategy. The results support the conclusions outlined in the paper. All assumptions were meet.	Good choice and implementation of the evaluation strategy. The results support the conclusions outlined in the paper. Most of the assumptions were meet.	A reasonable choice and implementation of the evaluation strategy. The generally support the conclusions outlined in the paper. Most of the assumptions were meet.	The implementation of the evaluation strategy was basic. The results could support the conclusions but they are not conclusive.	Little or no evaluation, or the results do not support the conclusions in the paper.
Presentation (10%)	The presentation was excellent, easy to follow and it was obvious the student understood the subject matter. All questions were answered reasonably.	The presentation was very good. It was obvious the student understood the subject matter. All questions were answered reasonably.	The presentation was good. It was obvious the student understood the subject matter. All questions were answered reasonably.	The presentation was good. The student has a reasonable grasp of the subject matter. Answered most the questions reasonably	The presentation was adequate. The student has a moderate grasp of the subject matter. Answered most the questions reasonably (4-<5)	The presentation was poor. The student did not demonstrate that they understood the subject matter. (<4)