ENGG5103 Techniques for Data Mining

Project Guidelines

✓ Project Goals

1. Help students to develop and understand data mining concepts.

2. Help students to get familiar with the entire process of data mining.

3. Improve students' skill in data analysis and model constructing for practical problems.

✓ Topics

There is no topic restriction on the course project. Given numerous real world data sources, you can find practical data mining problems from them, based on your own interests. You should use proper techniques learned from the course to solve it.

You are encouraged to include any data mining techniques that are useful for solving your problem. Please bear in mind the following points when you are doing your topic and data selection.

1. You should apply the techniques you learn in the course. If the method you use falls outside the scope of the course, marks will be significantly reduced.

2. This project must not be the same as any projects you have done or are doing in CUHK or other institutions. If the topic is similar to any of your projects, such as research project, you need to include a declaration and state clearly the **similarity and difference**. Fail to make such a declaration would lead to a zero mark in the project.

3. Include the source of information. Make clear reference to the data source and any others that benefit your project.

✓ Proposals

You are required to submit a project proposal before you start doing your project. The proposal should introduce your project topic briefly. Each group needs to submit one proposal.

1. It should contain appropriate backgrounds of the topic and the objective of your project.

2. Provide the data statistics of the project.

3. Briefly analyze the methods you may use. Out of all methods available at hand, tell that you plan to use which one and why. **In other words, answer question 1 and 2 in the "questions" section in your proposal.**

4. You will receive some comments after the deadline of submitting the proposal. Please take them into considerations when you are doing your project.

5. The total length of the proposal is limited to 2 pages. Please follow the online templates of proposals, and make scientific reports.

6. Groups who do not submit proposals are deemed as **abandoning** the project, and presentation would not be arranged.

7. Proposals would not be graded. However, comments could be important for your final reports.

✓ Reports

At the end of this semester, you are required to submit your project report. In this report, your projects should be presented thoroughly. Each group needs to submit one report. You are suggested to address the following points.

1. **Clearly show the problem you are solving.**
   a) Dataset statistics.
   b) Background.

2. **Clarify your data mining model.**
   a) Mining processes.
   b) Why you choose the method.

3. **Show the results of your experiments.**
   a) Tables.
   b) Figures.
   c) Other informative presenters.

4. **Draw conclusions based on your experiment results.**
   a) The knowledge you discovered.
   b) Why you draw these conclusions.

In addition, some important points you should remember are listed below.

1. The used datasets should be reasonably large so that you can extract meaningful statistical features.

2. Be aware of the weaknesses and strengths of popular data mining techniques. Choose the one that is best for solving your problem.

3. Extract knowledge through the whole data mining process. Be scientific and let the data speak.

4. When you finish your algorithm, the structure of the classifier or clustering results may unveil the knowledge behind data**. Analyze them and tell the hints.**

5. **Please answer all questions in the "questions" section in your report.**

Please make sure that you have put down all important materials (models, data, experimental results) in your report. Careful design of the article structure may help you present your project better.

✓ Questions

Please answer the following questions when you ae doing your project. A successful data mining project could provide answers to all questions.

1. **What's the data and task you are going to finish?**

2. **Which data mining algorithm you choose? Why?**

3. **What's the result of applying the technique to the data?**

4. **How does the result unveil the knowledge hidden behind the data? Can this help people do a better job in the future?**

There would be example project in the tutorials. Refer to the slice to see if you do a good job.

✓ Presentations

At the end of this semester, you are required to present your project (venues and dates will be released later). Please keep the following points in mind when you are preparing the presentations.

1. Show comprehensive understanding of current methodologies used in data mining.

2. You should explain to us that the techniques are suitable for solving your problem.

3. Make your conclusions convincing.

4. Highlight the key points (knowledge you discovered and some important statistical results).

5. Arrive at least **15** minutes before your presentation time slot.

6. Each group has limited time for presentation. If you spend longer time than expected, marks would be deducted.

Since time is limited, you need to learn how to present your job carefully. There would be tutorials on this.

✓ Submission Checklist

Please submit the files (hard copies and soft copies) before the deadline. The following points would help you to check if you have completed the submissions.

1. Submit a softcopy of your **project proposal** to blackboard (only pdf format is acceptable) before the submission deadline (no need for Veriguide check for proposals).

2. According to the policies of the university, students' works must go through plagiarism check to ensure academic honesty. Please submit your final **project report** to Veriguide (https://academic.veriguide.org/academic/login_CUHK.jspx). You have to submit the project report to the **course page** (not the tutorial page) and choose **assignment number 9** for the project report. Please choose the right assignment number. You will get a **declaration form** after submission.

3. Present your project, and submit a hardcopy of your **report** and the **declaration form** to us when you present your project.

4. Submit a softcopy of your **report** and the **declaration form** to blackboard (only pdf format is acceptable) before the submission deadline. (The final version submitted can differ slightly from the hardcopy you submit when you do presentations. If you make significant modifications, include a statement of changes and submit the final report to Veriguide again for plagiarism check).

✓ Deadlines

1. Form project groups (**2** students per group): Oct. 10th, 2016

2. Proposal (at most **2** pages): Nov. 9th, 2016

3. Final report: A few days after the presentation, (exact time to be determined).

4. Presentation: Early Dec, (exact time and venue to be determined).