

Data Mining Fall 2024

Survey Project

Project Content

This project asks you to survey **Data Mining literature reviews** about one application topic (more details on topic selection are given later). The purpose is to give you an up-to-date understanding and analyze state-of-the-art data mining techniques and applications.

Requirements

1. **Team Size:** two persons for each team.
 - Team registration form: <https://forms.gle/jJckmXYREctesF2DA>
2. **Topic Selection:** Six categories of topics are available with a limited team count for each topic, namely *Biomedicine*, *Finance*, *E-Commerce*, *Manufacturing*, *Mobile & Social Networks* (including all kinds of networks related to mobility/transportation of people/vehicles) and *Sports & Entertainment* (not limited to but must include **Sports**).
 - Please check <https://reurl.cc/vv19vl> to select and fill in your preference.
3. **Paper Selection:** Select at least 15 research papers published in journals/conferences. All surveyed papers should be published in recent five years (i.e., 2020-2024). Among them, at least 10 papers must be in recent three years (i.e., 2022-2024), with at least 3 papers in 2024. Please make the problems/applications covered within the selected topic as innovative and diverse as possible. For example, if you select the topic of “Biomedicine,” the covered papers should include different kinds of applications like intelligent disease diagnosis, disease risk prediction, drug discovery, etc., using data mining techniques.
4. **Deliver a report (in **WORD** and **PDF**) containing the following items:**

請注意，每一項均有分配分數；助教會根據下方準則為標準，並依據撰寫的

品質來給分（非有寫就滿分）。

- **Cover Page:** the basic information of your project (project title, team members (系級、姓名), etc.)
- **A list of all surveyed papers,** including complete information for each paper (Author, Paper Title, Source, Year, etc). Give a serial number to each paper for further citation (e.g., #1, #2, ..., etc).
- **Brief Introduction to each paper** (around 1-2 page for each paper), *including*:
 - ◆ Target Problem
 - ◆ Dataset (including the source [with URL link if available], characteristics, etc.)

- ◆ Data Mining workflow (better with a figure) and the main techniques used
 - ◆ Results (in particular, how data mining techniques help with the targeted problem, e.g., help speed up the execution performance by 10 times), etc.
 - A **Summarized Table** for the surveyed results, *including* columns of
 - ◆ Paper Number (e.g., #1)
 - ◆ Publication Year
 - ◆ Target Problem
 - ◆ Adopted Approach/Techniques
 - ◆ Datasets
 - ◆ Results & Contributions
 - **Comparative Study & Discussions** (at least 4 Pages; the more comprehensive, the better)
 - ◆ Pros & Cons of the surveyed papers
 - ◆ **Possible improvements and extensions on this topic (say, biomedicine) by taking all surveyed papers into consideration (note: this part is very important!)**
 - Name the report file as **“team_{team_ID}_survey_report.docx (or .pdf)”**
5. **Deliver paper files (in PDF)**
- Please name each file as the index you assigned in the list of all surveyed papers (e.g., 2.pdf).
 - Pack all paper files in a folder named **“team_{team_ID}_paper_archive”**.
6. **Pack the report file and the paper files folder in a zip file named “team_{team_ID}_survey_project.zip”**

Grading

- This survey project will take **10%** of the final score.

Important Date

- Team registration: **before 2024/11/11 (Mon.) a.m. 11:59:59**
- Topic registration: **2024/11/13 (Wed.) 10:00:00 ~ 2024/11/18 (Mon.) 23:59:59**
- Project deadline: **2024/12/13 (Fri.) 23:59:59**

Notes

- The report can be written in either English or Chinese.

Penalty

- Plagiarism and the use of any AI tools
 - Take any text from somebody else directly. (**Zero score** in this assignment)

- Use any AI composition software (e.g., ChatGPT) directly. (**Zero score** in this assignment)
- Format error or file name error
 - The report is not in pdf and docx. (-5%)
 - Any files have wrong name format. (each -5%)
- Late submission
 - If you submit your work within one day after the deadline, you will get only **80%** of the original score.
 - If you submit your work within two days after the deadline, you will get only **50%** of the original score.
 - If you submit your work more than two days after the deadline, you will get **zero score** in this assignment.