

Project 3

General Introduction:

In this project, you are free to select a topic of your own interest related to data mining. Each team should submit codes and a report via Canvas. The project can be finished individual or with a team no more than two people. A presentation will be scheduled in class.

Complete the following tasks:

1. Clearly describe the problem of your project, with a good motivation to study.
 - a. Examples: strategies in recommender system; disease prediction in medicine; ...
2. Design or apply proper methods to address the problem, use **at least two** different methods for comparison.
3. Evaluate these methods on **at least two** datasets, compare and analyze the performance difference between the methods.
4. Involve different settings (e.g., parameter settings, dataset distribution settings) and discuss how they would influence the methods, share any of your insights.
5. Be creative: explore something novel — e.g., identify a new problem, design your own approach, or improve upon existing methods with fresh insights.

Project Submission:

• Prepare your submission. **One team only needs to provide one submission.** Make a zipped folder named "**CaseID_CaseID_Proj3.zip**", where "CaseID" refers to your group members' Case IDs. In the folder, you should include:

1. **Report:** A pdf file named **Proj3_report.pdf**. Describe your selected problem, methods, and experiments. The report should **not exceed** 5 pages.
2. **Code:** A zipped folder named **code.zip**, which contains all codes used in this part. Inside the folder, please also provide a **README** file which describes how to run your code.

Note that copying code/results/report from another group or source is not allowed and may result in an F in the grades of all the team members.

Grades will be given based on the following criterion:

- Report (70 pts):
 - (10) Problem description: Proper problem selection and description
 - (15) Proper algorithm selection and clear description
 - (30 pts) Result evaluation and analysis
 - (10 pts) Overall coherence and clarity
 - (5 pts) Creativity
- Code (30 pts)
 - (25 pts) Correctness and Reproducibility of the results in the report.
 - (5) Readme: Readability and clarity