2020 Spring Ajou University

MACHIN LEARNING & DATA MINING PROJECT GUIDE

Term project

- Team of 5 members
  - Proposal (1~2 page)
  - Proposal presentation
  - Final presentation
  - Final report (max 10 pages)

Project evaluation

- Project score will be based on peer-review
- Evaluation Criteria:
  - Originality / Difficulty / Completeness / Presentation / Report
- Proportional to team contribution

팀명 (조번호)	이름 (본인포함)	1.아이디어	2. 구현	3. 보고서/ 발표자료	4. 팀워크	계
	계	**100**	**100**	**100**	**100**	**400**

* 1~4 각 기여도 항목에 대해 팀원들의 점수 합이 100이 되어야 합니다.
(예: 조원 3인의 경우, 1.아이디어: 30/30/40, 2. 구현: …)

* 추가하고 싶은 의견이 있는 경우, 아래에 작성하기 바랍니다.
기타의견:

Projects

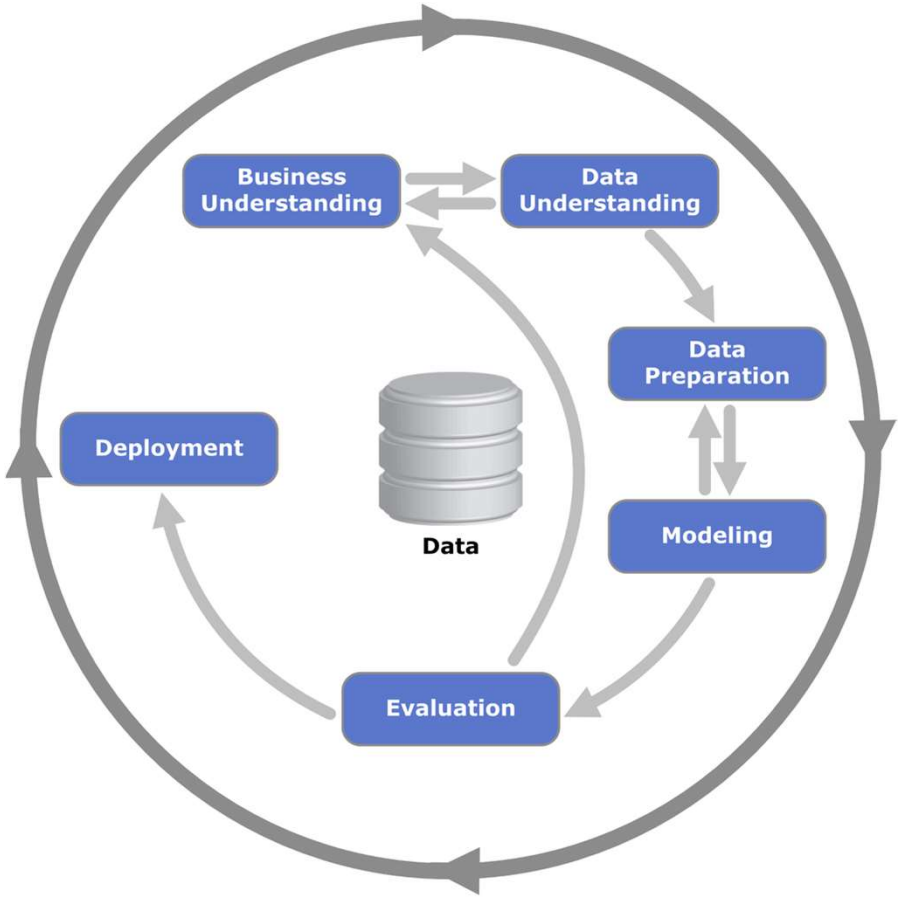
- Project proposal
  - May 15th : Project proposal and presentation file due (per team)
  - Proposal report: 1~2 page
  - (Online) Presentation: 5 minutes per team (5.19 ~ 5.22)
- Final report due(tentative)
  - June 18th: final report & final presentation file due
  - June 19th ~ 26th: final presentation
- Evaluation Criteria:
  - Originality / Difficulty / Completeness / Presentation / Report

Project proposal

- In your project proposal, you should write
  - The title of your project
  - Names of your group members
  - What kind of data you will use
  - What kind of problem you are solving
  - How you would solve your problem (main approaches, candidate algorithms)
  - How you will evaluate your solution

Project final report

- Standard process for data mining
  1. Business Understanding
  2. Data Understanding
  3. Data Preparation
  4. Modeling
  5. Evaluation
  6. (option) Deployment



Cross-Industry Standard Process for Data Mining(CRISP-DM) Overview

Project final report

1. **Business Understanding**
   - Background
   - Describe all requirements, assumptions, risks and constraints
   - Determine data mining goals and success criteria
   - Project plan
2. **Data Understanding**
   - Exploratory Data Analysis(EDA)
   - Describe the data you use or how you collect the data
   - Data quality report(Is there missing data? / Is it clearly identified? / …)
3. **Data Preparation**
   - Data cleaning / pre-processing
   - Describe in details if you construct the new data
4. **Modeling**
   - Model / Algorithm description
   - You may choose different models and compare with each other
   - How you set all the parameters of your model?
5. **Evaluation**
   - Clearly analyze your result
   - Specify some novel or unique findings, patterns, insights you revealed
   - Do the results of your model meet the goals you established?
   - Evaluate and validate the model

Midterm week (May 4-8)

- There will be no mid-term exam this semester
- Instead, you will do survey on possible project ideas and available datasets
- In your report, include at least three different datasets and for each of them
  - Briefly explain the major characteristics (e.g. what kind of features, labels, how many of them, EDA, etc.)
  - Discuss what kind of problems can be addressed using the dataset
  - It can be public datasets (such UCI data repository, or Kaggle sites), or you can think of collecting your own dataset (but this should be a doable one)
  - 3-5 pages