Project Guidelines and Grading Criteria

Updated on 27 August 2022

Project Description

The course project is aimed at allowing students to gain some hands-on experience in solving text mining / machine learning problems. In this project, you are tasked to apply relevant text mining / machine learning tools and techniques that you have acquired through the course (of course, you are welcome to go beyond the scope of the course) to mine useful insights and arrive at meaningful conclusions for a real-world application. You are required to complete this project in groups of 4-5, delivering in three stages, including Group Project Proposal and presentation (5 marks); Group Project Final Report together with codes and datasets (20 marks); Group Project Final Presentation (20 marks). The entire Group Project will contribute 45% to your overall grade. Details of each component, its requirements, and the respective grading criteria will be specified below.

Your final group project marks will be weighted by your project peer evaluation (0 indicates no contribution and 100% indicates full contribution). For example, if your group receives 5 marks for proposal, 18 marks for report, and 17 marks for presentation; you receive an average 90% of contribution for this project, your final project marks should be: (5+18+17)*90%=36. So you must team with your group members and try the best for contribution!

Importantly, regarding group project final presentation, due to our class size, it is infeasible to invite each group presenting 10 minutes within a 150-minute session. However, video recording is not the best way for communication and learning. So a mixing approach will be adopted. Each group will record a full presentation up to 10 minutes and submit by 10 Nov. (Thur.) At the same time each group will have a short live presentation (up to 3 minutes) in the class on 11 Nov. (Friday). Video recording will be shared to the whole class for the purpose of knowledge sharing.

Table of Contents

Project Deliverables

Please name your files in the following format, where XX is your group number.

1. Proposal_XX.pdf for Group Project Proposal
2. Proposal_Presentation_XX.pptx for Group Project Proposal Presentation slides
3. Final_Presentation_XX.pptx for Group Project Final Presentation slides
4. Final_video_XX.mp4 for Group Project Final Presentation video
5. Final_Report_XX.pdf for Group Project Final Report
6. Final_codes_XX.zip for Group Project Completed Codes and relevant datasets

Please submit the following items to the corresponding folder at
Canvas→BT4222→Assignments→Project

Project Component	Due Date and Time	Submission Items
Project Proposal Submission	06 Oct 2022 @ 23:59 (Thur.) online	- PDF Report - Presentation slides (prepared for a live presentation <= 2 mins)
Project Proposal *Presentation*	07 Oct 2022 @ 9:00 (Friday) In-class presentation	- <= 2 mins, highlights the key points in your proposal; it is not necessary for every group member to present.
Final Project Full Presentation Submission	10 Nov 2022 @ 23:59 (Thur.) online	- Presentation Slides (the same as the one used in your video recording) - Recorded Presentation ***Video*** (<=10 mins normal speed; highlight the analysis of your project, your important thoughts in feature engineering, model application or tuning, and key insights; every group member should present)
Final Project short *Presentation*	11 Nov 2022 @ 9:00 (Friday) In-class presentation	- <= 3 mins, a simplified live version of your full presentation submission. It should briefly introduce your machine learning project problem and outcome; and highlight the parts that you want to share with the whole class and the evaluation committee. it is not necessary for every group member to present.
Final Project Report and Codes Submission	10 Nov 2022 @ 23:59 (Thur.) online	- PDF Report - Completed Codes and relevant datasets

Format of Project Proposal and Report

- Cover page: Project Title, Group No., List of group members
- Line spacing: 1.5 lines
- Font: Arial, Calibri, or Times New Roman; Size 11
- Margins: Normal
- Columns: One
- Cover page, reference pages and index page (if applicable) will be excluded from page limit; appendixes (e.g., tables, figures) will be counted.

BT4222 Mining Web Data for Business Insights Academic Year 2022/2023
Semester 1

Project Proposal

You are to submit a PDF report (up to **3 pages**) that covers:

Bullet points and visualization using tables and diagrams are preferred.

1. A description of the problem that you want to address. e.g., "Using IMDB data to predict popular movie story". Why do you think it is an interesting issue and how you could contribute to its solution based on what is available in the market?
2. Some basic facts about the data, including the following:
    - Source of the dataset(s)
    - Data collection methods:
        - If you scrape the data, write about how you make use of scraping tools to get the desired data from sources and to comply with the relevant data acquisition policy.
        - If you download the data, gather the basic information on how the dataset was assembled by the creator.
    - Describe the dataset(s) - e.g., number of observations, unit of each observation (by user/item/page, by date/month/year, etc.), number of variables, type of variables (string, integer, etc.) [You may present the variables in table form.]
3. An explanation of what you think are the interesting aspects that you can look at in the dataset(s), in particular what domain knowledge may contribute to feature engineering? If it is supervised learning, introduce the label (or targets).
4. A discussion of what you hope to mine from the dataset(s)/some hypotheses that you might have in mind that you hope to verify - e.g. “we want to check if the box office, ratings and the profiles of movie list in the past three years could predict the box office of movies with different theme”.
5. A short discussion on how you are going to build a machine learning pipeline, including, how data will be organized and fed into model; possible machine learning methods; how to align between technical metrics and business metrics.

Group Project Proposal Grading Criteria

The Group Project Proposal weighs 5% of your overall final grade. Your proposal will be evaluated based on:

(1) Concision of the introduction of your motivation/problem/issue (1 mark)
(2) Your consideration on feature engineering incorporating domain knowledge (1 mark)
(3) Your preparation of data inputs (1 mark)
(4) Clarity in the statement of your hypotheses/“to-find-out-about” (1 mark)
(5) Correctness and thoughtfulness of the methods you foresee applying. (1 mark)

Group Project Final Report

What to submit: PDF Report (Please include URL to your Github Repo in the report)

You are to submit an **8-10 page** PDF report that extends/continues from your Group Project Proposal.

Group Project Final Report Grading Criteria

The Group Project Final Report weighs 20% of your overall final grade. Your report will be evaluated from the following components:

1. Writing: It should be a scientific essay that is well written and easily to follow, providing concise, precise and logic introduction, description, elaboration and discussion about the design and outcome of a machine learning project, facilitated with necessary tables, figures and diagrams. Please make sure that what is included in your essay is only necessary and meaningful to deliver useful information and achieve your goal. (5 marks)

2. A description of the problem (As stated in (1) of project proposal). (Hint: through model tuning, you should get a deeper understanding about your ML problem, your final report should reflect the reasonable adjustment or refinement compared to what is presented in the proposal) (2 marks)

3. Facts of the datasets and feature engineering including pre-processing steps to sanitize/manipulate/combine your dataset(s) (Hint: instead of replicating what is presented in your proposal, now you need to review the whole data collection process before your writing. Then you can deliver a clear data collection map with key facts and rationale behind.) (3 marks)
4. Model building and performance tuning
    - Machine learning models that you have used and highlight your measures that improve the model performance. Grading will consider the appropriateness of the models and methods applied for analysis, and the consistency with your problem statement (3 marks)
    - Comprehensive performance evaluation, considering reasonable benckmark, including computing cost and timing. (3 marks)
5. Discuss with evidence or references about what has been achieved through your machine learning project from the follow aspects (where they are applicable) (3 marks)
    - Your creativity in feature engineering in a specific domain (e.g., healthcare, telecom, movie industry, social media?
    - Your creativity in applying existing machine learning methodologies / models addressing a specific problem?
    - What has your machine learning models achieved that would be impossible or too costly for human being workforce?
    - What insights has been found through your project?
    - How can business decision making benefit from your project? Or the usefulness of your analysis in other similar real-world problems
6. Discuss the limitations of your machine learning project (1 marks)
    - What are the potential biases?
    - What are the tradeoffs?
    - Anything else…

Group Project Final Presentation

Guidelines about presentation video recording

1. You are to prepare 8-10 minutes worth of slides for presentation and record your group presentation. The flow and content (please select the highlights of your analysis) of the presentation can follow that of the Group Project Final Report.
2. You must include the title and group members at the beginning of the video, either as the first page of your slides or as an overlay text. Make sure that your video stays at the title page for at least 5 seconds.
3. The submitted video can be the recording video of your computer screen.

4. Record your presentation with reasonable image quality and high audio quality. Please ensure that your submitted video will be stored in .mp4 format and is able to play in either Windows or macOS operating systems.

Group Project Final Presentation Grading Criteria

The Group Project Presentation weighs 20% of your overall grade. You will submit your project final report, presentation slides, video (mp4 format compact size), together with codes (preferably in Jupyter notebook, including datasets) on 10 November 2022 by 23:59.

You will be awarded 5 marks as long as your slides and codes are complete and error-free, and the codes shall be able to run smoothly and produce the required results. Up to 15 marks will be awarded, by assessing your overall performance of the in-class presentation and the recorded full presentation, with aspects to the following rubrics:

(1) Clarity in introducing the background and motivating research question (2 marks)
(2) Logical and reasonable explanations of selected methods and analysis (5 marks)
(3) Convincing and intuitive explanation of model performance and key findings (5 marks)
(4) Self explainability of slides content (efficient delivery of message or ideas in each slide) (2 marks)
(5) Good time management during presentations (1 mark)

END