Course Project

STAT 2610SEF Data Analytics With Applications

2025 Spring

1. Introduction

   This project aims to apply what you have learnt in this course. You need to formulate a data analytics problem and solve the problem using data analysis tools developed in R Language.

2. Requirements

   - You should form a group with 6 members to complete the project
   - You must use R language in this project. Make sure the code you submitted can be run. If your application fails to run, your project will have zero marks.

3. Topic

   Report on trends in specific issues (e.g. AI, healthcare, robotics…)

   (you can make your own title in the poster and report)

4. Background

   You are a news reporter. You are now working with a group to report some trends in the world. You and your group need to make a report to show that your team can keep abreast of the world. So, your group plans to use text mining from a lot of articles to analyze the trends. Before submitting the report, your group needs to propose your ideas for the application.

5. Deliverables

   5.1 Proposal (10% of total course marks)

   - One-page A4 size poster to demonstrate your ideas for the application
   - Submit your poster in PDF or docx format to OLE
   - Content (suggested):
     - Background: describe your idea with the minimum of words and preferably with some numbers and figures
     - Features: briefly describe your method of text classification
     - Design: clear and attractive
   - Expectation

     Your group is expected to make an infographic to present an innovative

idea on text mining and classification tasks.

- Marking Criteria
  - Content 5%
  - Design 5%

5.2 Written Report (30% of total course marks)

- Minimum 5000 words report to summarize the features and finding of your application.
- Submit your report in PDF or docx format and the code in R to OLE
- Content (suggested):
  - Collect at least 10 text articles (each containing a minimum of 2000 words) and store each article in a plain text file (.txt) format and name them sequentially as "01.txt", "02.txt", etc. *(reminder: performance of the classification method may be affected due to fewer data samples)*
  - Implement text classification and visualization in R
  - Analyze results
- Report section
  - Introduction: background of the project
  - Dataset: a brief description with analysis of the dataset with a corresponding visualization
  - Method: a brief description of the text classification method you design
  - Results analysis: a brief summary of the classification results of your designed methods together with evaluation results
  - Conclusion: a brief description of the method and the results you obtained
  - References: a list of references
- Expectation

  Your group is expected to have an insight on data analysis
- Marking Criteria
  - Collection of dataset 5%
  - Implementation of an R program 10%
  - Result analysis and report writing 15%

6. Submission

    Each group is required to submit the deliverables by the deadline. Only one member for each group is required to submit. Group number, name and student id should be clearly written in the cover page (report), and in the bottom (poster).