CDS522 Project

2020/2021

Released Date: 18 November 2020 (Wed)
Submission Date: 30 December 2020 (Wed)
Submission method: Softcopy Pdf, Zipped Source Code (eLearn)

1. OVERVIEW

This project is a group project. Two in a group.

You are required to identify a problem and implement the solution of your chosen problem in text analytics.

You are required to write 4-6 pages (max is 8 pages) (including figures/tables/references) manuscript about your chosen topic of text analytics problem as well as solution.

You are required to use R or other similar text analytics tools, libraries or platforms to develop your solution.

2. TOPIC/PROBLEM SELECTION

Problem Requirements:

- The topic needs to present a real problem in a domain, e.g. education, healthcare, transportation, personal, finance, entertainment etc.
- Data sets in a domain can be like clinical record, trip advisor reviews, documents on your laptop etc.
- The problem must be related to text analytics or required text analytics solution.

Example problems/topics

- Restaurant Review Text Analytics, Analysis of Adjectives in Motivational Talk, Text Mining for Jobstreet Company Review, Application of Text Mining and Analytics on Corporate Resume etc.

3. DATA SETS

You will need to find data set based on your defined problem.

Public data set from sources like Kaggle etc. is allowed.

Private data set from companies can be used. If such dataset is used, you will be responsible to seek permission from your company and perform anonymization if the data is confidential.

Self-collected data set is welcome, especially if you are exploring a new domain or have creative problem that data set is available.

The size of the data sets is flexible, as long as you are able to prove the concept of your solution.

4. SOLUTION & ANALYTICS

The analytics performed needs to be related to the problem.

The analytics needs to base on real data (not synthetic) with reasonable scope depending on the problem.

The analytics outcome needs to be supported with technical details/methods/flows/issues etc.

Development of analytics solution requires usage of software libraries like R or similar (with scripting involved). Usage of web site demo tools is allowed for method explanation (e.g. using output of Stanford POS visual), but not counted as deliverable for Project Code component.

Brief evaluation of the analytics outcome using either simple empirical assessment or discussion.

5. SUBMISSION (eLearn)

Report

You will need to submit the report in softcopy format. Suggested length 4-6 pages (max 8 pages) (excluding cover). The report needs to follow manuscript (IEEE format)

Template: http://www.ieee.org/conferences_events/conferences/publishing/templates.html

Source Code

You will need to submit the source code in softcopy format (as single zipped file).

Infographic Poster

You will need to prepare a poster (or similar) for conveying the contents of the report.

The poster will be used during presentation.

A1 size (or have the quality of A1 size but no need to print out)

Softcopy submission only.

Presentation

5 minutes presentation each member based on the poster contents.

6. REPORT GUIDELINE

The report needs to include these sections. (These are guidelines only; you may modify the outlines accordingly to improve your report)

- Title + Abstract
- Introduction

- Problem
- Proposed solution (Solution can be demonstrated by examples, case study, flow of process, framework, lexicon, screen captures, technical details etc.)
- Evaluation criteria (at least one) to verify your solution (Data sets, platform/settings, performance metrics) You can use formal (e.g. empirical) or informal (e.g. expert feedback) evaluation.
- Discussions/Analysis/Findings
- Conclusion
- References
- Appendix of Scripts/Codes/Program (only show sample/main codes if the program is too long)

7. ASSESSMENT COMPONENTS AND CRITERIA

The project will be assessed based on the following breakdown.

- Project Report (50% of project mark)
- Project Codes (25% of project mark)
- Poster (15% of project mark)
- Presentation (10% of project mark).

All above-mentioned output of your project work is assessed by the following criteria:

- Problem selection
- Topic knowledge
- Technical soundness
- Originality and Creativity
- Organization/Readability
- Communication of contents (poster)
- Presentation.

8. SAMPLE

Sample poster

https://padlet.com/gankenghoon/cds522_2019

Sample report (converted to article format)

https://ieeexplore.ieee.org/document/9108706

https://www.europeanproceedings.com/article/10.15405/epsbs.2020.03.03.53