School of Information Technology

Department of Computer Science

COS802-2024

Course Final Project

Lecturer Prof. Vukosi Marivate

Last Revision: 20 Aug 2024 (Version 1.0)

Deadline: 7 and 11 November 2024

1. Project Expectations

One of the main goals of COS 802 is to get you comfortable with NLP and moving towards using state of the art approaches in NLP to solve real-world problems. For 2023 this is looking at wider applications of NLP in Data Science. If you have an interest in the future in doing research in this area, COS 802 also prepares you for this. The project is the main output of this class for you and as such do take time working on it. The project will be done **individually**.

Your first task will be choosing a project topic. For 2024, you will first bid on a dataset and or problem combo. Datasets have been made available on the course ClickUp. It is encouraged you use this to get your topics. You can set up a time to discuss your project (online or physical) ideas with me or Kayode Olaleye. Online office hours will be published.

You can have 3 types of approaches to a project :
- Experimental evaluation of algorithms and models on an interesting dataset;
- A theoretical project that considers a model, an algorithm or a property (measure) and derives a rigorous result about it;
- A scalable implementation of an algorithm.

A typical project will have a mix of the above.

What makes an interesting project?
- Given the topic, you ask questions that may lead to answers that get us to better understand the problem, problem area or the world.
- Provides an extension or possible extension to the current state of the art approaches in the area.

2. Deliverables

Overview

Deliverable	Score/Points
Bid on a dataset and/or problem combo **[2 September 2024]**	-
Project Proposal - Topic and Short Description(with research questions) **[18 September 2024]**	15
Project Report **[11 November 2024]** **6 - 8 pages KDD Format.**	60
Project Presentations **[7 November 2024]**	15
Project Files with Documentation **[15 Nov 2024]**	10
Total	**100**

Project Proposal [1 page, Arial size 11]

The proposal should answer the following questions:
- **Research Questions:**
  - What is the problem you are solving?
  - Why is it interesting?
- **Data:**
  - What data will you be using?
  - How big is the data? Attributes?
  - **Note:** You can use your own data, but make sure you have it by the time the proposal is submitted.
- **Approach:**
  - What methods, algorithms, techniques will you be using? [Be specific]
  - What to do you expect from them?
- **Evaluation:**
  - How will you measure success?
  - Are there baselines?
- **Expected Outputs:**
  - What do you expect your outputs to be at the end of the semester?

Project Report

Format

Template - KDD Explorations https://www.kdd.org/author-instructions
Length: 6 - 8 pages.

Expected Sections

- Introduction/Problem Definition [15%]
- Background/Literature Survey/Related Work [10%]

- Methodology/Model/Algorithm [25%]
- Experiments & Results [30%]
- Conclusion/Future Work/Discussion [10%]
- Writing [10%]

Project Presentation [7 slides, 3-4 minutes]

This is an opportunity to share your work with your other classmates.

3. Resources

See the Dataset Document

4. Rubrics

Available on Clickup.

5. Submission Instructions and Deadline

Submit your report and declaration of originality, in **pdf format**, on the course ClickUP by the due date. No submissions will be allowed after the due date.

6. Plagiarism

This department considers plagiarism to be a serious offence. Disciplinary action will be taken against students who commit plagiarism. For more information on plagiarism, please refer to http://www.library.up.ac.za/plagiarism/index.htm.

Plagiarism is a serious form of academic misconduct. It involves both appropriating someone else's work and passing it off as one's own work afterwards. Thus, you commit plagiarism when you present someone else's written or creative work (words, images, ideas, opinions, discoveries, artwork, music, recordings, computer-generated work, etc.) as your own.

As such you have to write this project report **in your own words**.

You will be required to submit, as part of this project report a declaration of originality.