NLP: Project Requirements

The aim of the project is to give you hands-on experience in NLP, and to help synthesize and apply what you are learning in the course. This course is about NLP techniques for finding hidden patterns in complicated data sets in order to solve real-world natural language related data analytic tasks, and your projects should reflect this. Projects will address either the application of NLP techniques to solve a particular task, or the development of new techniques, validated by experimental results. A typical project might involve:

- Selecting an interesting NLP applications/task to work on, e.g., with industrial significance, or motivated from computational biology, sociology, research, etc. Acquiring a dataset (often by downloading an existing dataset, but could instead involve web scraping, etc.).
- Exploratory analysis and/or visualization to get to know the data.
- Data preprocessing (i.e., Normalizing, feature extraction).
- Implementing and applying one or more NLP techniques to solve the task.
- Rigorous experimental evaluation. For most projects, this will involve cross-validation or train-tests splits, validation sets, and comparisons to baseline methods (**week 3 lecture**).

Note that you may have to read ahead to start your project. All readings are listed on the syllabus. Some important information:

- The project will be done in groups of 2 (expected)
- There are four deliverables for the project, worth 25% of the overall course grade.
    - Proposal 3%
    - Group formation 2%
    - Mid-term report 5%
    - Final Report & Presentation 15%

  Projects that are related to NLP task/research, you or your teammates may be working on, or your job, are, in fact, encouraged. I hope this course helps you solve NLP task you are interested in. However, your project work needs to be your own, and not that of your collaborators.

Project proposal:

Project proposals are to be sent to me by email (mkhan74@kennesaw.edu), and approved by the deadline. Please send me a short summary (< 1 page) of your proposed study, including the goals of the project, why this is important to tackle,

the data you plan to use, the type of models/algorithms you plan to use, and the names of the group members. If you need help selecting a project, please arrange to discuss it with me before the deadline or come to office hours. We'll share our group projects together in class on **Week 6 (09/20/2021- 09/24/2021)**.

Midterm progress report:

This a 1–2-page report. Please summarize the progress made, results obtained, and methods tried. Include a discussion of any challenges faced and plans to resolve them.

Final Report:

Think of the final report as a near-final draft of an academic research paper. The report must be no more than 15 pages, including references. If you plan to publish your work, you should follow the formatting guidelines of the publication venue. It must include a discussion of the relevant work in the literature, and how the work goes beyond this. The report will be evaluated on clarity in addition to technical merit. Late final reports will not be accepted. The report should begin with a title and author names, and include at least the following sections:

- abstract (500 words maximum),
- introduction,
- background and related work,
- methodology,
- experimental results,
- references.

Project ideas / guidelines:

Many types of projects are possible, subject to my approval. It is important that your methods are thoroughly evaluated, including comparisons with simpler approaches, and with competing methods from the literature (if possible). There should generally be an implementation component to the project. Note that research is challenging and uncertain, and it's ok if the project does not yield positive results.

Example topic areas include Entity recognition, question answering, chatbots, summarization, healthcare text mining (disease name recognition, chemical name recognition, etc.,), product name linking, product name searching, sentiment analysis, review analysis, social network analysis, recommender systems, etc.