Data Mining Project

**Instructor:** Ms. Marwa Alrehili

# Due Dates:

|  |  |
| --- | --- |
| **Milestone** | **Date** |
| Project Team Building (3 to 2 students) | 30/ 3/ 2023 |
| Final Project Proposal (Abstract) | 6 / 4 / 2023 |
| Progress Report | 23 / 4/ 2023 |
| Final Project First Draft and presentation | 7 / 5 / 2023 |
| Final Project Final Version | 14 / 5 / 2023 |

# Final Project Format

* Length: 6 to 10 pages, single spaced, not including references (longer papers permitted).
* Use IEEE style guidelines (suggested, not required)

o IEEE templates are available as an [Overleaf template](https://www.overleaf.com/latex/templates/ieee-conference-template/grfzhhncsfqn) .

Many of you will probably use Microsoft Word. However, here are some advantages of using Latex for technical papers:

* 1. automatically generated bibliography.
  2. easier to represent some math formulas; and
  3. automatic application of style sheets. If you do not write Latex with a text editor, graphic interfaces exist, e.g., [Overleaf](https://www.overleaf.com/). It compiles the latex quickly and allows group users to modify the same text.

# Guidelines for a Good Final Project

1. **Model your paper on published conference papers**
   * Check this website: ["How to write a computer science paper"](https://www.froihofer.net/en/students/how-to-write-a-computer-science-paper.html)

* Please note that it is not required to produce a state-of-the- art result, but you should include the following sections (Abstract, Introduction, Related Work, Data, Methodology, Experiments, Results, Discussion, Conclusion, Future Work).

1. **keep track of who did what** and provide this information in the final paper and the first draft.
2. **Use Machine Learning as a tool, but please write a Data Mining paper.**

# Project Objective

In this project, you are expected

1. Select a particular area of Data Mining that interests you,
2. Conduct a literature search on this area,
3. Focus on a specific problem in the area you selected, and Design and implement a novel learning scheme or
   1. Extend an existing scheme to deal with the problem you have identified.
   2. Compare the performance of different existing schemes on the specific problem you have identified.

# The topics are listed as follows:

* + Implement and design a system that recommends content based on the context of the article (Article Recommendation System - compare the performance of various machine learning tools on recommendation system)
  + Implement and design a system that that understands the purpose of your next trip (e.g., business, etc.) and recommends the best hotels based on the reviews and ratings of people who have stayed there for the same type of trip (Hotel Recommendation System - compare the performance of various machine learning tools on recommendation system).
  + Implement and design a system that recommends Product-based Amazon Reviews (Amazon Recommendation System - compare the performance of various machine learning tools on recommendation system).
  + Implement and design a system that recommends movie-based Netflix rating (movie Recommendation System - compare the performance of various machine learning tools on recommendation system).
  + Implement and design a system that classifies customer reviews as positive or negative (detects opinion).
  + Implement and design a system for analyzing Consumer’s Online Buying Behavior.
  + Design sentiment analysis in social media, for example, Twitter, Instagram, Facebook, etc.
  + Road Accidents Investigation and Forecasting Using Data Mining Techniques
  + Emerging topic detection in twitter stream based on pattern mining

# Resources Available for Download

[Kaggle](https://cs.nyu.edu/courses/fall19/CSCI-UA.0480-006/www.kaggle.com) provides data for many machine learning tasks including data mining tasks. For example,[the movie review](https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews)

* Software: Python