

CBD-3335

Data Mining and Analysis

Project

(30% of Final grade)

|  |  |
| --- | --- |
| **Instructor:** | Mohammad Saiful Islam |
| **Class:** | **CBD-3335** |

# Project Description

Students are expected to demonstrate their understanding of the learning outcomes through the Final Project by implementing a solution through proper Analysis of a real-life Data Mining problem.

1. **Objectives**

* Gain experience of collection and handling data from various sources like social media, databases (example, UCI database), NLTK Corpus or real-time data.
* Exploring data sources, processes for making sense of data and methods of relating data to observations and variables.
* Gain experience of exploring big data of real world
* Demonstrate association pattern mining utilizing a variety of algorithms.
* Gain experience of analyzing clusters and outliers using a variety of methods.
* Perform mining for data streams and text data using algorithms.
* Gain experience of applying data mining techniques for time series and explore Neural Networks.

1. **Important Notes**

* Work in groups of 4 students
* All reports must be submitted as a PDF, with source code as an archive (e.g., zip, tar.gz)
* Save the submission as “Course#\_Project\_Group-name.zip”

1. **Task Details**

First, you must submit a project proposal for your Final Project. It would be best if you were specific about the following sections in your project proposal.

**3.1 Motivation:**

What problem are you like to work with?

**3.2 Method:**

What learning techniques are you planning to apply? T to propose a method based on what you have learned in class and your preliminary research into the topic you have selected for your proposal. Please mention where the dataset should come from.

**3.3 Intended experiments:**

What experiments are you planning to run? I want to see if you have started thinking about these problems and provide you with helpful feedback.

**3.4 Planning and Milestones:**

How do you plan to divide the project into milestones and tasks? For projects with multiple group members, which member(s) is responsible for each task/milestone.

You should score well on the proposal if your submission follows the requirements above and the project appears to have been thought out with a good strategy. In the next class, at least two team members should present this. Your fellow mates and I will be able to give you some constructive feedback on the plan's viability.

The Project Proposal can be up to 2-pages, IEEE format, covering everything that is being asked.

The Final project should be based on the proposal, reflecting the feedback given at the proposal presentation. The final project report must be presented with quality visualizations, both for Exploratory Data Analysis (EDA) and for obtained result presentation.

1. **Deliverables:**

The deliverables for the final project should be:

1. Proposal for the project
2. Presentation with the results (all four group members must present).
3. A 7–10-page Project report in IEEE double-column format.
4. Python Codes