

TEAM PROJECT

Project Overview

The data mining project is designed to provide students with hands-on experience in exploring and analyzing real-world datasets using various data mining techniques. Students will work in teams of 3-4 members to tackle moderately complex datasets or a combination of public datasets.

The project will focus on solving at least two data mining problems, including but not limited to clustering, graph mining, recommendation engines, and RFM analysis. Proper Exploratory Data Analysis (EDA) techniques should be employed to understand the data and derive insights.

Project Requirements

1. Dataset Selection:

- Choose a moderately complex dataset or combine multiple public datasets to ensure richness and diversity in data.
- Ensure the dataset(s) provide ample opportunities for applying different data mining techniques.
- Define a high-level business problem you are trying to solve using this dataset.

2. Exploratory Data Analysis (EDA):

- Conduct thorough EDA to understand the structure, patterns, and characteristics of the dataset(s).
- Utilize appropriate visualizations and statistical methods to gain insights into the data.
- Document key findings and observations from the EDA process.

3. Data Mining Problems:

- Select at least 2 data mining problems to solve using the chosen dataset(s).
- Examples include:
 - Clustering: Grouping similar data points together using techniques taught in the class.
 - Graph Mining: Analyzing relationships between entities in a dataset.
 - Recommendation Engines: Suggesting items to users based on their preferences.
 - RFM Analysis: Segmentation of customers based on their behavior.
- Teams are encouraged to propose additional data mining problems if suitable for the dataset.

4. Implementation of Data Mining Techniques:

- Apply relevant data mining techniques to address the selected problems.
- Implement algorithms and methodologies discussed in class or found in literature.
- Ensure proper validation and evaluation of the results.

Project Material

- Prepare a comprehensive set of slides detailing the project's objectives, methodology, findings, and conclusions.
- Include code artifacts (e.g., Python notebooks, R scripts) used for data preprocessing, analysis, and model implementation.

- Document any libraries or frameworks utilized for the project.
- Provide clear explanations and annotations within the code for better understanding.

Submission Guidelines

- Submit the project slides and code artifacts by the specified deadline.
- Ensure all team members are listed and credited appropriately.
- Submissions should be organized and well-structured, following the provided template if available.
- Any additional documentation or supplementary materials should be included as needed.

Evaluation Criteria

- Quality and complexity of the dataset(s) chosen.
- Effectiveness of data mining techniques applied.
- Depth of Exploratory Data Analysis conducted.
- Clarity and coherence of project presentation.
- Completeness and quality of code artifacts.
- Creativity and originality in problem-solving approaches.

Note: It is recommended that teams regularly communicate and collaborate throughout the project to ensure successful completion and a well-integrated final submission.