Data Mining Report

Your Name

October 3, 2024

Abstract

This report details and highlights the work performed in the context of data mining. It includes motivations, thorough analysis, observations, and limitations for each task.

1 Introduction

Provide an overview of the report, including the main objectives and structure.

2 General Guidelines

This section provides a guide on the expected report's contents. Please keep in mind that the report has the goal of detailing and highlighting your work. As such, all analyses must contain:

- **Motivations**: Why did you perform this analysis, rather than another one? Why did you look to create one feature/representation, rather than another?
- **Thorough analysis**: Each analysis should be performed in a reasonably large set of settings, e.g., considering several hyperparameters for the algorithms you run, and, when appropriate, choosing a set of hyperparameters. Please justify your choices.
- **Observations**: What insight and/or information did you gain from each analysis you performed?
- **Limitations**: How strong are the analytical results, and observations you have found?

3 Task 1: Data Understanding

3.1 Assessing Data Quality

Analyze the dataset, including assessing data quality.

3.2 Data Distribution

Analyze the data distribution.

3.3 Relationships Between Features

Analyze the relationships between features.

4 Task 2: Data Transformation

The data transformation task includes three subtasks:

4.1 Feature Engineering and/or Novel Feature Definition

Improve the quality of your data by tackling eventually missing/incorrect values, either engineering or defining novel features of interest.

4.2 Outlier Detection

Detect and handle outliers in the dataset.

4.3 Revamped Data Understanding

A revamped data understanding task, now including the features of point 1, and eventual considerations of point 2.

5 Task 3: Clustering

Leverage clustering algorithms to identify and describe the groups of instances you have found. The section should consider all clustering algorithms tackled in the course:

- k-means clustering
- Density-based clustering
- Hierarchical clustering

5.1 Final Observations and Comparisons

Present final observations and comparisons on different clusterings. Additionally, the group can experiment with additional clustering algorithms available here. Additional algorithms can yield up to 2 bonus points in evaluation.

6 Conclusion

Summarize the main findings and insights from the report.