Machine Learning and Data Mining

Project Proposal

1. Group of students

Hoàng Văn Phương: 20200478

Trương Anh Huy: 20200287

1. Introduction

Clustering is an unsupervised machine learning technique with a lot of applications in the areas of pattern recognition, image analysis, customer analytics, market segmentation, social network analysis, and more. A broad range of industries use clustering, from airlines to healthcare and beyond.

1. Problem

The system needs to group the students based on a number of predefined attributes(e.g., age, gender, score, the number of registered courses for the semester, etc..).

Input: a file containing a set of vectors that present for students. Each element of a vector is equivalent to an attribute of a student.

Output: group of students after being clustered.

Approach: using K-means clustering or FCM clustering

Dataset: UCI dataset (**Turkiye Student Evaluation Data Set)**

1. Comparation and analysis

Analysis: give comments about the result of clustering.

Comparation: using some performance evaluation to appreciate the clustering result.

* RMSSTD (root-mean-square standard deviation)
* R-squared
* Dunn index
* Davies-Bouldin index