Descriptions

Database Files: SQL Scripts

Database Creation SQL: Initiates the creation of all tables in the Database. The result of Forward Engineering in MySQL Workbench.

Insert Test Data: Insertion Script to insert multiple test student records into the student’s databased

Machine Learning

KMC- Machine Learning Algorithm written in Python to Match students to potential roommates.

The algorithm begins with querying the data from the database.

During the data preprocessing the features necessary for roommate recommendations are selected and undergo a process to convert the ordinal variables to numeric

The features are then weighted based on their importance (Age has the biggest weight to prevent a large age gap in roommate suggestions)

Due to the initial data not having a target label. KMeans Clustering is initiated to group similar students together. Only 2 clusters were generated

Based on the cluster of the current student that is selected. All other students will be assigned a label of 1 if they are in the same cluster to indicate a likely match. Other students will be assigned a 0.

The KNN classifier is then invoked to predict additional roommates. The roommates are then provided with a compatibility score that is based upon their distance from the current student.

The accuracy rate is not as high as we would like due to the n-size of 106 we were only able to optimize it to yield an 87% accuracy score.