CS513C-KDD PROJECT PROPOSAL: Student Grade Pass or Fail Prediction

Project Group No: 2

Problem Statement:

There are many factors influence performance of students including demographic, social and school-related features and so on. The final grade of students will be affected by these lots of features. Take the students' scores in the third semester(G3) as the prediction standard and set a grade greater >=10 as pass.

Therefore, our task is predicting pass with all attributes except G3. Given, a set of features we want to predict the target variable as 1, which is defaulter or 0 which is non-defaulter.

Dataset:

This data approach student achievement in secondary education of two Portuguese schools. The dataset comprises 30 features in form of columns, out of which we may opt to use the essential features only, during implementation. We can use PCA and correlation for feature reduction, but we shall see.

Source of Dataset:

https://www.kaggle.com/datasets/dinhanhx/studentgradepassorfailprediction

Implementation Strategy and algorithms Used:

Group members have decided to implement and compare 8 different models among four different group members. We have chosen a few models from this course and few from outside the course. The following are the models selected by us:

a. Logistic Regression
b. Support Vector Machine (SVM)
c. K-Nearest Neighbor (KNN)
d. Random Forest
e. Naive Bayes
f. Decision Tree
g. Linear Discriminant Analysis (LDA)
h. Quadratic Discriminant Analysis (QDA)

Model metrics and Evaluation:

AUC-ROC, Confusion matrix, F1, Recall, Precision.

Team Members Names:

Group 2

a. Jiangrui Zheng
b. Junran Yang
c. Suqi Lu
d. Zhengyuan Han