The goal for the project of building a student intervention system is to identify students who might need early intervention before they fail to graduate. This problem is a supervised classification.

Data file is student-data.csv. We need to split the data into training and testing set for building the model. Three supervised classifiers are applied here including Logistic Regression, Gaussian Naive Bayes and Random Forests with different sizes of the training data. F1-score is measured to evaluate the model performance.