Abstract:

The goal of this project is to use classification models to predict students' performance levels to achieve the best results and reduce the failure rate, educational methods require innovative techniques to develop education quality by using machine learning algorithms.

Design:

This project is one of the T5 Data Science BootCamp requirements. Data obtained from Kaggle website. https://archive.ics.uci.edu/ml/datasets/student+performance

Data:

This data approach student achievement in secondary education of two Portuguese schools.The data attributes include student grades, demographic, social and school related features and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por).

Link: https://archive.ics.uci.edu/ml/datasets/student+performance

Objectives :

- Predict if students will pass the exams by their data
- Compare between the selected learning algorithms
- Discover the result behand students' failures

Feature Engineering:

- map each string to a numerical value
- use feature scaling method to normalize the range of independent variables or features of data

Models Used:

I will use different classifiers such as KNN and SVM classifier and compare between them and achieving a high result

Tools:

1. Python language
2. Numpy and Pandas for data manipulation.
3. Scikit-learn for modeling.
4. Matplotlib and Seaborn for plotting.
5. use library KNeighborsClassifier and SVC