Machine Learning Nanodegree Project 2

1. Classification vs Regression

Q: Which type of supervised machine learning problem is this, classification or regression? Why?

2. Exploring the Data

Q: Can you find out the following facts about the dataset?

Total number of students	
Number of students who passed	
Number of students who failed	
Graduation rate of the class (%)	
Number of features (excluding the label/target column)	

3. Preparing the Data

Q: Execute the following steps to prepare the data for modeling, training and testing:

1. Identify feature and target columns
2. Preprocess feature columns
3. Split data into training and test sets

4. Training and Evaluating Models

Q: Choose 3 supervised learning models that are available in scikit-learn, and appropriate for this problem. For each model:

1. What are the general applications of this model? What are its strengths and weaknesses?
2. Given what you know about the data so far, why did you choose this model to apply?
3. Fit this model to the training data, try to predict labels (for both training and test sets), and measure the F1 score. Repeat this process with different training set sizes (100, 200, 300), keeping test set constant.
4. Produce a table showing training time, prediction time, F1 score on training set and F1 score on test set, for each training set size.

Table 1: Model 1 (INSERT MODEL DETAILS)

Table 2 INSERT MODEL DETAILS

Table 3 INSERT MODEL DETAILS

5. Choosing the Best Model

Q: Based on the experiments you performed earlier, in 2-3 paragraphs explain to the board of supervisors what single model you choose as the best model.

- **Which model has the best test F1 score and time efficiency?**
- **Which model is generally the most appropriate based on the available data, limited resources, cost, and performance?**

Please directly compare and contrast the numerical values recorded to make your case.

Q: In 1-3 paragraphs explain to the board of supervisors in layman's terms how the final model chosen is supposed to work (for example if you chose a decision tree or support vector machine, how does it learn to make a prediction).

Fine-tune the model. Use gridsearch with at least one important parameter tuned and with at least 3 settings. Use the entire training set for this.
What is the model's final F1 score?

Final F1 Score	