

P2: BUILDING AN INTERVENTION SYSTEM

ANDY PEREZ

1. CLASSIFICATION VS REGRESSION

This is a classification problem since we are identifying what categories our students belong to. They are either at risk to fail, in which case they need assistance, or they are not at risk to fail, in which case they do not need assistance. Although internal to the project there may exist an algorithm that assigns a numerical value related to the likelihood of a student failing, the aim of the project, ultimately, is to spit out whether a student needs or does not need help in order to graduate.

2. EXPLORING THE DATA

Can you find out the following facts about the dataset?

- Total number of students
395
- Number of students who passed
265
- Number of students who failed
130
- Graduation rate of the class (%)
67.09%
- Number of features (excluding the label/target column)
30 (the last feature in student_data is whether the student passed, the label column)

3. PREPARING THE DATA

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

- ✓ Identify feature and target columns
- ✓ Preprocess feature columns
- ✓ Split data into training and test sets

4. TRAINING AND EVALUATING MODELS

4.1. AdaBoost.

- What are the general applications of this model? What are its strengths and weaknesses?
- Given what you know about the data so far, why did you choose this model to apply?

- Time consumption and F1 score table

4.2. **Bagging (with decision trees).**

- What are the general applications of this model? What are its strengths and weaknesses?
- Given what you know about the data so far, why did you choose this model to apply?
- Time consumption and F1 score table

4.3. **Naive Bayes.**

- What are the general applications of this model? What are its strengths and weaknesses?
- Given what you know about the data so far, why did you choose this model to apply?
- Time consumption and F1 score table