Predicting Whether a Student Will Fail Given Their Alcohol Consumption

Dataset

My data set has information on a person's traits and interests and familial background including age, sex, general health, family size, mother and father's relationship status, study time, extra curricular activities, weekday alcohol consumption, weekend alcohol consumption and more. The focal point of my data is the failures section. Failures is measured between 0 and 5. I will be making this binary by making a result over 0, a fail.

Proposal

Using these traits such as alcohol consumption on a weekday or weekend and study time, we are going to predict whether they will fail a class or not. I will only be testing IF they will fail not how many fails they will have. I will only be using these columns because these definitely carry the most weight.

Software

First I will need to merge the data by reading them in and using a pandas function "concat()". Then I will focus the data on certain key columns. The columns are referenced above: Weekday Alc. Consumption, Weekend Alc. Consumption and Study Time. These are obviously more stand out attributes detailed in this dataset. For example: Mother's occupation is definitely less important than alcohol consumption during the week.

Relevant Papers

1. A study I found on the Kaggle page for this dataset, done by Datai, is an in depth look and analysis at students chance of success. This is basically a rewording of my proposed study. They are also focusing on Weekday Alc Consumption and Weekend Alc. Consumption. This study also provides a lot of extra infographics using R. These infographics show a lot of extra interesting information like total alc consumption and a pie graph to visually represent the amount of students that drink at that level. [1]

2. Another study found on Kaggle, is done by a student that uses machine learning algorithms using python, pandas and scikit learn and more. This is incredibly close to my approach and library use. There is also a lot of mathplot library usage in order to create graphs for information display. [2]

References

1. https://www.kaggle.com/kanncaa1/does-alcohol-affect-success
2. https://www.kaggle.com/rmalshe/plots-correlations-predictions/notebook