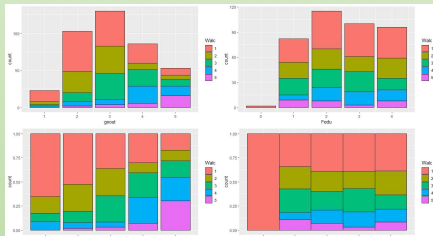# Determining Factors That Influence Student Alcohol Consumption

Aditya Bhattacharjee | Faculty Mentor: Dr. Chantal Larose
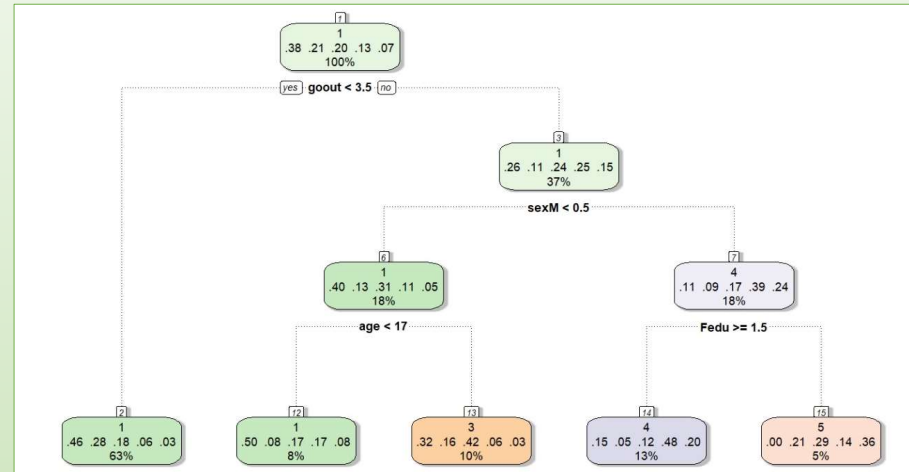Eastern Connecticut State University

Introduction

This project was determining the factors that influence a student's alcohol consumption on the weekend. This analysis can be utilized to fix potential problems with students who believe their alcohol consumption is negatively affecting their education.

Predictor Variable Analysis

- A higher percentage of students don't go out but when they do, alcohol consumption is much higher in all categories
- When a student's father education is low, they are more likely to not consume alcohol on the weekends

References
https://www.kaggle.com/uciml/student-alcohol-consumption

CART Model

Methodology

CART was used on this dataset of students taking math and Portuguese language courses in secondary school. Using a CART model which explains how an outcome variable's values can be predicted based on other values, the most important factors that influence student alcohol consumption were determined. Further calculations were made to determine the accuracy and sensitivity of the model.

The decision rules that can be harvested from this model are as follows:

- If goout is less than 3.5, then a student falls into category 1 46% of the time. 63% of the records fall into this leaf node.
- If goout is not less than 3.5 and sexM is less than 0.5 and age is less than 17, then a student is category 1 50% of the time or in category 3 42% of the time. 8% of the records fall into the left leaf node and 10% into the right leaf node.
- If goout is not less than 3.5 and sexM is less than 0.5 and Fedu is greater than or equal to 1.5, then a student is category 4 48% of the time or in category 5 36% of the time. 13% of the records fall into the left leaf node and 5% in the right leaf node.

Accuracy of the model on the training data is: $Accuracy = \frac{135}{298} = 45.30\%$

Accuracy of the model on the test data is: $Accuracy = \frac{40}{97} = 41.24\%$

Category 1:

Sensitivity: 83.78%

Specificity: 31.67%

Model Findings

- With a high sensitivity, we can correctly identify the students who consume alcohol on the weekend more frequently.
- With low specificity, there is a small risk of a higher False Positive Rate.

Future Work

- The model still has lower specificity than what I would like.
- The model could also be used to determine how alcohol consumption is affecting the grades of students.
- Further measures may be required in order to understand why the father's education matters more than the mother's education

Contact Information

Aditya Bhattacharjee
bhattacharjeea@my.easternct.edu