

Patrick O'Boyle
CSC 346 Introduction to Data Science
Project 5
1/24/2022

Problem 2:

1. The purpose of this problem was to demonstrate our understanding and use of Bayes Theorem. We used the theorem to predict red or white wine based on alcohol content and sugar content. We used the probabilities of the alcohol content and the sugar contents to train our model to make inferences about the data. We first took our prior probabilities and applied Bayes' theorem and got our posterior probabilities.
2. Methodology: First we imported our data from github and then converted the string labels of the column into numbers which is label encoding. After that contingency tables were made for type vs. alcohol content and sugar content. Then the probabilities are gathered for the alcohol contents we had to get the conditional probabilities for them. After that our data sets were tested on different ratios.
3. Conclusion: What I was able to conclude from the data is that the red wine had more of a higher alcohol content as well as lower sugar content. I found this through looking at my conditional probabilities and the Gaussian Naive Bayes models.
4. References: I used mainly the Naive bayes slides on some online sources.
https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.value_counts.html