Assignment 1

Overview

Your assignment is to design a simple, low-cost sensor that can distinguish between red wine and white wine.

Your sensor must correctly distinguish between red and white wine for at least 95% of the samples in a set of 6497 test samples of red and white wine.

Your technology is capable of sensing the following wine attributes:
- Fixed acidity - Free sulphur dioxide
- Volatile acidity - Total sulphur dioxide
- Citric acid - Sulphates
- Residual sugar - pH
- Chlorides - Alcohol
- Density

Tasks

Read WineQuality.pdf

Use the wine.csv file that is provided.

Build decision tree model to predict red vs white wine (you can use R or Python)

Submission

Submit the code file(s) or notebook used to implement your model along with a 1-page document (prefer PDF or an .ipynb notebook) outlining your solution and answer the following questions:

1. What is the percentage of correct classification results (using all attributes)?
2. What is the percentage of correct classification results (using a subset of the attributes of your choosing)?
3. What is the AUC of your model (all features vs feature subset)?

4. Visualize your decision tree
5. What is the best AUC that you can achieve? (include any lesser trials/experimental models you built while optimizing)
6. Which attributes are the most informative for this prediction task? How did you identify them?