2022年8月2日 星期二

紅酒品質預測

資料欄位

1 - fixed acidity:固定酸度

2 - volatile acidity：揮發性酸度

3 - citric acid：檸檬酸

4 - residual sugar：殘糖

5 - chlorides：氯化物

6 - free sulfur dioxide：游離二氧化硫

7 - total sulfur dioxide：總二氧化硫

8 - density：密度

9 - pH：酸鹼度

10 - sulphates：硫酸鹽

11 - alcohol Output variable (based on sensory data):酒精輸出變量

12 - quality (score between 0 and 10)：品質(區分6.5以上是好，以下是壞）

敘述性分析

敘述性分析發現，都沒有缺失值，但是每個數值欄位數字差異有點多，每個欄位數字差異有點大，數值範圍還要再做調整，還要將quality要區分好跟壞比較好做預測。

Fixed acidity、free sulfur dioxide、total sulfur dioxide、density、alcohol差異看起來比較大

quality在做區分好與壞之後，可以發現有資料不平衡的狀況，需要做調整

資料關聯性

由關聯性圖表可以發現跟紅酒品質比較有相關的係數是alcohol其他相關係數偏低

Fixed acidity跟citric acid關聯係數高，都同屬酸性，可能保留一個即可



Free sulfur dioxide跟total sulfur dioxide兩者關聯係數高，屬於同一種類係數，可能保留一個即可