



Creating a Regression Model to Predict Wine Quality

Group members: Claire Navara, Kelly Schultz, Rajiv Chaudhary

Introduction

Our dataset focused on the attributes of a white wine variant of the Portuguese “Vinho Verde” wine. We began our analysis with 4,898 rows and 12 attributes, quality being the target attribute.

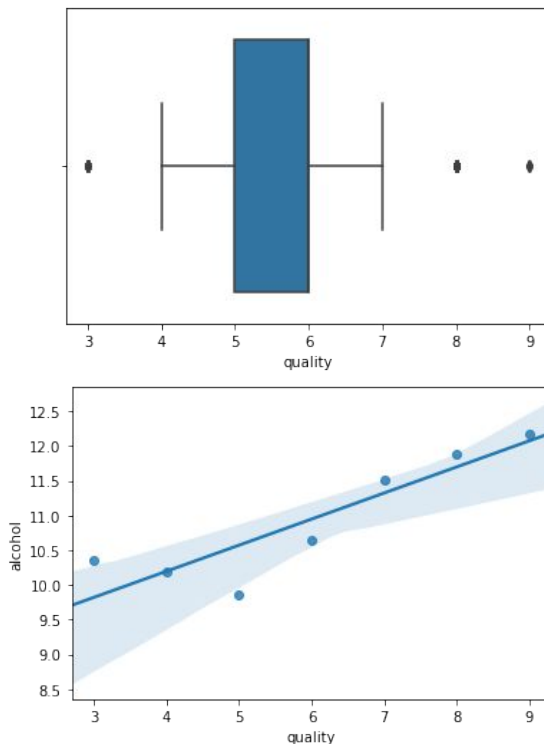
Problem

Given the chosen dataset, we wanted to create a regression model to predict white wine quality based on the composition of the wine.

Hypothesis

To better understand our attributes, we did some preliminary research. From this we expected attributes associated with acid, chlorides, and alcohol to have the largest impact on quality.

Visualizations



Model Results

	Results
Regression using alcohol	Training $R^2 = 0.210$ Testing $R^2 = 0.227$
Regression using all attributes	Training $R^2 = 0.282$ Testing $R^2 = 0.320$
Binary Classification Model (Tree-Based)	Training: 1.0 Test: 0.745
Multi-Label Classification Model (Tree-Based)	Training: 1.0 Test: 0.464

Conclusion

Based on our multiple models, we can conclude that this dataset was better suited for a classification problem.

We also concluded that attributes such as location, temperature, fermentation processes, etc. may be more effective when it comes to predicting wine quality.