

High - quality wine Prediction and Analysis



By Ruthy Yao

Summary

- Predictive and Descriptive analysis on red wine quality classification
- Random Forest Model with accuracy score of 93.4% and F1 score 71.2%
- Top five contributing factors and its value range:
 - Alcohol - to control between 10.8 and 12.2;
 - Sulphates - to control between 0.65 and 0.86;
 - Volatile acidity - to control between 0.3 and 0.49;
 - Density - to control between 0.9947 and 0.9974;
 - Total sulfur dioxide - to control between 17 and 43.

Outline

- Business Problem
- The data
- Method
- Results and Insights
- Next Steps



Business Problem



- Enlist all the products to the Industry Awards and Rating competition.
- Sales and marketing cost is hiking as we spend the money to support all the products – inefficient use of fund, low ROI.
- The management team looking to adopt a more “focused approach” by allocating the funding to only the high-quality products that are more likely to win the industry award.
- We also want to find out the drivers of high- quality wines, so we can target the parameters for measure and management in the fermentation process to enhance the quality of our wine products.

The data

The dataset comes from the red variants of the Portuguese "Vinho Verde" wine¹

- 1599 data entries
- 11 physiochemical properties
- Wine quality rated from 1 to 8

Analytical Method



- Three **machine learning models** are trained
 - Logistic Regression
 - Decision Tree
 - Random Forest
- **Feature Importance** technique to rank the parameters based on the contribution to the wine quality classification from the highest to the lowest
- **Quartile statistics** to compare and contrast the value distribution for each of the top five parameters between “high-quality” and “mediocre” wines.

Results and Insights

- Random Forest Model has been selected because it has the highest predictive power

Accuracy score

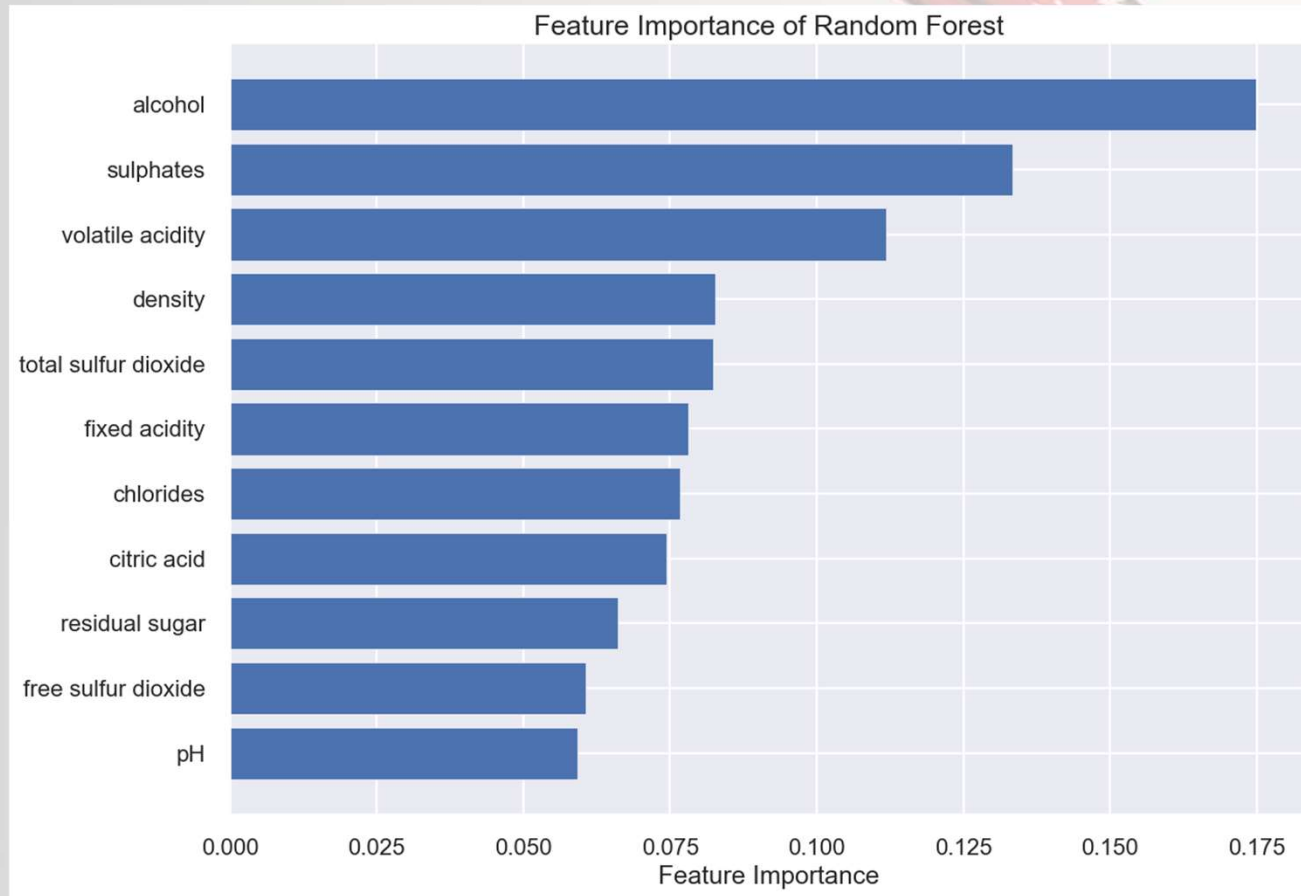
93.4%

F1 score¹

71.2%

Note: 1. F1 score is a balanced score that combines precision and recall

Results and Insights



- The parameters that contribute the most to a “high-quality” status are alcohol, sulphates, volatile acidity, density, total sulfur dioxide

Results and Insights

- The desirable value range for each parameter

Physiochemical Properties	Value Range
Alcohol	10.8 - 12.2
Sulphates	0.65 - 0.86
Volatile acidity	0.3 - 0.49
Density	0.9947 - 0.9974
Total sulfur dioxide	17 - 43

Next Steps

- **Enlarge the sample size** - to improve the predicative accuracy.
- **Analyse the interacted features** - Build interaction features to improve the predicative power.
- **Prepare for the model deployment** - Further analysis on what fermentation techniques and how to implement the techniques to control the physio-chemical properties in the desirable range will prepare us for the implementation of those analytical insights in the real world practice.

A close-up, artistic photograph of a glass of red wine. The glass is tilted, and a stream of red wine is being poured from it, creating a dynamic, flowing shape. The background is a soft, out-of-focus light color.

Thank you!

Email: zejia.yao@gmail.com

GitHub: <https://github.com/RuthyYao>