Understanding Sommelier Judging Criteria

Christian Wood

1 Introduction

Sommeliers are wine experts trained to evaluate subtle differences in wines, assessing quality through sight, smell, and taste. This project investigates whether sommeliers consistently identify these differences and explores the physiochemical features that influence their quality ratings. Understanding these criteria is valuable for winemakers aiming to refine production and for consumers seeking higher-quality wines.

2 The Dataset

The dataset [4] contains 4,898 ratings of "Vinho Verde" wine from northern Portugal, split into red and white varieties. Each wine is described by 11 physiochemical features, alongside a quality rating determined through sensory evaluation (sight, smell, taste). These ratings range from 0 (lowest quality) to 10 (highest quality).

These physicochemical features vary according to the winemaking process each wine has undergone resulting in each wine's different sensory experience, and thus the quality score provided by the sommeliers. These physiochemical features can be broken down into three broad aspects of a wine's profile; Body, flavour, and preservation, determining the mouthfeel, texture, and fullness of the wine, the flavour profile and balance of the wine, and the longevity, stability, and preservation of the wine respectively.

Given the extremely high quality of the dataset used, no data processing is required to clean the dataset ahead of use. However, before training our models, the data was split into three sets using an industry-standard 70:15:15 ratio for training, validation, and test sets. The validation set was reserved for tuning hyper-parameters and assessing model generalisability independently of the k-fold cross-validation process.

3 Regression Models

Although the scale is a set of discrete values which would suggest a classification algorithm would be appropriate, its linear property as a scale results in a regression model being the most appropriate due to its ability to factor in the linear relationship between numeric scores. We will therefore explore two different regression approaches outlined below.

The first approach will be using linear regression, which we expect will enable the prediction of a given wine's quality according to its physiochemical components. We will leverage sklearn's "sklearn.linear_model.LinearRegression" model, which we will train on the training set.

The second approach uses multiple decision trees[2] generated on different subsets of the training data known as random forests [3]. We expect the random forest approach to perform better on this task than linear regression as we believe the scoring of a wine's quality is influenced by many complex, non-linear relationships between features, which random forests can more effectively capture and therefore should result in more accurate quality predictions, although they may sacrifice explainability when compared to linear regression. For this model, we will employ sklearn's "sklearn.ensemble.RandomForestRegressor", trained on the aforementioned training set, and then tuned on the validation set, with the optimal parameters identified (max_depth=None, min_samples_split=2, min_samples_leaf=1).

4 Analysis of Results

Random forests significantly outperformed linear regression, with k-fold cross-validation revealing consistently better performance across folds. Validation metrics confirmed lower Mean Squared Error (MSE), Mean Absolute Error (MAE) and higher R^2 values (average $R^2 = 0.498$ for red and $R^2 = 0.545$ white wines) with standard deviations across these in these in an acceptable range ($<0.25\%$) for both red and white wines.

In *Figure 1* we see two subplots showing feature importance by feature groupings, and Shap value by feature[1], where Shap value is normalised between -1 and 1 for each feature to enable comparison between white and red wines and between features. We additionally highlight the distribution of wine feature values through the shape of each plot.

In our feature importance subplot *(Figure 1, top subplot)*, we observe the varying importance of the different features, with flavour for both white and red varieties clearly being the highest importance group of features. This is followed by preservation and then body. We can theorise preservation features may be important due to their role in ensuring the wines' flavour is conserved. Body by contrast affects the mouthfeel of a wine so is a key feature in defining its quality, but is a secondary aspect to a wine when compared to a wine's preserved flavour.

By analysing the results in our Shap values by feature *(Figure 1, bottom subplot)*, we can identify how different features impact a wine's quality and how this may vary between white and red varieties. For example, when looking at alcohol, we observe a clear preference for higher-alcohol wines, possibly due to alcohol's nature as a flavour enhancer. We can however also observe that the distribution density (not the wine's density feature) of wines in the dataset decreases as the alcohol level increases, likely due to the higher difficulty in producing higher alcohol levels without distillation. Further observations can be derived similarly from this second subplot, and can also include analysis of inter-feature relationships offering tailored strategies for optimising production.

5 Limitations

This study is limited by its focus on a single wine variety (Vinho Verde) and only incorporates 11 physiochemical features. Broader datasets with diverse wines could reveal varying quality criteria across styles and regions. Additionally, factors like subjective preferences or external conditions during tastings are not accounted for, potentially influencing ratings. Future work should explore other wine varieties, expand feature sets, and consider sommelier biases.

6 Conclusion

The study highlights sommeliers' consistency in quality assessment and their clear preferences for specific features, such as high alcohol. These findings offer actionable insights for winemakers aiming to enhance product quality. Future research could extend this approach to other wine varieties and incorporate more advanced predictive techniques, further bridging the gap between sensory evaluation and data-driven winemaking.

References

[1] An introduction to shap values and machine learning interpretability. https://tex.stackexchange.com/questions/3587/how-can-i-use-bibtex-to-cite-a-web-page. Accessed: 2024-12-02.

[2] What is a decision tree? https://www.ibm.com/topics/decision-trees. Accessed: 2024-12-02.

[3] What is random forest? https://www.ibm.com/topics/random-forest. Accessed: 2024-12-02.

[4] Cerdeira A. Almeida F. Matos T. Cortez, Paulo and J. Reis. Wine Quality. UCI Machine Learning Repository, 2009. DOI: https://doi.org/10.24432/C56S3T.

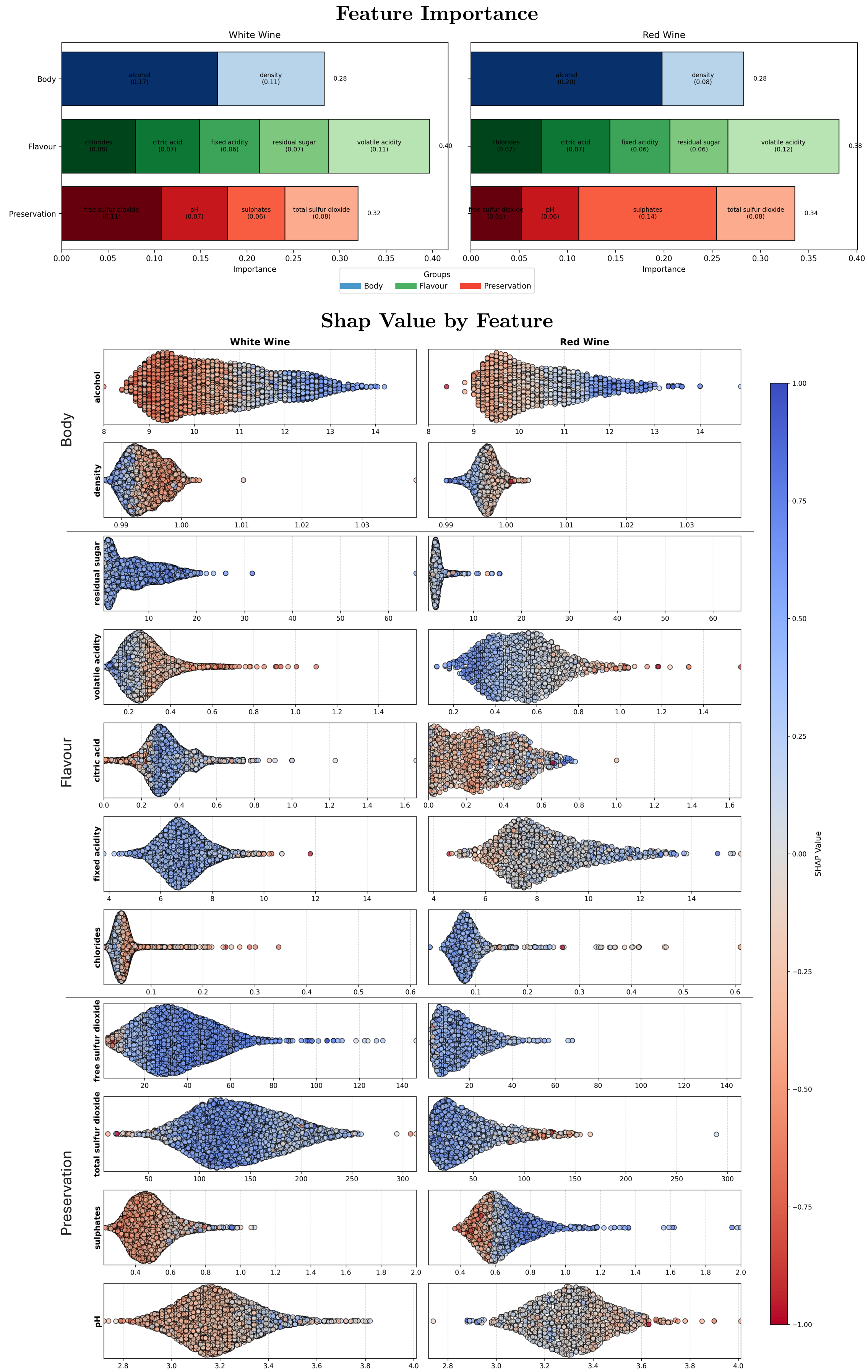

Figure 1: Feature Importance and Shap Values for Random Forest Models of White and Red varieties of Vinho Verde.

Appendix

Metric	Mean	Std Dev
Linear Regression - MSE	0.429	0.051
Linear Regression - MAE	0.507	0.037
Linear Regression - R^2	0.342	0.064
Random Forest - MSE	0.327	0.024
Random Forest - MAE	0.413	0.016
Random Forest - R^2	0.498	0.028

Figure 2: Red Wine Model K-fold Cross Validation Results

Metric	Mean	Std Dev
Linear Regression - MSE	0.568	0.026
Linear Regression - MAE	0.586	0.011
Linear Regression - R^2	0.275	0.018
Random Forest - MSE	0.356	0.026
Random Forest - MAE	0.425	0.013
Random Forest - R^2	0.545	0.023

Figure 3: White Wine Model K-fold Cross Validation Results

Generative AI Statement

AI-supported/AI-integrated use is permitted in this assessment. I acknowledge the following uses of GenAI tools in this assessment:

- (YES / NO) I have used GenAI tools for developing ideas.
- (YES / NO) I have used GenAI tools to assist with research or gathering information.
- (YES / NO) I have used GenAI tools to help me understand key theories and concepts.
- (YES / NO) I have used GenAI tools to identify trends and themes as part of my data
- analysis.
- (YES / NO) I have used GenAI tools to suggest a plan or structure for my assessment.
- (YES / NO) I have used GenAI tools to give me feedback on a draft.
- (YES / NO) I have used GenAI tool to generate image, figures or diagrams.
- (YES / NO) I have used GenAI tools to proofread and correct grammar or spelling errors.
- (YES / NO) I have used GenAI tools to generate citations or references.
- (YES / NO) Other: [please specify]
- I have not used any GenAI tools in preparing this assessment.

I declare that I have referenced use of GenAI outputs within my assessment in line with the University referencing guidelines.