HORIBA

Fluorescence

Predicting Wine Mouthfeel using Machine Learning and Simple Chemical Variables

Application Note
Life Sciences
FL-2025-05-01

Abstract

This application note details the use of bagging (random forest, RF) and boosting (extreme gradient boosting, XGBoost) machine learning (ML) algorithms to predict red wine mouthfeel attributes from simple chemical measurements. A panel of 15 wine experts assessed 30 commercial red wines (Australian and Spanish) using rate-all-that-apply sensory methodology and chemical data, including linear sweep voltammetry, excitation-emission matrix, and absorbance, were collected. Principal component analysis simplified the sensory data, revealing four independent mouthfeel dimensions: 'drying', 'full body', 'velvety', and 'gummy'. RF and XGBoost models outperformed classical partial least squares regression, achieving over 80% accuracy on test data. This study demonstrates the potential of ML coupled with simple chemical measurements for rapid and cost-effective prediction of wine sensory properties.

Introduction

Machine learning (ML) is revolutionizing agricultural and food science, offering powerful tools for modeling complex datasets. Compared to traditional chemometric methods like partial least squares (PLS) linear regression, ML algorithms offer advantages when handling limited sensory data and high-dimensional chemical data, with the development of non-linear models. This note explores the application of random forest (RF) and extreme gradient boosting (XGBoost) modeling to predict scores for wine mouthfeel attributes, as crucial wine quality attributes, from voltammetric and absorbance-transmittance and fluorescence excitation-emission matrix (A-TEEM) analyses. These chemical approaches could offer rapid and cost-effective alternatives to complex sensory techniques.

Materials and Methods

Wine Samples:

Thirty commercial red wines (18 Tempranillo Tinto, 12 Garnacha Tinta) from Australia and Spain were selected, varying in origin, region, and vintage.

Chemical Analysis

- Basic wine chemical parameters (pH, titratable acidity, alcohol) were measured.
- **Voltammetry:** Linear sweep voltammetry (0-1200 mV) was performed using Edel Therapeutics equipment, recording 120 voltammetric signals in triplicate.
- **Absorbance and Fluorescence:** Absorbance-transmittance and fluorescence excitation-emission matrix (A-TEEM™) data were acquired using an Aqualog® UV-800 spectrometer (precursor to the new Veloci™ Wine Analyzer). Absorbance was recorded at λ = 240–700 nm and EEMs at λ_{Ex} = 250–700 nm and λ_{Em} = 250–800.
- **Data Analysis:** For voltammetric data, the first derivative was calculated. For A-TEEM spectra, they were normalized and corrected for the influence of inner filter effects (IFE), solvent background, dark detector signals, first (±16 nm filter) and second (±32 nm filter) order Rayleigh masking. Three-dimensional EEM data were transformed by reshaping to a 2D matrix.

Sensory Analysis

- **Panelists:** 15 Australian wine experts (winemakers and researchers).
- **Methodology:** Rate-all-that-apply (RATA) on a 7-point scale for 31 mouthfeel and taste attributes. Global sensory attributes (astringency, bitterness, flavor intensity) were rated on a 10-cm scale.
- **Data Analysis:** Two-way ANOVA, PCA with varimax rotation, and Pearson correlation coefficients were used to analyze sensory data.

Prediction Modeling

Prediction by regressing of the sensory variables from: voltammetric, absorbance and 2D EEM signals by PLS and ML algorithms.

- **PLS Regression:** Single Y-variable PLS models were developed for each sensory attribute using voltammetric, absorbance, and EEM data.
- **Machine Learning (RF and XGBoost):**
  - Data were split into 80% training and 20% test sets
  - Hyperparameters were optimized using cross-validation
  - Variable importance was assessed

Data and Results

Secondary Data Analysis

- PCA revealed four independent sensory dimensions:
  - Dimension 1: 'Drying'
  - Dimension 2: 'Full body'
  - Dimension 3: 'Velvety'
  - Dimension 4: 'Gummy'
- Significant differences in sensory attributes were observed among wines.

Prediction Modeling

- RF and XGBoost models significantly outperformed PLS regression in terms of number of modeled sensory attributes and model performance.
- **Accuracy:** Both ML algorithms achieved average accuracy exceeding 0.80 on test data across voltammetric, absorbance, and EEM datasets compared to ≤ 0.42 for PLS.
- **XGBoost vs RF:** XGBoost showed slightly better performance (higher accuracy and lower error), but RF offered advantages in computational efficiency.
- **Variable Importance:** The models identified key chemical variables contributing to mouthfeel perception.
- **Overfitting Control:** Careful hyperparameter optimization and validation were crucial to prevent overfitting.

Sensory Dimension XGBoost R2 RF R2		
Drying	0.85	0.82
Full Body	0.88	0.84
Velvety	0.83	0.80
Gummy	0.86	0.81

Table 1: Example Results: R^2 of XGBoost and RF models predicting sensory dimensions from absorbance data

(Note: Data values are examples and may vary slightly from the original study)

The combination of Abs+EEM datasets (i.e., full A-TEEM data) yielded slightly better explained variances than EEM or Abs alone, but the errors were comparable.

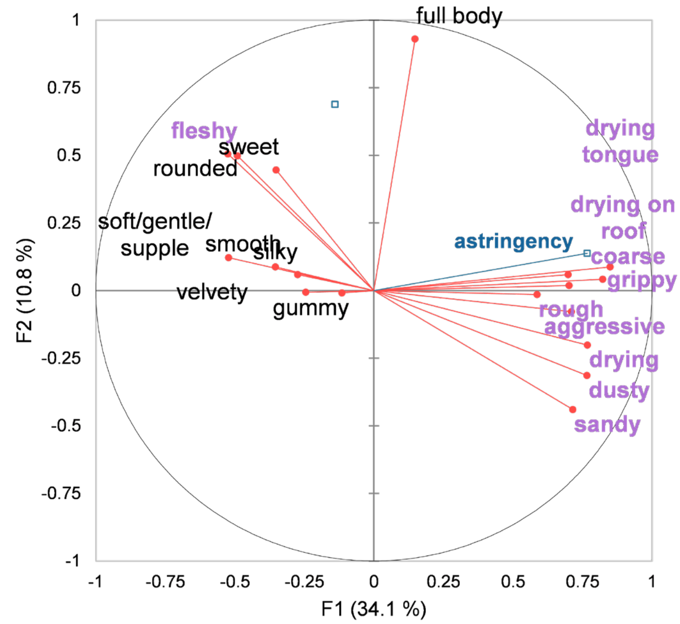

Figure 1: PCA Biplot of Sensory Attributes and Wines. Active variables in purple contribute significantly to F1. Variables in blue are supplementary variables.

(Note: A representative PCA biplot showing the distribution of wines and sensory attributes along the principal components should be included here)

Conclusion

This study demonstrated the efficacy of RF and XGBoost algorithms for predicting wine mouthfeel traits from simple chemical measurements. Both ML methods outperformed PLS regression, achieving high predictive accuracy and low error. The use of voltammetric, absorbance, or EEM data provides the foundation for a rapid and cost-effective alternative approach to wine sensory analysis. These findings have significant implications for the wine industry in relation to efficient quality control and product development.

Acknowledgements

The authors acknowledge the original study authors: María-Pilar Sáenz-Navajas, Chelo Ferreira, Susan E.P. Bastian, David W. Jeffery.

Veloci™ Wine Analyzer is a new A-TEEM spectrometer that replaces Aqualog for wine analysis.

info.sci@horiba.com **www.horiba.com/scientific**

USA: +1 732 494 8660 **France:** +33 (0)1 69 74 72 00 **Germany:** +49 (0) 6172 1396 0
UK: +44 (0)1604 542 500 **Italy:** +39 06 51 59 22 1 **Japan:** +81(75)313-8121
China: +86 (0)21 6289 6060 **India:** +91 80 41273637 **Singapore:** +65 (0)6 745 8300
Taiwan: +886 3 5600606 **Brazil:** +55 (0)11 2923 5400 **Other:** +33 (0)1 69 74 72 00