

Optimizing Red and White Wine Quality Prediction Using Hybrid Ensemble Learning and Interpretable Feature Engineering

Dissanayake D.K.H.^{1*}, Kumara B.T.G.S.²

¹Department of Computing and Information Systems, Faculty of Computing,
Sabaragamuwa University of Sri Lanka, Sri Lanka

²Department of Software Engineering, Faculty of Computing,
Sabaragamuwa University of Sri Lanka, Sri Lanka

*dkhdissanayake@std.appsc.sab.ac.lk

The conventional method of wine quality prediction has always been based on sensory testing by human analyst and this has been subjective as well as time consuming, and this type of evaluation is hard to scale to the modern wine production system. The current development of machine learning allowed predicting wine quality based on physicochemical data; however, the current models are often characterized by the class imbalance, low interpretability, and poor consistency in working with red and white wine samples. This paper hypothesizes a hybrid ensemble learning model, which combines classification and regression models with interpretable feature engineering to enhance predictive performance and interpretability. The dataset of red and white wines in the region of Vinho Verde was utilized, with comprehensive preparations and feature scaling, feature selection, and resampling, including SMOTE and SMOTE-ENN to overcome an uneven distribution of classes. Several models of base line, such as Support Vector Machines, Naïve Bayes, Ridge Regression, Artificial Neural Networks, Random Forest, and Gradient Boosting, were compared and tested against ensemble and stacked hybrid models. The performance of hybrid ensemble models is shown to be more accurate, has a higher recall and a larger F1-score, especially when compared to single baseline models. SHAP and LIME analysis of interpretability found alcohol content to be a positively strong predictor of wine quality, whereas volatile acidity was significantly negative. The explanations present practical recommendations that can be used by the winemakers in an attempt to make better production choices. Altogether, the suggested framework provides the best predictive results with reasonable interpretability, justifying the application of machine learning-based wine quality assessment in the wine industry practically.

Keywords: *Feature Engineering, Hybrid Ensemble Learning, Machine Learning, Model Interpretability, Wine Quality Prediction.*