WINE TASTING PROJECT

Using only partial least squares regression for generalized linear models (the plsRglm package in R), build an ensemble model to predict the quality score given to each wine from the Vinho Verde region of Portugal. The data consists of chemical tests on wines and an overall assessment score averaged from at least three professional wine tasters.

The prediction model should be trained and will be tested on both red and white wine after joining the two datasets. We will split the supplied data into an 80% training set and a 20% hold-out validation set before running your training script with a random seed. We will use mean absolute error as a performance metric.

You will submit:

Your script for training the model or ensemble in a .R file

A short and informal paper explaining your exploratory analysis, findings, and reasoning for: data splitting, feature engineering, pre-processing, model building, hyperparameter optimization, model stacking, and withholding set validation