

# Capstone 3: Project Proposal

**Problem Statement:** Can we predict wine quality scores from the physicochemical properties of wines?

## **Project Description:**

Vinho verde is a medium-alcohol wine from the Minho (northwest) region of Portugal particularly appreciated due to its freshness.

For this project I am predicting the quality of Vinho Verde wines to be of either low or high quality given the physicochemical components of wines that are introduced into the product during the wine-making process.

My hypothesis is that the physicochemical properties of a wine (e.g. pH and alcohol content) which can be controlled at the winemaking stage, can be used to predict the overall quality of the wine. If this hypothesis is correct, then we can figure out which of the wine's properties are most important in optimizing the output of high quality wines by a winemaker.



## **Proposal:**

Frank Ceballos has a well-written and comprehensive article on stacking classifiers and their uses in boosting performance. They are best used when an increase in performance can't be achieved through any further data acquisition, feature engineering, or preprocessing and all typical classification model approaches have been exhausted. I propose to use this blending approach to predict the quality of red and white wines to get a performance boost since the data are not too large.

I then will look at feature importance by doing Permutation Feature Importance on the best stacking classification models. This would allow me to determine which of the physicochemical properties in the wines dataset can contribute to high quality wines, which can further be used by wine makers in developing high quality wines, assessing why some of their wines perform the way they do, or in predicting how well a given wine should perform.