Wine Quality Prediction

- In good wine we trust

Yucheng Zhu
June 10, 2019

Motivation + Demo

- Conventional methods of evaluating wine quality are primarily based on the subjective experience of tasters.

- This project not only explores the relationship between physiochemical properties of wines and their taste, but also intends to utilize machine learning methods to figure out which set of physiochemical properties make a wine taste "good."

Data

- Wine Quality Dataset from UCI Machine Learning Repository

- Red variant of the Portuguese "Vinho Verde" wine.

- 1599 Instances

- Variables Highlight:

  - Fixed acidity, volatile acidity, citric acid, residual sugar, free sulfur dioxide, total sulfur dioxide, density, alcohol Output variable

  - Quality (score between 0 and 10)

Model

- Random Forest (n_estimators = 200)

- Binary classification: buy or not buy (cutoff: quality score = 6.5)

- A train/test split of 80/20 was applied.

- Success:
  - ML: Prediction accuracy should be at least 80%
  - Business: User retention rate should be at least 60%

- Result:
  - ML: 0.93 Testing Accuracy

Insight

- Chemical-wise, different types of wine generally have very similar levels of "Density" and "pH".
- Compared with low quality wine (quality <=4), high quality wine (quality >=6) generally:
  - Higher in "Fixed Acidity," "Citric Acid," "Free Sulfur Dioxide"
  - Lower in "Volatile Acidity" and "Chlorides"
- Measurable physicochemical properties of wines can help people differentiate good and bad wines.

Thank you!

- Yucheng Zhu
- Yucheng.Zhu@u.northwestern.edu
- github.com/yzelac/MSiA423-Project/