Wine Quality Dataset

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis.
Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.

Problem Identification

How can we use physicochemical properties of wine to predict human taste preference (quality) of the wine?

Context

Wine currently requires a human to taste and rate its quality. We want to see if using a machine learning approach to predict wine quality can be beneficial to wine companies to help increase consumer satisfaction.

Criteria for Success

Building a successful model that can capture the relationship between physicochemical properties of the wine and wine quality.

Scope of Solution Space

Clean dataset of missing or erroneous values. Apply machine learning algorithms to test for best predictive accuracy. Use feature selection to check if all features are relevant to the output.

Constraints

Possible constraints to solving the business problem of ‘using physicochemical properties of wine to predict human taste preference’ include limited wine data; The wines from this dataset are from the northern region of portugal and so our models would not work with other wine varieties. The output, wine quality on a scale from 0-10, is determined from the median value of the subjective ratings of 3 wine experts. This wine quality rating may not be representative of the tastes of people from other markets.

Stakeholders

Viticulture Commission of the Vinho Verde Region

Data Sources

Paulo Cortez, University of Minho, Guimarães, Portugal, http://www3.dsi.uminho.pt/pcortez
A. Cerdeira, F. Almeida, T. Matos and J. Reis, Viticulture Commission of the Vinho Verde Region(CVRVV), Porto, Portugal
@2009