

Capstone Two: Project Proposal

Problem statement (Hypothesis)

How could a vinho verde wine company produce more than 30% of quality above 7 wines through quality-physicochemical features modeling and implement corresponding quality control in the coming three years?

Context

Vinho verde is a unique wine product from the Minho (northwest) region of Portugal. The quality of wine rate from 0 (very bad) to 10 (very good). The problem of vinho verde wine is the classes are not balanced, much more normal wines than excellent or poor ones. One solution for any local Vinho verde wine in Minho region is to implement the data driven strategy to improve 30% of wines to quality above 7 in the coming three years.

Criteria for Success

A vinho verde wine company(Hypothesis) produces more than 30% of quality above 7 wines in the coming three years through quality-physicochemical features modeling and implementing corresponding quality control.

Scope of the Solution

The data scientist team will build models and predict the wine quality. The corresponding company will implement quality control to produce more than 30% of quality above 7 wines in the coming three years.

Constrains

Other factors that have an impact on wine quality such as grape type, wine storage temperature and time were not provided. Therefore, the quality prediction based on physicochemical features only has unknown limitations on accuracy.

Stakeholders

The owner of the dataset, Any Vinho verde wine companies that like to implement this data driven strategy to their business. The data scientist team.

Data sources

<https://archive.ics.uci.edu/dataset/186/wine+quality>

The dataset was provided by Cortez et al in 2009, it contains the information of white wine and red wine physicochemical tests.

Approach

1. Data Wrangling-Collecting, cleaning and transforming the data.
2. Exploratory Data Analysis and Feature Engineering - data visualization to investigate trends and tell a story.
3. Feature Engineering, Modeling and Machine Learning
4. Data Story

Deliverable

A GitHub repo containing the work of each step including: A slide deck and A project report