

**Presentation of the**

# Wine Quality Analytics System

**Determining the quality of white vinho verde wines based on their  
physicochemical composition**

A cooperation of



# Agenda

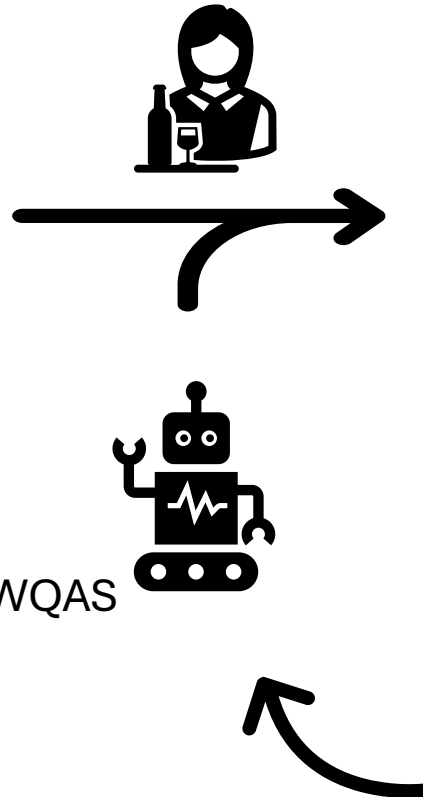
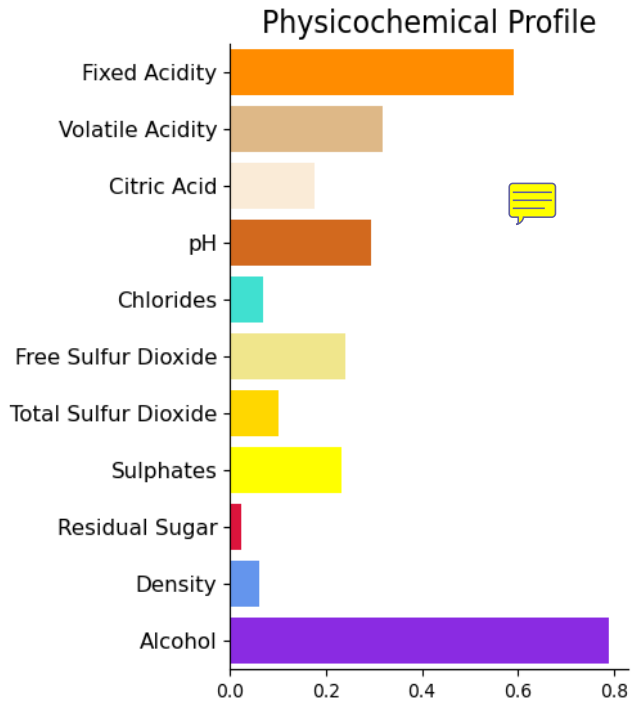
- The Wine Quality System
- The Challenges
- Data Insights: Alcohol and Density
- The Machine Learning Model
- Results
- Conclusions

# The Wine Quality System

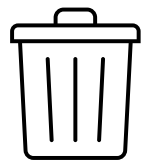
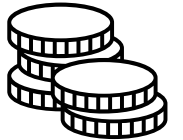
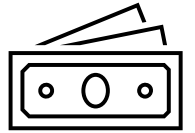
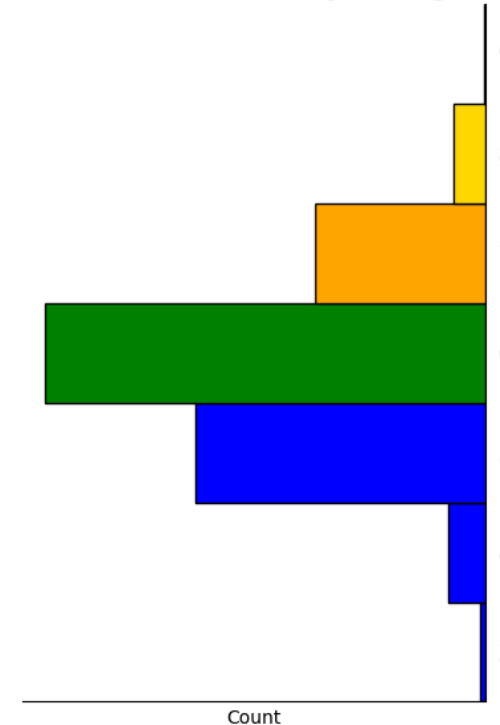


Marketing

Tasters

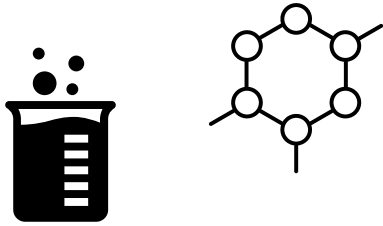


White Wine Quality Rating



- Wine Quality Analytics System supports the tasters with finding high quality wines
- The system can be continuously improved.

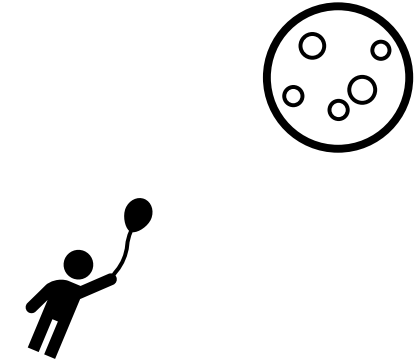
# Challenges



complex mixture of  
>1000 substances  
resulting from a  
complicated process



- The taste is the least understood human sense and highly individual
- Each chemical composition has it's own appeal

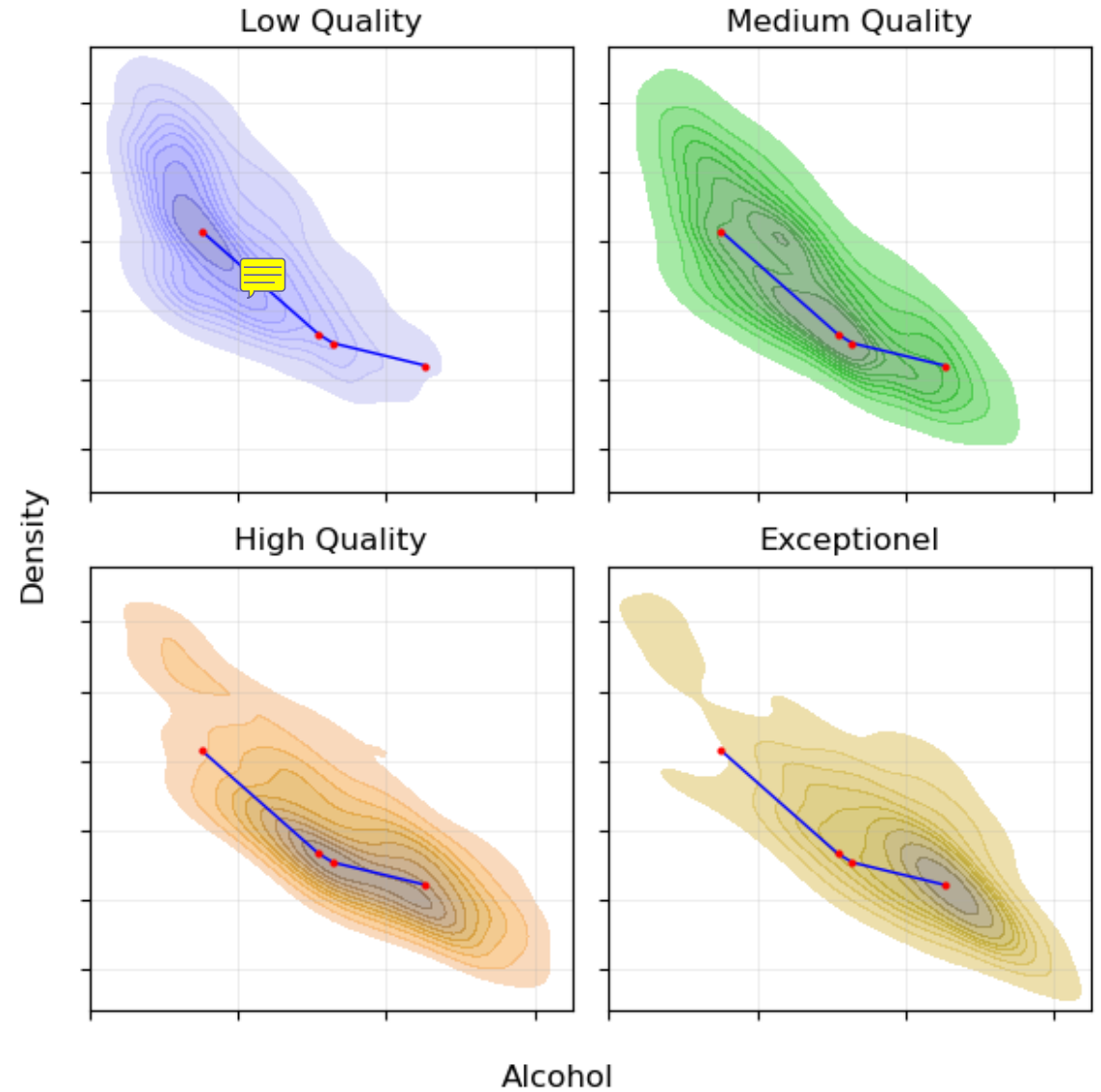
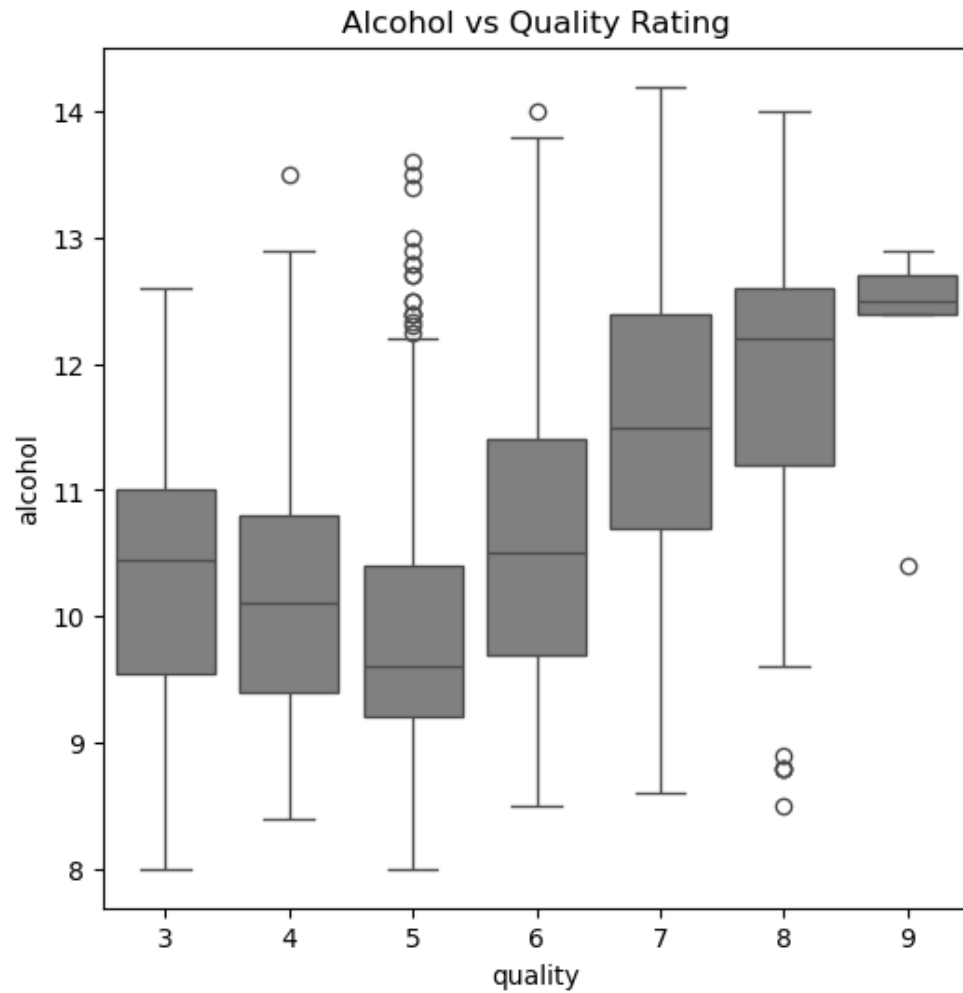


- Low data coverage
- 11 components
- 3500 samples
- 1 quality rating
- Only a few high quality wines

Noisy, high-dimensional, non-linear, interdependent  
relation between Quality and Features

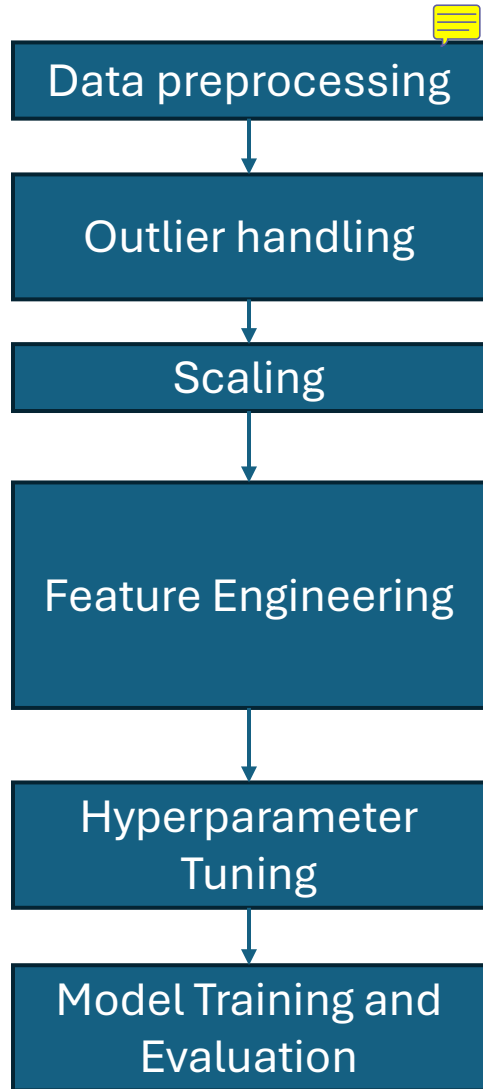
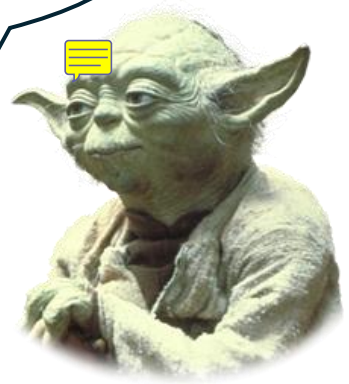
Weak representation

# Examples: Alcohol, the main Quality Indicator



# Machine Learning Model

Random Forest Classifier with high recall we need, hmmm!

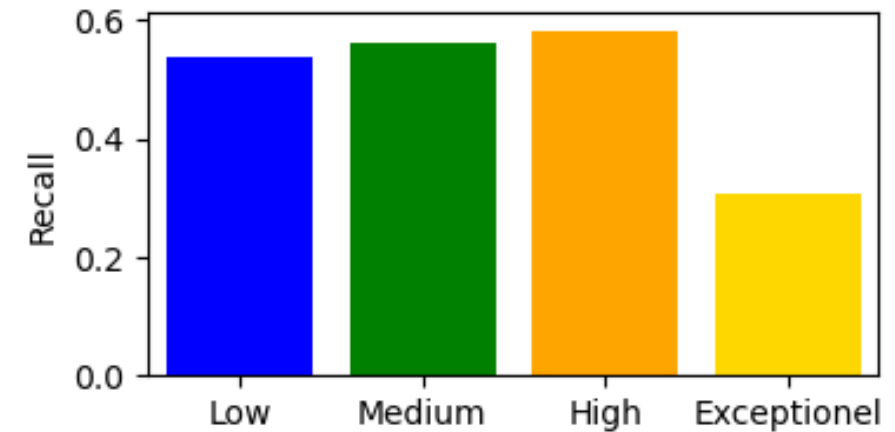
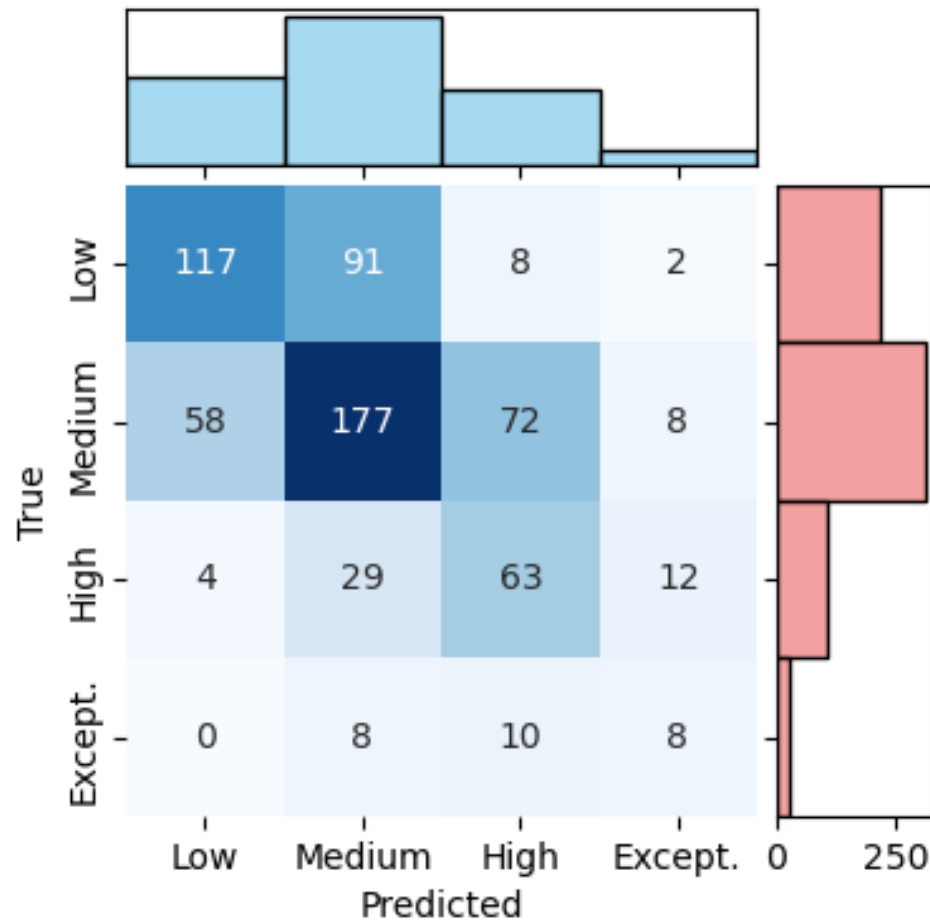


- Remove Duplicates
- Use Univariate and multivariate methods
- **Drop low quality outliers**
- **Introduce Quality Label**
- **Incorporate Features Interactions**
  - 15 most important interactions
  - Avoid interactions of noisy features
- Optimize for high recall
- **Fix class Imbalance: Over- and Undersampling**
- Pay attention Overfitting



# Results

Confusion Matrix and Label Distributions



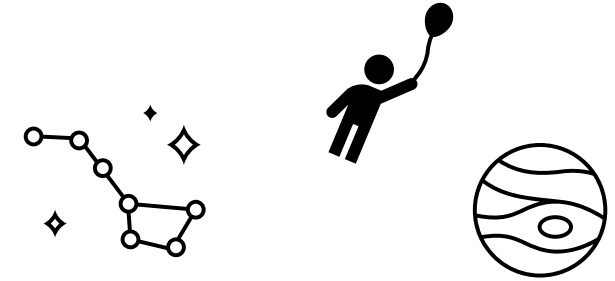
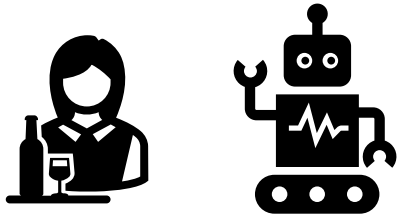
## Pros

- Model predicts both high quality categories pretty well
- Reasonable parameters and param. behaviour
- Feature Importances

## Cons

- Accuracy 0.54
  - Low
  - Big gap to training accuracy

# Conclusions



## WQS in practice

- The model can support the Tasters in selecting a good wine with zero effort.
- It has to be used with caution, wine quality still has to be finally determined by an expert.
- Ensure a good fermentation process.

## More Data!

- Ensure high data quality
- Continue collecting data
- Measure more quantities
- Standardize quality measurements

## Model Refinement

- Multilabel classification
- Advanced algorithms can be applied



Thank you for your attention!

Are there any Questions?