Projekt Portfolio

- **Multiclass Classification of wine quality**
- **Code and Models**
- **Wine type: red / white**
- **Evalution and Results**
- **Repo in GitHub**

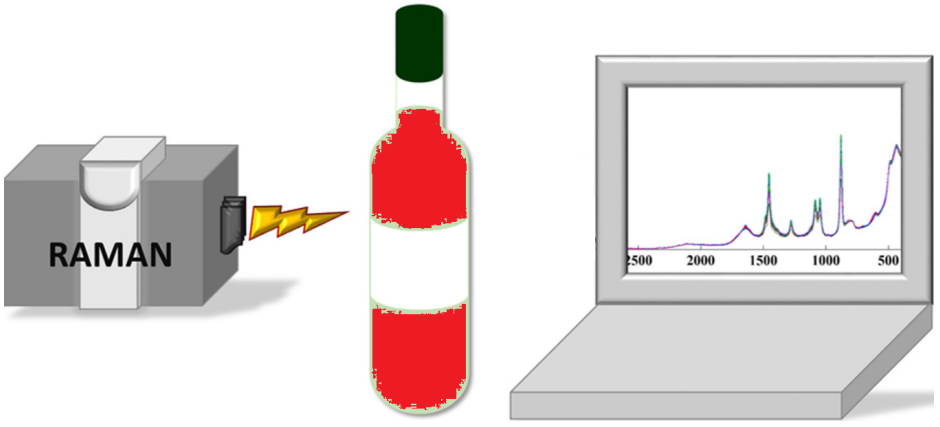
Mamoun Naim

Multiclass Classification of wine quality

Excellent production depends on quality control: function, dimensions, composition, ... of the products.

Quality control

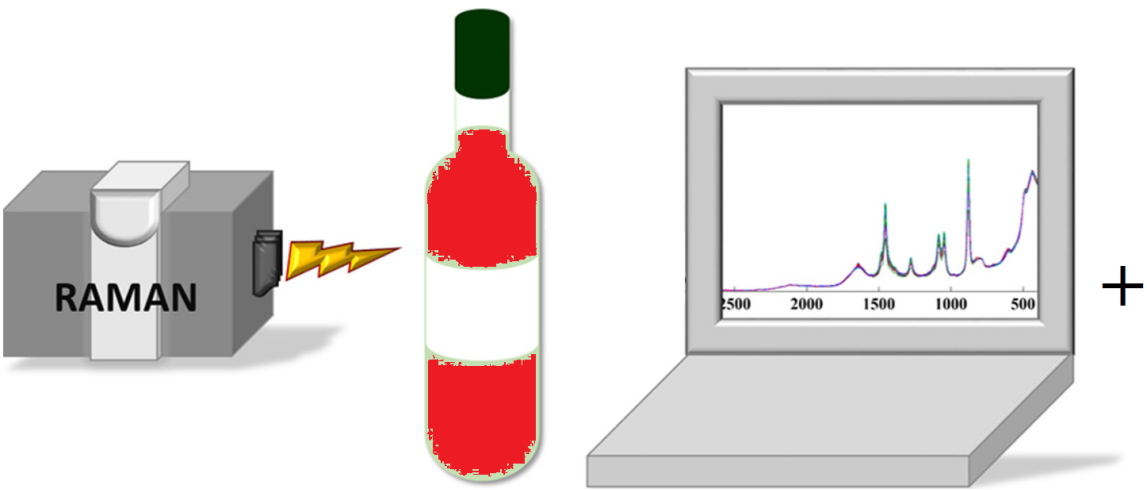
None-destrctive measurements: e.g. Raman-Spectrpometer



For example: analyze the contents of wine **without** opening the bottle!

From Measurements to quality

Measurements (features) → math. Formula /machine lernen → quality



+ **Classification** ⇨ **quality**

wine quality

Wine type: red / white

Contents measurement  11 features / inputs

Quality {0 .. 10}  Target / output of classification

Quality-value is a **subjective** measurement by wine tasters

Models

Binary classification: Wine ist good or bad (0 or 1)

multiclass classification: Quality in the range [0 .. 10]

Tested Models: 2 classifications and 3 Regressions

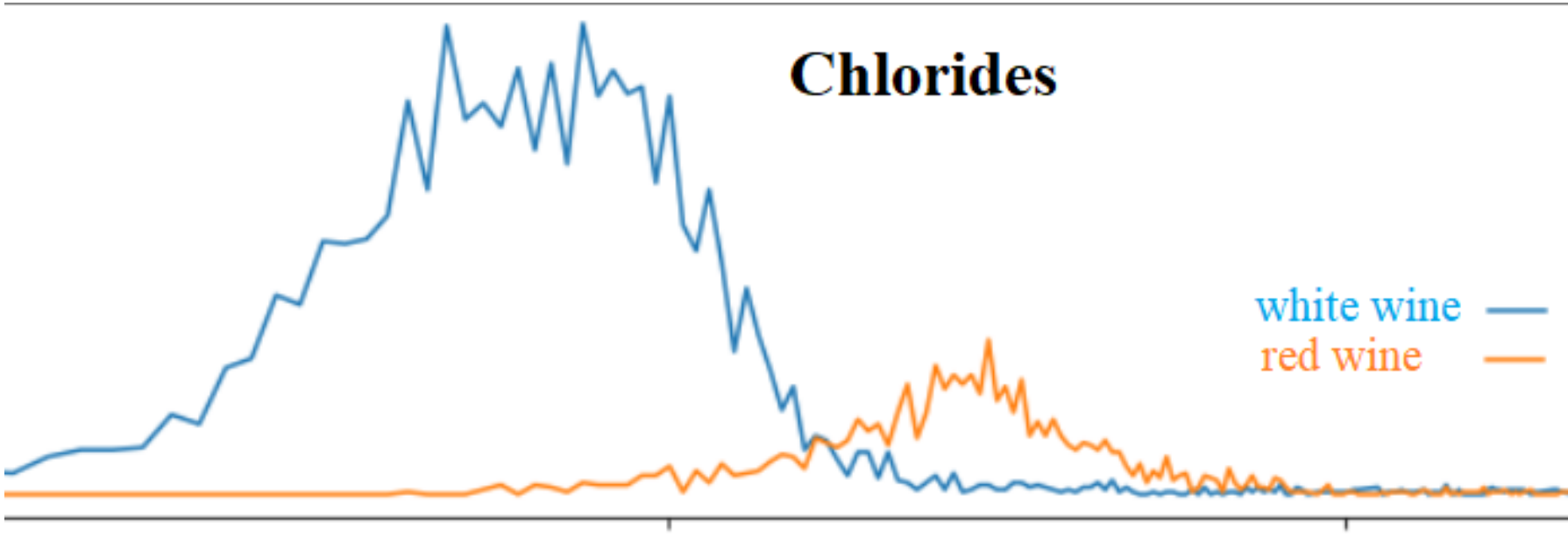
The best model "RandomForestClassifier" fine-tunded and saved as a binary pickle file

Wine type: red / white

red or white wine data. importance of wine type for classification?

EDA

Type as extra feature



Classification analysis

Surprisingly, the type of wine hardly plays a role in the classification!

Results: Classification accuracy

Das best Model: RandomForestclassifeir

Multiclass : 64 %
Binary classification: more than 92%

In all Data: Wine taster evaluate quality only in [3 .. 9]

7 classes = min 14% (random classification)

Binary = min 50% (random classification)

Subjective measurement - Classification error

There is no fixed rule for evalution
A taster could simple give 6 instead 5 as quality

Wine quality multiclassifier with tolerance of 1
(prediction – real target) <=1

Classification accuracy = 95%

Tolerance =2  Classification accuracy >99%

For example, **never** 4 instead of 7.

final results

Integration of machine learning algorithms in measuring devices.
Alle Infos über Produkt as input. The ouput is go/ no go

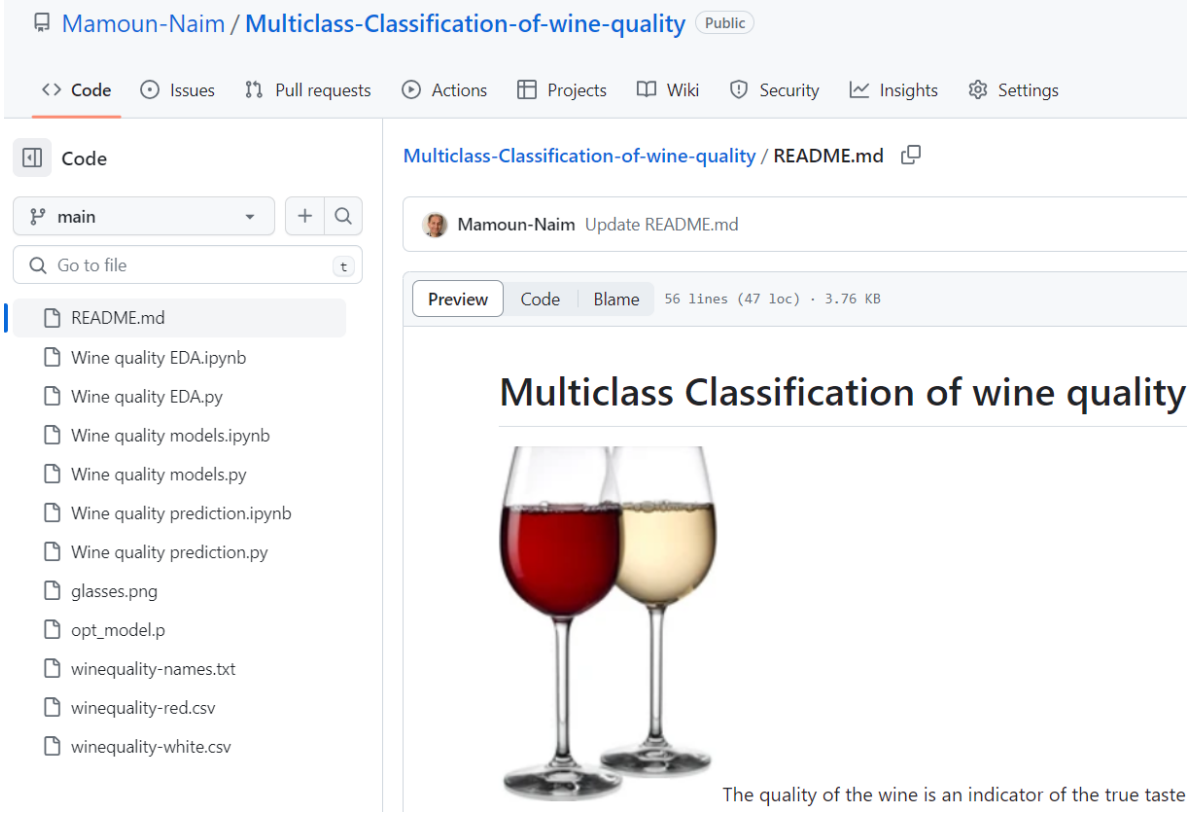
The type-feature of wine does not play a role in the classification!

Muticlass classification accuracy 64 % (or 95% with 1-tolerace)
Binary class classification accuracy 92 %

Repo in GitHub

README.md

winequality-red.csv
winequality-white.csv



code as Jupyter Notebooks and script.
1- "Wine quality EDA.py"
2- "Wine quality models.py"
3- "Wine quality prediction.py"

Best model
opt_model.p