Let Machine Learning choose your wine!

Rasmus Lund, Lukas Kezic, Frederik Thomsen & Jakob Aarestrup - Aarhus University

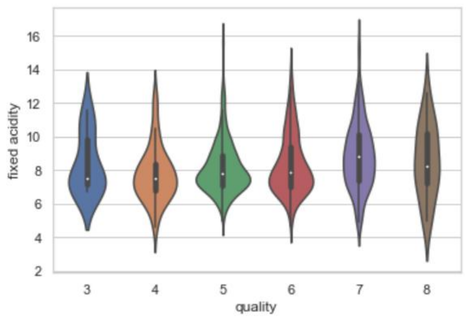
Introduction

Have you ever wondered if it is possible to determine a wines quality by looking at its chemical properties? With the help of Machine Learning algorithms this have now become possible.

Take a look!

Chemical properties

Fixed acidity vs. Quality:



Results

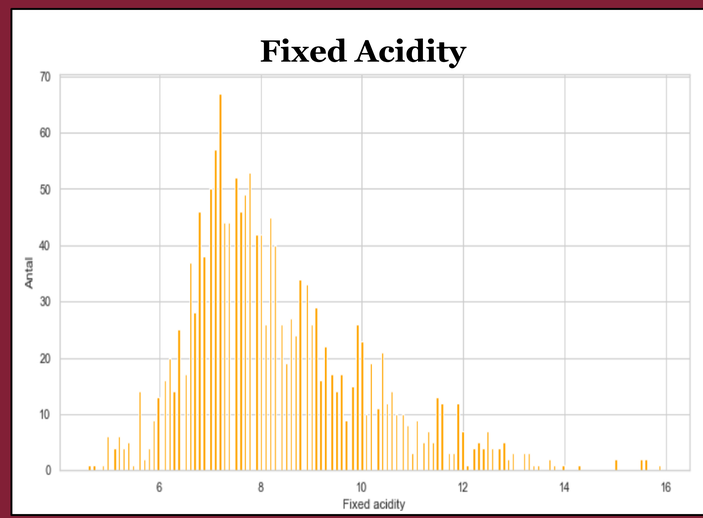
There is a very small number of outliers in this set.

The Linear regression algorithm produced the best results with a precision of 99%.

Based on the violin-plots we couldn't determine any specific parameters being the sole reason for a good or bad wine.

The majority of wine scored an average of 6 in quality.

	fixed acidity	volatile acidity	pH	sulphates	alcohol	quality
count	1599.00	1599.00	1599.00	1599.00	1599.00	1599.00
mean	8.32	0.53	3.31	0.66	10.42	5.64
std	1.74	0.18	0.15	0.17	1.07	0.81
min	4.60	0.12	2.74	0.33	8.40	3.00
25%	7.10	0.39	3.21	0.55	9.50	5.00
50%	7.90	0.52	3.31	0.62	10.20	6.00
75%	9.20	0.64	3.40	0.73	11.10	6.00
max	15.90	1.58	4.01	2.00	14.90	8.00



Conclusions

The Machine Learning algorithm can successfully determine what is good wine and what is bad wine.

With the small number of outliers, it was a no brainer to use PCA-test to determine which parameters were the most significant.

The parameters “fixed acidity” and "volatile acidity", has the biggest influence in the quality of wine.

Linear regression was the most efficient algorithm in wine detection.

Dataset

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis.
Modeling wine preferences by data mining from physicochemical properties.

Choice of algorithm

- Linear regression ☺
- Decision trees 😐
- SGD classifier ☹

Further information

- Wine type in dataset: “Vinho Verde”.
- Dataset is ordered not balanced.
- Number of chemical properties: 11 + output.