Abstract

White Wines are very specific goods. The quality depends from many factors. Barrels, kind of grapes, location etc. but also from ingredients and some of them do not depends from kind of grapes but from recipe and the activities carried over production process. The level of preservatives, level of alcohol and ph ratio and many other depends is the manufacturer of wines will get desired certification and will not loose on quality of wine.

The Quality management is very important in many industries. Data mining is a nowadays a necessary tool in this field.

We used the classification technique – clustering to find the similarities between attributes and decision tree, generalised linear regression, Random forest, Gradient Boosted Tree to see the dependency of attributes from the target attributes and predict the target variables from our dataset of white wines variables.

Keywords: wine quality, data mining, classification, clustering, k-means, x-means, decision tree, gradient boosted tree, random forest, generalised linear regression

Introduction

Problem statement:

The level of preservatives in wines starts to be a big problem but also necessary ingredient.

How much level of preservatives influences the other attributes of wine? Is that wine still a balanced and healthy for us? How about level of alcohol in wine?

How about answering that question just because of study of already made wines? Very expensive process of developing a new quality wines can be set aside. There is literature about manufacturing wines and this expertise is necessary but we would like to show just small scope for improvement.

And hopefully this project will help little bit more understand the dependencies between attributes of wines.

Conclusions

New technology can be used in majority of industries. Perfect example here is manufactory of white wines.

We can see from the work above that the attributes do not work in isolation. All of them have impact on each other.

Although it is still hard to predict the taste of wine and quality. We have to remember that attribute quality it is a subjective attribute when compare to others – scientific.

The Clustering analysis showed us that we can deal with 4 groups of clusters, showed us dependencies and correlations Total sulfur dioxide and free sulfur dioxide, Fixed acid and PH, Residual sugar and density.

Target attributes analysis of targets such us level of alcohol, sugar/Ph ratio, level of preservatives and Key Performance Indicator – Balanced wine prove that other ingredients of wine influence the important attribute and they cannot be checked in isolation.

By eliminating some attributes we just created our model weaker. They all happen to be necessary.