



Advanced Business Analytics with R

Mihir Raikar msr180000

Project Summary

The Dataset is about white wine variant of the Portuguese Vinho Verde wine.

Goal: Predict what makes the best wine in terms of quality Data at a glance:

Variables: "fixed_acidity", "volatile_acidity", "citric_acid", "residual_sugar", "chlorides", "free_sulfur_dioxide", "total_sulfur_dioxide", "density", "pH", "sulphates", "alcohol", "quality"

One more variable is created, called Quality Level. It is the bucket in which the quality falls:

Quality is 4 or less, then Bad

Quality is 5 or 6, then Medium

Quality is 6 or more, then Good

I have tried Running various Models for classification. The Accuracy is as below:

Model	Accuracy on Quality	Accuracy on Quality Level Quality level
SVM	74.29	100%
Naïve Bayes	70.2%	97.81
Ridge	71.99%	99.69%
Lasso	73.42%	100%
Elastic Net	-	-

We Can say that the accuracy at Quality is not that great but at quality level is Very high.

This tells us that our model can be used to classify the wines of the Vinho Verde Winery.

This can be generalized to other wines as well.

Various variables and their weights can be used to determine the class of a wine. Considering that the maximum observations were from quality level class: Medium, It can be implied that the good and wines are few. Maximum wines are medium. Vinho Verde or any other wine maker needs to focus on this segment of wines for maximum sales.

If there was a way to obtain sales data, there can be an analysis to link the quality of wine to sales generated and profit earned.

