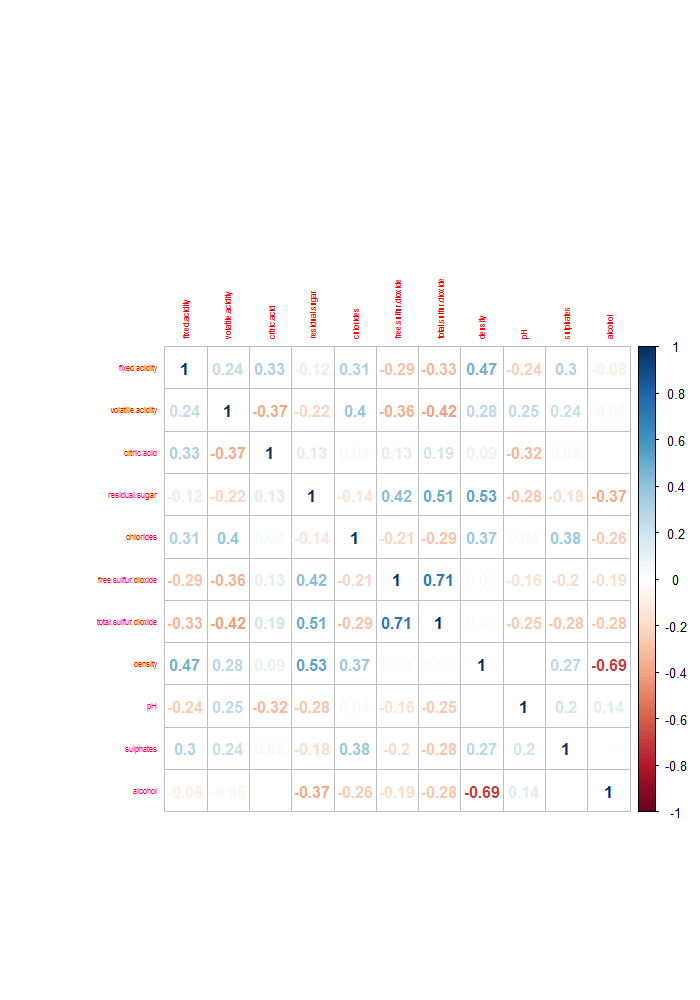
**AIM: -** Build classification model which will classify the quality of wine depend on multiple factors.

**INSIGHTS: -**

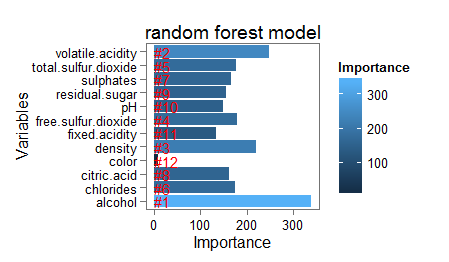
* **Following is the correlation plot.**



* I selected acid, sugar and alcohol levels as the main features. However, from the above observations, it appears that sugar (residual sugar) does not have a clear relationship with quality.
* Free sulfur dioxide and total dioxide appear to have strong correlation (0.71). This is expected as these two factors are related. The implication is we may need to choose only one of them for model building.
* The relationship between density and quality is strong (-0.69).

**Random Forest Model:-**

* According to my analysis the best accuracy is given by random forest model.
* The accuracy of random forest model is 94.74%.



* As from above figure we can extract that alcohol is the factor that explains best the quality of wines and so on.
* As in this random forest with accuracy 94.74% I have classified the quality of wines as GOOD and BAD.
* But when I classify the quality of wines as GOOD NORMAL and BAD the quality of random forest model decreases to 89.1%
* After some research about wine quality, it appears to me that acid, sugar and alcohol levels are the most important features when tasting and deciding wine quality.
* From the samples analyzed, the wines with higher alcohol content exhibited lower SO2 content as compared with samples with lower alcohol content.
* Variable importance from this model is as follows:

|  |  |  |
| --- | --- | --- |
| **Rank** | **variable** | **MeanDecreaseGini** |
| #1 | alcohol | 326.45484 |
| #2 | volatile.acidity | 238.46427 |
| #3 | density | 226.53733 |
| #4 | free.sulfur.dioxide | 178.79974 |
| #5 | total.sulfur.dioxide | 176.83455 |
| #6 | chlorides | 176.57641 |
| #7 | sulphates | 167.03343 |
| #8 | citric.acid | 166.74997 |
| #9 | residual.sugar | 163.74586 |
| #10 | pH | 148.41871 |
| #11 | fixed.acidity | 141.35692 |
| #12 | color | 8.52188 |

**Decision Tree Model:-**

**Using CART:**

* The accuracy using this model is 93.27%.
* Variable importance from this model is as follows:

|  |  |  |
| --- | --- | --- |
| **Rank** | **variable** | **Overall** |
| #1 | volatile.acidity | 778.8844 |
| #2 | alcohol | 744.6663 |
| #3 | density | 729.8779 |
| #4 | chlorides | 710.8186 |
| #5 | citric.acid | 671.498 |
| #6 | residual.sugar | 607.2978 |
| #7 | free.sulfur.dioxide | 594.8586 |
| #8 | fixed.acidity | 580.7995 |
| #9 | total.sulfur.dioxide | 572.7123 |
| #10 | sulphates | 500.2955 |
| #11 | pH | 499.2313 |
| #12 | color | 26.84979 |

**Using C50:**

* The accuracy using this model is 83.54%.
* Following are the rules given by this model which are important :

For judging Quality of wine as **Bad** I took three rules that are

* Rule 1: (41, lift 2.7)

fixed.acidity > 8.9

fixed.acidity <= 10

volatile.acidity > 0.26

citric.acid > 0.32

residual.sugar > 1.8

Alcohol <= 10.3

Class BAD [0.977]

* Rule 2: (36, lift 2.7)

fixed.acidity <= 8.9

volatile.acidity > 0.665

citric.acid <= 0.03

Alcohol <= 10.3

Class BAD [0.974]

* Rule 11: (23/1, lift 2.5)

volatile.acidity > 0.495

sulphates <= 0.37

Class BAD [0.920]

For judging Quality of wine as **GOOD** I took three rules that are

* Rule 36: (30, lift 1.5)

fixed.acidity <= 7.1

residual.sugar <= 1.65

free.sulfur.dioxide > 16

total.sulfur.dioxide > 86

total.sulfur.dioxide <= 127

Alcohol > 10.3

Class GOOD [0.969]

* Rule 37: (124/6, lift 1.5)

volatile.acidity <= 0.205

residual.sugar > 12

Chlorides > 0.033

Class GOOD [0.944]

* Rule 38: (652/38, lift 1.5)

sulphates > 0.37

Alcohol > 11.7

Class GOOD [0.940]

Insights from these rules are:

* **BAD WINE** we can classify when the content of alcohol <=10.3 ,

Volatile acidity >0.26 and the residual sugar >1.8 as if the content of residual sugar increases the taste of wine will be BAD.

* **GOOD WINE** we can classify when the content of alcohol >10.3 ,

Volatile acidity < 0.21 and the residual sugar <=1.65 as if the content of residual sugar should be less than 1.65 than taste of wine will be GOOD.