**Wine Modeling**

Given the task to improve the wine production which can help for better market demand. Also asked to identify the relationship between different features given in the dataset.

**The following list my analysis**

**Dataset**: - total dataset given is 12 measurements on 6,497 samples of white and red styles of *Vinho Verde* wines. There are NO missing data.

**Details Regarding relationship among features.**

As we can see in the attached file **Wine\_EDA\_Preprocessing**, features such as alcohol, sulphates, density, total\_sulfur\_dioxide, free\_sulfur\_dioxide, chlorides as similar behaviour while rest of the features such as fixed\_acidity, volatile\_acidity, citric\_acid, residual\_sugar, pH doesn't have unique behaviour between red wine and white wine

***fixed\_acidity***- Not unique behaviour, fixed\_acidity is increasing against quality for red wine but vice-versa for white wine

***volatile\_acidity***- volatile acidity is low for good quality for both red and white wine, but it is low for all categories in white wine and high for bad and ok categoreis in red wine.

***citric\_acid***- Citric acid also has distinct behaviour compared between red and white wine. low for white wine for good quality but high for red wine

***residual\_sugar***- residual sugar also has distinct behaviour compared between red and white wine. low for white wine for good quality but high for red wine

***pH*** - Not unique behaviour for white wine having more pH score is good for better product but for red wine need lesser score for good quality

***free\_sulfur\_dioxide***- Kind of unique behaviour but white wine for good quality uses more amount of free\_sugar\_dioxide but not red wine. The lesser the good quality

***total\_sulfur\_dioxide***- similar to free\_sulfur dioxide , but not entirly unique behaviour for both red and white wine. The lesser the value the better the quality

***chlorides*** - feature decreased or increased with quality for both red wine and white wine. Both having unique beaviour and for good quality wine we need less chloride

***density*** - very straight feature the lesser the better for quality for both red and white wine. Unique characteristics

***sulphates*** - very straight feature the more the better for quality for both red and white wine. Unique characteristics

***alcohol*** - very straight feature the more the better for quality for both red and white wine. Unique characteristics

**Some other preprocessing: -**

Converted the given 10 category quality column into 3 category quality column by having **good** has better quality, **ok** as average quality and **bad** as low quality of wine. Ratings greater than or equal to 7 falls into good category, ratings 6 into ok category (as more data is in this category, so to keep balanced distribution I kept only 1 value in this category) and 5 and less than 5 into low category. Please go through **Wine\_EDA\_Preprocessing** file in the attached.

Converted the style into 0 and 1, where 0 means red wine and 1 means white wine

Normalized the features except style and quality, so that scale range of all feature falls into same range.

Label encoding of style features also improves the accuracy but I haven’t used this encoding because my analysis of the dataset is to understand the relationship than improving the accuracy .

**Model Building:**

I used different models like XGBOOST, Random Forest, Light GBM to train and test the data preprocessed data. The enrite code of model building is in the attached file **Wine\_Final\_Model\_Building**.

**Accuracy details after train and test split 90-10 ratio**

Random Forest - train accuracy - 84.8% - test accuracy - 66%   
XGBoost - train accuracy - 93.6% - test accuracy - 72%   
Light GBM - train accuracy - 83.8% - test accuracy - 67%

Conclusion: -

The better model overall is XGBoost and still the all the model can be improved with good amount of data. Relationship among red and white wine and features importance regarding the style is given above. By considering all these analyses will improve the quality of wine and improved market for the wine.