

Wine Quality

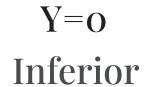


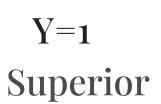
Predictive Model



Confusion Matrix

 $\hat{Y}=o$ Negative $\hat{Y}=1$ Positive



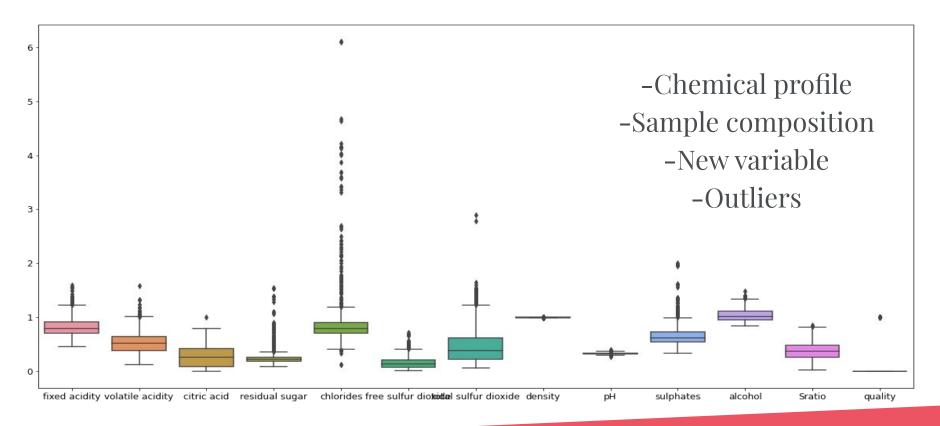




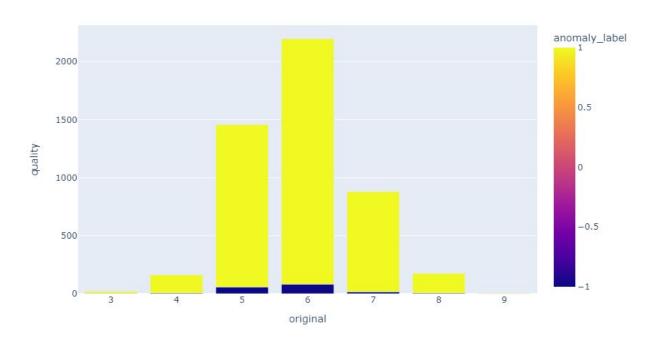




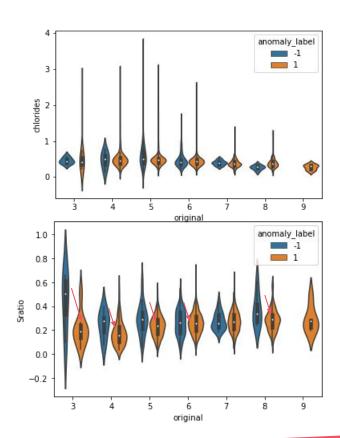
EDA

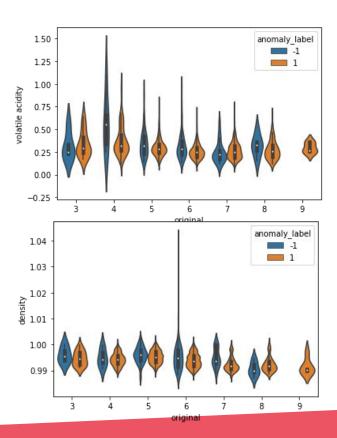


Outliers



Outliers





Correlation Heatmaps

	Correlation Heatmap												
fixed acidity -		-0.023	0.29		0.023	-0.049		0.27	-0.43	-0.017	-0.12	-0.14	-0.056
volatile acidity -	-0.023	1	-0.15	0.064	0.071	-0.097		0.027	-0.032	-0.036	0.068	-0.2	-0.052
citric acid -	0.29	-0.15	1	0.094					-0.16	0.062	-0.076	0.016	-0.022
residual sugar -		0.064		1	0.089	0.3	0.4	0.84	-0.19	-0.027	-0.45	0.051	-0.1
chlorides -	0.023	0.071			1	0.1		0.26	-0.09	0.017	-0.36	-0.033	-0.17
free sulfur dioxide -	-0.049	-0.097		0.3		1	0.62	0.29	-0.00062	0.059	-0.25	0.74	-0.021
total sulfur dioxide -				0.4		0.62		0.53	0.0023		-0.45	-0.013	-0.14
density -	0.27	0.027		0.84	0.26	0.29	0.53		-0.094	0.074	-0.78	-0.066	-0.25
pH -	-0.43	-0.032	-0.16	-0.19	-0.09	-0.00062	0.0023	-0.094	1			0.0008	0.09
sulphates -	-0.017	-0.036	0.062	-0.027	0.017	0.059		0.074	0.16		-0.017	-0.022	0.036
alcohol -	-0.12	0.068	-0.076	-0.45	-0.36	-0.25	-0.45	-0.78		-0.017		0.064	0.34
Sratio -	-0.14	-0.2	0.016	0.051	-0.033	0.74	-0.013	-0.066	0.0008	-0.022	0.064	1	0.11
quality -	-0.056	-0.052	-0.022	-0.1	-0.17	-0.021	-0.14	-0.25		0.036	0.34	0.11	1
	fixed acidity -	volatile acidity –	citric acid -	residual sugar –	chlorides -	ee sulfur dioxide –	tal sulfur dioxide –	density -	¥	sulphates -	alcohol -	Sratio -	quality –

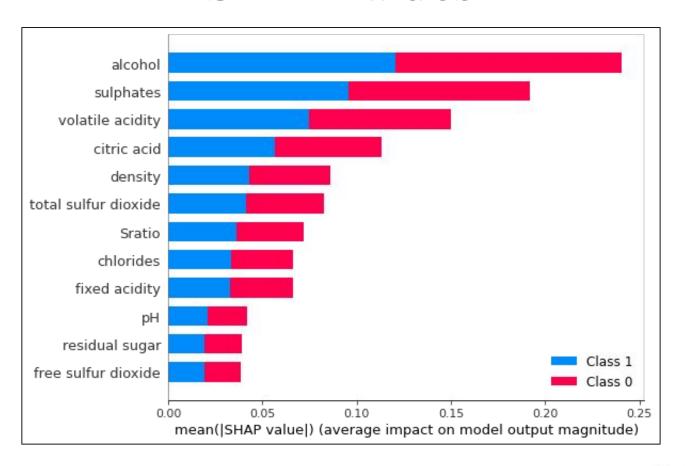
Correlation Heatmaps

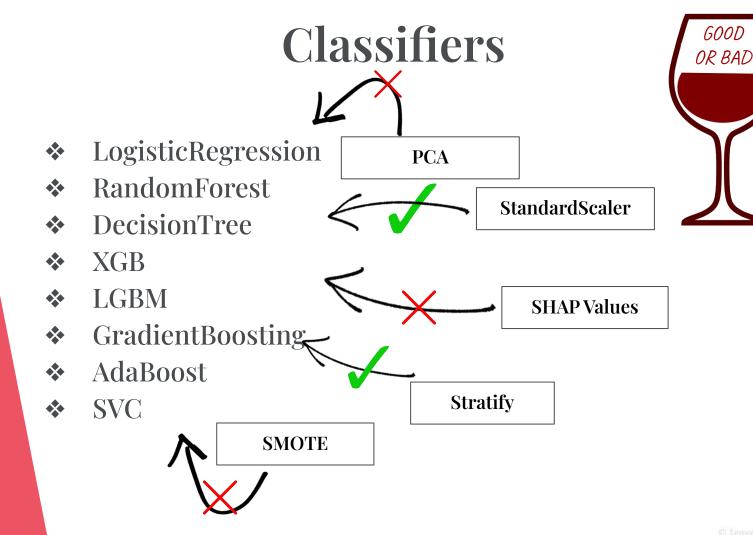
						Corre	elation Hea	tmap					
fixed acidity -		-0.26	0.67			-0.15	-0.11	0.67	-0.68	0.18	-0.062	-0.13	0.12
volatile acidity -	-0.26	1	-0.55							-0.26	-0.2	-0.073	-0.27
citric acid -	0.67	-0.55	1	0.14		-0.061		0.36	-0.54	0.31		-0.17	0.21
residual sugar -								0.36	-0.086			-0.071	0.048
chlorides -					1	0.0056	0.047		-0.27	0.37	-0.22	-0.11	-0.097
free sulfur dioxide -	-0.15		-0.061				0.67	-0.022			-0.069	0.33	-0.072
total sulfur dioxide -	-0.11					0.67	1	0.071	-0.066		-0.21	-0.37	-0.14
density -	0.67	0.022	0.36	0.36				1	-0.34		-0.5	-0.26	-0.15
pH -	-0.68	0.23	-0.54	-0.086	-0.27		-0.066	-0.34		-0.2			-0.057
sulphates -		-0.26	0.31		0.37				-0.2	1	0.094		0.2
alcohol -	-0.062	-0.2			-0.22	-0.069	-0.21	-0.5			1	0.25	0.41
Sratio -	-0.13	-0.073	-0.17	-0.071	-0.11	0.33	-0.37	-0.26				1	0.13
quality -		-0.27	0.21		-0.097	-0.072	-0.14	-0.15	-0.057		0.41	0.13	1
	fixed acidity -	volatile acidity –	citric acid –	residual sugar –	chlorides -	free sulfur dioxide –	otal sulfur dioxide –	density -	Ŧ	sulphates -	alcohol -	Sratio -	quality -

Scallers

Classifier	Scaller	Stratify	Score	F1 1	Precision 1	Recal 1	test average precision	test balanced accuracy	test roc auc
LogisticRegression		√	0,8	0,31	0,59	0,21	0,54	0,59	0,79
LogisticRegression	StandardScaler	\checkmark	0,8	0,38	0,59	0,28	0,54	0,61	0,8
LogisticRegression	MinMaxScaler	\checkmark	0,8	0,33	0,58	0,23	0,54	0,6	0,79
LogisticRegression	RobustScaler	\checkmark	0,8	0,36	0,58	0,26	0,54	0,61	0,8
LogisticRegression	Normalizer	√	0,79	0,15	0,58	0,08	0,52	0,54	0,78
LogisticRegression	PowerTransformer	\checkmark	0,8	0,37	0,58	0,27	0,54	0,61	0,8
LogisticRegression	QuantileTransformer	√	0,8	0,36	0,57	0,26	0,54	0,61	0,8

SHAP Values





White Wines

Score

0,88

0,78

0,77

0,77

0.75

0,62

0,61

0,61

0.69

0,49

0,48

0,48

Test

F1 1

Test

Precision 1

Test

Recal 1

0.81

0,83

0,84

0,84

test balanced

accuracy

0.6

0,74

0,74

0,74

test roc auc

0,81

0,81

0,81

SMOTE

- (898->6140)

Classifier

SVC

SVC

SVC

Scaller

Standard

Standard

Standard

Standard

PCA

Stratify

 $\sqrt{}$

 $\sqrt{}$

√ √

 $\sqrt{}$

Local

Outlier

RandomForest	Standard		\checkmark			0,9	0,73	0,86	0,64	0,64	0,83
RandomForest											
	Standard	$\sqrt{}$	$\sqrt{}$			0,89	0,72	0,84	0,63	0,63	0,82
RandomForest											
	Standard	$\sqrt{}$	\checkmark	\checkmark		0,88	0,66	0,82	0,56	0,62	0,81
RandomForest					√0,35						
	Standard		\checkmark		- (4898->4144)	0,9	0,74	0,82	0,67	0,64	0,83
RandomForest					√0,5						
	Standard		\checkmark		- (4898->4698)	0,89	0,73	0,77	0,7	0,65	0,83
RandomForest											
	Standard		\checkmark	\checkmark	$\sqrt{}$	0,87	0,67	0,74	0,61	0,64	0,84
SVC	Standard		\checkmark	\checkmark		0,78	0,62	0,5	0,81	0,75	0,81
RandomForest					√0,5						0.84

Red Wines

	Methodology	Score	F1 0	F1 1	Precision 0	Precision 1	Recal 0	Reca	best params	precision 1
0	GrindLogisticRegression	0.89	0.94	0.43	0.90	0.72	0.98	0.3	{'classifierC': 4.0, 'classifiersolver':	0.72
1	GrindRandomForestClassifier	0.94	0.97	0.73	0.94	0.93	0.99	0.6	('classifiercriterion': 'gini', 'classifier	0.93
2	GrindSVC	0.82	0.89	0.53	0.96	0.41	0.83	0.7	{"classifierC": 2.0, 7 'classifiergamma': 1	0.41
0	GrindLGBMClassifier	0.93	0.96	0.73	0.95	0.82	0.98	0.65	('classifier_boosting_type': 'dart', 'classif	
1	GrindXGBClassifier	0.95	0.97	0.80	0.96	0.82	0.97	0.77	{'classifier_booster': 'gbtree'}	
2	GrindGradientBoostingClassifier	0.93	0.96	0.72	0.94	0.84	0.98	0.63	('classifierlearning_rate': 1, 'classifier	
3	GrindAdaBoostClassifier	0.90	0.94	0.54	0.92	0.70	0.97	0.44	{'classifier_algorithm': 'SAMME.R', 'classifi	

Sources

UCI Machine Learning Repository: Wine Quality Data Set

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547–553, 2009.

https://www.extension.iastate.edu/wine/total-sulfur-dioxide-why-it-matters-too/

Total Sulfur Dioxide - Why it Matters, Too!



Thank you for your attention!





