https://github.com/garsonbyte/Machine-Learning-Project

Predicting Wine Quality and Type

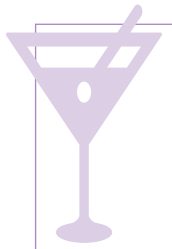
FINDING THE DIFFERENCES BETWEEN RED AND WHITE WINE

GARSON CHOW

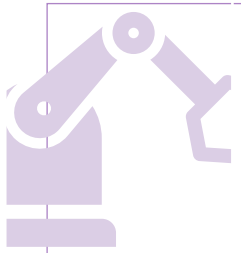
MEI SHIN LEE

Goals:

DETERMINE WINE QUALITY OF WHITE WINES BASED OFF ELEVEN PREDICTIVE FEATURES



PREDICT WHETHER A WINE SAMPLE IS RED OR WHITE



Find the best algorithm

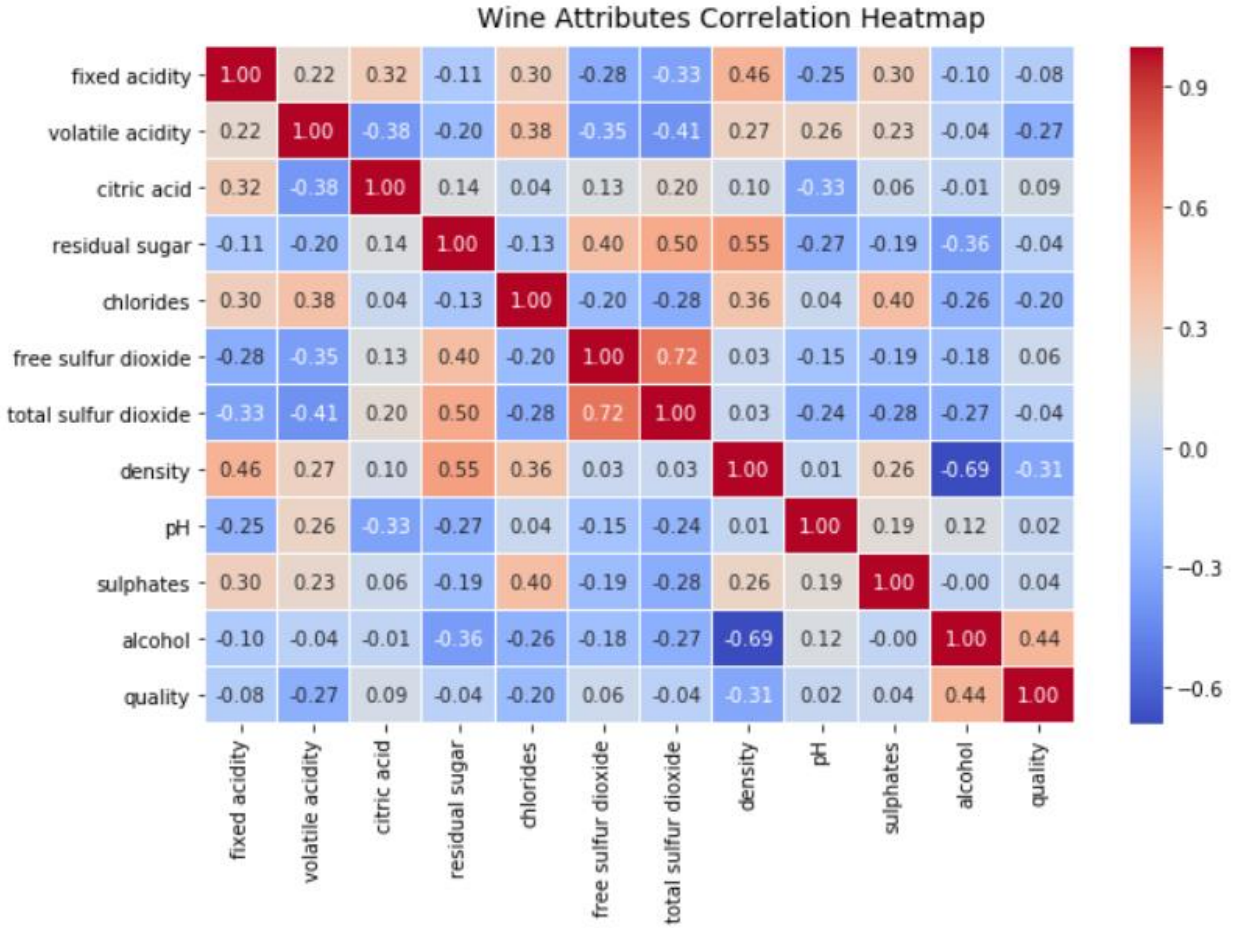
Methodology

1. *Unsupervised Learning:*
   - Principal Component Analysis
2. *Supervised Learning (Wine Quality):*
   - Linear Regression
3. Supervised Learning (Wine Type):
   - Logistic Regression
   - Support Vector Machine

Overview of Data

Wine Type	# of Samples
Red Wine	1599
White Wine	4898
Total	6497

*NOTE THAT WE HAVE ROUGHLY 3 TIMES MORE WHITE WINE SAMPLES THAN RED WINE SAMPLES.



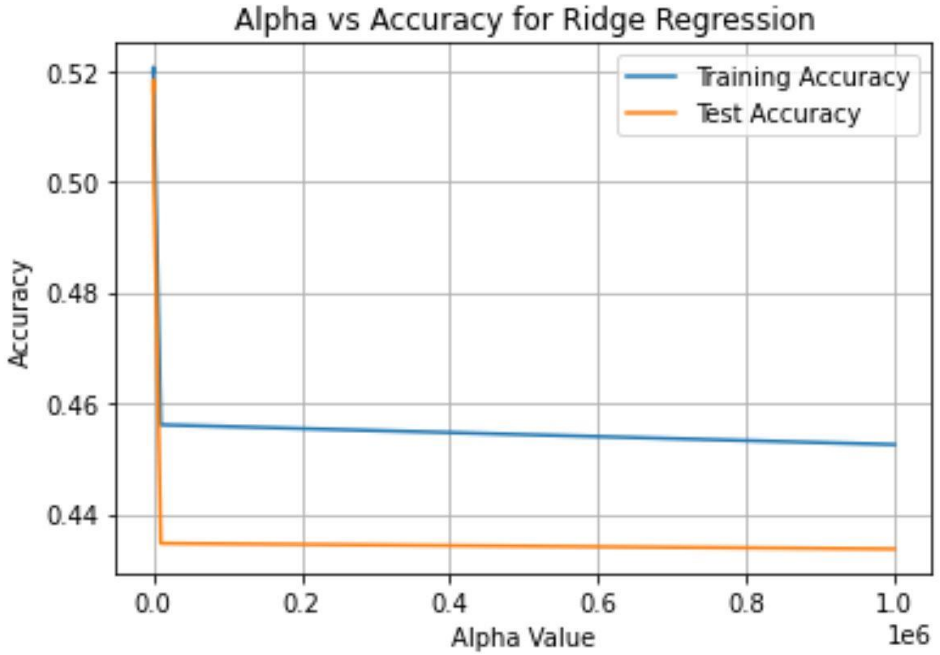
Predicting Wine Quality

- PREDICTED WHITE WINE QUALITY USING 11 DIFFERENT FEATURES.
- TRAIN/TEST SPLIT RATIO OF 80%:20%

	# of Samples
Train	3918
Test	980
Total	4898

RESULTS:

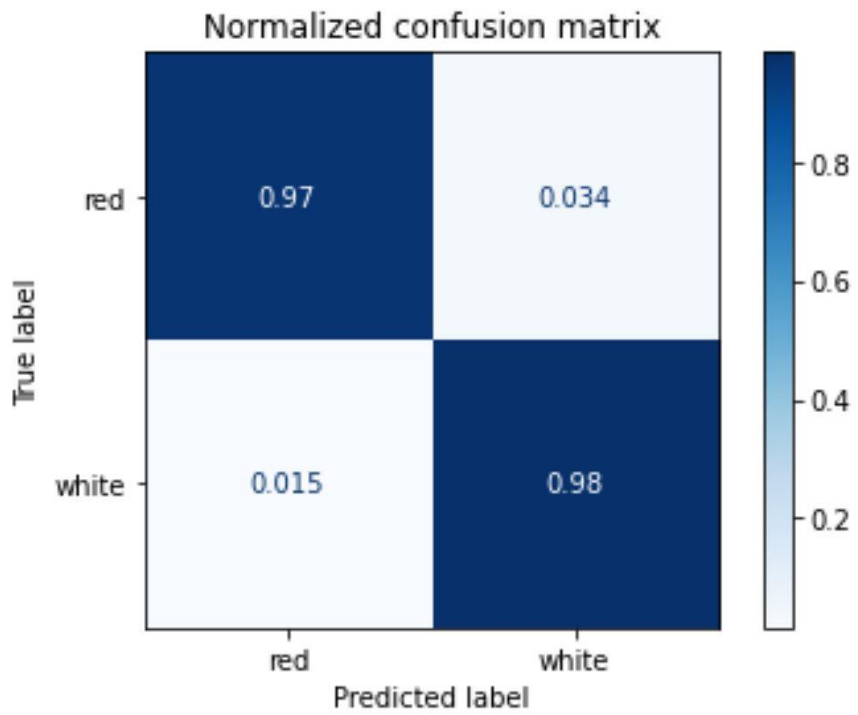
RSS	TSS	R^2
558.38351	757.31734	0.26268



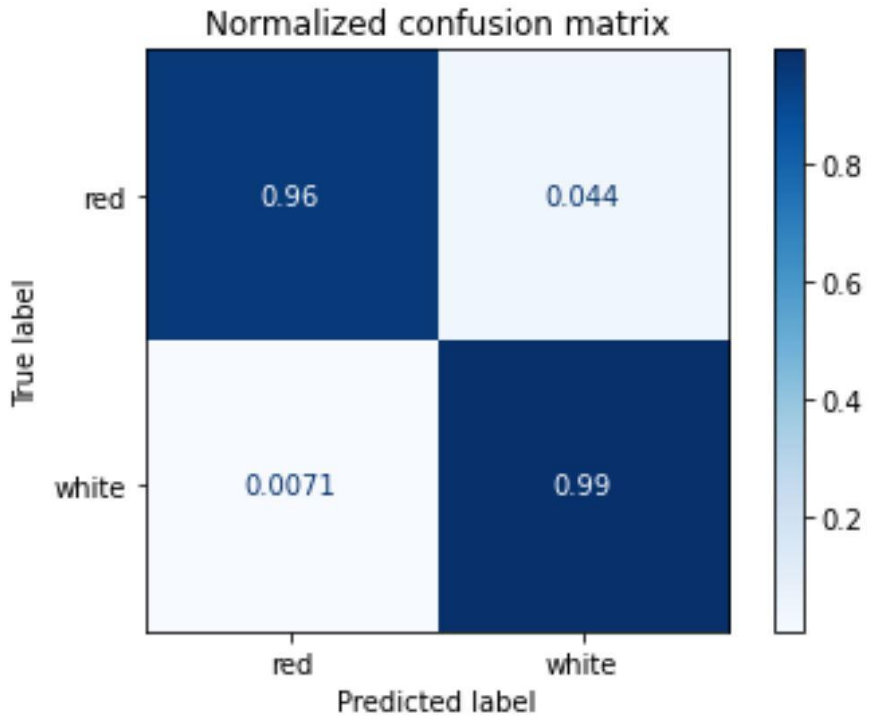
WHEN REGULARIZATION WAS ADDED TO OUR MODEL, THE OVERALL ACCURACY RATE DECREASED.

Predicting Wine Type

- LOGISTIC REGRESSION

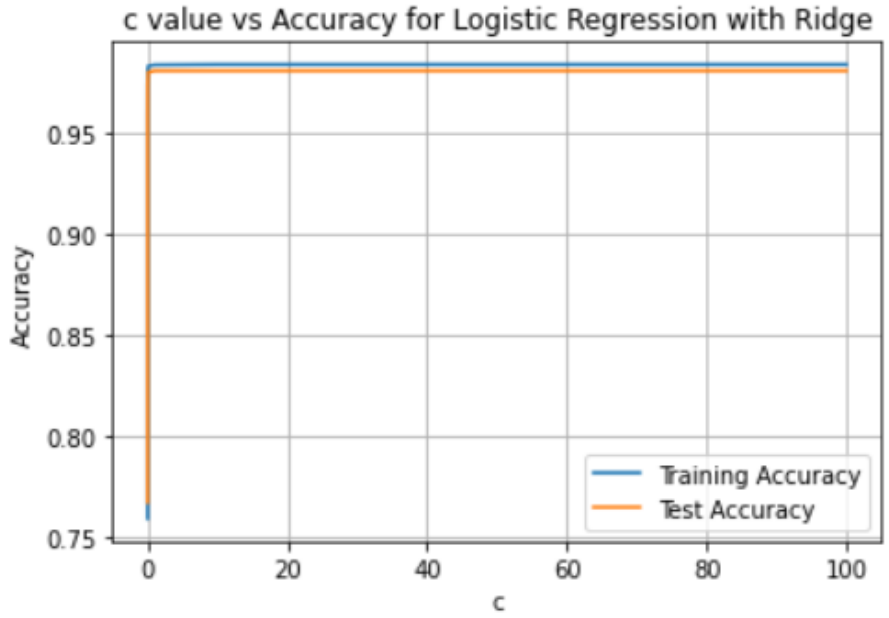


- SVM



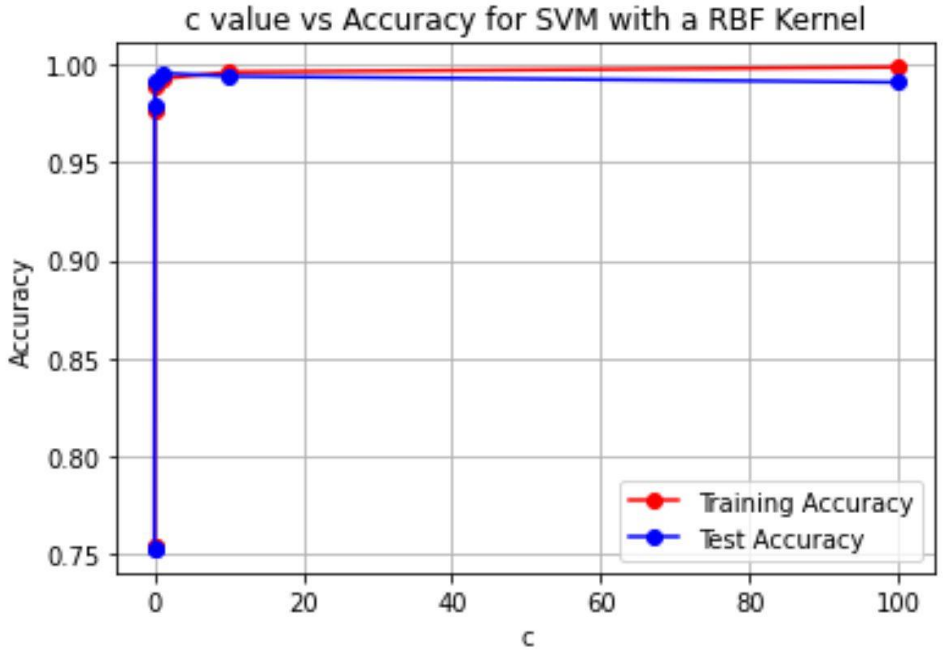
Predicting Wine Type (cont.)

- LOGISTIC REGRESSION WITH REGULARIZATION



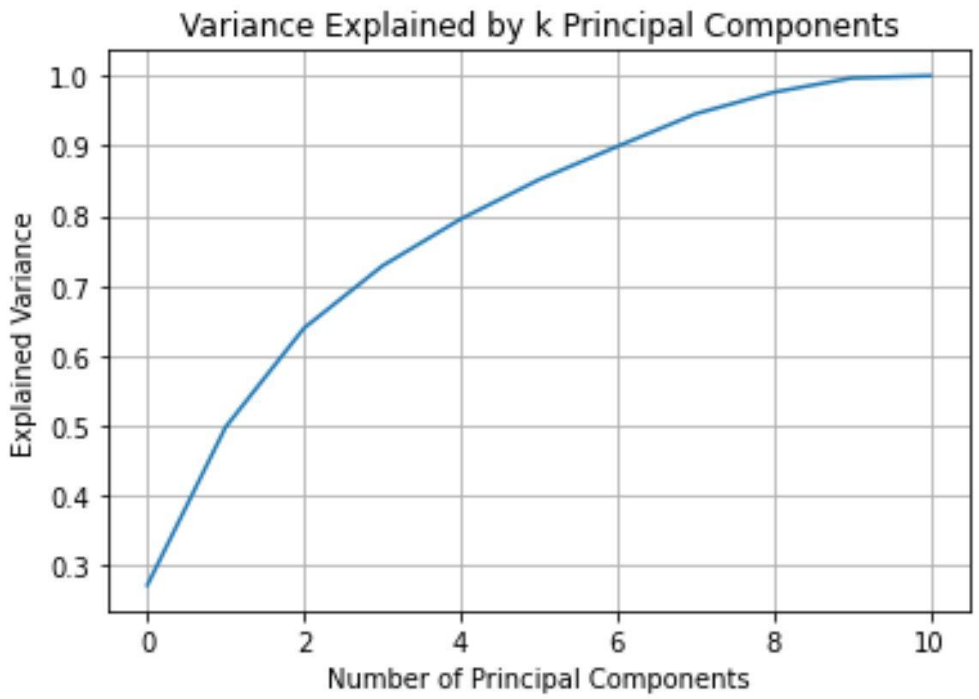
ADDING REGULARIZATION WORSENED THE PERFORMANCE OF MODEL. THE PIVOT POINT OF THE GRAPH OCCURRED AT C = 1 (DEFAULT VALUE).

- SVM WITH RBF KERNEL

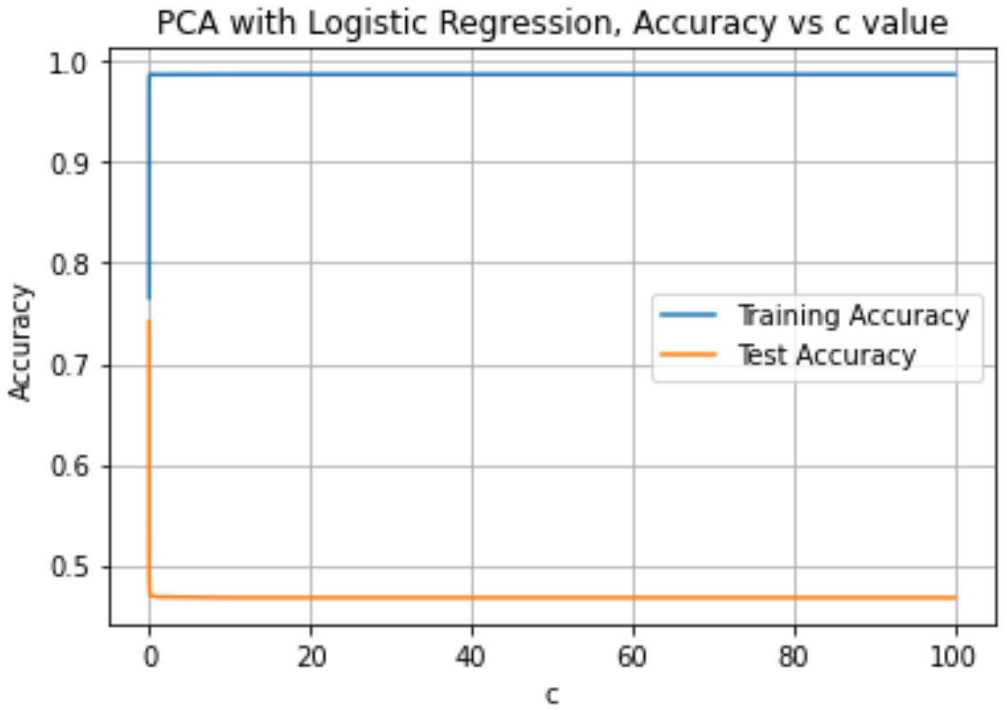


ALL KERNELS FOR SVM PRESENTED EQUALLY GOOD RESULTS. AS REGULARIZATION WAS INCREASED, THE ACCURACY OF THE MODEL WAS DECREASED.

PCA



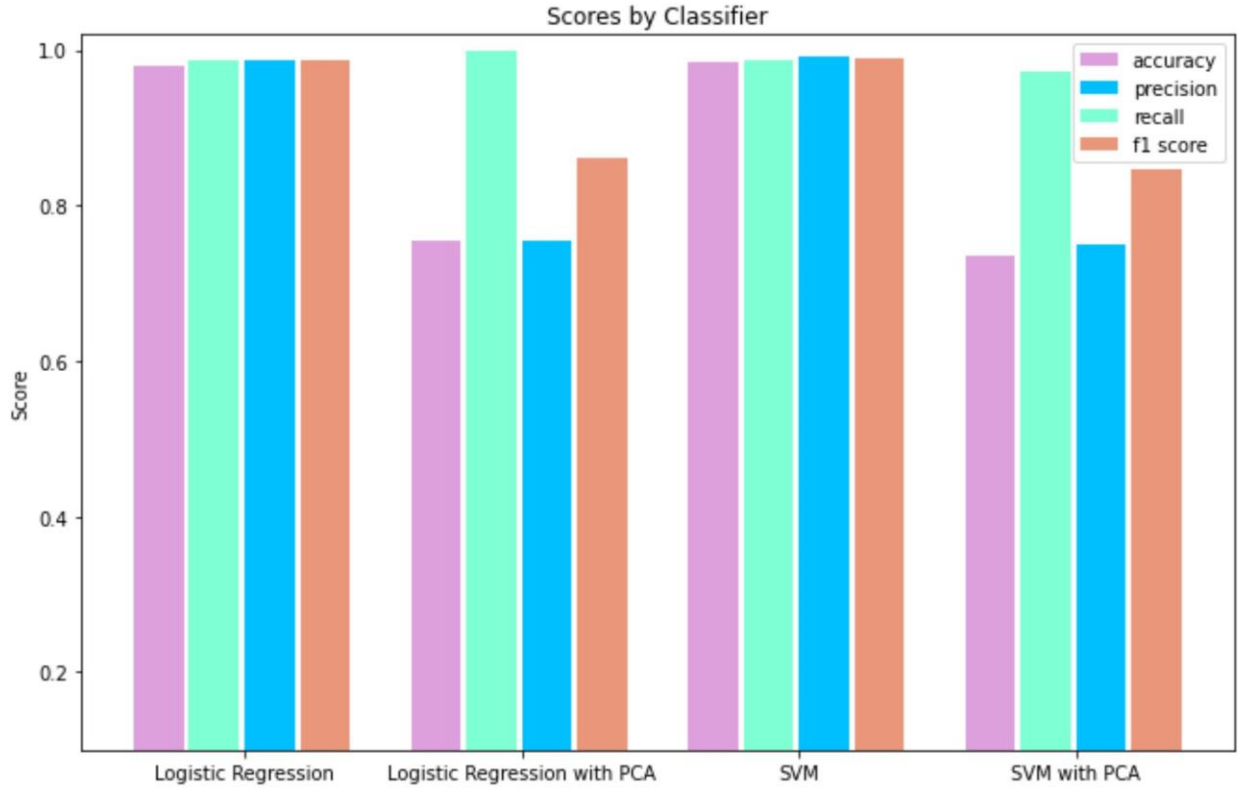
NO EXPLICIT STRUCTURE WAS FOUND AFTER APPLYING PCA



THE ACCURACY TEST SCORES WERE MUCH LOWER THAN EXPECTED

Overall Results for Classification

- LOGISTIC REGRESSION AND SVM SHOWED PROMISING RESULTS
- APPLYING PCA LOWERED THE ACCURACY AND PRECISION SCORES
- 98.92 % ACCURACY (SVM)



References

- HTTPS://ARCHIVE.ICS.UCI.EDU/ML/DATASETS/WINE+QUALITY
- HTTPS://TOWARDSDATASCIENCE.COM/THE-ART-OF-EFFECTIVE-VISUALIZATION-OF-MULTI-DIMENSIONAL-DATA-6C7202990C57#:~:TEXT=THE%20BEST%20WAY%20TO%20GO,A%20DIMENSION%20IN%20THE%20DATA