For this wine quality dataset from the UCI Machine Learning Repository, data from many samples of red and white vinho verde wine from Portugal is presented. The results of physiochemical tests for fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol are presented. The results of a taste test for quality by domain experts is also included on a scale of 0 to 10. Using this data, classification needs to be performed to ensure that the quality of wine being produced is high and also to make sure that products sold as high quality wine are not actually wines which have been illegally adulterated [1]. Performing classification on wine is difficult because taste is the least understood of the human senses, making classification a difficult and subjective task. For this particular dataset, several classification techniques have already been applied to it. First, the creators of the data performed a multivariate linear regression on it. They then used a neural network to perform classification on it and lastly used a support vector machine to classify it [1]. The multivariate linear regression provided results which were easiest to understand and the support vector machine provided the best results. One difficulty that was found with using neural networks and support vector machines to classify the data was that, while no a priori information is required, several hyperparameters such as the number of hidden nodes in a neural network or the kernel parameter in a support vector machine must be correctly set to get good results.

# References

|  |  |
| --- | --- |
| [1] | P. Cortez et al., "Modeling wine preferences by data mining from physiochemical properties," *Decision Support Systems,* pp. 547-533, 2009. |