



Project Proposal

My Nordqvist (myno1600) & Herman Jansson (heja1600)



Wine Quality Data Set

Red wine (1599 instances)

White wine (4898 instances)

The unnormalized attributes will be normalized

Fixed acidity	Volatile acidity	Citric acid	Residual sugar	Chlorides	Free sulfur dioxide	Total sulfur dioxide	Density	PH	Sulphates	Alcohol	Quality
7	0.27	0.36	20.7	0.05	45	170	1	3	0.45	8.8	6
6.3	0.3	0.34	1.6	0.05	14	132	0.99	3.3	0.49	9.5	6
8.1	0.28	0.4	6.9	0.05	30	97	1	3.26	0.44	10.1	6
7.2	0.23	0.32	8.5	0.06	47	186	1	3.19	0.4	9.9	6
7.2	0.23	0.32	8.5	0.06	47	186	1	3.19	0.4	9.9	6
8.1	0.28	0.4	6.9	0.05	30	97	1	3.26	0.44	10.1	6
6.2	0.32	0.16	7	0.05	30	136	0.99	3.18	0.47	9.6	6
7	0.27	0.36	20.7	0.05	45	170	1	3	0.45	8.8	6
6.3	0.3	0.34	1.6	0.05	14	132	0.99	3.3	0.49	9.5	6
8.1	0.22	0.43	1.5	0.04	28	129	0.99	3.22	0.45	11	6
8.1	0.27	0.41	1.45	0.03	11	63	0.99	2.99	0.56	12	5
8.6	0.23	0.4	4.2	0.04	17	109	0.99	3.14	0.53	9.7	5



Method 1

How can we know based on input whether its white or red wine?

Dataset

- Red wine (1599 instances)
- White wine (4898 instances)

Data mining method

- Decision tree (Predict yes or no outcome of varying variables/attributes)

Training and Testing

- Bootstrap datasets and use holdout for dividing training and test sets

Output

- Whether its white or red wine based on attributes



Method 1

How can we know based on input whether its white or red wine?

Evaluation Methods

- leave-one-out cross validation (Prediction accuracy for decision tree algorithm)
- Confusion matrix
 - Whether the predicted value of white/red wine actually is white/red wine
 - Matthews correlation matrix to evaluate the confusion matrix, determine the accuracy.



Method 2

How could we know the quality of newly produced wines?

Dataset

- White wine (4898 instances)

Data mining method

- Multiple Linear Regression
 - Serves multiple variables to predict outcome of different variables

Training and Testing

- Bootstrap datasets and use holdout for dividing training and test sets

Output

- Wine quality learned attributes



Method 2

How could we know the quality of newly produced wines?

Evaluation Method

- Backward elimination (eliminates all non-significant variables)