## Wine Data SAS Analysis: Rita Dicarlo, Katie Clewett, Chang Guo
## Classification

### The HPSPLIT Procedure

| Performance Information | |
|---|---|
| Execution Mode | Single-Machine |
| Number of Threads | 2 |

Data Access Information

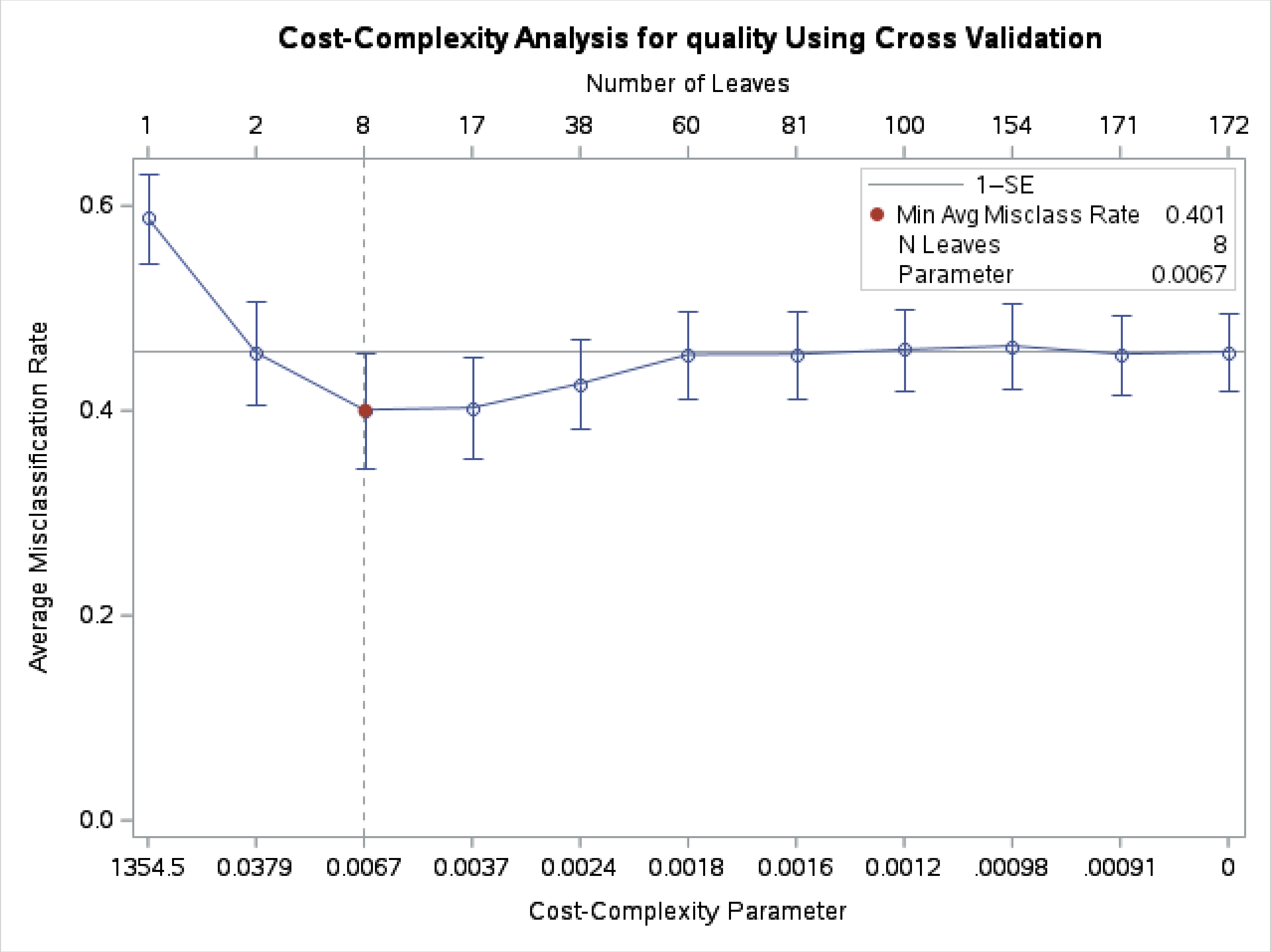
| Data | Engine | Role | Path |
|---|---|---|---|
| WORK.TRAIN | V9 | Input | On Client |

| Model Information | |
|---|---|
| Split Criterion Used | Entropy |
| Pruning Method | Cost-Complexity |
| Subtree Evaluation Criterion | Cost-Complexity |
| Number of Branches | 2 |
| Maximum Tree Depth Requested | 10 |
| Maximum Tree Depth Achieved | 10 |
| Tree Depth | 5 |
| Number of Leaves Before Pruning | 197 |
| Number of Leaves After Pruning | 8 |

| | |
|---|---|
| Number of Observations Read | 945 |
| Number of Observations Used | 945 |

## Wine Data SAS Analysis: Rita Dicarlo, Katie Clewett, Chang Guo
## Classification

### The HPSPLIT Procedure



**10-Fold Cross Validation Assessment of Model**

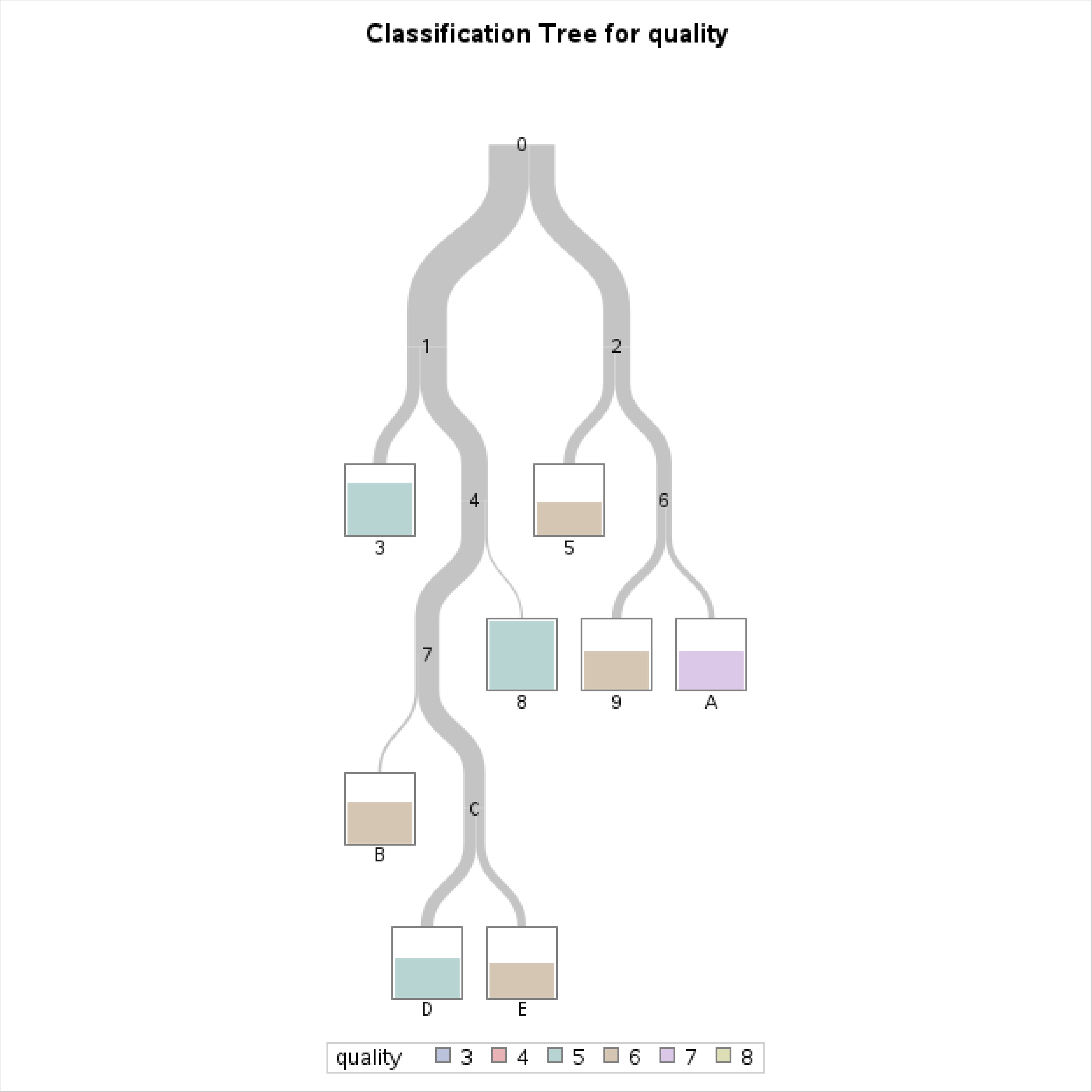
|  | Average Square Error |  |  |  | Number of Leaves |  |  | Misclassification Rate |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|
| N Leaves | Min | Avg | Standard Error | Max | Min | Median | Max | Min | Avg | Standard Error | Max |
| 7 | 0.0782 | 0.0948 | 0.00715 | 0.1039 | 5 | 7.5 | 11 | 0.2959 | 0.4323 | 0.0601 | 0.5000 |

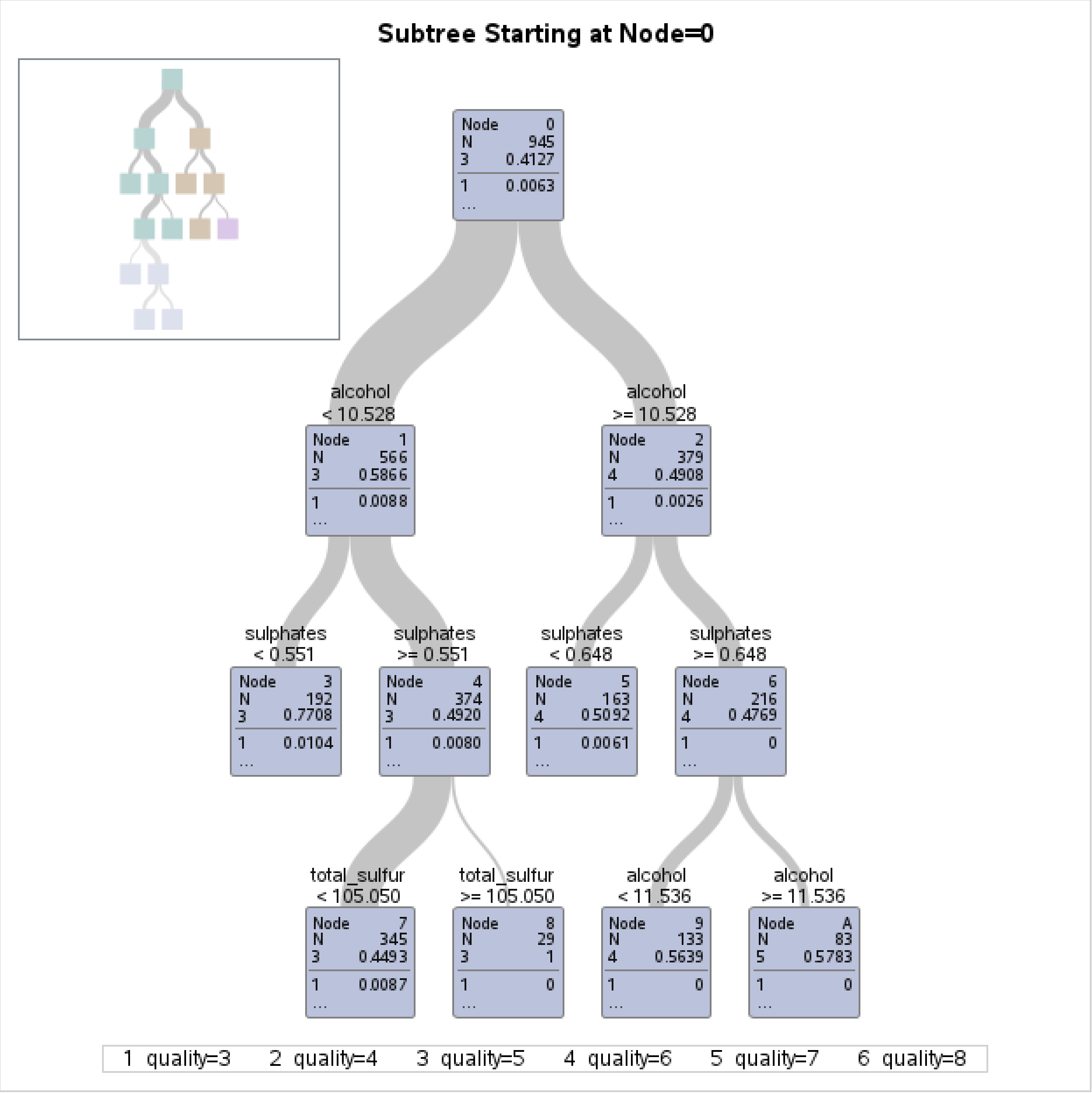
**10-Fold Cross Validation Confusion Matrix**

|  | Predicted |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| Actual | 3 | 4 | 5 | 6 | 7 | 8 | Error Rate |
| 3 | 0 | 0 | 3 | 3 | 0 | 0 | 1.0000 |
| 4 | 0 | 0 | 16 | 17 | 0 | 0 | 1.0000 |
| 5 | 0 | 0 | 269 | 120 | 1 | 0 | 0.3103 |
| 6 | 0 | 0 | 125 | 220 | 28 | 0 | 0.4102 |
| 7 | 0 | 0 | 11 | 72 | 49 | 0 | 0.6288 |
| 8 | 0 | 0 | 0 | 5 | 6 | 0 | 1.0000 |

## Wine Data SAS Analysis: Rita Dicarlo, Katie Clewett, Chang Guo
## Classification

### The HPSPLIT Procedure


Classification Tree for quality
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
quality
3
4
5
6
7
8



## Wine Data SAS Analysis: Rita Dicarlo, Katie Clewett, Chang Guo
## Classification

### The HPSPLIT Procedure

**Confusion Matrices**

| | | Predicted | | | | | | Error |
|---|---|---|---|---|---|---|---|---|
| | Actual | 3 | 4 | 5 | 6 | 7 | 8 | Rate |
| Model Based | 3 | 0 | 0 | 3 | 3 | 0 | 0 | 1.0000 |
| | 4 | 0 | 0 | 15 | 18 | 0 | 0 | 1.0000 |
| | 5 | 0 | 0 | 282 | 107 | 1 | 0 | 0.2769 |
| | 6 | 0 | 0 | 97 | 248 | 28 | 0 | 0.3351 |
| | 7 | 0 | 0 | 3 | 81 | 48 | 0 | 0.6364 |
| | 8 | 0 | 0 | 0 | 5 | 6 | 0 | 1.0000 |
| Cross Validation | 3 | 0 | 0 | 3 | 3 | 0 | 0 | 1.0000 |
| | 4 | 0 | 0 | 16 | 17 | 0 | 0 | 1.0000 |
| | 5 | 0 | 0 | 269 | 120 | 1 | 0 | 0.3103 |
| | 6 | 0 | 0 | 125 | 220 | 28 | 0 | 0.4102 |

| Confusion Matrices | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Predicted | | | | | | Error |
| | Actual | 3 | 4 | 5 | 6 | 7 | 8 | Rate |
| | 7 | 0 | 0 | 11 | 72 | 49 | 0 | 0.6288 |
| | 8 | 0 | 0 | 0 | 5 | 6 | 0 | 1.0000 |

| Fit Statistics for Selected Tree | | | | | | |
|---|---|---|---|---|---|---|
| | N Leaves | ASE | Mis-class | Entropy | Gini | RSS |
| Model Based | 8 | 0.0872 | 0.3884 | 1.3597 | 0.5233 | 494.5 |
| Cross Validation | 7 | 0.0948 | 0.4323 | | | |

| Variable Importance | | | |
|---|---|---|---|
| | Training | | |
| Variable | Relative | Importance | Count |
| alcohol | 1.0000 | 8.8686 | 3 |
| sulphates | 0.5490 | 4.8685 | 2 |
| total_sulfur | 0.4143 | 3.6742 | 1 |
| vol_acidity | 0.2918 | 2.5882 | 1 |

# Wine Data SAS Analysis: Rita Dicarlo, Katie Clewett, Chang Guo
# Classification

## The HPFOREST Procedure

| Performance Information | |
|---|---|
| Execution Mode | Single-Machine |
| Number of Threads | 2 |

| Data Access Information | | | |
|---|---|---|---|
| Data | Engine | Role | Path |
| WORK.TRAIN | V9 | Input | On Client |

| Model Information | | |
|---|---|---|
| Parameter | Value | |
| Variables to Try | 3 | (Default) |
| Maximum Trees | 100 | |
| Actual Trees | 100 | |
| Inbag Fraction | 0.3 | |
| Prune Fraction | 0 | (Default) |
| Prune Threshold | 0.1 | (Default) |
| Leaf Fraction | 0.00001 | (Default) |
| Leaf Size Setting | 1 | (Default) |
| Leaf Size Used | 1 | |
| Category Bins | 30 | (Default) |
| Interval Bins | 100 | |
| Minimum Category Size | 5 | (Default) |
| Node Size | 100000 | (Default) |
| Maximum Depth | 20 | (Default) |
| Alpha | 1 | (Default) |

| Model Information | | |
|---|---|---|
| **Parameter** | **Value** | |
| **Exhaustive** | 5000 | (Default) |
| **Rows of Sequence to Skip** | 5 | (Default) |
| **Split Criterion** | . | Gini |
| **Preselection Method** | . | BinnedSearch |
| **Missing Value Handling** | . | Valid value |

| Number of Observations | |
|---|---|
| **Type** | **N** |
| **Number of Observations Read** | 945 |
| **Number of Observations Used** | 945 |

| Baseline Fit Statistics | |
|---|---|
| **Statistic** | **Value** |
| **Average Square Error** | 0.109 |
| **Misclassification Rate** | 0.587 |
| **Log Loss** | 1.208 |

**Fit Statistics**

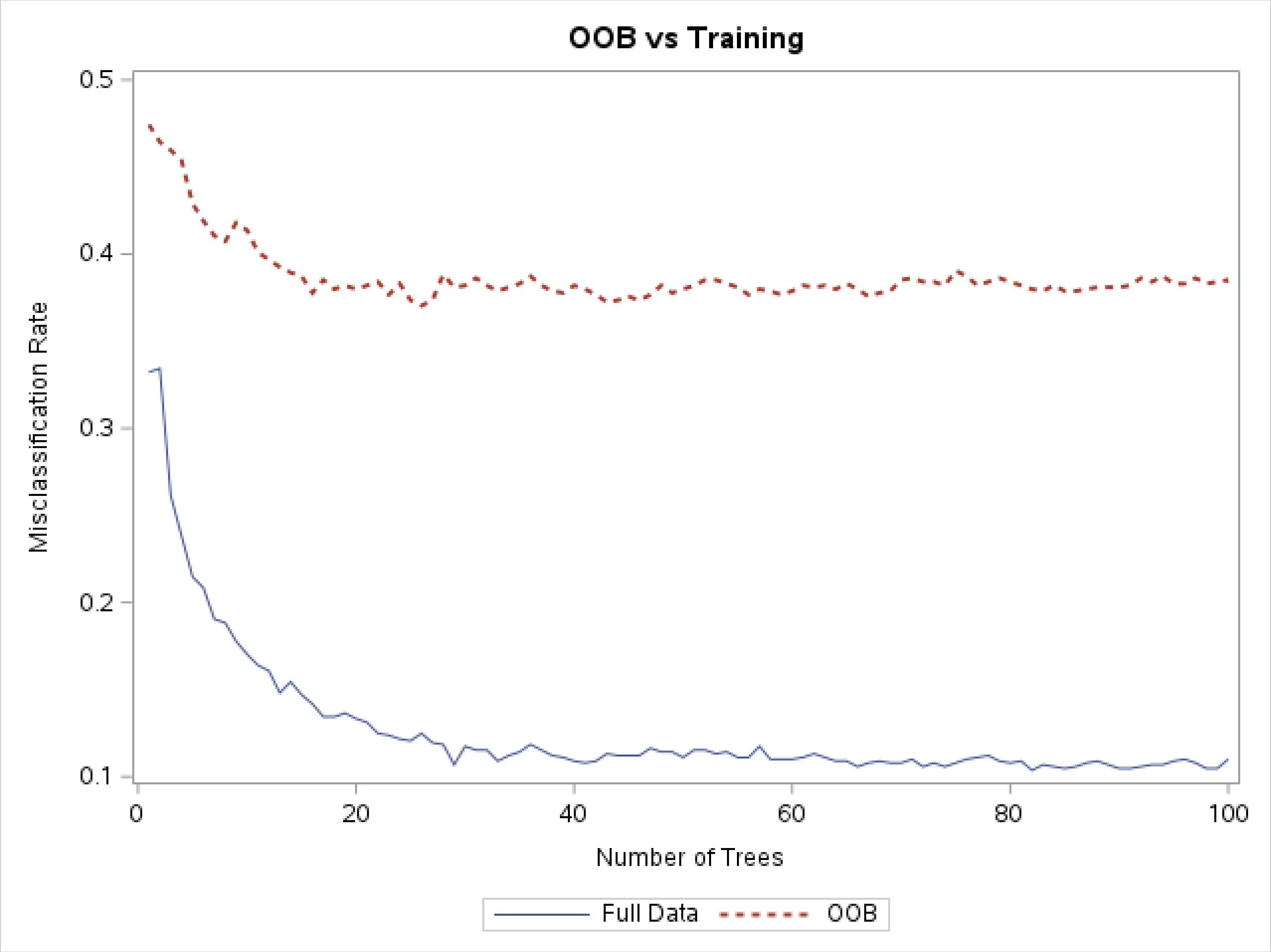
| Number of Trees | Number of Leaves | Average Square Error (Train) | Average Square Error (OOB) | Misclassification Rate (Train) | Misclassification Rate (OOB) | Log Loss (Train) | Log Loss (OOB) |
|---|---|---|---|---|---|---|---|
| **1** | **108** | 0.1108 | 0.1581 | 0.332 | 0.474 | 7.651 | 10.92 |
| **2** | **207** | 0.0732 | 0.1323 | 0.334 | 0.464 | 3.342 | 8.30 |
| **3** | **322** | 0.0609 | 0.1175 | 0.261 | 0.460 | 1.800 | 6.55 |
| **4** | **420** | 0.0577 | 0.1110 | 0.238 | 0.454 | 1.294 | 5.40 |
| **5** | **517** | 0.0535 | 0.1041 | 0.215 | 0.429 | 0.930 | 4.58 |
| **6** | **624** | 0.0503 | 0.0991 | 0.208 | 0.419 | 0.674 | 3.75 |
| **7** | **736** | 0.0485 | 0.0966 | 0.190 | 0.411 | 0.580 | 3.39 |
| **8** | **832** | 0.0482 | 0.0954 | 0.188 | 0.407 | 0.540 | 3.00 |
| **9** | **925** | 0.0476 | 0.0940 | 0.178 | 0.418 | 0.538 | 2.70 |
| **10** | **1034** | 0.0467 | 0.0921 | 0.170 | 0.414 | 0.512 | 2.50 |
| **11** | **1153** | 0.0452 | 0.0892 | 0.164 | 0.401 | 0.502 | 2.20 |
| **12** | **1260** | 0.0449 | 0.0891 | 0.161 | 0.397 | 0.503 | 2.16 |
| **13** | **1359** | 0.0445 | 0.0885 | 0.148 | 0.393 | 0.501 | 2.09 |
| **14** | **1475** | 0.0438 | 0.0876 | 0.154 | 0.389 | 0.475 | 2.02 |
| **15** | **1582** | 0.0432 | 0.0864 | 0.147 | 0.387 | 0.472 | 1.93 |
| **16** | **1686** | 0.0429 | 0.0859 | 0.142 | 0.378 | 0.469 | 1.88 |
| **17** | **1798** | 0.0425 | 0.0853 | 0.134 | 0.385 | 0.469 | 1.84 |
| **18** | **1900** | 0.0421 | 0.0846 | 0.134 | 0.380 | 0.466 | 1.79 |
| **19** | **1996** | 0.0419 | 0.0842 | 0.137 | 0.382 | 0.465 | 1.77 |
| **20** | **2111** | 0.0415 | 0.0836 | 0.133 | 0.380 | 0.463 | 1.74 |
| **21** | **2214** | 0.0412 | 0.0834 | 0.131 | 0.382 | 0.462 | 1.72 |
| **22** | **2312** | 0.0411 | 0.0834 | 0.125 | 0.384 | 0.462 | 1.68 |
| **23** | **2419** | 0.0411 | 0.0834 | 0.124 | 0.377 | 0.462 | 1.60 |
| **24** | **2524** | 0.0410 | 0.0831 | 0.122 | 0.383 | 0.460 | 1.60 |
| **25** | **2625** | 0.0408 | 0.0828 | 0.121 | 0.374 | 0.459 | 1.55 |
| **26** | **2729** | 0.0408 | 0.0826 | 0.125 | 0.370 | 0.459 | 1.53 |
| **27** | **2829** | 0.0409 | 0.0830 | 0.120 | 0.374 | 0.462 | 1.53 |

**Fit Statistics**

| Number of Trees | Number of Leaves | Average Square Error (Train) | Average Square Error (OOB) | Misclassification Rate (Train) | Misclassification Rate (OOB) | Log Loss (Train) | Log Loss (OOB) |
|---:|---:|---:|---:|---:|---:|---:|---:|
| 28 | 2944 | 0.0408 | 0.0829 | 0.119 | 0.388 | 0.462 | 1.51 |
| 29 | 3044 | 0.0407 | 0.0828 | 0.107 | 0.381 | 0.462 | 1.51 |
| 30 | 3152 | 0.0407 | 0.0827 | 0.117 | 0.382 | 0.461 | 1.51 |
| 31 | 3251 | 0.0408 | 0.0827 | 0.115 | 0.386 | 0.462 | 1.51 |
| 32 | 3347 | 0.0409 | 0.0827 | 0.115 | 0.382 | 0.463 | 1.51 |
| 33 | 3463 | 0.0407 | 0.0826 | 0.109 | 0.379 | 0.462 | 1.51 |
| 34 | 3574 | 0.0406 | 0.0824 | 0.112 | 0.381 | 0.462 | 1.51 |
| 35 | 3680 | 0.0406 | 0.0824 | 0.114 | 0.383 | 0.462 | 1.49 |
| 36 | 3790 | 0.0406 | 0.0825 | 0.119 | 0.387 | 0.463 | 1.49 |
| 37 | 3888 | 0.0406 | 0.0824 | 0.115 | 0.382 | 0.463 | 1.49 |
| 38 | 3995 | 0.0406 | 0.0823 | 0.112 | 0.379 | 0.463 | 1.47 |
| 39 | 4091 | 0.0405 | 0.0822 | 0.111 | 0.378 | 0.463 | 1.47 |
| 40 | 4200 | 0.0405 | 0.0823 | 0.109 | 0.382 | 0.463 | 1.47 |
| 41 | 4305 | 0.0405 | 0.0822 | 0.108 | 0.380 | 0.463 | 1.41 |
| 42 | 4412 | 0.0405 | 0.0823 | 0.109 | 0.377 | 0.464 | 1.41 |
| 43 | 4522 | 0.0404 | 0.0822 | 0.113 | 0.372 | 0.463 | 1.41 |
| 44 | 4623 | 0.0404 | 0.0822 | 0.112 | 0.374 | 0.464 | 1.41 |
| 45 | 4726 | 0.0403 | 0.0821 | 0.112 | 0.376 | 0.463 | 1.39 |
| 46 | 4826 | 0.0404 | 0.0821 | 0.112 | 0.374 | 0.463 | 1.39 |
| 47 | 4929 | 0.0403 | 0.0820 | 0.116 | 0.377 | 0.463 | 1.39 |
| 48 | 5025 | 0.0404 | 0.0821 | 0.114 | 0.382 | 0.463 | 1.39 |
| 49 | 5126 | 0.0405 | 0.0822 | 0.114 | 0.378 | 0.463 | 1.39 |
| 50 | 5236 | 0.0405 | 0.0822 | 0.111 | 0.380 | 0.463 | 1.37 |
| 51 | 5344 | 0.0404 | 0.0822 | 0.115 | 0.382 | 0.463 | 1.35 |
| 52 | 5453 | 0.0404 | 0.0823 | 0.115 | 0.385 | 0.464 | 1.35 |
| 53 | 5555 | 0.0403 | 0.0821 | 0.113 | 0.385 | 0.463 | 1.35 |
| 54 | 5653 | 0.0403 | 0.0821 | 0.114 | 0.383 | 0.464 | 1.35 |
| 55 | 5756 | 0.0403 | 0.0819 | 0.111 | 0.381 | 0.463 | 1.35 |
| 56 | 5860 | 0.0402 | 0.0818 | 0.111 | 0.377 | 0.463 | 1.35 |
| 57 | 5966 | 0.0402 | 0.0818 | 0.117 | 0.380 | 0.463 | 1.35 |
| 58 | 6069 | 0.0402 | 0.0818 | 0.110 | 0.379 | 0.462 | 1.33 |
| 59 | 6166 | 0.0402 | 0.0818 | 0.110 | 0.377 | 0.463 | 1.31 |
| 60 | 6275 | 0.0401 | 0.0816 | 0.110 | 0.379 | 0.462 | 1.27 |
| 61 | 6363 | 0.0401 | 0.0817 | 0.111 | 0.382 | 0.462 | 1.27 |
| 62 | 6471 | 0.0402 | 0.0817 | 0.113 | 0.381 | 0.463 | 1.27 |
| 63 | 6578 | 0.0402 | 0.0817 | 0.111 | 0.382 | 0.463 | 1.27 |
| 64 | 6686 | 0.0402 | 0.0817 | 0.109 | 0.380 | 0.462 | 1.27 |
| 65 | 6790 | 0.0401 | 0.0815 | 0.109 | 0.383 | 0.462 | 1.27 |
| 66 | 6891 | 0.0401 | 0.0815 | 0.106 | 0.380 | 0.462 | 1.27 |
| 67 | 6992 | 0.0401 | 0.0816 | 0.108 | 0.376 | 0.462 | 1.27 |
| 68 | 7097 | 0.0400 | 0.0814 | 0.109 | 0.378 | 0.462 | 1.27 |
| 69 | 7207 | 0.0400 | 0.0815 | 0.108 | 0.379 | 0.462 | 1.27 |
| 70 | 7306 | 0.0401 | 0.0814 | 0.108 | 0.385 | 0.462 | 1.26 |
| 71 | 7410 | 0.0400 | 0.0814 | 0.110 | 0.386 | 0.461 | 1.26 |
| 72 | 7523 | 0.0400 | 0.0814 | 0.106 | 0.384 | 0.461 | 1.24 |
| 73 | 7624 | 0.0400 | 0.0814 | 0.108 | 0.384 | 0.461 | 1.25 |

| Fit Statistics | | | | | | | |
|---|---|---|---|---|---|---|---|
| Number of Trees | Number of Leaves | Average Square Error (Train) | Average Square Error (OOB) | Misclassification Rate (Train) | Misclassification Rate (OOB) | Log Loss (Train) | Log Loss (OOB) |
| 74 | 7724 | 0.0400 | 0.0814 | 0.106 | 0.382 | 0.461 | 1.25 |
| 75 | 7823 | 0.0401 | 0.0814 | 0.108 | 0.390 | 0.462 | 1.25 |
| 76 | 7926 | 0.0401 | 0.0816 | 0.110 | 0.387 | 0.462 | 1.25 |
| 77 | 8028 | 0.0401 | 0.0815 | 0.111 | 0.382 | 0.462 | 1.25 |
| 78 | 8130 | 0.0401 | 0.0815 | 0.112 | 0.384 | 0.462 | 1.25 |
| 79 | 8232 | 0.0401 | 0.0815 | 0.109 | 0.386 | 0.462 | 1.25 |
| 80 | 8332 | 0.0401 | 0.0815 | 0.108 | 0.384 | 0.462 | 1.25 |
| 81 | 8436 | 0.0400 | 0.0813 | 0.109 | 0.382 | 0.462 | 1.22 |
| 82 | 8553 | 0.0400 | 0.0812 | 0.104 | 0.380 | 0.462 | 1.22 |
| 83 | 8659 | 0.0400 | 0.0812 | 0.107 | 0.379 | 0.462 | 1.21 |
| 84 | 8763 | 0.0400 | 0.0813 | 0.106 | 0.382 | 0.462 | 1.21 |
| 85 | 8863 | 0.0400 | 0.0813 | 0.105 | 0.379 | 0.462 | 1.21 |
| 86 | 8964 | 0.0400 | 0.0813 | 0.106 | 0.379 | 0.462 | 1.21 |
| 87 | 9072 | 0.0400 | 0.0813 | 0.108 | 0.380 | 0.462 | 1.21 |
| 88 | 9186 | 0.0399 | 0.0813 | 0.109 | 0.381 | 0.462 | 1.21 |
| 89 | 9295 | 0.0399 | 0.0813 | 0.107 | 0.381 | 0.462 | 1.21 |
| 90 | 9396 | 0.0399 | 0.0812 | 0.105 | 0.381 | 0.461 | 1.21 |
| 91 | 9485 | 0.0399 | 0.0811 | 0.105 | 0.382 | 0.461 | 1.20 |
| 92 | 9573 | 0.0399 | 0.0812 | 0.106 | 0.386 | 0.462 | 1.20 |
| 93 | 9674 | 0.0400 | 0.0812 | 0.107 | 0.384 | 0.462 | 1.20 |
| 94 | 9782 | 0.0399 | 0.0812 | 0.107 | 0.387 | 0.462 | 1.20 |
| 95 | 9879 | 0.0399 | 0.0811 | 0.109 | 0.383 | 0.462 | 1.18 |
| 96 | 9980 | 0.0399 | 0.0811 | 0.110 | 0.383 | 0.461 | 1.18 |
| 97 | 10078 | 0.0399 | 0.0810 | 0.108 | 0.386 | 0.461 | 1.16 |
| 98 | 10186 | 0.0398 | 0.0810 | 0.105 | 0.383 | 0.461 | 1.16 |
| 99 | 10289 | 0.0398 | 0.0810 | 0.105 | 0.384 | 0.461 | 1.16 |
| 100 | 10389 | 0.0398 | 0.0810 | 0.110 | 0.385 | 0.461 | 1.16 |

| Loss Reduction Variable Importance | | | | | |
|---|---|---|---|---|---|
| Variable | Number of Rules | Gini | OOB Gini | Margin | OOB Margin |
| alcohol | 2178 | 0.163513 | -0.01866 | 0.239715 | 0.05579 |
| chlorides | 626 | 0.035312 | -0.02884 | 0.058962 | -0.00307 |
| vol_acidity | 1135 | 0.081046 | -0.04207 | 0.121102 | 0.00161 |
| free_sulfur | 1050 | 0.058708 | -0.04628 | 0.103072 | -0.00129 |
| total_sulfur | 1553 | 0.099693 | -0.05383 | 0.171029 | 0.01710 |
| sulphates | 2059 | 0.124944 | -0.06134 | 0.211574 | 0.02173 |
| pH | 1688 | 0.086853 | -0.07573 | 0.151529 | -0.01331 |



## High Quality Wine
## Classification using validation data set

### The LOGISTIC Procedure

| Model Information | |
|---|---|
| Data Set | WORK.VALIDATION_SET |
| Response Variable | high_quality |
| Number of Response Levels | 2 |
| Model | binary logit |
| Optimization Technique | Fisher's scoring |

| | |
|---|---|
| Number of Observations Read | 481 |
| Number of Observations Used | 481 |

| Response Profile | | |
|---|---|---|
| Ordered Value | high_quality | Total Frequency |
| 1 | 0 | 432 |
| 2 | 1 | 49 |

**Probability modeled is high_quality=0.**

| Model Convergence Status |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

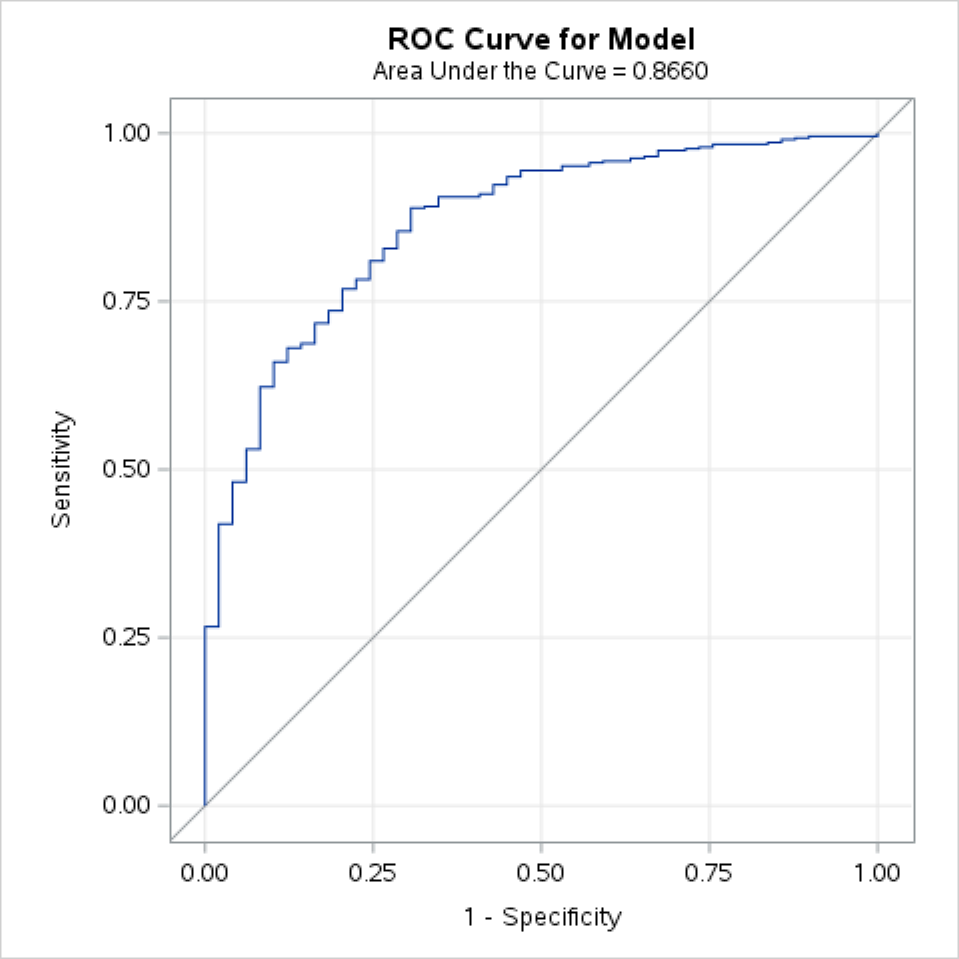
| Model Fit Statistics | | |
|---|---|---|
| Criterion | Intercept Only | Intercept and Covariates |
| AIC | 318.666 | 238.956 |
| SC | 322.842 | 259.835 |
| -2 Log L | 316.666 | 228.956 |

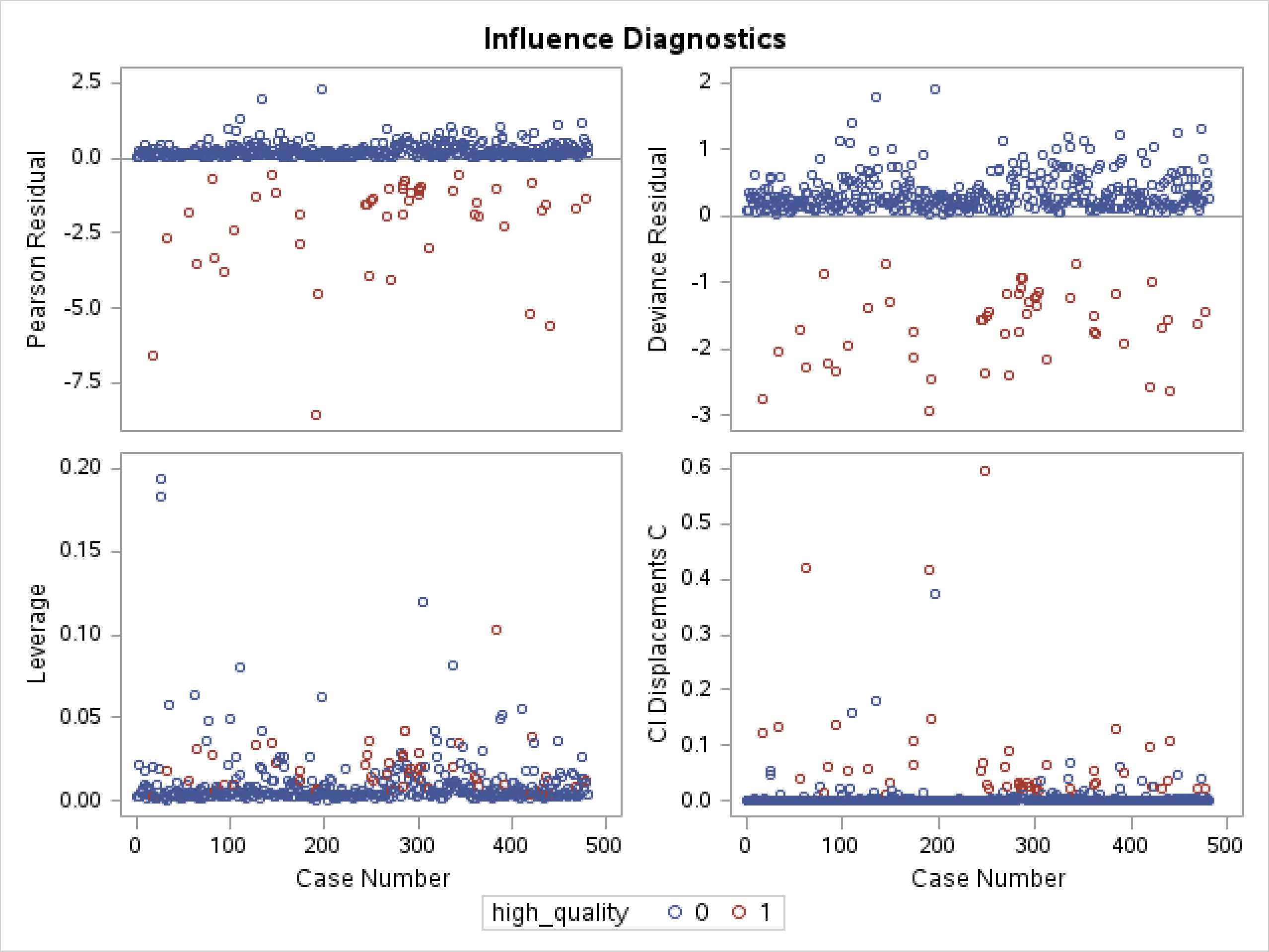
| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 87.7105 | 4 | <.0001 |
| Score | 87.9760 | 4 | <.0001 |
| Wald | 57.6247 | 4 | <.0001 |

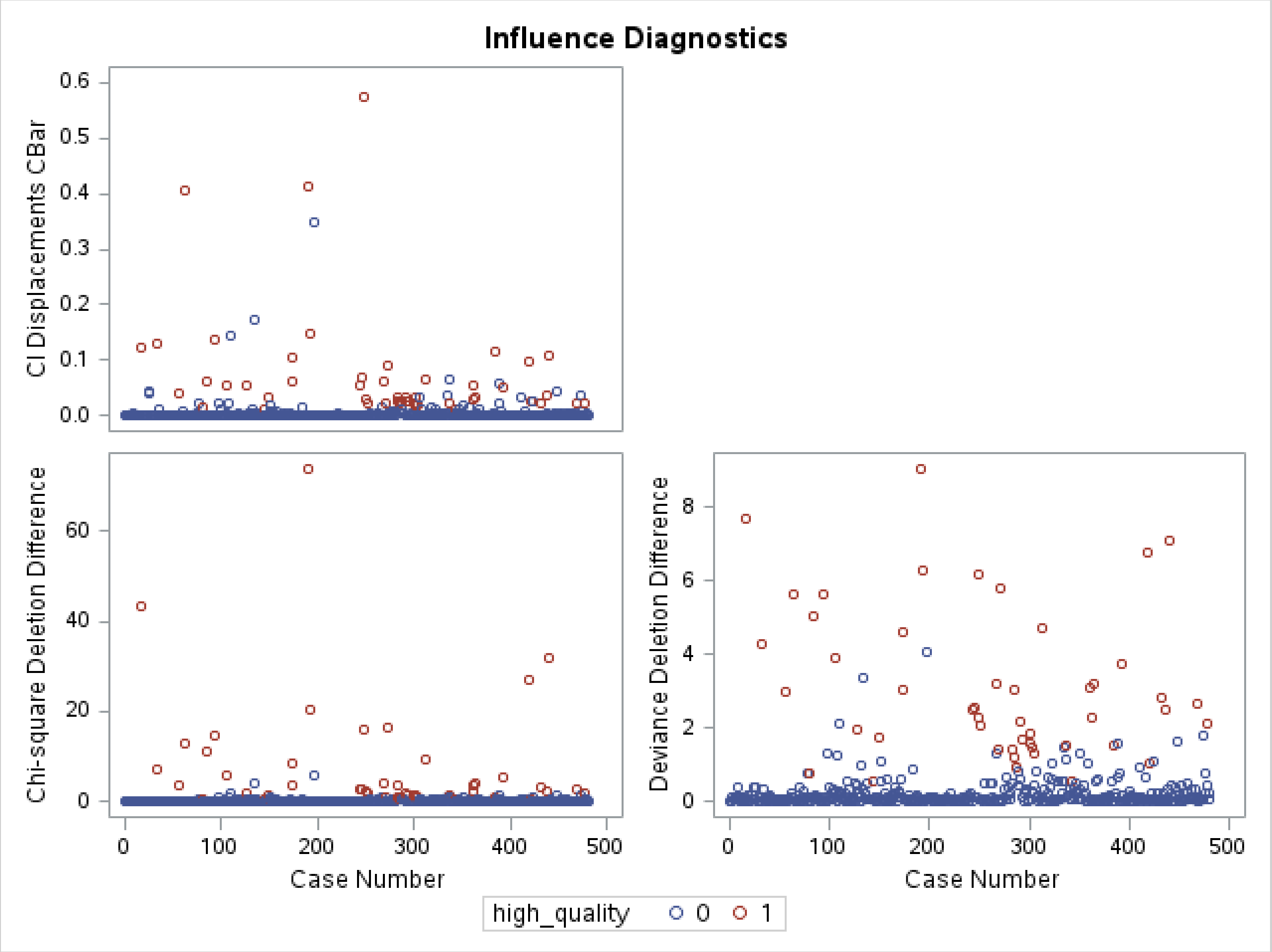
| Analysis of Maximum Likelihood Estimates | | | | | |
|---|---|---|---|---|---|
| | | | Standard | Wald | |
| Parameter | DF | Estimate | Error | Chi-Square | Pr > ChiSq |
| Intercept | 1 | 11.5202 | 2.1183 | 29.5752 | <.0001 |
| alcohol | 1 | -0.9193 | 0.1596 | 33.1670 | <.0001 |
| sulphates | 1 | -2.6453 | 0.9729 | 7.3933 | 0.0065 |
| total_sulfur | 1 | 0.0196 | 0.00820 | 5.7066 | 0.0169 |
| vol_acidity | 1 | 3.7199 | 1.2526 | 8.8197 | 0.0030 |

| Odds Ratio Estimates | | | |
|---|---|---|---|
| Effect | Point Estimate | 95% Wald Confidence Limits | |
| alcohol | 0.399 | 0.292 | 0.545 |
| sulphates | 0.071 | 0.011 | 0.478 |
| total_sulfur | 1.020 | 1.004 | 1.036 |
| vol_acidity | 41.258 | 3.543 | 480.502 |

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 86.6 | Somers' D | 0.732 |
| Percent Discordant | 13.4 | Gamma | 0.732 |
| Percent Tied | 0.0 | Tau-a | 0.134 |
| Pairs | 21168 | c | 0.866 |


ROC Curve for Model
Area Under the Curve = 0.8660
Sensitivity
1.00
0.75
0.50
0.25
0.00
1 - Specificity
0.00
0.25
0.50
0.75
1.00


Influence Diagnostics
Pearson Residual
2.5
0.0
-2.5
-5.0
-7.5
Deviance Residual
2
1
0
-1
-2
-3
Leverage
0.20
0.15
0.10
0.05
0.00
CI Displacements C
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Case Number
0
100
200
300
400
500
high_quality ○ 0 ○ 1



## High Quality Wine
## Classification using validation data set

### The HPSPLIT Procedure

| Performance Information | |
|---|---|
| **Execution Mode** | Single-Machine |
| **Number of Threads** | 2 |

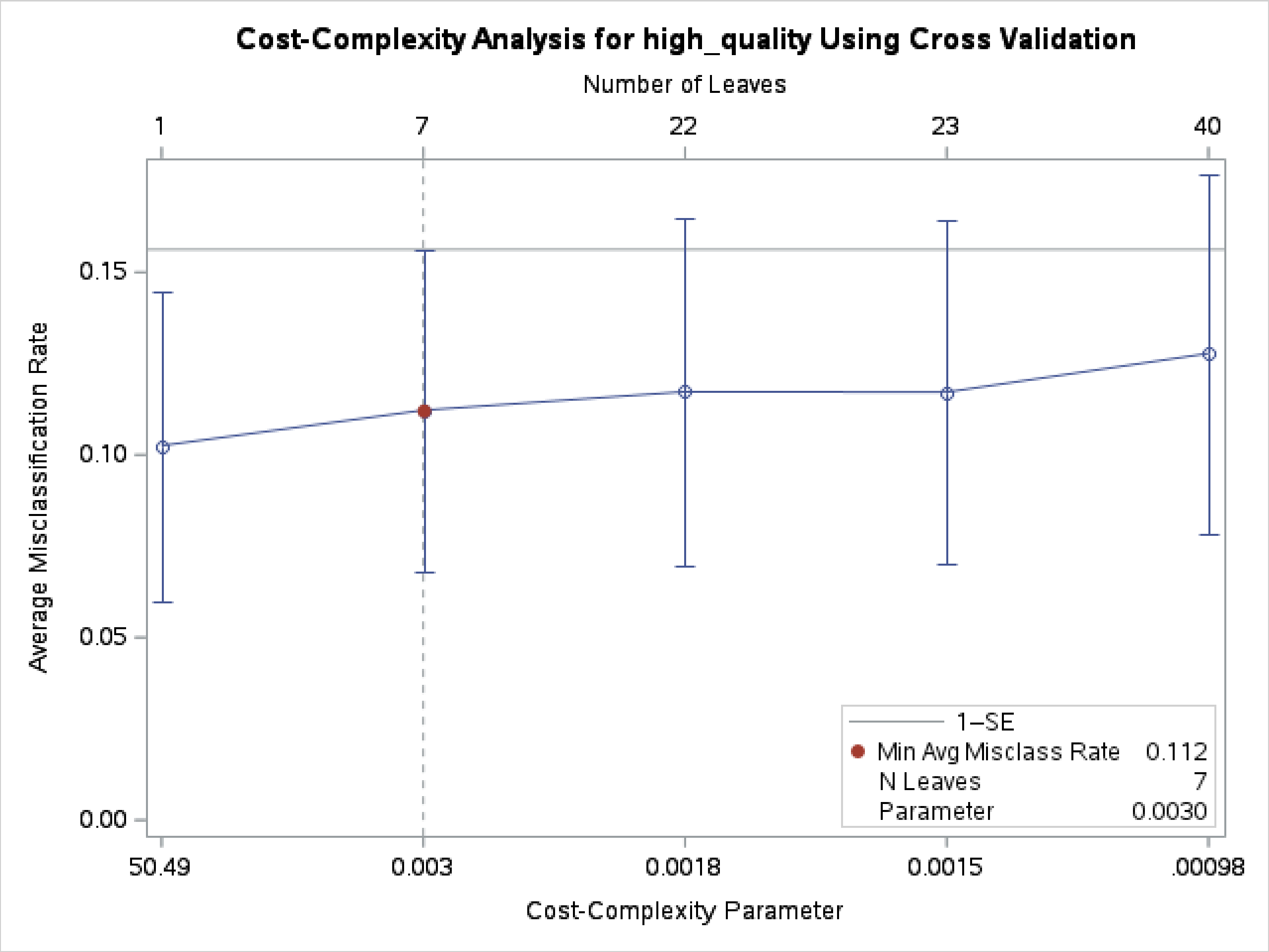
| Data Access Information | | | |
|---|---|---|---|
| **Data** | **Engine** | **Role** | **Path** |
| **WORK.VALIDATION_SET** | V9 | Input | On Client |

| Model Information | |
|---|---|
| **Split Criterion Used** | Entropy |
| **Pruning Method** | Cost-Complexity |
| **Subtree Evaluation Criterion** | Cost-Complexity |
| **Number of Branches** | 2 |
| **Maximum Tree Depth Requested** | 10 |
| **Maximum Tree Depth Achieved** | 10 |
| **Tree Depth** | 6 |
| **Number of Leaves Before Pruning** | 40 |
| **Number of Leaves After Pruning** | 8 |
| **Model Event Level** | 0 |

| | |
|---|---|
| **Number of Observations Read** | 481 |
| **Number of Observations Used** | 481 |

## High Quality Wine
## Classification using validation data set

### The HPSPLIT Procedure



**10-Fold Cross Validation Assessment of Model**

| | Average Square Error | | | | Number of Leaves | | | Misclassification Rate | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| N Leaves | Min | Avg | Standard Error | Max | Min | Median | Max | Min | Avg | Standard Error | Max |
| 8 | 0.0306 | 0.0801 | 0.0325 | 0.1407 | 6 | 8.5 | 12 | 0.0278 | 0.0945 | 0.0391 | 0.1636 |

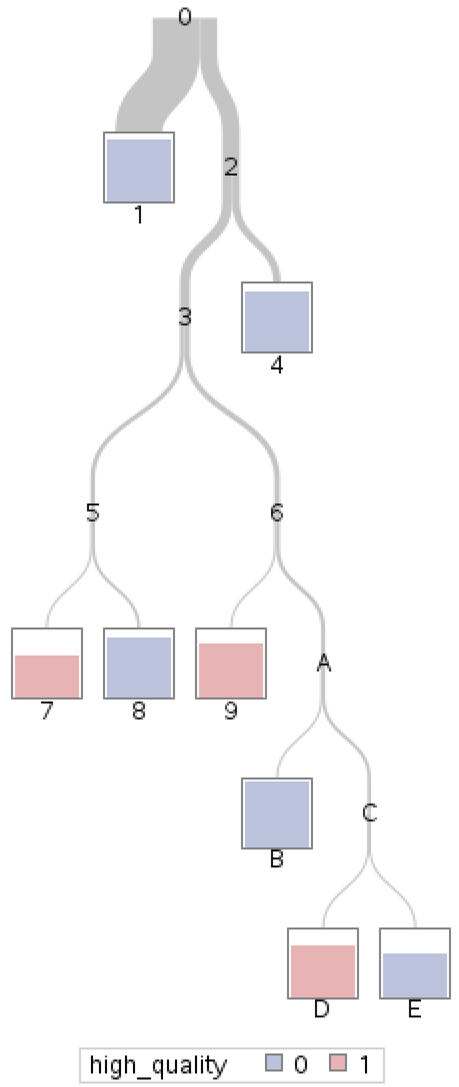
**10-Fold Cross Validation Confusion Matrix**

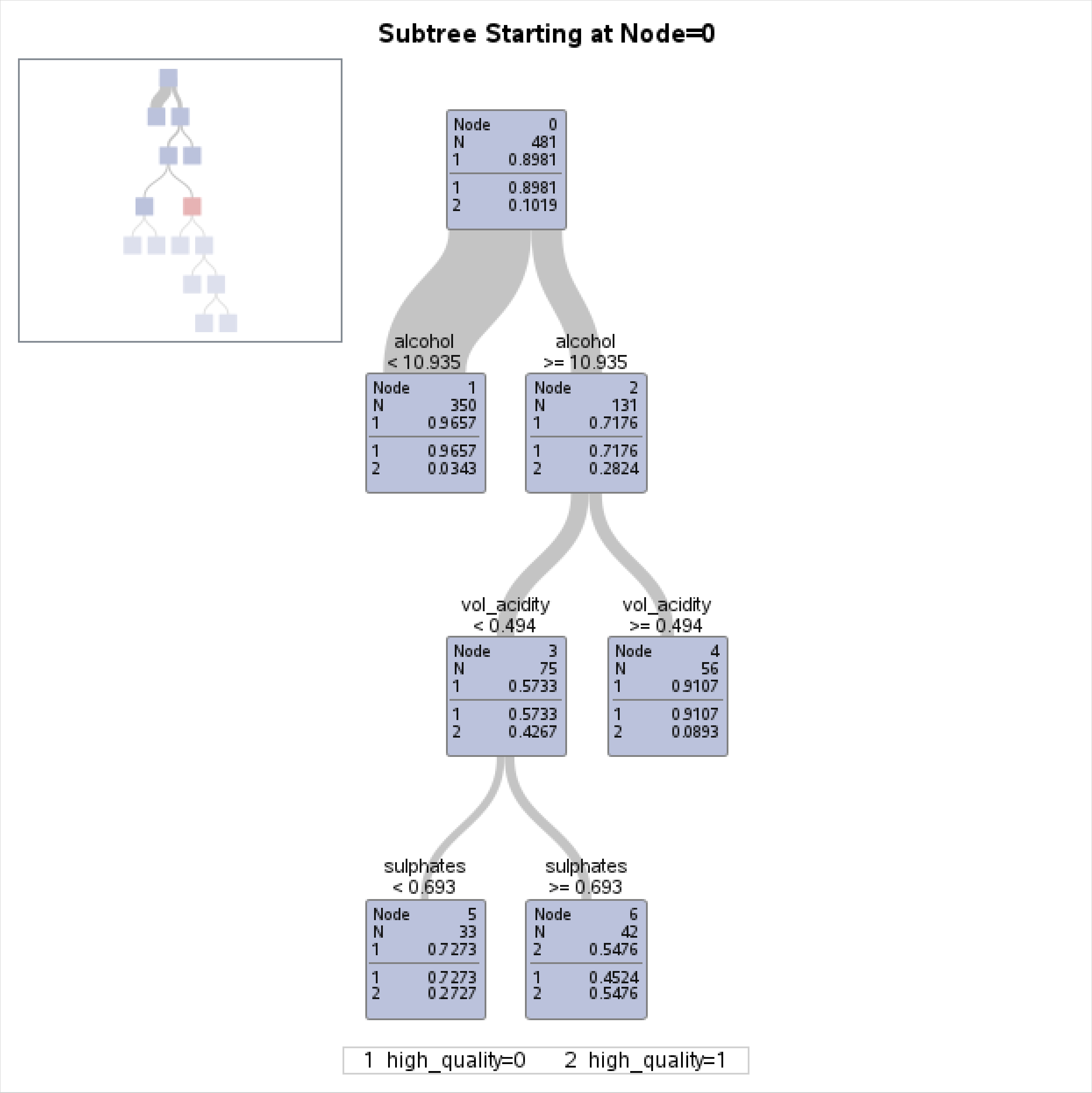
| | Predicted | | |
|---|---|---|---|
| Actual | 0 | 1 | Error Rate |
| 0 | 415 | 17 | 0.0394 |
| 1 | 30 | 19 | 0.6122 |

## High Quality Wine
## Classification using validation data set

### The HPSPLIT Procedure

### Classification Tree for high_quality





## High Quality Wine
## Classification using validation data set

### The HPSPLIT Procedure

**Confusion Matrices**

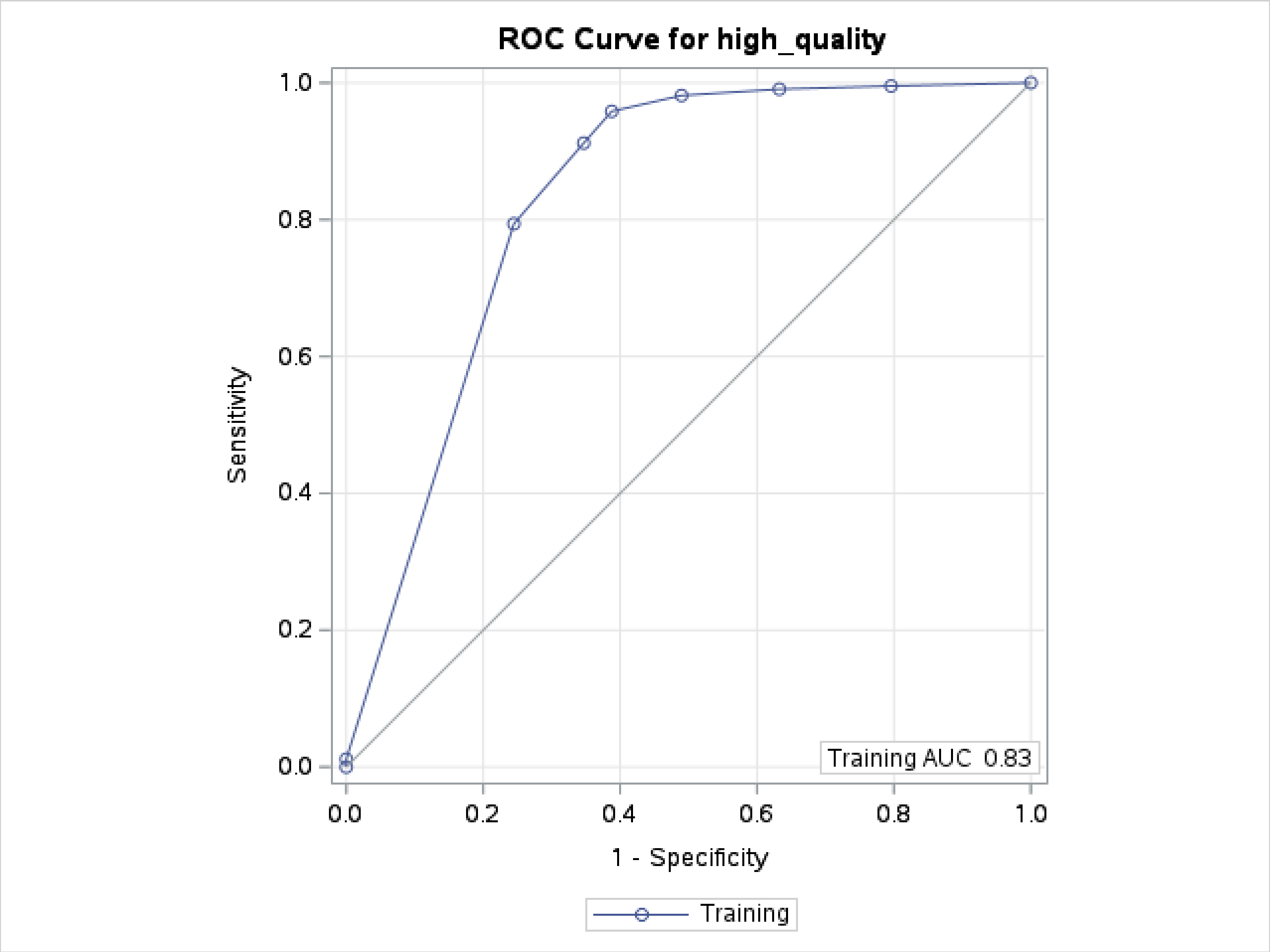
| | | Predicted | | Error |
|---|---|---|---|---|
| | Actual | 0 | 1 | Rate |
| Model Based | 0 | 424 | 8 | 0.0185 |
| | 1 | 24 | 25 | 0.4898 |
| Cross Validation | 0 | 415 | 17 | 0.0394 |
| | 1 | 30 | 19 | 0.6122 |

**Fit Statistics for Selected Tree**

| | N Leaves | ASE | Mis-class | Sensitivity | Specificity | Entropy | Gini | RSS | AUC |
|---|---|---|---|---|---|---|---|---|---|
| Model Based | 8 | 0.0564 | 0.0665 | 0.9815 | 0.5102 | 0.3089 | 0.1127 | 54.2116 | 0.8294 |

Fit Statistics for Selected Tree

|  | N<br>Leaves | ASE | Mis-<br>class | Sensitivity | Specificity | Entropy | Gini | RSS | AUC |
|---|---|---|---|---|---|---|---|---|---|
| Cross Validation | 8 | 0.0801 | 0.0945 | 0.9606 | 0.3878 |  |  |  |  |



Variable Importance

|  | Training |  |  |
|---|---|---|---|
| Variable | Relative | Importance | Count |
| alcohol | 1.0000 | 3.4264 | 1 |
| vol_acidity | 0.9248 | 3.1689 | 2 |
| total_sulfur | 0.8867 | 3.0381 | 3 |
| sulphates | 0.4877 | 1.6712 | 1 |