IS 467 – Knowledge Discovery Technologies

Homework 3

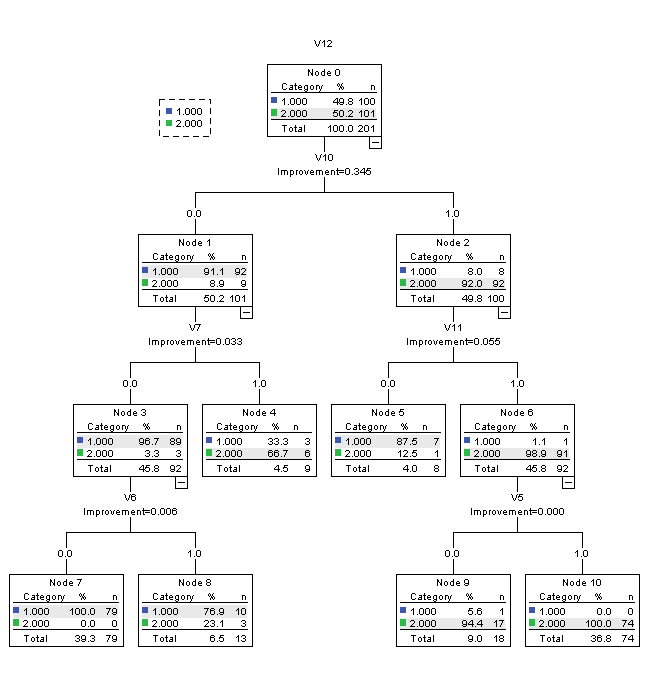
SPSS Print out

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| V1 | 300 | 0 | 1 | .28 | .451 |
| V2 | 300 | 0 | 1 | .09 | .291 |
| V3 | 300 | 0 | 1 | .27 | .446 |
| V4 | 300 | 0 | 1 | .14 | .344 |
| V5 | 300 | 0 | 1 | .69 | .462 |
| V6 | 300 | 0 | 1 | .35 | .478 |
| V7 | 300 | 0 | 1 | .33 | .470 |
| V8 | 300 | 0 | 1 | .10 | .296 |
| V9 | 300 | 0 | 1 | .38 | .485 |
| V10 | 300 | 0 | 1 | .47 | .500 |
| V11 | 300 | 0 | 1 | .73 | .445 |
| V12 | 300 | 1 | 2 | 1.50 | .501 |
| Valid N (listwise) | 300 |  |  |  |  |

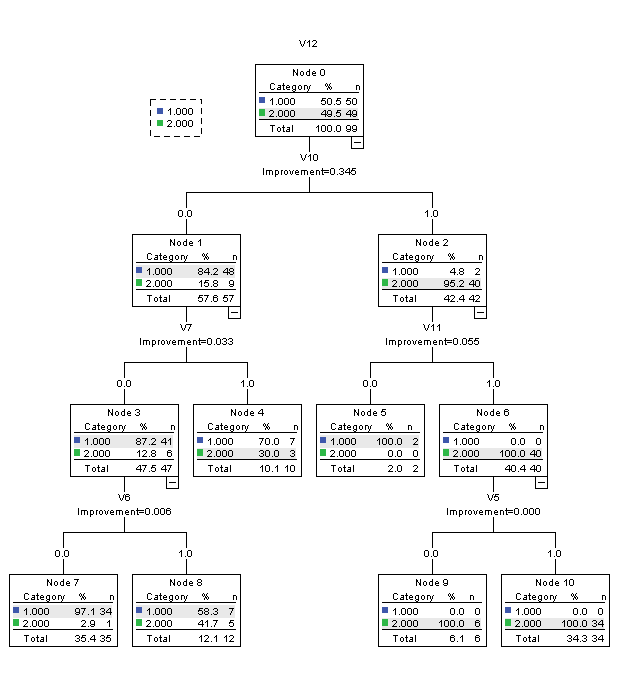
|  |
| --- |
| **Warnings** |
| Gain summary Tables are not displayed because profits are undefined. |
| Target category gains tables are not displayed because target categories are undefined. |

|  |  |  |
| --- | --- | --- |
| **Model Summary** | | |
| Specifications | Growing Method | CRT |
| Dependent Variable | V12 |
| Independent Variables | V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11 |
| Validation | Split Sample |
| Maximum Tree Depth | 20 |
| Minimum Cases in Parent Node | 10 |
| Minimum Cases in Child Node | 5 |
| Results | Independent Variables Included | V10, V7, V6, V1, V9, V11, V3, V5, V4, V8, V2 |
| Number of Nodes | 11 |
| Number of Terminal Nodes | 6 |
| Depth | 3 |

**Training Sample**



**Test Sample**



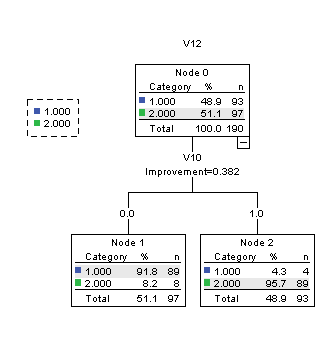
|  |  |  |
| --- | --- | --- |
| **Risk** | | |
| Sample | Estimate | Std. Error |
| Training | .040 | .014 |
| Test | .131 | .034 |
| Growing Method: CRT  Dependent Variable: V12 | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification** | | | | |
| Sample | Observed | Predicted | | |
| 1 | 2 | Percent Correct |
| Training | 1 | 96 | 4 | 96.0% |
| 2 | 4 | 97 | 96.0% |
| Overall Percentage | 49.8% | 50.2% | 96.0% |
| Test | 1 | 43 | 7 | 86.0% |
| 2 | 6 | 43 | 87.8% |
| Overall Percentage | 49.5% | 50.5% | 86.9% |
| Growing Method: CRT  Dependent Variable: V12 | | | | |

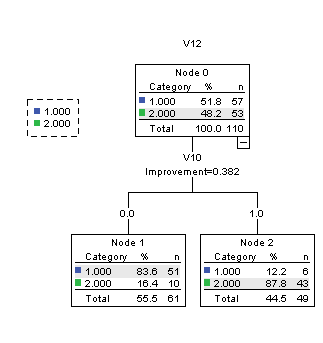
------------------------------------ Increase number of cases for each parent and child -------------------s

|  |  |  |
| --- | --- | --- |
| **Model Summary** | | |
| Specifications | Growing Method | CRT |
| Dependent Variable | V12 |
| Independent Variables | V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11 |
| Validation | Split Sample |
| Maximum Tree Depth | 30 |
| Minimum Cases in Parent Node | 20 |
| Minimum Cases in Child Node | 10 |
| Results | Independent Variables Included | V10, V7, V9, V11, V1, V6, V3, V8, V5, V2, V4 |
| Number of Nodes | 3 |
| Number of Terminal Nodes | 2 |
| Depth | 1 |

**Training Sample**



**Test Sample**



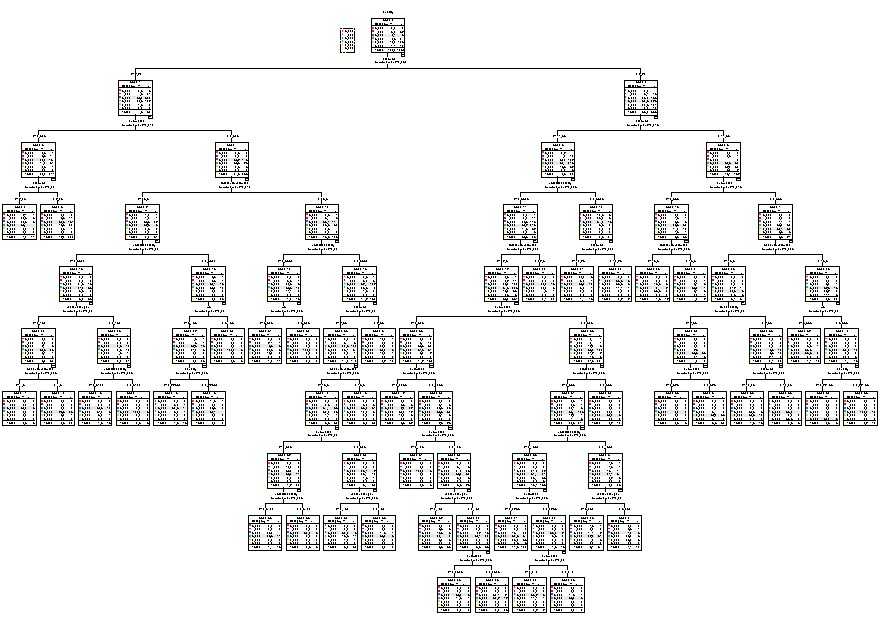
|  |  |  |
| --- | --- | --- |
| **Risk** | | |
| Sample | Estimate | Std. Error |
| Training | .063 | .018 |
| Test | .145 | .034 |
| Growing Method: CRT  Dependent Variable: V12 | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification** | | | | |
| Sample | Observed | Predicted | | |
| 1 | 2 | Percent Correct |
| Training | 1 | 89 | 4 | 95.7% |
| 2 | 8 | 89 | 91.8% |
| Overall Percentage | 51.1% | 48.9% | 93.7% |
| Test | 1 | 51 | 6 | 89.5% |
| 2 | 10 | 43 | 81.1% |
| Overall Percentage | 55.5% | 44.5% | 85.5% |
| Growing Method: CRT  Dependent Variable: V12 | | | | |

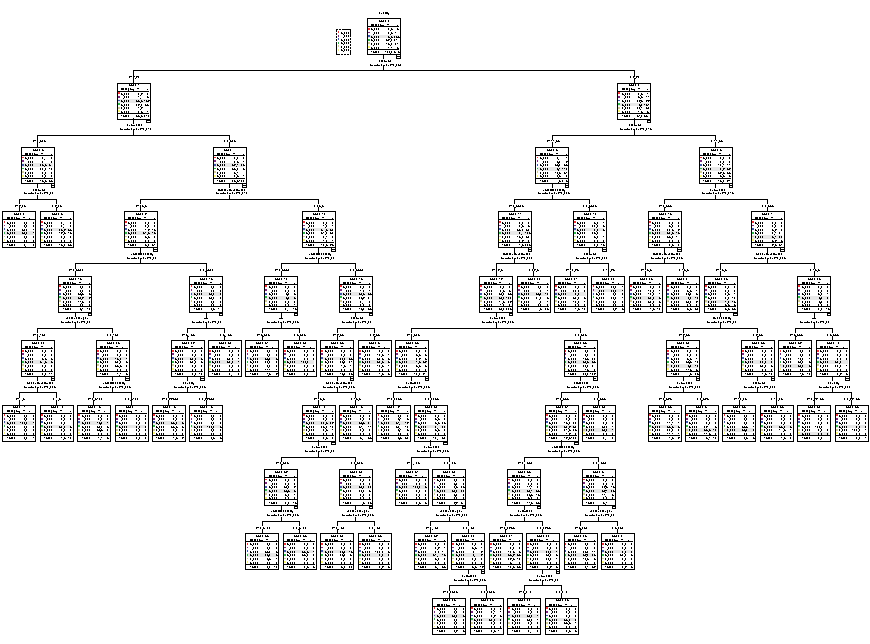
------------------------- Question 2 --------------------------- Red wine

|  |  |  |
| --- | --- | --- |
| **Model Summary** | | |
| Specifications | Growing Method | CRT |
| Dependent Variable | quality |
| Independent Variables | fixedacidity, volatileacidity, citricacid, residualsugar, chlorides, freesulfurdioxide, totalsulfurdioxide, density, pH, sulphates, alcohol |
| Validation | Split Sample |
| Maximum Tree Depth | 20 |
| Minimum Cases in Parent Node | 10 |
| Minimum Cases in Child Node | 5 |
| Results | Independent Variables Included | alcohol, density, totalsulfurdioxide, sulphates, volatileacidity, chlorides, citricacid, pH, residualsugar, freesulfurdioxide, fixedacidity |
| Number of Nodes | 79 |
| Number of Terminal Nodes | 40 |
| Depth | 9 |

**Training Sample**



**Test Sample**



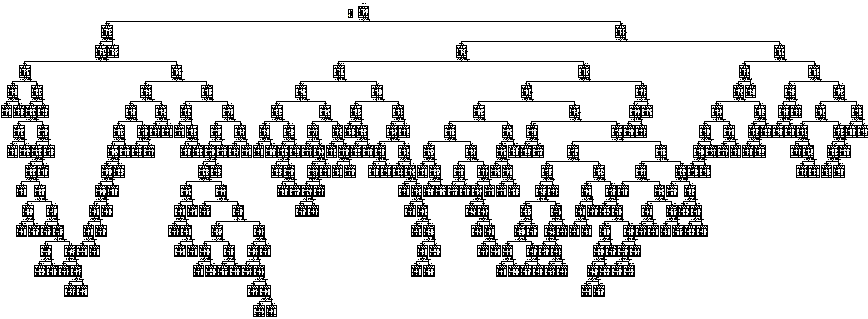
|  |  |  |
| --- | --- | --- |
| **Risk** | | |
| Sample | Estimate | Std. Error |
| Training | .276 | .014 |
| Test | .421 | .021 |
| Growing Method: CRT  Dependent Variable: quality | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Classification** | | | | | | | | |
| Sample | Observed | Predicted | | | | | | |
| 3 | 4 | 5 | 6 | 7 | 8 | Percent Correct |
| Training | 3 | 0 | 0 | 3 | 4 | 0 | 0 | 0.0% |
| 4 | 0 | 5 | 25 | 9 | 0 | 0 | 12.8% |
| 5 | 0 | 3 | 374 | 65 | 1 | 0 | 84.4% |
| 6 | 0 | 1 | 85 | 322 | 14 | 2 | 75.9% |
| 7 | 0 | 0 | 15 | 55 | 58 | 0 | 45.3% |
| 8 | 0 | 0 | 0 | 7 | 2 | 3 | 25.0% |
| Overall Percentage | 0.0% | 0.9% | 47.7% | 43.9% | 7.1% | 0.5% | 72.4% |
| Test | 3 | 0 | 1 | 2 | 0 | 0 | 0 | 0.0% |
| 4 | 0 | 1 | 8 | 5 | 0 | 0 | 7.1% |
| 5 | 0 | 2 | 166 | 69 | 1 | 0 | 69.7% |
| 6 | 0 | 1 | 69 | 130 | 10 | 4 | 60.7% |
| 7 | 0 | 0 | 7 | 45 | 19 | 0 | 26.8% |
| 8 | 0 | 0 | 1 | 3 | 2 | 0 | 0.0% |
| Overall Percentage | 0.0% | 0.9% | 46.3% | 46.2% | 5.9% | 0.7% | 57.9% |
| Growing Method: CRT  Dependent Variable: quality | | | | | | | | |

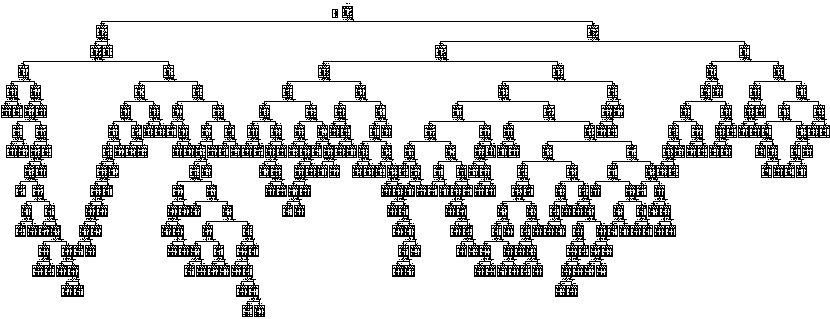
------- decrease bin size for each -----------------

|  |  |  |
| --- | --- | --- |
| **Model Summary** | | |
| Specifications | Growing Method | CRT |
| Dependent Variable | quality |
| Independent Variables | fixedacidity, volatileacidity, citricacid, residualsugar, chlorides, freesulfurdioxide, totalsulfurdioxide, density, pH, sulphates, alcohol |
| Validation | Split Sample |
| Maximum Tree Depth | 20 |
| Minimum Cases in Parent Node | 5 |
| Minimum Cases in Child Node | 2 |
| Results | Independent Variables Included | alcohol, totalsulfurdioxide, density, sulphates, chlorides, volatileacidity, pH, freesulfurdioxide, residualsugar, citricacid, fixedacidity |
| Number of Nodes | 289 |
| Number of Terminal Nodes | 145 |
| Depth | 15 |

**Training Sample**



**Test Sample**



|  |  |  |
| --- | --- | --- |
| **Risk** | | |
| Sample | Estimate | Std. Error |
| Training | .118 | .010 |
| Test | .389 | .022 |
| Growing Method: CRT  Dependent Variable: quality | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Classification** | | | | | | | | |
| Sample | Observed | Predicted | | | | | | |
| 3 | 4 | 5 | 6 | 7 | 8 | Percent Correct |
| Training | 3 | 2 | 2 | 1 | 3 | 0 | 0 | 25.0% |
| 4 | 1 | 7 | 11 | 9 | 3 | 0 | 22.6% |
| 5 | 0 | 0 | 449 | 23 | 6 | 0 | 93.9% |
| 6 | 1 | 1 | 26 | 388 | 7 | 0 | 91.7% |
| 7 | 0 | 0 | 4 | 23 | 116 | 0 | 81.1% |
| 8 | 0 | 0 | 2 | 3 | 3 | 4 | 33.3% |
| Overall Percentage | 0.4% | 0.9% | 45.0% | 41.0% | 12.3% | 0.4% | 88.2% |
| Test | 3 | 0 | 0 | 1 | 0 | 1 | 0 | 0.0% |
| 4 | 2 | 2 | 11 | 4 | 3 | 0 | 9.1% |
| 5 | 2 | 2 | 145 | 48 | 6 | 0 | 71.4% |
| 6 | 0 | 0 | 59 | 131 | 24 | 1 | 60.9% |
| 7 | 0 | 0 | 7 | 17 | 30 | 2 | 53.6% |
| 8 | 0 | 0 | 0 | 3 | 3 | 0 | 0.0% |
| Overall Percentage | 0.8% | 0.8% | 44.2% | 40.3% | 13.3% | 0.6% | 61.1% |
| Growing Method: CRT  Dependent Variable: quality | | | | | | | | |