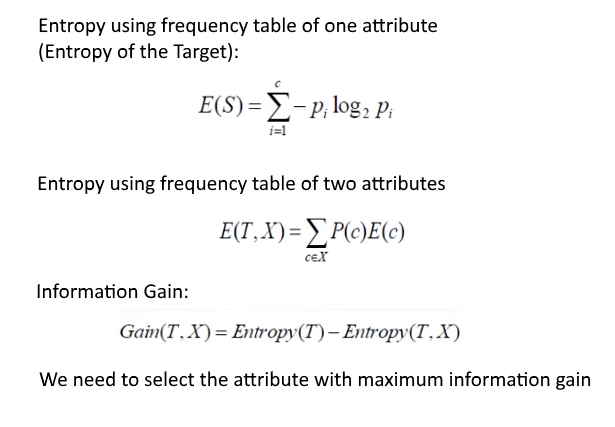
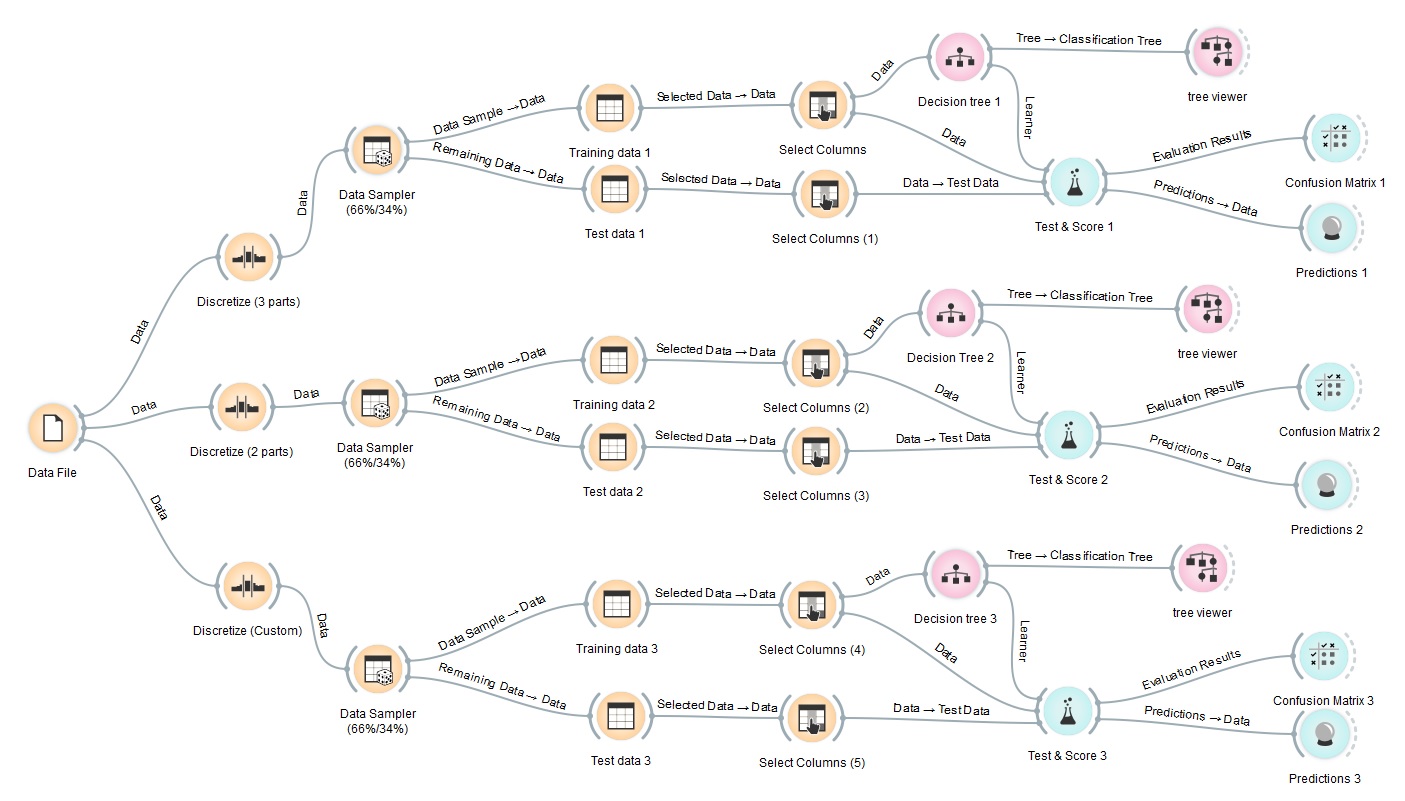
The following formulas were used to calculate the entropy and gain. The attribute with the max gain is considered as the node to split.



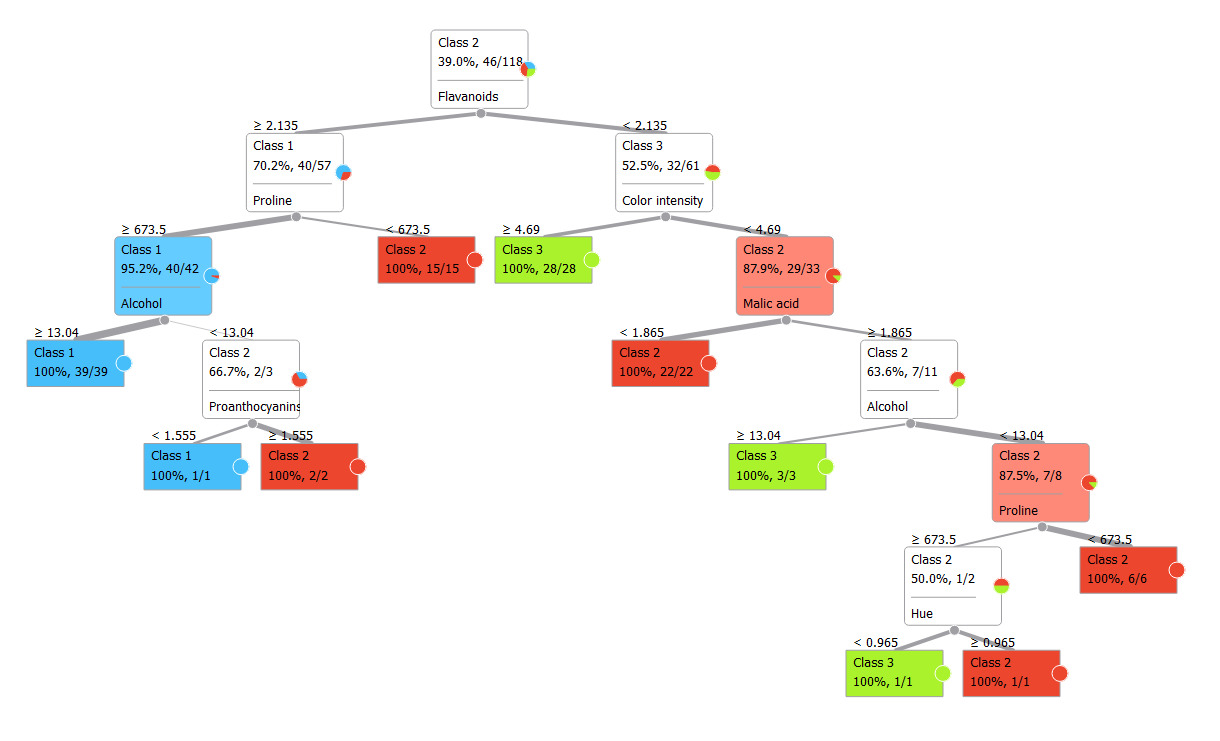
Orange Worksheet screenshot:

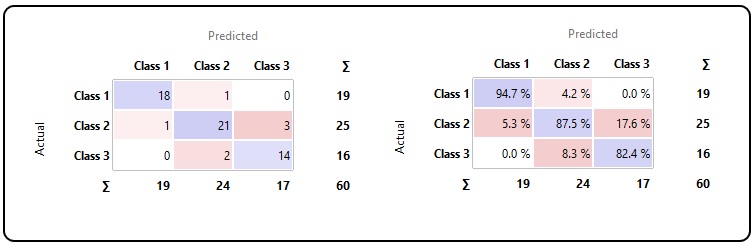


1. Test 2 :

|  |  |  |
| --- | --- | --- |
| 1) Alcohol | < 13.04 | ≥ 13.04 |
| 2) Malic acid | < 1.865 | ≥ 1.865 |
| 3) Ash | < 2.355 | ≥ 2.355 |
| 4) Alkalinity of ash | < 19.45 | ≥ 19.45 |
| 5) Magnesium | < 97.5 | < 97.5 |
| 6) Total phenols | < 2.355 | ≥ 2.355 |
| 7) Flavonoids | < 2.135 | ≥ 2.135 |
| 8) Non Flavonoid phenols | < 0.335 | ≥ 0.335 |
| 9) Proanthocyanidins | < 1.555 | ≥ 1.555 |
| 10)Color intensity | < 4.69 | ≥ 4.69 |
| 11)Hue | < 0.965 | ≥ 0.965 |
| 12)OD280/OD315 of diluted wines | < 2.775 | ≥ 2.775 |
| 13)Proline | < 673.5 | ≥ 673.5 |

1. Screenshot of the decision tree:



1. Confusion Matrix:
2. 

Initially we calculate entropy and gain for all the attributes and select the node with the max gain.

Level 1: we select first node Flavonoids as it has the max gain

|  |  |  |  |
| --- | --- | --- | --- |
|  | E(T) | E(T,X) | Gain |
| 1) Alcohol | 1.56941 | 1.05489 | 0.51452 |
| 2) Malic acid | 1.56941 | 1.37114 | 0.19828 |
| 3) Ash | 1.56941 | 1.52547 | 0.04395 |
| 4) Alkalinity of ash | 1.56941 | 1.34844 | 0.22097 |
| 5) Magnesium | 1.56941 | 1.42359 | 0.14583 |
| 6) Total phenols | 1.56941 | 1.06623 | 0.50318 |
| 7) Flavonoids | 1.56941 | 0.94071 | 0.62870 |
| 8) Non Flavonoid phenols | 1.56941 | 1.39462 | 0.17480 |
| 9) Proanthocyanidins | 1.56941 | 1.30253 | 0.26689 |
| 10)Color intensity | 1.56941 | 1.02127 | 0.54814 |
| 11)Hue | 1.56941 | 1.19282 | 0.37659 |
| 12)OD280/OD315 of diluted wines | 1.56941 | 1.08616 | 0.48326 |
| 13)Proline | 1.56941 | 1.03204 | 0.53737 |

Level 2: flavonoids < 2.135 : select Color intensity as it has max gain

|  |  |  |  |
| --- | --- | --- | --- |
|  | E(T) | E(T,X) | Gain |
| 1) Alcohol | 0.99826 | 0.77113 | 0.22713 |
| 2) Malic acid | 0.99826 | 0.63637 | 0.36189 |
| 3) Ash | 0.99826 | 0.96360 | 0.03466 |
| 4) Alkalinity of ash | 0.99826 | 0.91838 | 0.07988 |
| 5) Magnesium | 0.99826 | 0.96941 | 0.02884 |
| 6) Total phenols | 0.99826 | 0.95501 | 0.04324 |
| 7) Flavonoids | 0.99826 | 0.94071 | 0.05754 |
| 8) Non Flavonoid phenols | 0.99826 | 0.99031 | 0.00795 |
| 9) Proanthocyanidins | 0.99826 | 0.96914 | 0.02911 |
| 10)Color intensity | 0.99826 | 0.28826 | 0.71000 |
| 11)Hue | 0.99826 | 0.43409 | 0.56417 |
| 12)OD280/OD315 of diluted wines | 0.99826 | 0.72259 | 0.27566 |
| 13)Proline | 0.99826 | 0.96632 | 0.03194 |

Level 2: flavonoids >= 2.135 : select Proline as it has the max gain

|  |  |  |  |
| --- | --- | --- | --- |
|  | E(T) | E(T,X) | Gain |
| 1) Alcohol | 0.87914 | 0.21462 | 0.66452 |
| 2) Malic acid | 0.87914 | 0.85916 | 0.01998 |
| 3) Ash | 0.87914 | 0.84419 | 0.03494 |
| 4) Alkalinity of ash | 0.87914 | 0.80412 | 0.07502 |
| 5) Magnesium | 0.87914 | 0.69153 | 0.18760 |
| 6) Total phenols | 0.87914 | 0.85587 | 0.02326 |
| 7) Flavonoids | 0.87914 | 0.94071 | -0.06158 |
| 8) Non Flavonoid phenols | 0.87914 | 0.83729 | 0.04185 |
| 9) Proanthocyanidins | 0.87914 | 0.86679 | 0.01234 |
| 10)Color intensity | 0.87914 | 0.50007 | 0.37906 |
| 11)Hue | 0.87914 | 0.87911 | 0.00002 |
| 12)OD280/OD315 of diluted wines | 0.87914 | 0.86388 | 0.01526 |
| 13)Proline | 0.87914 | 0.20351 | 0.67562 |

Level 3 flavonoids >= 2.135 , Proline >=673.5

Select Alcohol as it has max gain.

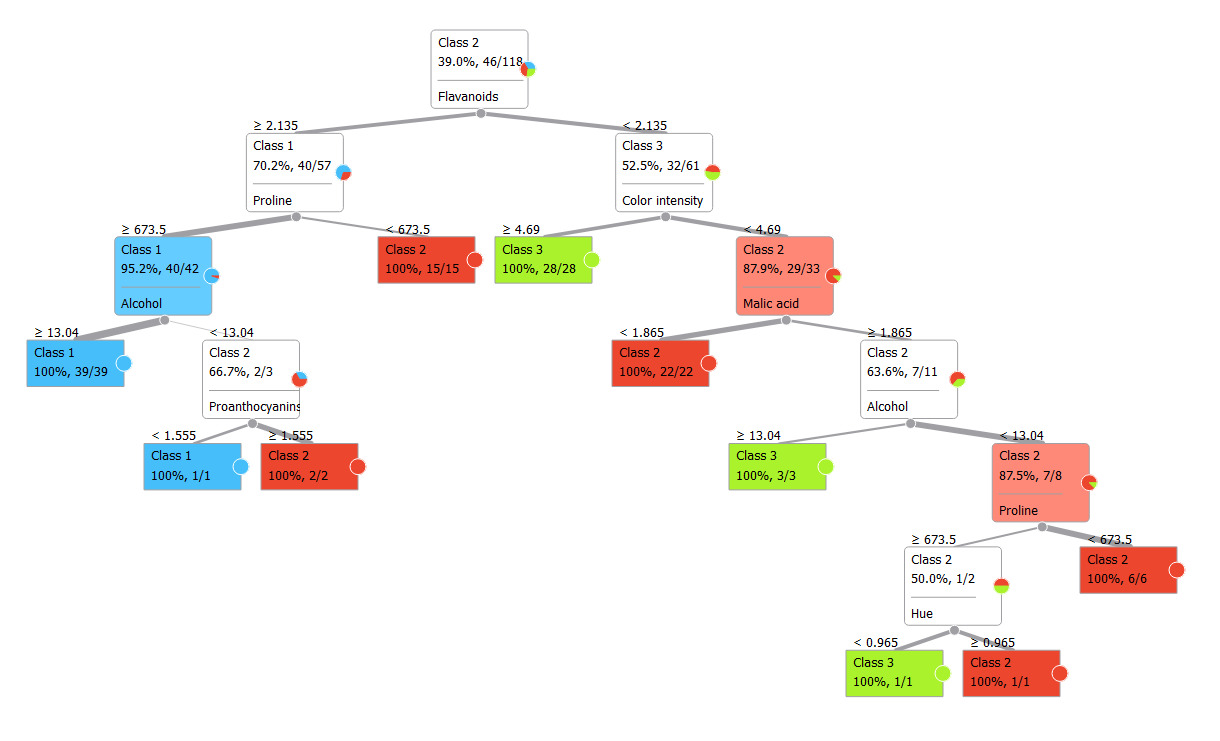
|  |  |  |  |
| --- | --- | --- | --- |
|  | E(T) | E(T,X) | Gain |
| 1) Alcohol | 0.27620 | 0.06559 | 0.21060 |
| 2) Malic acid | 0.27620 | 0.24749 | 0.02871 |
| 3) Ash | 0.27620 | 0.27312 | 0.00308 |
| 4) Alkalinity of ash | 0.27620 | 0.26128 | 0.01492 |
| 5) Magnesium | 0.27620 | 0.26452 | 0.01167 |
| 6) Total phenols | 0.27620 | 0.20825 | 0.06795 |
| 7) Flavonoids | 0.27620 | 0.94071 | -0.66452 |
| 8) Non Flavonoid phenols | 0.27620 | 0.26452 | 0.01167 |
| 9) Proanthocyanidins | 0.27620 | 0.26333 | 0.01286 |
| 10)Color intensity | 0.27620 | 0.17189 | 0.10431 |
| 11)Hue | 0.27620 | 0.25240 | 0.02380 |
| 12)OD280/OD315 of diluted wines | 0.27620 | 0.23611 | 0.04009 |
| 13)Proline | 0.27620 | 0.20351 | 0.07268 |

Level 3 flavonoids >= 2.135 , Proline >=673.5, Alcohol < 1.555

Select Proanthocyanidins as it has the max gain.

|  |  |  |  |
| --- | --- | --- | --- |
|  | E(T) | E(T,X) | Gain |
| 1) Alcohol | 0.91830 | 0.06559 | 0.85270 |
| 2) Malic acid | 0.91830 | 0.24749 | 0.67081 |
| 3) Ash | 0.91830 | 0.66667 | 0.25163 |
| 4) Alkalinity of ash | 0.91830 | 0.26128 | 0.65702 |
| 5) Magnesium | 0.91830 | 0.66667 | 0.25163 |
| 6) Total phenols | 0.91830 | 0.66667 | 0.25163 |
| 7) Flavonoids | 0.91830 | 0.94071 | -0.02242 |
| 8) Non Flavonoid phenols | 0.91830 | 0.66667 | 0.25163 |
| 9) Proanthocyanidins | 0.91830 | 0.00000 | 0.91830 |
| 10)Color intensity | 0.91830 | 0.17189 | 0.74641 |
| 11)Hue | 0.91830 | 0.25240 | 0.66590 |
| 12)OD280/OD315 of diluted wines | 0.91830 | 0.66667 | 0.25163 |
| 13)Proline | 0.91830 | 0.20351 | 0.71478 |

This process continues till we are left over with only leaf nodes and we get our decision tree.



Instances table for all the three tests:

