Name: Pratik Jasani

Course: EN 605.645

**Decision Tree Self Check - ID3 Algorithm**

Suppose we have the following training data where Shape, Size and Color are the features (attributes) and Safe? is the class label:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Shape** | **Size** | **Color** | **Safe?** |
| 1 | Round | Large | Blue | No |
| 2 | Square | Large | Green | Yes |
| 3 | Square | Small | Red | No |
| 4 | Round | Large | Red | Yes |
| 5 | Square | Small | Blue | No |
| 6 | Round | Small | Blue | No |
| 7 | Round | Small | Red | Yes |
| 8 | Square | Small | Green | No |
| 9 | Round | Large | Green | Yes |
| 10 | Square | Large | Green | Yes |
| 11 | Square | Large | Red | No |
| 12 | Square | Large | Green | Yes |
| 13 | Round | Large | Red | Yes |
| 14 | Square | Small | Red | No |
| 15 | Round | Small | Green | No |

Initial entropy: -(8/15)log2(8/15) – (7/15)log2(7/15) = 0.996791632 (8 = No, 7 = Yes, 15 = Total)

Pick the best attribute(information gain for each attribute and choose the once that lowers the entropy the most):

**Shape = Round:** (No=3, Yes=4, Total=7)=>-(3/7)log2(3/7) – (4/7)log2(4/7) = 0.985228

**Shape = Square:** (No=5, Yes=3, Total=8) => -(5/8)log2(5/8) – (3/8)log2(3/8) = 0.954434

**Shape:** (Round total= 7, Square total=8, total = 15) =>(7/15)\* 0.985228 +(8/15)\* 0.954434 **= 0.968805**

**Size = Large:** (No=2, Yes=6, Total=8) => -(2/8)log2(2/8) – (6/8)log2(6/8) = 0.811278

**Size = Small:** (No=6, Yes=1, Total=7)=> -(6/7)log2(6/7) – (1/7)log2(1/7) = 0.591673

**Size:** (Large total = 8, Small total = 7, total = 15) => (8/15)\* 0.811278 + (7/15)\*0.591673 = **0.708796**

**Color = Red :** (No=3, Yes=3, Total=6) => -(3/6)log2(3/6) – (3/6)log2(3/6) = 1

**Color = Green:** (No=2, Yes=4, Total=6) => -(2/6)log2(2/6) – (4/6)log2(4/6) = 0.918296

**Color = Blue:** (No=3, Yes=0, Total=3) => -(3/3)log2(3/3) – (0/3)log2(0/3) = 0

**Color :** (B total=3, R total=6, G total=6, total=15) => **0.767318**

Now we know the best attribute is **Shape** which gives us (initial entropy – shape entropy) =

0.996791632 - 0.968805 = 0.027986632, so we have the following tree now.

Shape

Square

Round

Now we choose the next best attribute the same logic as above.

**Shape = Square, Size = Large:** (No=1, Yes=3, Total=4) => 0.811278

**Shape = Square, Size = Small:** (No=4, Yes=0, Total=4) => 0

**Shape = Square, Size:** (Large total = 4, small total = 4, total = 8) = (4/8)\* 0.811278 = 0.405639

**Shape = Square, Color = Red:** (No=3, Yes=0, Total=3) => 0

**Shape = Square, Color = Green:** (No=1, Yes=3, Total=4) => 0.811278

**Shape = Square, Color = Blue**: (No=1, Yes=0, Total=1) => 0

**Shape = Square, Color:** (Red total=3, green total=4, blue total = 1) => = (4/8)\* 0.811278 = 0.405639

Since both have the same entropy, we can choose one randomly. Now we have,

Shape

Round

Square

Size

Large

Small

No

**Shape = Square, Size = Large, Color = Red:** (No=1, Yes=0, Total=1) => 0

**Shape = Square, Size = Large, Color = Green:** (No=0, Yes=3, Total=3) => 0

**Shape = Square, Size = Large, Color = Blue**: (No=0, Yes=0, Total=1) => 0

Shape

Round

Square

Size

Large

Small

Green

Red

Yes

No

Color

No

**Shape = Round, Size = Large:** (No=1, Yes=3, Total=4) => 0.811278

**Shape = Round, Size = Small:** (No=2, Yes=1, Total=3) => 0.918296

**Shape = Round, Size:** (Large total = 4, small total = 4, total = 8) = (4/8)\* 0.811278 = 0.4

**Shape = Round, Color = Red:** (No=0, Yes=1, Total=3) => 0

**Shape = Round, Color = Green:** (No=1, Yes=1, Total=2) => 1

**Shape = Round, Color = Blue**: (No=1, Yes=0, Total=1) => 0

**Shape = Round, Color:** (Red total=3, green total=2, blue total = 1) => (2/7)\* 1 = 0.115897

Shape

Round

Square

Size

Size

Small

Large

Large

Small

Green

Red

Yes

No

Color

No

**Shape = Round, Size = Small, Color = Red:** (No=0, Yes=1, Total=1) => 0

**Shape = Round, Size = Small, Color = Green:** (No=1, Yes=0, Total=1) => 0

**Shape = Round, Size = Small, Color = Blue**: (No=1, Yes=0, Total=1) => 0

**Shape = Round, Size = Large, Color = Red:** (No=0, Yes=2, Total=2) => 0

**Shape = Round, Size = Large, Color = Green:** (No=0, Yes=1, Total=1) => 0

**Shape = Round, Size = Large, Color = Blue**: (No=1, Yes=0, Total=1) => 0

The final decision tree:

Shape

Round

Square

Size

Size

Small

Large

Large

Small

Color

Green

Red

No

Color

No

Red

Yes

Green

No

Blue

Yes

Yes

Red

Color

Blue

No

Green

No

Yes