ID3 Decision Tree Testing

Expected Output Tests

Test Case 1

Input data:

10	a	x	red
10	a	y	red
10	b	x	blue
20	b	y	red
20	b	z	green
30	a	z	green

Test:

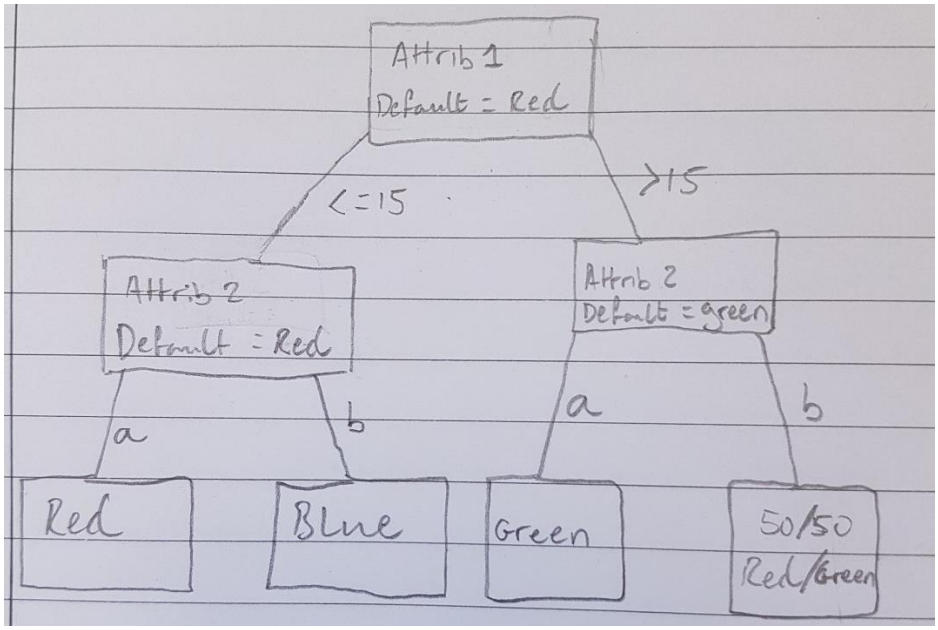
Create decision tree from input data – Non-Anonymised data with the fourth attribute as the class, using attributes 1 & 2 for classification

Pass Conditions:

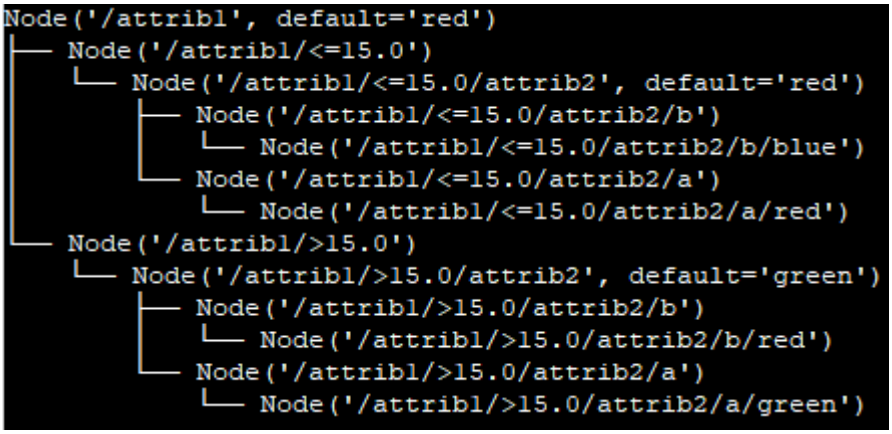
- Decision tree created that matches expected output

Expected output:

Random elements involved – expected output shows equal chance between two classes with a '/'



Actual output:


Node('/attrib1', default='red')
Node('/attrib1/<=15.0')
Node('/attrib1/<=15.0/attrib2', default='red')
Node('/attrib1/<=15.0/attrib2/b')
Node('/attrib1/<=15.0/attrib2/b/blue')
Node('/attrib1/<=15.0/attrib2/a')
Node('/attrib1/<=15.0/attrib2/a/red')
Node('/attrib1/>15.0')
Node('/attrib1/>15.0/attrib2', default='green')
Node('/attrib1/>15.0/attrib2/b')
Node('/attrib1/>15.0/attrib2/b/red')
Node('/attrib1/>15.0/attrib2/a')
Node('/attrib1/>15.0/attrib2/a/green')

Passed: Yes

Test Case 2

Input data:

10	a	x	red
10	a	y	red
10	b	x	blue
20	b	y	green
20	b	z	green
30	a	z	green

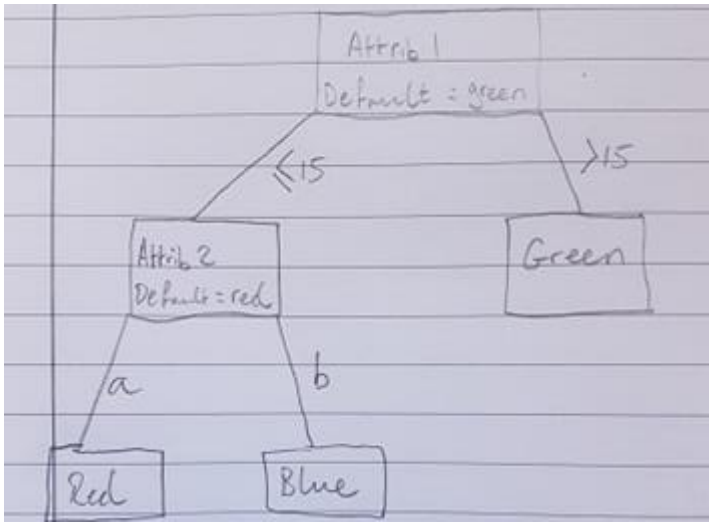
Test:

Create decision tree from input data – Non-Anonymised data with the fourth attribute as the class, using attributes 1 & 2 for classification. Note similarity to Test Case 1, with change in class of one record – should result in a trimmed tree

Pass Conditions:

- Decision tree created that matches expected output

Expected output:



Actual output:

```
Node('/attrib1', default='green')
├── Node('/attrib1/<=15.0')
│   └── Node('/attrib1/<=15.0/attrib2', default='red')
│       ├── Node('/attrib1/<=15.0/attrib2/a')
│       │   └── Node('/attrib1/<=15.0/attrib2/a/red')
│       └── Node('/attrib1/<=15.0/attrib2/b')
│           └── Node('/attrib1/<=15.0/attrib2/b/blue')
└── Node('/attrib1/>15.0')
    └── Node('/attrib1/>15.0/green')
```

Passed: Yes

Test Case 3

Input data:

10	['a', 'b']	x	blue
10	['a', 'b']	x	red
10	['a', 'b']	y	red
[20, 30]	['a', 'b']	y	red
[20, 30]	['a', 'b']	z	green
[20, 30]	['a', 'b']	z	green

Test:

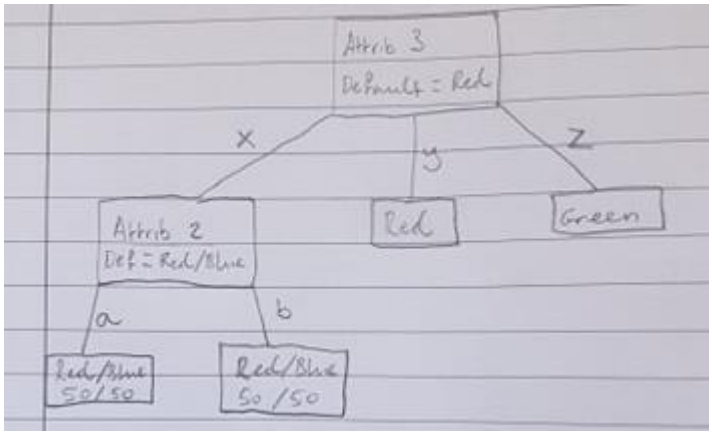
Create decision tree from input data –Anonymised data with the fourth attribute as the class, using attributes 2 & 3 for classification. Note attribute 1 is anonymised but should be ignored, attribute 3 is specified and should be dealt with differently to attribute 2

Pass Conditions:

- Decision tree created that matches expected output

Expected output:

Random elements involved – expected output shows equal chance between two classes with a '/'



Actual output:

```
Node('/attrib3', default='red')
├── Node('/attrib3/y')
│   └── Node('/attrib3/y/red')
├── Node('/attrib3/x')
│   └── Node('/attrib3/x/attrib2', default='red')
│       ├── Node('/attrib3/x/attrib2/b')
│       │   └── Node('/attrib3/x/attrib2/b/red')
│       └── Node('/attrib3/x/attrib2/a')
│           └── Node('/attrib3/x/attrib2/a/blue')
└── Node('/attrib3/z')
    └── Node('/attrib3/z/green')
```

Passed: Yes

Test Case 4

Input data:

10	a	x	green
10	a	y	red
10	b	x	blue
20	b	y	green
20	b	z	green
30	a	z	blue
40	a	x	blue
20	c	y	red

Test:

Classify input data using decision tree from Test Case 2. Note – this data set contains previously unseen values, however it should default to attribute default classification

Pass Conditions:

- Actual classification accuracy matches expected accuracy

Expected accuracy:

50%

Actual accuracy:

```
TOTAL:  8
CORRECT:  4
INCORRECT:  4
UNCLASSIFIED:  0
% ACCURACY:  50.0
```

Passed: Yes

Test Case 5

Input data:

10	a	x	green
10	a	y	red
10	b	x	blue
20	b	y	green
20	b	z	green
30	a	z	blue
40	a	x	blue
20	c	y	red

Test:

Classify input data using decision tree from Test Case 1. Note – this data set contains previously unseen values, however it should default to attribute default classification

Pass Conditions:

- Actual classification accuracy matches expected accuracy

Expected accuracy:

25%

Actual accuracy:

25%

Passed: Yes

Test Case 6

Input data:

10	a	x	green
10	[‘a’,’b’]	y	red
10	b	x	blue
20	b	y	green
20	b	z	green
30	a	z	blue
40	a	x	blue
20	c	y	red

Test:

Classify input data using decision tree from Test Case 1. Note – the anonymised value in record two will be mapped back at random. Therefore, there will be two different potential classification accuracies.

Pass Conditions:

- Actual classification accuracy matches one of the expected classification accuracies
- Repeat tests show one of two accuracies depending on mapping back of attribute 2 in record 2

Expected accuracies:

If [‘a’,’b’] maps to ‘a’: 25%

If [‘a’,’b’] maps to ‘b’: 12.5%

Output (two cases):

```
TOTAL:  8
CORRECT:  2
INCORRECT:  6
UNCLASSIFIED:  0
% ACCURACY:  25.0
```

```
TOTAL:  8
CORRECT:  1
INCORRECT:  7
UNCLASSIFIED:  0
% ACCURACY:  12.5
```

Passed: Yes