Classification Trees

$$gini = \sum_i f_i(1-f_i)$$

A	B
35	65

$f_1 = 0.35 \quad f_2 = 0.65$

$$gini = f_1(1-f_1) + f_2(1-f_2)$$
$$= 0.35(0.65) + 0.65(0.35)$$
$$= 0.455$$

A	B
25	75

$f_1 = 0.25 \quad f_2 = 0.75$

$$gini = 0.375$$

A	B
10	90

$gini = 0.18$

$f_1 = \dfrac{47}{92}$

$f_2 = \dfrac{45}{92}$

R9 <= 1.925
gini = 0.5
samples = 92
value = [47, 45]

$$gini = f_1(1-f_1) + f_2(1-f_2)$$

— no. of observations

Count: 0 1

R24 <= 0.095
gini = 0.328
samples = 58
value = [46, 12]

R17 <= 0.015
gini = 0.057
samples = 34
value = [1, 33]

gini = 0.142
samples = 39
value = [36, 3]

gini = 0.499
samples = 19
value = [10, 9]

gini = 0.0
samples = 1
value = [1, 0]

gini = 0.0
samples = 33
value = [0, 33]

Tree 1:

```
R9 <= 1.925
gini = 0.5
samples = 92
value = [47, 45]
```
T →
```
R24 <= 0.095
gini = 0.328
samples = 58
value = [46, 12]
```
F →
```
R24 <= 0.04
gini = 0.057
samples = 34
value = [1, 33]
```

From R24 <= 0.095:
T →
```
R8 <= 0.175
gini = 0.142
samples = 39
value = [36, 3]
```
F →
```
R21 <= 1.57
gini = 0.499
samples = 19
value = [10, 9]
```

From R24 <= 0.04:
T →
```
gini = 0.0
samples = 1
value = [1, 0]
```
F →
```
gini = 0.0
samples = 33
value = [0, 33]
```

From R8 <= 0.175:
T →
```
gini = 0.0
samples = 1
value = [0, 1]
```
F →
```
gini = 0.1
samples = 38
value = [36, 2]
```

count: 0 1
pred = 0

From R21 <= 1.57:
T →
```
gini = 0.355
samples = 13
value = [10, 3]
```
F →
```
gini = 0.0
samples = 6
value = [0, 6]
```

Tree 2:

```
R9 <= 1.925
gini = 0.5
samples = 92
value = [47, 45]
```
→
```
R24 <= 0.095
gini = 0.328
samples = 58
value = [46, 12]
```
→
```
R24 <= 0.04
gini = 0.057
samples = 34
value = [1, 33]
```

From R24 <= 0.095:
```
R8 <= 0.175
gini = 0.142
samples = 39
value = [36, 3]
```
```
R21 <= 1.57
gini = 0.499
samples = 19
value = [10, 9]
```

From R24 <= 0.04:
```
gini = 0.0
samples = 1
value = [1, 0]
```
```
gini = 0.0
samples = 33
value = [0, 33]
```

count: 0, 1
pred = 1

From R8 <= 0.175:
```
gini = 0.0
samples = 1
value = [0, 1]
```
```
gini = 0.1
samples = 38
value = [36, 2]
```

From R21 <= 1.57:
```
gini = 0.355
samples = 13
value = [10, 3]
```
```
gini = 0.0
samples = 6
value = [0, 6]
```

Feature Importance Plot