Task 6

Instance	a_1	a_2	a_3	Class
1	M	X	A	YES
2	F	Y	B	YES
3	M	Y	C	YES
4	F	Y	C	YES
5	M	X	C	YES
6	F	Y	D	NO
7	M	Y	A	NO
8	F	X	A	NO
9	M	Y	A	NO
10	F	X	C	NO

Start by checking Gini for table:

$$GINI = 1 - \frac{5}{10}^2 - \frac{5}{10}^2 = 0.5$$

Check for a_1

	Yes	No
M	3	2
F	2	3

$$GINI_M = 1 - \frac{3}{5}^2 - \frac{2}{5}^2 = 0.48$$
$$GINI_F = 1 - \frac{2}{5}^2 - \frac{3}{5}^2 = 0.48$$
$$GINI_{a_1} = \frac{5}{10} * 0.48 + \frac{5}{10} * 0.48 = 0.48$$
$$GAIN_{a_1} = 0.5 - 0.48 = 0.02$$

Check for a_2

	Yes	No
X	2	2
Y	3	3

$$GINI_X = 1 - \frac{2}{4}^2 - \frac{2}{4}^2 = 0.5$$
$$GINI_F = 1 - \frac{3}{6}^2 - \frac{3}{6}^2 = 0.5$$
$$GINI_{a_2} = \frac{4}{10} * 0.5 + \frac{6}{10} * 0.5 = 0.5$$
$$GAIN_{a_2} = 0.5 - 0.5 = 0.0$$

Check for a_3

	Yes	No
A	1	3
B	1	0
C	3	1
D	0	1

$GINI_A = 1 - \frac{1}{4}^2 - \frac{3}{4}^2 = 0.375$

$GINI_B = 1 - \frac{1}{1}^2 - \frac{0}{1}^2 = 0.0$

$GINI_B = 1 - \frac{3}{4}^2 - \frac{1}{4}^2 = 0.375$

$GINI_D = 1 - \frac{0}{1}^2 - \frac{1}{1}^2 = 0.0$

$GINI_{a_3} = \frac{4}{10} * 0.375 + \frac{1}{10} * 0.0 + \frac{4}{10} * 0.375 + \frac{1}{10} * 0.0 = 0.3$

$GAIN_{a_3} = 0.5 - 0.3 = 0.2$

Best gain is a_3

Check multiway:

$GINI = 0.3$

Check binary:

	Yes	No
A	1	3
B, C, D	4	2

$GINI_A = 1 - \frac{1}{4}^2 - \frac{3}{4}^2 = 0.375$

$GINI_{BCD} = 1 - \frac{4}{6}^2 - \frac{2}{6}^2 = 0.44$

$GAIN = \frac{4}{10} * 0.375 + \frac{6}{10} * 0.44 = 0.414$

	Yes	No
A, B	2	3
C, D	3	2

$GINI_{AB} = 1 - \frac{2}{5}^2 - \frac{3}{5}^2 = 0.48$

$GINI_{CD} = 1 - \frac{3}{5}^2 - \frac{2}{5}^2 = 0.48$

$GAIN = \frac{5}{10} * 0.48 + \frac{5}{10} * 0.48 = 0.48$

Multiway split is better!!

Now check for A

Check for a_1

A	Yes	No
M	1	2
F	0	1

A	Yes	No
X	1	1
Y	0	2

Now check for C

Check for a_1

Check for a_2

C	Yes	No	C	Yes	No
M	2	0	X	1	1
F	1	1	Y	2	0

$GINI_M = 1 - \frac{2}{2}^2 - \frac{0}{2}^2 = 0.0$

$GINI_F = 1 - \frac{1}{2}^2 - \frac{1}{2}^2 = 0.5$

$GINI_{a_1} = \frac{2}{4} * 0.0 + \frac{2}{4} * 0.5 = 0.25$

$GINI_X = 1 - \frac{1}{2}^2 - \frac{1}{2}^2 = 0.5$

$GINI_F = 1 - \frac{0}{2}^2 - \frac{2}{2}^2 = 0.0$

$GINI_{a_2} = \frac{2}{4} * 0.5 + \frac{2}{2} * 0.0 = 0.25$

As GINI for C is equal we select best GINI from A that is a_2

Finally we splitt on a_1