Task 6

Instance	a_1	\mathfrak{a}_2	a ₃	Class
1	M	X	A	YES
2	F	Y	В	YES
3	M	Y	С	YES
4	F	Y	С	YES
5	M	X	С	YES
6	F	Y	D	NO
7	M	Y	A	NO
8	F	X	A	NO
9	M	Y	A	NO
10	F	X	С	NO

Start by checking Gini for table:

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

Check for a_1

Check for a_2

	Yes	No		Yes	No
М	3	2	X	2	2
F	2	3	Υ	3	3

$$GINI_{M}=1-rac{3}{5}^{2}-rac{2}{5}^{2}=0.48$$
 $GINI_{X}=1-rac{2}{4}^{2}-rac{2}{4}^{2}=0.5$ $GINI_{F}=1-rac{2}{5}^{2}-rac{3}{5}^{2}=0.48$ $GINI_{F}=1-rac{3}{6}^{2}-rac{3}{6}^{2}=0.5$ $GINI_{a_{1}}=rac{5}{10}*0.48+rac{5}{10}*0.48=0.48$ $GINI_{a_{2}}=rac{4}{10}*0.5+rac{6}{10}*0.5=0.5$ $GAIN_{a_{1}}=0.5-0.48=0.02$ $GAIN_{a_{2}}=0.5-0.5=0.0$

Check for a_3

	Yes	No
А	1	3
В	1	0
С	3	1
D	0	1

$$\begin{split} GINI_A &= 1 - \frac{1}{4}^2 - \frac{3}{4}^2 = 0.375 \\ GINI_B &= 1 - \frac{1}{1}^2 - \frac{0}{1}^2 = 0.0 \\ GINI_B &= 1 - \frac{3}{4}^2 - \frac{1}{4}^2 = 0.375 \\ GINI_D &= 1 - \frac{0}{1}^2 - \frac{1}{1}^2 = 0.0 \\ GINI_{a_3} &= \frac{4}{10} * 0.375 + \frac{1}{10} * 0.0 + \frac{4}{10} * 0.375 + \frac{1}{10} * 0.0 = 0.3 \\ GAIN_{a_3} &= 0.5 - 0.3 = 0.2 \end{split}$$

Best gain is a_3

Check multiway:

$$GINI = 0.3$$

Check binary:

	Yes	No		Yes	No
Α	1	3	А, В	2	3
B, C, D	4	2	C, D	3	2

$$GINI_{A} = 1 - \frac{1}{4}^2 - \frac{3}{4}^2 = 0.375$$
 $GINI_{AB} = 1 - \frac{2}{5}^2 - \frac{3}{5}^2 = 0.48$ $GINI_{BCD} = 1 - \frac{4}{6}^2 - \frac{2}{6}^2 = 0.44$ $GINI_{CD} = 1 - \frac{3}{5}^2 - \frac{2}{5}^2 = 0.48$ $GAIN = \frac{4}{10} * 0.375 + \frac{6}{10} * 0.44 = GAIN = \frac{5}{10} * 0.48 + \frac{5}{10} * 0.48 = 0.48$ 0.414

Multiway split is better!!

Now check for A

Check for a_1			Α	Yes	No	
	Α	Yes	No	X	1	1
	М	1	2	Υ	0	2
	F	0	1			

Now check for C

Check for a_1 Check for a_2

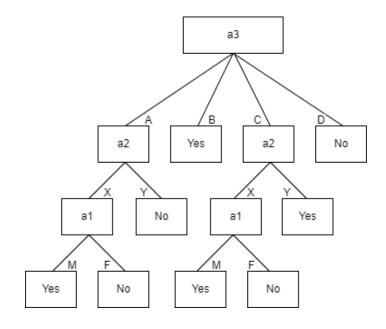
Task 6

С	Yes	No	С	Yes	No
М	2	0	X	1	1
F	1	1	Υ	2	0

$$GINI_{M} = 1 - rac{2}{2}^{2} - rac{0}{2}^{2} = 0.0 \hspace{1cm} GINI_{X} = 1 - rac{1}{2}^{2} - rac{1}{2}^{2} = 0.5 \ GINI_{F} = 1 - rac{1}{2}^{2} - rac{1}{2}^{2} = 0.5 \ GINI_{F} = 1 - rac{0}{2}^{2} - rac{2}{2}^{2} = 0.0 \ GINI_{a_{1}} = rac{2}{4} * 0.0 + rac{2}{4} * 0.5 = 0.25 \hspace{1cm} GINI_{a_{2}} = rac{2}{4} * 0.5 + rac{2}{2} * 0.0 = 0.25$$

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

Finally we splitt on a_1



Task 6