

- ① doc 2 (1, 2, 3 | 26, 27, 28)
 doc 5 (18, 19, 20 | 156, 157, 158)
- ② doc 2 (15, 16, 17)
 doc 5 (35, 36, 37)

Exo $n^o = 1$

Exo $n^o = 2$

number of documents = 4

	df	idf
one	3	$\log_{10}(\frac{4}{3}) = 0,12$
two	4	$\log_{10}(\frac{4}{4}) = 0$
three	2	$\log_{10}(\frac{4}{2}) = 0,30$
four	1	$\log_{10}(\frac{4}{1}) = 0,60$

LNN

Query	tf _{ra}	tf _w	idf	final _{vec}
one	1	1	1	1
three	2	1,3	1	1,3

$$tf_w = 1 + \log_{10}(tf_{ra})$$

$$\text{Sim}(\text{query}, \text{doc}_3) = (0,19 * 1) + (0,98 * 1,3) = 1,464$$

doc 3	tf _{ra}	idf	final _{vec}	norm _{vec}
one	1	0,12	0,12	$\frac{0,12}{0,611} = 0,19$
two	1	0	0	0
tree	1	0,60	0,60	$\frac{0,60}{0,611} = 0,98$

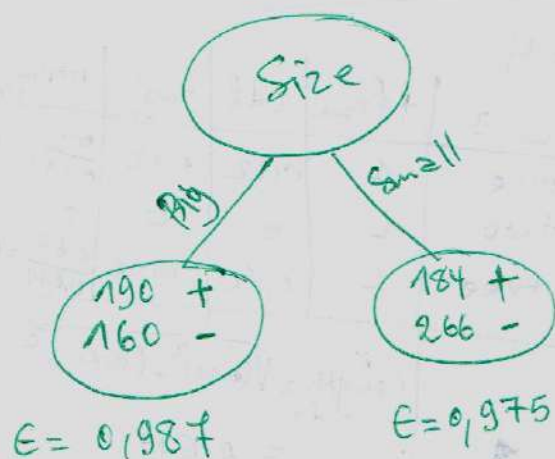
Length_{d3} = $\sqrt{(0,12)^2 + (0,60)^2} = 0,611$

NTC

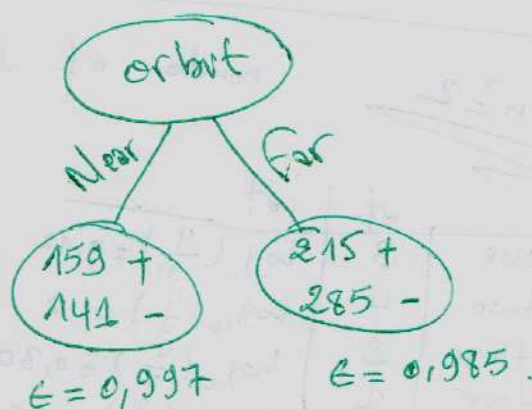
Exo n=3.

Entropy of the collection.

$$\begin{aligned}
 E(s) &= -P_+ \log(P_+) - P_- \log(P_-) \\
 &= \frac{374}{800} \log_2\left(\frac{374}{800}\right) - \frac{426}{800} \log_2\left(\frac{426}{800}\right) \\
 &= 0,509 + 0,487 \\
 &= 0,996.
 \end{aligned}$$

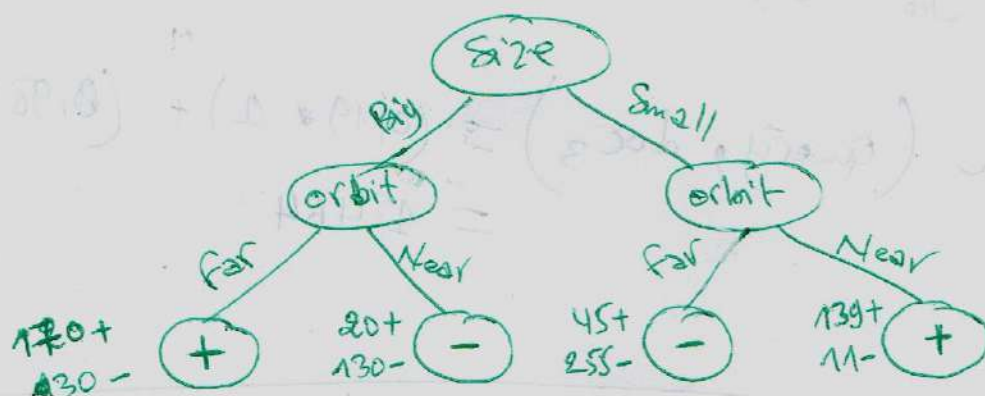


$$\begin{aligned}
 IG(Size) &= E(s) - \sum P_i \cdot E(i) \\
 &= 0,996 - \left(\frac{350}{800} \cdot 0,987 + \frac{450}{800} \cdot 0,975 \right) \\
 &= 0,015
 \end{aligned}$$



$$IG(orbit) = 0,0065$$

we choose size as root for D.T.



you do not need to do any other calculation.

All rest data will be choosing with majority voting.

②

Ero $n^0 = 4$

①	points	Center			new cluster to belong
		c_1	c_2	c_3	
		(8,4)	(7,5)	(4,9)	
	$A_1(2,10)$	8,48	7,07	2,23	C_3
	$A_2(2,5)$	6,08	5	4,47	C_3
	$A_3(8,4)$	0	1,41	6,40	C_1
	$A_4(5,8)$	5	3,60	1,41	C_3
	$A_5(7,5)$	1,41	0	5	C_2
	$A_6(6,4)$	2	1,41	5,38	C_2
	$A_7(1,2)$	7,28	6,70	7,61	C_2
	$A_8(4,9)$	6,40	5	0	C_3

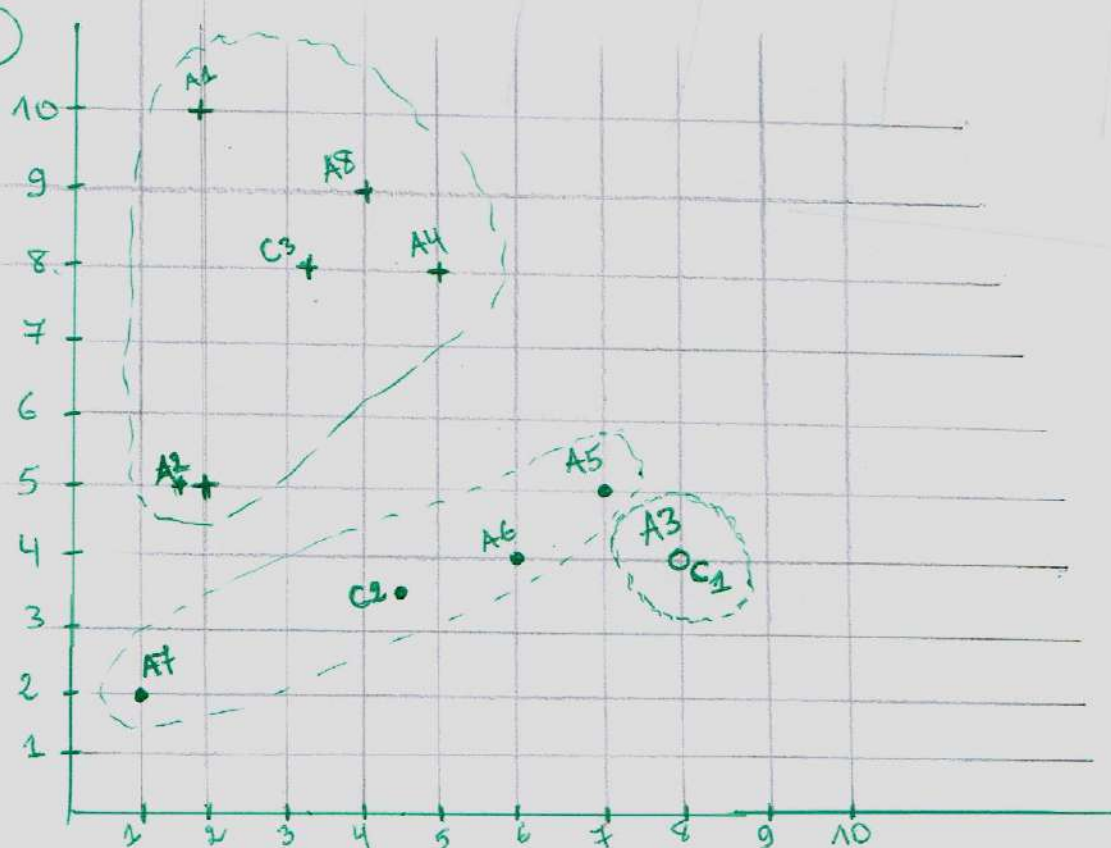
②

$$C_1 = (8,4)$$

$$C_2 = \left(\frac{7+6+1}{3}, \frac{5+4+2}{3} \right) = (\overline{4,66}, 3,66)$$

$$C_3 = \left(\frac{2+2+5+4}{4}, \frac{10+5+8+9}{4} \right) = (3,25, 8)$$

③



③