

Assignment 4

Name : Usama Tufail

Reg. No. : FA21-Bse-053 (section C)

Course : Intro to data science

Bow:

	S1	S2	S3
Analysis	0	0	1
best	0	1	0
computer	1	0	0
courses	1	1	0
data	1	1	2
important	1	0	0
in	1	0	0
is	1	1	0
most	1	0	0
of	1	1	0
one	1	1	0
perform	0	0	1
science	2	1	0
scientists	0	0	1
the	1	1	1
this	0	1	0

TF:

	S1	S2	S3
Data	$\frac{1}{11}$	$\frac{1}{8}$	$\frac{2}{15}$
science	$\frac{2}{11}$	$\frac{1}{8}$	0
is	$\frac{1}{11}$	$\frac{1}{8}$	0
one	$\frac{1}{11}$	$\frac{1}{8}$	0
of	$\frac{1}{11}$	$\frac{1}{8}$	0
the	$\frac{1}{11}$	$\frac{1}{8}$	$\frac{1}{15}$
most	$\frac{1}{11}$	0	0
important	$\frac{1}{11}$	0	0
courses	$\frac{1}{11}$	$\frac{1}{8}$	0
in	$\frac{1}{11}$	0	0
computer	$\frac{1}{11}$	0	0
this	0	$\frac{1}{8}$	0
best	0	$\frac{1}{8}$	0
scientists	0	0	$\frac{1}{15}$
perform	0	0	$\frac{1}{15}$
analysis	0	0	$\frac{1}{15}$

ADF:

computer

$$\log(3/1) = 0.48$$

is

$$\log(3/2) = 0.18$$

data

$$\log(3/3) = 0$$

the

$$\log(3/3) = 0$$

important

$$\log(3/1) = 0.48$$

one

$$\log(3/2) = 0.18$$

of

$$\log(3/2) = 0.18$$

science

$$\log(3/2) = 0.18$$

in

$$\log(3/1) = 0.48$$

courses

$$\log(3/2) = 0.18$$

most

$$\log(3/1) = 0.48$$

best

$$\log(3/1) = 0.48$$

this

$$\log(3/1) = 0.48$$

perform

$$\log(3/1) = 0.48$$

scientists

$$\log(3/1) = 0.48$$

analysis

$$\log(3/1) = 0.48$$

TF-IDF:

	S1	S2	S3
data	0	0	0
Science	0.033	0.0225	0
is	0.0163	0.0225	0
one	0.0163	0.0225	0
of	0.0163	0.0225	0
the	0	0	0
most	0.0436	0	0
important	0.0436	0	0
Courses	0.0163	0.0225	0
in	0.0436	0	0
computer	0.0436	0	0
this	0	0.06	0
best	0	0.06	0
scientists	0	0	0.096
perform	0	0	0.096
Analysis	0	0	0.096

cosine:

$$\rightarrow \cos(S_1, S_2) = \frac{S_1 \cdot S_2}{|S_1| |S_2|}$$

$$S_1 \cdot S_2 = 1+2+1+1+1+1+0+0+1+0+0+0+0+0+0+0$$
$$= 8$$

$$|S_1| = \sqrt{1+4+1+1+1+1+1+1}$$
$$= \sqrt{14}$$

$$|S_2| = \sqrt{1+1+1+1+1+1+1+1+1}$$
$$= \sqrt{9} = 3$$

$$\rightarrow \cos(S_1, S_3) = \frac{S_1 \cdot S_3}{|S_1| |S_3|}$$

$$S_1 \cdot S_3 = 3$$

$$|S_1| = \sqrt{14}$$

$$|S_3| = \sqrt{8}$$

$$\cos(S_1, S_3) = \frac{3}{\sqrt{14} \times \sqrt{8}} = 0.28$$

$$\rightarrow \cos(S_2, S_3) = \frac{S_2 \cdot S_3}{|S_2| |S_3|}$$

$$S_2 \cdot S_3 = 3$$

$$|S_2| = \sqrt{9} = 3$$

$$|S_3| = \sqrt{8}$$

$$\cos(S_2, S_3) = \frac{3}{3\sqrt{8}}$$

$$= 0.35$$

Manhattan:

$$\rightarrow \text{manhattan}(s_1, s_2) = \sum_{i=1}^{16} |s_{1[i]} - s_{2[i]}|$$

$$= 1 + 1 + 1 + 1 + 1 + 1 + 1$$

$$= 7$$

$$\rightarrow \text{manhattan}(s_1, s_3)$$

$$= 1 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$$

$$= 14$$

$$\rightarrow \text{manhattan}(s_2, s_3)$$

$$= 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$$

$$= 11$$

Euclidean:

$$\text{euclidean}(s_1, s_2) = \sqrt{\sum_{i=1}^{16} (s_{1[i]} - s_{2[i]})^2}$$

$$= 0.92$$

$$\text{euclidean}(s_2, s_3) = \sqrt{\sum_{i=1}^{16} (s_{2[i]} - s_{3[i]})^2}$$

$$= 1.26$$

$$\text{euclidean}(s_1, s_3) = \sqrt{\sum_{i=1}^{16} (s_{1[i]} - s_{3[i]})^2}$$

$$= 1.3$$