

NAME :-

DILAWAR HUSSAIN

REG No :-

SP20-BCS-120

GROUP :-

II

ASSIGNMENT :-

04

COURSE :- Intro to DATA Science

Q \Rightarrow 01

BOW :-

	Sunshine	State	enjoy	brown	fox	jump	high	run	fast	T. length
S1	2	1	1	0	0	0	0	0	0	4
S2	0	0	0	2	2	1	1	1	0	7
S3	1	1	0	0	1	0	0	1	1	5

T-F

	Sunshine	State	enjoy	brown	fox	jump	high	run	fast
S1	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	0	0	0
S2	0	0	0	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	0
S3	$\frac{1}{5}$	$\frac{1}{5}$	0	0	$\frac{1}{5}$	0	0	$\frac{1}{5}$	$\frac{1}{5}$

I.d.f

	Sunshine	State	enjoy	brown	fox	Jump	high	run	fast
S1	0.17	0.17	0.47	0.47	0.17	0.47	0.47	0.17	0.47
S2									
S3									

Tf-idf:-

	Sunshine	State	enjoy	brown	fox	Jump	high	you	fast
S_1	0.08	0.04	0.11	0	0	0	0	0	0
S_2	0	0	0	0.13	0.04	0.06	0.06	0.02	0
S_3	0.03	0.03	0	0	0.03	0	0	0.03	0.09

QUESTION \Rightarrow 2

Cosine Similarity

$$S_1 = \langle 0.08, 0.04, 0.11, 0, 0, 0, 0, 0, 0, 0 \rangle$$

$$S_3 = \langle 0.03, 0.03, 0, 0, 0.03, 0, 0, 0.03, 0.09 \rangle$$

$$S_1 = \langle 2, 1, 1, 0, 0, 0, 0, 0, 0, 0 \rangle$$

$$S_3 = \langle 1, 1, 0, 0, 1, 0, 0, 1, 1, 1 \rangle$$

$$S_1 \cdot S_3 = (2 \times 1) + (1 \times 1) + (1 \times 0) + (0 \times 0) + (0 \times 1) + (0 \times 0) + (0 \times 0) + (0 \times 1) + (0 \times 1) + (0 \times 1)$$

$$S_1 \cdot S_3 = 3$$

$$|S_1| = (2^2 + 1^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2)^{1/2} = \sqrt{6} = 2.45$$

$$|S_3| = (1^2 + 1^2 + 0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 1^2 + 1^2 + 1^2)^{1/2} = \sqrt{5} = 2.24$$

$$\cos(S_1, S_3) = \frac{S_1 \cdot S_3}{|S_1| |S_3|} = \frac{3}{(2.45)(2.24)} = 0.55$$