

NAME : Aneeq DuraiZ

Roll No : SP20-BCS-063

Group : 4

BAG OF WORDS

	Sunshile	state	enjoy	brown	fox	Jump	high	run	fast	Total
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

$$S1 = [2, 1, 1, 0, 0, 0, 0, 0, 0]$$

$$S2 = [0, 0, 0, 2, 2, 1, 1, 1, 0]$$

$$S3 = [1, 1, 0, 0, 1, 0, 0, 1, 1]$$

TF MODEL

	Sunshine	state	enjoy	brown	fox	Jump	high	run	fast	Total
S1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	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	$\frac{1}{5}$	$\frac{1}{5}$	

IDF MODEL

$$IDF = \log \left(\frac{\text{Total documents}}{\text{number of docs containing : term}} \right)$$

IDF-VALUE

Sunshine	$\log(3/2) = 0.18$
state	$\log(3/2) = 0.18$
enjoy	$\log(3/1) = 0.47$
brown	$\log(3/2) = 0.47$
fox	$\log(3/2) = 0.18$
Jump	$\log(3/1) = 0.47$
high	$\log(3/1) = 0.47$
run	$\log(3/2) = 0.18$
fast	$\log(3/1) = 0.47$

TF-IDF MODEL

	Sunshine	state	enjoy	brown	fox	Jump	high	run	fast
S1	0.083	0.0425	0.1175	0	0	0	0	0	0
S2	0	0	0	0.137	0.048	0.067	0.067	0.034	0
S3	0.034	0.034	0	0	0.034	0	0	0.034	0.094

Q2: Cosine Similarity b/w S_1 & S_3

$$\text{cosine similarity} = \frac{(S_1 \cdot S_3)}{|S_1| |S_3|}$$

$$S_1 = [2, 1, 1, 0, 0, 0, 0, 0, 0]$$

$$S_3 = [1, 1, 0, 0, 1, 0, 0, 1, 1]$$

$$S_1 \cdot S_3 = 3$$

$$|S_1| = \sqrt{2^2 + 1 + 1} = 2.45$$

$$|S_3| = \sqrt{1 + 1 + 1 + 1 + 1} = 2.23$$

$$\text{Cosine Similarity} = \frac{3}{2.45 \times 2.23} = 0.59$$