Name: Anas Khalid   Roll No: SP20-BCS-102

Group: G4   Assignment: IDS

8:1

Bog of words:-

Sentences	Sunshine	state	enjoy	brown	fox	jump	high	run	fast
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:

Sentences	sunshine	state	enjoy	brown	fox	jump	high	run	fast
S1	1/2	1/4	1/4	0	0	0	0	0	0
S2	0	0	0	2/7	2/7	1/7	1/7	1/7	0
S3	1/5	1/5	0	0	1/5	0	0	1/5	1/5

Idf:

Sunshine	0.176
state	0.176
enjoy	0.477
brown	0.477
fox	0.176
jump	0.477
high	0.477
run	0.176
fast	0.477

TF-Idf :

Sentences	sunshine	state	enjoy	brown	fox	jump	high	run	fast
S1	0.088	0.044	0.1192	0	0	0	0	0	0
S2	0	0	0	0.1362	0.050	0.068	0.088	0.025	0
S3	0.0352	0.0353	0	0	0.0352	0	0	0.0328	0.954

Cosine Similarity

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

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

Cosine similarity $= \dfrac{S1 \cdot S3}{|S1||S3|}$

$S1 \cdot S3 = 2\times1 + 1\times1 + 1\times0 + 0\times1 + 0\times1 + 0\times1 = 2+1 = 3$

$|S1| = \sqrt{2\times2 + 1\times1 + 1\times1 + 0\times0 + 0\times0 + 0\times0}$

$= \sqrt{4+1+1} = \sqrt{6}$

$|S3| = \sqrt{1\times1 + 1\times1 + 0\times0 + 1\times1 + 1\times1 + 1\times1} = \sqrt{5}$

Consine similarity $= \dfrac{3}{(\sqrt{6})(\sqrt{5})} = 0.54772$