Muhammad Asad
SP20-BCS-158
Group - 4

Intro to Data Science

Assignment 5

S1 = "sunshine state enjoy sunshine"
S2 = "brown fox jump high, brown fox run."
S3 = "sunshine state fox run fast"

BOW Model :-

Vocabulary : brown, enjoy, fast, fox, high, jump, run, state, sunshine

BOW:

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

Vector S1 = [0, 1, 0, 0, 0, 0, 0, 1, 2]
Vector S2 = [2, 0, 0, 2, 1, 1, 1, 0, 0]
Vector S3 = [0, 0, 1, 1, 0, 0, 1, 1, 1]

TF Model :-

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

IDF Model:-

	brown	enjoy	fast	fox	high	jump	run	state	sunshine
IDF	0.47	0.47	0.47	0.17	0.47	0.47	0.17	0.17	0.17

TF - IDF Values :-

tf.idf = tf x idf

	tf-idf (S1)	tf-idf (S2)	tf-idf (S3)
sunshine	0.77	0	0.417
state	0.38	0	0.417
enjoy	0.50	0	0
brown	0	0.67	0
fox	0	0.51	0.417
jump	0	0.33	0
high	0	0.33	0
run	0	0.25	0.417
fast	0	0	0.54

Cosine Similarity b/w S1 and S3 :

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

S2 = [0, 0, 1, 1, 0, 0, 0, 1, 1]

$$\cos(S1, S3) = \frac{(S1 \cdot S3)}{|S1| \cdot |S3|}$$

$(S1 \cdot S3) = (0 + 0 + 0 + 0 + 0 + 0 + 0 + 1 + 2) \Rightarrow 3$

$|S1| = (1 + 1 + 4)^{0.5} = 2.45$

$|S2| = (1 + 1 + 1 + 1 + 1)^{0.5} = 2.236$

$$\cos(S1, S3) = \frac{3}{(2.45)(2.236)} \Rightarrow 0.5476$$

Cosine Similarity = 0.5476