Assignment #05

Introduction to Data Science

Question No. 1

Solution:-

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

Box Model:-

	sunshine	state	enjoy	sunshine	brown	fox	jump	high
S1	2	1	1	1	0	0	0	0
S2	0	0	0	0	2	2	1	1
S3	1	1	0	0	0	1	0	0

	run	fast	Total
S1	0	0	4
S2	1	0	7
S3	1	1	5

$S1 \{ 2, 1, 1, 1, 0, 0, 0, 0, 0, 0 \}$
$S2 \{ 0, 0, 0, 0, 2, 2, 1, 1, 1, 0 \}$
$S3 \{ 1, 1, 0, 0, 0, 1, 0, 0, 1, 1 \}$

TF model:-

$$tf\ (\text{'sunshine'}) / =$$
$$tf\ (\text{'state'}) / =$$
$$tf$$

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

	jump	high	sun	fast
tf-S1	0	0	0	0
tf-S2	1/7	1/7	1/7	0
tf-S3	0	0	1/5	1/5

IDF model:-

$$IDF\ (\text{sunshine}) = \log(3/2) = 0.405$$
$$IDF\ (\text{state}) = \log(3/2) = 0.405$$
$$IDF\ (\text{enjoy}) = \log(3/1) = 0.693$$
$$\text{//}\ (\text{brown}) = \log(3/1) = 0.693$$
$$\text{//}\ (\text{fox}) = \log(3/2) = 0.405$$
$$\text{//}\ (\text{jump}) = \log(3/1) \fallingdotseq 0.693$$
$$\text{//}\ (\text{high}) = \log(3/1) = 0.693$$
$$\text{//}\ (\text{sun}) = \log(3/2) = 0.405$$
$$\text{//}\ (\text{fast}) = \log(3/1) = 0.693$$

TF-IDF values :-

	sunshine	state	enjoy	brown	fox	jump
Tf-IDF S1	0.202	0.101	0.173	0	0	0
Tf-IDF S2	0	0	0	0.198	0.115	0.099
Tf-IDF S3	0.081	0.081	0	0	0.081	0

	high	run	fast	
Tf-IDF S1	0	0	0	
Tf-IDF S2	0.99	0.057	0	
Tf-IDF S3	0	0.081	0.138	

Question 2:-

Cosine Similarity :-

$$S1 = [2, 1, 1, 0, 0, 0, 0, 0, 0]$$
$$S2 = [1, 1, 0, 0, 1, 0, 0, 1, 1]$$

$$S1 \cdot S2 = 2 \times 1 + 1 \times 1 + 1 \times 0 + 0 \times 0 + 0 \times 1 + 0 \times 0 + 0 \times 0 +$$
$$0 \times 1 + 0 \times 1 \quad = 2 + 1 \quad = 3$$

$$|S1| = \sqrt{2 \times 2 + 1 \times 1 + 1 \times 1} = \sqrt{4+1+1} = \sqrt{6} = 2.44$$

$$|S2| = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2} = \sqrt{5} = 2.23$$

$$\cos(S1, S2) = \frac{(S1 \cdot S2)}{|S1| |S2|} = \frac{3}{2.44 \times 2.23} = 0.55$$

$$\boxed{\cos(S1, S2) = 0.55}$$

Ans :-