Name: Romesa Ramzan

Registration No: FA19-BCS-140

Question # 01:

Solution

BoW:

S1 = "sunshine state enjoy sunshine"

S2 = "brown fox jump high, brown fox run"

S3 = "sunshine state fox run fast"

9 unique words

"sunshine", "state", "enjoy", "brown", "fox", "jump", "high", "run", "fast"

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

vector S1 : [0, 0, 2, 1, 1, 0, 0, 0, 0]

vector S2 : [0, 0, 0, 0, 0, 2, 2, 1, 1]

vector S3 : [1, 1, 1, 1, 0, 0, 1, 0, 0, 5]

TF

Total no of words in S1 = 4

Total no of words in S2 = 7

Total no of words in S3 = 5

On S1

tf("sunshine") = 2/4 = 1/2

tf("state") = 1/4

tf("enjoy") = 1/4

On S2

tf('brown") = 2/7

tf('fox') = 2/7

tf('jump') = 1/7

tf('high') = 1/7

tf('run') = 1/7

On S3

tf('sunshine') = 1/5

tf('state') = 1/5

tf('fox') = 1/5

tf('run') = 1/5

tf('fast') = 1/5

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

IDF

$$idf = \log\left(\frac{\text{Total no of documents}}{\text{no of documents containing terms}}\right)$$

On S1

idf ('sunshine') = log (3/2) = 0.18

idf ('state') = log (3/1) = 0.48

idf ('enjoy') = log (3/1) = 0.48

On S2

idf('brown') = log(3/2) = 0.18
idf('fox') = log(3/2) = 0.18
idf('jump') = log(3/1) = 0.48
idf('high') = log(3/1) = 0.48
idf('run') = log(3/1) = 0.48

On S3

idf('sunshine') = log(3/1) = 0.48
idf('state') = log(3/1) = 0.48
idf('for') = log(3/1) = 0.48
idf('run') = log(3/1) = 0.48
idf('fast') = log(3/1) = 0.48

	S1	S2	S3
run	0	0.48	0.48
sunshine	0.18	0	0.48
State	0.48	0	0.48
enjoy	0.48	0	0
brown	0	0.18	0
fox	0	0.18	0.48
jump	0	0.48	0
high	0	0.48	0
fast	0	0.48	0.48

tf x idf

	tf x idf		
	S1	S2	S3
sunshine	0.09	0	0.096
state	0.12	0	0.096
enjoy	0.12	0	0
brown	0	0.051	0
fox	0	0.025	0.096
jump	0	0.068	0
high	0	0.068	0
run	0	0.068	0.096
fast	0	0	0.096

Question # 02

Compute the cosine similarity b/w S1 and S3

Solution:

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

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

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

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

$(S1 \cdot S3) = 3$

$|S1| = (2\times2 + 1\times1 + 1\times1)^{1/2} = 2.44$

$|S3| = (1\times1 + 1\times1 + 1\times1 + 1\times1 + 1\times1)^{1/2} = 2.23$

$$\cos(S1, S3) = \frac{3}{(2.44)(2.23)} = \frac{3}{5.4412}$$

$$\cos(S1, S3) = 0.5513$$