M T W T F S S

Date: 00-00-0000

Name: M. Aun Raza
Reg. no: SP20-BCS-092
Group: (2)
Semester: 6th
Assignment: 5.

Question #1

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

Bag of words :-

S1 => Sunshine, State, enjoy.
S2 => brown, fox, Jump, high, run.
S3 => sunshine, state, fox, run, fast.

	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

Term frequencies :

S1 => Sunshine => 2/4
State => 1/4
Enjoy => 1/4

S2 => Brown => 2/7
fox => 2/7
Jump => 1/7
high => 1/7
run => 1/7

Mon Tue Wed Thu Fri Sat Sun Date: ___/___/____

S3 ⇒ sunshine ⇒ 1/5
state ⇒ 1/5
fox ⇒ 1/5
run ⇒ 1/5
fast ⇒ 1/5

IDF

Sunshine ⇒ $\log \frac{3}{2}$ ⇒ 0.176
State ⇒ $\log \frac{3}{2}$ ⇒ 0.176
Enjoy ⇒ $\log \frac{3}{1}$ ⇒ 0.477
brown ⇒ $\log \frac{3}{1}$ ⇒ 0.477
fox ⇒ $\log \frac{3}{2}$ ⇒ 0.176
Jump ⇒ $\log \frac{3}{1}$ ⇒ 0.477
High ⇒ $\log \frac{3}{1}$ ⇒ 0.477
run ⇒ $\log \frac{3}{2}$ ⇒ 0.176
fast ⇒ $\log \frac{3}{1}$ ⇒ 0.477

TF.IDF

S1

Sunshine ⇒ $\frac{2}{4} \times 0.176 \Rightarrow 0.088$
State ⇒ $\frac{1}{4} \times 0.176 \Rightarrow 0.044$
Enjoy ⇒ $\frac{1}{4} \times 0.477 \Rightarrow 0.11925$

S2

Brown ⇒ $\frac{2}{7} \times 0.477 = 0.136$
Fox ⇒ $2/7 \times 0.176 = 0.051$
Jump ⇒ $\frac{1}{7} \times 0.477 = 0.068$
High ⇒ $\frac{1}{7} \times 0.477 = 0.068$
Run ⇒ $\frac{1}{7} \times 0.176 = 0.025$

S3

Sunshine ⇒ 0.0352
State ⇒ 0.0352
Fox ⇒ 0.0352
Run ⇒ 0.0352
Fast ⇒ 0.0954

Mon Tue Wed Thu Fri Sat Sun

Date: ___/___/____

Q.2 Cosine Similarity b/w S_1 and S_3.

$$\frac{\vec{S_1} \cdot \vec{S_3}}{|\vec{S_1}||\vec{S_3}|} \Rightarrow \cos\theta = \frac{\vec{S_1} \cdot \vec{S_3}}{|\vec{S_1}|\,|\vec{S_3}|}$$

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

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

$$\vec{S_1} \cdot \vec{S_3} = (2\times1) + (1\times1) + (1\times0) + (0\times0) + (0\times1) + (0\times0) + (0\times0) + (0\times1) + (0\times1)$$

$$= 2 + 1 = 3$$

$$|\vec{S_1}| = \sqrt{2^2 + 1^2 + 1^2} = \sqrt{4+1+1} = 2.45$$

$$|\vec{S_3}| = \sqrt{1^2 + 1^2 + 1^2 + 1^2} = \sqrt{4} = 2$$

Putting values in the formula,

$$= \frac{\vec{S_1} \cdot \vec{S_3}}{|\vec{S_1}| \cdot |\vec{S_3}|}$$

$$= \frac{3}{3.2}$$

$$= \frac{3}{6} = \frac{1}{2}$$

$$\cos(S_1, S_3) = 0.5$$