

Date: .....

Sun Mon Tue Wed Thu Fri Sat

MUAZ IJAZ

SP20-BCS-056 - B

GROUP 4

## Assignment no 5

∴ Question no 1:-

S<sub>1</sub> ⇒ "sunshine state enjoy sunshine".

S<sub>2</sub> ⇒ "brown fox jump high, brown fox run".

S<sub>3</sub> ⇒ "sunshine state fox run fast".

Bag of words:-

S<sub>1</sub> ⇒ sunshine, state, enjoy.

S<sub>2</sub> ⇒ brown, fox, jump, high, run.

S<sub>3</sub> ⇒ sunshine, state, fox, run, fast.

Document	Sunshine	State	enjoy	brown	fox	Jump	high	run	fast
S <sub>1</sub>	2	1	1	0	0	0	0	0	0
S <sub>2</sub>	0	0	0	2	2	1	1	1	0
S <sub>3</sub>	1	1	0	0	1	0	0	1	1

Term Frequency :-

S<sub>1</sub>:-

Sunshine ⇒ 2/4

State ⇒ 1/4

enjoy ⇒ 1/4

Date: .....

Sun Mon Tue Wed Thu Fri Sat

S2:-

Brown	$\Rightarrow$	$2/7$
fox	$\Rightarrow$	$2/7$
Jump	$\Rightarrow$	$1/7$
high	$\Rightarrow$	$1/7$
Run	$\Rightarrow$	$1/7$

S3:-

Sunshine	$\Rightarrow$	$1/5$
State	$\Rightarrow$	$1/5$
fox	$\Rightarrow$	$1/5$
run	$\Rightarrow$	$1/5$
fast	$\Rightarrow$	$1/5$

Inverse document frequency :-

Sunshine	$\Rightarrow \log(3/2)$	$\Rightarrow 0.176$
State	$\Rightarrow \log(3/2)$	$\Rightarrow 0.176$
Enjoy	$\Rightarrow \log(3/1)$	$\Rightarrow 0.477$
brown	$\Rightarrow \log(3/1)$	$\Rightarrow 0.477$
fox	$\Rightarrow \log(3/2)$	$\Rightarrow 0.176$
Jump	$\Rightarrow \log(3/1)$	$\Rightarrow 0.477$
High	$\Rightarrow \log(3/1)$	$\Rightarrow 0.477$
run	$\Rightarrow \log(3/2)$	$\Rightarrow 0.176$
fast	$\Rightarrow \log(3/1)$	$\Rightarrow 0.477$

Date: .....

Sun Mon Tue Wed Thu Fri Sat

Term frequency - Inverse document frequency:-

S1:-

$$\text{Sunshine} \Rightarrow \frac{2}{4} \times 0.176 \Rightarrow 0.088$$

$$\text{State} \Rightarrow \frac{1}{4} \times 0.176 \Rightarrow 0.044$$

$$\text{Enjoy} \Rightarrow \frac{1}{4} \times 0.477 \Rightarrow 0.11925$$

S2:-

$$\text{Brown} \Rightarrow \frac{2}{7} \times 0.477 \Rightarrow 0.136$$

$$\text{Fox} \Rightarrow \frac{2}{7} \times 0.176 \Rightarrow 0.051$$

$$\text{Jump} \Rightarrow \frac{1}{7} \times 0.477 \Rightarrow 0.068$$

$$\text{High} \Rightarrow \frac{1}{7} \times 0.477 \Rightarrow 0.068$$

$$\text{Run} \Rightarrow \frac{1}{7} \times 0.176 \Rightarrow 0.025$$

S3:-

$$\text{Sunshine} \Rightarrow \frac{1}{5} \times 0.176 \Rightarrow 0.0352$$

$$\text{State} \Rightarrow \frac{1}{5} \times 0.176 \Rightarrow 0.0352$$

$$\text{fox} \Rightarrow \frac{1}{5} \times 0.176 \Rightarrow 0.0352$$

$$\text{Run} \Rightarrow \frac{1}{5} \times 0.176 \Rightarrow 0.0352$$

COPY

Date: .....

Sun Mon Tue Wed Thu Fri Sat

$$\text{Fast} \Rightarrow \frac{1}{5} \times 0.477 \Rightarrow 0.0954$$

### -: TF-IDF TABLE :-

Vocabulary	TF-IDF (S1)	TF-IDF (S2)	TF-IDF (S3)
Sunshine	0.088	0	0.0352
State	0.044	0	0.0352
Enjoy	0.11925	0	0
brown	0	0.136	0
fox	0	0.051	0.0352
Jump	0	0.068	0
high	0	0.068	0
Run	0	0.025	0.0352
fast	0	0	0.0954

-: Question no 2 :-

-: Cosine Similarity btw S1 and S3 :-

$$\frac{\vec{S1} \cdot \vec{S3}}{|\vec{S1}| |\vec{S3}|}$$

$$\Rightarrow \cos \theta = \frac{\vec{S1} \cdot \vec{S3}}{|\vec{S1}| |\vec{S3}|}$$

Date: \_\_\_\_\_

Sun Mon Tue Wed Thu Fri Sat

Create document vector:-

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

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

$$\vec{S1} \cdot \vec{S3} \Rightarrow (2 \times 1) + (1 \times 1) + (1 \times 0) + (0 \times 0) + (0 \times 1) + (0 \times 0) + (0 \times 0) + (0 \times 1) + (0 \times 1)$$

$$\Rightarrow 2 + 1$$

$$\Rightarrow 3$$

$$|\vec{S1}| \Rightarrow \sqrt{2 \times 2 + 1 \times 1 + 1 \times 1}$$

$$\Rightarrow \sqrt{4 + 1 + 1}$$

$$\Rightarrow 2.45$$

$$|\vec{S2}| \Rightarrow \sqrt{1 \times 1 + 1 \times 1 + 1 \times 1 + 1 \times 1 + 1 \times 1}$$

$$\Rightarrow \sqrt{5}$$

$$\Rightarrow 2.24$$

Now put values in the formula:-

$$\frac{\vec{S1} \cdot \vec{S3}}{|\vec{S1}| \cdot |\vec{S3}|}$$

$$\Rightarrow \frac{3}{2.47}$$

$$\Rightarrow 0.547$$

$$\cos(S1, S3) \Rightarrow 0.547$$