

IDS-PDF Report

Ass 1.5

Name:- Sarmad Wakeed Ahmed.
Reg:- SP20-BCS-117.
Group:- 4.

Q1:-

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

BOG:-

S1 => Sunshine, state, enjoy.

S2 => brown, fox, jump, high, run.

S3 => Sunshine, state, fox, run, fast.

	Sunshine	State	enjoy	brown	fox	Jump
S1	2	1	1	0	0	0
S2	0	0	0	2	2	1
S3	1	1	0	0	1	0

	High	run	Fast
S1	0	0	0
S2	1	1	0
S3	0	1	1

Term Frequency:-

S1:-

Sunshine $\Rightarrow 2/4$

state $\Rightarrow 1/4$

enjoy $\Rightarrow 1/4$

...

...

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	$\Rightarrow \log(3/2) \Rightarrow 0.176$
Sitate	$\Rightarrow \log(3/2) \Rightarrow 0.176$
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	$\Rightarrow \log(3/1) \Rightarrow 0.477$

IF-IDF:-

S1:-

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

$$\text{Sitate} :- \frac{1}{4} \times 0.176 \Rightarrow 0.044$$

$$\text{Enjoy} :- \frac{1}{4} \times 0.477 \Rightarrow 0.1192$$

S2:-

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

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

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

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

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

S3:-

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

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

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

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

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

TF-IDF TABLE:-

	(S1)	(S2)	(S3)
Sunshine	0.088	0	0.0352
S. state	0.044	0	0.0352
Enjoy	0.11925	0	0
brown	0	0.136	0
Fox	0	0.051	0
Jump	0	0.068	0.0352
High	0	0.068	0
Run	0	0.025	0.0352
Fast	0	0	0.0954

Q2:-

Cosine Similarity b/w S1 and S3:-

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

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

Document Vector:-

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

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

$$\begin{aligned}\vec{S1} \cdot \vec{S3} &\Rightarrow (2 \times 1) + (1 \times 1) + (1 \times 0) + (0 \times 0) \\ &\quad + (0 \times 1) + (0 \times 0) + (0 \times 0) + (0 \times 1) \\ &\quad + (0 \times 1) \\ &\Rightarrow 2 + 1 \\ &\Rightarrow 3\end{aligned}$$

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

$$\Rightarrow \sqrt{6} \quad \Rightarrow 2.45$$

$$|\vec{S3}| \Rightarrow \sqrt{1 \times 1 + 1 \times 1 + 1 \times 1 + 1 \times 1 + 1 \times 1}$$

$$\Rightarrow \sqrt{5} \quad \Rightarrow 2.24$$

Now:- 3

~~(6)(5)~~

34

~~3010~~

$\cos(S1, S3) = 7$ ~~0.51~~ Amb:-

3

(2.45) (2.24)

1

~~10~~

3

5.488