

M HASSAN LATIEF  
SP20-BCS-155

## Assignment

TF - IDF - Cosine Similarity

### Question-1

#### BOW

	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

T. length

S1	4
S2	7
S3	5

#### T.F

	Sunshine	state	enjoy	brown	fox	jump	high	run	fast
S1	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	0	0	0
S2	0	0	0	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	0
S3	$\frac{1}{5}$	$\frac{1}{5}$	0	0	$\frac{1}{5}$	0	0	$\frac{1}{5}$	$\frac{1}{5}$

## Idf :-

Sunshine	0.17
State	0.47
enjoy	0.47
brown	0.17
fox	0.47
jump	0.47
high	0.47
run	0.17
fast	0.47

## TF. IDF :-

	Sunshine	State	enjoy	brown	fox	jump	high	run	fast
S1	0.08	0.04	0.11	0	0	0	0	0	0
S2	0	0	0	0.13	0.04	0.06	0.06	0.02	0
S3	0.03	0.03	0	0	0.03	0	0	0.03	0.09

## Cosine Similarity :-

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

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

$$s1 \cdot s3 = (2+1) = 3$$

$$s1 = \sqrt{2^2 + 1^2 + 1^2} = 2.44$$

$$s3 = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2} = 2.23$$

$$\text{cosine similarity} = 0.55$$

---