

## IDS - Assign No: 5

G7-II

M T W T F S

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

FA19-BCS-043

Rafay Farrukh

## QUESTION NO: 1

Compute the BOW, TF, IDF and  
TF-IDF

S1 "Sunshine state enjoy sunshine"

S2 "Brown fox jump high , brown fox run"

S3 " Sunshine state fox run fast

For BOW document term matrix

	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

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

Vector S2: [0 0 0 2 2 1 1 1 0]

Vector S3: [1 1 0 0 1 0 0 1 1]

Now for term frequency

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

M T W T F S

## Term frequency

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

## IDF Inverse document frequency

$$idf(\text{sun} \text{shine}) = \log(3/2) = 0.18$$

$$idf(\text{state}) = \log(3/2) = 0.18$$

$$idf(\text{enjoy}) = \log(3/1) = 0.48$$

$$idf(\text{brown}) = \log(3/1) = 0.48$$

$$idf(\text{fox}) = \log(3/2) = 0.18$$

$$idf(\text{jump}) = \log(3/1) = 0.48$$

$$idf(\text{high}) = \log(3/1) = 0.48$$

$$idf(\text{run}) = \log(3/2) = 0.18$$

$$idf(\text{fast}) = \log(3/1) = 0.48$$

M	T	W	T	F	S
---	---	---	---	---	---

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

## TF-IDF

S1 → Sunshine State enjoy sunshine.

$$tf \cdot idf(\text{Sunshine}) = \frac{2}{4} \times 0.18 = 0.09$$

$$tf \cdot idf(\text{state}) = \frac{1}{4} \times 0.18 = 0.045$$

$$tf \cdot idf(\text{enjoy}) = \frac{1}{4} \times 0.48 = 0.12$$

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

$$tf \cdot idf(\text{brown}) = \frac{2}{7} \times 0.48 = 0.14$$

$$tf \cdot idf(\text{fox}) = \frac{2}{7} \times 0.18 = 0.05$$

$$tf \cdot idf(\text{jump}) = \frac{1}{7} \times 0.48 = 0.07$$

$$tf \cdot idf(\text{high}) = \frac{1}{7} \times 0.48 = 0.07$$

$$tf \cdot idf(\text{run}) = \frac{1}{7} \times 0.18 = 0.03$$

S3 → "Sunshine state fox run fast"

$$tf \cdot idf(\text{sunshine}) = \frac{1}{5} \times 0.18 = 0.04$$

$$tf \cdot idf(\text{state}) = \frac{1}{5} \times 0.18 = 0.04$$

$$tf \cdot idf(\text{fox}) = \frac{1}{5} \times 0.18 = 0.04$$

$$tf \cdot idf(\text{run}) = \frac{1}{5} \times 0.18 = 0.04$$

$$tf \cdot idf(\text{fast}) = \frac{1}{5} \times 0.48 = 0.096$$

Q: 2

## Cosine Similarity

$S_1 \sim S_3$

TF-IDF vectors:-

$$S_1 = [0.09, 0.045, 0.12, 0, 0, 0, 0, 0, 0]$$

$$S_3 = [0.04, 0.04, 0, 0, 0.04, 0, 0, 0.04, 0.096]$$

$$\cos(S_1, S_3) = \frac{S_1 \cdot S_3}{|S_1| |S_3|}$$

$$S_1 \cdot S_3 = (0.09 \times 0.04) + (0.045 \times 0.04) + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0.0054$$

$$|S_1| = \sqrt{(0.09 \times 0.09) + (0.045 \times 0.045) + 0 + 0 + (0.12 \times 0.12)}$$
$$= \sqrt{0.0081} = 0.1566$$

$$|S_3| = \sqrt{(0.04 \times 0.04) + (0.04 \times 0.04) + (0.096 \times 0.096) + (0.04 \times 0.04) + (0.096 \times 0.096)}$$
$$= \sqrt{0.0016 + 0.0016 + 0.0016 + 0.0016} = 0.0092$$

$$|S_3| = 0.125$$

$$\cos(S_1, S_3) = 0.0054$$

$$0.1566 \times 0.125 = 0.0195$$

$$0.2759 = 0.4277$$