IDS - Assignment # 5

GII.

Name: Ammar Asif
Reg no: FA19-BCS-088

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 for run fast

For Bow document term matrix.

	sunshine	state	enjoy	brown	fox	jump	high	run	fa
S1	2	1	1	0	0	0	0	b	0
S2	0	0	0	2	2	1	1	1	0
S3	1	1	0	0	1	0	0	0	1

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

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 (sunshine) = log (3/2) = 0.18

idf (state) = log (3/2) = 0.18

idf (enjoy) = log (3/1) = 0.48

idf (brown) = log (3/1) = 0.48

idf (fox) = log (3/2) = 0.18

idf (jump) = log (3/1) = 0.48

idf (high) = log (3/1) = 0.48

idf (run) = log (3/2) = 0.18

idf (fast) = log (3/1) = 0.48

TF.IDF

S1 → sunshine state enjoy sunshine.

tf.idf (sunshine) = $\frac{2}{4}$ x 0.18 = 0.09

tf.idf (state) = 1/4 x 0.18 = 0.045

tf.idf (enjoy) = 1/4 x 0.48 = 0.12

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

tf · idf (brown) = 2/7 × 0.48 = 0.14
tf · idf (fox) = 2/7 × 0.18 = 0.05
tf · idf (jump) = 1/7 × 0.48 = 0.07
tf · idf (high) = 1/7 × 0.48 = 0.07
tf · idf (run) = 1/7 × 0.18 = 0.03

S3 → " sunshine state fox run fast "

tf · idf (sunshine) = 1/5 × 0.18 = 0.04
tf · idf (state) = (1/5 × 0.18) = 0.04
tf · idf (fox) = 1/5 × 0.18 = 0.04
tf · idf (run) = 1/5 × 0.18 = 0.04
tf · idf (fast) = 1/5 × 0.48 = 0.096

Question NO # 2

Cosine - Similarity

S1 and S3

TF - IDF vectors :-

$S1 = [0.09, 0.045\ 0.12\ 0\ 0\ 0\ 0\ 0\ 0]$

$S3 = [0.04\ \ 0.04\ \ 0\ 0\ 0\ 0\ 4\ 0\ 0\ 0\ 0\ 4\ 0\ 0\ 96]$

$$\cos(S1, S3) = \frac{S1 \cdot S3}{|S1|\,|S3|}$$

$$S1 \cdot S3 = (0.09 \times 0.04) + (0.045 \times 0.04) + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0.0054$$

$$|S1| = \sqrt{(0.09 \times 0.09) + (0.045 \times 0.045) + (0.12 \times 0.12) + 0 + 0} = 0.15660$$

$$|S3| = \sqrt{(0.04)(0.04) + (0..0410.04) + (0.04 \times 0.04) + (0.04 \times 0.04) + (0.096 \times 0.096)}$$

$$= \sqrt{0.0016 + 0.0016 + 0.00\$16 + 0.0016 + 0.0092}$$

$|S3| = 0.125$

$$\cos(S1, S3) = \frac{0.0054}{0.15660 \times 0.125}$$

$$= 0.275$$