ASSIGNMENT 5

Subject : Introduction To Data Science

Submitted TO : Muhammad Sharjeel

Submitted By : Fatima Naqvi

SP20-BCS-145

Section : B

Q1

S1 " Sunshine state enjoy sunshine "
S2 " Brown fox jump high, brown fox run "
S3 " Sunshine state fox run fast "

BOW Model

Output	Input: Sunshine	state	enjoy	Brown	fox	jump	high	run	fast	Total length
S1	2	1	1	0	0	0	0	0	0	4
S2	0	0	0	2	2	1	1	1	0	7
S3	1	1	0	0	1	0	0	1	1	5

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]

Term Frequency Model

	Sunshine	state	enjoy	Brown	fox	jump	high	run	fast
Tf-S1	1/2	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
TS S3	1/5	1/5	0	0	1/5	0	0	1/5	1/5

IDF Model

	idf	
Sunshine	0.18	idf (sunshine) = log(3/2) = 0.18
state	0.18	idf (state) = log (3/2) = 0.18
enjoy	0.48	idf (enjoy) = log (3/1) = 0.48
Brown	0.48	idf (Brown) = log (3/1) = 0.48
fox	0.48	idf (fox) = log (3/2) = 0.18
jump	0.48	idf (jump) = log (3/1) = 0.48
high	0.48	idf (high) = log (3/1) = 0.48
run	0.18	idf (run) = log (3/2) = 0.18
fast	0.48	idf (fast) = log (3/1) = 0.48

TF IDF

S1

tf-idf (sunshine) = tf * idf = 1/2 * 0.18 = 0.09

tf-idf (state) = tf * idf = 1/4 * 0.18 = 0.045

tf-idf (enjoy) = tf * idf = 1/4 * 0.48 = 0.12

S2

tf-idf (Brown) = 2/7 * 0.48 = 0.137

tf-idf (fox) = 2/7 * 0.18 = 0.051

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 = 0.026

S3

tf-idf (Sunshine) = 1/5 * 0.18 = 0.036

tf-~~state~~ idf (state) = 1/5 * 0.18 = 0.036

tf-~~fo~~ idf (fox) = 1/5 * 0.18 = 0.036

tf-idf (run) = 1/5 * 0.18 = 0.036

tf-idf (fast) = 1/5 * 0.48 = 0.096

	idf(S1)	idf(S2)	idf(S3)
Sunshine	0.09	0	0.036
state	0.045	0	0.036
enjoy	0.12	0	0
Brown	0	0.137	0
fox	0	0.051	0.036
jump	0	0.07	0
high	0	0.07	0
run	0	0.026	0.036
fast	0	0	0.096

Q2

Similarity between S1 and S3.

$$\cos(S1, S3) = (S1 \cdot S3) / |S1|\,|S3|$$

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

$S3 = [1\ 1\ 0\ 0\ 1\ 0\ 0\ 1]$

$$S1 \cdot S3 = 2*1 + 1*1 + 1*0 + 0*0 + 0*1 + 0*0 + 0*0 + 0*1 + 0*1$$
$$= 3$$

$$|S1| = 2^2 + 1*1 + 1*1 + 0*0 + 0*0 + 0*0 + 0*0 + 0*0 + 0*0 + 0*0$$
$$= 6^{0.5} = 2.45$$

$$|S3| = 1*1 + 1*1 + 0*0 + 0*0 + 1*1 + 0*0 + 0*0 + 1*1 + 1*1$$
$$= 5^{0.5} = 2.24$$

$$\cos(S1, S3) = 3 / (2.45)(2.24)$$

$$\cos(S1, S3) = 0.55$$