COMSATS UNIVERSITY LAHORE

Assignment-5

Name: _Muhammad Shibbal Farooq_

Reg No: _SP20-BCS-091_

Course: _Introduction to Data Science_

Submitted to: _Dr. Muhammad Sharjeel Sb_

Department of Computer Science

S1 = "sunshine state enjoy sunshine"

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

S3 = "sunshine state fox run fast"

Bag of words

	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

TF Table

	brown	enjoy	fast	fox	high	jump	run	state	sunshine
S1	0.000000	0.25	0.0	0.000000	0.000000	0.000000	0.000000	0.25	0.5
S2	0.285714	0.00	0.0	0.285714	0.142857	0.142857	0.142857	0.00	0.0
S3	0.000000	0.00	0.2	0.200000	0.000000	0.000000	0.200000	0.20	0.2

IDF Table

sunshine	0.176
state	0.176
enjoy	0.477
brown	0.477
fox	0.176
jump	0.477
high	0.477
run	0.176
fast	0.477

Cosine Similarity:

Cosine Similarity between S1 and S3 : 0.5477225575051661