SPdO- BCS-071 (NLP) Assignment (5)

Manually NLP Report

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	fox	jump	brown	run	fast	high
S_1	2	1	1	0	0	0	0	0	0
S_2	0	0	0	2	1	2	1	0	0
S_3	1	1	0	1	0	0	1	1	0

Term Frequency Table:-

S_1	2/3	1/3	1/3	0	0	0	0	0	0
S_2	0	0	0	2/4	1/4	2/4	1/4	0	0
S_3	1/5	1/5	0	1/5	0	0	1/5	1/5	0

Inverse Document Frequency :-

S_1 => log(3/2) => sunshine => 0.176 => 0.18
log(3/2) => stale => 0.18
log(3/1) => enjoy => 0.48
log(3/2) => fox => 0.18

log(3/1) => jump => 0.48
log(3/1) => brown => 0.48
log(3/2) => run => 0.18
log(3/1) => fast => 0.18
log(3/0) => high => 0

TF-IDF Table :-

	Brown	enjoy	fast	fox	high	jump	run	stale	sunshine
S_1	0	0.12	0	0	0	0	0	0.04	0.08
S_2	0.14	0	0	0.1	0.01	0.06	0.02	0	0
S_3	0	0	0.09	0.03	0	0	0.03	0.03	0.03

(Cosine Similarity)

$S_1 \cdot S_3 \Rightarrow 2\times1 + 1(1) + 1(0) + 0 + 0 + 0 + 0 + 0 + 0$

$= 3$

$|S_1| \Rightarrow \sqrt{4+1+1+0+0+0+0+0+0+0} = \sqrt{6} \Rightarrow 2.449$

$|S_3| \Rightarrow \sqrt{1+1+0+0+1+0+0+1+1} = \sqrt{5} \Rightarrow 0.236$

$\cos(S_1, S_3) \Rightarrow \dfrac{S_1 S_3}{|S_1||S_3|}$

$\text{cosine} \Rightarrow \dfrac{3}{(2.449)(2.238)} \Rightarrow 0.54735$