Rohaan Yousuf
FA19-BCS-045
IDS ___ G-4

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

BoW model:

	sunshine	state	enjoy	brown	fox	jump	high	run	fast	Total length
S_1	2	1	1	0	0	0	0	0	0	4
S_2	0	0	0	2	2	1	1	1	0	7
S_3	1	1	0	0	1	0	0	1	1	5

Vector S_1 = [2 1 1 0 0 0 0 0 0]
Vector S_2 = [0 0 0 2 2 1 1 1 0]
Vector S_3 = [1 1 0 0 1 0 0 1 1]

TF model:

	sunshine	state	enjoy	brown	fox	jump	high	run	fast	Total
S_1	2/4	1/4	1/4	0	0	0	0	0	0	4
S_2	0	0	0	2/7	2/7	1/7	1/7	1/7	0	7
S_3	1/5	1/5	0	0	1/5	0	0	1/5	1/5	5

Vector S_1 = [1/2 1/4 1/4 0 0 0 0 0 0]
Vector S_2 = [0 0 0 2/7 2/7 1/7 1/7 1/7 0]
Vector S_3 = [1/5 1/5 0 0 1/5 0 0 1/5 1/5]

IDF model:

	sunshine	state	enjoy	beaun	for	jmp	high	run	fast
S_1	0.088	0.044	0.119	0	0	0	0	0	0
S_2	0	0	0	0.136	0.050	0.068	0.068	0.025	0
S_3	0.035	0.035	0	0	0.035	0	0	0.035	0.095

Cosine Similarity:

S_1 = "sunshine state enjoy sunshine"

S_3 = "sunshine state for run fast"

$$\text{cosine similarity} = \frac{S_1 \cdot S_3}{|S_1||S_3|}$$

$$S_1 \cdot S_3 = 2\times1 + 1\times1 + 1\times0 + 0 + 0\times1 + 0 + 0 + 0\times1 + 0\times1$$
$$= 3$$

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

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

so,

$$\text{cosine similarity} = \frac{3}{(\sqrt{6})(\sqrt{5})}$$
$$= 0.54735$$