

ASSIGNMENT-5

NAME: ARSAL SHAHZAIB BIN MUSHTAQ

REG # SP20-BLS-109

GROUP: G IV

INTRO TO DATA SCIENCE

SOLUTION 1

Row

	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

	Sunshine	state	enjoy	brown	fox	jump	high	run	fast
S1	2/4	1/4	1/4	0	0	0	0	0	0
S2	0	0	0	2/7	2/7	1/7	1/7	1/7	0
S3	1/5	1/5	0	0	1/5	0	0	1/5	1/5

IDF

Sunshine = $\log(3/2) = 0.176$
state = $\log(3/2) = 0.176$
enjoy = $\log(3/1) = 0.477$
brown = $\log(3/1) = 0.477$
fox = $\log(3/2) = 0.176$
jump = $\log(3/1) = 0.477$
high = $\log(3/1) = 0.477$
run = $\log(3/2) = 0.176$
fast = $\log(3/1) = 0.477$

TF-IDF

	Sunshine	state	enjoy	brown	fox	jump	high	run	fast
S1	0.088	0.044	0.119	0	0	0	0	0	0
S2	0	0	0	0.136	0.05	0.068	0.068	0.025	0
S3	0.035	0.035	0	0	0.035	0	0	0.035	0.035

SOLUTION 2

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

$$S2 = [0, 0, 0, 2, 2, 1, 1, 1, 0]$$

$$S3 = [1, 1, 0, 0, 1, 0, 0, 1, 1]$$

COSINE-SIMILARITY B/W $S1$ and $S3$

$$\begin{aligned} S1 \cdot S3 &= (2 \times 1) + (1 \times 1) + (1 \times 0) + (0 \times 0) + (0 \times 1) + (0 \times 0) + (0 \times 0) \\ &\quad + (0 \times 1) + (0 \times 1) = 2 + 1 = 3 \end{aligned}$$

$$\begin{aligned} |S1| &= (2^2 + 1^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2)^{1/2} \\ &= (4 + 1 + 1)^{1/2} \\ &= (6)^{1/2} \\ &= 2.45 \end{aligned}$$

$$\begin{aligned} |S3| &= (1^2 + 1^2 + 0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 1^2 + 1^2)^{1/2} \\ &= (1 + 1 + 1 + 1 + 1)^{1/2} \\ &= (5)^{1/2} \\ &= 2.24 \end{aligned}$$

$$\cos(S1, S3) = \frac{S1 \cdot S3}{|S1||S3|} = \frac{3}{(2.45)(2.24)} = 0.55$$
