Introduction To Data Science

Name: Muhammad Hashir

Reg. no: SP20-BCS-104

Group: G4

ASSIGNMENT # 5

Question - 1)

Ans:-

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

BoW Model

output	Sunshine	State	enjoy	Brown	fox	jump	high	run	fast	Total
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
Tfreq - S1	½	¼	¼	0	0	0	0	0	0
Tf - S2	0	0	0	3/7	2/7	1/7	1/7	1/7	0
Tf - S3	1/5	1/5	0	0	1/5	0	0	1/5	1/5

IDF Model

	idf			
Sunshine	idf (sunshine)	=	$\log(3/2)$	= 0.18
State	idf (State)	=	$\log(3/2)$	= 0.18
Enjoy	idf (Enjoy)	=	$\log(3/1)$	= 0.48
Brown	idf (Brown)	=	$\log(3/1)$	= 0.48
Fox	idf (Fox)	=	$\log(3/2)$	= 0.18
Jump	idf (Jump)	=	$\log(3/1)$	= 0.48
High	idf (High)	=	$\log(3/1)$	= 0.48
Run	idf (Run)	=	$\log(3/2)$	= 0.18
Fast	idf (Fast)	=	$\log(3/1)$	= 0.48

TF - IDF :-

S1)

$$\text{tf-idf (sunshine)} = tf * idf = \frac{1}{2} * 0.18 = 0.09$$
$$\text{tf-idf (State)} = tf * idf = \frac{1}{4} * 0.18 = 0.045$$
$$\text{tf-idf (enjoy)} = tf * idf = \frac{1}{4} * 0.48 = 0.12$$

S2)

$$\text{tf-idf (Brown)} = \frac{2}{7} * 0.48 = 0.137$$
$$\text{tf-idf (Fox)} = \frac{2}{7} * 0.18 = 0.051$$
$$\text{tf-idf (Jump)} = \frac{1}{7} * 0.48 = 0.7$$
$$\text{tf-idf (High)} = \frac{1}{7} * 0.48 = 0.07$$
$$\text{tf-idf (Run)} = \frac{1}{7} * 0.18 = 0.026$$

S3)

$$\text{tf-idf (sunshine)} = \frac{1}{5} * 0.18 = 0.036$$
$$\text{tf-idf (State)} = \frac{1}{5} * 0.18 = 0.036$$
$$\text{tf-idf (Fox)} = \frac{1}{5} * 0.18 = 0.036$$
$$\text{tf-idf (Run)} = \frac{1}{5} * 0.18 = 0.036$$
$$\text{tf-idf (Fast)} = \frac{1}{5} * 0.48 = 0.096$$

	idf(S1)	idf(S2)	idf(S3)
Sunshine	0.09	0	0.36
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

Question - 2)

Ans :-

Cosine Similarity b/w S1 & S3

$$Cos(S1, S3) = \frac{(S1 \cdot S3)}{|S1| \, |S3|}$$

S1 = [2 1 1 0 0 0 0 0 0]
S3 = [1 1 0 0 1 0 0 1 1]

S1 · S3 = 2×1 + 1×1 + 1×0 + 0×0 + 0×1 + 0×0

  0×0 + 0×1 + 0×1

| S1 · S3 = 3 |

|S1| = 2×2 + 1×1 + 1×1 + 0×0 + 0×0 + 0×0 + 0×0 + 0×0 + 0×0

  = $\sqrt{6}$ = | 2.45 |

|S3| = 1×1 + 1×1 + 0×0 + 0×0 + 1×1 + 0×0 + 0×0 + 1×1 + 1×1

  = $\sqrt{5}$ = | 2.24 |

$$Cos(S1, S3) = \frac{3}{(2.45)(2.24)}$$

Cos(S1, S3) = **0·55**