



INTRODUCTION TO DATA SCIENCE



ASSIGNMENT 5:

NAME: HAMMAD ZAFAR

REG#: SP20-BCS-136-B

SUBMITTED TO: Dr. Muhammad Sharjeel

IDS assignment - 5

Q1:-

Compute the BoW model, and IDF model for each of the terms in the following three sentences.

Then calculate the TF.IDF values.

S1 "sunshine state enjoy sunshine"

S2 "brown fox jump high, brown fox run"

S3 "sunshine state fox run fast"

BoW:-

	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	1	0	1	0	0	1	1

TF:-

	sunshine	state	enjoy	brown	fox	jump	high	run	fast
S1	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	0	0	0
S2	0	0	0	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	0
S3	$\frac{1}{5}$	$\frac{1}{5}$	0	0	0	0	0	$\frac{1}{5}$	$\frac{1}{5}$

IDF:-

$$\text{sunshine} = \log(3/2) = 0.176$$

$$\text{state} = \log(3/2) = 0.176$$

$$\text{enjoy} = \log(3/1) = 0.477$$

$$\text{brown} = \log(3/1) = 0.477$$

$$\text{fox} = \log(3/2) = 0.176$$

$$\text{jump} = \log(3/1) = 0.477$$

$$\text{run} = \log(3/2) = 0.176$$

$$\text{fast} = \log(3/1) = 0.477$$

$$\text{high} = \log(3/1) = 0.477$$

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	0.0355	0	0	0.035	0	0	0.035	0.095

Q2:-

Compute the cosine similarity between S1 and S3.

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

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

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

Cosine-similarity between $S1$ and $S3$:-

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

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

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

$$\begin{aligned} \cos(S1, S3) &= \frac{S1 \cdot S3}{|S1| |S3|} = \frac{3}{(2.45)(2.24)} \\ &= \mathbf{0.55} \end{aligned}$$