IDS ASSIGNMENT-5

BoW model, TF model, IDF model, TF.IDF, cosine similarity

DECEMBER 30, 2022

SANA IMRAN
SP20-BCS-087
Group-4

Q-1. Compute the BoW model, TF model, and IDF model for each of the terms in the following three sentences. Then calculate the TF.IDF values.

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

BoW = count of a term in a single document.

$$TF = \frac{\text{count of a term in a single document}}{\text{length of the document}}$$

$$IDF = \log\left(\frac{\text{number of documents in corpus}}{\text{total documents in which the term occured}}\right)$$

Terms	BoW Model			TF Model			IDF Model	TF.IDF		
	S1	S2	S3	S1	S2	S3		S1	S2	S3
sunshine	2	0	1	0.5	0	0.2	0.176	0.088	0	0.035
state	1	0	1	0.25	0	0.2	0.176	0.044	0	0.035
enjoy	1	0	0	0.25	0	0	0.4771	0.119	0	0
brown	0	2	0	0	0.285	0	0.4771	0	0.136	0
fox	0	2	1	0	0.285	0.2	0.176	0	0.05	0.035
jump	0	1	0	0	0.142	0	0.4771	0	0.067	0
high	0	1	0	0	0.142	0	0.4771	0	0.067	0
run	0	1	1	0	0.142	0.2	0.176	0	0.025	0.035
fast	0	0	1	0	0	0.2	0.4771	0	0	0.095
weight	4	7	5	1	≈1	1				

Q-2. Compute the cosine similarity between S1 & S3.

Solution.

Vectors of S1 & S3 are;

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

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

The formula for cosine similarity is the dot product of two vectors divided by the product of their lengths. i-e;

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

$$\Rightarrow S1 \cdot S3 = (2*1) + (1*1) + (1*0) + (0*0) + (0*1) + (0*0) + (0*0) + (0*1) + (0*1)$$

$$= 2 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0$$

$$S1 \cdot S3 = 3$$

$$\Rightarrow |S1| = \sqrt{(2*2) + (1*1) + (1*1) + (0*0) + (0*0) + (0*0) + (0*0) + (0*0) + (0*0)}$$

$$= \sqrt{4+1+1} = \sqrt{6}$$

$$\Rightarrow |S1| = 2.45$$

$$\Rightarrow |S3| = \sqrt{(1*1) + (1*1) + (0*0) + (0*0) + (1*1) + (0*0) + (0*0) + (1*1) + (1*1)}$$

$$= \sqrt{1+1+1+1+1} = \sqrt{5}$$

$$\Rightarrow |S3| = 2.24$$

Now; cosine similarity;

$$\cos(S1, S3) = \frac{S1 \cdot S3}{|S1| \times |S3|} = \frac{3}{(2.45)(2.24)} = \frac{3}{5.488}$$

$$\cos(S1, S3) = 0.55$$