



Vidyavardhini's College of Engineering and Technology  
Department of Artificial Intelligence & Data Science

AY: 2025-26

Class:	AI	Semester:	VII
Course Code:	CSC702	Course Name:	BDA

Name of Student:	BARI ANKIT VINOD
Roll No. :	61
Assignment No.:	5
Title of Assignment:	
Date of Submission:	29/9/25
Date of Correction:	30/9/25

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Completeness	5	4
Demonstrated Knowledge	3	2
Legibility	2	2
Total	10	8

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Completeness	5	3-4	1-2
Demonstrated Knowledge	3	2	1
Legibility	2	1	0

Checked by

Name of Faculty : Ms. Sweetsy Patil

Signature :

Date : 30/9/25

# Assignment No. 5

BDA

Q. 1) Find Jaccard similarity and Jaccard distance for following data -

$$A = [0, 1, 2, 5, 6, 8, 9] \quad , \quad B = [0, 2, 3, 4, 5, 7, 9]$$

→ S1- Find the intersection  $A \cap B$

$$|A \cap B| = \{0, 2, 5, 9\} = 4$$

S2- Find the union  $A \cup B$

$$|A \cup B| = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} = 10$$

S3- Jaccard distance (Sim<sup>ly</sup>)

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{4}{10} = \underline{\underline{0.4}}$$

S4- Jaccard distance

$$D(A, B) = 1 - J(A, B) = 1 - 0.4 = \underline{\underline{0.6}}$$

Q. 2) Consider the following two vectors in the euclidean space -

$$x = [1, 2, -1] \text{ and } y = [2, 1, 1] \text{ , calculate the cosine distance. (x.d)}$$

→ S1- dot product  $x$  and  $y$ .

$$\begin{aligned} x \cdot y &= (1)(2) + (2)(1) + (-1)(1) \\ &= 2 + 2 - 1 = \underline{\underline{3}} \end{aligned}$$

S2- Norms of  $x$  and  $y$

$$\begin{aligned} \|x\| &= \sqrt{1^2 + 2^2 + (-1)^2} \\ &= \sqrt{1 + 4 + 1} = \sqrt{6} \\ \|y\| &= \sqrt{2^2 + 1^2 + 1^2} \\ &= \sqrt{4 + 1 + 1} = \sqrt{6} \end{aligned}$$

S3- Cosine similarity -

$$\text{Cos Sim} = \frac{x \cdot y}{\|x\| \|y\|} = \frac{3}{\sqrt{6} \cdot \sqrt{6}} = \frac{3}{6} = \underline{\underline{0.5}}$$

S4- Cosine distance -

$$\text{Cos dis} = 1 - \text{Cos sim} = 1 - 0.5 = \underline{\underline{0.5}}$$

Q. 3) By using edit distance method, transform "SEA" into "ATE" and calculate edit distance bet<sup>n</sup> two strings.

→ Given,

Source = "SEA", Target = "ATE"

S1. DP Table (1) -

		A	T	E
	0	1	2	3
S	1	1	2	3
E	2	2	2	2
A	3	2	2	<u>3</u>

S2. Result - edit distance = 3

S3. Possible transformation - total edits = 3

Q. 4) Consider the following two vectors,  $x = [100011]$ ,  $y = [011010]$ , calculate the hamming distance bet<sup>n</sup> two vectors.

→ Given,  $x = [1, 0, 0, 0, 1, 1]$ ,  $y = [1, 1, 1, 0, 1, 0]$

S1. Position	x	y	Same / different
1	1	1	Same
2	0	1	different
3	0	1	different
4	0	0	Same
5	1	1	Same
6	1	0	different

S2. Count differences -

different at positions 2, 3 and 6 → total = 3.

S3. final answer -

The hamming distance = 3.

FOR EDUCATIONAL USE