Assignment: K-Means Clustering

Course Code: CS3151

Topic: K-Means Clustering

CLO: CLO3 – Analyze artificial intelligence techniques for practical problem-solving

Total Marks: 20

Submission Deadline: 28-06-2025

Background:

K-means clustering is an unsupervised learning algorithm used to group data into clusters based on similarity. It is widely used for data segmentation, pattern recognition, and image compression. In this assignment, you will explore how the K-means algorithm works and apply it to a small dataset.

Dataset:

Point	X	Y
A	1.0	2.0
B	1.5	1.8
C	5.0	8.0
D	8.0	8.0
E	1.0	0.6
F	9.0	11.0

Use this dataset to perform K-means clustering manually or using a tool like Python/Excel.

Instructions and Questions:

Q1. Understanding the Algorithm (3 marks)
- a. What is the purpose of K-means clustering?
- b. Explain how the K-means algorithm works, including initialization, assignment, and update steps.

Q2. Manual Clustering (6 marks)

- a. Using K=2, randomly initialize two centroids and assign each point to the nearest cluster.
- b. Update the centroids and repeat the assignment. Show the process for two iterations.

Q3. Visualization (3 marks)

- a. Plot the dataset points and show the clusters after the final iteration (can be hand-drawn or software-based).
- b. Label the centroids on your plot.

Q4. Evaluation and Discussion (5 marks)

- a. What is the within-cluster sum of squares (WCSS), and how is it used to evaluate clustering?
- b. Discuss the effect of different initial centroids on the final clustering result.
- c. What strategies can be used to choose the optimal number of clusters (K)?

Q5. Reflection (3 marks)

- a. Mention a real-world scenario where K-means clustering can be applied.
- b. Describe a limitation of the K-means algorithm and suggest a possible solution.