Assignment-5

Similarity & Dissimilarity of Attributes

1. You are working as a data analyst in an e-commerce company. The company wants to compare customer preferences to recommend products. Each customer's preferences are represented as a vector of features (e.g., ratings of different product categories).

Customer ratings for 5 product categories (scale 1-5)
customer_A = [4, 5, 2, 3, 4]
customer_B = [5, 3, 2, 4, 5]
# Binary preferences (1 = liked, 0 = disliked)
customer_A_binary = [1, 0, 1, 1, 0, 1]
customer_B_binary = [1, 1, 1, 0, 0, 1]
To measure how similar or different two customers are, write a Python program that computes the following measures between their preference vectors:
   1. Euclidean Distance – to measure overall difference in ratings.
   2. Manhattan Distance – to measure absolute deviation in preferences.
   3. Cosine Similarity – to check alignment of customer interests regardless of magnitude.
   4. Hamming Distance – if preferences are represented in binary (liked/disliked).
   5. Jaccard Similarity – to measure similarity in items both customers liked.

2. An online movie platform wants to analyze how similar two users are based on their movie ratings. Each user rates movies on a scale of 1 to 5.

Write a Python program to compute the following between two users:

   1. Chebyshev Distance – to measure the maximum difference in their ratings.

   2. Minkowski Distance (with p=3) – a generalized distance measure.

Use the following data:

Movie ratings (scale 1-5)

user1 = [5, 3, 4, 4, 2]

user2 = [4, 2, 5, 4, 3]