Computer Science and Engineering Department, SVNIT, Surat
M.Tech.- I DS (Semester - 1)
Foundations of Data Science
(CSDS111)

Lab Assignment: 7

Distance Measures:

Perform the following without using the built-in functions of distance measures:

1. Use the following values to calculate **various distance measures** for them and prepare a visualization to compare them for each scenario and describe their sensitivity with given values.
   a. {3000, 2000, 1000}, {4000, 5000, 6000}
   b. {300, 200, 100}, {400, 500, 600}
   c. {30, 20, 10}, {40, 50, 60}
   d. {3, 2, 1}, {4, 5, 6}
   e. {0.3, 0.2., 0.1}, {0.4, 0.5, 0.6}
   f. {0.03, 0.02, 0.01}, {0.04, 0.05, 0.06}
   g. {0.003, 0.002, 0.001}, {0.004, 0.005, 0.006}

2. Use string values as follows to calculate **various distance measures** for them and prepare a visualization to compare them for each scenario and describe their sensitivity with given values.
   a. "1000 0000 0000 0000", "0000 0000 0000 00001"
   b. "0000 0000 0000 0000", "1111 1111 1111 1111"
   c. "1111 0000 0000 0000", "0000 0000 0000 1111"

3. Implement the Levenshtein Distance Matrix to transform Your name into Your Friend's Name.

Use cases of Distance Measures using built-in functions:

Use a Multi-dimensional matrix or Dataframe to perform the following and explain your selection of distance measures used. Utilize some preprocessing, if required.

4. To identify similar customers and similar movies:

   Use the Netflix records of movie ratings provided by customers. Consider movies as similar if they were rented or rated highly by many of the same customers, and see customers as similar if they rented or rated highly many of the same movies.

   Discuss the data frame/matrix rows and columns and also discuss the impact of swapping rows or columns on similarity identification.

5. To identify similar flower IDs of the iris dataset before and after changing the scale.

6. Take the weather / stock / sensor dataset to identify the outlier record IDs.

7. Consider your FDS laboratory assignments as documents, and identify similarities among all these 1-7 assignment documents.