Course No.: CSE 2102
Course Title: Sessional based on CSE 2101
Experiment No. 1
Name of the Experiment: Implementation of Nearest Neighbor classification algorithms with and without distorted pattern.

Course Outcomes: CO1
Learning Domain with Level: Cognitive (Applying, Analyzing, Evaluating & Creating)

Background:

In **statistics**, the k-nearest neighbors' algorithm (k-NN) is a **non-parametric supervised learning** method first developed by **Evelyn Fix** and **Joseph Hodges** in **1951**[1]. It is used for both classification and regression.

K-nearest Neighbor Classification algorithm:

- Begin by defining the value of k, which represents the number of nearest neighbors to consider.
- Next, gather and organize the data that will be used for the analysis. This data should include a set of labeled training examples and a set of unlabeled test examples.
- For each test example, calculate the distance between the test example and each training example using a distance metric, such as Euclidean distance.
- Sort the training examples by their distance to the test example, with the closest training examples at the top of the list.
- Select the k training examples that are closest to the test example.
- Determine the majority label among the k training examples and assign that label to the test example.
- Repeat steps 3-6 for each test example, then evaluate the accuracy of the model by comparing the predicted labels to the true labels.
- If necessary, adjust the value of k or other parameters to improve the accuracy of the model.
- Once the algorithm is deemed accurate, it can be used to classify new examples.

Your task is to solve the following activities:

- Describe the characteristic of your dataset.
- Define the ratio of training set and test dataset
- Design a program
  - To Apply the K-nearest neighbor algorithm as classifier,
  - Analyze the accuracy of the classifier.
  - Justify for which dataset, this classifier can not perform as we expect and why.

Reference:

1. Fix, Evelyn; Hodges, Joseph L. (1951), Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties (Report). USAF School of Aviation Medicine, Randolph Field, Texas.