## **Assignment No: 01**

# Assignment Name: Classification Using k-Nearest Neighbors (k-NN) Algorithm

## **Introduction:**

k-Nearest Neighbors (k-NN) is a popular supervised machine learning algorithm used for both classification and regression tasks. The main idea behind k-NN is to classify a new data point based on the majority class of its k nearest neighbors from the training dataset.

## **Dataset Preparation:**

The dataset contains three columns: point, location, and label. The location column stores coordinates in the format (x, y). We extracted the x and y values from this column and used them as features for the model, while the label column served as the target variable.

## **Model Training and Testing:**

The dataset was split into training (80%) and testing (20%) subsets. A k-NN classifier was built with k = 3, and the training set was used to train the model. The model was then evaluated on the test set to measure its performance.

## **Results:**

The model achieved 100% accuracy on the test data. This perfect score is due to the pattern-based dataset, where each label’s coordinate range was distinct with no overlap between classes. As a result, the k-NN algorithm could perfectly classify every test point. The performance was verified using both accuracy score and a detailed classification report.

## **Conclusion:**

The k-NN algorithm is a simple yet powerful classification method. In this experiment, the dataset was highly structured, making classification easy for the model. In real-world scenarios, where data often has overlapping features and noise, accuracy may be lower. Nevertheless, k-NN remains an effective method when the data is well-separated and properly scaled.

## **References:**

1. Scikit-learn Documentation:

https://scikit-learn.org/stable/modules/neighbors.html

2. Python Official Documentation: https://docs.python.org/3/

3. Pandas Documentation: https://pandas.pydata.org/docs/