Marwadi University Marwadi Chandarana Group	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning	**Aim: KNN**	
Experiment No:04	**Date:**	**Enrolment No: 92301733051**

Aim : KNN

1. Impact of Selecting K Value Over Different Natured Datasets

The value of K in the K-Nearest Neighbors (KNN) algorithm plays a very important role in the performance of the model.

- If K is too small (like 1 or 3), the model becomes highly sensitive to noise and may overfit the data. This happens because the prediction depends only on a few neighbors, which may include outliers or noisy samples.
- If K is too large (like 20 or 50), the model becomes too smooth and may underfit the data, as it averages over many neighbors and loses the ability to capture local patterns.

Impact based on dataset nature:

- Low-noise and well-separated datasets: Small K works well because nearest neighbors are reliable.
- High-noise or overlapping datasets: Large K performs better to average out the noise.
- Imbalanced datasets: A large K can cause bias toward the majority class, so K must be chosen carefully.

2. Process of Applying KNN in Regression Problems

KNN can also be used for regression problems where the goal is to predict a continuous value.

Steps to apply KNN Regression:

1. Select the value of K (number of neighbors).
2. Compute the distance (such as Euclidean distance) between the new input sample and all training samples.
3. Select K nearest neighbors based on the smallest distance.
4. Predict the output by calculating the average (or weighted average) of the target values of those K neighbors.

Example:

If K = 3 and the target values of neighbors are [5, 7, 9], the predicted output is: Prediction = (5 + 7 + 9) ÷ 3 = 7.0

Marwadi University Marwadi Chandarana Group	**Marwadi University** **Faculty of Engineering and Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning	**Aim: KNN**	
Experiment No:04	**Date:**	**Enrolment No: 92301733051**

3. Can KNN Be Applied on Unlabelled (Unsupervised Learning) Dataset?

KNN is a supervised learning algorithm, which means it requires labeled data to predict class labels (in classification) or target values (in regression).

- In unsupervised learning, there are no labels available, so KNN cannot be directly applied for prediction.

But there are some unsupervised applications using KNN concept:

- Anomaly Detection:
  Calculate the average distance of a data point to its K nearest neighbors. If the distance is very large compared to other points, it is treated as an anomaly.

- KNN in Clustering:
  Build a graph where each node connects to its K nearest neighbors. Then use clustering methods like DBSCAN or spectral clustering.

Conclusion:

KNN cannot directly solve unsupervised problems but its concept of "neighbor distance" is useful in tasks like anomaly detection or clustering.