

# **ASSIGNMENT NO: 3**

## **Problem Statement -**

Apply appropriate ML algorithm on a dataset collected in a cosmetics shop showing details of customers to predict customer response for special offers.

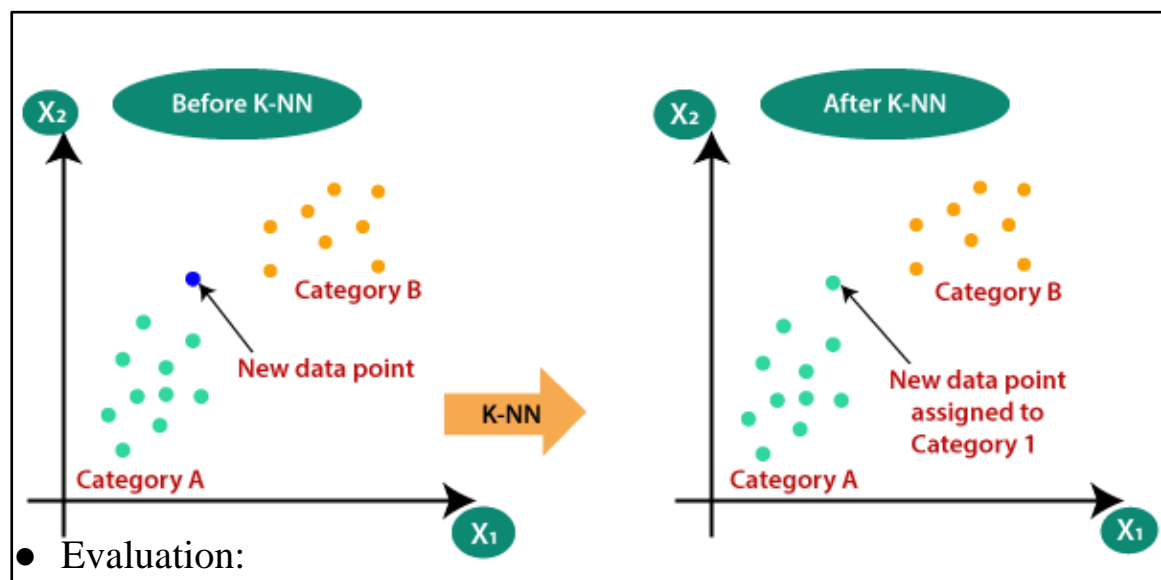
## **S/W Packages and Libraries used:**

For the following assignment, the interpreter used was Google Collab and the Primary Library used was-

- Scikit-learn: Scikit-learn offers a wide range of machine learning algorithms and evaluation metrics, enabling the implementation and evaluation of the KNN classifier for predicting customer responses to special offers.
- K-Nearest Neighbors (KNN) is a simple but effective supervised machine-learning technique used for classification and regression. The basic idea of KNN is based on the assumption that similar data points tend to be close together in the feature space.
- Here's a brief overview of how KNN works:
  - Training Phase:
    - During the training phase, KNN stores all available data points as well as their corresponding class labels (for

classification) or output values (for regression). KNN is a lazy learner, hence there is no explicit model training during this phase.

- Prediction Phase:
  - For each new data point that needs to be classified or predicted, the algorithm calculates its distance to all other data points in the training set.
  - The distance metric commonly used is Euclidean distance, but other distance metrics like Manhattan distance or cosine similarity can also be used.
  - For regression tasks, KNN predicts the output value as the average (or weighted average) of the output values of the K nearest neighbors.



- After making predictions on the test set using KNN, the performance of the model is evaluated using appropriate evaluation metrics such as accuracy, precision, recall, F1-score

(for classification), mean squared error, and mean absolute error (for regression).

## **Theory-**

- Data Preparation:
  - Load the dataset using Pandas to a DataFrame, ensuring it contains relevant customer details and responses to special offers.
  - Preprocess the data by handling missing values, encoding categorical variables, and splitting the dataset into features (input variables) and target variables (customer response).
- Feature Scaling:
  - Standardize the features using StandardScaler from Scikit-learn to ensure all features are on the same scale, preventing any particular feature from dominating the model training process.
- Train-Test Split:
  - Split the dataset into training and testing sets using `train_test_split` from Scikit-learn. This ensures the model is trained on one portion of the data and evaluated on another to assess its generalization performance.
- Model Selection and Training:

- Apply the K-Nearest Neighbors (KNN) algorithm from Scikit-learn to predict customer responses to special offers. KNN is chosen for its simplicity and effectiveness in classification tasks.
- Train the KNN classifier on the training data, tuning hyperparameters such as the number of neighbors (K) if necessary.

### **Applications:**

- Customer Relationship Management: Predicting customer responses to special offers can aid in personalized marketing strategies, enhancing customer satisfaction and loyalty.
- Sales Optimization: Understanding customer preferences enables the optimization of product offerings and promotional campaigns, leading to increased sales and revenue.

### **Limitations:**

- Data Quality: The effectiveness of the predictive model heavily relies on the quality and representativeness of the dataset. Inaccurate or biased data may lead to unreliable predictions.
- Model Complexity: While KNN is simple and intuitive, it may not perform optimally with large datasets or in high-dimensional feature spaces. Other algorithms like Decision Trees or Random Forests could be explored for better performance in such scenarios.

## **Conclusion:**

By following this methodology and understanding the applications and limitations of the assignment, practitioners can effectively apply machine learning techniques to predict customer responses in a cosmetics shop setting, leveraging the capabilities of libraries like Pandas, NumPy, and Scikit-learn.