**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.

**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.

**Algorithm-**

Random Forest Classifier:

* Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

**Working-**

A diagram of a confusion matrix

Description automatically generated

A screenshot of a computer

Description automatically generated

**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.