**ASSIGNMENT: - 03**

**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, Library, Package: -**

1. **Software**: You can use either R or Python for this task. Both languages offer a range of libraries and tools for machine learning tasks.
2. **Library**: scikit-learn (sklearn)
3. **Packages**:
4. pandas: For data manipulation and preprocessing.
5. numpy: For numerical computations and array operations.
6. matplotlib/seaborn: For data visualization.
7. sklearn: For machine learning algorithms and model evaluation.

**Theory:**

**Support Vector Machine (SVM):**

* Supervised Learning Algorithm: SVM is a supervised learning algorithm, meaning it requires labelled data for training. It learns to classify data into different categories based on the features provided.
* Classification and Regression: SVM can be used for both classification and regression tasks. In classification, it separates data points into different classes, while in regression, it predicts a continuous outcome.
* Hyperplane Separation: The primary goal of SVM is to find the hyperplane that best separates different classes in the feature space. This hyperplane has the maximum margin, which is the distance between the hyperplane and the nearest data points of each class.
* Effective in High-dimensional Spaces: SVM is effective in high-dimensional spaces, where the number of dimensions exceeds the number of samples. It can handle complex datasets with many features.
* Kernel Trick: SVM uses a kernel function to map the input data into a high-dimensional feature space. This allows it to find nonlinear decision boundaries in the original feature space**.**

**Advantages**:

1. **Statistical Analysis**: R offers a comprehensive set of statistical functions and algorithms for data analysis, hypothesis testing, regression, clustering, and more.
2. **Graphics and Visualization**: R has powerful graphics capabilities with packages like ggplot2 for creating high-quality plots, charts, and visualizations.
3. **Community and Packages**: The R community is large and active, contributing thousands of packages to CRAN, providing solutions for diverse analytical tasks.
4. **Integration and Reproducibility**: R integrates well with other languages and tools, and tools like RMarkdown support reproducible research by combining code, visualizations, and narrative text in interactive documents.

**Applications**:

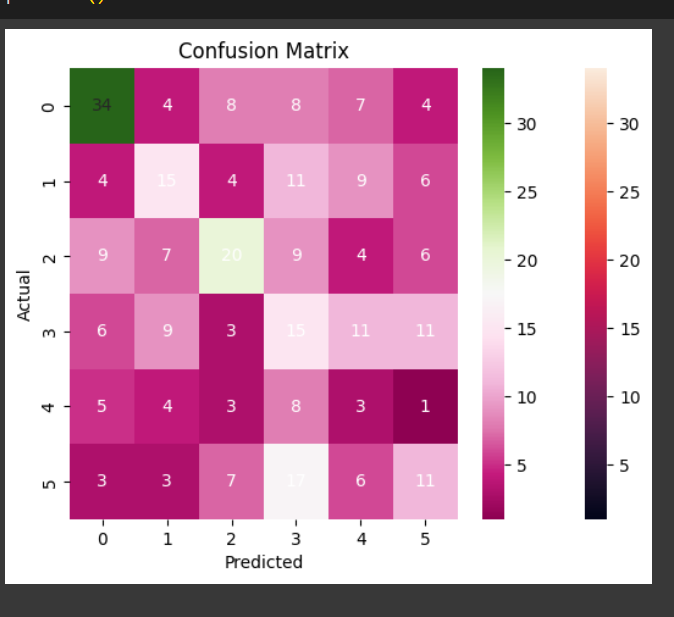
* **Data Analysis**: R is widely used for data analysis in various domains such as finance, healthcare, marketing, and academia.
* **Statistical Modeling**: R is used for statistical modeling tasks like linear regression, logistic regression, time series analysis, and machine learning.
* **Data Visualization**: R is popular for creating visualizations ranging from simple plots to complex interactive dashboards.
* **Research and Academia**: R is extensively used in research, academic studies, and data-driven scientific research.

**Limitations**:

1. **Learning Curve**: R has a steep learning curve for beginners, especially those without a programming or statistical background.
2. **Performance**: Certain operations in R can be slower compared to other languages like Python, especially for large datasets or computationally intensive tasks.
3. **Production Deployment**: While R is excellent for prototyping and analysis, it may not be the best choice for large-scale production systems.

**Working/ Algorithm:**

1. Import necessary libraries (pandas, sklearn, seaborn, matplotlib).
2. Read cosmetic dataset into a pandas DataFrame (df).
3. Split data into features (X: Price, Rank) and target variable (y: Label).
4. Split data into training and testing sets (X\_train, X\_test, y\_train, y\_test).
5. Standardize features using StandardScaler (scaler).
6. Fit and transform training set, and transform testing set using scaler.
7. Create RandomForestClassifier model with n\_estimators=100 and random\_state=42.
8. Fit model to training data (X\_train, y\_train).
9. Predict target variable for testing set (X\_test) using model.predict.
10. Calculate accuracy score using accuracy\_score.
11. Generate confusion matrix using confusion\_matrix.
12. Generate classification report using classification\_report.
13. Print confusion matrix and classification report.
14. Create heatmap of confusion matrix using seaborn's heatmap.
15. Customize heatmap and display using matplotlib's plt.show().



**Conclusion:**

Applying a machine learning algorithm on a dataset from cosmetics shop to predict customer responses for special offers involves steps like data preprocessing, feature selection, model selection, training, evaluation, and prediction. A suitable algorithm such as logistic regression, decision trees, or random forest can be chosen based on the dataset characteristics. The model's performance is evaluated using metrics like accuracy, precision, recall, or F1-score. In conclusion, the ML algorithm helps in predicting customer responses effectively, aiding in targeted marketing strategies and offer optimizations.