🔹 1. Logistic Regression

A statistical model used for binary classification. It uses the sigmoid function to map linear combinations of input features to probabilities between 0 and 1.

🔹 2. Binary Classification

A task that involves classifying data into two distinct classes (e.g., purchase or not). Logistic regression is commonly used for such tasks.

🔹 3. Supervised Learning

A machine learning paradigm where the model learns from labeled data. In this case, the label is whether the user purchased the product or not.

🔹 4. Feature Scaling (StandardScaler)

Standardization of data to have mean = 0 and standard deviation = 1. This ensures all features contribute equally during model training.

🔹 5. Model Training (`fit()` method)

The process where the algorithm learns the relationship between features and labels from the training data.

🔹 6. Model Prediction (`predict()` method)

Using the trained model to predict labels for new/unseen data.

🔹 7. Train-Test Split

Dividing data into training and testing subsets to evaluate the model's performance on unseen data.

🔹 8. Sigmoid Function

Used in logistic regression to convert the linear output into a probability value.
Formula:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

🔹 9. Confusion Matrix

A matrix showing the counts of:

- **True Positives (TP)**
- **False Positives (FP)**
- **True Negatives (TN)**
- **False Negatives (FN)**
  Used to evaluate classification performance.

🔹 10. Accuracy

Proportion of total predictions that are correct.
Formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

🔹 11. Precision

Indicates how many predicted positives were actually positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

🔹 12. Recall (Sensitivity / TPR)

Measures how many actual positives were captured.

$$\text{Recall} = \frac{TP}{TP + FN}$$

🔹 13. Specificity (TNR)

Measures how many actual negatives were correctly identified.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

🔹 14. Error Rate

Shows how many predictions were wrong.

$$\text{Error Rate} = \frac{FP + FN}{Total}$$

🔹 15. False Positive Rate (FPR)

Shows the proportion of actual negatives incorrectly classified as positive.

$$\text{FPR} = \frac{FP}{FP + TN}$$

🔹 16. Classification Report

Summarizes precision, recall, f1-score, and support for each class. Helpful for quick evaluation of classification performance.

🔹 17. Evaluation Metrics in ML

Includes precision, recall, F1-score, accuracy, ROC-AUC, etc., to assess model quality beyond just accuracy.

🔹 18. Data Preprocessing

Cleaning data (like checking for nulls), selecting features, encoding labels, and scaling to prepare for ML modeling.

🔹 19. scikit-learn Library

A powerful Python ML library used for data preprocessing, model training, prediction, and evaluation.

🔹 20. Data Visualization (Matplotlib)

Used for visualizing the confusion matrix and performance insights for easier interpretation.