

Home Assignment: Customer Purchase Prediction using Logistic Regression

Learning Objective

The objective of this assignment is to apply **Logistic Regression** to solve a **binary classification problem**.

Students will learn how to:

- Build a Logistic Regression model
 - Predict class labels and probabilities
 - Evaluate model performance using **Confusion Matrix, Accuracy, Precision, Recall**
-

Expected Completion Time

- **Best Case:** 45 minutes
 - **Average Case:** 75 minutes
-

Problem Statement

You are provided with a **Customer Purchase dataset** that contains demographic information about customers and whether they purchased a product.

Your task is to build a **Logistic Regression model** that predicts whether a customer will purchase the product based on their profile.

Dataset Description

Dataset file: `customer_purchase_data.csv`

Column Name	Description
Age	Age of the customer
EstimatedSalary	Estimated annual salary

Column Name	Description
Purchased	Target variable (0 = No, 1 = Yes)

Assignment Requirements

Step 1: Data Loading and Exploration

- Load the dataset using Pandas
 - Display:
 - First few rows
 - Dataset shape
 - Summary statistics
 - Check for missing values
-

Step 2: Feature and Target Selection

- Independent Variables (X):
 - Age
 - EstimatedSalary
 - Dependent Variable (y):
 - Purchased
-

Step 3: Train–Test Split

- Split the dataset into:
 - **80% Training**
 - **20% Testing**
 - Use a fixed `random_state` for reproducibility
-

Step 4: Model Building – Logistic Regression

- Create a **Logistic Regression** model using `sklearn`
 - Train the model using training data
 - Predict:
 - Class labels for test data
 - Probabilities for test data
-

Step 5: Model Evaluation (Important)

Evaluate the model using the following metrics:

1. **Confusion Matrix**
 - Display the matrix and explain:
 - True Positives
 - True Negatives
 - False Positives
 - False Negatives
 2. **Accuracy Score**
 - Measure overall correctness of the model
 3. **Precision**
 - Measure how many predicted purchases were actually correct
 4. **Recall**
 - Measure how many actual purchases were correctly identified
-

Step 6: Visualization

- Plot the **Confusion Matrix** using Matplotlib
 - Clearly label:
 - Axes
 - Class names (Purchased / Not Purchased)
-

Step 7: User Input Prediction

- Ask the user to enter:
 - Age
 - Estimated Salary
 - Predict:
 - Purchase probability
 - Final class label (Yes / No)
 - Display the result clearly
-

Step 8: Model Interpretation

Answer the following questions (in comments or markdown):

1. Why is Logistic Regression suitable for this problem?
 2. What does Precision indicate in a business context?
 3. What does Recall indicate in a business context?
 4. If Precision is high but Recall is low, what does it mean?
-

Hints for Learners

- Logistic Regression predicts **probabilities**, not just labels
 - Use:
`predict_proba()`
to see probabilities
 - Confusion Matrix helps understand **type of errors**, not just accuracy
 - Accuracy alone can be misleading for classification problems
-

Expected Outcome

After completing this assignment, you should be able to:

- Apply Logistic Regression to a real classification problem
 - Evaluate models using Confusion Matrix, Accuracy, Precision, and Recall
 - Interpret classification metrics in a business context
 - Predict outcomes for new customer inputs
 - Understand trade-offs between Precision and Recall
-

💡 Industry Perspective

Logistic Regression is widely used in:

- Marketing campaign response prediction
- Customer conversion analysis
- Fraud detection
- Credit risk assessment