

A screenshot of a math equation

Description automatically generated

A screenshot of a white paper with black text

Description automatically generated

**Explanation of the Code:**

**1.Load Libraries:** dplyr for data manipulation and caret for model training and evaluation.

**2.Load Data:** Reads the CSV file into a data frame.

**3.Data Preprocessing:** Converts categorical variables to numeric format.

**4.Define Features and Target:** Separates features (X) from the target variable (y).

**5.Split Dataset:** Splits the data into training and testing sets (80% training, 20% testing).

**6.Fit Model:** Fits a logistic regression model using the training data.

**7.Predictions:** Makes predictions on the test set.

**8.Evaluation:** Evaluates the model using a confusion matrix and calculates the ROC-AUC.

**About Dataset**

**Context**

When a bank receives a loan application, based on the applicant’s profile the bank has to make a decision regarding whether to go ahead with the loan approval or not. Two types of risks are associated with the bank’s decision."

"If the applicant is a good credit risk, i.e. is likely to repay the loan, then not approving the loan to the person results in a loss of business to the bank

If the applicant is a bad credit risk, i.e. is not likely to repay the loan, then approving the loan to the person results in a financial loss to the bank."