SCHOOL OF SCIENCE & TECHNOLOGY
DEPARTMENT OF DATA SCIENCE & ANALYTICS
MATHEMATICAL FINANCE (MTH3010B)
PRACTICAL 2
FALL 2025
DUE DATE: 1^{st} December, 2025 (Time 11:59 PM)

INSTRUCTOR: Carolyne Macharia

Instructions

You are provided with a credit risk dataset containing borrower characteristics and loan information. Your task is to perform exploratory data analysis (EDA) and build a model to estimate the Probability of Default (PD) using python.

Assignment Tasks

Task 1. Data Understanding and Cleaning

(a) Load the dataset and provide a summary of its structure (number of observations, variables, types).

(b) Identify and handle missing values appropriately.

(c) Detect and comment on outliers in key numerical variables.

Task 2. Exploratory Data Analysis (EDA)

(a) Provide univariate analysis for numerical and categorical variables.

(b) Provide bivariate analysis between predictors and the target variable (Default).

(c) Plot at least three visualizations (e.g., histograms, boxplots, bar charts, correlation heatmap) and interpret them.

Task 3. Feature Engineering

(a) Create at least two new features that may improve model performance if necessary.

(b) Justify why these new features may be useful.

(c) Encode categorical variables using one-hot encoding.

(d) Normalize or scale numerical features if needed.

Task 4. Probability of Default (PD) Modeling

(a) Split the data into training and testing sets. (use a ratio of 80:20)

(b) Fit a Logistic Regression model to predict the probability of default.

(c) Interpret the sign and magnitude of at least two coefficients.

(d) Evaluate the model using metrics such as AUC, accuracy, confusion matrix, and KS statistic.

Task 5. Reporting

(a) Summarize your findings and insights from the EDA.

(b) Comment on the model's strengths and weaknesses.

(c) Provide recommendations for improving the model.