Business Analytics – 2024

By Farzad Minooei, PhD, MBA

Homework #5: Logistic Regression

Q1: To answer the assignment, use hw_5.csv.

The data set contains information on 1,000 loan applicants of a bank. Default is the dependent variable in the data, where 1 denotes presence of loan default and 0 denotes absence of loan default. There are 15 predicting variables; 12 variables are numeric and 3 variables are categorical as follows:

- Checking _Amount (Numeric)
- Term (displayed in months (Numeric))
- Credit_score(Numeric)
- Gender (Categorical)
- Marital_status(Categorical)
- Car_loan (1- Own car loan, 0- Does not own car loan – Numeric)
- Personal_loan (1- Own Personal loan, 0- Does not own Personal loan – Numeric)
- Home_loan (1- Own Home loan, 0- Does not own Home loan – Numeric)
- Education_loan (1- Own Education loan, 0- Does not own Education loan – Numeric)
- Emp_status (Categorical)
- Amount (Numeric)
- Saving_amoun (Numeric)
- Emp_duration (which is displayed in months (Numeric))
- Age (which is displayed in years (Numeric))
- No_of_credit_account (Numeric)

The goal of the assignment is to build a model to predict bank-loan default.

a) Do exploratory analysis to understand your data.
b) Split data into test and train. Build a model to predict bank-loan default.
c) Calculate the best cutoff value for your model in part b.
d) Evaluate the performance of your model on test based on accuracy, precision, recall, and F-beta score.