

FinTech-Bank Credit Risk

Credit Classification

Motivation

FinTech-Bank serves hundreds of thousands of clients in a variety of ways. Some use FinTech-Bank as a chequing account, some as a budgeting app, but many use FinTech-Bank for our loan-related products.

The credit risk team at FinTech-Bank has a critical role in determining whether each user qualifies for our loan products or not. There are two types of risks involved:

1. Not approving a loan to a person who is likely to repay the loan
2. Approving a loan to a person who is unlikely to repay the loan

Objective

Your goal is to build a model that minimizes the two risks above and ultimately maximizes profit. Specifically, your goal is to build a model that best predicts whether a user is creditworthy or not using the target variable, Credit_Risk. *We recommend that you create a model that determines a probability or a score for each user, and then determine the best score cutoff to classify each user.*

You will use the dataset provided, **credit_data.csv**. In addition to building the model, you are required to answer the following questions:

1. What variables are the most indicative of whether a person should be approved for a loan or not? How did you determine this?
2. What metric did you use to evaluate your model, and why did you choose that metric in particular? How did your model perform given your metric of choice?
3. Assume the following profit matrix, where giving a loan to a creditworthy user results in 40% profit, while giving a loan to an uncreditworthy user results in a 100% loss:

		Predicted	
		Creditworthy	Non-creditworthy
Actual	Creditworthy	40%	0
	Non-creditworthy	-100%	0

What is the per-user profit of your model given this cost matrix? What final recommendation(s) would you make after conducting this analysis?

4. If you had access to any additional data that you wanted, what other features would you add to your model?

Data

The dataset provided has 1000 rows by 21 columns (20 features and a target variable):

1. **Chequings_Amount:** \$ amount in the user's chequing account
2. **Duration_Months:** Length of loan in months
3. **Credit_History:** How well the user has paid back credit
4. **Purpose_Of_Loan**
5. **Loan_Amount**
6. **Savings_Amount:** \$ amount in savings
7. **Employment_Status**
8. **Installment_Rate:** installment rate in percentage of disposable income
9. **Personal_Status_Sex:** male or female, single/divorced/separated/married/widowed
10. **Other_Debtor_Guarantor:** whether or not the user has another guarantor
11. **Resident_Since:** how long the user has been at their present residence
12. **Property_Type:**
13. **Age:** how old the user is
14. **Other_Installments:** other installment plans
15. **Housing_Payment_Type:** rent, own, free
16. **Number_Of_Loans:** Number of existing credits at this bank
17. **Job:** status of employment
18. **Number_Of_Dependents:** Number of people being liable to provide maintenance for
19. **Own_Telephone**
20. **Foreign_Worker**
21. **Credit_Risk:** 1 if uncreditworthy, 2 if creditworthy

Deliverables

1. Code with your final model and any exploratory data analysis that you conducted. Please submit a single **Jupyter Notebook file** with your EDA component as well as your modeling component.
2. A **csv file** containing the features of your model, the predicted value, and the actual value for your test data.
3. A **pdf** with your answers to the three questions above. Please clearly state any assumptions that you have made.