Credit risk prediction

Creditworthiness: Accurate detection for high-risk credit Applicants

Marc Bourleau,
Julian Changanaqui,
Charlène Khairallah

Credit assessment for better decision making

As a bank responsible for managing the accounts of multiple clients, caution is needed when extending credit to certain individuals.

How can we ensure that credit applicants will repay their debts reliably and on time?

Our goal is to:

- Provide actionable insights for informed decision-making
- Understand the key elements influencing the credit risk
- Most importantly, correctly classify clients that default

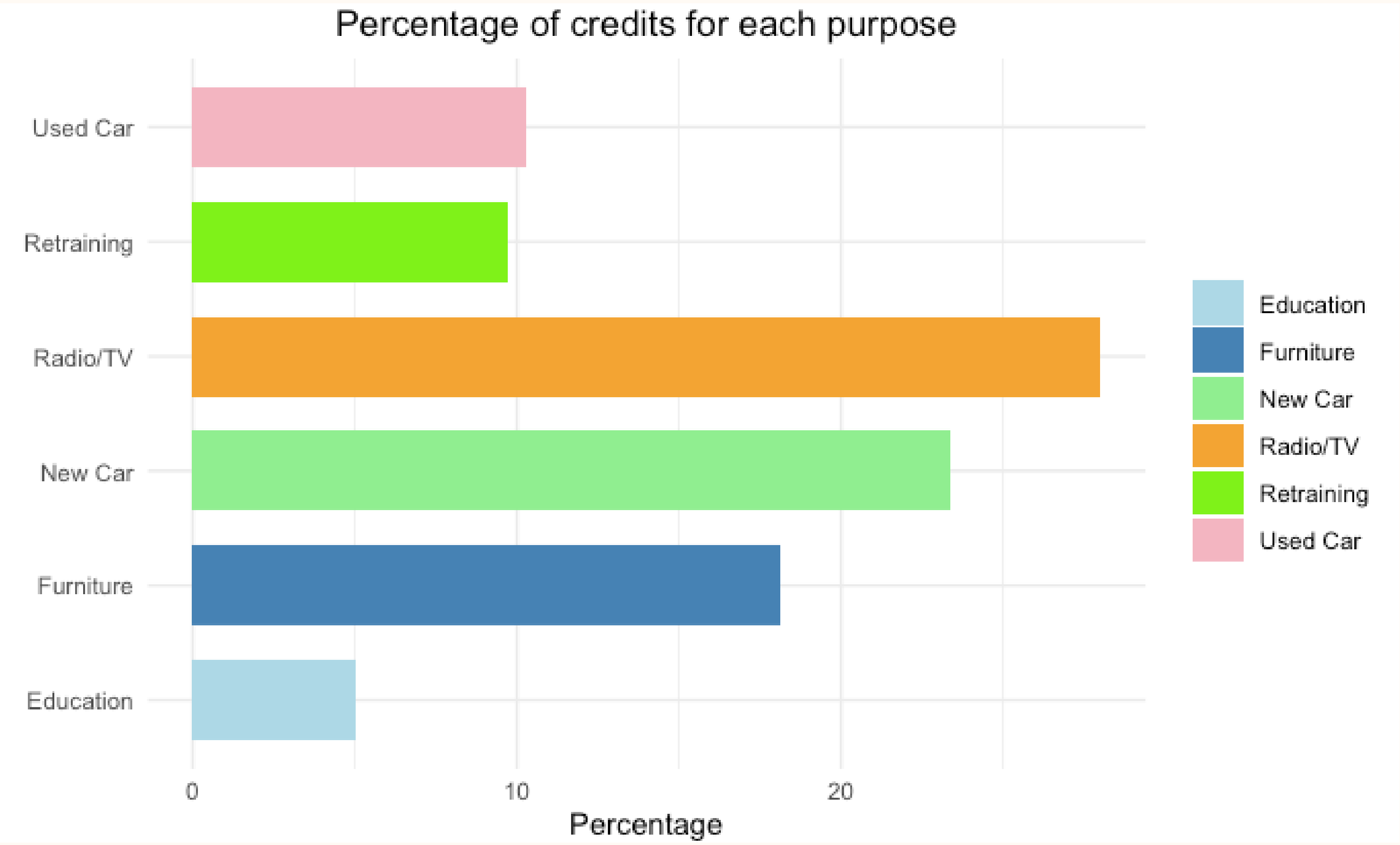
Uncovering Insights for Enhanced Credit Assessment

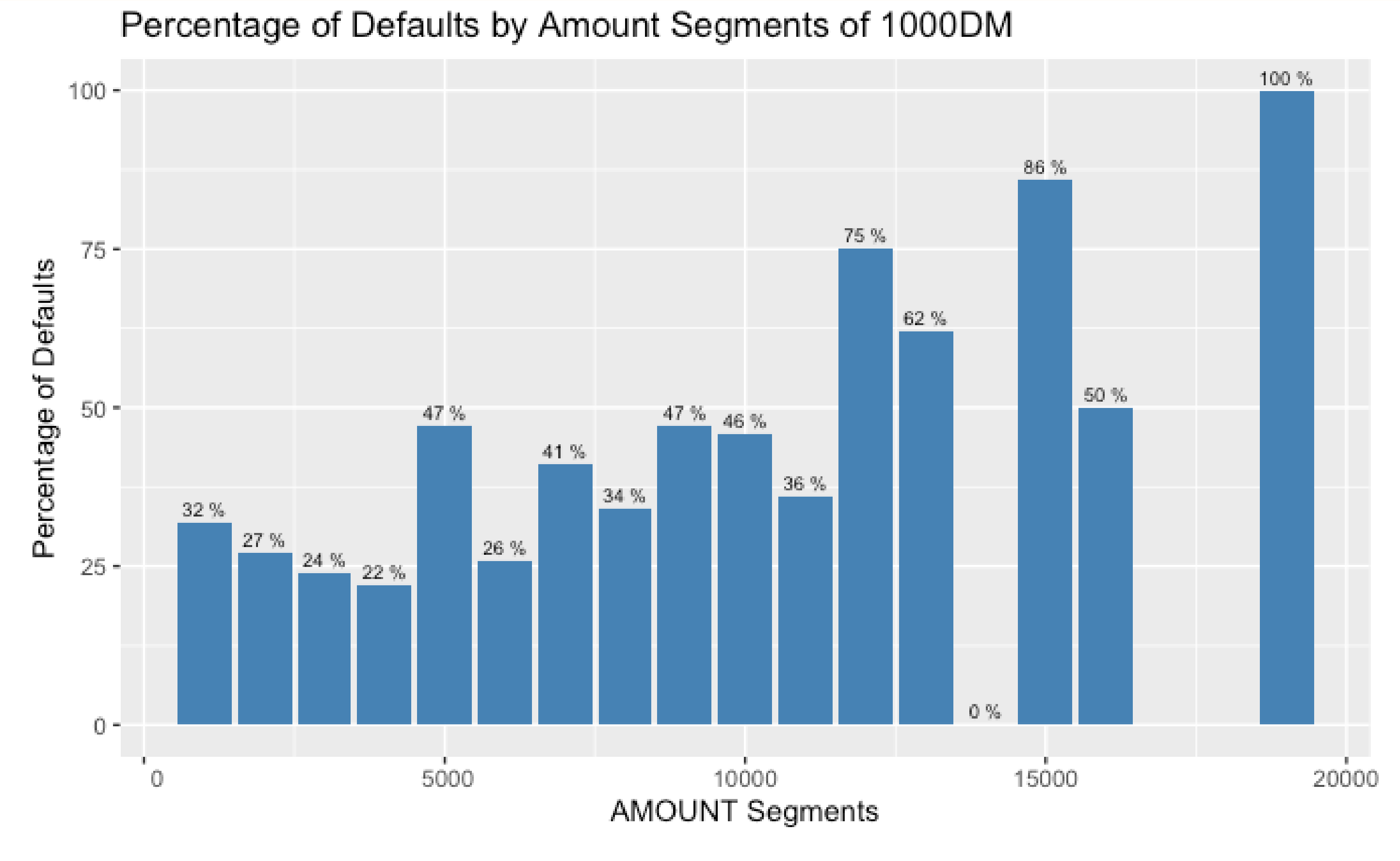
Who are they ?

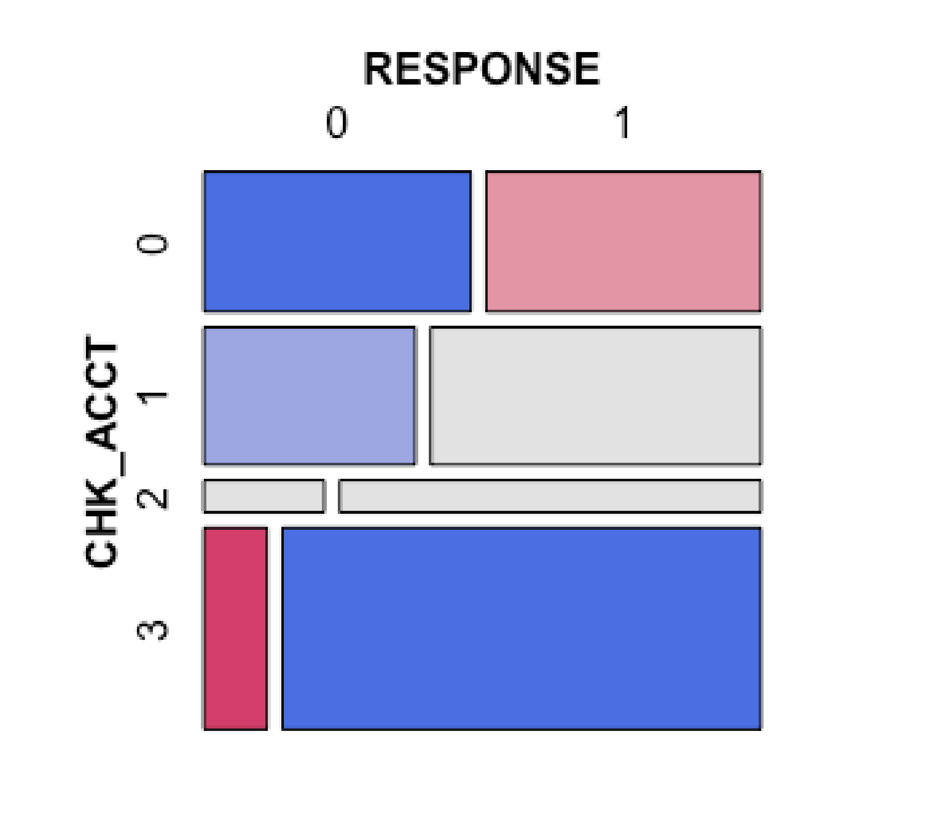
We have analyzed a database of 1000 credit applicants, which encompasses 30 variables, providing us with valuable information and insights into their financial profiles and creditworthiness.

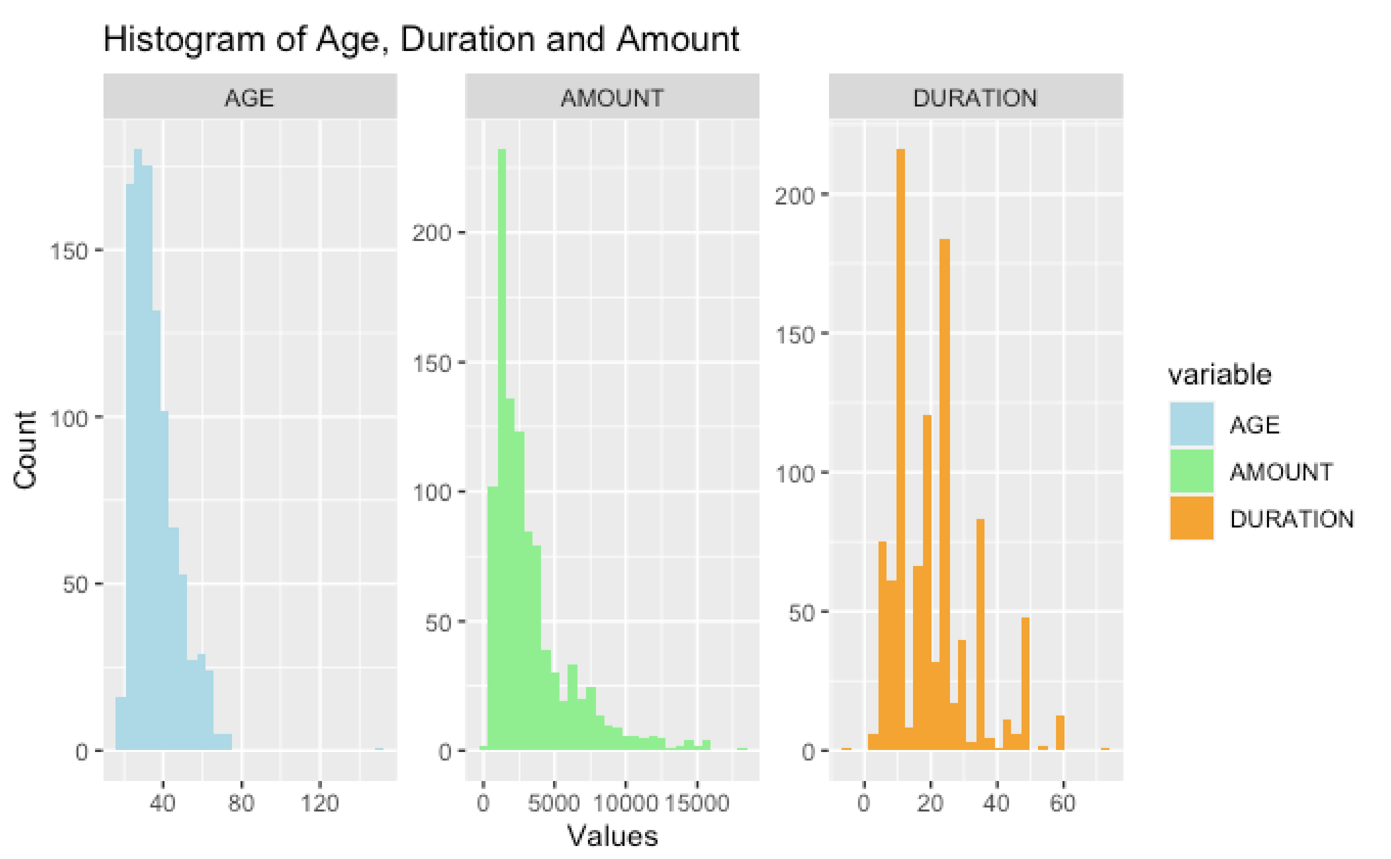
- Number: 300 credit-default applicants
- Average age: 35 years old
- Average credit demand : 3271DM (Deutsche marks)
- Purpose of the credits: mostly for radio/TV and new cars
- Job situation of the majority : skilled employee

- Higher loan amount = higher percentage of default
- 310 female applicants not defined in a proper way
- 700 applicants considered as good credit applicant
- No checking account leads very often to bad credit rating


Percentage of credits for each purpose
Used Car
Retraining
Radio/TV
New Car
Furniture
Education
0
10
20
Percentage
Education
Furniture
New Car
Radio/TV
Retraining
Used Car


Percentage of Defaults by Amount Segments of 1000DM
Percentage of Defaults
100
75
50
25
0
32 %
27 %
24 %
22 %
47 %
26 %
41 %
34 %
47 %
46 %
36 %
75 %
62 %
0 %
86 %
50 %
100 %
0
5000
10000
15000
20000
AMOUNT Segments


RESPONSE
0
1
CHK_ACCT
0
1
2
3


Histogram of Age, Duration and Amount
AGE
AMOUNT
DURATION
Count
Values
variable
AGE
AMOUNT
DURATION

Data Quality for Effective Credit risk Analysis

Data cleaning is vital for model quality. We performed essential modifications such as :

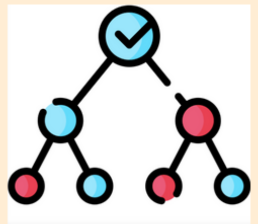
- eliminating outliers (e.g., Age=151)
- removing missing values (14 NA from Age)
- converting categorical variables to factors
- transforming them into binary variables (0 or 1)
- creating a new variable called "female"
- establishing a new dataset named "bank_grouped." In "bank_grouped," variables related to the material and education purposes of credit were appropriately regrouped
- scaling techniques to standardize the variables
- reversed order of categories:

One-hot encoding

Positive 1: client defaulted
Negative 0: no-default

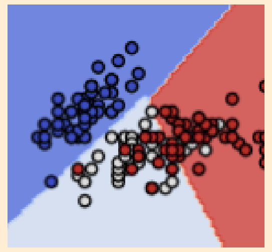
Powerful Models to Identify High-Risk Credit Applicants

Classification tree



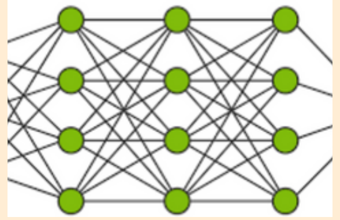
+ Improves by pruning
− Low sensitivity (max 55%)

Support Vector Machine



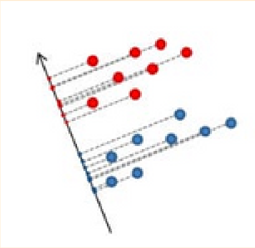
+ Radial basis Kernel
− Linear Kernel

Neural Network



+ Non-linearity
− Interpretation

Linear Discriminant Analysis



+ Normal sensitivity: 62%
− Could not do feature selection

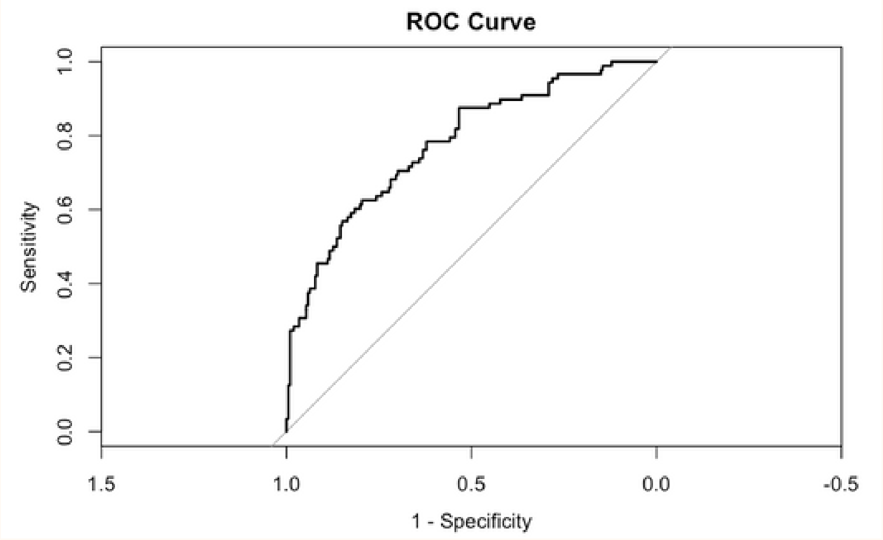
Modeling for variable significance : Key factors in credit risk assessment

Model used → Logistic Regression, Random Forest, Glmnet

- Status of Checking Account: This variable provides insights into the client's financial stability and liquidity, which significantly influences their creditworthiness.
- Duration: The duration of the loan is a key factor in assessing the client's ability to repay their obligations over a specific period.
- Loan Amount: The magnitude of the loan amount serves as an indicator of the client's financial capability and impacts their creditworthiness.

Recommendations

- The best model is the RandomForest: 95% of sensitivity with bank_grouped and without CV and imbalanced
- Top 3 variables
- Increase the threshold: cost of False Negative is high (classifying no-default when the client was default), should lower the threshold to have more positive instances, even if False Positives will increase



Limitations

- Small data set (less than 1000 observations)
- Unbalanced data: almost 70% no-default (No) and 30% default (Yes)

Checking account >= 200 DM

Duration = 3 months

Amount = 3'000DM

Julie

Checking account: none

Duration = 18 months

Amount = 15'000DM

Henry

Credit risk prediction

Thank You For Attention