Case: German Credit Applications

The German Credit data set contains observations on 21 variables for 1000 past applicants for credit. Each applicant was rated as "good credit" (700 cases) or "bad credit" (300 cases). New applicants for credit can also be evaluated on these 21 "predictor" variables. We want to develop a credit scoring rule that can be used to determine if a new applicant is a good credit risk or a bad credit risk, based on values for one or more of the predictor variables.

Data Available

1. CHK_ACCT – This variable gives the Checking Account Status and has been classified as 0, 1, 2 & 3 based on the balance. This is a categorical variable.
2. DURATION – This variable signifies the duration period of the credit taken and is quantitative.l
3. HISTORY – This variable signified past credit history and is categorical in nature. This has been mapped to appropriate numbers.
4. PURPOSE OF CREDIT - This variable signifies the purpose of taking credit and is categorical in nature. This has been mapped to appropriate numbers.
5. AMOUNT – This variable signifies the amount of credit taken and is numerical in nature.
6. SAV_ACCT – This variable signifies the average balance in saving account and is categorical in description. This has been mapped to appropriate numbers as classification.
7. EMPLOYMENT – This variable signifies the present employment of the person who is seeking the credit and is categorical in nature. This is mapped to an appropriate numbers as classification.
8. INSTALLMENT RATE – This is a percentage of the disposable income of the person who is seeking credit and is quantitative in nature.
9. MARTIAL STATUS – This indicates the marital status of the applicant and has been mapped to appropriate numbers
10. CO- APPLICANT – This indicates whether a co-applicant or guarantor exists for the credit. This is binary in nature.
11. PRESENT RESIDENCE – This indicates the number of years the applicant is residing in the given address. This is a categorical variable and is mapped to appropriate numbers as classification.
12. REAL ESTATE – This indicates whether the applicant owns any real estate and is binary in nature.
13. AGE – This indicates the age of the applicant and is a numerical.
14. OTHER INSTALLMENT – This indicates whether the applicant has any other installments and is binary in nature.
15. RESIDENCE – This indicates whether the applicant owns a residence and is binary in nature.
16. NUMBER OF CREDITS – This indicates the number of existing credits with the bank and is numerical in nature.
17. JOB – This indicates the type of job the applicant performs and is categorical in nature and is mapped to numbers as classification.
18. NUMBER OF DEPENDENTS – This indicates the number of people for whom the applicant is liable for maintenance.
19. TELEPHONE – This is binary which indicates whether the applicant has Phone or not.
20. FOREIGN – This is a binary which indicates whether the applicant is a foreigner or not.
21. RESPONSE – The final response is based on the aforementioned twenty inputs and classifies as GOOD or BAD Credit.

All the variables are explained further in Table 1.

Table 1:

S.No	Variable Name	Description	Variable Type	Code Description
1	CHK_ACCT	Checking Account status	Categorical	0: < 0 DM 1: Greater than 0 less than 200 DM 2: Greater than 200 DM 3: No checking Account
2	Duration	Duration of credit in months	Numerical	
3	History	Credit History	Categorical	0: No credits taken 1: All credits at this bank paid back duly 2: Existing credits paid back duly 3: delay in paying off 4: critical account
4	Purpose of Credit	New Car Used Car Furniture Radio/TV Education Retraining	Binary Binary Binary Binary Binary Binary	
5	Amount	Credit Amount	Numerical	
6	Sav_Acct	Average Balance in Savings Account	Categorical	0: < 100 DM 1: Greater than 100 less than 500 DM 2: Greater than 500 and less than 1000 DM 3: Greater than 1000 DM 4: unknown/no savings account
7	Employment	Present employment since	Categorical	0: unemployed 1: less than one year 2: between 1 and 4 years 3: between 4 and 7 years 4: Greater than 7 years
8	Installment Rate	Installment rate as a % of disposable income	numerical	
9	Marital Status	Male and Divorced Male and Single Male Married Female and Divorced	Binary Binary Binary Binary	

		Female and single Female and married	Binary Binary	
10	Co-applicant	Applicant has a co-applicant Applicant has a guarantor	Binary Binary	
11	Present resident	Present resident since	Categorical	0: less than 1 year 1: between 1 and 2 years 2: between 2 and 3 years 3: More than 3 years
12	Real_estate	Applicant owns a real estate Applicant owns no property	Binary Binary	
13	Age	Age in years	Numerical	
14	Other installment	Applicant has other installment plan credit	Binary	
15	Residence	Own Rented	Binary Binary	
16	Num_credits	Number of existing credits at this bank	Numerical	
17	Job	Nature of job	Categorical	0: unemployed/unskilled 1: unskilled/resident 2: skilled employee/official 3: Management/Self-employed/highly qualified/officer
18	Num_dependent	Number of people for whom liable to provide maintenance	Numerical	
19	Telephone	Has telephone connection in his/her name	Binary	
20	Foreign	Foreign worker	Binary	
21	Response	Credit rating good Credit rating bad	Binary Binary	

The consequences of misclassification have been assessed as follows: the costs of a false positive (incorrectly saying an applicant is a good credit risk) outweigh the cost of a false negative (incorrectly saying an applicant is a bad credit risk) by a factor of five.

This can be summarized in the following table.

Table 1.2

	Predicted		
		Good	Bad
	Good	0	100 DM
Actual	Bad	500 DM	0

Opportunity cost table:

Table 1.3

	Predicted		
		Good (Accept)	Bad (Reject)
	Good	100 DM	0
Actual	Bad	-500 DM	0

Case questions

1. Use logistic regression which can help the bank to classify the future applicants into good or bad creditors.
2. How you can use the information given in table 1.2 and 1.3 to decide whether an applicant should be given credit or not.