

German Credit

The German Credit data set contains observations on 30 variables for 1000 past applicants for credit. Each applicant was rated as “good” (700 cases fulfilled terms of credit agreement) or “bad” (300 cases defaulted on loan payments).

New applicants for credit can also be evaluated on the 30 "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. All the variables are explained in the table below.

(Note: The original data set had a number of categorical variables, some of which have been transformed into a series of binary variables so that they can be appropriately handled by XLMiner. Several ordered categorical variables have been left as is, to be treated by XLMiner as numerical. The data have been organized in the spreadsheet German.xls)

Var. #	Variable Name	Description	Variable Type	Code Description
1.	OBS#	Observation No.	Categorical	
2.	CHK_ACCT	Checking account status	Categorical	0 : < 0 DM 1: 0 < ... < 200 DM 2 : => 200 DM 3: unknown
3.	DURATION	Duration of credit in months	Numerical	
4.	HISTORY	Credit history	Categorical	0: no credits taken 1: all credits at this bank paid back duly 2: existing credits paid back duly till now 3: delay in paying off in the past 4: critical account
5.	NEW_CAR	Purpose of credit	Binary	car (new) 0: No, 1: Yes
6.	USED_CAR	Purpose of credit	Binary	car (used) 0: No, 1: Yes
7.	FURNITURE	Purpose of credit	Binary	furniture/equipment 0: No, 1: Yes
8.	RADIO/TV	Purpose of credit	Binary	radio/television 0: No, 1: Yes
9.	EDUCATION	Purpose of credit	Binary	education 0: No, 1: Yes
10.	RETRAINING	Purpose of credit	Binary	retraining 0: No, 1: Yes
11.	AMOUNT	Credit amount	Numerical	
12.	SAV_ACCT	Average balance in savings account	Categorical	0 : < 100 DM 1 : 100<= ... < 500 DM 2 : 500<= ... < 1000 DM 3 : =>1000 DM 4 : unknown
13.	EMPLOYMENT	Present employment since	Categorical	0 : unemployed 1: < 1 year 2 : 1 <= ... < 4 years 3 : 4 <=... < 7 years 4 : >= 7 years
14.	INSTALL_RATE	Installment rate as % of disposable income	Numerical	
15.	MALE_DIV	Applicant is male and divorced	Binary	0: No, 1: Yes
16.	MALE_SINGLE	Applicant is male and single	Binary	0: No, 1: Yes
17.	MALE_MAR	Applicant is male and married or widower	Binary	0: No, 1: Yes

Var. #	Variable Name	Description	Variable Type	Code Description
18.	CO-APPLICANT	Application has a co-applicant	Binary	0: No, 1: Yes
19.	GUARANTOR	Applicant has a guarantor	Binary	0: No, 1: Yes
20.	TIME_RES	Present resident since - years	Categorical	0: <= 1 year 1<...<=2 years 2<...<=3 years 3:>4years
21.	REAL_ESTATE	Applicant owns real estate	Binary	0: No, 1: Yes
22.	PROP_NONE	Applicant owns no property (or unknown)	Binary	0: No, 1: Yes
23.	AGE	Age in years	Numerical	
24.	OTHER_INSTALL	Applicant has other installment plan credit	Binary	0: No, 1: Yes
25.	RENT	Applicant rents	Binary	0: No, 1: Yes
26.	OWN_RES	Applicant owns residence	Binary	0: No, 1: Yes
27.	NUM_CREDITS	Number of existing credits at this bank	Numerical	
28.	JOB	Nature of job	Categorical	0 : unemployed/ unskilled - non-resident 1 : unskilled - resident 2 : skilled employee / official 3 : management/ self-employed/highly qualified employee/ officer
29.	NUM_DEPEND	Number of dependents	Numerical	
30.	TELEPHONE	Applicant has phone in his or her name	Binary	0: No, 1: Yes
31.	FOREIGN	Foreign worker	Binary	0: No, 1: Yes
32.	RESPONSE	Fulfilled terms of credit agreement	Binary	0: No, 1: Yes

The table below shows the values of these variables for the first several records in the case.

OBS#	CHK_ACCT	DURATION	HISTORY	NEW_CAR	USED_CAR	FURNITURE	RADIO/TV	EDUCATION	RETRAINING	AMOUNT	SAV_ACCT	EMPLOYMENT	INSTALL_RATE	MALE_DIV	MALE_SINGLE
1	0	6	4	0	0	0	1	0	0	1169	4	4	4	0	1
2	1	48	2	0	0	0	1	0	0	5951	0	2	2	0	0
3	3	12	4	0	0	0	0	1	0	2096	0	3	2	0	1
4	0	42	2	0	0	1	0	0	0	7882	0	3	2	0	1

MALE_MAR	CO-APPLICANT	GUARANTOR	TIME_RES	REAL_ESTATE	PROP_NONE	AGE	OTHER_INSTALL	RENT	OWN_RES	NUM_CREDITS	JOB	NUM_DEPEND	TELEPHONE	FOREIGN	RESPONSE
0	0	0	4	1	0	67	0	0	1	2	2	1	1	0	1
0	0	0	2	1	0	22	0	0	1	1	2	1	0	0	0
0	0	0	3	1	0	49	0	0	1	1	1	2	0	0	1
0	0	1	4	0	0	45	0	0	0	1	2	2	0	0	1

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 benefits of a true positive (correctly saying an applicant is a good credit risk) by a factor of 5 to 1.5. This can be summarized in the following table.

Opportunity Cost Table (in Deutsch-marks)

	Predicted (Decision)		
		Good (Accept)	Bad (Reject)
Actual	Good	0	150 DM
	Bad	500 DM	0

The opportunity cost table was derived from the average net profit per loan as shown below.

Average Net Profit Table

	Predicted (Decision)		
		Good (Accept)	Bad (Reject)
Actual	Good	150 DM	0
	Bad	-500 DM	0

We'll use this table in assessing the performance of the various models because it is simpler to explain to decision-makers who are used to thinking of their decision in terms of net profits.

In particular, we'll investigate decision trees (homework 5) and neural networks (homework 6) to see if we can improve the profitability of the loans department (currently, the department is running at a loss because too many customers offered loans are defaulting).