

# Description of the German credit dataset

[https://archive.ics.uci.edu/ml/datasets/statlog+\(german+credit+data\)](https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data))

## 1. Title: German Credit data

## 2. Source Information

Professor Dr. Hans Hofmann  
Institut f"ur Statistik und "Okonometrie  
Universit"at Hamburg  
FB Wirtschaftswissenschaften  
Von-Melle-Park 5  
2000 Hamburg 13

## 3. Number of Instances: 1000

Two datasets are provided. The original dataset, in the form provided by Prof. Hofmann, contains categorical/symbolic attributes and is in the file "german.data".

For algorithms that need numerical attributes, Strathclyde University produced the file "german.data-numeric". This file has been edited and several indicator variables added to make it suitable for algorithms which cannot cope with categorical variables. Several attributes that are ordered categorical (such as attribute 17) have been coded as integer. This was the form used by StatLog.

## 4. Number of Attributes german: 20 (7 numerical, 13 categorical)

Number of Attributes german.numer: 24 (24 numerical)

## 5. Attribute description for german (modified by Eliana Lambrou to match *GermanCreditData.csv*)

### Attribute 1: (qualitative)

*Status of existing checking account*

- 1: ... < 0 DM
- 2: 0 <= ... < 200 DM
- 3: ... >= 200 DM /salary assignments for at least 1 year
- 4: no checking account

### Attribute 2: (numerical)

*Duration in month*

### Attribute 3: (qualitative)

*Credit history*

- 0: no credits taken/all credits paid back duly
- 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/other credits existing (not at this bank)

### Attribute 4: (qualitative)

*Purpose*

- 0: car (new)

- 1: car (used)
- 2: furniture/equipment
- 3: radio/television
- 4: domestic appliances
- 5: repairs
- 6: education
- 7: (vacation - does not exist?)
- 8: retraining
- 9: business
- 10: others

**Attribute 5: (numerical)**

*Credit amount*

**Attribute 6: (qualitative)**

*Savings account/bonds*

- 1: ... < 100 DM
- 2: 100 <= ... < 500 DM
- 3: 500 <= ... < 1000 DM
- 4: .. >= 1000 DM
- 5: unknown/ no savings account

**Attribute 7: (qualitative)**

*Present employment since*

- 1: unemployed
- 2: ... < 1 year
- 3: 1 <= ... < 4 years
- 4: 4 <= ... < 7 years
- 5: .. >= 7 years

**Attribute 8: (numerical)**

*Instalment rate in percentage of disposable income*

**Attribute 9: (qualitative)**

*Personal status and sex*

- 1: male: divorced/separated
- 2: female: divorced/separated/married
- 3: male: single
- 4: male: married/widowed
- 5: female: single

**Attribute 10: (qualitative)**

*Other debtors / guarantors*

- 1: none
- 2: co-applicant
- 3: guarantor

**Attribute 11: (numerical)**

*Present residence since*

**Attribute 12: (qualitative)**

*Property*

- 1: real estate
- 2: if not 1 : building society savings agreement/life insurance
- 3: if not 1/2 : car or other, not in attribute 6
- 4: unknown / no property

**Attribute 13: (numerical)**

*Age in years*

**Attribute 14: (qualitative)**

*Other instalment plans*

- 1: bank
- 2: stores
- 3: none

**Attribute 15: (qualitative)**

*Housing*

- 1: rent
- 2: own
- 3: for free

**Attribute 16: (numerical)**

*Number of existing credits at this bank*

**Attribute 17: (qualitative)**

*Job*

- 1: unemployed/unskilled - non-resident
- 2: unskilled - resident
- 3: skilled employee / official
- 4: management/ self-employed/ highly qualified employee/ officer

**Attribute 18: (numerical)**

*Number of people being liable to provide maintenance for*

**Attribute 19: (qualitative)**

*Telephone*

- 1: none
- 2: yes, registered under the customers name

**Attribute 20: (qualitative)**

*foreign worker*

- 1: yes
- 2: no

## 6. Cost Matrix

This dataset requires use of a cost matrix (see below)

	1	2
1	0	1
2	5	0

(1 = Good, 2 = Bad)

the rows represent the actual classification and the columns the predicted classification.

It is worse to class a customer as good when they are bad (5), than it is to class a customer as bad when they are good (1).