

Statlog (German Credit Data) Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: This dataset classifies people described by a set of attributes as good or bad credit risks. Comes in two formats (one all numeric). Also comes with a cost matrix

Data Set Characteristics:	Multivariate	Number of Instances:	1000	Area:	Financial
Attribute Characteristics:	Categorical, Integer	Number of Attributes:	20	Date Donated	1994-11-17
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	706435

Source

Professor Dr. Hans Hofmann
Institut für Statistik und Ökonometrie
Universität Hamburg
FB Wirtschaftswissenschaften
Von-Melle-Park 5
2000 Hamburg 13

Data Set Information

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.

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

		Predicted	
		1	2
Actual	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).

Attribute Information

Attribute 1: (qualitative)

Status of existing checking account

A11 : ... < 0 DM

A12 : 0 <= ... < 200 DM

A13 : ... >= 200 DM / salary assignments for at least 1 year

A14 : no checking account

Attribute 2: (numerical)

Duration in month

Attribute 3: (qualitative)

Credit history

A30 : no credits taken/ all credits paid back duly

A31 : all credits at this bank paid back duly

A32 : existing credits paid back duly till now

A33 : delay in paying off in the past

A34 : critical account/ other credits existing (not at this bank)

Attribute 4: (qualitative)

Purpose

A40 : car (new)

A41 : car (used)

A42 : furniture/equipment

A43 : radio/television

A44 : domestic appliances

A45 : repairs

A46 : education

A47 : (vacation - does not exist?)

A48 : retraining

A49 : business

A410 : others

Attribute 5: (numerical)

Credit amount

Attribute 6: (qualitative)

Savings account/bonds

A61 : ... < 100 DM

A62 : 100 <= ... < 500 DM

A63 : 500 <= ... < 1000 DM

A64 : .. >= 1000 DM

A65 : unknown/ no savings account

Attribute 7: (qualitative)

Present employment since

A71 : unemployed

A72 : ... < 1 year

A73 : 1 <= ... < 4 years

A74 : 4 <= ... < 7 years

A75 : .. >= 7 years

Attribute 8: (numerical)

Installment rate in percentage of disposable income

Attribute 9: (qualitative)

Personal status and sex

A91 : male : divorced/separated

A92 : female : divorced/separated/married

A93 : male : single

A94 : male : married/widowed

A95 : female : single

Attribute 10: (qualitative)
Other debtors / guarantors
A101 : none
A102 : co-applicant
A103 : guarantor

Attribute 11: (numerical)
Present residence since

Attribute 12: (qualitative)
Property
A121 : real estate
A122 : if not A121 : building society savings agreement/ life insurance
A123 : if not A121/A122 : car or other, not in attribute 6
A124 : unknown / no property

Attribute 13: (numerical)
Age in years

Attribute 14: (qualitative)
Other installment plans
A141 : bank
A142 : stores
A143 : none

Attribute 15: (qualitative)
Housing
A151 : rent
A152 : own
A153 : for free

Attribute 16: (numerical)
Number of existing credits at this bank

Attribute 17: (qualitative)
Job
A171 : unemployed/ unskilled - non-resident
A172 : unskilled - resident
A173 : skilled employee / official
A174 : management/ self-employed/
highly qualified employee/ officer

Attribute 18: (numerical)
Number of people being liable to provide maintenance for

Attribute 19: (qualitative)
Telephone
A191 : none
A192 : yes, registered under the customers name

Attribute 20: (qualitative)
foreign worker
A201 : yes
A202 : no

Relevant Papers:

N/A

Papers That Cite This Data Set¹:

Jeroen Eggermont and Joost N. Kok and Walter A. Kusters. [Genetic Programming for data classification: partitioning the search space](#). SAC. 2004. [[View Context](#)].

Ke Wang and Shiyu Zhou and Ada Wai-Chee Fu and Jeffrey Xu Yu. [Mining Changes of Classification by Correspondence Tracing](#). SDM. 2003. [[View Context](#)].

Avelino J. Gonzalez and Lawrence B. Holder and Diane J. Cook. [Graph-Based Concept Learning](#). FLAIRS Conference. 2001. [[View Context](#)].

Oya Ekin and Peter L. Hammer and Alexander Kogan and Pawel Winter. [Distance-Based Classification Methods](#). e p o r t RUTCOR ffl Rutgers Center for Operations Research ffl Rutgers University. 1996. [[View Context](#)].

Chotirat Ann and Dimitrios Gunopulos. [Scaling up the Naive Bayesian Classifier: Using Decision Trees for Feature Selection](#). Computer Science Department University of California. [[View Context](#)].

Paul O' Dea and David Griffith and Colm O' Riordan. [DEPARTMENT OF INFORMATION TECHNOLOGY](#). P. O'Dea (NUI). [[View Context](#)].

Paul O' Dea and Josephine Griffith and Colm O' Riordan. [Combining Feature Selection and Neural Networks for Solving Classification Problems](#). Information Technology Department, National University of Ireland. [[View Context](#)].

Citation Request:

Please refer to the Machine Learning Repository's [citation policy](#)