

credit



PREREQUISITES

MACHINE LEARNING CASE STUDY: CREDIT DATA SET

- German Credit Dataset is a research dataset from the University of Hamburg from 1994 and donated by Prof. Hans Hoffman.
- Description (and manual download) at the UCI repository.

Importing the Data

We use OpenML (R-Package) to download the dataset in a machine-readable format and convert it into a `data.frame`:

```
# load the dataset from OpenML Library  
d = OpenML::getOMLDataSet(data.id = 31)  
  
# convert the OpenML object to a regular data.frame  
credit = as.data.frame(d)
```

EXPLORATORY DATA ANALYSIS

Factor Variables

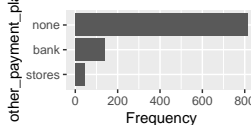
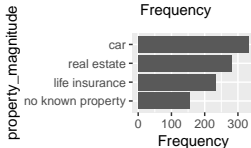
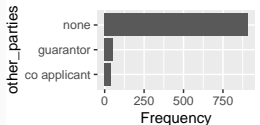
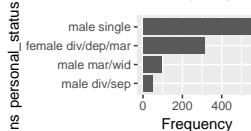
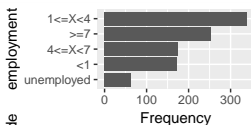
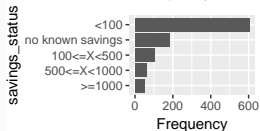
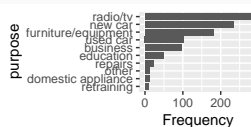
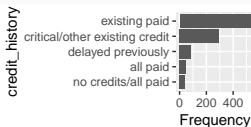
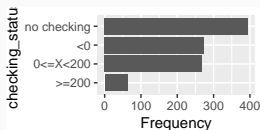
```
skimr::skim_to_list(credit) %>%  
  .$factor %>%  
  knitr::kable(format = 'latex', booktabs = TRUE) %>%  
  kableExtra::kable_styling(latex_options = 'HOLD_position', font_size = 5)
```

variable	missing	n_unique	top_counts
checking_status	0	4	no : 394, <0: 274, 0<=: 269, >=2: 63
class	0	2	goo: 700, bad: 300, NA: 0
credit_history	0	5	exi: 530, cri: 293, del: 88, all: 49
employment	0	5	1<=: 339, >=7: 253, 4<=: 174, <1: 172
foreign_worker	0	2	yes: 963, no: 37, NA: 0
housing	0	3	own: 713, ren: 179, for: 108, NA: 0
job	0	4	ski: 630, uns: 200, hig: 148, une: 22
other_parties	0	3	non: 907, gua: 52, co : 41, NA: 0
other_payment_plans	0	3	non: 814, ban: 139, sto: 47, NA: 0
own_telephone	0	2	non: 596, yes: 404, NA: 0
personal_status	0	4	mal: 548, fem: 310, mal: 92, mal: 50
property_magnitude	0	4	car: 332, rea: 282, lif: 232, no : 154
purpose	0	10	rad: 280, new: 234, fur: 181, use: 103
savings_status	0	5	<10: 603, no : 183, 100: 103, 500: 63

EXPLORATORY DATA ANALYSIS

Barplots of discrete features

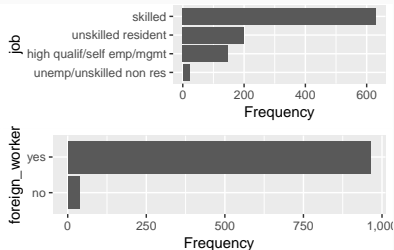
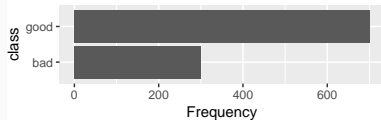
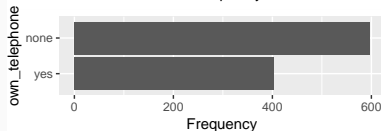
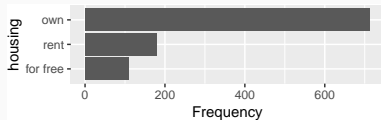
```
plot_bar(credit[, 1:14])
```



EXPLORATORY DATA ANALYSIS

Barplots of discrete features

```
plot_bar(credit[, 15:21])
```



Numerical Variables

```
skimr::skim_to_list(credit) %>%  
  .$numeric %>%  
  knitr::kable(format = 'latex', booktabs = TRUE) %>%  
  kableExtra::kable_styling(latex_options = 'HOLD_position', font_size = 6)
```

variable	missing	mean	sd
age	0	35.55	11.38
credit_amount	0	3271.26	2822.74
duration	0	20.9	12.06
existing_credits	0	1.41	0.58
installment_commitment	0	2.97	1.12
num_dependents	0	1.16	0.36
residence_since	0	2.85	1.1

EXPLORATORY DATA ANALYSIS

Histograms of numerical features

```
plot_histogram(credit)
```

