Data Analytics for Decision Making Homework : Proposal 2

Description and relevant informations

a) Title : credit data ; name of the file : credit.csv

b) Abstract :

The credit data has data on 1000 past credit applicants, described by 30 variables. Each applicant is rated as "Good" or "Bad" credit (encoded as 1 and 0 respectively in the response variable). We want to obtain a model that may be used to determine if new applicants present a good or bad credit risk. This dataset was donated by Professor Dr. Hans Hofmann.

c) Number of instances : 1000

d) Number of attributes : 30

e) Attribute Information :

Var.#	Variable Name	Description	Variable Type	Description
1.	OBS#	Observation No.	Categorical	
2.	CHK_ACCT	Checking account status	Categorical	$0 : < 0\,\text{DM}$ $1 : 0 < \cdots < 200\,\text{DM}$ $2 : \geq 200\,\text{DM}$ 3 : no checking account
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	$0 : < 100\,\text{DM}$ $1 : 100 \leq \cdots < 500\,\text{DM}$ $2 : 500 \leq \cdots < 1000\,\text{DM}$ $3 : \geq 1000\,\text{DM}$ 4 : unknown/no savings account
13.	EMPLOYMENT	Present employment since	Categorical	0 : unemployed $1 : < 1\,\text{year}$ $2 : 1 \leq \cdots < 4\,\text{years}$ $3 : 4 \leq \cdots < 7\,\text{years}$ $4 : \geq 7\,\text{years}$
14.	INSTALL_RATE	Installment rate as % of disposable income	Numerical	
15.	MALE_DIV	Applicant is male and divorced	Binary	0 : No, 1 : Yes

Var.#	Variable Name	Description	Variable Type	Description
16.	MALE_SINGLE	Applicant is male and single	Binary	0 : No, 1 : Yes
17.	MALE_MAR_WID	Applicant is male and married or a widower	Binary	0 : No, 1 : Yes
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.	PRESENT_RESIDENT	Present resident since - years	Categorical	$0 : \leq 1$ year $1 : 1 < \cdots \leq 2$ years $2 : 2 < \cdots \leq 3$ years $3 : > 4$ years
21.	REAL_ESTATE	Applicant owns real estate	Binary	0 : No, 1 : Yes
22.	PROP_UNKN_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_DEPENDENTS	Number of people for whom liable to provide maintenance	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

f) Missing attribute values : no

g) Response variable : credit rating is good
   1. 0 : No
   2. 1 : Yes

h) **Goal** : we want to obtain a model that may be used to determine if new applicants present a good or bad credit risk

You must use the CRISP-DM model to solve this problem. In addition to the models described in the course, each group will have to study another particular model. In the phase of Data understanding, if you think it makes sense to introduce variables to summarise a group of variables, you can do so with full justification.