BUILDING A MODEL TO PREDICT THE STATUS OF CREDIT

Data Source:
UCI Machine Learning Repository, "South German Credit" Dataset:
https://archive.ics.uci.edu/ml/datasets/South+German+Credit+%28UPDATE%29

Source:
Ulrike Grömping
Beuth University of Applied Sciences Berlin
Website with contact information: https://prof.beuth-hochschule.de/groemping/

Objective
 The objective of this analysis is to build a model to predict the status of credit (good or bad)

Variables:

Variable name: status
Content: status of the debtor's checking account with the bank (categorical)

Variable name: duration
Content: credit duration in months (quantitative)

Variable name: credit_history
Content: history of compliance with previous or concurrent credit contracts (categorical)

Variable name: purpose
Content: purpose for which the credit is needed (categorical)

Variable name: amount
Content: credit amount in DM (quantitative; result of monotonic transformation; actual data and type of transformation unknown)

Variable name: savings
Content: debtor's savings (categorical)

Variable name: employment_duration
Content: duration of debtor's employment with current employer (ordinal; discretized quantitative)

Variable name: installment_rate
Content: credit installments as a percentage of debtor's disposable income (ordinal; discretized quantitative)

Variable name: personal_status_sex
Content: combined information on sex and marital status; categorical; sex cannot be recovered from the variable, because male singles and female non-singles are coded with the same code (2); female widows cannot be easily classified, because the code table does not list them in any of the female categories

Variable name: other_debtors
Content: Is there another debtor or a guarantor for the credit? (categorical)

Variable name: present_residence
Content: length of time (in years) the debtor lives in the present residence (ordinal; discretized quantitative)

Variable name: property
Content: the debtor's most valuable property, i.e. the highest possible code is used. Code 2 is used, if codes 3 or 4 are not applicable and there is a car or any other relevant property that does not fall under variable sparkont. (ordinal)

Variable name: age
Content: age in years (quantitative)

Variable name: other_installment_plans
Content: installment plans from providers other than the credit-giving bank (categorical)

Variable name: housing
Content: type of housing the debtor lives in (categorical)

Variable name: number_credits
Content: number of credits including the current one the debtor has (or had) at this bank (ordinal, discretized quantitative); contrary to Fahrmeir and HamerleÃ¢‚¬â„¢s (1984) statement, the original data values are not available.

Variable name: job
Content: quality of debtor's job (ordinal)

Variable name: people_liable
Content: number of persons who financially depend on the debtor (i.e., are entitled to maintenance) (binary, discretized quantitative)

Variable name: telephone
Content: Is there a telephone landline registered on the debtor's name? (binary; remember that the data are from the 1970s)

Variable name: foreign_worker
Content: Is the debtor a foreign worker? (binary)

Variable name: credit_risk
Content: Has the credit contract been complied with (good) or not (bad) ? (binary)

Relevant Papers:

Fahrmeir, L. and Hamerle, A. (1981, in German). Kategoriale Regression in der betrieblichen Planung. *Zeitschrift für Operations Research* **25**, B63-B78.

Fahrmeir, L. and Hamerle, A. (1984, in German). *Multivariate Statistische Verfahren* (1st ed., Ch.8 and Appendix C). De Gruyter, Berlin.

Grömping, U. (2019). South German Credit Data: Correcting a Widely Used Data Set. Report 4/2019, Reports in Mathematics, Physics and Chemistry, Department II, Beuth University of Applied Sciences Berlin. URL: [[Web Link]].

Häußler, W.M. (1979, in German). Empirische Ergebnisse zu Diskriminationsverfahren bei Kreditscoringsystemen. *Zeitschrift für Operations Research* **23**, B191-B210.

Hofmann, H.J. (1990, in German). Die Anwendung des CART-Verfahrens zur statistischen Bonitätsanalyse von Konsumentenkrediten. *Zeitschrift für Betriebswirtschaft* **60**, 941-962.

Open data LMU (2010; accessed Nov 27 2019; in German). Kreditscoring zur Klassifikation von Kreditnehmern. URL: [[Web Link]].