CREDIT RISK ASSESSMENT

Description: The business of banks is making loans. Assessing the credit worthiness of an applicant is of crucial importance. You have to develop a system to help a loan officer decide whether the credit of a customer is good, or bad. A bank's business rules regarding loans must consider two opposing factors. On the one hand, a bank wants to make as many loans as possible. Interest on these loans is the ban's profit source. On the other hand, a bank cannot afford to make too many bad loans. Too many bad loans could lead to the collapse of the bank. The bank's loan policy must involve a compromise not too strict, and not too lenient.

To do the assignment, you first and foremost need some knowledge about the world of credit. You can acquire such knowledge in a number of ways.

1. Knowledge Engineering. Find a loan officer who is willing to talk. Interview her and try to represent her knowledge in the form of production rules.
2. Books. Find some training manuals for loan officers or perhaps a suitable textbook on finance. Translate this knowledge from text form to production rule form.
3. Common sense. Imagine yourself as a loan officer and make up reasonable rules which can be used to judge the credit worthiness of a loan applicant.
4. Case histories. Find records of actual cases where competent loan officers correctly judged when not to, approve a loan application.

The German Credit Data:

Actual historical credit data is not always easy to come by because of confidentiality rules. Here is one such dataset (original) Excel spreadsheet version of the German credit data (download from web).

In spite of the fact that the data is German, you should probably make use of it for this assignment, (Unless you really can consult a real loan officer!)

A few notes on the German dataset:

- DM stands for Deutsche Mark, the unit of currency, worth about 90 cents Canadian (but looks and acts like a quarter).
- Owns_telephone. German phone rates are much higher than in Canada so fewer people own telephones.
- Foreign_worker. There are millions of these in Germany (many from Turkey). It is very hard to get German citizenship if you were not born of German parents.
- There are 20 attributes used in judging a loan applicant. The goal is the classify the applicant into one of two categories, good or bad.

Description of the German credit dataset.

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)
   Attribute description for German

Attribute description for German

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 customer's name

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

Relabeled values of attribute

Relabeled values in attribute checking_status

From: A11	To: '<0'
From: A12	To: '0<=X<200'
From: A13	To: '>=200'
From: A14	To: 'no checking'

Relabeled values in attribute credit_history

From: A30	To: 'no credits/all paid'
From: A31	To: 'all paid'
From: A32	To: 'existing paid'
From: A33	To: 'delayed previously'
From: A34	To: 'critical/other existing credit'

Relabeled values in attribute purpose

From: A40	To: 'new car'
From: A41	To: 'used car'
From: A42	To: furniture/equipment
From: A43	To: radio/tv
From: A44	To: 'domestic appliance'
From: A45	To: repairs
From: A46	To: education
From: A47	To: vacation
From: A48	To: retraining
From: A49	To: business
From: A410	To: other

Relabeled values in attribute savings_status

From: A61	To: '<100'
From: A62	To: '100<=X<500'
From: A63	To: '500<=X<1000'
From: A64	To: '>=1000'
From: A65	To: 'no known savings'

Relabeled values in attribute employment

From: A71	To: unemployed
From: A72	To: '<1'
From: A73	To: '1<=X<4'
From: A74	To: '4<=X<7'
From: A75	To: '>=7'

Relabeled values in attribute personal_status

From: A91	To: 'male div/sep'
From: A92	To: 'female div/dep/mar'
From: A93	To: 'male single'
From: A94	To: 'male mar/wid'
From: A95	To: 'female single'

elabeled values in attribute other_parties

From: A101	To: none
From: A102	To: 'co applicant'
From: A103	To: guarantor

Relabeled values in attribute property_magnitude

From: A121	To: 'real estate'

From: A122	To: 'life insurance'
From: A123	To: car
From: A124	To: 'no known property'

Relabeled values in attribute other_payment_plans

From: A141	To: bank
From: A142	To: stores
From: A143	To: none

Relabeled values in attribute housing

From: A151	To: rent
From: A152	To: own
From: A153	To: 'for free'

Relabeled values in attribute job

From: A171	To: 'unemp/unskilled non res'
From: A172	To: 'unskilled resident'
From: A173	To: skilled
From: A174	To: 'high qualif/self emp/mgmt'

Relabeled values in attribute own_telephone

From: A191	To: none
From: A192	To: yes

Relabeled values in attribute foreign_worker

From: A201	To: yes
From: A202	To: no

Relabeled values in attribute class

From: 1	To: good
From: 2	To: bad