# Aprendizagem Computacional *

banking - case description

Task description

The bank wants to improve their services. For instance, the bank managers have only vague idea, who is a good client (whom to offer some additional services) and who is a bad client (whom to watch carefully to minimize the bank loses). Fortunately, the bank stores data about their clients, the accounts (transactions within several months), the loans already granted, the credit cards issued. The bank managers hope to improve their understanding of customers and seek specific actions to improve services. A mere application of a discovery tool will not be convincing for them.

To test a data mining approach to help the bank managers, it was decided to address two problems, a descriptive and a predictive one. While the descriptive problem was left open, the predictive problem is the prediction of whether a loan will end successfuly.

Data description
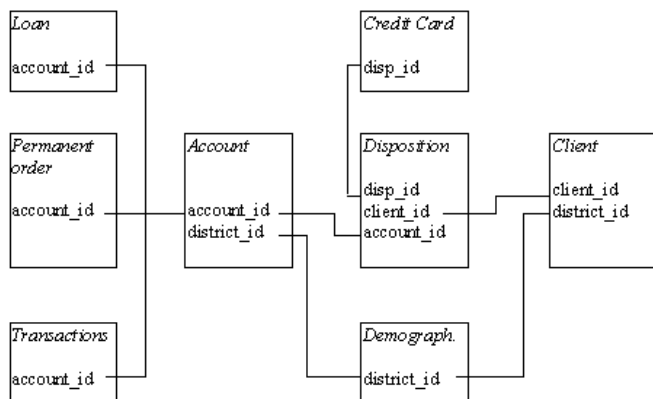
The data about the clients and their accounts consist of following relations:

- relation **account** (4500 objects – each record describes static characteristics of an account,
- relation **client** (5369 objects) – each record describes characteristics of a client,
- relation **disposition** (5369 objects) – each record relates together a client with an account i.e. this relation describes the rights of clients to operate accounts,
- relation **permanent order** (6471 objects) – each record describes characteristics of a payment order,
- relation **transaction** (1056320 objects) – each record describes one transaction on an account,
- relation **loan** (682 objects) – each record describes a loan granted for a given account,
- relation **credit card** (892 objects) – each record describes a credit card issued to an account,
- relation **demographic data** (77 objects) – each record describes demographic characteristics of a district.

Each account has both static characteristics (e.g. date of creation, address of the branch) given in relation "account" and dynamic characteristics (e.g. payments debited or credited, balances) given in relations "permanent order" and "transaction". Relation "client" describes characteristics of persons who can manipulate with the accounts. One client can have more accounts, more clients can manipulate with single account; clients and accounts are related together in relation "disposition". Relations "loan" and "credit card" describe some services which the bank offers to its clients; more credit cards can be issued to an account, at most one loan can be granted for an account. Relation "demographic data" gives some publicly available information about the districts (e.g. the unemployment rate); additional information about the clients can be deduced from this.

Relation account

item	meaning	remark
account_id	identification of the account	
district_id	location of the branch	
date	date of creating of the account	in the form YYMMDD
frequency	frequency of issuance of statements	

Relation client

item	meaning	remark
client_id	client identifier	
birth number	birthday and sex	the number is in the form YYMMDD for men, the number is in the form YYMM+50DD for women, where YYMMDD is the date of birth
district_id	address of the client	

Relation disposition

item	meaning	remark
disp_id	record identifier	
client_id	identification of a client	
account_id	identification of an account	
type	type of disposition (owner/user)	only owner can issue permanent orders and ask for a loan

Relation permanent order (debits only)

item	meaning	remark
order_id	record identifier	
account_id	account, the order is issued for	
bank_to	bank of the recipient	each bank has unique two-letter code
account_to	account of the recipient	
amount	debited amount	
K_symbol	characterization of the payment	

Relation Transaction

item	meaning	remark
trans_id	record identifier	
account_id	account, the transation deals with	
date	date of transaction	in the form YYMMDD
type	+/- transaction	
operation	mode of transaction	
amount	amount of money	
balance	balance after transaction	
k_symbol	characterization of the transaction	

bank	bank of the partner	each bank has unique two-letter code
account	account of the partner	

Relation Loan

item	meaning	remark
loan_id	record identifier	
account_id	identification of the account	
date	date when the loan was granted	in the form YYMMDD
amount	amount of money	
duration	duration of the loan	
payments	monthly payments	
status	status of paying off the loan	'A' stands for contract finished, no problems, 'B' stands for contract finished, loan not payed, 'C' stands for running contract, OK so far, 'D' stands for running contract, client in debt

Relation Credit card

item	meaning	remark
card_id	record identifier	
disp_id	disposition to an account	
type	type of card	possible values are "junior", "classic", "gold"
issued	issue date	in the form YYMMDD

Relation Demographic data

item	meaning	remark
A1 = district_id	district code	
A2	district name	
A3	region	
A4	no. of inhabitants	
A5	no. of municipalities with inhabitants < 499	
A6	no. of municipalities with inhabitants 500-1999	
A7	no. of municipalities with inhabitants 2000-9999	
A8	no. of municipalities with inhabitants >10000	
A9	no. of cities	
A10	ratio of urban inhabitants	
A11	average salary	
A12	unemploymant rate '95	
A13	unemploymant rate '96	
A14	no. of enterpreneurs per 1000 inhabitants	
A15	no. of commited crimes '95	
A16	no. of commited crimes '96	