Early Prepayment of Mortgage Loans

Win.
The Right
Way.
Together.

UniCredit

OVERVIEW

Data: The data provided for this use-case is organized in three files:

- **balances:** Contains dynamic (monthly) information on the outstanding exposure to each mortgage loan, remaining maturity, the customer total exposure with the Bank, etc. Further, indication of full prepayment of the loan is included. Partial prepayment events could also be derived from information in this file.
- **customers**: Contains static information on certain customer's characteristics as recorded at the inception of the loan.
- **collateral**: Contains static information on certain characteristics of the real estate property pledged as collateral recorded at the inception of the loan.

Any additional relevant data could be added at the modeler's discretion – market interest rates, inflation, level of wages, real estate prices, etc.

Problem: Build a model that, given new input data (as above) could predict full (and possibly partial) prepayments of mortgage loans over a certain time horizon (i.e. 1 year, 6 months, 3 months).

DATA

File balances.csv

- reporting_date: end-of-month for which the data in the below fields are reported
- customer_id
- total_cust_exposure: total exposure to the customer across all products active at the reporting date (in BGN)
- overdraft: number of active overdrafts in the current month
- consumer_loan: number of active consumer loans in the current month
- credit_card: number of active credit cards in the current month
- customer_rating: the borrower rating according to the Bank's internal model (lower value indicates better creditworthiness)
- account_id: the mortgage loan id
- open_date: start date of the contract
- maturity_date: end date of the contract
- original_principal: the contractual granted amount of the loan (in BGN)
- current_principal: the remaining, unpaid part of the loan as of the current month (in BGN)
- interest_rate: interests rate of the loan
- prepayment_status: indicates whether the loan was fully prepaid and the source of funds used for prepayment; this value is not month-specific and it is given for all reporting dates for which info on this loan exists (this filed will be examined further below)

File customers.csv

- customer_id
- age*: encoded age ranges
- maritial_status*: encoded marital status
- town: customer's town of residence
- empl_type*: encoded employment type (i.e. self-employed)
- profession*: encoded profession
- tot_mnth_income: total monthly income of the customer (salary + other income)
- gross_salary: customer's monthly gross salary

File collateral.csv

- account_id: the mortgage loan id
- collateral_id: the id of the residential real estate property pledged as collateral
- market_value: market value of the property (in BGN)
- collateral_type*: encoded type of collateral (i.e. flat)
- building_type*: encoded building type
- mortgage_utilization*: encoded use of the property
- town: town where the property is located
- area: area of the property in m^2

* description of encoded fields is not provided due to data protection requirements

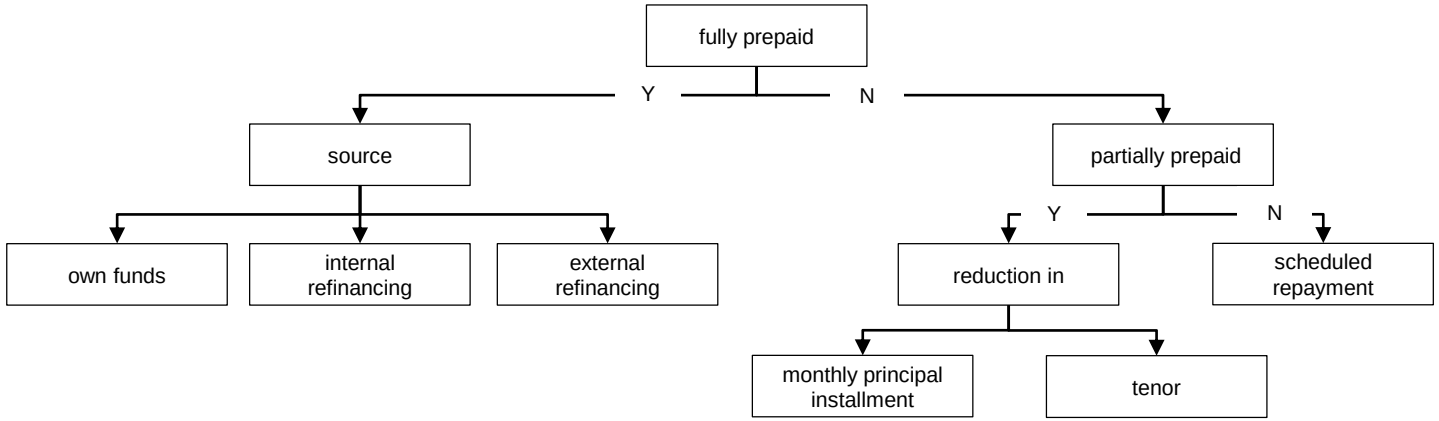
TARGET DEFINITION 1/2

The target should be defined using the field “prepayment_status” from the “balances” table.

This filed contains the following possible values:

- **no**: The loan is not fully prepaid prior to its maturity. These loans are:
  - still active (being repaid regularly according to their repayment plans) or already repaid (as expected) with the last installment on maturity;
  - there has been a partial early prepayment which resulted either in:
    - reduction in the tenor (could be identified when the balance drops suddenly in a given month and the maturity_date changes), or
    - reduction in the monthly installment amount (could be identified when the balance drops suddenly in a given month and the maturity_date does not change)
- **own**: The loans was fully prepaid with the client’s own funds.
- **refinanced-UC**: The loans was fully prepaid via refinancing with another loan by the Bank.
- **refinanced other banks**: The loans was fully prepaid via refinancing with a loan by another bank.
- There are also combinations of the above sources of prepayment (i.e. own funds + loan by another bank); these are relatively rare and we leave it to the modeler to decide how to handle them (exclude, overwrite, etc.). The timing of the prepayment event (for all sources of prepayment) is identified by the fact that the loan_id disappears in months following the prepayment.

The below chart visualizes the above prepayment options:



TARGET DEFINITION 2/2

The specific prediction that the model should make (hence the target definition) is left as a modelling choice. You might consider:

- Binary classification – full prepayment vs. no full prepayment
- Multiclass classification – consider also partial prepayment and possibly the type of source for full prepayment
- other

Suggestions (you might of course find other approaches more plausible):

- Treat monthly observation of a loan as independent, i.e. as separate loans.
- Possible choice is to consider predicting the occurrence of an event over the next 1-year horizon (identify a loan as prepaid in the 1-year window starting from the current month).