

Likelihood of payment

At each point in time it is important for Alfa Co. to understand which debtors are likely to repay their debts within a near future, this enables the company to plan its collection efforts in a cost-effective way.

We would like for you to construct a model to predict the likelihood of a debtor making a payment during the coming months. You have a few potential target variables to pick from (all starting with IS_).

Here are the field descriptions of data set.

Column name	Description
DEBTOR_NO,	Id representing a unique debtor
RAND_NO,	random number
SNAPSHOT_DATE,	the time stamp of the data
IS_PAID_12M,	1 if a payment is made 12 months after the snapshot_date, 0 if no payment within the same period
IS_PAID_6M,	1 if a payment is made 6 months after the snapshot_date, 0 if no payment within the same period
IS_PAID_3M,	1 if a payment is made 3 months after the snapshot_date, 0 if no payment within the same period
IS_PAID_AMT_12M,	Amount paid within 12 months after the snapshot_date
IS_PAID_AMT_6M,	Amount paid within 6 months after the snapshot_date
IS_PAID_AMT_3M,	Amount paid within 3 months after the snapshot_date
CASE_AGE_MONTHS,	Age of debt
INTEREST_RATE,	Interest rate of the debt
REM_PRINCIPAL_AMT,	The size of the principal
REM_PRINCIPAL_INTEREST_AMT,	The size of the accrued interest + principal
REM_TOTAL_AMT,	Total debt size
DEBTOR_AGE,	Age of debtor
PAID_AMT_12M,	Historical payments during past 12 months
PAID_AMT_3M,	Historical payments during past 3 months
PAID_AMT_6M,	Historical payments during past 6 months
INCOME_TOTAL_LAST_YEAR,	Income earned last year
TOTAL_DEBT_AMT	Total debt at the bailiff. For a debt to be with the bailiff it must be sent there by a creditor, hence, a person can have debts with Alfa Co. or other creditors that isn't sent to the bailiff.

NULL values in any column means that the information isn't available.

In our discussion we will want you to reason around the technical aspects of your work, see it as an opportunity to highlight your knowledge in various ML techniques. We are also keen to know your thoughts about data preparation and preprocessing.

And please remember that being able to reason around the problem is as important as the technical aspects.