Completely Automated Credit Scoring
From Scorecard Development to its Implementation

D. Berestnev
O. Travkin
Scorecard development process

**Manual process**

- Data selection
- Partitioning
- Interactive grouping
  - Manual analysis of factors splitting received (stability, volume)
  - Business logics checks
  - Correlation analysis
- Multifactoral analysis
  - Manual selection of machine learning algorithms and their options

The most time is taken up by the process of forming and splitting factors.

**Automated process**

- Data selection
- Partitioning
- Automatic grouping
  - Increasing the number of splitting algorithms
  - Applying statistical bootstrap and testing for the estimation of stability and the quality of splittings
- Multifactoral analysis
  - Automation of the search process of optimal machine learning

Automation of the given part of development process enables us to cut time significantly.

The most time is taken up by the process of forming and splitting factors.
Automatic development process: Interactive grouping process

Number of iterations

= 

Number of algorithms

X

Number of leaves

X

Number of group size
Automatic development process: business logic checks

We have developed a special script language assigning the requirements for automatic business logics checks of each factor.

Value 50 is a global minimum.

After 50 the direction of dependency can be altered.

After 50 the direction of dependency can be altered, but 60 is a global extremum.
Automatic development process: Using cross variables

Using cross-variables instead of excluding in case of correlation pairs or low performance of factor enables us to detect hidden dependencies and increase performance of the future model. Automation also enables us to significantly increase the number of analyzing variables.

\[ R^2 = 0.8557 \]

Building decision tree as a method of making cross-variable
Automatic development process: Logistic Regression parameters

The criterion of the optimal number of model factors:
- Akaike information criterion
- Schwarz bayesian criterion

\[ y = \frac{1}{1 + \exp(-z)} + \epsilon \]

The algorithm of adding factors to the model:
- forward
- backward
- stepwise

Regularization type:
- Lasso
- Ridge
- Elastic-net

The value of regularization constant
The Bank’s advantages for the creation of such instrument:

- Significant number of credit applications per day (50-60 thousand)
- Good practical experience of data scientists (average experience more than 5 years)
- The same platform for model development and implementation into the online process

Our robot:

Internal hackathon:

Data scientist:

- t = 2 weeks
- Gini = 66%

Our robot:

- t = 4 hours
- Gini = 67%

The number of models used increased from 10 to 25
Automatic validation process

5 – DATA QUALITY TESTS
5 – MODEL PERFORMANCE TESTS
3 – MODEL SPECIFICATION TEST
2 – MODEL STABILITY TEST
Automatic monitoring approval rate process

CUSUM TEST

\[ S_{hi}(i) = \max(0, S_{hi}(i - 1) + x_i - \hat{\mu}_0 - k) \]
\[ S_{lo}(i) = \max(0, S_{lo}(i - 1) + \hat{\mu}_0 - k - x_i), \]

where \( S_{hi}(0) \) and \( S_{lo}(0) \) are 0. When either \( S_{hi}(i) \) and \( S_{lo}(i) \) exceeds \( h \), the process is out of control.
AUTOMATIC STRATEGY ADJUSTMENT SYSTEM (ASAS)

- Cutoff settings
- Monitoring of approval rate
- Model rebuilding
- Monitoring of approval rate
- Model validation
- Implementation into On-line credit process
- RTDM AUTOMATIC STRATEGY ADJUSTMENT SYSTEM (ASAS)
THANK YOU!