

Completely Automated Credit Scoring From Scorecard Development to its Implementation

D. Berestnev
O. Travkin

Scorecard development process

Manual process



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

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

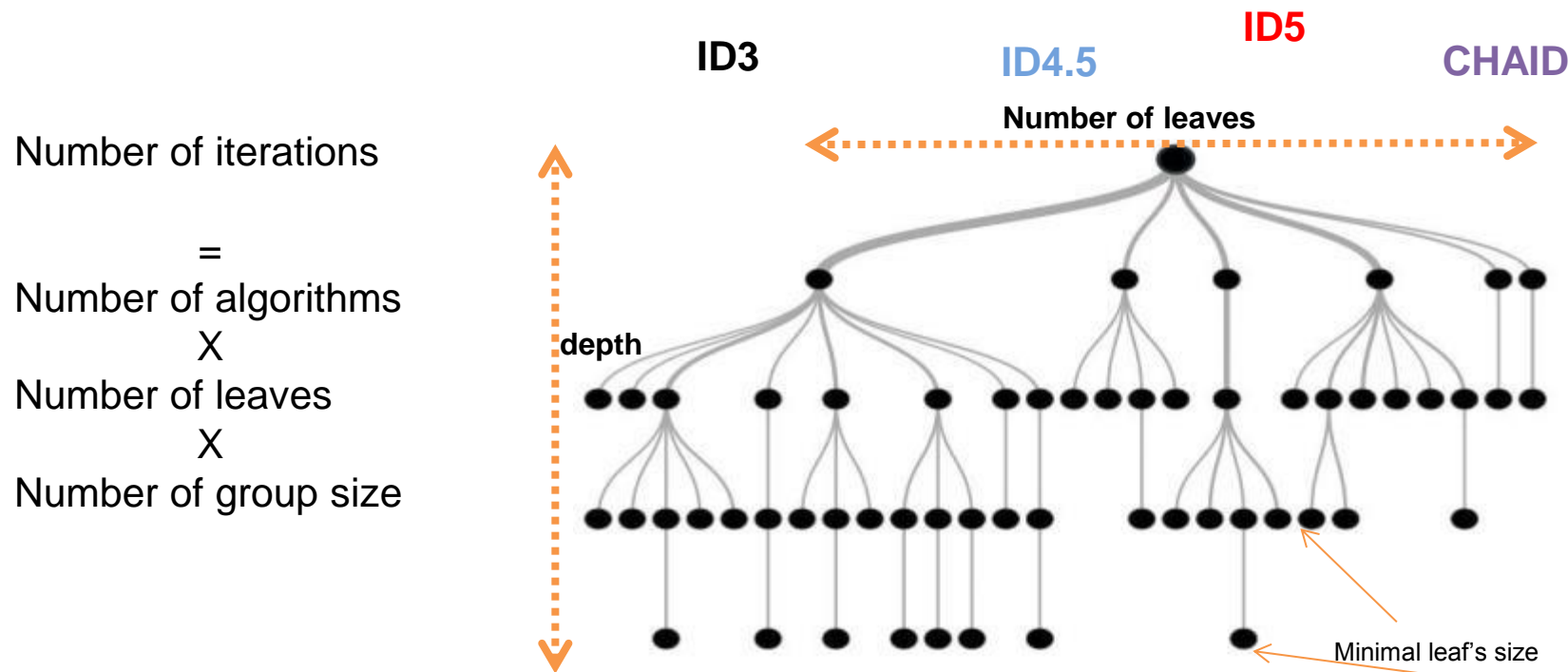
Automated process



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

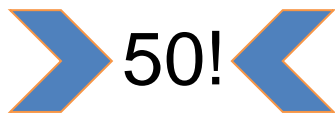
- Data selection
- Partitioning
- Automatic grouping
 - Increasing the number of splitting algorithms
 - Applying statistical bootstrap and testing for the estimation of stability and the quality of splittings
- Multifactorial analysis
 - Automation of the search process of optimal machine learning

Automatic development process: Interactive grouping process

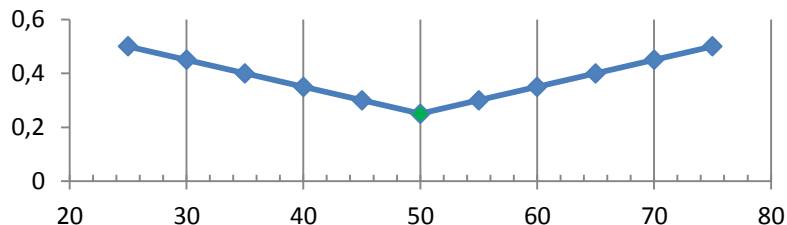


Automatic development process: business logic checks

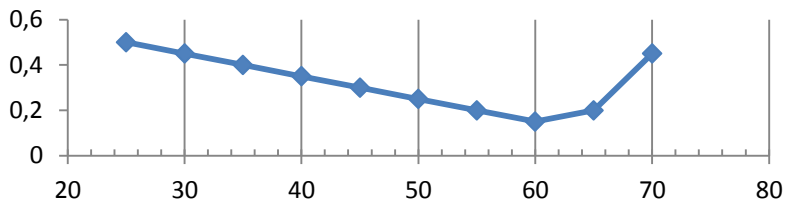
We have developed a special script language assigning the requirements for automatic business logics checks of each factor



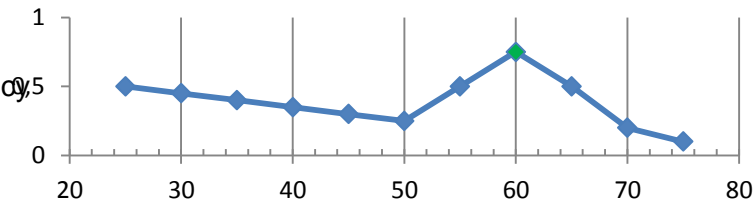
Value 50 is a global minimum



After 50 the direction of dependency can be altered

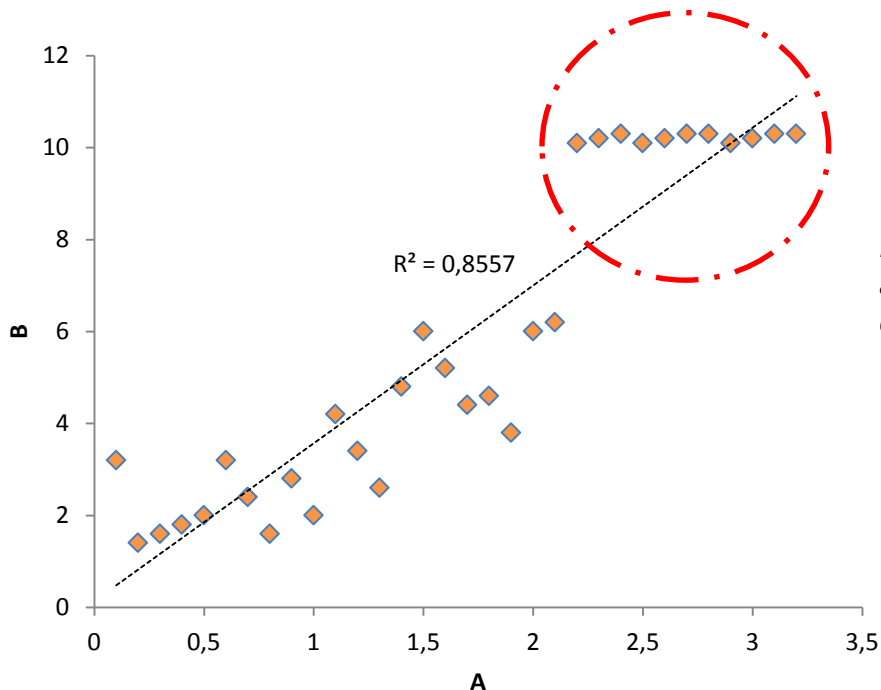


After 50 the direction of dependency can be altered, but 60 is a global extremum

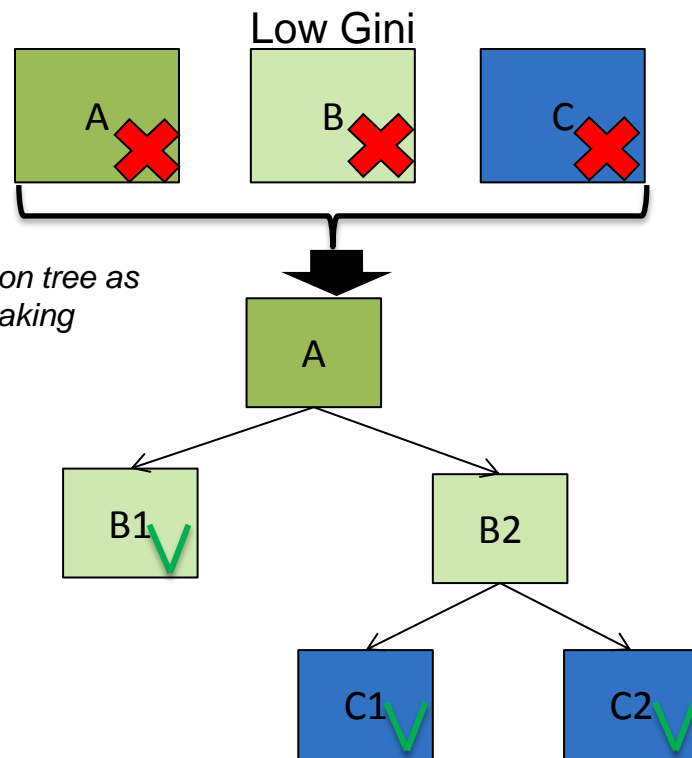
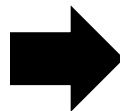


Automatic development process: Using cross variables

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.



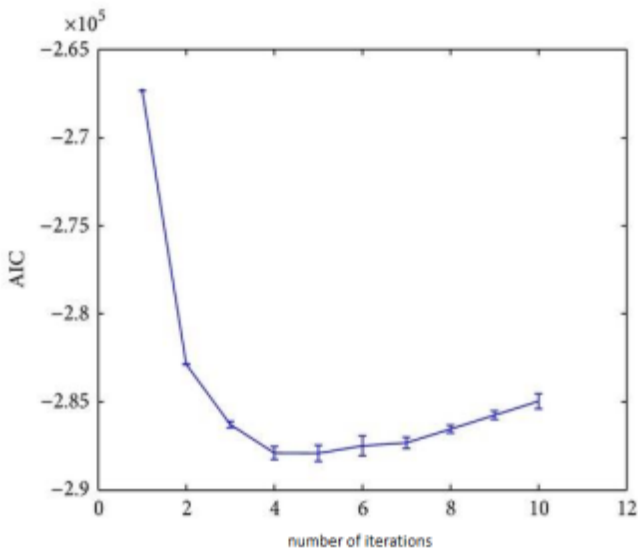
Building decision tree as a method of making cross-variable



Automatic development process: Logistic Regression parameters

The criterion of the optimal number of model factors:

- *Akaike information criterion*
- *Schwarz bayesian criterion*



$$y = \frac{1}{1 + \exp(-z)} + \varepsilon$$

The algorithm of adding factors to the model:

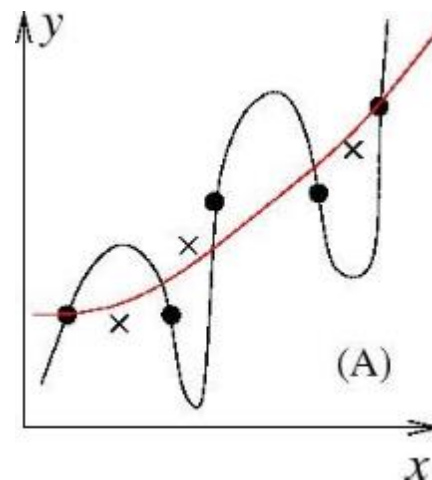
- *forward*
- *backward*
- *stepwise*

Regularization type:

- *Lasso*
- *Ridge*
- *Elastic-net*

+

The value of regularization constant



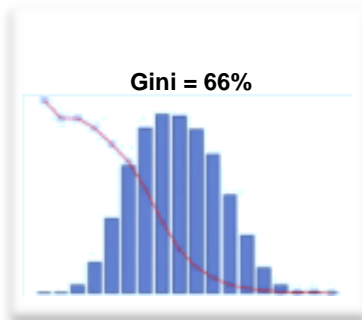
Automatic development process: results

Internal hackathon:

Data scientist



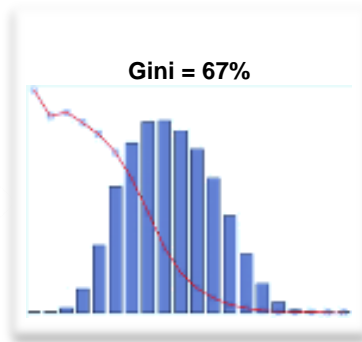
t = 2 weeks



OUR ROBOT



t = 4 hours



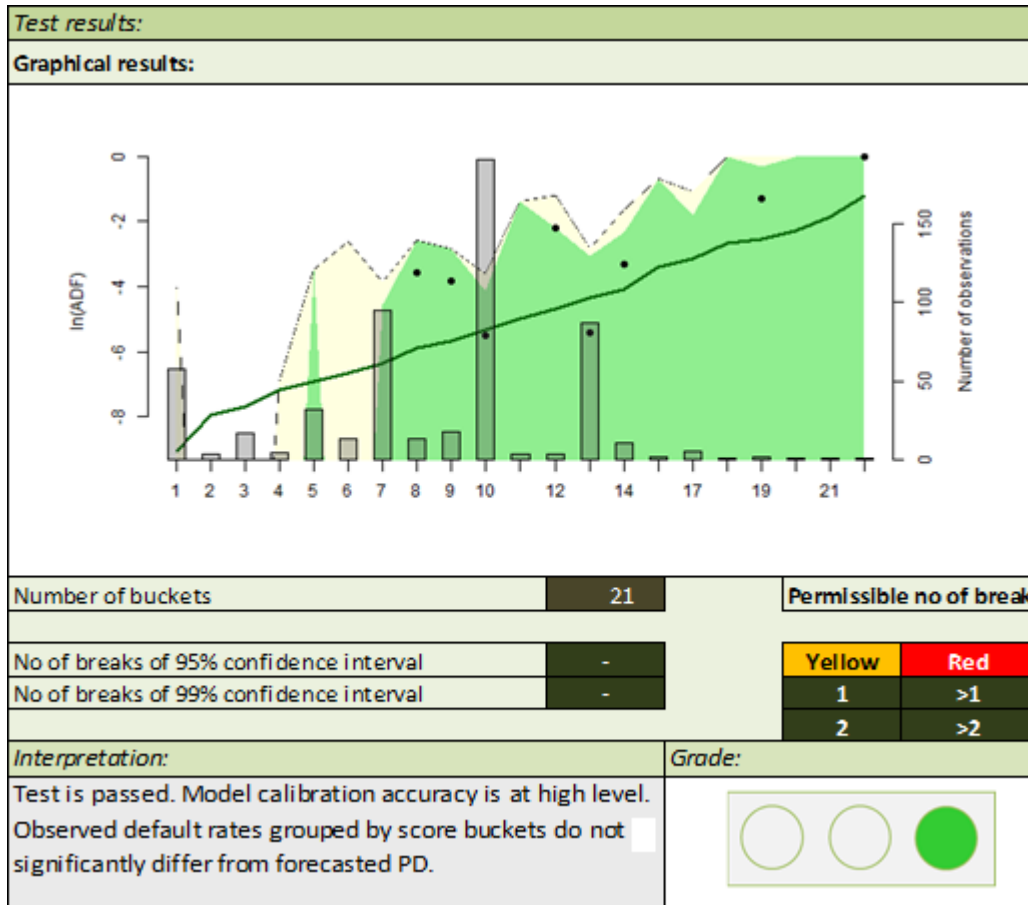
The Bank's advantages for the creation of such instrument:

- ❑ 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

The number of models used increased from 10 to 25



Automatic validation process



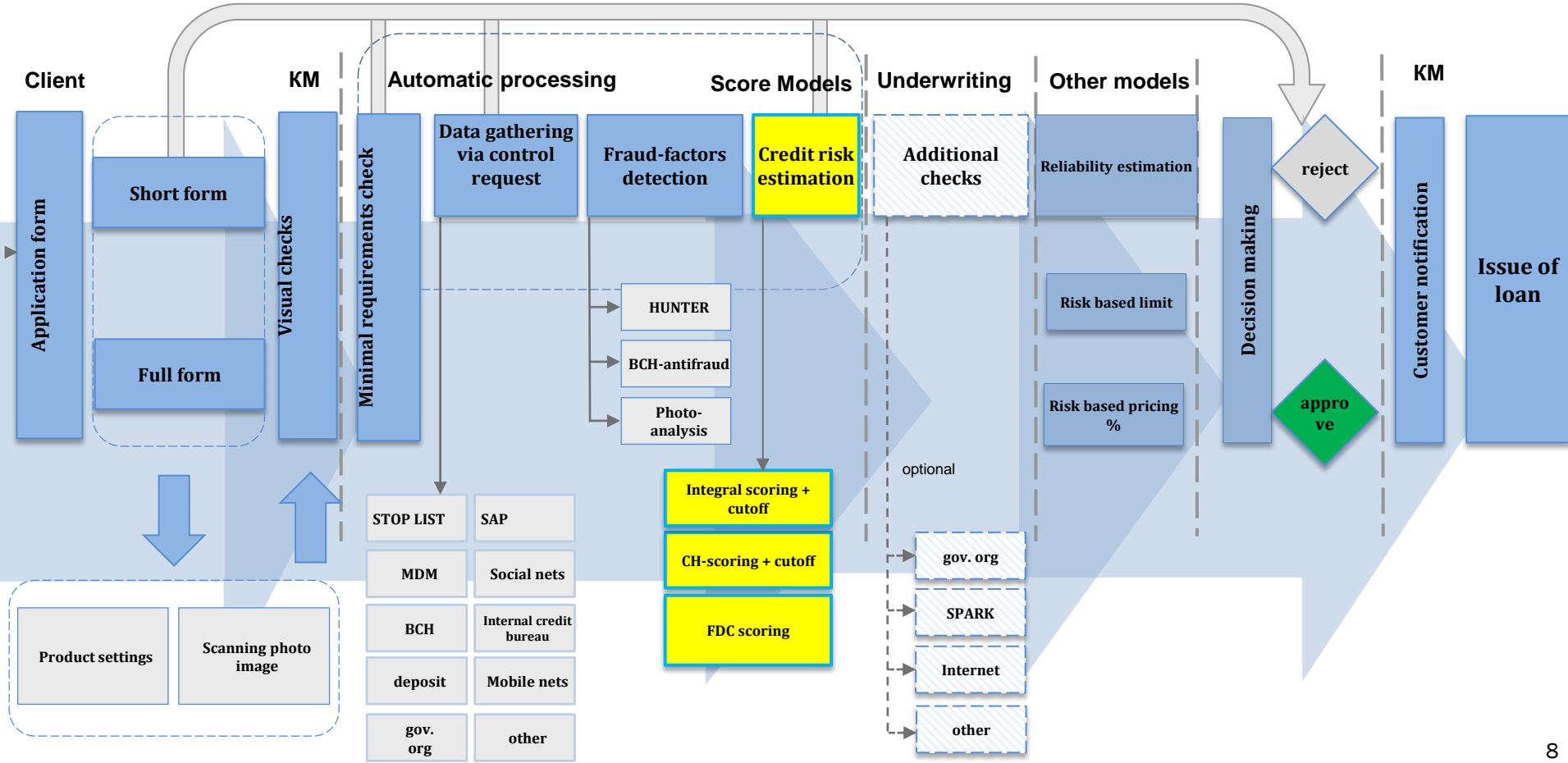
5 – DATA QUALITY TESTS

5 – MODEL PERFORMANCE TESTS

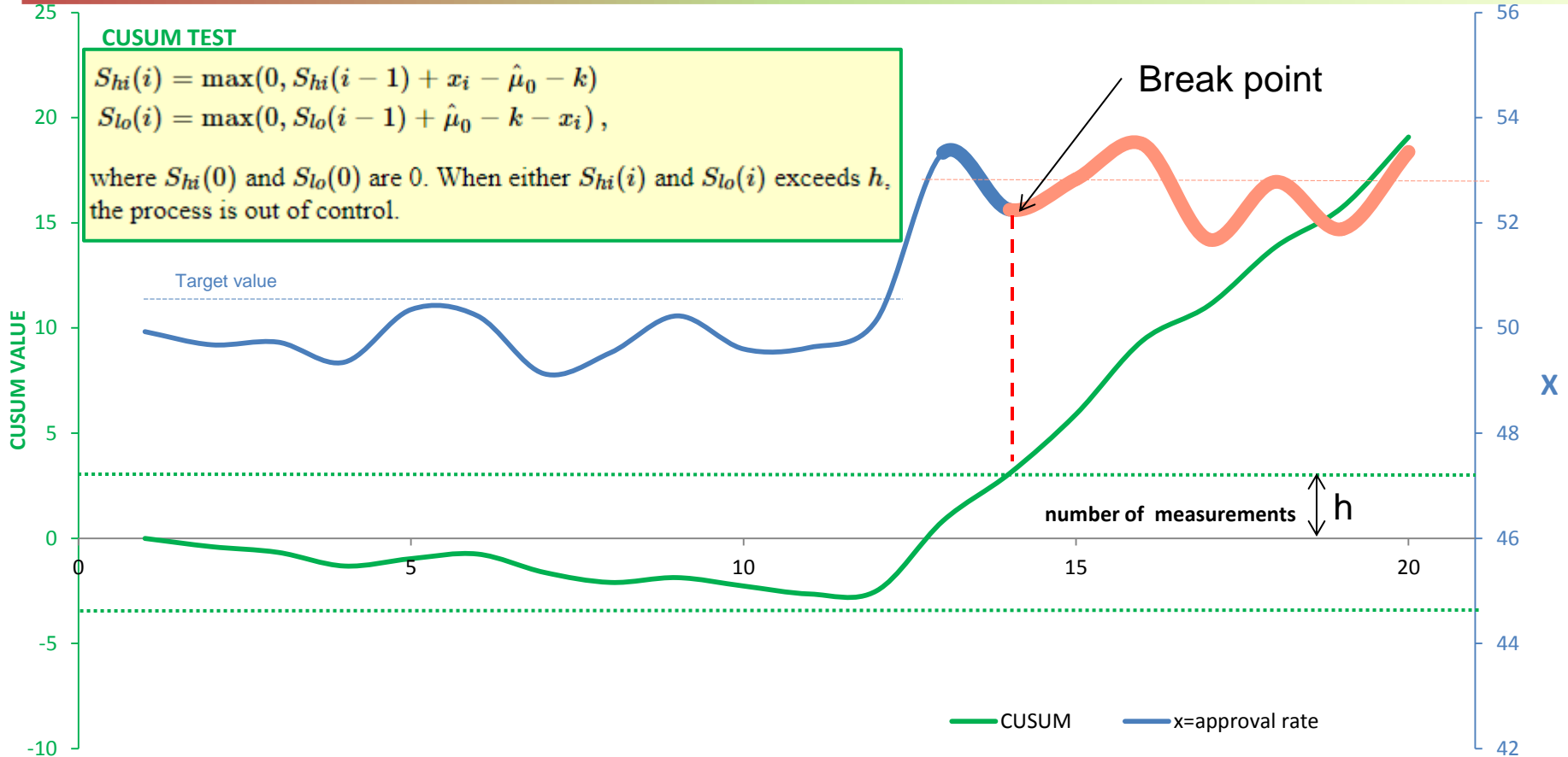
3 – MODEL SPECIFICATION TEST

2 – MODEL STABILITY TEST

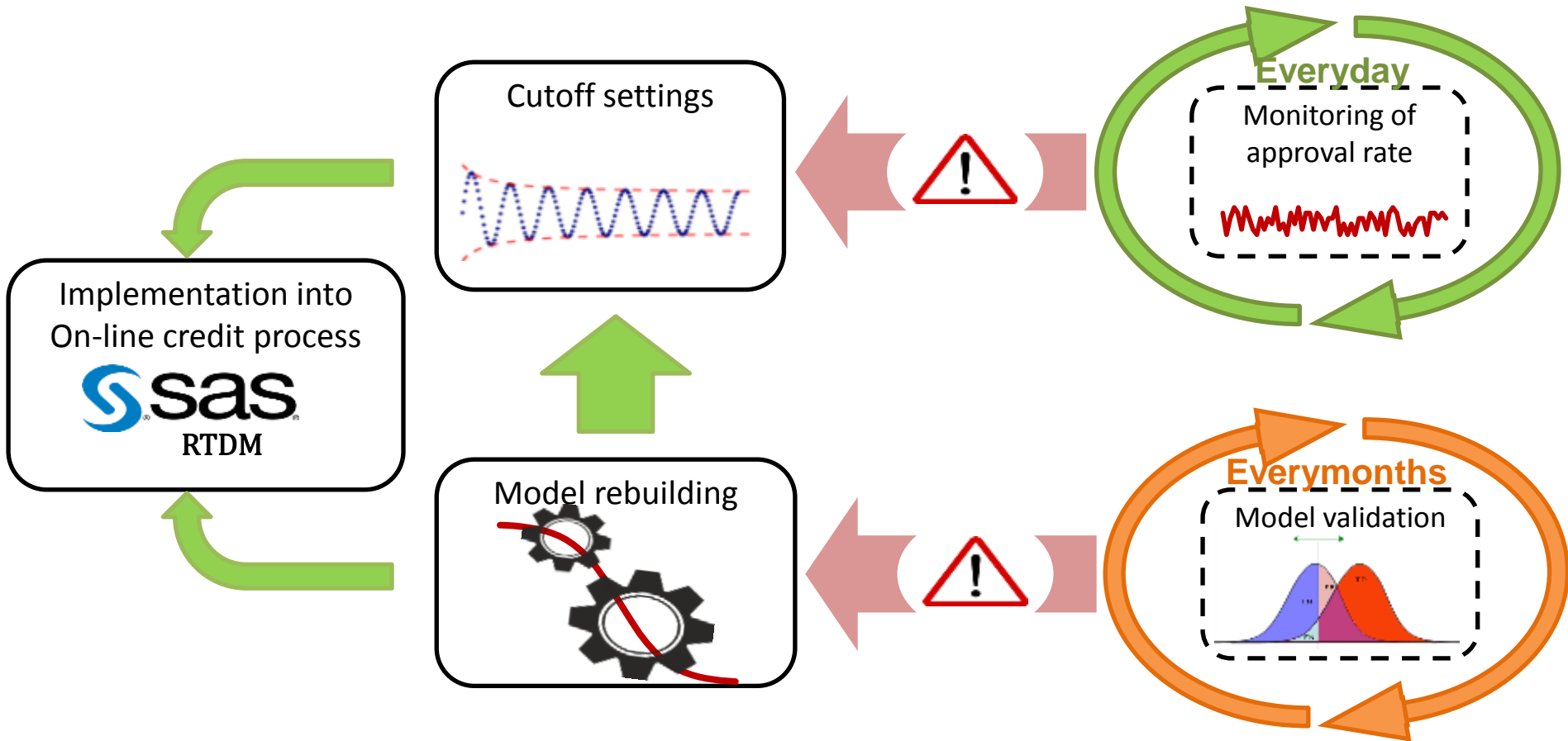
Automatic Implementation model process and setting cutoffs



Automatic monitoring approval rate process



AUTOMATIC STRATEGY ADJUSTMENT SYSTEM (ASAS)



THANK YOU!