Next steps (ideas)

Changes in the solution

Deal with outliers (there are clients that have unusual amount of applications)
- For each client take only last application (and previous if available), so the model is trained on distinct clients
- Try different scaling for numeric features
- Try more simple but more explainable models

Deal more carefully with date
- We will not be able to use a month of application for the future applications
- Add day of the week

What info do clients provide in application?

What can be added?

What can be changed or made more precise?

What kind of info we do not use but can add?

Geo data for defining neighborhood like in house/apartment prices problems (city, ip, gps, device)?

Target customers with an e-mail – clusterize / classify clients based on their historic profile
- All customers who are close to maturity date
- All that have good history
- All that accept offers with some rate
- That declined recent offer, try to close them one more time

What kind of info about similar task is available?

Competitions

Kaggle Home Credit

Rosbank

Raiff

Rosbank new year (geo)

Telecom

Validation

Kfold

Blending

EarlyStopping

Out of sample

Add predicts where the model works badly to the training set