

Karina Serikova

For several past weeks we've been trying to evaluate what the target is. It depends on which amount of risk (lose in money) bank is ready to take. They usually target on the number of days with default (просрочка). For some models it could be suitable to predict 30+ days default in, for example, 3 months (that more suits the fraud detection). But our case is more farseeing, so we want to predict the probability of client having default of 90+ days in 12 month. To do so we need all the historical information on clients who had been our clients for longer than 12 months, because only they have the flag of 90+ in 12 months. Taking this info is easy, the hardest part is to evaluate the features that would have any relation with target, by relation we mean any sort of dependency between these two. Also we need to make sure that any extra unnecessary features did not attended in model building. These features are those who take part in internal rules of bank, so they usually checked on the levels before scorecard, and this does not make sense to include them in scorecard building. Also we can calculate some features that could have good impact on scorecard quality. For example we could count the number of closed contracts of client, the open ones, divide one on another, play with features. Because we cannot rely only on what we have, we have to create something new. About the dataset, the bank has the privacy and information security, so we cannot fully show what we got, we will try to encrypt the info, so it would not be identified.

Syzdykbayev Makhmut

I studied machine learning algorithms for several weeks. After a few papers, researches, I started exploring decision trees. Decision trees has tendency to overfitting. We can solve this problems with reduce nodes, leafs but better solution is use ensembles

random forests or decision tree boosting

Both of algorithms good in prediction and has strengths and weaknesses

Forest uses randomize method for train data and take random data for trees, boosting take only certain data and learn from mistakes

I experiment by decreasing or increasing the number of trees, take the better parameters

And etc. This algorithms works poorly on high-dimensional sparse data, and maybe I will learn other algorithms like as SVM. So far I have not started writing code with bank data because I have not yet chosen which algorithm I will use. Using bank data is a matter of time

I also need the final dataset from karina in which we will discuss all the details

Rysbek Beibarys

In the period of some weeks, I was reading and exploring information about score cards in banking system to determine What it is, for what and what kind of dates are necessary. Main definition of score card is a measuring borrower's financial ability. Important variables to measure score card of a borrower are age, status of marriage, education degree, presence of workplace, salary and length of employment and etc. Each bank has own algorithm to examine score card of a client. And the result of the examine will be some percentage or quantity. If score card result of a borrower reaches this percentage or quantity, bank will approve loan. Resource taken from the website <https://www.banki.ru/wikibank/application-scoring>. Also, during this time I was considering how can We demonstrate final score. There were variants such to show Accepted/Denied status, to show in percentage the minimum threshold and the result of client's score card. And for demo version I choose to show the result in amcharts. For this reason, I was

looking for some amcharts for demonstration of result of score card on the resources such as getbootstrap.com and keenthemes.com websites. Also, I've construct demo version of the frontend design. However, it is not the final version. It could be changed by working process and after the discussion with teammates.

PROFILE PAGE

Personal Business About us +777 777 7777

PROFILE

Please fill out all required fields

First Name

Last Name

Unique identification number

Phone

Email

Address

Workplace

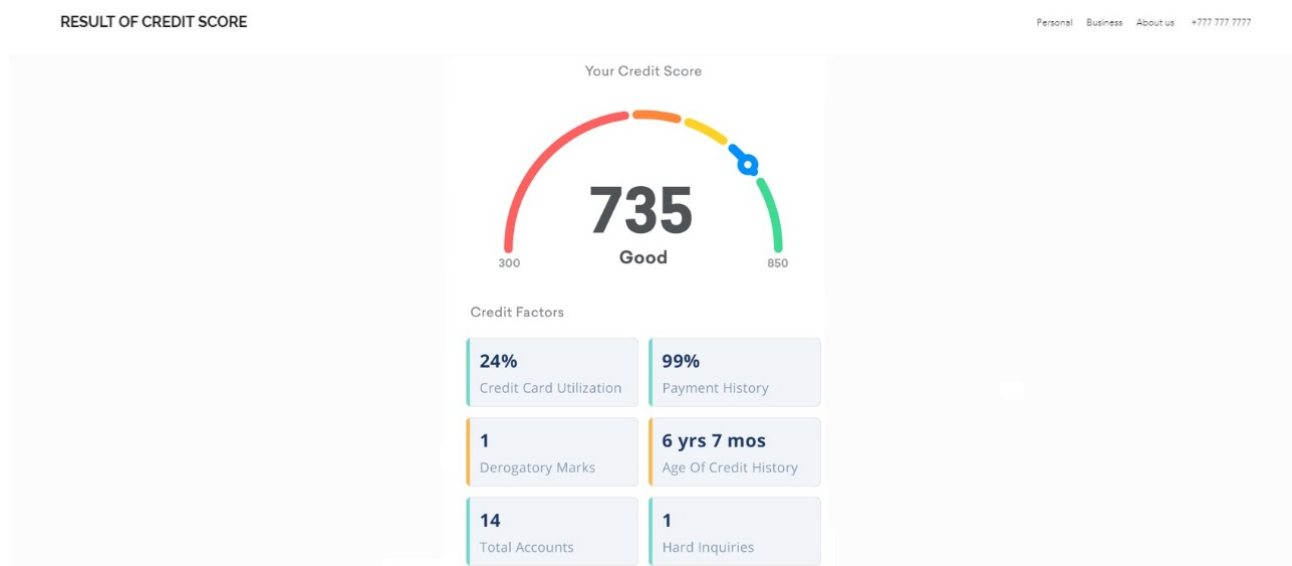
Trustworthy person

Phone number of trustworthy person

Degree of kinship

Submit

Thanks for submitting!



Zhansaya Alpyspay

The construction of scoring cards is based on statistical models. For their construction there must be sufficient and high-quality information about the borrowers of the bank. The quality of the initial statistical data for constructing a statistical model determines its accuracy of forecasting and

the success of developing a scoring system as a whole. These three weeks I tried to construct an efficient and useful database referring on our researches and interviews that we did before.

So, the first table should contain the personal data of a borrower that is taken from the filled application.

The development of a scoring model is based on an analysis of previous credit experience, so this must be the second table.

The next type of data is the borrower's internal credit history: current account balance, current debt, number of accounts, availability and volume of savings, number of previous loans at the bank, the largest value of the amount of debt on previous credit accounts, the presence of late payments, availability other banking products and services, regularity of payment of previous debts for all obligations, etc.

The information about the request results should also be saved in a separate table.

So I've outlined the tables and fields trying to break down the information into the smallest useful pieces, but it may change during the development stage. There are also some specific cases that should be taken into account before filling the database, such as, data on a specific type of customer must be excluded from the source database(for ex., atypical customers - scammers, bank employees, VIP customers, deceased customers, minors, double applications, loans with stolen cards, etc.)

The database management system that we will use is probably will be PostgreSQL.

I also started a research on what backend development framework is better to use, in order to make the data exchange process faster and reliable, so I study and compare the trending backend frameworks.