

Methods and techniques to assess creditworthiness of individuals

Tatiana Chernysheva
Tomsk Polytechnic University
Tomsk, Russia
tatch@list.ru

Maria Milovanova¹, Sergei Min'kov²
¹Tomsk Polytechnic University
²Tomsk State University
Tomsk, Russia
tatch@list.ru

Abstract—This article surveys methods and techniques used to assess creditworthiness of individuals. We pay particular attention to the scoring method. The classification of bank's customers is shown. The decision tree analysis is proposed for use, a decision tree example is shown using computer calculations.

Keywords—creditworthiness of individuals, assessment methods, decision tree method

I. INTRODUCTION

Creditworthiness is the assessed likelihood of a borrower to fully pay back a loan and interests thereon to a bank. In the Russian and international financial practice this category is an object for researches and evaluation.

When determining creditworthiness, two key factors are taken into account: the income of a borrower allowing him to pay back a loan within the confines of the loan agreement and holding of real property that can be pledged as collateral for the loan. Also, bank officials should allow for general market conditions, economic trends and various risks that the bank and its clients may face [1-3].

To analyze a borrowing capacity of individuals, financial institutions use information provided by a loan applicant on accommodation, employment and other details. They correlate an amount of the requested loan and the following factors:

- Income of the loan applicant;
- Household make-up;
- Real property ownership;
- Personal characteristics;
- Credit history.

Banks evaluate this information about individuals according to their policies and internal guidelines. On the market today, there is a variety of methods and techniques to determine the creditworthiness of borrowers. Below are the most popular ones.

II. SCORING METHOD OF ASSESSING CREDITWORTHINESS

This procedure of risk assessment with regard to a loan is based on a science-backed statistical model. The subject-matter of it is as follows: a client is assigned a credit score, derived on the basis of the analysis considering a number of characteristics, which allows the determination whether it is appropriate to grant a loan in a given case.

The scoring model allows us to bring a large amount of information on customers to "a common denominator". On frequent occasions, even an experienced loan officer may experience a difficulty in determining which of the loan applicants is more risky: a single man, running its own business, or a woman lawyer, divorced and having two children, and both of them have the same level of income. Scoring analysis enables the definition what sort of data are of great importance and what can be neglected.

To construct an analytical model, at first, European banks examine the selected customers who has already taken loans and expressed themselves in a certain way at the time of paying back the loans. This selection can include tens of thousands of people, with a set of distinct features to describe everyone in this selection. All clients are divided into two categories: "good" or "bad." The criteria used for attributing an individual to one of the two groups are rather vague and depend on the policy of a particular bank. Thus, the "bad" may refer to both an individual with a loan overdue for more than three months and an individual paying off the loan ahead of schedule (a credit institution could not earn anything from this loan).

Each characteristic of a client profile is converted into a variable; with its values corresponding to the ratio of a number of "bad" borrowers to the number of "good" borrowers with the same quality. In more complex cases, the logarithm of this ratio is taken as a base. Anyway, each feature receives a numeric value that reflects the level of risk [4-6].

There are a lot of techniques proposed to classify bank customers.

1. Regression analyses - building a regression for estimating a number of variables that are not correlated with each other;

2. Linear programming analyses – development of a linear scoring model;

3. Neural networks - the division of customers into groups, where the risk is about the same within the group and very different from the risk in other groups. This method is characterized by the ability to identify unusual situations, that is why it has been supposed as a particularly effective in protecting against credit card fraud.

4. Genetic algorithm – an analogue to the process of natural selection. Different models “mutate”, crossing with each other, as a result "the strongest" is chosen among of them.

5. The k-Nearest Neighbors algorithm - each client is assigned to a certain position in space. The classification is made by a majority of neighboring borrowers: "bad" or "good."



Fig. 1. Working layout of the Credit Scoring Method

The final evaluation of individual creditworthiness is determined by the formula:

$$Z = 0.15 \cdot X_1 + 0.3 \cdot X_2 + 0.25 \cdot X_3 + 0.3 \cdot X_4,$$

where Z is the creditworthiness;

X₁ is the social standing;

X₂ is the economic status;

X₃ is the accommodation details;

X₄ is the assessment of goodwill;

0.15, 0.3, 0.25, 0.3 are the weights of the relevant risk factors that determine the creditworthiness of the borrower.

Meanwhile, the scoring analysis has some serious limitations. Firstly, only those customers who have been given a loan are classified. Banking officials have no information to analyze the behavior of individuals to be denied a loan. It is possible that they would be good borrowers.

The second limitation to use the scoring evaluation is that the socio-economic situation is constantly changing, causing people to have new visions and interests, and the foregone model becomes less effective over time. The evaluation system must be constantly updated to reflect the objective economic realities. In Western Europe, the method is updated every 18 months on average; and in the case of Russia's institutions, a shorter interval is required for updating: at least once every six months.

III. THE DECISION TREE METHOD

A decision tree is a decision support tool applied in a scoring system context based on logical algorithms with a clear hierarchy. Due to the building of logical chains, the program is completely transparent, and decisions can be argued and tracked at any stage of calculation [7-8]. The aim of the system is to build a tree on the basis of individual input variable data with the purpose to classify loan applicants into two classes (to

grant a loan / to deny a loan). An example of the decision tree algorithm in a context of loan applicant classification is as follows (Figure 2).

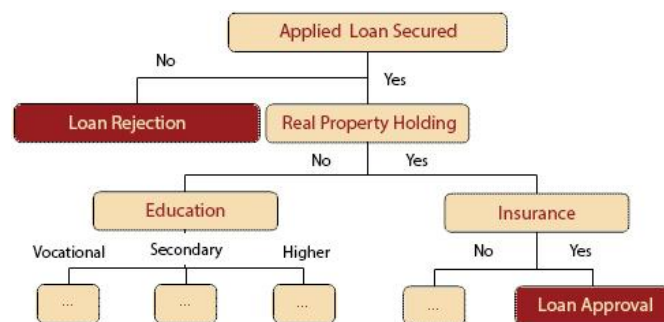


Fig. 2. A part of the decision tree example

In contrast to the previous model, in the process of classification, the decision tree includes only basic statistical data, rejecting intermediates, because at a certain moment of calculation, when it is required to input new information into the algorithm, some individual data become irrelevant and outdated. And only two resulting options remain for choosing: Loan Rejection or Loan Approval.

This method works according to the following pattern [9-12]:

1. The tree is based on the past periods, the classification for each case is known in advance (constant). When analyzing, the program chooses the root nodes, then splits data into subsets, which, in turn, hereinafter also down to leaf nodes, and so on until the resulting segment.

2. New nodes appear only during the introduction of a new key parameter. The decision tree is built using a criterion called entropy (uncertainty).

3. A node with a higher level of uncertainty is preferred. This algorithm is used in making the final decision whether to deny or approve a loan.

4. If an event occurs that affects constant values, the model rebuilds and adapts.

Advantages of the decision tree are as follows:

5. The building algorithm is simple to operate and easy to understand. A bank employee is able to explain the work and answer why this decision is made.

6. It can work perfectly with any number of inputs. They do not need to be structured or normalized for the program

7. At the same time, a decision tree can be viewed as both the data by category and intermediate factors.

8. Based on the statistical data, it is possible to check the accuracy of the decision tree.

9. If a hierarchy has been broken while chain building, the decision tree is able to give accurate results, because of more important factors considered.

10. The main thing is that almost every bank employee can run this program. Without special training, since the tree

uses the concept of "white box", all decisions can be explained by the logic, moving from one block to another.

Notwithstanding, the decision tree has its disadvantages:

1. It is difficult to derive an optimum chain, reflecting the optimal level of competence but without overloading with variable factors at the same time.

2. An exact algorithm of choice is required; otherwise the system will be confusing and the selection in the system will be difficult to understand.

3. In some cases, the system can decide that the choice is appropriate where a larger number of nodes are used. This can be connected with the wrong building algorithm as well.

To solve all the above problems, it is enough to optimally adjust the tree depth setting by removing nodes that are insignificant or in conflict with more important blocks.

IV. AN EXAMPLE OF USING A DECISION TREE

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

For making an example, we apply the Tree Analyzer software tool. The initial information represents a 1000 - number selection set, which is used to numerically express physical, financial and personal details (behavior at a time of repaying the loan in full) of a person. Firstly, for the tree building, such personalized input data relative to a bank customer are used as Passport number, Full name, Residence permit, Amount of loan requested, Loan terms, Lending purpose, Average income and expenditure in a month, Main expenditure, Real property and motor vehicle ownership, Bank accounts, Insurance, Place of employment, Employer's industry, Work record, Borrower's job position, Time with current employer, Sex, Marital status, Age, Number of Dependents, Time at current address, Security for loan.

It is worth noting that these personal data inputs as Passport number, Full name, Residence permit, Place of employment and Employer's industry are immediately rejected by the program, because such data has no practical value in view of its unique character and is not used for constructing algorithms.

The target node is always "Loan Approval" with two possible optional decisions "Yes" (true) and "No" (false). "No" is in the case provided a customer is not considered as a fit to comply with the specified parameters; there is a wide range of reasons to reject a loan application: overdue credit, bad credit history and so on. The "Yes" decision is its opposite.

After setting up the input parameters and breaking them down into dominant and destination blocks, a hierarchical model is built for assessing the individual's creditworthiness. This structure is called as a decision tree (Figure 3).

After analyzing the decision tree the following conclusions can be drawn.

Since the program is based on the criteria intended to eliminate most errors, this allows the program to define that the node "Security for loan" is more important than "Time with current employer" or "Time at current address" in the hierarchy. It is also important that some input parameters are not used in the process of building a decision tree; this indicates that these parameters are not relevant or dominant in responding with regard to the destination block "Loan Approval".

Such fields as 'Amount of loan requested', "Loan terms", "Lending purpose" are out of picture in the decision tree. All these parameters are replaced by "Security for loan", as it is regarded as a key block by the system. This can be explained by the fact that this block is able to generalize all of the above fields and, at the same time, simplify the system to avoid its overloading with duplicate blocks.

The decision tree system is capable of memorizing and summarizing input parameters, and this enables us to determine the most likely model of a borrower's behavior. That is, if a potential borrower has input parameters similar to those already been processed by the program, the probability that the borrower's model of behavior will be the same as his foregoer is quite high.

Based on previously built hierarchical models, the program is able to classify a borrower; it is enough to input the responds to the key building blocks of a decision tree, without constructing a complete logical chain and respond to the destination block "Loan Approval".

V. ASSESSMENT OF CREDITWORTHINESS IN A CONTEXT OF PAYMENT CAPACITY

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

Banks use the data on person's income to analyze a degree of risk for a borrower to lose income. For this purpose, such documents are required by banks as: a proof of income indicating a wage/salary over the last six months, a loan application form issued by the bank and authorized by employer's signature and seal or income tax return [13]. An amount of average monthly income is reduced by the amount of taxes and mandatory payments and is then multiplied by a correction factor, which is determined by banks according to their internal regulations and policies (in different credit institutions the correction factor can vary within 0.3-0.6).

The technique of assessing creditworthiness according to payment capacity is used by the Russia's biggest credit organization - Sberbank. When applying for a loan, borrowers are required to furnish the following information:

1. Passport
2. Duly completed loan application form

3. Income proof

4. Certificates confirming the security for a loan – property documents.

With regard to the submitted documents and information, bank's officials shall make a decision on the advisability of granting a loan. For this purpose, the amount of income tax paid is subtracted from the person's average monthly income. The resulting figure in the application form is reduced by the size of payment required to be paid monthly (including any other loan repayments). This is how to calculate a net income.

The borrower's payment capacity is determined according to the formula:

$$P = NI * K * t,$$

where NI is the average monthly income (net) over the last six months minus all mandatory payments;

K is the factor depending on a value NI: K = 0,3 when NI is equivalent to 500 USD,

K = 0,4 when NI is equivalent to 501 to 1000 USD,

K = 0,5 when NI is equivalent to 1001 to 2000 USD,

K = 0,6 when NI is equivalent to 2000 USD,

t is the period of loan (in months).

When a loan officer has some doubts about maintaining the borrower's income over the estimated period of loan, the value NI can be adjusted downwards with appropriate explanations in the conclusion made by the loan officer.

When a loan is in rubles, the payment capacity is calculated in rubles. When a loan in foreign currency is applied, the payment capacity is calculated in US dollars. The payment capacity of guarantors is defined similarly to the payment capacity of the borrower with the difference that K = 0.3, is regardless of the value of NI.

Nemesysco, an Israeli company, developed HR1-Credit software for banks and the system is used for credit risk assessment.

All banks want to determine as quickly as possible the verity of information provided by a potential bank's customer, to assess a client's financial situation in order to protect themselves against the risk of loss in assets. The assessment of customers' creditworthiness is of great importance today. A large number of loan defaults in Russia leads to an increase in interest rates for loans. Debts from defaulters are reimbursed by bona fide borrowers. Some credit organizations reject to grant loans without collateral in an effort to reduce the risk of loss. Currently, the scoring analysis becomes one of the main tools available to lenders to assess the creditworthiness of borrowers. But the scoring system still skips a number of unscrupulous borrowers. Also scoring may not admit a bona fide customer. The HR1-Credit technology serves to the purpose of identifying customers with high risks [14].

The clients are classified to the high-risk group, who:

- have no intention of paying a loan before applying for it;

- intentionally submit incorrect data on their credit history.

The HR1-Credit program is an automatic test system, taking 10 minutes to check the potential client and assess his creditworthiness and measure the risk of default.

Voice analysis technology can conduct a deep analysis at different layers of voice and a range of emotion. SENSE can detect various emotional states such as excitement, embarrassment, stress, desire to share information, undistracted attention.

For a more accurate analysis of creditworthiness, there is a special algorithm used for calculating creditworthiness. This algorithm builds an assessment on the basis of changes in the parameters during the test, including a psychological background. The test is designed to check individuals who apply for a loan to a bank. In this connection, it is important to be certain that it is a person who really needs a loan, but not that acting by proxy, is under checking.

The borrower is assessed and evaluated using the testing system against the following:

- Personal responsibility and intention to pay a loan;
- Financial capacity to pay a loan.
- Borrower's credit history for the last two years;
- Tension state during the test.

Analysis of results to assess a borrower (options available):

"Overall tension" is the assessment to use the results of the test. The assessment should be in a certain range.

If a credit history and payment capacity indicates a low risk, it means there are reasons for the denial of a loan after analyzing the voice.

When both tests show a high risk, a manager decides to refuse to grant a loan.

This procedure takes 5-7 minutes. To pass this test, the potential client communicates only with a computer in a separate quiet room without fear that someone could hear him. Then the manager makes a decision relating to test results.

CONCLUSION

There are different techniques, tools and models to determine the client's creditworthiness. Banks are constantly improving their methodologies in an effort to take a more impersonal and balanced decisions as to granting a loan. The decision tree is much more advantageous than the scoring system because the decision tree modeling is the self-adaptive software, capable to change its behavior by means of new inputs to their algorithms, with requiring minimum interference. In building hierarchical chains, the decision tree enables the rejection of those values that do not affect the final result, thus, the program is more accurate in assessing the borrower's creditworthiness. Different approaches to creditworthiness always have one goal - to ensure stable operation of banks and minimize their risks.

REFERENCES

- [1] L. Quirini, L. Vannucci, A new index of creditworthiness for retail credit products, *Journal of the Operational Research Society*, 2010, 61(3) pp.455-461.
- [2] H. Rottmann, T. Wollmershäuser, A micro data approach to the identification of credit crunches, *Applied Economics*, 2013, 45(17), pp.2423-2441.
- [3] A.A. Zakharova, V.V. Ostanin, Formalization model of expert knowledge about a technical index level of engineering products, *IOP Conference Series: Materials Science and Engineering*, 2015, Vol. 91, Article number 012070. - p. 1-8.
- [4] A.A. Zakharova, Decision making models on the basis of expert knowledge for an engineering enterprise strategic management // *Applied Mechanics and Materials*, 2015, Vol. 770, pp. 645-650.
- [5] T.Yu. Chernysheva, E.V. Gnedash, T.Y. Zorina, N.V. Lenskaya, Information systems project risk assessment: expert approach, *Applied Mechanics and Materials*, 2014, Vol. 682, p. 539-543.
- [6] M. Chrzanowska, E. Alfaro, D. Witkowska, The individual borrowers recognition: Single and ensemble trees, *Expert Systems with Applications* 36 (3 part 2), pp.6409-6414.
- [7] L. Rokach, Decision forest: Twenty years of research, 2016 *Information Fusion* 26, pp.11-125.
- [8] B. Srdjevic, M. Pipan, Z. Srdjevic, B. Blagojevic, T. Zoranovic, Virtually combining the analytical hierarchy process and voting methods in order to make group decisions, *Universal Access in the Information Society*, 2015 V. 14, Issue 2, pp. 231-245.
- [9] Sok, H.K., Ooi, M.P.-L., Kuang, Y.C., Demidenko, S. Multivariate alternating decision trees, *Pattern Recognition*, 2016, 50, pp.195-209.
- [10] I. Ben-Gal, A. Dana, N. Shkolnik, G. Singer, Efficient construction of decision trees by the dual information distance method, *Quality Technology and Quantitative Management*, 2014, 11(1), pp.133-147.
- [11] V. Purdilă, Ș.-G. Pentiuc, Fast decision tree algorithm, *Advances in Electrical and Computer Engineering*, 2014, 14(1), pp.65-68.
- [12] V. Silich, M. Silich, Hierarchical Object Model of Variant Choice, *Proceedings The 7th Korea-Russia International Symposium on Science and Technology KORUS 2003*, June 28-July 6, 2003, pp. 378-381.
- [13] P.-F. Pai, Y.-S. Tan, M.-F. Hsu, Credit Rating Analysis by the Decision-Tree Support Vector Machine with Ensemble Strategies, *International Journal of Fuzzy Systems*, 2015, 17(4), pp.521-530.
- [14] T. Yu. Chernysheva, N.V. Lenskaia, A Study of methods for assessing the creditworthiness of potential borrowers - individual person, *Economics and entrepreneurship*, 2014, no. 1-1., pp. 209-213.

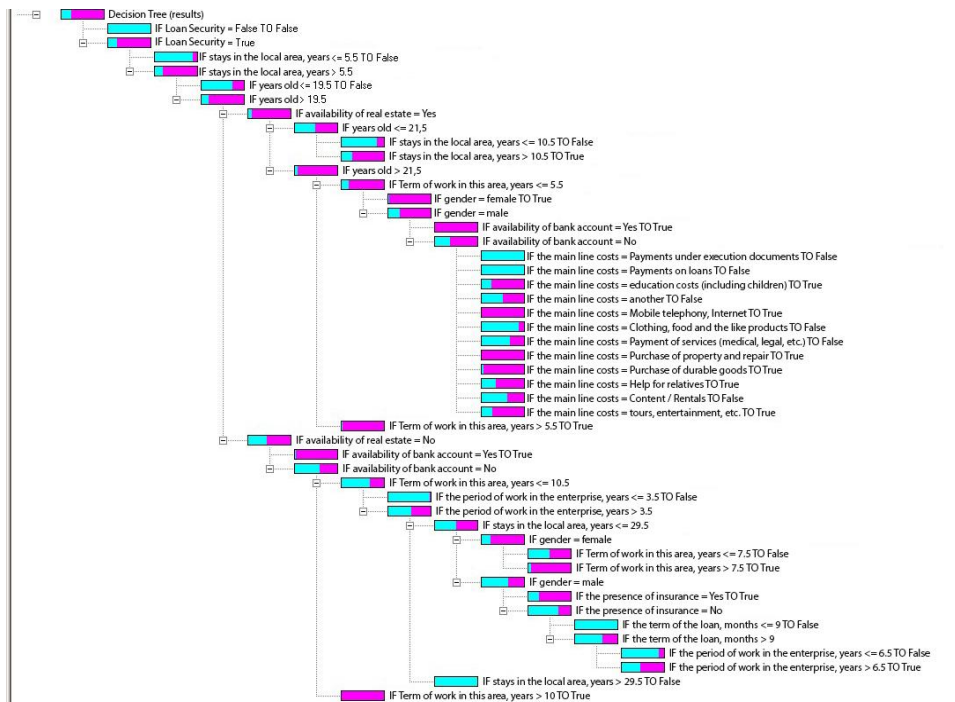


Fig. 3. A decision tree