UDC 004.9

IMPROVING THE QUALITY OF CREDIT ACTIVITY BY USING SCORING MODEL

Melnyk K. V. – PhD, Associate Professor of the Department of Software Engineering and Management Information Technologies, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine.

Borysova N. V. – PhD, Associate professor of the Department of Intelligent Computer Systems, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine.

ABSTRACT

Context. The problem of credit assessment of a client is considered. It is a simultaneous processing of lender's data of different nature with further definition of the credit rating. The object of this study was the process of lending to individuals by credit institutions.

Objective. The purpose of the work is to study the process of improving the quality of lending through the development and use of a scorecard model.

Method. An analytical review of the domain area was conducted. A business process model for assessing clients' creditworthiness in the form of an IDEF0 diagram is developed. Dedicated groups of indicators characterizing a potential lender from different directions. Selected sets of values for each indicator of credit separately. The methods of solving the problem of clients' creditworthiness are analyzed. Selected Bayesian naive classifier as a method for solving the problem of classification of potential lenders. The existing information systems for assessing the creditworthiness of clients are analyzed. A scoring model for assessing credit ratings by the client in the form of an algorithm is developed. The list of functional requirements of the information system, which is presented in the form of a use case diagram is determined. Three-level architecture for the information system is proposed. A database model has been developed to preserve customer information. An information system was developed for determining the credit rating of a client based on the developed scoring model. Numerous studies have been conducted to determine the class of a potential creditor. The process of determining the quality of credit activity is analyzed. Quality indicators for assessing the creditworthiness of clients are selected. The method of calculating the quality of credit activity is offered.

Results. The scoring model was developed, which was used in solving the credit assessment of clients through the help of the proposed information system. The process of improving the quality of credit rating is investigated.

Conclusions. The conducted experiments have confirmed the proposed scoring model and allow recommending it for use in practice for assessment process of client creditworthiness. Scientific novelty is to improve the process of credit activity by automating the use of naïve Bayes classifier, which reduces the human factor in decision-making.

KEYWORDS: scoring model, classification task, naive Bayesian classifier, credit score assessment, lending, lender, borrower, and creditworthiness.

ABBREVIATIONS

NBC – Naive Bayes classifier;
IS – information system;
IT – information technologies.

NOMENCLATURE

K – set of clients;
Q – set of output classes;
q_l – l-th output class, element of the set Q;
q_{lk} – l-th output class of the k-th client, $k \in K$;
K_l – set of clients from l-th output class;
I – set of client's characteristics;
$J_i, i \in I$ – set of i-th characteristic meaning;
f_{ij} – j-th value of i-th characteristics, $j \in J_i, i \in I$;
f_{ij}^k – j-th value of i-th characteristics from k-th client, $j \in J_i, i \in I, k \in K$;
x_{ij}^l – number of j-th value of i-th characteristics l-th output class;
y_l – number of l-th output class from set of clients K_l;
$P(f_{ij}/q_l)$ – conditional probability of occurrence j-th value of i-th characteristic f_{ij} from output class q_l;
$P(q_l)$ – probability of assignment to a client output class q_l;
$P(q_l/\{f_{ij}^{k+1}\})$ – conditional probability of assignment to a client l-th output class based on conditionals $\{f_{ij}^{k+1}\}$;
$R(q_l)$ – probability of assignment to a $k+1$-th client output class q_l;
P – precision of NBC;
R – recall of NBC;
a – the number of clients of a particular class that has been allocated IS and which are really related to this class;
b – the number of clients who were mistakenly attributed to a particular group;
c – the number of customers who mistakenly was not in a certain class.

INTRODUCTION

Lending is the main activity of credit institutions, banks and credit unions. It provides almost 50% of the

DOI 10.15588/1607-3274-2019-2-7

profitability of these institutions [1]. The basic concepts in the process of lending are a loan, a creditor and a borrower. The credit is the relationship between two subjects of credit relations the creditor and the borrower, in which the creditor transfers the borrower's money, and the borrower undertakes to return the same amount of money within a certain period [2, 3].

In the process of conducting credit operations, the institution works with a credit risk. It is the risk of non-payment by the borrower of the amount of principal and interests. Non-repayment of loans, especially large ones, can lead to a bankruptcy of a credit institution, and this may be the cause of bankruptcy of other enterprises, banks and individuals associated with this bank or credit union.

Therefore, credit risk management is an indispensable part of the strategy and tactics of survival and development of any bank or credit institution. Risk reduction in the implementation of loan operations can be achieved on the basis of a comprehensive examination of the creditworthiness of the bank's clients. Credibility of a client is the ability of a person to complete and to pay off in time for his debt obligations [3, 4].

This topic is of great interest for the research of not only Ukrainian but also foreign economists [2–8]. The scientific literature focuses on lending to enterprises and legal entities. In this case, there are many financial indicators of the company's work, which allow to assess the degree of credit risk.

However, there is no effective method for determining the individual's creditworthiness, therefore, a great interest is the increase in the quality of the credit assessment process when working with the people. When it comes to lending to the population, an important role in determining the creditworthiness is not so much the ability to repay the borrower's debt, but the willingness to pay the loan and pay interest on time. Readiness for this is different one; it depends on the personal characteristics of each person. These characteristics may include education, age, social class, gender, family status, etc. [7–9]. To analyze personal and financial indicators of an individual is a complex process for a financial consultant, which will result in the adoption of sound financial decisions. In order to minimize the subjectivity of a financial consultant and reducing the time to take decisions on lending to individuals use scoring [10, 11].

Scoring is a way to quickly evaluate a potential client. The result of it a client score. The received score can be used to solve various financial issues.

Scoring model of credit assessment is a mathematical model of the borrower's behavior based on the accumulated statistics. The result of the using is integral assessment of the risk in the view of the probability of repayment of the loan.

Therefore, the actual task is to develop a scoring model for assessing the creditworthiness of individuals, which will increase the speed of analysis of information about the client, reduce the time for approval of the decision on the loan and reduce the subjectivity of the assessment. And all of this will increase the quality of credit activity.

The object of study is the process of lending to individuals by credit institutions.

The subject of the study is the theoretical and methodological tool for assessing the client's creditworthiness by the use of a scorecard model.

The purpose of the work is to improve the quality of lending through the development and use of a scoring model that will satisfy the interests of the lending institution associated with the risk of non-repayment of credit.

1 PROBLEM STATEMENT

The problem of the assessment of creditworthiness is the task of classification, where the classes are customers from the database, which need to be classified according to their indicators into two groups: reliable clients, that is, the borrowers are creditworthy and unreliable when borrowers are uncreditworthy.

The formal formulation of the task of classifying clients can be presented in the following way. Let X is a set of clients of the bank, at the same time $X = \{x_1,...x_m\}$. Let Y is a set of classes that need to be broken up by clients, that is $Y = \{y_1,...y_k\}$. The task of classification is a reproduction of one set in another, in which each element of the first set becomes unambiguous with the particular element of the second set $a : X \to Y$.

To solve the problem of assessing the creditworthiness of a client it is necessary:

– to allocate a set of characteristics of the client – information about the borrower, which will be taken from the application form and documents provided by the borrower;

– to choose a method that will allow to determine the dependence between the characteristics of the client and the level of his creditworthiness;

– define the client class.

2 REVIEW OF THE LITERATURE

At the present time almost all credit institutions use IT in solving many financial problems, including lending. Using IT can improve the quality of the results. An overview of IT used to assess the creditworthiness of individuals allows to emphases following IS:

1. The Ukrainian Bureau of Credit Histories (UBCI) [12] is an IS that provides such services: collection, processing, analysis, keeping of information, which is a credit history; providing legal and individual person with advisory services; providing credit reports

2. Mobile app "Credit History" is app from the Google Play service [13]. Functional abilities of IS: calculate a credit rating; see your credit history online and find out how to improve it; receive reminders for next payments; analyze personal finances and make informed financial decisions.

3. Online-banking Privat24 [14] is the IS for assessment creditworthiness of clients. To calculate the credit rating you need to go to the menu item "My accounts" → "Credit rating" and fill in the necessary information.

DOI 10.15588/1607-3274-2019-2-7

An analysis of existing IS for assessing creditworthiness in Ukraine shows that the main disadvantage is the secrecy of methods used by credit institutions to calculate credit score. Also, systems do not allow to change the indicators for the calculation of the rating credit assessment. Therefore, the results indicate the relevance of this study.

There are many methods to determine the credit rating, which have their advantages and disadvantages. All methods can be divided into two main groups: expert methods and scorecard models that use mathematical methods for processing information, such as Data Mining [15–19].

Expert methods are use only the experience and knowledge of a financial advisor, so this approach is characterized by a high degree of subjectivity and a high probability of error in deciding about issue a loan. The second group of methods is scoring. The scoring procedure using a scorecard model includes: information gathering, the construction of a mathematical model (the choice of the classification method and the definition of criteria for risk categories) and the distribution of creditors by risk category.

Among the main advantages of scoring systems, one can distinguish:

– lowering the loan non-repayment level;

– increasing speed and impartiality in decision making;

– possibility of effective management of a loan portfolio;

– absence of long-term training of the employee.

The main disadvantages of a scoring system for assessment of client's creditworthiness:

– high cost of adaptation of the used model under the current situation;

– the probability of a model error in determining the creditworthiness of a potential borrower is due to the subjective opinion of a specialist.

The methods and approaches in the scoring systems are quite diverse, they can be use both individually and in different combinations. An analytical review of the main known and currently used methods for assessing creditworthiness is presented in Table 1.

Comparing the methods of assessing the creditworthiness of the clients presented in the table above, one can conclude that the "Naïve" Bayesian classifier is a good classifier in the scoring model for processing information about a potential lender.

Table 1 – Review of methods for assessing creditworthiness

Method	Advantages	Disadvantages
Expert method	1. Experience and knowledge gained by credit experts. 2. Possibility of obtaining quantitative estimates in cases when there is no statistical information. 3. The speed of obtaining results due to the lack of complex mathematical calculations.	1. Big probability of model error in determining the creditworthiness of a potential borrower is due to the subjective opinion of a specialist. 2. Necessity of long-term training of the employee.
"Naive" classifier Bayes	1. Good results of classification. 2. Simple calculations. 3. Adaptation to new data. 4. High speed. 5. Modest memory requirements. 6. Small amount of training data. 7. Effectively works with categorical data.	1. Requires a sample containing all possible combinations of variables. 2. Assumptions for the independence of variables. 3. The need to convert different types of data to a category.
Discriminant analysis	1. Finding the strongest differences between classes – predictors. 2. Ability to restore missed values. 3. Possibility of solving the problem of classification of new data using past learning.	1. Works with interval data or with a scale of relationships. 2. There is a rather rough method for scoring of the assumptions of the linearity of the discriminatory function.
Logistic regression	1. Ability to choose the number of classes. 2. Ability to find an assessment of the likelihood of non-return of the loan.	1. Complex calculations for obtaining weight coefficients, so it needs a more powerful computer base and advanced computer security. 2. Sensitivity to correlation between characteristics. 3. Low resistance to errors. 4. Dependency on the data set.
Cluster analysis	1. Ability to classify multidimensional observations. 2. Ability to work with indicators that may have non-numeric character. 3. Ability to work on small sample sizes. 4. Possibility of calculations in the conditions of non-fulfillment of normal distribution of random variables.	1. The problem of choosing the distance metric in the space between the centers of the classes. 2. The problem of checking the adequacy of the results.
Decision trees	1. Simple to understand and interpret. 2. Does not require thorough preparation of data. 3. Ability to work both with categorical and interval variables. 4. Allows you to work with a large amount of information without special training procedures.	1. The problem of obtaining an optimal decision tree is NP-task. 2. The problem of retraining is the problem of creating too complex structures that do not sufficiently represent the data.
Neural Networks	1. Good results of classification. 2. Ability to change the structure of the network for new observations. 3. Ability to explain the rather complicated relationship between the values of risk factors and their level.	1. Great statistics for network learning are required. 2. The problem of choosing a network architecture is exist. 3. The difficulty of choosing a learning method is exist. 4 The complexity of calculations of weight coefficients between separate layers during training of a network is exist.

DOI 10.15588/1607-3274-2019-2-7

3 MATERIALS AND METHODS

The first step of assessing the client's creditworthiness with the help of the NBC is the choice of customer characteristics. Consider the most significant indicators that characterize the creditworthiness of clients, and their possible meaning:

– personal indicators;

– indicators on labor activity;

– financial indicators.

Personal indicators of client creditworthiness:

1) Sex = {Man, Woman}

2) Age = {< 25, 25–30, 30–35, 35–45, 45–50, 50–55, >55}

3) Family status = {Unmarried, Married}

4) Number of dependents = {0, 1, ≥ 2}

5) Education = {Secondary, Specialized Secondary, Higher}

6) Period of residence in the region = {< 1 year, 1–3, 3–5, 5–7, 7–10, 10–15, ≥ 15 years}

7) Location = {Center, Region}

Indicators on labor activity of client creditworthiness:

1) Branch of the company = {Banking and financial activity, Information Technology, Non-state medical services, Industry, Construction, Service, Transport and communications, Science or education or culture, State and social organizations, Policy, Agriculture, Other}

2) Enterprise class = {Small, Medium, Great}

3) Professional experience = {< 3, 3–5, 5–10, 10–15, > 15 years}

4) Experience in an enterprise = {<1, 1–3, 3–5, 5–7, 7–10, >10 years}

5) Position at an enterprise = {Manager, not a manager}

Financial indicators of client creditworthiness:

1) Own land = {Have, Have not}

2) Own country house = {Have, Have not}

3) Private property = {Have, Have not}

4) Garage = {Have, Have not}

5) Car = {Domestic, Foreign, Have not}

6) An apartment = {Have, Have not}

7) The area of the apartment = { Have not, from 18 to 25 square meters, 25–32, 32–50, 50–70, ≥ 70}

8) How did obtain an apartment = {Have not, Purchase, Gift}

The second stage of solving the problem of assessing the creditworthiness of a client with the help of NBC is the development of a scoring model. To do this, first of all, it is necessary to build a model of the business process of assessing the creditworthiness of the client. The notation IDEF0 was chosen for modeling process. It let to create the context level of the business process (Fig. 1). The input for the client's creditworthiness assessment process is a questionnaire. It contains general information about the client, financial and social. With the help of a scorecard model, a certain class of creditworthiness is assigned to the client. After IS make a decision to issue a loan. This process is a step-by-step process, therefore, for the implementation of this business process, are needed a loan expert and a financial advisor who is responsible for part of the procedures.

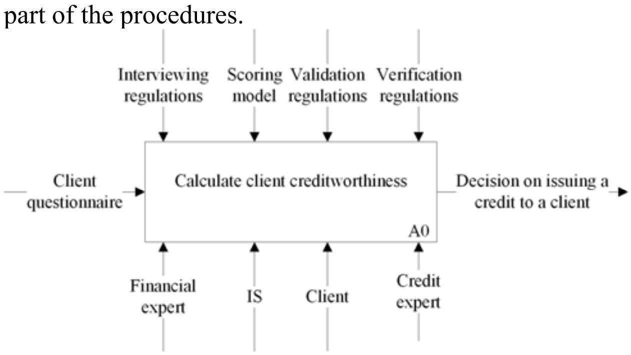

Figure1 – Context level of assessment of client creditworthiness

The list of procedures for evaluating the client's creditworthiness is presented at the next level of detail of the business process (Fig. 2).

The customer credit assessment process consists of four steps:

– an interview: at this stage, the financial consultant conducts a preliminary selection of clients by clarifying all conditions of credit;

– validation and verification of the questionnaire: at this stage, a credit expert with the help of the IS checks the correctness of filling in the questionnaire, and then the correspondence of the personal data to the real information of the client;

– scoring of the questionnaire: the result of this process is the assessment of the client's creditworthiness, which is calculated using a scoring model based on the Bayesian Classifier;

– decision on granting a loan: the final decision on the issue of a loan is taken at this stage.

A business process model for creditworthiness assessment is the basis for developing a scorecard model [15, 16, 20–22].

The NBC is a probabilistic classifier that uses the Bayesian theorem to determine the probability of belonging to the sample element to one of the classes assuming the independence of the variables. The Bayesian Classifier can be applied to any data set that can be represented as categorical data or list of features. A feature is any property that may be present or absent from the sample. The NBC is often very effective when working with the data despite assumptions about the independence of the features. Another important feature of the NBC is that classifier can be built on a new sample with missing values. The reason of it the assumption that the presence or absence of values of variables is completely random. In practice, the NBC, regardless of a number of shortcomings, has proved itself to be quite good due to the high speed of operation, simplicity, scalability and moderate memory requirements. Examples of using are:

– data classification in real time [19, 20];

– classification of texts, for example, spam filtering: Google Analytics and Yandex metrics confirm that Bayesian approach to classification has proven to be very good when it detects spam emails [23];

DOI 10.15588/1607-3274-2019-2-7

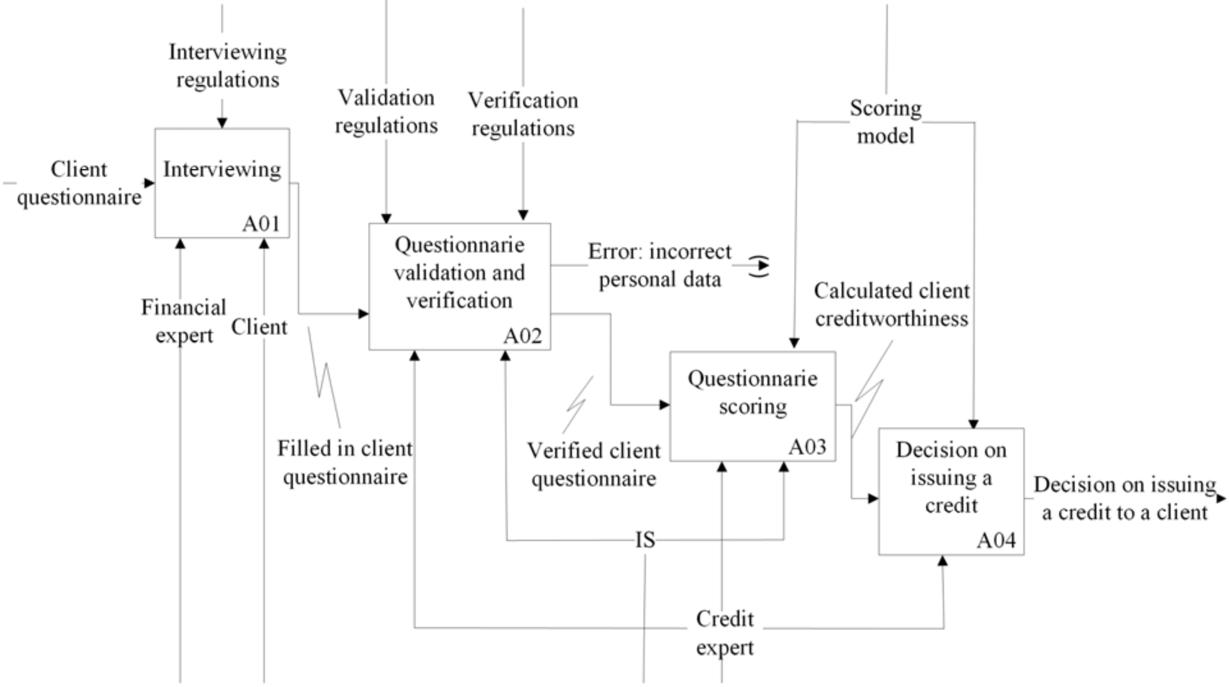

Figure 2 – Assessment process of client creditworthiness

– analysis of the tonality of the text: analysis of social media, identification of positive and negative clients; this approach, in conjunction with collaborative filtering, is implemented in advisory systems [24].

Let consider more detailed using of the NBC for calculating the credit rating. The algorithm of using the Bayesian Classifier to determine the client's creditworthiness consists of two steps:

1) training of the NBC – based on data of clients who have already been issued loans, and who in fact make a monthly contribution;

2) using of the NBC – credit rating is calculated for potential lenders.

Each new potential lender should be classified into one of two classes $Q=\{q_l\}$, $l=\overline{1,2}$: the first class q_1 includes those customers who can give a loan, the second one q_2 includes all other customers.

Taking into account the introduced notation, the customer classification algorithm using the Bayesian Classifier consists of the following steps.

1. Stage of training of the NBC:

I Select set of indicators I . Define set of each i -th characteristic meaning $J_i, i \in I$. Find value for $f_{ij}^k, k \in K, K = \bigcup_{l=1}^{2} K_l$ for each client and its output class $q_{lk}, l=\overline{1,2}$.

II Calculate number of j -th value of i -th indicator f_{ij}^k from all clients for each output class separately

$$x_{ij}^l = \sum_{j \in J_i} \sum_{k \in K_l} f_{ij}^k, \quad K_l \in K, \; l=\overline{1,2} .$$

III Calculate number of l -th output class

$$y_l = \sum_{k \in K_l} q_{lk} .$$

IV Calculate conditional probability $P(f_{ij}/q_l)$ of j -th value of i -th indicator f_{ij} in output class q_l

$$P(f_{ij}/q_l) = \frac{x_{ij}^l}{\sum_{j \in J_i} x_{ij}^l} .$$

V Calculate probability $P(q_l)$ of client's output class

$$P(q_l) = \frac{y_l}{\sum_{l=1}^{2} y_l} .$$

2. Stage of using of the NBC:

VI Enter set of values $\{f_{ij}^{k+1}\}$, $j \in J_i, i \in I$ for checking $k+1$ -th client.

VII Calculate conditional probability $P(q_l/\{f_{ij}^{k+1}\})$ of l -th client's output class according to the conditions $\{f_{ij}^{k+1}\}$, $j \in J_i, i \in I$

$$P(q_l/\{f_{ij}^{k+1}\}) = P(q_l) \cdot \prod_{i,j} P(f_{ij}/q_l), \; i \in I, j \in J_i .$$

VIII Assign of conditional probability to the probability $R(q_l)$ of the output class of $k+1$ -th client

$$R(q_l) = P(q_l/\{f_{ij}^{k+1}\}) .$$

DOI 10.15588/1607-3274-2019-2-7

IX Define output class of $k+1$-th client

$$q_{lk+1} = \arg\max_{l=1,2} R(q_l).$$

For the automation of the calculating process of the credit rating is using a scoring model. The following IS functionality is proposed, which is presented in the form of a use-case diagram (Fig. 3). This diagram shows the user of the IS and all the actions that he can perform. Consider these options for use in more detail way.

In order to work with customer data: to view, to edit, to enter new data, the option “Management client’s data” is provided. By analogy, the option “Management indicators” allows to work with the characteristics used in the scoring model.

The main task of the IS is to assess the creditworthiness of a client. The use case “Calculate creditworthiness” is responsible for it. This functionality involves the use of a scorecard model and the generation of a relevant report on the creditworthiness assessment of clients. The report can be saved to file or can be showed on the screen. The use cases “Use Scoring Model”, “Generate Report”, “Save to File”, “Display to Screen” are responsible for these functionality accordingly. The use case “Get some help” is showing information about the program and FAQ.

The classical three-level architecture is proposed for IS. Each element of architecture is at its own level and responsible for the implementation of a limited set of functions. IS architecture is shown in Fig. 4 in the form of a deployment diagram where it is possible to separate 3 main nodes presented by devices of different purposes.

The client part is presented with the graphical interface of the web application. For efficient operation of IS, customer information is stored in a database that should be located on one of the servers of the credit institution. Also this database can has information about authorizing and accessing the database as a separate entities in the database. In order to access the databases, it is necessary to make requests from client part, which is presented as a web application. The web server retrieves the requests to database. The results of the request (sampling and processing) the web server will return to the client part.

The following model of the database is proposed for storing information on existing and potential lenders (Fig. 5).

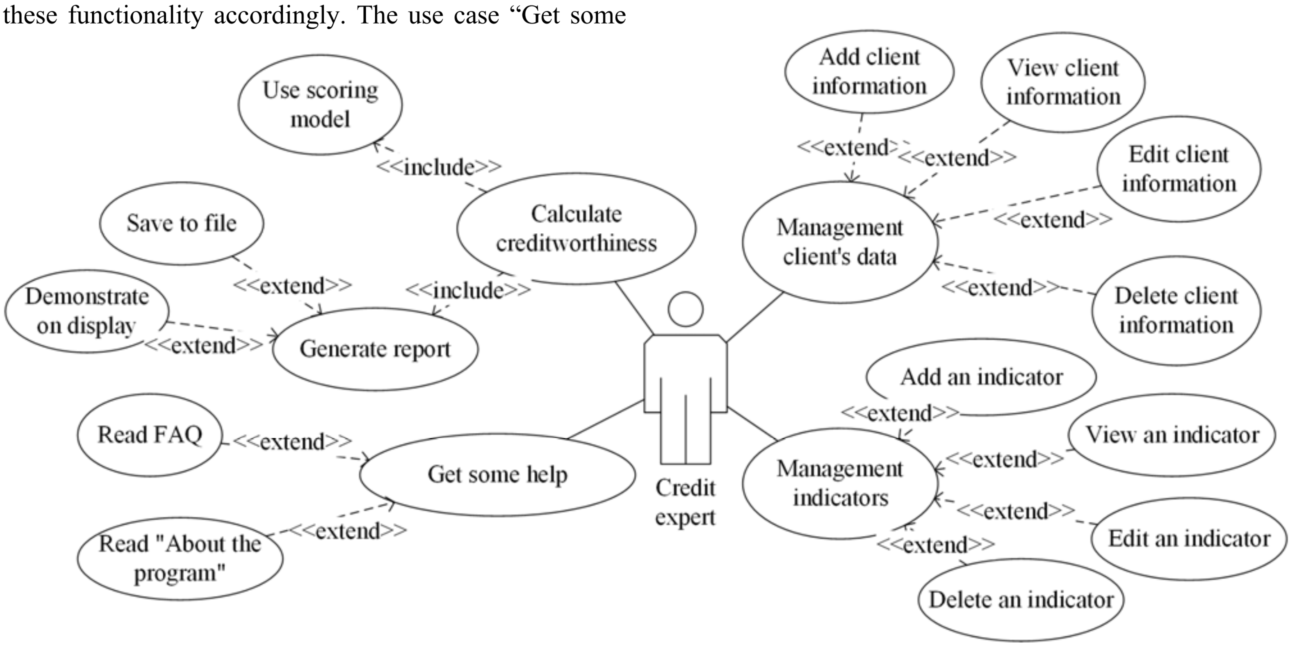

Figure 3 – Use case diagram

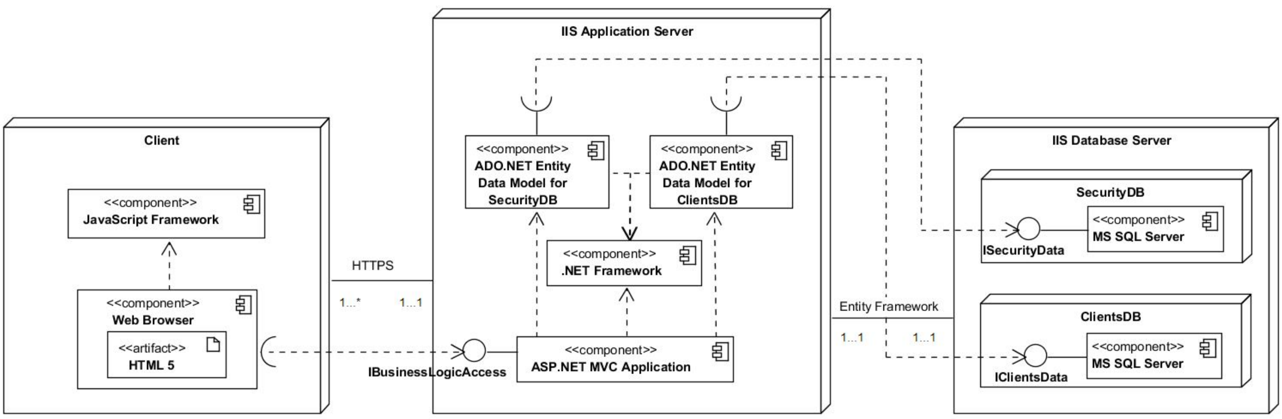

Figure 4 – Deployment diagram

DOI 10.15588/1607-3274-2019-2-7

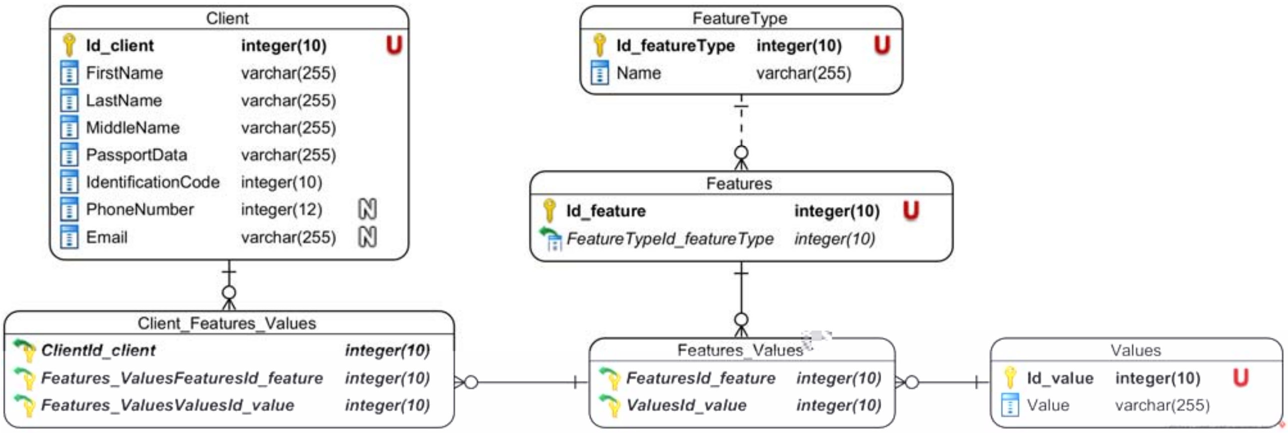

Figure 5 – Database model

The structure of Database model is consist of 6-th entities (Table 2):

– the entity of “Client” corresponds to the main data about clients of a credit institution;

– the entity “FeatureType” describes a group of indicators used to assess the creditworthiness of clients;

– the entity “Features” is a vocabulary for the indicators used for the scorecard model;

– the entity “Values” describes all possible values of all indicators;

– the associative entity “Features_Values” describes the relationship between the entities «Features» and «Values»;

– the associative entity “Client_Features_Values” creates an interconnection between the entities “Client” and “Features_Values”.

Table 2 – Description of client database model

Attribute	Attribute description
Entity “Client”	
Id_client	Client’s Id
FirstName	Client’s name
LastName	Surname
MiddleName	Third name
PassportData	Passport
IdentificationCode	Code of Physical Entity
PhoneNumber	Phone Number
Email	Email
Entity “FeatureType”	
Id_featureType	Id of Feature Group
Name	Group name
Entity “Features”	
Id_feature	Id of Feature
FeatureTypeId_featureType	Type of feature
Entity “Values”	
Id_value	Id of feature value
Value	Value of feature
Entity “Features_Values”	
FeaturesId_feature	Id of Feature
ValuesId_value	Id of feature value
Entity “Client_Features_Values”	
ClientId_client	Client’s Id
Features_ValuesFeaturesId_feature	Id of Feature
Features_ValuesValuesId_value	Id of feature value

So, a scorecard model based on NBC and IS was proposed, which automates the process of solving the client’s creditworthiness assessment task.

4 EXPERIMENTS

For the training of the Bayesian Classifier the data was used for 113 clients who have already been issued a loan or have been denied. Granting a loan was the case if the last line received the value of “1”. An example of input data is presented in Table 3.

Table 3 – Fragment of set with training data

	Set 1	Set 2	....	Set 113
Loan amount	7000	7500	....	10000
Loan term	6	6	....	12
Sex	woman	man	...	woman
Age	37	38	...	41
Family status	Unmarried	Unmarried	....	Married
Number of dependents	2	2	....	2
Education	Higher	Secondary	....	Secondary
Period of residence in the region	22	12	....	21
Location	Region	Center	....	Center
Branch of the company	Industry	Service		Other
Enterprise class	Medium	Small	....	Medium
Professional experience	5	8	....	5
Experience in an enterprise	>10	>10	....	9
Position at an enterprise	not a manager	not a manager	....	not a manager
Own land	Have not	Have not	....	Have
Own country house	Have	Have	....	Have not
Private property	Have not	Have	....	Have not
Garage	Have	Have not	....	Have not
Car	Domestic	Foreign	....	Domestic
An apartment	Have not	Have	....	Have
The area of the apartment	37	29	....	38
How did obtain an apartment	Have not	Have not	....	Have not
To give a loan	1	1	....	0

DOI 10.15588/1607-3274-2019-2-7

All information from the Table 3 was divided into three types of indicators: personal indicators, indicators on labor activity, and financial indicators. After that each indicator obtained appropriate meaning.

5 RESULTS

The calculation of the creditworthiness of clients is based on the trained Bayesian classifier. Using the NBC was conducted on the example of three clients, one of them was denied in the loan, for the second and third one scoring model calculated the value of the credit assessment, which corresponds to the economic expediency of the loan. Example of calculated results and client's information according to table 3 are presented below.

Client 143 = {25500; 12; man; 25 – 30; unmarried; 0; higher; 1 – 3; Center; Service; Medium; < 3; <1; not a manager; Have not; Have not; Have not; Have not; Have not; Have not; Have not; Have not; 0}

Client 97 = {14500; 12; man; >55; married; 1; higher; ≥ 15 years; Center; Banking and financial activity; Great; > 15 years; 7 – 10; Manager; Have; Have; Have; Have; Foreign; Have not; 32 – 50; Have not; 1}

Client 136 = {17000; 24; man; 30–35; married; 1; higher; ≥ 15 years; Center; Information Technology; Medium; 5 – 10; 3 – 5; Manager; Have not; Have not; Have; Have; Foreign; Have; 18 – 25; Gift; 1}

Clients 97 and 136 have a good credit assessment, so they can obtain the loan. Another one doesn't have any financial support, so his credit value is very small for issuing the loan.

The issuance of money to the client, which the scoring model has calculated the positive value of the credit assessment, occurs after signing by the bank and the client of the loan agreement and other agreements (mortgages, guarantees of commercial pledge, etc.).

6 DISCUSSION

To use the developed IS in real conditions, it is necessary to overcome the adequacy of the scoring model. To do this, it is necessary to check the quality of the proposed classifier, which is the basis for making decisions in the credit activity. For the checking NBC we used client's data, which did not use in the training classifier.

The quality of credit activity is an integral characteristic of a process that shows the degree of suitability of a process for achieving certain goals. To assess the quality of the classifier, we can use the following quality indicators: precision, recall, measure of Van Risbergen [25, 26].

Precision is a criterion that shows the proportion of clients that really belong to a particular class with respect to all clients that IS has identified about this class: $P = \frac{a}{a+b}$. The more precision, the fewer the number of false alarms.

Recall is a quality metric that shows the proportion of clients that really belong to a particular class with respect to all clients from this class: $R = \frac{a}{a+c}$. The more recall of the data obtained from the IS, the greater the benefit of the received information.

A good metric for a joint assessment of precision and recall is the F-measure or measure of Van Risbergen, which is defined as the weighted harmonic average of precision and recall: $F = \frac{2PR}{P+R}$.

To check the quality of the work of the NBC 61 clients were taken, which were not used in the training. A series of experiments were carried out with different conditions. Firstly, the NBC obtained 61 cards for verification, then some of these cards were used for training, and the remainder was again checked for the definition of customer classes. The process was repeated until all data was used. The results of the experiments are shown in table 4. We can also evaluate the client class definition error: $Error = \frac{a+c}{number\ of\ clients}$.

Table 4 – Result of checking NBC

№	a	b	c	Number of clients	P	R	F	Error
1	33	5	3	61	0,87	0,92	0,89	0,13
2	25	4	2	52	0,86	0,93	0,89	0,12
3	23	2	2	43	0,92	0,92	0,92	0,09
4	23	2	1	37	0,92	0,96	0,94	0,08
5	15	1	1	20	0,94	0,94	0,94	0,10
6	10	1	0	16	0,91	1,00	0,95	0,06

As can be seen from Table 4, as the training pattern increases the Van Riesbergen measure increases (Fig. 6). The error in determining the client's credit rating decreases.

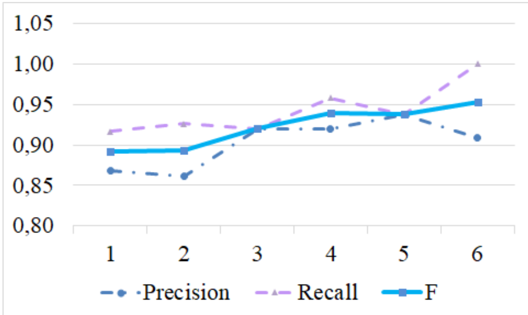

Figure 6 – Quality metrics

The IS considered earlier [12–14] are characterized by the closeness of the scoring model used for decision making. This restriction does not allow doing many experiments to test the performance of these IS.

The conducted studies and the obtained results show the feasibility of using the proposed scoring model and the developed IS in real conditions.

CONCLUSIONS

In the course of this research, the task of assessing the creditworthiness of a potential creditor was solved. To do this, an analytical review of the main methods for assessing creditworthiness was conducted. Result of it is the method of scoring-analysis based on Data Mining tech-

DOI 10.15588/1607-3274-2019-2-7

nology. A scorecard model for calculating the credit rating based on the use of the NBC was developed.

A design stage of IS was conducted, during which a diagram of activity was developed for a clear presentation of the work of the classification algorithm, a use-case diagram was developed for allocating functional capabilities, a deployment diagram was developed for representing the architecture of the IS. A database model was developed that allows the knowledge domain to be structured. The conducted pilot studies have shown **the practical significance** of the results for credit institutions, as evidenced by the calculated assessment of the quality of credit activity.

The scientific novelty of the obtained results consists in improving the client's credit rating process by using the scoring model based on NBC, which reduced the subjective factor in decision making and also reduced the time for processing information about a potential creditor.

The results obtained in the course of this research work indicate the feasibility of using the proposed solution under real conditions in credit institutions to increase the quality of lending activities.

Prospects for further research are to automate the process of selecting informative indicators of customer creditworthiness.

REFERENCES

1. Bank Lending Survey [Electronic resource] / Access mode: https://bank.gov.ua/control/en/publish/category;jsessionid=4AEC639346EC2333E34BA6A8D94B48CA?cat_id=20741795, 03.12.2018.
2. Carlson Mark, Shan Hui, Warusawitharana Missaka Capital ratios and bank lending: A matched bank approach, *Journal of Financial Intermediation*, 2013, Volume 22, Issue 4, pp. 663–687. https://doi.org/10.1016/j.jfi.2013.06.003
3. Ling Kock Sheng, Teh Ying Wah A comparative study of data mining techniques in predicting consumers' credit card risk in banks, *African Journal of Business Management,* 2011, Vol. 5 (20), pp. 8307–8312.
4. Polozhennya pro kredituvannya, zatv. Postanovoyu Pravlinnya NBU 28.09.1995 № 246, *Pravove regulyuvannya kreditnix vidnosin v Ukraïni: 36 normat. Aktiv*, Kiev, Yurinkom Inter, 2001, pp. 53–66.
5. Kevin Johnston. How to Evaluate a Firm's Credit Worthiness [Electronic resource], Access mode: https://smallbusiness.chron.com/evaluate-firms-credit-worthiness-25925.html, 30.11.2018.
6. Celan Bryant. How to Determine The Creditworthiness of a Customer [Electronic resource], Access mode: https://blog.apruve.com/how-to-determine-the-creditworthiness-of-a-customer. – 03.12.2018.
7. Motwani A., Chaurasiya P., Bajaj G. Predicting Credit Worthiness of Bank Customer with Machine Learning Over Cloud, *International journal of computer sciences and engineering*, 2018, No. 6(7), pp. 1471–1477. DOI: 10.26438/ijcse/v6i7.14711477
8. Shvidkij A. I., Miroshnichenko A. A. Metody ocenki kreditosposobnosti korporativnyx klientov kommercheskogo banka: rossijskij i zarubezhnyj opyt, *Mezhdunarodnyj zhurnal prikladnyx i fundamental'nyx issledovanij*, 2016, No. 7–4, pp. 667–672.
9. Gotovchikov I. F. Prakticheskij metod e'kspress-ocenki finansovyx vozmozhnostej fizicheskix i yuridicheskix lic, *Bankovskoe kreditovanie,* 2009, No. 3, P. 115.
10. Pramod S. Pati, Aghav Dr. J. V., Sareen Vikram An Overview of Classification Algorithms and Ensemble Methods in Personal Credit Scoring, *International Journal of Computer Science and technology*, 2016, Vol. 7, Issue 2, pp. 183–188.
11. Thabiso Peter Mpofu, Mukosera Macdonald Credit Scoring Techniques: A Survey, *International Journal of Computer Science and technology*, 2014, Vol. 3, Issue 8, pp. 165–168.
12. Ukrainian bureau credit history [Electronic resource] Access mode: https://ubki.ua/ua . 05.12.2018.
13. Mobile app "Credit history" [Electronic resource] Access mode:https://play.google.com/store/apps/details?id=ua.ubki&hl=uk . 05.12.2018.
14. Internet-bank Privat24 [Electronic resource], Access mode: https://www.privat24.ua/ . 10.12.2018.
15. Eibe Frank, Mark A. Hall, Christopher J. Palestro and Ian H. Witten Data Mining: Practical Machine Learning Tools and Techniques, *Elsevier Science & Technology Books*, 2016, 654 p.
16. Kesavaraj G., Sukumaran S. A study on classification techniques in data mining, *Fourth International Conference on Computing, Communications and Networking Technologies, Tiruchengode,* 2013, pp. 1–7. DOI: 10.1109/ICCCNT.2013.6726842
17. Mel'nik K. V., Єrshova S. I. Problemy i osnovnye podxody k resheniyu zadachi medicinskoj diagnostiki, *Sistemi obrobki informacii*, 2011, No. 2 (92), pp. 244–248.
18. Mariya Yao, Adelyn Zhou, Marlene Jia Applied Artificial Intelligence: A Handbook For Business Leaders Paperback, *Topbots Inc*, 2018, 228 p.
19. Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning, MIT Press, 2016, 800 p.
20. Naive Bayes Classifiers [Electronic resource], Access mode: https://www.geeksforgeeks.org/naive-bayes-classifiers/, 04.12.2018.
21. Mel'nik K. V., Glushko V. N. Primenenie apparata Bajesovyx setej pri obrabotke dannyx iz medicinskix kartochek, *Science and Education a New Dimension: Natural and Technical Sciences,* 2013, I (2), Issue 15, pp. 126–129. Vengriya, Budapesht.
22. Mel'nik K. V., Goloskokov A. E. Ispol'zovanie setej doveriya dlya zadachi skrininga, T*ezi dopovidej mizhnarodnoï naukovo-praktichnoï konferencii «Informacijni texnologii: nauka, texnika, texnologiya, osvita, zdorov'ya»*. Xarkiv, NTU «XPI», 2014, P. 14.
23. Google Analytics Referral Spam [Electronic resource], Access mode: https://medium.com/@lenguyenthedat/google-analytics-referral-spam-85bb6b7aed2b, 02.11.2018.
24. Koji Miyahara, Michael J. Pazzani Improvement of Collaborative Filtering with the Simple Bayesian Classifier [Electronic resource], University of California, Irvine. Access mode: https://www.ics.uci.edu/~pazzani/Publications/IPSJ.pdf. 01.12.2018.
25. Mel'nik K. V. Ocinka yakosti medichnoï informacii, *Materiali Mizhnarodnoï naukovo-praktichnoï konferencii "Problemi i perspektivi rozvitku IT industrii*". Xar'kov, XNEU imeni Semena Kuznecya, 2018, P. 70.
26. Woodall P. M., Oberhofer M., & Borek A. A Classification of Data Quality Assessment and Improvement Methods, *International Journal of Information Quality,* 2014, No. 3. https://doi.org/10.1504/IJIQ.2014.068656.

Received 10.12.2018.
Accepted 12.02.2019.

DOI 10.15588/1607-3274-2019-2-7

e-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2019. № 2
p-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2019. № 2

УДК 004.9

ПІДВИЩЕННЯ ЯКОСТІ КРЕДИТНОЇ ДІЯЛЬНОСТІ ЗА РАХУНОК ВИКОРИСТАННЯ СКОРИНГОВОЇ МОДЕЛІ

Мельник К. В. – кандидат технічних наук, доцент кафедри Програмної інженерії та інформаційних технологій управління, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна.

Борисова Н. В. – кандидат технічних наук, доцент кафедри Інтелектуальних комп'ютерних систем, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто проблему оцінки кредитоспроможності клієнта. Вона являє собою одночасну обробку даних кредитора різної природи з подальшим визначенням кредитного рейтингу. Об'єктом даного дослідження виступав процес кредитування фізичних осіб кредитними установами.

Ціль. Ціль роботи є дослідження процесу підвищення якості кредитної діяльності за рахунок розробки і використання скорингової моделі.

Метод. Проведено аналітичний огляд предметної області. Розроблено модель бізнес-процесу оцінювання кредитоспроможності клієнтів у вигляді IDEF0 діаграми. Виділені групи показників, які характеризують потенційного кредитора з різних боків. Обрані набори значень для кожного показника кредитоспроможності окремо. Проаналізовано методи вирішення задачі оцінки кредитоспроможності клієнтів. Обрано наївний класифікатор Байєсу як метод вирішення задачі класифікації потенційних кредиторів. Проаналізовано існуючі інформаційні системи для оцінки кредитоспроможності клієнтів. Розроблено скорингову модель оцінки кредитного рейтингу клієнту у вигляді алгоритму. Визначено перелік функціональних можливостей інформаційної системи, які представлено у вигляді діаграми варіантів використання. Запропоновано трирівневу архітектуру для інформаційної системи. Розроблено модель бази даних для збереження інформації щодо клієнтів. Розроблено інформаційну систему для визначення кредитного рейтингу клієнту на основі використання розробленої скорингової моделі. Проведено чисельні дослідження визначення класу потенційного кредитора. Проаналізовано процес визначення якості кредитної діяльності. Обрано показники якості визначення оцінки кредитоспроможності клієнтів. Запропоновано спосіб розрахунку якості кредитної діяльності.

Результати. Розроблено скорингову модель, яка була використана при вирішенні оцінки кредитоспроможності клієнтів за допомогою запропонованої інформаційної системи. Досліджено процес підвищення якості оцінки кредитного рейтингу.

Висновки. Проведені експерименти підтвердили значимість запропонованої скорингової моделі та дозволяють рекомендувати її для використання на практиці для аналізу процесу кредитоспроможності клієнтів. Наукова новизна полягає в удосконаленні процесу кредитної діяльності за рахунок автоматизації використання наївного класифікатора Байєсу, що дозволяє зменшити людський фактор при прийнятті рішень.

КЛЮЧОВІ СЛОВА: скорингова модель, задача класифікації, наївний байєсівський класифікатор, оцінка кредитного рейтингу, кредитування, позичальник, кредитор, кредитоспроможність.

УДК 004.9

ПОВЫШЕНИЕ КАЧЕСТВА КРЕДИТНОЙ ДЕЯТЕЛЬНОСТИ ЗА СЧЕТ ИСПОЛЬЗОВАНИЯ СКОРИНГОВОЙ МОДЕЛИ

Мельник К. В. – кандидат технических наук, доцент кафедры Программной инженерии и информационных технологий управления, Национальный технический университет «Харьковский политехнический институт», Харьков, Украина.

Борисова Н. В. – кандидат технических наук, доцент кафедры интеллектуальных компьютерных систем, Национальный технический университет «Харьковский политехнический институт», Харьков, Украина.

АННОТАЦИЯ

Актуальность. Рассмотрена проблема оценки кредитоспособности клиента, которая представляет собой одновременную обработку данных кредитора различной природы с последующим определением кредитного рейтинга. Объектом данного исследования выступал процесс кредитования физических лиц кредитными учреждениями.

Цель. Целью работы является исследование процесса повышения качества кредитной деятельности за счет разработки и использования скоринговой модели.

Метод. Проведен аналитический обзор предметной области. Разработана модель бизнес-процесса оценки кредитоспособности клиентов в виде IDEF0 диаграммы. Выделены группы показателей, характеризующих потенциального кредитора с разных сторон. Определены наборы значений для каждого показателя кредитоспособности отдельно. Проанализированы методы решения задачи оценки кредитоспособности клиентов. Избран наивный классификатор Байеса как метод решения задачи классификации потенциальных кредиторов. Проанализированы существующие информационные системы для оценки кредитоспособности клиентов. Разработана скоринговая модель оценки кредитного рейтинга клиента в виде алгоритма. Определен перечень функциональных возможностей информационной системы, представленных в виде диаграммы вариантов использования. Предложена трехуровневая архитектура информационной системы. Разработана модель базы данных для хранения информации о клиентах. Разработана информационная система для определения кредитного рейтинга клиенту на основе использования предложенной скоринговой модели. Проведены численные исследования определения класса потенциального кредитора. Проанализирован процесс определения качества кредитной деятельности. Выбраны показатели качества определения оценки кредитоспособности клиентов. Предложен способ расчета качества кредитной деятельности.

Результаты. Разработана скоринговая модель, которая была использована при оценке кредитоспособности клиентов с помощью предложенной информационной системы. Исследован процесс повышения качества оценки кредитного рейтинга.

DOI 10.15588/1607-3274-2019-2-7

Выводы. Проведенные эксперименты подтвердили значимость предложенной скоринговой модели и позволяют рекомендовать ее для использования на практике для анализа процесса кредитоспособности клиентов. Научная новизна заключается в усовершенствовании процесса кредитной деятельности за счет автоматизации использования наивного классификатора Байеса, что позволяет уменьшить человеческий фактор при принятии решений.

КЛЮЧЕВЫЕ СЛОВА: скоринговая модель, задача классификации, наивный байесовский классификатор, оценка кредитного рейтингу, кредитование, заемщик, кредитор, кредитоспособность.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Bank Lending Survey [Electronic resource] / Access mode: https://bank.gov.ua/control/en/publish/category;jsessionid=4AEC639346EC2333E34BA6A8D94B48CA?cat_id=20741795 . – 03.12.2018.
2. Mark Carlson. Capital ratios and bank lending: A matched bank approach / Mark Carlson, Hui Shan, Missaka Warusawitharana // Journal of Financial Intermediation. – 2013. – Volume 22, Issue 4. – P. 663–687. https://doi.org/10.1016/j.jfi.2013.06.003
3. Ling Kock Sheng. A comparative study of data mining techniques in predicting consumers' credit card risk in banks / Ling Kock Sheng and Teh Ying Wah // African Journal of Business Management. – 2011. – Vol. 5 (20). – P. 8307–8312.
4. Положення про кредитування / затв. Постановою Правління НБУ 28.09.1995 № 246 // Правове регулювання кредитних відносин в Україні: 36 нормат. Актів. – К. : Юрінком Інтер, 2001. – С. 53–66.
5. Kevin Johnston. How to Evaluate a Firm's Credit Worthiness [Electronic resource] / Kevin Johnston. – Access mode: https://smallbusiness.chron.com/evaluate-firms-credit-worthiness-25925.html . – 30.11.2018.
6. Celan Bryant. How to Determine The Creditworthiness of a Customer [Electronic resource] / Celan Bryant. – Access mode: https://blog.apruve.com/how-to-determine-the-creditworthiness-of-a-customer. – 03.12.2018.
7. Motwani A. Predicting Credit Worthiness of Bank Customer with Machine Learning Over Cloud / A. Motwani, P. Chaurasiya, G. Bajaj // International journal of computer sciences and engineering. – 2018. – № 6(7). – P. 1471–1477. DOI: 10.26438/ijcse/v6i7.14711477
8. Швидкий А. И. Методы оценки кредитоспособности корпоративных клиентов коммерческого банка: российский и зарубежный опыт / А. И. Швидкий, А. А. Мирошниченко // Международный журнал прикладных и фундаментальных исследований. – 2016. – № 7–4. – С. 667–672.
9. Готовчиков И. Ф. Практический метод экспресс-оценки финансовых возможностей физических и юридических лиц / И. Ф. Готовчиков // Банковское кредитование. – 2009. – № 3. – С. 115.
10. Pramod S. Patil. An Overview of Classification Algorithms and Ensemble Methods in Personal Credit Scoring / Pramod S. Patil, Dr. J. V. Aghav, Vikram Sareen // International Journal of Computer Science and technology. – Vol. 7, Issue 2. – 2016. – P. 183–188.
11. Thabiso Peter Mpofu. Credit Scoring Techniques: A Survey / Thabiso Peter Mpofu, Macdonald Mukosera // International Journal of Computer Science and technology. – 2014. – Vol. 3, Issue 8. – P. 165–168.
12. Ukrainian bureau credit history [Electronic resource] / Access mode: https://ubki.ua/ua . – 05.12.2018.
13. Mobile app "Credit history" [Electronic resource] / Access mode: https://play.google.com/store/apps/details?id=ua.ubki&hl=uk. – 05.12.2018.
14. Internet-bank Privat24 [Electronic resource] / Access mode: https://www.privat24.ua/ . – 10.12.2018.
15. Data Mining: Practical Machine Learning Tools and Techniques / [Eibe Frank, Mark A. Hall, Christopher J. Palestro and Ian H. Witten] // Elsevier Science & Technology Books, 2016. – 654 p.
16. Kesavaraj G. A study on classification techniques in data mining / G. Kesavaraj and S. Sukumaran // Fourth International Conference on Computing, Communications and Networking Technologies, Tiruchengode, 2013. – P. 1–7. doi: 10.1109/ICCCNT.2013.6726842
17. Мельник К. В. Проблемы и основные подходы к решению задачи медицинской диагностики / К. В. Мельник, С. І. Єршова // Системи обробки інформації. – 2011. – № 2 (92). – С. 244–248.
18. Mariya Yao. Applied Artificial Intelligence: A Handbook For Business Leaders Paperback / Mariya Yao, Adelyn Zhou, Marlene Jia // Topbots Inc. – 2018. – 228 p.
19. Ian Goodfellow. Deep Learning / Ian Goodfellow, Yoshua Bengio, Aaron Courville // MIT Press, 2016. – 800 p.
20. Naive Bayes Classifiers [Electronic resource] / Access mode: https://www.geeksforgeeks.org/naive-bayes-classifiers/ . – 04.12.2018.
21. Мельник К. В. Применение аппарата Байесовых сетей при обработке данных из медицинских карточек / К. В. Мельник, В. Н. Глушко // Science and Education a New Dimension: Natural and Technical Sciences. – 2013. – I (2), Issue 15. – P. 126–129. Венгрия, Будапешт.
22. Мельник К. В. Использование сетей доверия для задачи скрининга / К. В. Мельник, А. Е. Голоскоков // Тези доповідей міжнародної науково-практичної конференції «Інформаційні технології: наука, техніка, технологія, освіта, здоров'я». – Харків : НТУ «ХПІ», 2014. – С. 14.
23. Google Analytics Referral Spam [Electronic resource] / Access mode: https://medium.com/@lenguyenthedat/google-analytics-referral-spam-85bb6b7aed2b . – 02.11.2018.
24. Koji Miyahara. Improvement of Collaborative Filtering with the Simple Bayesian Classifier [Electronic resource] / Koji Miyahara and Michael J. Pazzani. – University of California, Irvine. – Access mode: https://www.ics.uci.edu/~pazzani/Publications/IPSJ.pdf. – 01.12.2018.
25. Мельник К. В. Оцінка якості медичної інформації / К. В. Мельник // Матеріали Міжнародної науково-практичної конференції «Проблеми і перспективи розвитку ІТ індустрії». – Х. : ХНЕУ імені Семена Кузнеця, 2018. – С. 70.
26. Woodall P. M. A Classification of Data Quality Assessment and Improvement Methods / P. M. Woodall, M. Oberhofer, & A. Borek // International Journal of Information Quality. – 2014. – № 3. https://doi.org/10.1504/IJIQ.2014.068656.

DOI 10.15588/1607-3274-2019-2-7