Customer Classification and Decision Making in the Digital Economy Based on Scoring Models

Article *in* International Journal of Management Cases · August 2020

DOI: 10.34218/IJM.11.6.2020.134

6 authors, including:

Yurii Popovskyi
Vasyl` Stus Donetsk National University
7 PUBLICATIONS **25** CITATIONS
SEE PROFILE

Oleksandr Korniichuk
National Academy of Sciences of Ukraine
7 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Serhii Kozlovskyi
Vasyl' Stus Donetsk National University
53 PUBLICATIONS **489** CITATIONS
SEE PROFILE

International Journal of Management (IJM)
Volume 11, Issue 6, June 2020, pp. 1463-1481, Article ID: IJM_11_06_134
Available online at http://www.iaeme.com/IJM/issues.asp?JType=IJM&VType=11&IType=6
ISSN Print: 0976-6502 and ISSN Online: 0976-6510
DOI: 10.34218/IJM.11.6.2020.134

CUSTOMER CLASSIFICATION AND DECISION MAKING IN THE DIGITAL ECONOMY BASED ON SCORING MODELS

Ruslan Lavrov
Department of Finance, Banking and Insurance,
Chernihiv National University of Technology, Chernihiv, Ukraine

Natalia Burkina
Department of Business Statistics and Economic Cybernetics,
Vasyl' Stus Donetsk National University, Vinnytsia, Ukraine

Yurii Popovskyi
Department of Business Statistics and Economic Cybernetics,
Vasyl' Stus Donetsk National University, Vinnytsia, Ukraine

Serhii Vitvitskyi
Department of Legal Disciplines, Donetsk Law Institute of the Ministry of Internal Affairs of Ukraine, Kryvyi Rih, Ukraine

Oleksandr Korniichuk
Institute of Feed and Agriculture of Podillya National Academy of Agrarian Sciences of Ukraine, Vinnytsia, Ukraine

Serhii Kozlovskyi
Department of Entrepreneurship, Corporate and Spatial Economics,
Vasyl' Stus Donetsk National University, Vinnytsia, Ukraine

ABSTRACT

The article presents how cluster models works to create customer classification and to make managerial decision for saving clients and founding new target auditory. The objective of research is to find out the relevant techniques for building scoring models in different fields. The main hypothesis of research was checking the quantity of scoring models in different fields. It was applied k-nearest neighbors support vector method for decision making in the digital economy based on scoring models. In order to realize the principle of customer classification and revealing the client categories with risk of leaving the company it was created the client's classification model. Moreover, risk issue was shown on the example of fraud dynamic. It was researched different categories of fraud and pointed out their features. According the results of

the building models it was proposed some recommendation about decision making in the risk situation. The model shows how to save existing clients and how to share client base through the finding of client groups portraits and how to be carefully in the risk situation.

Jel Classification: *A1, C1, C5, C6, K1, K33.*

Key words: modelling, decision making, algorithms, scoring models, customer classification, digital economy, cluster analysis

Cite this Article: Ruslan Lavrov, Natalia Burkina, Yurii Popovskyi, Serhii Vitvitskyi, Oleksandr Korniichuk and Serhii Kozlovskyi, Customer Classification and Decision Making in the Digital Economy Based on Scoring Models, *International Journal of Management*, 11(6), 2020, pp. 1463-1481.
http://www.iaeme.com/IJM/issues.asp?JType=IJM&VType=11&IType=6

1. INTRODUCTION

With the development and widespread use of computer technologies, there are significant changes in the economy. Many activities are increasingly being transferred to the Internet. In many areas (for example, insurance, health, science, agriculture), past information is heavily digitized [1]. All them allows to form large volumes of data (Big Data), processing of which becomes an additional direction of socio-economic, technical, scientific analysis, allows to establish new logical patterns and to make management decisions based on them [2].

The development and visual effectiveness of digital technologies have a significant impact on the trajectories of the economy and society. The electronic toolkit leads to radical changes in people's lives and is one of the priorities for most economic leaders, including the United States, Germany, China, Japan etc. The creation of new technologies and the introduction of their business brings about global changes that are associated with the emergence of new digital infrastructures, the rapid development of digital communications and the improvement of computer technology.

The integration of these technologies into the economic and socio-political life of society testifies to the formation of a new system of the world economy – digital [3]. Devices for IT technologies every day are technically improved and relatively cheaper. The service market is replenished with new technologies for interacting with customers and makes it possible to respond in a timely manner to various changes in business relations. Lagging the technology market means taking risks and becoming uncompetitive or even squeezed out of the market. Scoring, what is it and what features does the scoring procedure have? Making a bank profit directly depends on the quality of the loan portfolio. The smaller the financial risks, the greater the likelihood of a quick return of borrowed funds with an additional profit from the payment of interest. That’s why, considering applications for a loan, the bank carries out a thorough examination of potential customers, analyzing possible financial risks. For each client, the entrepreneur estimates the likelihood of retaining the person as a client for a maximum possible time. To realize this idea, the risk of losing a client for the future is again assessed. Introducing new types of products to the market, an entrepreneur assesses the risk of successful introduction of this product in a new market, and searches for certain categories of target audiences that are most suitable for consumers of their products.

In all these cases, as well as in many others, economists use scoring models that help reduce risks during economic activity, find, maintain and increase customers, optimize the production and supply of new products, as well as increase profits. Scoring is a heuristic way of developing ratings and classifying different objects into groups. It assumes that people with

similar social indicators behave the same way. Today it is used in banking, marketing, insurance and many others economical and legal areas. Literally, "scoring" means the counting of scores. This article shows what kind of points do modern analysts consider and what do they need it for.

2. LITERATURE REVIEW

The term “digital economy” is very popular nowadays. But there is no single definition of it. The digital economy is the creation, distribution and use of digital technologies and related products and services. Digital technologies are technologies for the collection, storage, processing, retrieval, transmission and submission of data electronically [4]. Richard Heeks, in his research on Information and Communication Technologies for Development, points out that these technologies create new opportunities in the digital sphere: an entrepreneur or a company can optionally use the digital system in their activities [3]. This process may include datafication (implementation of storage technologies for large arrays), digitization (conversion of all parts of the value chain from analogue to digital format), virtualization (physical decomposition of processes), and generativity (use of data and technology in a new, different from the original , assignments by reprogramming and recombination) [3]. Thus, by the generalizations of R. Bucht [5], the term "digital economy" refers exclusively to the events that are currently underway and the unfinished transformation of all sectors of the economy due to the digitization of information by using computer technology. Because with the development and availability of information systems and technologies, the Internet, microcomputers (payment terminals, mobile payment points), smartphones, netbooks, etc. entrepreneurs and buyers have found it convenient to make financial and economic agreements.

The digital economy is the most important topic of scientific debate and especially of the legal aspects of interaction with economic processes. In today's environment, the digital economy is an unrestricted way of doing business and requires the establishment of sound state control mechanisms for the activity of economic entities in the legislation. The study of the regulation of digital relations is a strategic task of national and interstate policy aimed at ensuring the security of the entire modern world. Undoubtedly the root cause of the emergence of a new organization form of the economy was the emergence of a new technological way, as well as not covered types of economic activity occurring in the forms of interaction between the state and its citizens. Because the Internet has virtually endless access to electronic resources, including online stores, which in turn allows you to shop. This approach requires the definition of concepts and conditions at the regulatory level in different countries of the world. In most cases, misunderstandings and disputes arise in unforeseen situations. For example, buying real estate with the possibility of electronic signature, which in turn is a risk of e-commerce.

Certainly, often controversial situations that arise in the digital environment can be resolved through current legislation. The difficulty is usually caused by the fact that many legal provisions, of course, do not explicitly provide for their application to Internet relations. In such circumstances, an interpretation of the legal provisions is required. In studies of Cherdantsev V., Kobelev P. it is noted that the digital economy encompasses a complex of electronic business operations and e-commerce, and the corresponding infrastructure is included in this concept. The process of informatization affects the regulation of production, financial circulation and e-commerce. Where to use the global information network environment for communication [6]. Mishko F., Vasilyeva K., Popov V. researching "Trends in the Legal Regulation of Civil and Competitive Relations in the Digital Economy" determine that the implementation of digital agreements requires the separation of concepts of

digital offer and digital acceptance and expanding the list of civil law objects by including the terms "information" and "digital financial assets" [7]. The digital economy is evolving not only to pay for products or to organize relationships in the economic sphere, but also as a tool for market research and decision-making based on analysis and neural networks. Based on the smart technologies of economic processes interaction, it would be advisable to develop the direction of smart contract models.

A. Vashkevich. describes in detail in his work “Smart Contracts: What and Why” that a significant part of the norms can be algorithmized and the regulation become automated. Smart contracts have become known along with blockchain and cryptocurrency technology. Now it is a part of mythology with high expectations or unreasonable fears. But smart contracts are much wider than algorithms that translate tokens. They will change the surrounding legal reality, the work of lawyers and the life of business [8]. Access to data to digital economy became a key factor in development of products and innovations, and data collection and their use by the third parties generates a set of the controversial issues concerning protection of the competition and the rights of businessmen. An important factor of successful international business in the field of digital economy is ensuring the honest competition. Rapid growth of innovations and use of new technologies within digital economy sometimes advances traditional models of regulation therefore state policy can consider not fully the growing competition in various industries.

In scientific research of "Feature of legal regulation of digital intellectual economy" of K.M. Belikov, claims, - "to guarantee the open markets, innovations, quality and efficiency and also freedom of choice for consumers, the effective competition has to be protected from restrictions". [9]

As instruments of protection of the competition in the conditions of digital economy allocate the following: ban on the conclusion of anti-competitive agreements; the ban on abuse of a dominant position in the market; control of merges for prevention of domination in the market and prevention of creation of essential obstacles for the effective competition [10].

Different approaches to legal regulation in the sphere of digital economy, as a rule, meet that it is necessary to provide such legal regime within which innovations, on the one hand, will freely develop, and, on the other hand, possible risks will be prevented. At the same time as one of risks most often mark out impossibility to precisely predict a way which development of innovations in the sphere of digital economy therefore the adopted legislation has to be rather flexible will go and be formed taking into account as it is possible the bigger number of data [11]. Around the world the global world financial institutions face various problems among which digital fraud takes not the last place. Fast development of communications and information technologies, and their active application in the sphere of financial services, promotes distribution of various types of fraud, losses from which are estimated at billions of dollars. For the purpose of prevention of fraud in the sphere of business by the company of the Lab neurodat it is developed technology of assessment of emotions of people on the basis of the set parameters. Specially under this task developed the Emotion Miner platform which continues work and allows to analyze video. Collected data formed the basis of methods of training of neuronets in recognition of human emotions. Algorithms pay attention to a voice (tone height, a timbre, loudness, pauses in language), emotional coloring and semantics of the text, a mimicry the person, speed and the direction of movements of a body and position of separate extremities, heart rate on the basis of changes of skin color, breath on the movement of a thorax and also a sex, age of the person and presence at it on a face of points, moustaches, beards.

The result of work received multimodal architecture which at the same time can analyze audio, video, gestures and physiological parameters. Development is planned to be used branches of business, advertizing, spheres of safety and medicine [12]. Different approaches to legal regulation in the sphere of digital economy, as a rule, meet that it is necessary to provide such legal regime within which innovations, on the one hand, will freely develop, and, on the other hand, possible risks will be prevented. At the same time as one of risks most often mark out impossibility to precisely predict a way which development of innovations in the sphere of digital economy therefore the adopted legislation has to be rather flexible will go and be formed taking into account as it is possible the bigger number of data. Initially, scoring was designed to automate the process of deciding on a loan. Prior to the introduction of scoring, the decision on who to extend the loan to was made by a credit expert. He decided, based on experience and his own opinion, guided by the client's parameters affecting his creditworthiness. In the 1940s, the implementation of scoring systems began. In 1941, David Durant published the first credit scoring research to evaluate the role of various factors in the forecasting system. After the end of World War II, the demand for credit products increased sharply, and it became clear that traditional decision-making methods were performing poorly for large numbers of customers. The explosion in demand for loans, driven in part by the implementation of credit cards, motivated lenders to introduce automated systems for deciding on lending. The development of computer technology made it possible to process large amounts of financial data. In 1956, FICO was established to develop consumer loans. In the 60's, the implementation of computer technology in the scoring area began. In 1963, it was proposed to use discriminant data analysis for credit scoring. In 1975 with the adoption of the "US Equal Credit Opportunity Act I", the scoring was finally recognized. An important step in the development of credit scoring was the emergence of behavior scoring in the early 90's. Its purpose is to predict payments to existing customers.

Recently, the development of scoring systems has been driven by external regulation. As part of the capital adequacy requirements for banks following the entry into force of the second Basel Committee for Banking Supervision 2001, institutions should closely monitor the risks associated with their loan portfolios. Credit scoring methods allow to do this. In Christina Bolton's dissertation "Logistic regressions and their application in credit scoring", (2009) it’s considered the concept of credit scoring for banking in South Africa. It’s stressed the methods of constructing a scoring model with emphasis on the logistic regression method [22]. Thesis of Matthias Kremlin "Adaptive models and their application in credit scoring", (2011) emphases methods of constructing predictive models in the conditions of drift and data retention. A new method for building scoring models based on the decision tree method is presented. It’s used to estimate drift in two sets of real financial data [24].

3. PROBLEM STATEMENT AND HYPOTHESES OF RESEARCH

3.1. Research Objectives

The purpose of the research is theoretical justification and development of cluster models working for creating customer classification and making managerial decision. It might direct to save clients and found new target auditory.

According to the research purpose, the following research tasks have been formulated:

- to analyze different approaches to defining the digital economics and digital environment;
- to demonstrate features of the legal regulation in the digital economics;
- to consider scoring models in different spheres of economics and law;
- to define their types and advantages using in different fields of economics and law;

- to adopt k-nearest neighbors support vector method for implementing the principle of customer classification and revealing the client categories with risk of leaving the company;
- to discuss different ways for found out risk groups of clients;
- to show the work of mathematical models on the example of the company.
- to propose some recommendation how to save existing clients and how to share client base through the finding the client groups portraits.

3.2. Purpose of the Study

This paper is aimed to find out the relevant techniques for building scoring models in different fields of economic and law. It was discussed some basic scoring models, their types and advantages using in different fields of economics. For practical realization it was used k-nearest neighbors support vector method in order to implement the principle of customer classification and to reveal the client categories with risk of leaving the company.

3.3. The Research Hypothesis

The main hypothesis of research was checking the quantity of scoring models in different fields. Hypothetically, among the variety of scoring models there are the set of more effective for the decision makink in different domains. The main attention was paid for the finding risk groups of clients. The question under study is: “What scoring models categories are the most effective for decision making working with risk clients according comparative analysis”. Thus, thanks to results of the working model it was found the set of the most relevant models and it was proposed some recommendation how to save existing clients and how to share client base through the finding the client groups portraits. Scoring is a whole customer distribution system based on statistics. It is an important assistant in determining the potential solvency or the future activity of the client as well as the reliable helper of prompt assessment, which is widely used in the economical and legal sector today. The main goal of traditional scoring is to classify bank customers into two categories - “good” and “bad”, based on which the lender can choose the appropriate action in relation to this client. A “bad” client, for example, can be defined as a client with a low empirical probability of loan repayment. To decide, the financial structure of providing a borrower with a loan or material assets uses a system for calculating points. Data processing for decision making is assigned to algorithms using scoring. Test tasks are being developed, as it were, to sort out risk zones and automatically calculate the borrower's potential solvency. If decision-making algorithms are in the risk zone, then it is possible for the client to offer a smaller amount or other conditions. The decision will be made based on many factors. The introduction of artificial intelligence, which laid down the conditions of economic processes and developed a mechanism for calculating the criteria, allows you to train the assessment system.

Scoring is a complex mathematical algorithm that can draw conclusions based on processed data, analyze social factors on an existing client base in a few years. For example, a scoring program can process data on defaulters or debtors over the past some years and identify typical social, age, or behavioral factors. Based on these data, the evaluation will be adjusted and, when analyzing the next clients, the program will consider these new factors. Obviously, in banking databases, you can use algorithms that will look for similar characteristics in new loan applications relative to past similar contacts. It should be noted that scoring is not an ideal financial risk analysis program but helps to quickly and accurately make management decisions when working with big data.

4. METHODOLOGY

According the scoring tasks, all scoring models are divided into three categories, demonstrated on the Figure 1.

Application scoring – scoring credit rating of an individual. If, after entering all the answers in the program, the loan officer replies that the scoring has been completed, this means that the bulk of the analytical verification has been completed. Next, the application of the individual goes to the security service, where bank experts check the client according to several criteria. Conducting a scoring assessment can eliminate the human factor – both specialist's bias towards a client and overly loyal attitude as well as intentional concealment of some factors that indicate an increased financial risk for the bank [28]. The financial scoring algorithm is quite complicated and considers many factors when setting a general assessment of financial risks. Each bank has its own algorithm for verifying customer solvency and discipline regarding loan repayments.

Credit scoring is an automatic scoring system for a borrower. Each client of the bank passes a questionnaire – leaves detailed information about himself. Any of its characteristics has own value in points. After checking the reliability of these data and summing up the scores, a decision is made on the solvency of the potential borrower and, based on this value, on the issuance or non-granting of a loan. The value of the "passing" score depends on the loan product.

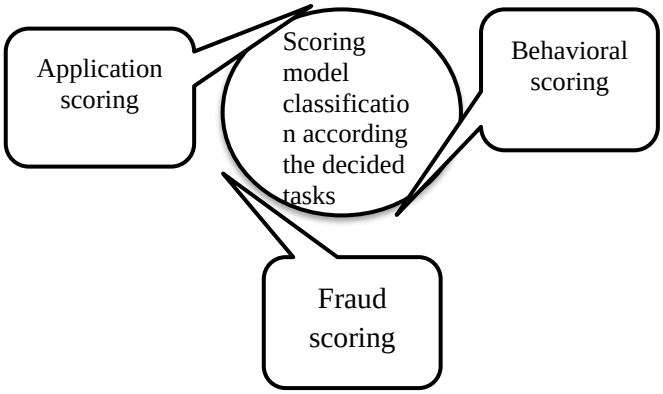

Figure 1 Scoring model classification according the decided tasks

Source: Compiled by the authors

Scoring cards consist of hundreds of positions that are constantly updated and changed. They are created based on processing large amounts of data on credit precedents: repaid and outstanding loans. For example, statistics show that women are more disciplined in financial matters and therefore have a higher scoring score. The factors of a person's residence in the area, as well as his employment in an industry, have their own values. Their value depends on the current economic depression of the region and the growth or decline of production. Significantly lower the final score of a conviction record, administrative offenses, non-payment of fines or alimony. In addition to points, there are so-called stop- and go-factors - circumstances that clearly block the consideration of the borrower's application or, on the contrary, immediately give it a "green light". For example, the first may be the age of the applicant (too young or too old), the second – work in a prestigious international company or in a company that has been served at the bank for many years.

Fraud scoring. This type of scoring is a complex system for detecting any inconsistencies or matches that are also detected through cross-checks. Its goal is to identify anything that might arouse suspicion. When the loan application arrives, the client's personal data is first checked

for authenticity. They are checked against various databases that banks purchase from law enforcement agencies, credit bureaus, collect themselves or find them in some other way. Bankers can also use data that is publicly available. For example - a database of invalid, stolen or lost identity documents. If the authentication is passed, the rules of cross-checks for identifying suspicious situations begin to work. A lot of factors are analyzed and compared: phone numbers, customer addresses, addresses of bank branches, names of tank managers who arrange loans, age of borrowers and others. For example, the system will respond if a new loan application contains a business phone number, which in several previous ones was indicated as home, because it considers this to be suspicious. It will report that the applicant is registered at the same address as the person listed by the bank in the "blacklist". Scoring considers that one of the family members with a not good past inclines relatives to fraud. The system found something suspicious. Two scenarios are possible further. The first is automatic failure. It is issued if there are clear signs of fraud. For example, the application contains a passport, which is listed in the list of stolen items, or a contact phone number, which is on the bank's blacklist. The second – the application is submitted to the risk managers of the bank for "manual" verification. It's happens if a circumstance is found without obvious criminal signs, but it requires explanation. For example, two loan applications have the same address of residence and home phone number. Perhaps these are people who live in a civil marriage, or maybe it's a dummy phone and address. In this case, the verifier calls one of the clients and finds out whether these people know one of one, clarifies some parameters in order to understand whether a person is lying or not. Risk management constantly monitors changes in the quality of the bank's loan product portfolios and develops new audit rules. Each bank maintains its blacklist of customers, which is constantly updated. Security services cooperate with each other and with law enforcement agencies. Such scoring helps identify fraudsters by a variety of signs that they often don't even know about. However, it cannot foresee all situations.

Another classification of scoring models bases on the mathematical methods for building of these models. Among the statistical methods are popular discriminant analysis, linear regression, logistic regression and decision tree. Other methods came from mathematics: mathematical programming, neural networks, genetic algorithms and expert systems. Let's analyze the most common methods represented on the Figure 2.

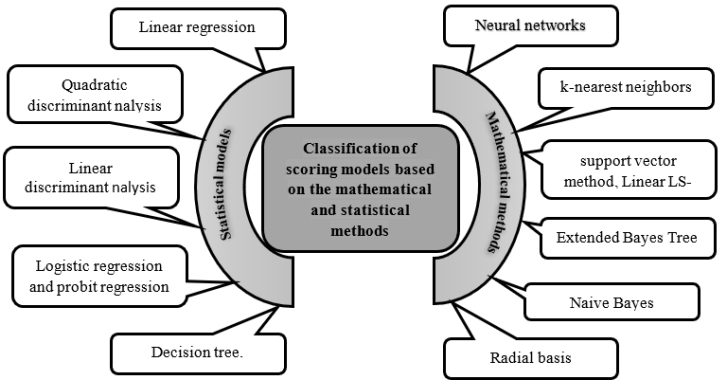

Figure 2 Classification of scoring models based on the mathematical and statistical methods

Source: Compiled by the authors

Linear discriminant analysis (LDA). Linear discriminant analysis is a method for classifying objects into predefined categories. Its main idea is to find a linear combination of explanatory variables that would best categorize objects. By separation, it is best understood as one that

ensures the maximum distance between the average of these categories. The score is calculated as a linear function of the client's attributes values:

$$Z = \beta^T x = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k, \quad (1)$$

where $x=(x_1,...,x_k)$ – customer attribute values, $\beta=(\beta_1,..,\beta_k)$ – model parameters that maximize the ratio,

$$M = \frac{\beta^T (m_G - m_B)}{\sqrt{\beta^T \Sigma \beta}} \quad (2)$$

m_G, m_B is the vector of means for good and bad customers, Σ is the general covariance matrix.

The linear discriminant method involves the fulfillment of two conditions:

- the covariance matrices of independent variables for both groups must coincide.
- independent variables should be distributed normally.
- The main advantage of this method is the possibility to use it even in case of normality violation.

Quadratic discriminant analysis (QDA) is a nonlinear generalization of the LDA. It's a method that does not use the assumption of homogeneity of the covariance matrix. As a decision rule, a quadratic function (3) is applied:

$$d_k = -0.5(x-\mu_k)^T C_k^{-1}(x-\mu_k) - 0.5\ln|C_k| + \ln \pi_k \quad (3)$$

where |Ck| is the determinant of the covariance matrix of the k-th class, C_k^{-1} its inverse matrix; π_k is the priori probability of observing objects of the k-th class. The test object also belongs to the class with the maximum value d_k.

A quadratic discriminant analysis is very effective when the dividing surface between the classes has a pronounced nonlinear character (for example, a paraboloid or an ellipsoid in the 3D case). However, it retains most of the shortcomings of the LDA: it uses the assumption that the distribution is normal and does not work when covariance matrices are degenerate (for example, with many variables). Another disadvantage of QDA is disability to "explain" the results because of the equation of the separating hypersurface is expressed implicitly.

In marketing, discriminant analysis is often used to identify factors that differentiate between different types of customers and/or products based on surveys or other forms of data collection. The use of discriminant analysis in marketing is usually described by the following steps:

- Formulating a task and collection data. Identification the most significant features by which buyers evaluate a product in this category. Use quantitative marketing research methods (such as surveys) to collect data from a slice of potential buyers regarding their preferences for the characteristics of the goods. Data for various products is encoded and entered a statistical system, such as R, SPSS, or SAS.
- Estimate the coefficients of the discriminant function and determine the statistical significance and consistency. Choose the appropriate discriminant analysis method. Direct methods evaluate discriminant function with simultaneously evaluation of all attributes. A step-by-step method introduces the features sequentially. Two-class methods should be used when the dependent variable has only two states. The multiple discriminant method is used when the dependent quantity has three or more qualitative states. SPSS uses Wilk's Lambda or F-stat in SAS to test significance. The most

common method of evaluating solvency is to divide the available data into estimates and verification or deferred data. Evaluation data is used to construct the discriminant function. The deferred data is used to construct a classification matrix, which indicates the number of correctly and incorrectly classified objects.

- Drawing a two-dimensional picture, determination the dimensions, interpretation the results. A statistical program marks the results. In two-dimensional space, each object is displayed. The distance between products characterizes the degree of difference between them. Dimensions must be determined by the researcher. This requires subjective judgment and is often a daunting task.

Linear regression. A linear regression method is the simplest scoring method. In the case of two categories, it is equivalent to the linear discriminant analysis method and expresses the dependence of one variable (dependent) on the other (independent). In general, it's represented by formula (4):

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon \quad (4)$$

where Y – dependent variable; X_i – explaining independent variables; β_i – unknown regression coefficients that are found by the least squares method; ε – error.

It requires the following assumption: the relationship between the dependent and independent variables must be linear; errors should be independent and distributed normally.

Logistic regression and probit regression. The logit regression model is binary model. It allows to model and to forecast simple categorical data. The logistic regression model is defined as follows:

$$\log(\frac{p}{1-p}) = \beta^T x = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \quad (5)$$

where p is the estimate of the probability that the client is "bad", β is the vector of unknown regression parameters, which is calculated as maximizing the likelihood ratio.

The logistic regression model is based on the logarithm function. In turn, probit regression is based on a normal distribution and is defined as follows:

$$N^{-1}(p) = \beta^T x = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \quad (6)$$

where $N(x) = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$; the vector β is calculated like the logistic regression model.

Since logistic regression and probit regression use similar distribution shapes, the results of using these models are also similar. Logistic regression is highly preferred, since calculations are simpler than in probit regression and there are more tools to work with it. Due to its binary nature, logistic regression is preferable to linear regression in use for building scoring models. In practice, it was found that the difference in the accuracy of the predicted results is insignificant. However, there is a predominance of logistic regression in scoring systems.

Neural networks. Artificial neural networks are simulations of neural networks found in nature. Neural networks consist of layers which, in turn, consist of nodes. There are 3 types of layers in networks: input, hidden, output. Customer attributes, such as gender, age, etc., form the input layer. The output y_k for the k-th node with m inputs is represented as follows:

$$y_k = \varphi(v_k) = \varphi\left(\sum_{j=0}^{m} \omega_j x_j\right) = \varphi(\omega^T x) \quad (7)$$

where $\varphi(x)$is the activation function, χ is the input data vector, ω is the weight vector which indicates the strength of the connection between nodes.

Despite the possibility of achieving high accuracy of the forecast, it is impossible to understand the reasons why a decision was made. It's the main disadvantage of the use of neural networks for scoring models development.

Method k nearest neighbors. Nonparametric method for classifying objects. Based on a metric that determines the similarity between the data. Initially, training data is divided into classes. Then the evaluated data is entered and the similarity between the entered and training data is determined. Based on the metric, k nearest neighbors are selected. The new element belongs to the class with the most of its neighbors. The number of neighbors k is determined by a compromise between compensation and dispersion. The smaller the class, the less k is chosen. Moreover, it is not necessary that for large k the result will be better. One of the advantages of this method is easy possibility to add new data without changing the model. The nonparametric nature of this method allows to work with irrationalities in risk functions in the attribute space. The absence of a formal method for choosing k and the impossibility of a probabilistic interpretation of the result, are the main disadvantages of the method. These difficulties can be solved using the Bayesian approximation method.

Comparison of various methods. A series of comparative studies have been conducted for scoring methods. The ranking criteria were the percentage of classification errors and the ROC curve. It was studied eight data sets (Table 1).

Table 1 Comparison of various methods

Method	Average rating
Neural networks	3.2
Support Vector Method	3.7
Logistic Regression	4.3
Linear discriminant analysis	5.3
Linear LS-SVM	5.5
Extended Bayes Tree	5.6
Naive Bayes Classifier	7.8
Radial basis functions	9.1
k-nearest neighbors (k = 100)	9.5
Linear SVM	10.1
Quadratic discriminant analysis	10.8
Decision tree	10.8
Linear programming	11.9
Decision tree	13.7
k-nearest neighbors (k = 10)	14.1

Source: Compiled by the authors

The Table 1 shows that Neural Networks, Support Vector Method and Logistic Regression were the best in the studied eight data sets. There is no optimal scoring model for any situation. The choice of model depends on the data and the purpose for which the creation of the model is directed. In addition, the best rating method will not necessarily be the best in this situation.

5. STATISTICAL ANALYSIS

5.1. Reliability and Validity

Aiming to achieve high reliability and validity of the research, in the calculations it was applied the set of scoring models as well as a probability statistical model. The first model using the data of real gym company help to build the system of decision making for risk clients. And the second one describing the fraud processes was based on the official open data statistics. It shows the level of different types of frauds and the dynamic of their changes. It helps to account this factor working with unknown clients.

5.1.1. Data Collection

To develop Customers Classification Scoring Model, it was considered sample of the gym clients, which was consist of the next features:

Age – age of client;

Income – average client's month income, thousand $;

Children – the number of children or grandsons in the age less than 15;

Sex – male (1) or female (0);

Education – school (1); Bachelor's degree (2); Master's degree (3); graduate school or second diploma (4);

Visit_Count – the number of visiting of gym during last month;

Is_Client – if person continue to be the client of the gym (1) or he leave this gym (0).

For the fraud statistical analyses, it was used official open data [13, 2020].

5.2. Data Analysis and Results

The mathematical model of Behavioral scoring for customers classification according k-nearest neighbors support vector method is demonstrated on the Figure 3. The model is created at the StatSoft Statistica Enterprise 10.0 with the help of module k-means clustering after normalization of the entering sample.

Figure 3 shows what categories of clients have the biggest probability to leave this gym next month. So, there are two risk groups at this gym. The riskiest group is represented by the Cluster 1 and the next risk group is shown in the Cluster 5.

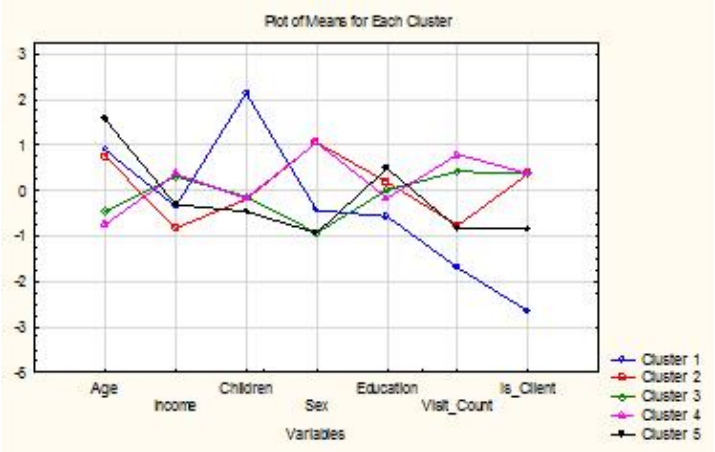

Figure 3 Customer Classification Scoring Model

Source: Compiled by the authors

These cluster consists of clients with the following characteristics (Figure. 4). So, it was found some risk client categories. Among them it was emphasized group which characterized as the group of people with middle income, over 50 years with little grandsons, without high school diploma. The next set of probably risk clients is the group of women with middle income over 60 years with some diplomas and/or degrees without little grandsons or probably only with one grandson.

Finally, it was stressed on the set of elderly gym visitors. It is highlighted on the necessity of the realization marketing company directed on the elderly people. It was pointed out that the considered company need at least two advertising programs: for elder people and their grandsons at the same time or together and for elder clients without little grandsons. The next

analysis was hold in the law according fraud statistics. There are a lot of methods fraud detection. Among them are: identity validation potential risks associated with the borrower's individual characteristics; phone and address check to validate borrower's info; Income and Employment Analysis; variety of automatic systems like MERS (Mortgage Electronic Registration System) or NFPB (National Fraud Protection Database) and others. The aim of this research was to understand, who and where taking place frauds and to find new and emerging trends of them.

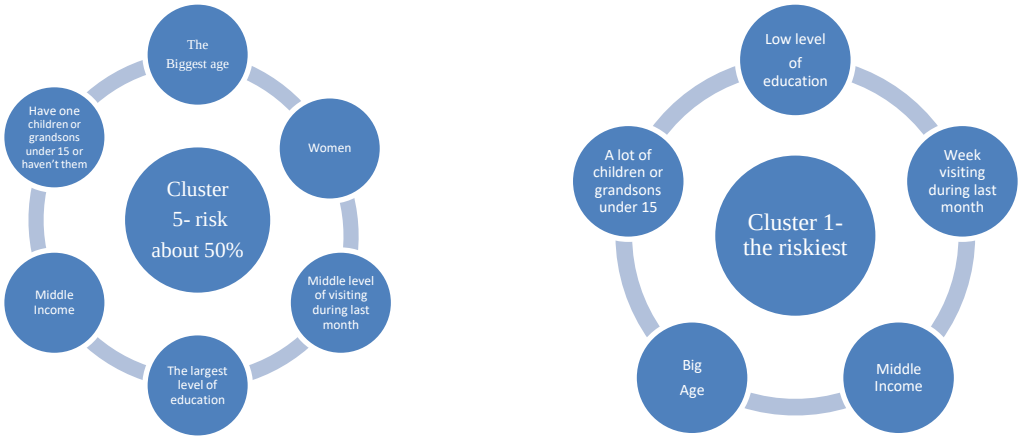

Figure 4 Risk clusters client's characteristics

Source: Compiled by the authors

Let's consider the official statistics of fraud by financial products, demonstrated on the Figure 5.

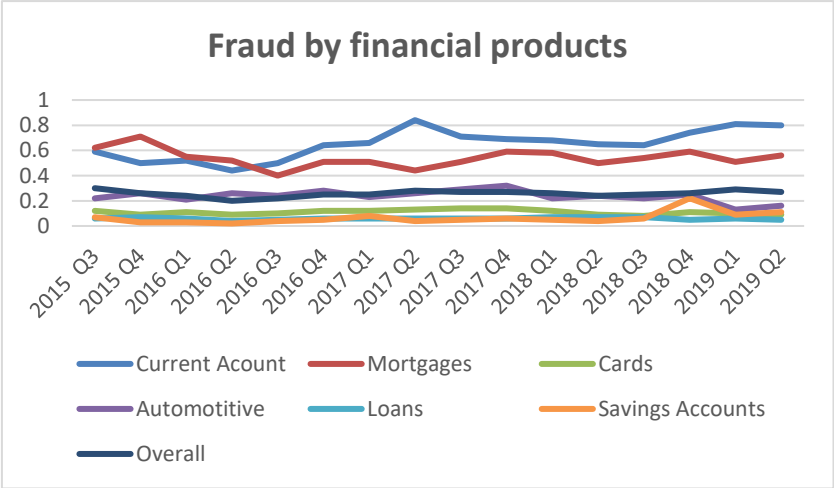

Figure 5 Fraud by financial products

Source: Compiled by the authors

Given graphs demonstrate that the level of fraud is not very high. Different kinds of fraud have different variation, but all of them are less then 1%. The riskiest activity connects with Current Accounts. Moreover, this kind of frauds have the highest grow dynamic. The less risky operations connected with Saving Accounts. However, it also shows rise dynamic during last year.

Figure 6 shows Means and Standard Deviations different kinds of frauds. Mean value demonstrate two riskiest operations: Current Account and Mortgages and operations almost without risk: Loans and Saving Account. Standard Deviation show the same trend – the

biggest variation for the first two operations and the littlest – for the last two kinds of fraud. Moreover, the deviation for all operations except Current Account and Mortgages are less than error of calculation.

Table 2 Means and Standard Deviations different kinds of fraud

Variable	**Means and Standard Deviations Casewise deletion of MD N=16**	
Current Acount	0,650625	0,117840
Mortgages	0,540000	0,071926
Cards	0,109375	0,018786
Automotitive	0,236875	0,046435
Loans	0,059375	0,008539
Savings Accounts	0,065000	0,047610

Source: Compiled by the authors

In order to define dependence between all given kinds of fraud it was realized correlation analysis. Correlation matrix shown on the Table 3 consists of the Pearson correlation coefficients and demonstrates the level of dependence between different frauds.

Table 3 Correlation matrix for different kinds of fraud

Variable	**Current Acount**	**Mortgages**	**Cards**	**Automotitive**	**Loans**	**Savings Accounts**	**Overall**
Current Acount	1,00	-0,17	0,33	-0,26	0,13	**0,49**	**0,68**
Mortgages	-0,17	1,00	-0,08	0,04	0,30	0,18	0,32
Cards	0,33	-0,08	1,00	**0,51**	0,04	0,01	0,42
Automotitive	-0,26	0,04	**0,51**	1,00	0,01	-0,22	-0,23
Loans	0,13	0,30	0,04	0,01	1,00	-0,27	0,37
Savings Accounts	**0,49**	0,18	0,01	-0,22	-0,27	1,00	0,32
Overall	**0,68**	0,32	0,42	-0,23	0,37	0,32	1,00

Source: Compiled by the authors

Correlation matrix helps to understand relationships between different types of fraud. Thus, it highlights that frauds on Current Accounts have the most influence on the Overall frauds with the correlation level r=0,68. And the Saving Accounts fraud influence on the Current Account fraud with the r=0,49. This fact explain the trend of the last year to raising both types of financial fraud. So, despite of the minor level of Saving Account fraud, considering its raising trend, there is a growth risk of the biggest level of fraud – Current Account.

In order to detail the character of fraud's changing it's interesting to research Frequency Distribution for each kind of fraud. They are demonstrated on figures 6-11. To identify their features, it was conduct comparing analysis of their graphs. According given graphs all of them have normal distribution, but heir own parameters and properties.

Figure 8 demonstrates Frequency Distribution for Current Account.

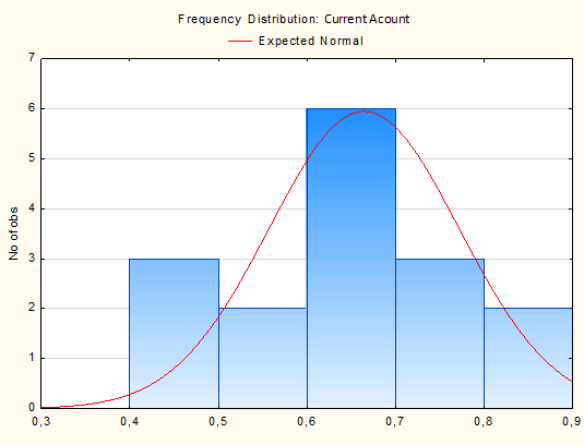

Figure 6 Frequency Distribution for Current Account

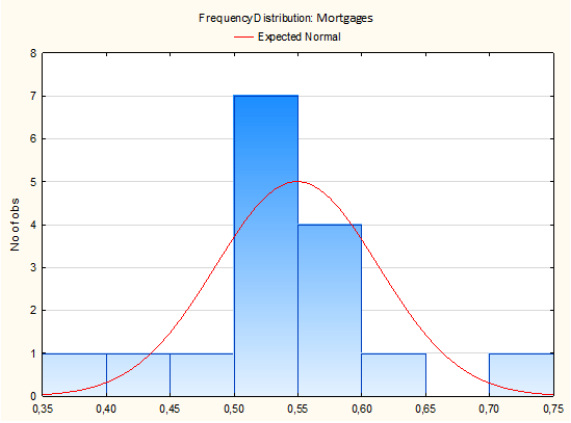

Figure 7 Frequency Distribution for Mortgages

Source: Compiled by the authors

Usually normal curves are symmetric about the mean μ. But practically it has some variations in its parameters. So, the Frequency Distribution of Current Account is left-skewed or negatively-skewed distribution because it has a little longer left tail in the negative direction on the number line. The mean in this direction is also to the left of the peak. Moreover, it has the mean to the left of the median. Thus, frauds in Current Account are extremely important to research because they have the biggest value, and more than half considered years have the percent more than mean. Moreover, it has the growth trend. To conclude, it's the riskiest kind of financial operation considering frauds (Figure 7).

Frequency Distribution for Mortgages is almost symmetric Normal Distribution. But it has a positive value of Kurtosis, which tells that it has heavy-tails or a lot of data in its tails. It means that the mean value isn't characterizes Mortgages fraud for all considering years. It was random high peak value that maybe need additional research. It's more typical to have low value for this distribution. So, Mortgages fraud is not so danger, but sometimes it appears high unknown activity that need to be learnt (Figure 8).

In the case of Cards operation, it's interesting to note Bimodal character of Normal Distribution. It has not one pick. There are two picks in it. And it means that there is no clear value of cards fraud. So, the risk of frauds in these operations is rather high because it's difficult to forecast the level of such frauds at every time. Despite of the stochastic character of this distribution the level of danger in this situation isn't so high because of not critical level of mean according to statistical values (Figure 9).

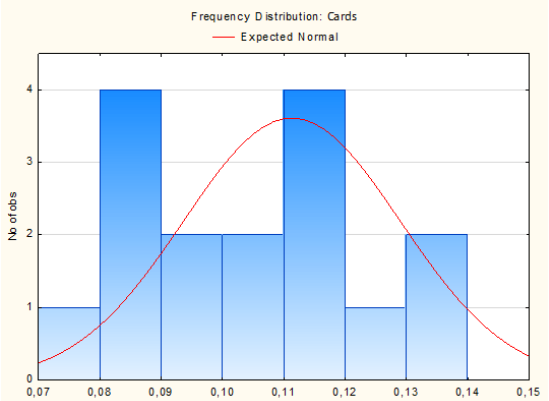

Figure 8 Frequency Distribution for Cards

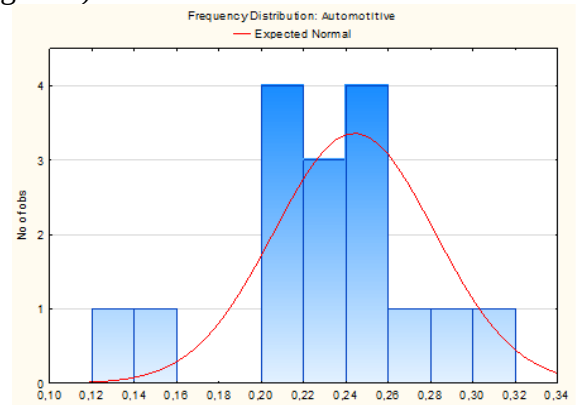

Figure 9 Frequency Distribution for Automotitive

Source: Compiled by the authors

Distribution of Automotitive frauds is partly like to Frequency Distribution for Current Account. But with lower high pick. So, it's not so danger as Current Account fraud, but it's recommended to pay attention on it. Moreover, according to the statistical analysis, demonstrated on the figure 6, this type of frauds has the third place among all kinds of all financial frauds (Figure 10).

Saving Accounts has right-skewed distribution with a long right tail. Right-skewed distributions are also called positive-skew distributions. That's because there is a long tail in the positive direction on the number line. The mean is also to the right of the peak. Because this histogram's tail has the biggest positive skew to the right, Saving Accounts fraud have light-tails or little data in their tails. Especially in its right tail. This fact once more improve that this kind of frauds are the minimum risky (Figure 11).

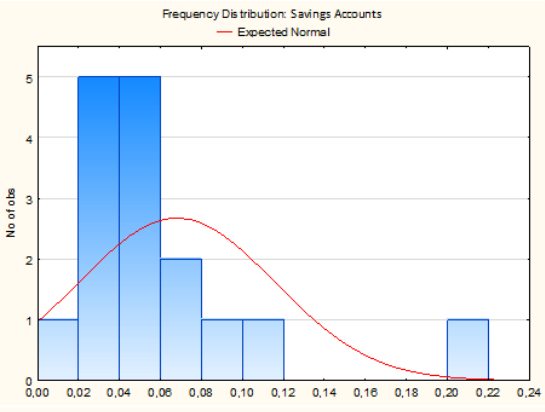

Figure 10 Frequency Distribution for Saving Account

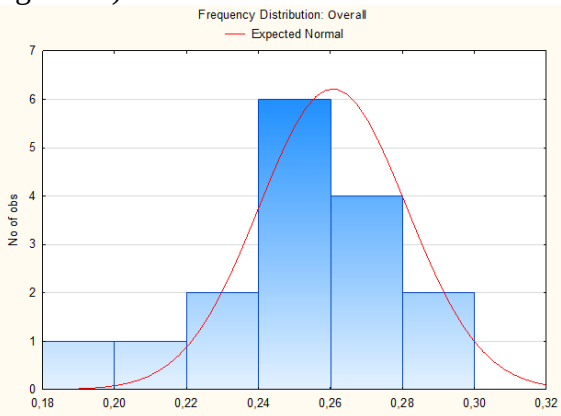

Figure 11 Frequency Distribution for Overall

Source: Compiled by the authors

The overall distribution of financial frauds has similar form to Current Account Distribution. And it's not surprisingly, because the current account according the correlation analysis have rather high influence on the overall frauds.

So, statistical and visual analysis of distributions give the opportunity to find out some features of the kinds of frauds character. It was found the most and the less risky financial operations and made the hypotheses about their further behavior.

6. DISCUSSION

6.1. Key Findings and Results

The results of the study are summarized in the following statements:

- Methods of building scoring models in different fields of economy and law have been analyzed and investigated in order to create classifications of client groups.
- Classifications of client groups for financial management decisions have been developed and are theoretically justified.
- Considered and practically analyzed, evaluation models, which are assistants for managers in determining potential solvency or future activity of the client, as well as for rapid assessment of financial characteristics of clients.
- Scoring models, their types and advantages of use in different areas of economy are considered, as well as methods of reference vectors of k-nearest neighbors are specified for implementation of the principle of classification of clients and identification of clients with risk of leaving the company.

- The risk groups of clients are determined based on the use of scoring models and are practically confirmed by the mathematical model on the example of the company.
- Based on the results of the model, a recommendation was proposed on how to preserve existing clients and how to divide the client base by looking for portraits of client groups.
- Factors of impact of new information products and technologies on the modern economic market are considered. A comparative analysis of the services market using new technologies and rapid interaction with customers is provided.
- Consider risks in economic structures, which depend on timely response to various changes in business relations and lag from the technology market, which leads to the risk of uncompetitivity or displacement from the market.

6.2. Prospects for Further Research

The relevance of the results of the study is confirmed by the fact that the development of information technologies is emerging new tools of business management. Modern Internet technologies are developing very rapidly and give frauds new opportunities for financial fraud. In turn, business and financial structures should step up to the times and look for new methods of early warning of financial risks. Information technology, combined with mathematical statistical techniques, enables decision-making algorithms to be constructed well in advance of bid processing before applying for financial transactions.

Given the availability of complete statistical information, further research should be related to Internet things, socio-economic decision-making technologies. It is also planned to develop computer models, risk and consumption loss algorithms, investment analytical technologies.

7. CONCLUSION

The better scoring system is developed, the more objective it is, the more correctly and quickly will evaluate the risks of the bank, saving it from possible losses. That is why the scoring model is kept in strict confidence and is developed individually by each enterprise. To mislead it, you need to know how to answer specific questions in the questionnaire. And this is the main reason why enterprises almost never report the reasons for the refusal to customers. Scoring has its own strengths and weaknesses. It helps to identify potential defaulters and fraudsters, while not eliminating the risks of issuing a loan to an unreliable client or refusing to conscientious. In this paper, the most used methods for constructing scoring models were described. Currently, scoring is widespread in the world and has proven to be an effective decision-making tool in the digital economy. In many areas of the economy, they have already abandoned the use of expert assessment in favor of scoring systems. However, despite the wide distribution, scoring is poorly covered in Ukrainian literature, although many foreign works have been devoted to it. Scoring has great potential for use but is still a “black box” for people using it. Scoring systems should continue to be studied and improved.

REFERENCES

[1] Dahlman, C., Mealy S., Wermelinger M. (2016). *Harnessing the Digital Economy for Developing Countries*. Paris: OECD. Available at: http://www.oecd-ilibrary.org/docserver/download/4adffb24-en.pdf

[2] Haltiwanger J., Jarmin R.S. (2000). Measuring the Digital Economy. Understanding the Digital Economy (E. Brynjolfsson, B. Kahin (eds)). Cambridge: MIT Press, MA, pp. 13–33.

[3] Bucht, R., Hicks, R. (2018). Definition, concept and measurement of the digital economy. *Bulletin of international organizations*. Vol.13, pp.143-172.

[4] Kozlovskyi, S., Nikolenko, L., Peresada, O., Pokhyliuk, O., Yatchuk, O., Bolgarova, N., Kulhanik, O. (2020). Estimation level of public welfare on the basis of methods of intellectual analysis. *Global Journal of Environmental Science and Management*, Vol. 6(3), pp. 355-372. http://dx.doi.org/10.22034/gjesm.2020.03.06

[5] Bucht R., Hicks R. Definition (2019). *Concept and measurement digital economy*. Available at: https://iorj.hse.ru/data/2018/08/30/1154589879/Бухт Хикс Определение Концепция и измерение цифровой экономики.pdf

[6] Cherdantsev, V., Kobelev, P. (2010). Formation of a single information space. *Agrarian Gazette of the Ural*. Vol. 11-1 (77). pp. 102-103.

[7] Myshko, F., Vasileva, K., Popov, V., Strelnikov, I. (2019). Trends in legal regulation of civil and competitive relations in the sphere of digital economy. *Journal of Economic Security*, Vol. 1. pp. 155-159.

[8] Kozlovskyi, S., Grynyuk, R., Baidala, V., Burdiak, V., Bakun, Y. (2019). Economic Security Management of Ukraine in Conditions of European Integration. *Montenegrin Journal of Economics*, Vol. 15(3), pp. 137-153. http://dx.doi.org/10.14254/1800-5845/2019.15-3.10

[9] Belikova, K. (2018). *Features of the legal regulation of the digital intellectual economy*. Available at: https://cyberleninka.ru/article/n/osobennosti-pravovogo-regulirovaniya-tsifrovoy-intellektualnoy-ekonomiki. DOI 10.24411/2073-3313-2018-10090

[10] Kozlovskyi, S., Baidala, V., Tkachuk, O., Kozyrska, T. (2018). Managament of the sustainable development of the agrarian sector of the regions of Ukraine. *Montenegrin Journal of Economics*, Vol. 14. № 4, pp. 175-190. http://dx.doi.org/10.14254/1800-5845/2018.14-4.12

[11] Kozlovskyi, S., Bilenko, D., Kuzheliev, M., Lavrov, R., Kozlovskyi, V., Mazur, H., Taranych, A. (2020), «The system dynamic model of the labor migrant policy in economic growth affected by COVID-19», Global Journal of Environmental Science and Management, Vol. 6 (Special Issue (Covid-19)), pp. 95-106. http://dx.doi.org/10.22034/GJESM.2019.06.SI.09

[12] Bakhmat, N., Maksymchuk, B., Voloshyna, O., Kuzmenko, V., Matviichuk, T., Kovalchuk, A. Maksymchuk, I. (2019). Designing cloud-oriented university environment in teacher training of future physical education teachers. *Journal of Physical Education and Sport*, 19 (4), 1323-1332.

[13] Sitovskyi A., Maksymchuk B., Kuzmenko V., Nosko Y., Korytko Z., Bahinska O.Maksymchuk, I. (2019). Differentiated approach to physical education of adolescents with different speed of biological development (2019). *Journal of Physical Education and Sport*, Vol.19 (3), Art 222, pp. 1532-1543.

[14] Churchill G. A., Nevin J. R., Watson R. R. (1977). *The role of credit scoring in the loan decision*. Credit World. March.

[15] Myers J. H., Forgy E. W. (1963). The development of numerical credit evaluation systems. *Journal of American Statistical Association*, September, 15000 p.

[16] Okwonu, F. (2012) A Model Classification Technique for Linear Discriminant Analysis for Two Groups. *International Journal of Computer Science Issues*, Vol. 9, Issue 3(2), pp.125-128.

[17] Koziuk V., Dluhopolskyi O., Petruk, V. (2019). Globalization, innovation and fragility of optimal fiscal zones: secessions risks of Belgium, lessons for Ukraine. *Ideology and Politics Journal*, №1(12), 60-90.

[18] Kozlovskyi, S., Shaulska, L., Butyrskyi, A., Burkina, N., Popovskyi, Y. (2018). The marketing strategy for making optimal managerial decisions by means of smart analytics. *Innovative Marketing*, Vol. 14, No. 4, pp. 1-18. https://doi.org/10.21511/im.14(4).2018.01

[19] Kuan, C., White, H. (1994). Artificial neural networks: an econometric perspective. *Econometric Reviews*, Vol. 13, pp. 1-91.

[20] Lavrov, R., Beschastnyi, V., Nikolenko, L., Yousuf, A., Kozlovskyi, S., Sadchykova, I. (2019), Special aspects of the banking institutions rating: a case for Ukraine, Banks and Bank Systems, Vol. 14 № 1, pp. 48-63. http://dx.doi.org/10.21511/bbs.14(3).2019.05

[21] Koziuk, V., Dluhopolskyi, O., Hayda, Y., Shymanska, O. (2018). Typology of welfare states: quality criteria for governance and ecology. Problems and Perspectives in Management, Vol. 16 (4), pp. 235-245. https://doi.org/10.21511/ppm.16(4).2018.20

[22] Yousuf, A., Haddad, H., Pakurar, M., Kozlovskyi, S., Mohylova, A., Shlapak, O., & Janos, F. (2019). The effect of operational flexibility on performance: a field study on small and medium-sized industrial companies in Jordan. *Montenegrin Journal of Economics*, Vol. 15(1), pp. 47-60. https://ideas.repec.org/a/mje/mjejnl/v15y2019i1p47-60.html

[23] Bogachev, V., Kolesnikov, A. (2012). The task of Monja-Kantorovich: achievements, connections and prospects. *Success of mathematical sciences*, Vol. 5(407), pp. 3-110.

[24] Mudrecova, E. (2013*) Diploma work on the topic: Evaluation of customer ratings based on the mathematical model of scoring*. MIEM, Moscow.

[25] Bucht R., Hicks R. (2018). Definition, concept and measurement of the digital economy. *Bulletin of international organizations*, Vol. 13, pp. 143-172.

[26] European Parliament. *Challenges for Competition Policy in a Digitalised Economy*. Brussels: Euro- pean Parliament. Available at: http://www.europarl.europa.eu/RegData/etudes/STUD/2015/542235/IPOL_ STU(2015)542235_EN.pdf/.

[27] Salchenberger L., Cinar E. & Lash N. (1992). Neural networks: a new tool for predicting thrift failures. *Decision Sciences*, Vol. 23, pp. 899-916.

[28] Marhasova, V., Kovalenko, Y., Bereslavska, O., Muravskyi, O., Fedyshyn, M., Kolesnik, O. (2020). Instruments of monetary-and-credit policy in terms of economic instability. *International Journal of Management*, Vol. 11(5), pp. 43-53.

[29] K P Tripathi, (2011) Decision Making as a Component of Problem Solving, *International Journal of Information Technology & Management Information System*, 1(1), pp. 55-59

[30] Jandel S Yadav, Anshul Gangele and Dharam Buddhi, (2018) Evaluation of Product Quality In QFD Using Multi Attribute Decision Making (MADM) Techniques In Manufacturing Industry, *International Journal of Production Technology and Management*, 9(2), pp. 74–86.

[31] Dr. Giriraj Kiradoo, (2011) Evaluating the Significance of Management Information Systems on Strategic Decision Making for Gaining Competitive Advantage, *International Journal of Information Technology & Management Information System*, 2(1), pp. 5–12.

[32] Indranil Ghosh, (2014) An Intelligent Hybrid Multi Criteria Decision Making Technique to Solve a Plant Layout Problem, *International Journal of Industrial Engineering Research and Development*, 5(3), pp. 13–23.