**System Analysis of Financial Monitoring Subjects’ Activities
for the Country’s Economic Security Ensuring**

 **Prikazchikova A. S.1, Prikazchikova G. S.2**

*1NRNU "MEPhI" post-graduate student, leading expert of the Rosfinmonitoring Information*

*Systems Development Department, Moscow Russia*

*2Associate Professor of the Russian Customs Academy, Moscow, Russia*

Prikazchikova A. S.: ORCiD: 0000-0003-0975-5939

Prikazchikova G. S.: ORCiD: 0000-0002-9679-6423

Corresponding Author: Prikazchikova A. S., E-mail: aska4.92@mail.ru

**Abstract:** the article considers the binary classification problem of economic security objects on the credit institutions example, for which it is proposed to use machine learning methods. In the study process the expediency of one of the methods of machine learning — the method of k-nearest neighbors — was proved to solve this problem, its efficiency amounted
to 84 %.

**Key words:** machine learning methods, financial statements, performance indicators, credit institutions, binary classification, k-nearest neighbors method.

**Introduction**

The most important condition for the country’s national security ensuring is the stability of its economic system. One of the elements of such system is credit institutions whose reputational risks are to be assessed.

To date, the task of automated (on-line) monitoring of the credit institutions activities, qualitative and professional assessment of banks financial condition is especially topical. As a result of new regulatory and legal acts in the credit and financial sphere coming into legal force, as well as amendments to the current legislation, the number of credit institutions has significantly decreased by 2017. However, currently their number exceeds 500, as a result there is a growing need for scientific and practical development in the evaluation of economic and information security subjects capable of automatically monitoring banks activities and predicting the risk of license revocation.

A comprehensive analysis of credit institutions activities is a very urgent task of financial monitoring, the solution of which is based on the use of large (several terabytes) information arrays of heterogeneous data. These arrays are fed into the information system (hereinafter - IS) of financial monitoring. In order to assess the situation, which is the initial and the most important stage of decision-making, the processing of incoming data is carried out in the IS, which allows to present the subjects with the vectors of indicators. The essence of the assessment of the situation is the convolution of the indicators into a single (several) rating assessment, which characterizes the measure of the deviant activity of the object. The approach to assessing the situation existing in the financial monitoring is that the experts assign the weighting factors to the components of the vector and add the results obtained. At the same time, in accordance with the established practice, experts prepare information on the situation in a natural language [1].

However, the results thus obtained are accompanied by expert subjectivism and socio-political motivation of experts. In addition, the increasing volume of incoming data (approximately 20% annually) leads to a decrease in the speed of the data processing at the stage of assessing the situation. Rosfinmonitoring’s management is forced to work with subjective assessments and face a significantly increased deadlines for obtaining the result.

Under these conditions, the contradiction between Rosfinmonitoring management's need and the existing practice of processing data, accompanied by high time and resource costs, and expert subjectivism, is growing. This contradiction is due to the insufficient level of systemacity of the existing approach and can be eliminated by developing a comprehensive methodology for the process of assessing the situation in the financial monitoring IS [1].

**Material and Theoretical Bases of Research**

The analysis of approaches to the task of assessing credit institutions activity by their vector indicators in Financial Reporting Form No. 101, which contains more than 100 fields, showed that one of the machine learning methods, the method of k-nearest neighbors is more promising.

The method of k-nearest neighbors is a method of solving a classification problem that relates objects to a class that owns most of its k nearest neighbors in a multidimensional feature space.

In the learning process, the algorithm remembers all feature vectors and their corresponding class labels. When working with real data, i.e. observations, whose class labels are unknown, the distance between the new observation vector and those previously stored is calculated. Then k-nearest vectors are selected, and the new object belongs to the class that owns most of them [2].

We formalized the statement of the k-nearest neighbors method problem:

Let be the set of objects specified as ;

 – set of admissible answers.

A training sample is specified .

It is required to find a lot of answers  for objects .

Algorithm  of weighted nearest neighbors: on the set of objects the Euclidean distance function 



                                   (1)

For an arbitrary object , we arrange the objects of the training sample  in the order of increasing distances to :



where through  denotes that object of the training sample, which is the  -th neighbor of the object . We introduce a similar notation for the answer to the -th neighbor: .

Thus, an arbitrary object  generates its renumbering of the sample.

In the most general form, the k-nearest neighbor algorithm can be represented as follows (2):



 (2)

where  is a given weight function that estimates the degree of the -neighbor importance to classify the object . It is natural to assume that this function is nonnegative and does not increase by  [3].

After bringing the necessary information to the appropriate type, five main factors were identified by the principal components method of factor analysis on the basis of the financial reporting No. 101 indicators, which characterize the credit institutions activity: F1, F2, F3, F4, F5, the contribution to the total variance of which amounted to more than 90%. As a result, it became possible to carry out further research on selected components.

The example of the working table for the analysis is shown in Fig. 1.



Figure 1: Initial data for analysis

One of the stages of multidimensional objects classification using the method of k-nearest neighbors is the determination of the correct number of k-nearest neighbors of the analyzed object.

In the case of optimizing the number of k-nearest neighbors on condition the algorithm of the k-nearest neighbor is unstable to noise emissions.

On the contrary, the algorithm is excessively stable and degenerates into a constant. Thus, extreme values are undesirable. In practice, the optimal value of parameter is determined by the sliding control criterion, most often by the leave-one-out-cross-validation method.

**Results**

In the STATISTICA software package (STATISTICA), the optimal number of k-nearest neighbors was analyzed, which showed that this value (K Optimal) should be equal to 6 (Fig. 2).



Figure 2: Optimal number of k-nearest neighbors in the binary classification

In the course of further research in STATISTICA, the following results were obtained.

So, the total number of objects involved in the analysis amounted to 242 units, of which 181 units are training samples, 61 units are tested objects. Based on the fact that we had 2 classes at our disposal: a class of trustworthy credit organizations and a class of organizations with a high financial stability risk, we obtained the diagram of the classification results of the k-nearest neighbors shown in Fig. 3.

The correctness of the conducted binary classification was 84%, which confirms the expediency of using the method under investigation when analyzing credit institutions’s characteristic space in practice.



Figure 3: Results of binary classification by k-nearest neighbors method

**Conclusion**

An expert application of k-nearest neighbors method for the financial and credit sector objects analysis will allow to assess a bank reliability with a high degree of probability, to predict the risk of license revokation, and to take preventive or other measures in case of establishing the fact of the credit organization's involvement in schemes for the legalization of criminal proceeds and the financing of terrorism.

**Acknowledgements**This work was supported by Competitiveness Growth Program of the Federal Autonomous Educational Institution of Higher Education National Research Nuclear University MEPhI (Moscow Engineering Physics Institute).

**References**

[1] Denisenko, A.S. Information system for processing financial monitoring data in the field of bioresources [Text]: author's abstract. dis. on comp. of the Cand.Tech.Sci scientific degree. (05.13.01) / Denisenko Andrey Sergeevich; NRNU "MEPI". – Moscow, 2017. – 29 p.

[2] Abidin, T. and Perrizo, W. SMART-TV: A Fast and Scalable Nearest Neighbor Based Classifier for Data Mining. Proceedings of ACM SAC-06, Dijon, France, April 23-27, 2006. ACM Press, New York, NY, pp.536-540.

[3] Apostolos Papadopoulos, Yannis Manolopoulos. Performance of Nearest Neighbor Queries in R-Trees Source // Proceedings of the 6th International Conference on Database Theory. 1997. P. 394 – 408.