Knowledge E
Engaging minds

Conference Paper

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

Prikazchikova A. S.[1] and Prikazchikova G. S.[2]

[1]NRNU "MEPhI" post-graduate student, leading expert of the Rosfinmonitoring Information Systems Development Department, Moscow Russia
[2]Associate Professor of the Russian Customs Academy, Moscow, Russia

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 %.

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

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

Received: 11 December 2017
Accepted: 20 January 2018
Published: 13 February 2018

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the FinTech and RegTech: Possibilities, Threats and Risks of Financial Technologies Conference Committee.

1. 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.

How to cite this article: Prikazchikova A. S. and Prikazchikova G. S., (2018), "System Analysis of Financial Monitoring Subjects' Activities for the Country's Economic Security Ensuring" in *III Network AML/CFT Institute International Scientific and Research Conference "FinTech and RegTech: Possibilities, Threats and Risks of Financial Technologies"*, KnE Social Sciences, pages 444–449. DOI 10.18502/kss.v3i2.1575

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].

2. 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 $X \in \mathbb{R}^n$ be the set of objects specified as $X^m = \{x_i\}_{i=1}^m$;

Y – set of admissible answers.

A training sample is specified $\{(x_i, y_i)\}_{i=1}^{\ell}$.

It is required to find a lot of answers $\{y_i\}_{i=1}^m$ for objects $\{x_i\}_{i=1}^m$.

Algorithm K of weighted nearest neighbors: on the set of objects the Euclidean distance function $\rho(x, x')$:

$$\rho(x, x') = \sum_{i=1}^{n} (x_i - x_i')^2. \quad (1)$$

For an arbitrary object $x \in \{x_i\}_{i=1}^m$, we arrange the objects of the training sample x_i in the order of increasing distances to x:

$$\rho(x, x_{1;x}) \le \rho(x, x_{2;x}) \le \cdots \le \rho(x, x_{m;x}),$$

where through $x_{i;x}$ denotes that object of the training sample, which is the i-th neighbor of the object x. We introduce a similar notation for the answer to the i-th neighbor: $y_{i;x}$.

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

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

$$a(x) = \arg\max_{y \in Y} \sum_{i=1}^{m} [x_{i;x} = y] w(i, x), \quad (2)$$

where $w(i, u)$ is a given weight function that estimates the degree of the i-neighbor importance to classify the object u. It is natural to assume that this function is nonnegative and does not increase by i [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.

1 Наименование КО	2 Лицензия	3 Отзыв	4 Регион	5 FACTOR1	6 FACTOR2	7 FACTOR3	8 FACTOR4	9 FACTOR5
Викинг	2	0	Санкт-Петербург и обл.	-0,04595	-0,02622	-0,07261	-0,06115	-0,14258
Кредит Москва	5	1	Москва и обл.	-0,03943	-0,02472	-0,06996	-0,06208	-0,14460
Примтеркомбанк	21	0	Приморский край	-0,04745	-0,02937	-0,07280	-0,06126	-0,14402
Александровский	53	0	Санкт-Петербург и обл.	-0,04210	-0,02630	-0,06625	-0,05812	-0,14537
Темпбанк	55	0	Москва и обл.	-0,03132	-0,04281	-0,06538	-0,06034	-0,11299
Кольцо Урала	65	0	Свердловская обл.	-0,02103	-0,01510	-0,06795	-0,06888	-0,14985
Саммит Банк	85	0	Приморский край	-0,04550	-0,03230	-0,07292	-0,05965	-0,14334
Кемсоцинбанк	96	0	Кемеровская обл.	-0,04518	-0,02981	-0,07217	-0,06075	-0,14372
БКС — Инвестиционный Банк	101	0	Новосибирская обл.	-0,07789	-0,05677	0,15797	-0,09631	-0,17604
Расчетно-Кредитный Банк	103	0	Москва и обл.	-0,03356	-0,06842	-0,05557	-0,07636	-0,00482
Сельмашбанк	106	0	Ростовская обл.	-0,04644	-0,03237	-0,07260	-0,06055	-0,14349
Континент Финанс	149	1	Москва и обл.	-0,04677	-0,03147	-0,07250	-0,06051	-0,14295
РН Банк	170	0	Москва и обл.	-0,00385	-0,00918	-0,09316	-0,00829	-0,17592
Международный Банк Санкт-Петербурга (МБСП)	197	0	Санкт-Петербург и обл.	-0,05816	0,02314	0,03316	-0,07526	-0,11814
Экспресс-Кредит	210	1	Москва и обл.	-0,03922	-0,02845	-0,07249	-0,06124	-0,14567
Зенит Сочи	232	0	Краснодарский край	-0,04501	-0,02969	-0,07202	-0,06107	-0,14468
Спецстройбанк	236	0	Москва и обл.	-0,04597	-0,03022	-0,07249	-0,06089	-0,14340
Саратов	330	0	Саратовская обл.	-0,04506	-0,03193	-0,07257	-0,06043	-0,14311
Единая Расчетная Палата	384	0	Москва и обл.	-0,04693	-0,03130	-0,07256	-0,06048	-0,14291
Идея Банк	430	1	Краснодарский край	-0,04102	-0,03306	-0,06523	-0,06298	-0,14490
Бест Эффортс Банк	435	0	Москва и обл.	-0,05168	-0,01362	-0,04705	-0,06842	-0,14691
Камский Коммерческий Банк	438	0	Татарстан	-0,03667	-0,03647	-0,07356	-0,05728	-0,13459
Балаково-Банк	444	0	Саратовская обл.	-0,04655	-0,03122	-0,07279	-0,05998	-0,14332
Крыловский	456	1	Краснодарский край	-0,04189	-0,03348	-0,06991	-0,06111	-0,14377
Новопокровский	467	0	Краснодарский край	-0,04214	-0,03556	-0,07703	-0,04678	-0,13768
Русский Национальный Банк	469	0	Ростовская обл.	-0,03855	-0,03500	-0,07469	-0,05149	-0,14489
Енисей	474	1	Красноярский край	-0,04529	-0,02334	0,01927	-0,07908	-0,14357
Кубаньторгбанк	478	0	Краснодарский край	-0,04985	-0,03552	-0,07205	-0,06252	-0,14692
Татсоцбанк	480	0	Татарстан	-0,04790	-0,02538	-0,07417	-0,06411	-0,15393
Ростфинанс	481	0	Ростовская обл.	-0,04290	-0,03028	-0,07298	-0,05556	-0,14221
Донхлеббанк	492	0	Ростовская обл.	-0,04279	-0,03326	-0,07403	-0,06375	-0,14623

Figure 1: Initial data for analysis.

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

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.

3. 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).

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

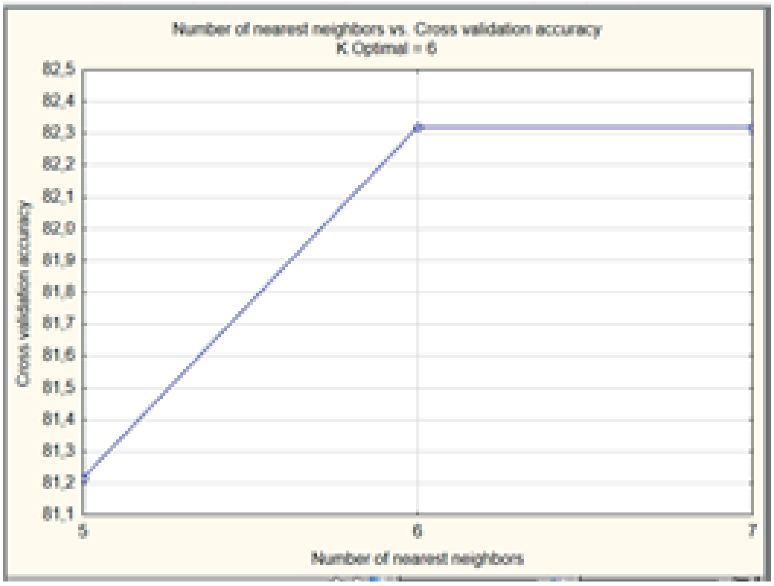

Figure 2: Optimal number of k-nearest neighbors in the binary classification.

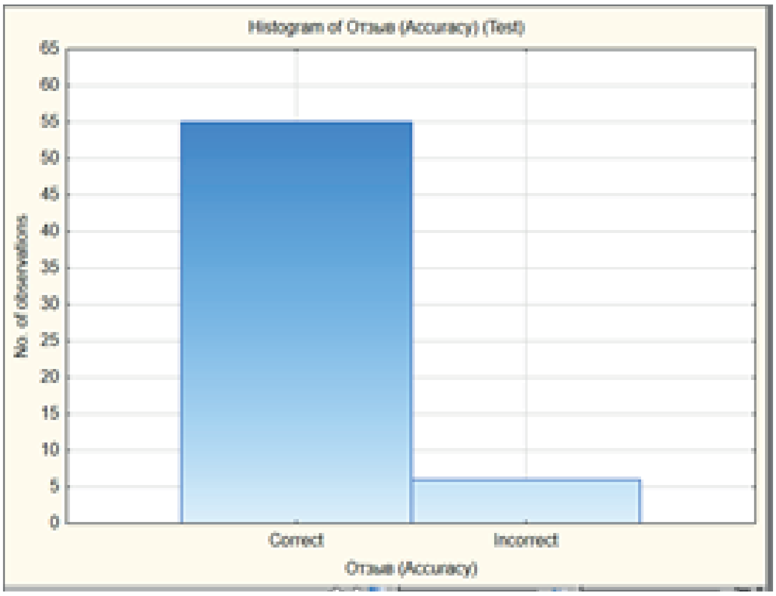

Figure 3: Results of binary classification by k-nearest neighbors method.

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.

4. 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.