COMPUTER RESEARCH AND MODELING
2020 VOL. 12 NO. 1 P. e201–e215
DOI: 10.20537/2076-7633-2020-12-1-201-215

MODELS OF ECONOMIC AND SOCIAL SYSTEMS

UDC: 330.46

Comparison of Arctic zone RF companies with different Polar Index ratings by economic criteria with the help of machine learning tools

L. R. Borisova[1,2], A. V. Kuznetsova[3,4,a], N. V. Sergeeva[4], O. V. Senko[5]

[1] Financial University under the Government of RF, 49 Leningrasky prosp., Moscow, 125993, Russia

[2] Moscow Institute of physics and technology, 9 Institutsky lane, Dolgprudny, Moscow region, 141700, Russia

[3] Institute for Biochemical Physics (IBCP), Russian Academy of Sciences (RAS), 4 Kosygina st., Moscow, 119334, Russia

[4] Azforus Ltd., 4 Kosygina st., Moscow, 119334, Russia

[5] Federal Research Center "Informatics and Control" of RAS, 44/2 Vavilova st., Moscow, 119333, Russia

E-mail: [a] azforus@yandex.ru

Received 13.09.2019, after completion — 01.11.2019.
Accepted for publication 14.11.2019.

The paper presents a comparative analysis of the enterprises of the Arctic Zone of the Russian Federation (AZ RF) on economic indicators in accordance with the rating of the Polar index. This study includes numerical data of 193 enterprises located in the AZ RF. Machine learning methods are applied, both standard, from open source, and own original methods — the method of Optimally Reliable Partitions (ORP), the method of Statistically Weighted Syndromes (SWS). Held split, indicating the maximum value of the functional quality, this study used the simplest family of different one-dimensional partition with a single boundary point, as well as a collection of different two-dimensional partition with one boundary point on each of the two combining variables. Permutation tests allow not only to evaluate the reliability of the data of the revealed regularities, but also to exclude partitions with excessive complexity from the set of the revealed regularities. Patterns connected the class number and economic indicators are revealed using the SDT method on one-dimensional indicators. The regularities which are revealed within the framework of the simplest one-dimensional model with one boundary point and with significance not worse than $p < 0.001$ are also presented in the given study. The so-called sliding control method was used for reliable evaluation of such diagnostic ability. As a result of these studies, a set of methods that had sufficient effectiveness was identified. The collective method based on the results of several machine learning methods showed the high importance of economic indicators for the division of enterprises in accordance with the rating of the Polar index. Our study proved and showed that those companies that entered the top Rating of the Polar index are generally recognized by financial indicators among all companies in the Arctic Zone. However it would be useful to supplement the list of indicators with ecological and social criteria.

Keywords: machine learning methods, sustainable development, Arctic Zone of the Russian Federation, economic criteria, the Polar Index of companies

Citation: *Computer Research and Modeling*, 2020, vol. 12, no. 1, pp. e201–e215 (Russian).

**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2020 Т. 12 № 1 С. e201–e215**

DOI: 10.20537/2076-7633-2020-12-1-201-215

Ки&М

МОДЕЛИ ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ

УДК: 330.46

Применение методов машинного обучения для сравнения компаний Арктической зоны РФ по экономическим критериям в соответствии с рейтингом Полярного индекса

Л. Р. Борисова[1,2], А. В. Кузнецова[3,4,a], Н. В. Сергеева[4], О. В. Сенько[5]

[1] Финансовый университет при Правительстве РФ, Россия, 125993, Ленинградский просп., д. 49

[2] Московский физико-технический институт (государственный университет), Россия, 141700, Московская обл., г. Долгопрудный, Институтский пер., д. 9

[3] Институт биохимической физики им. Н. М. Эмануэля, Россия, 119334, г. Москва, ул. Косыгина, д. 4

[4] ООО «Азфорус», Россия, 119334, г. Москва, ул. Косыгина, д. 4

[5] Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Россия, 119333, г. Москва, ул. Вавилова, д. 44, кор. 2

E-mail: [a] azforus@yandex.ru

Получено 13.09.2019, после доработки — 01.11.2019.
Принято к публикации 14.11.2019.

В работе проведен сравнительный анализ предприятий Арктической зоны Российской Федерации (АЗ РФ) по экономическим показателям в соответствии с рейтингом Полярного индекса. В исследование включены числовые данные 193 предприятий, находящихся в АЗ РФ. Применены методы машинного обучения, как стандартные, из открытых ресурсов, так и собственные оригинальные методы — метод оптимально достоверных разбиений (ОДР), метод статистически взвешенных синдромов (СВС). Проведено разбиение с указанием максимального значения функционала качества, в данном исследовании использовалось простейшее семейство разнообразных одномерных разбиений с одной-единственной граничной точкой, а также семейство различных двумерных разбиений с одной граничной точкой по каждой из двух объединяющих переменных. Перестановочные тесты позволяют не только оценивать достоверность данных выявленных закономерностей, но и исключать из множества выявленных закономерностей разбиения с избыточной сложностью.

Использование метода ОДР на одномерных показателях выявило закономерности, которые связывают номер класса с экономическими показателями. Также в приведенном исследовании представлены закономерности, которые выявлены в рамках простейшей одномерной модели с одной граничной точкой и со значимостью не хуже чем $p < 0.001$.

Для достоверной оценки подобной диагностической способности использовали так называемый метод скользящего контроля. В результате этих исследований был выделен целый набор методов, которые обладали достаточной эффективностью.

Коллективный метод по результатам нескольких методов машинного обучения показал высокую значимость экономических показателей для разделения предприятий в соответствии с рейтингом Полярного индекса.

Наше исследование доказало и показало, что те предприятия, которые вошли в топ рейтинга Полярного индекса, в целом распознаются по финансовым показателям среди всех компаний Арктической зоны. Вместе с тем представляется целесообразным включение в анализ также экологических и социальных факторов.

Ключевые слова: методы машинного обучения, устойчивое развитие, Арктическая зона РФ, экономические критерии, Полярный индекс компаний

1. Introduction

The purpose of this work was to provide an objective assessment of the sustainable development of companies in the Arctics zone of the Russian Federation (AZ RF) according to economic indicators from the point of view of the rating of the Polar Index.

The rating is a joint project of the expert center “PORA” and the Department of Environmental Economics of the Faculty of Economics of Lomonosov Moscow State University [Nikonorov, 2018; Nikonorov, Utkina, 2018]. “Polar index. Version 1.0” has become the first specialized rating of companies whose geography of activity includes the Arctic Zone of Russia. 17 large companies represented in the Arctic participated in the rating. Among them were 6 oil and gas, 3 metallurgical, 3 petrochemical, 2 energy concerns, one diamond mining and one transport company, and the State Atomic Energy Corporation. According to the Presidential Decree (No. 296, 2014) “On land territories of the Arctic Zone of the Russian Federation”, it includes the Murmansk Region, the Nenets, Yamalo-Nenets and Chukotka Autonomous Districts, and separate territories of the Krasnoyarsk Territory, Arkhangelsk Region and the Republics of Sakha (Yakutia), Karelia and Komi.

The criteria defining the Polar Index include three components: economic, environmental and social criteria. Only by giving equal attention to all three components is the enterprise able to increase the stability of the company under the difficult conditions of the development of the Arctic zone of the Russian Federation.

The Arctic is a strategic resource base, a zone of industrial development with high potential, which requires strict adherence to the principles of sustainable development. The concept of sustainable development is based on the coordination and balancing of the economic, social and environmental components of the development process of society itself, orienting the economic growth, in the long run, toward achieving social and environmental goals that ensure the growth of the level and quality of the life of people.

The complexity of the development of the Arctic Zone has the following prerequisites: weak development of territories, low population density, socio-economic problems associated with high cost and low standard of living, poorly developed transport and logistics links and poor infrastructure.

Solving these problems requires a systematic application of new modern technologies and methods. This approach is seen in many projects aimed at improving life in the Arctic Zone of the Russian Federation.

First of all, single-industry towns that completely depend on a single city-forming enterprise, the state of which fully determines the fate of the city, attract attention. The “resource potential” (human, intellectual, technological, infrastructural), focused on natural resources and the military industry, as it was established before, is now coming to an end.

New projects are emerging in the form of an adaptive concept of “Smart Arctic City”, which allows one to form stable distributed autonomous units in the AZ of the Russian Federation. It is assumed that such cities will be organically integrated into the economic structure of the region, forming unique cluster-type network ecosystems. An obligatory condition for the success of solving such projects is adaptation to the conditions and peculiarities of the AZ of the Russian Federation, taking into account the specifics of the Arctic.

It is necessary to incorporate an ecological approach into the condition of sustainable development of enterprises from the very beginning. A powerful economy with a high level of technology allows one to take a comprehensive approach to the use of alternative electric power in the AZ of the Russian Federation, to model the mechanism of public-private partnership, to preserve specially protected natural territories in the AZ of the Russian Federation with the help of concession mechanisms. By analogy with other countries, it is possible to use the recreational potential of the northern territories, attracting tens of thousands of tourists from all over the world. Substantial resources are needed to eliminate the already accumulated environmental damage. The creation and use of ecotechnoparks can help solve the problem of disposal of household and industrial waste.

The geopolitical tension in the Arctic zone dictates the need for socio-economic development of the Arctic territories of the Russian Federation [Dodin, 2005]. Only the use of the latest multiparametric methods of data analysis can help in justifying commercial efficiency and economic profitability, in assessing the socio-economic effect of the introduction of modern technologies, and will help give recommendations on the location of production, its equipment with technological equipment and personnel, provision of the necessary engineering infrastructure, as well as construction materials.

A new stage in the development of the Arctic dictates the use of artificial intelligence (AI) systems under the conditions of the flow of digital information collected about the multifaceted components of management and environmental management in the Far North. Collecting and analyzing information, processing it, creating a visual final analytical report on the conducted research are necessary components of the sustainable development of the region, improving the quality of life of Northerners and a new round of industrial development of the macroregion. The introduction of technologies with the possibility of their practical use in the economic activities of economic entities is more than relevant. It is necessary to promote the education and development of intellectual and technical creativity of young people in order to facilitate the development of these approaches not only by rare specialists in the field of AI.

The new financing instruments include such concepts as “responsible investment” and “green financing”. That is, of importance to the beneficiary of the project is not only the profit from its implementation, but also social benefits and, above all else, ensuring that the project brings about positive changes in the environmental situation rather than causing harm to the environment . It is extremely important to raise funds for projects related to renewable energy, energy efficiency, environmentally friendly transport or a low-carbon economy. The tools for evaluating investment objects cannot be applied without using new methods of analysis in the developing market. These are also relevant for the selection of municipalities for the implementation of investment projects. The desire to achieve maximum transparency in the selection of a region for pilot, infrastructure and innovative projects implemented by investment funds requires the creation of special objective rules for the implementation of the project in the region that has won the competition. In this matter, new methods of data analysis are also indispensable.

The tasks requiring the use of AI also include the issue of the health of people living in the Arctic region. There are studies in which it is planned to identify the factors that most strongly affect health and chronic diseases that occur under living conditions in the Far North. It is also planned to identify the genes involved in adaptation to living conditions in high latitudes. The data obtained will be used for the diagnosis and prevention of diseases, including cancer, in people living in the Arctic region.

All of these problems and tasks require a strict systematic analytical approach using the most modern information technologies, including the most up-to-date domestic methods of data analysis.

The elaborated methodology concerning the sustainable development of the enterprise systematizes well the criteria necessary for this. In the review by Kokin and Yakovleva, the following definition is proposed: “The financial stability of any economic entity is the ability to carry out basic and other types of activities under conditions of entrepreneurial risk and a changing business environment in order to maximize the welfare of owners” [Kokin, Yakovleva, 2010; Savitskaya, 2005].

Along with the criteria of financial stability (1st type: inflation, creditors' demands, bankruptcy of debtors, changes in the tax system, economic policy of the state, quality of products, fluctuations in exchange rates, seasonality of cash flows), other types are also considered:

type 2 is marketing sustainability (advertising, consumer income level, availability of supply and demand);

type 3 is production stability (stocks of raw materials, volume of production and launch of products, use of new technologies, and sale of products);

type 4 is social stability (the level of education of the team, labor market, wages, working conditions, and demographic problems);

type 5 is environmental sustainability (environmental protection activities, man-made disasters).

A strategy for the development of the Arctic zone of the Russian Federation and ensuring national security for the period up to 2020 has been developed and approved by the President of the Russian Federation on February 20, 2013 (No. Pr-232).

RAS specialists are actively involved in the substantiation and creation of the multipurpose space system “Arctic” for monitoring the situation in the northern latitudes. They are working on the creation of a system for remote (space) monitoring of natural and technological hydrocarbon releases to the surface of the waters of Russia.

In the future, it would be interesting to analyze the data of this monitoring to study the environmental sustainability of AZ enterprises.

Within the framework of cooperation between the Center for the Study of Economic Problems of Developing the Arctic at the Faculty of Economics of Lomonosov Moscow State University and the Arctic Development Project Office (PORA), the “Polar Index of the Barents Region” has been created. Within its framework, the following ratings were created:

1) the first rating of sustainable development of intrastate territorial units (provinces) Barents Region;
2) rating of sustainable development of companies operating in the Barents Region [Nikonorov, Utkina, 2018].

This paper focuses on the study of the possibilities of sustainable development of companies in the Arctic zone of the Russian Federation.

The use of innovative technologies in the Arctic can bring broad social and economic benefits both in the Arctic region and beyond. In [Nikonorov, Utkina, 2018], the “traffic light” principle was used, which made it possible to rank 26 analyzed companies of the Barents Arctic region from the Russian Federation, Finland, Sweden, Norway according to the sustainable development rating, considering only integral indicators according to the original methodology.

Modern computer technologies and methods make it possible to analyze multiparametric databases (Big Data). The proposed work uses the reporting data of Russian companies obtained from the Bureau-van Dyke database (Ruslana format — detailed global). A comparative analysis of economic indicators of enterprises of the Arctic Zone of the Russian Federation is carried out. The comparison groups included enterprises from the TOP list with a high rating of the Polar Index and enterprises that were not included in it.

The methods of machine learning on the training sample revealed a set of the most informative indicators that influence, to a greater or lesser extent, the forecast as to which class the enterprise will be assigned to: the first class (high Polar index) or the second. In the course of research, use was made of various machine learning methods, including original logical-statistical methods (Data Science) [Kuznetsova, Senko, 2005]. A recognition model that calculates collective estimates was used [Senko et al., 2011]. The logical-statistical approach allows for analysis without making a priori assumptions about the type of probability distributions, and this approach is also effective under the conditions of selection of any size and a large number of poorly structured features.

2. Materials and methods

Numerical data of 193 enterprises of the Russian Arctic were included in the study. The number of objects of the first class: 42 — subsidiaries of 17 companies included in the TOP list with a high rating of the Polar Index (see [Nikonorov, Utkina, 2018]). The number of objects of the second class: 151 companies not included in this list. The number of indicators in the initial training sample was 64. The indicators are taken from the Bureau-van Dyke database (Ruslana format — detailed global), kindly provided to the Financial University under the Government of the Russian Federation. Financial indicators were taken from the sections: “Balance Sheet”, “Profit and Loss Statement”, “Exchange information” — “Annual assessment of the company”.

To predict the TOP rating of the Polar Index by economic (financial) indicators, the “Recognition” system was used, which includes a set of machine learning methods. The following well-proven approaches were also used: the original recognition method “Method of statistically weighted syndromes” [Kuznetsova, Senko, 2005; Senko et al., 2011; Senko, Kuznetsova, 2010; Kuznetsova et al., 2011; Kuznetsova et al., 2011; Kirilyuk et al., 2017; Kuznetsova et al., 2018; Kuznetsova et al., 2018]. This method is based on making collective decisions on the areas of optimal partitions found using the method of optimally reliable partitions (ORP) [Senko et al., 2011]. The predictive ability was evaluated using the sliding control method, which gives an objective estimate of accuracy. Predictive ability in this case was understood as the correct assignment of an enterprise to a class from the analyzed sample according to a set of economic indicators.

Partitions with the maximum value of the quality functional are searched for within several families of various levels of complexity. In this study, we used the simplest family of all possible one-dimensional partitions with one boundary point, as well as a family of all possible two-dimensional partitions with one boundary point for each of the two explanatory variables. In other words, the latter family can be described as a set of two-dimensional partitions with two boundaries, each of which is parallel to only one of the coordinate axes. The optimal partition is considered to be a law only after a positive result of statistical verification. The verification is performed using permutation tests based on comparing the quality of optimal partitions on the initial analyzed sample with the quality of partitions on a large number (2-3 thousand) of random samples obtained from the initial training sample using random permutations of the values of the predicted value relative to fixed positions of vectors of predictive variables. Adjustment tests make it possible not only to assess the reliability of the revealed regularities, but also to exclude from the set of output regularities partitions with real complexity.

3. Results and discussion

At the first stage, studies were conducted to identify the economic indicators that are the most informative from the point of view of separating the two classes being compared. The use of the ORP method on one-dimensional indicators revealed patterns linking the class number with economic indicators.

3.1. One-dimensional patterns

Table 1 shows, in descending order of informativeness, the economic indicators that make the greatest contribution to the separation of the two classes of enterprises under study, that is, they are informative in recognition.

Table 1. One-dimensional partitions obtained by the method of optimal reliable partitions

N*	Indicator	Border	lower		higher		F	$p <$
			group 1**	group 2	group 1	group 2		
1	Basic funds FIAS	14.28	12 (28.6 %)	0 (0 %)	30 (71.4 %)	151 (100 %)	45.76	0.0005
2	Total assets	15.14	11 (26.2 %)	0 (0 %)	31 (73.8 %)	151 (100 %)	41.72	0.0005
3	Total equity and liabilities	15.14	11 (26.2 %)	0 (0 %)	31 (73.8 %)	151 (100 %)	41.72	0.0005
4	Sales revenue / turnover	18.09	25 (59.5 %)	145 (96 %)	17 (40.5 %)	6 (4 %)	41.5	0.0005
5	Cost of production	18.56	31 (73.8 %)	150 (99.3 %)	11 (26.2 %)	1 (0.7 %)	36.53	0.0005
6	Other equity (including reserves)	17.08	22 (52.4 %)	137 (90.7 %)	20 (47.6 %)	14 (9.3 %)	33.12	0.0005
7	Other equity	16.08	32 (76.2 %)	150 (99.3 %)	10 (23.8 %)	1 (0.7 %)	32.59	0.0005
8	Equity	17.23	20 (47.6 %)	133 (88.1 %)	22 (52.4 %)	18 (11.9 %)	32.57	0.0005
9	Other operating expenses	15.98	28 (66.7 %)	145 (96 %)	14 (33.3 %)	6 (4 %)	30.34	0.0005
10	Retained earnings	18.42	33 (78.6 %)	150 (99.3 %)	9 (21.4 %)	1 (0.7 %)	28.69	0.0005
11	Reserves for upcoming expenses	14.25	31 (73.8 %)	148 (98 %)	11 (26.2 %)	3 (2 %)	28.46	0.0005
12	Other current liabilities	14.87	30 (71.4 %)	146 (96.7 %)	12 (28.6 %)	5 (3.3 %)	25.97	0.0005
13	Reserves	13,57	32 (76.2 %)	148 (98 %)	10 (23.8 %)	3 (2 %)	24.78	0.0005
14	Creditors (suppliers and contractors)	16.88	29 (69 %)	144 (95.4 %)	13 (31 %)	7 (4.6 %)	24.37	0.0005
15	Material fixed assets	17.68	24 (57.1 %)	135 (89.4 %)	18 (42.9 %)	16 (10.6 %)	23.44	0.0005
16	Interest paid	9.752	27 (64.3 %)	37 (24.5 %)	15 (35.7 %)	114 (75.5 %)	23.34	0.0005
17	Taxes	14.38	27 (64.3 %)	140 (92.7 %)	15 (35.7 %)	11 (7.3 %)	22.67	0.001
18	Gross profit	16.25	22 (52.4 %)	130 (86.1 %)	20 (47.6 %)	21 (13.9 %)	22.21	0.000333
19	Management expenses	15.83	36 (85.7 %)	151 (100 %)	6 (14.3 %)	0 (0 %)	22.15	0.0005
20	Other financial expenses	11.09	25 (69.4 %)	42 (28.6 %)	11 (30.6 %)	105 (71.4 %)	20.7	0.0005
21	Working capital	17.18	26 (61.9 %)	136 (90.1 %)	16 (38.1 %)	15 (9.9 %)	19.23	0.000333
22	Loans	3,841	25 (59.5 %)	37 (24.5 %)	17 (40.5 %)	114 (75.5 %)	18.39	0.001

* The border is represented in a logarithmic scale, that is, the table shows the natural logarithm of the corresponding indicator in thousands of rubles.

** Class 1 is a group of enterprises included in the top rating of the Polar Index; Class 2 is a group of enterprises not included in it.

The regularities revealed within the framework of the simplest one-dimensional model with one boundary point [Senko, Kuznetsova, 2010] are presented, with a value no worse than $p < 0.001$. For all indicators taking real numerical values, the boundary was set automatically according to the database using the chi-square method. The informativeness and statistic reliability of the relationship with the predicted value (high or low level of the polar index) in both cases were evaluated using the method of optimal reliable partitions.

The F-coefficient shows the degree of reliability of the pattern. It indicates how many times the value of the functional describing the quality of the partition on the original sample exceeds the maximum value of the functional of the quality of partitions on random samples (permutation test). The F-coefficients allow us to compare the reliability for two patterns in which the value of the quality functional on random samples has never exceeded the value on real data.

As can be seen from Table 1, the following indicators have become the most informative: "Fixed assets" — the source of FIAS, "Total assets", "Total capital and liabilities". For them, the values of the first class are below the boundary, and there are no values of the second class at all. But for the following indicators, the values of the 1st class prevail above the border: Revenue from sales / Turnover, Cost of production, Other own capital (including reserves), Other equity, Equity, Other operating expenses, Retained earnings, Reserves of upcoming expenses, Other current liabilities, Reserves, Creditors (suppliers and contractors), Material fixed assets, etc.

The scattering diagrams (Figs. 1 and 2) clearly show the values for enterprises of Class 1 (red crosses) and Class 2 (green circles) according to the corresponding indicators. The boundaries of the division are set in such a way that objects of the same class predominate on one side of the border.

For the indicator "Total assets" Fig. 1 shows that below the boundary equal to 15.14. objects of the first class predominate — 11 (26.2 %), objects of the 2nd class are absent. Above the border, respectively, objects of class 2 predominate — 151 objects (100 %) against 31 objects from the first class (73.8 %).

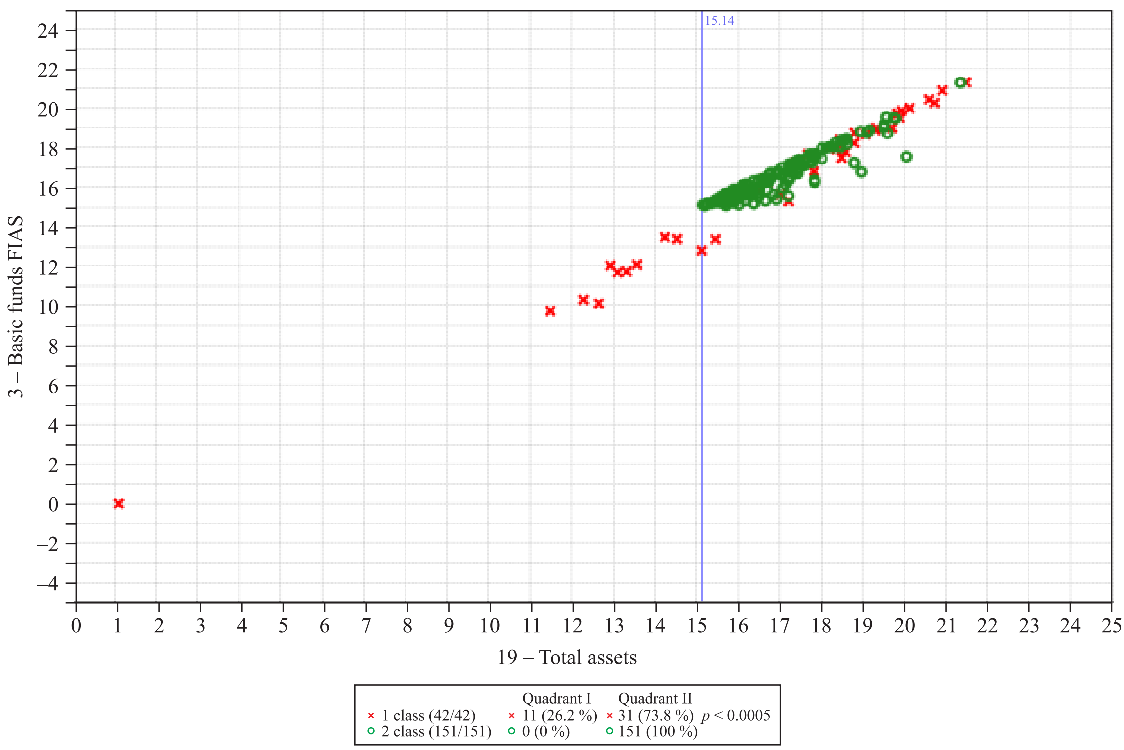

Fig. 1. The indicator "Total assets" on the abscissa axis. On the ordinate axis is an auxiliary indicator. In the first quadrant, the values of the first class prevail. In the second quadrant, the values of the second class prevail.

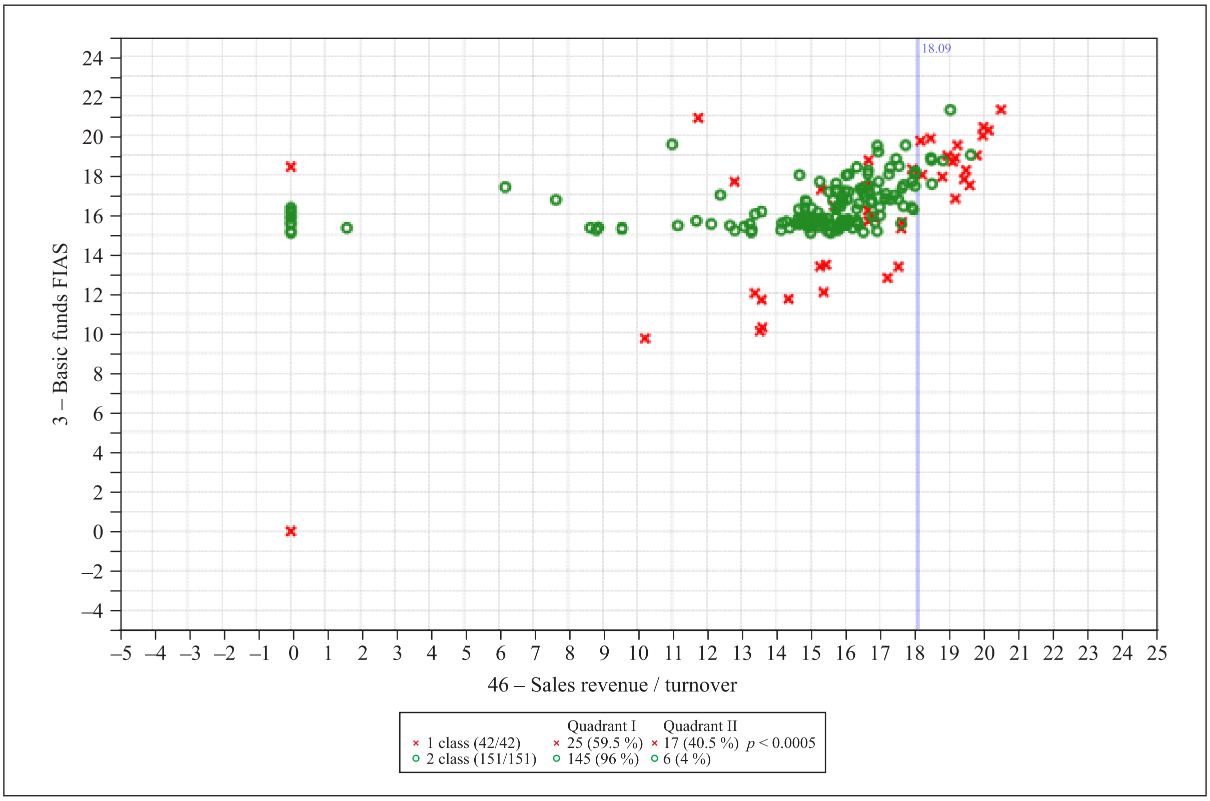

Fig. 2. The natural logarithm of the indicator "Sales revenue / turnover", expressed in thousands of rubles, is plotted on the abscissa axis. The second indicator is auxiliary. The first quadrant is dominated by the second class (green circles). The second quadrant is dominated by the first class (red crosses).

It can be seen in Figure 2 that objects of the second class predominate to the left of the 18.09 border: 145 objects (96 %). Above the border, 17 values prevail — 40.5 % — of objects of the high rating class of the Polar Index. While the values of the second class are only 4 %. The reliability of this pattern at the level of $p < 0.0005$ was obtained using a permutation test.

3.2. Two-dimensional regularities

Far more reliable splits were revealed on pairs of indicators. As can be seen in the scattering diagram (Fig. 3), the boundaries of the partitions are located parallel to the coordinate axes, forming 4 quadrants numbered clockwise. In this case, the upper left quadrant has number 1.

Figure 3 shows that the first class prevails in the second and fourth quadrants. Enterprises with the values of the fixed assets logarithm (FIAS) below 14.277 and the production cost logarithm above 17.985 have a chance to get into the top of the Polar Index rankings.

3.3. Application of machine learning methods (*recognition methods*)

Recognition methods were used to investigate the relationship of economic indicators of companies with their position in the rating of the Polar Index (Companies). For a reliable assessment of the diagnostic ability, the sliding control method was used. As a result, a set of methods with sufficient efficiency was identified (Table 2). The structure of the recognition algorithm allows for diagnostics in the sense of assigning an enterprise to one of the groups [Kuznetsova, Senko, 2005; Senko et al., 2011; Senko, Kuznetsova, 2010; Kuznetsova et al., 2011; Kuznetsova et al., 2011; Kirilyuk et al., 2017; Kuznetsova et al., 2018; Kuznetsova et al., 2018]. The accuracy of the algorithm is determined by the total number of matches of the automatic diagnosis and the real class number.

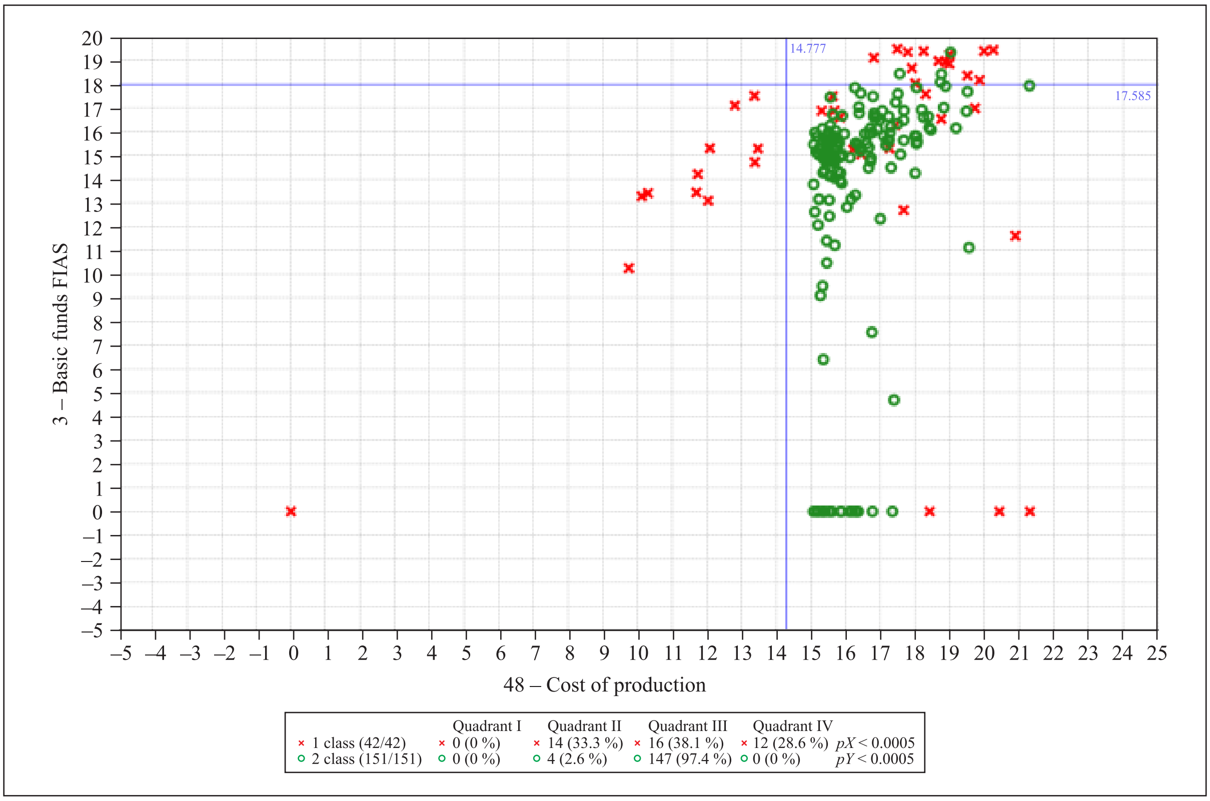

Fig. 3. Indicators: abscissa axis — “Fixed assets, FIAS source”, ordinate axis — “Production cost”. There are no values in the first quadrant (upper left). In the second quadrant (upper right), the first class prevails. In the third quadrant (lower right), the first class prevails. In the fourth quadrant (lower left), the first class prevails.

In any method, the recognition of an arbitrary object can be represented as the sequential execution of two operations. At the first step, values are calculated that reflect the measure of the object's proximity to one of the classes. This value is called “grade per class”. In the second step, the actual recognition is performed. Usually a certain threshold ***d*** is used. For many machine learning methods, estimates are usually calculated in such a way that they belong to the segment [0, 1]. If the “grade for the class” of an object is below the threshold ***d***, then they are assigned to the first class, and if the score is above the threshold, then it is rated in the second class. All performance indicators, except ROC AUC, depend on the threshold value ***d***. Table 2 shows the performance indicators for two values of the threshold ***d***: 0.4 and 0.5.

Table 2 shows that the best and most balanced solution is a collective solution with the maximum value for Accuracy and F-measure, as well as high values for other indicators.

As can be seen from Table 2, the best result has been obtained when using the “Decisive Forest” method, the specificity (recognition of the second class) is especially high — 96.7 %. The method of statistically weighted syndromes (SWS) has also yielded a good result. The method of support vectors (MSV) showed the best result in the categories of “accuracy” and “specificity”.

To obtain a collective solution, the average value of the estimates of the probability of objects falling into classes computed by various methods was calculated. The collective method compensates for the errors of individual algorithms included in the collective. It is theoretically proved that the error of the collective solution in forecasting problems does not exceed the average error calculated for the methods separately. All this suggests that the collective solution [Senko, Kuznetsova, 2010] is highly reliable.

Table 2. The results of machine learning methods and a collective solution

Method	Threshold	Accuracy	Sensitivity	Specificity	Precision	F-measure
Collective solution	0.4	0.886	0.619	0.96	0.813	0.702
SVM	0.4	0.808	0.738	0.828	0.544	0.626
SWS	0.4	0.725	0.762	0.715	0.427	0.548
Decision Forest	0.4	0.87	0.619	0.94	0.743	0.676
Collective solution	0.5	0.87	0.429	0.993	0.948	0.590
SVM	0.5	0.881	0.524	0.98	0.88	0.656
SWS	0.5	0.839	0.643	0.894	0.628	0.636
The Decision Forest	0.5	0.876	0.548	0.967	0.821	0.658

The better the recognition results obtained using the machine learning method or a collective of MO methods (ensemble), the more convex is the shape of the ROC curve. In this case, we have a high recognition result. The area under the AUC curve is close to 0.9.

The training sample has turned out to be informative enough to separate one class under study from another.

Based on the results of recognition of the second class, a conclusion can be drawn about enterprises that were mistakenly recognized as Class 1: they have a great potential to occupy a high position in the rating of the Polar Index.

Table 3. ROC AUC results for machine learning methods

Machine learning method	ROC AUC value
Decision forest	0.895
SVM	0.872
SWS	0.847
Nearest Neighbor method (KNN)	0.748

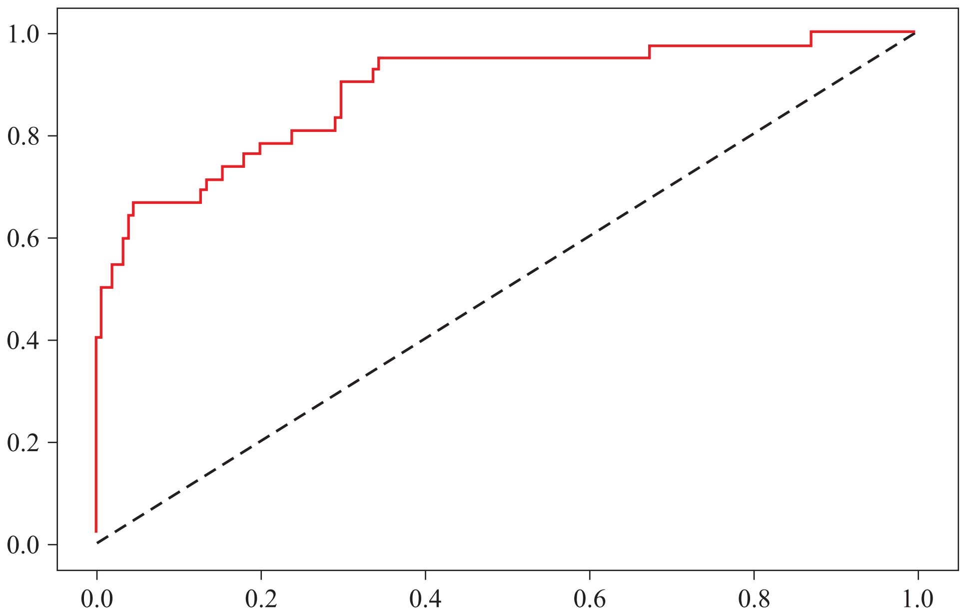

Fig. 4. ROC curve for a collective solution, AUC = 0.888

At the same time, for some part of the companies that experts attribute to Class 1 (high Polar Index), an error was received during recognition. The latter is indicative of certain deviations in the reporting indicators of such companies from the typical values for enterprises of the top rating of the Polar Index. For example, from Fig. 5 it can be seen that to the values of the logarithm of retained earnings above 15.14 (quadrant II, earnings logarithm) there correspond 9 objects of class 1 and only one object of class 2. The contribution of the "Retained earnings" attribute for all objects from quadrant II will correspond to assignment to the first class. For the only one of the enterprises from class 2 in quadrant II, it is advisable to conduct an additional analysis of the reasons why it did not get into the top of the rating.

The method of statistically weighted syndromes is based on voting on the base sets determined by the boundaries of the partitions obtained by the SDT methods. If the value of an object of one class falls into the base set with the predominance of the values of objects of another class, there is a high probability that it will be assigned to another class. Hence, recognition errors occur in this method.

Figure 6 shows a two-dimensional scattering diagram. It is possible to single out a relatively small number of companies that are not included in the top rating (Class 2), for which financial indicators correspond to the first class. In Fig. 6, in the second quadrant with high values of total assets and production costs, the values of Class 1 enterprises (a high rating of the Polar Index) prevail — 14 objects. But at the same time there are 4 enterprises from the 2nd class. At the same time, in the third quadrant, among the majority of values of the second class (147, 97.4 %), there are 17 objects of the first class.

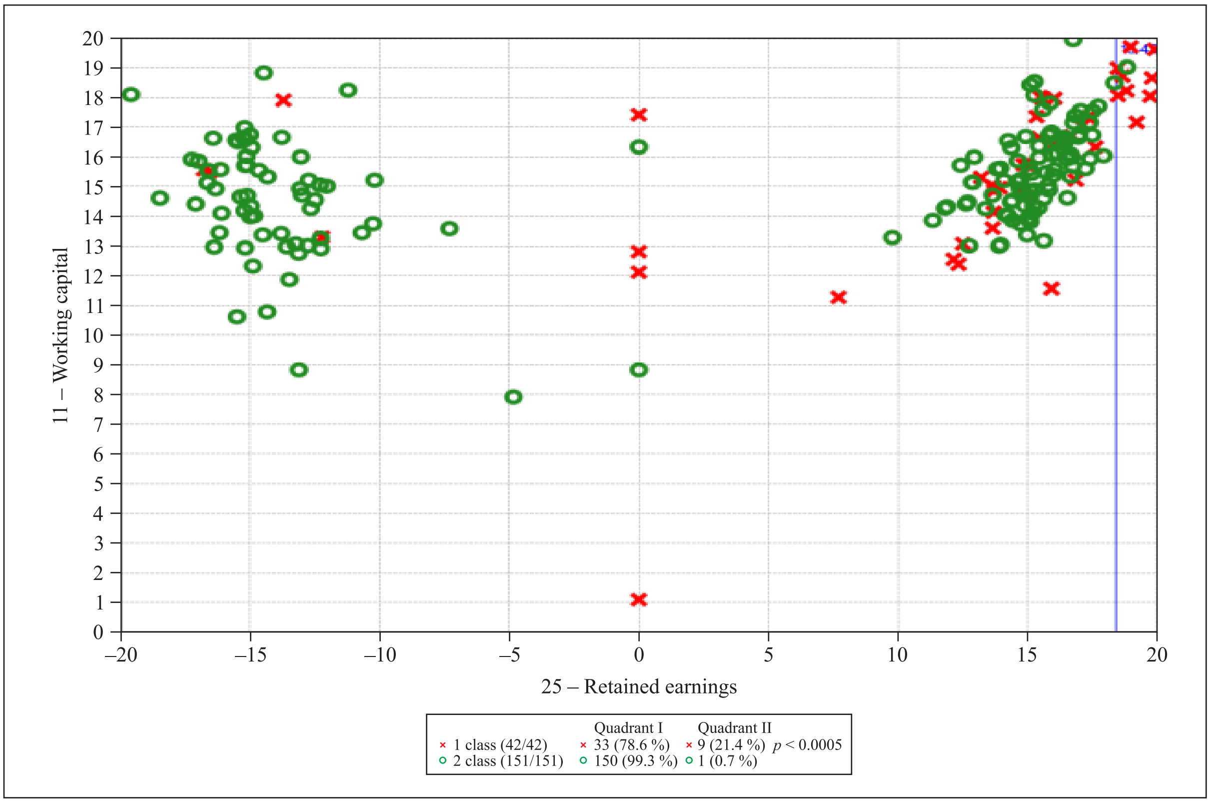

Fig. 5. An example of a possible error during recognition. According to the indicator "Retained earnings" to the left of the border equal to 18.42, 33 objects of Class 1 (x) are among 150 objects of class 2 (o). To the right of the border, one enterprise of the 2nd class is located among 9 objects of the 1st class. In this case, there is a high probability that an object of class 2 will be mistakenly assigned to class 1 by the recognition algorithm

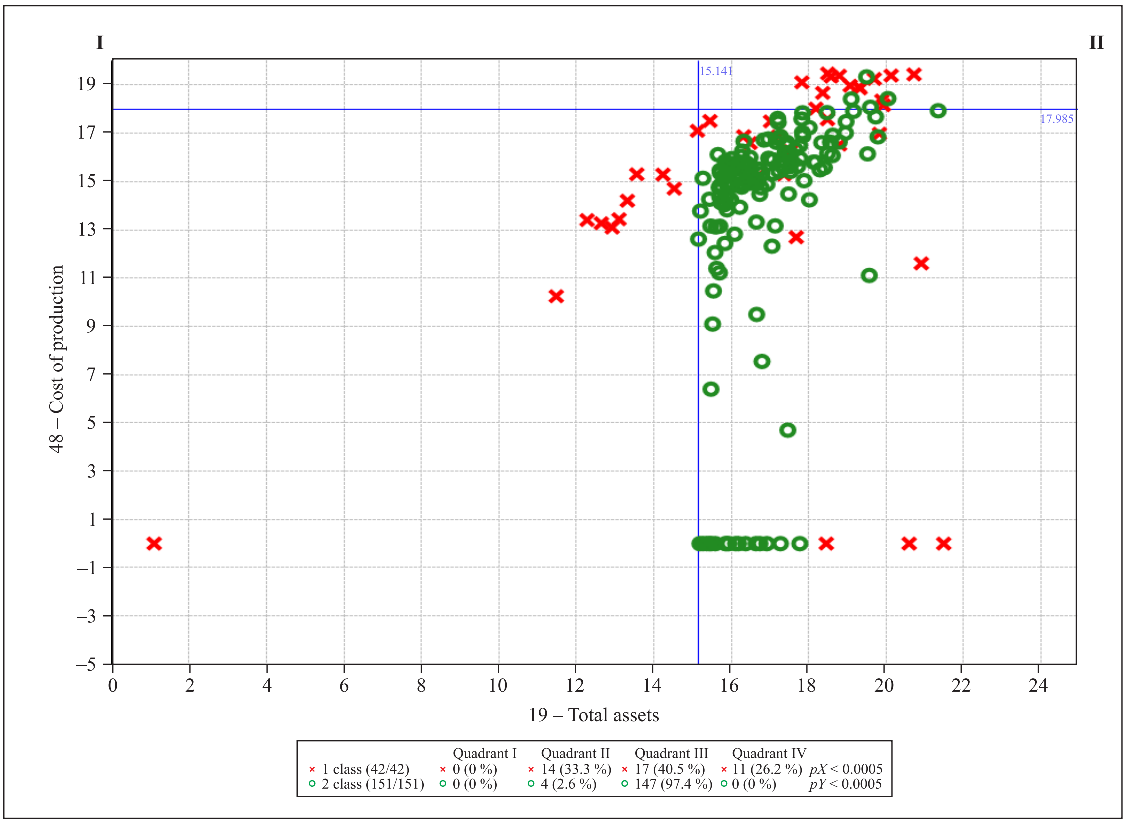

Fig. 6. Two-dimensional scattering diagram. The X-axis is the logarithms of total assets; the Y-axis is the logarithms of the cost of production

For such companies, you can try to conduct an additional analysis of factors other than the financial reporting indicators used and which determined a low rating on the rating of the Polar Index. Correction of such factors potentially allow I would like to increase the efficiency of companies' activities in terms of its sustainability.

4. Conclusions

Our research has shown that the enterprises included in the top rating of the Polar Index are generally recognized by financial indicators among the companies of the Arctic zone. For a collective solution, AUC = 0.888, the sensitivity is 62 %, and the specificity is 96 %.

Thus, the resulting decisive rule can be used as a prognostic algorithm to assess the potential of the company in terms of the rating of the Polar Index (company).

Due to the fact that the pilot version of the program for assessing the sustainability of an enterprise by economic criteria has shown its effectiveness, it can be recommended to add numerical information on social and environmental criteria to the database. This will create a decisive rule for predicting the sustainable development of the enterprise according to all three criteria of the Polar Index: economic, social and environmental.

5. Thanks

Many thanks to ANO “EXPERT CENTER OF TIME” for supporting this research Project in the form of grant No.153-G.

We express our gratitude to Igor Leonidovich Kirilyuk, Junior Researcher of the Institute of Economics of the Russian Academy of Sciences, for his help in discussing the materials of this article.

References

Додин Д. А. Устойчивое развитие Арктики: проблемы и перспективы. — СПб.: Наука, 2005. — 283 с.

Dodin D. A. Ustoichivoye razvitie Arktiki: problemy i perspektivy [Sustained development of Arctic region: problems and prospects]. — Saint Petersburg: Nauka, 2005. — 283 p. (in Russian).

Кокин А. С., Яковлева Г. Н. Показатели устойчивости организации // Экономические науки. Вестник Нижегородского университета им. Н. И. Лобачевского. — 2010. — № 3 (1). — С. 256– 261.

Kokin A. S., Yakovleva G. N. Pokazateli ustoychivosti organizatsii // Ekonomicheskiye nauki. Vestnik Nizhegorodskogo universityeta im. N. I. Lobachevskogo [Organizational sustainability indicators. Economic sciences. Bulletin of the Nizhny Novgorod University N. I. Lobachevsky]. — 2010. — No. 3 (1). — P. 256–261 (in Russian).

Кузнецова А. В., Сенько О. В. Возможности использования методов Data Mining при медико-лабораторных исследованиях для выявления закономерностей в массивах данных // Врач и информационные технологии. — 2005. — Т. 1. — С. 49–56.

Kuznetsova A. V., Senko O. V. Vozmozhnosti ispolzovaniya metodov Data Mining pri mediko-laboratornykh issledovaniyakh dlya vyiavleniya zakonomernostey v massivakh dannykh [Possibilities of Data Mining methods in medical lab research to identify patterns in data arrays] // Vrach i informatsionnyie tehnologii. — 2005. — Vol. 1. — P. 49–56 (in Russian).

Кузнецова В., Костомарова И. В., Водолагина Н. Н., Малыгина Н. А., Сенько О. В. Изучение влияния клинико-генетических факторов на течение дисциркуляторной энцефалопатии с использованием методов распознавания // Матем. биология и биоинформ. — 2011. — Т. 6, № 1. — С. 115–146. — URL: http://www.matbio.org/2011/ Senko2011(6_115).pdf (дата обращения: 25.03.2013).

Kuznetsova V., Kostomarova I. V., Vodolagina N. N., Malygina N. A., Sen'ko O. V. Izucheniye vliyaniya kliniko-geneticheskikh faktorov na techeniye distsirkulyatornoy entsefalopatii s ispolzovaniyem metodov raspoznavaniya [Study of effects of clinical and genetic factors on severity of discirculatory encephalopathy with the help of pattern recognition methods] // Matem. Biologia i bioinform. — 2011. — Vol. 6, No. 1. — P. 115–146 (in Russian). — Avialable at: URL: http://www.matbio.org/2011/ Senko2011(6_115).pdf (accessed 25.03.2013).

Кузнецова А. В., Сенько О. В., Кузнецова Ю. О. Преодоление проблемы черного ящика при использовании методов машинного обучения в медицине // Врач и информационные технологии. Искусственный интеллект в здравоохранении. — 2018. — № 7 (октябрь). — С. 74–80. — URL: http://www.idmz.ru/media/vit_ru/2018/4/vititm_2018_07pdf.pdf

Kuznetsova A. V., Senko O. V., Kuznetsova Yu. O. Overcoming the black box problem when using machine learning methods in medicine // Preodoleniye problemy chornogo yashchika pri ispolzovanii metodov mashinnogo obucheniya v meditsine [Doctor and information technology. Artificial Intelligence in Health Care] // Vrach i informatsionnyye tekhnologii. Iskusstvenny intellect v zdravookhranenii. — 2018. — No. 7 (Oct). — P. 74–80 (in Russian).

Кузнецова Ю. О., Борисова Л. Р., Кузнецова А. В., Сенько О. В. Прозрачный интерфейс для прогноза в машинном обучении // Сборник научных трудов XIX Международной конференции DAMDID/RCDL. — 2017. — С. 493–495.

Kuznetsova Yu. O., Borisova L. R., Kuznetsova A. V., Senko O. V. Prozrachnyiy interfeys dlya prognoza v mashinnom obuchenii [A transparent interface for forecasting in machine learning] // M. MGU. Sbornik nauchnyih trudov XIX Mezhdunarodnoy konferentsii DAMDID/RCDL'. — 2017. — P. 493–495 (in Russian).

Никоноров С. М., Уткина Е. Э. От адаптации целей устойчивого развития в Арктике к разработке Полярного индекса Баренц-региона (регионы и компании) // Экономика устойчивого развития. — 2019. — Т. 38, № 2. — С. 272–277.

Nikonorov S. M., Utkina E. E. Ot adaptacii celei ustoichivogo razvitia v Arktike k razrabotke poliarnogo indexa Barents Regiona (regiony i kompanii) [From the adjustment of the Sustainable Development Goals in the Arctic to the development of the Polar Index of the Barents Region (regions and companies)] // Economics of stable development. — 2019. — Vol. 38, No. 2. — P. 272–277 (in Russian).

НИР «Рейтинг устойчивого развития регионов и компаний Баренцева Евро-Арктического региона — Полярный индекс Баренц-региона» — рук-ль проекта профессор МГУ С. М. Никоноров. — 2018.

NIR “Reiting ustoichivogo razvitia regionov i kompaniy Barenceva Evro-Arkticheskogo regiona — Poliarny index Berants-regiona” — rukovoditel proyekta prefessor MGU S. M. Nikonorov [R&D “Sustained development rates of regions of Barents see Euro-Arctic region — Barents region Polar Index” — project manager MSU prof. S. M. Nikonorov]. — 2018 (in Russian).

Савицкая Г. В. Анализ хозяйственной деятельности предприятия: учебник. — 2-е изд., испр. и доп. — М.: Новое знание, 2005.

Savitskaya G. V. Analiz hozyaystvennoy deyatelnosti predpriyatiya: uchebnik. — 2-ye izd., ispr. i dop. [Analysis of enterprise economic activity: a textbook. 2nd ed., Rev. and add.]. — Moscow: Novoye znaniye, 2005 (in Russian).

Kirilyuk I. L., Kuznetsova A. V., Senko O. V., Morozov A. M. Method for detecting significant patterns in panel data analysis // Pattern Recognition and Image Analysis. — 2017. — Vol. 27, No. 1. — P. 94–104.

Senko O., Kuznetsova A. A recognition method based on collective decision making using systems of regularities of various types // Pattern Recognition and Image Analysis. — 2010. — Vol. 20, No. 2. — P. 152–162.

Senko O. V., Kuznetsova A. V., Malygina N. A., Kostomarova I. V. Method for Evaluating of Discrepancy between Regularities Systems in Different Groups // Information Technologies & Knowledge. — 2011. — Vol. 5, No. 1. — P. 46–54.