

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. Сергеева⁴, О. В. Сенько⁵

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

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. Sergeeva⁴, O. V. Senko⁵

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

² Moscow Institute of physics and technology, 9 Institutsky lane, Dolgoprudny, 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. 201–215 (Russian).

1. Введение

Целью данной работы было создание объективной оценки устойчивого развития компаний Арктической зоны Российской Федерации (АЗ РФ) по экономическим показателям с точки зрения рейтинга Полярного индекса.

Рейтинг является совместным проектом экспертного центра «ПОРА» и кафедры экономики природопользования экономического факультета МГУ имени М. В. Ломоносова [Никонов, 2018; Никонов, Уткина, 2018]. «Полярный индекс. Версия 1.0» стал первым специализированным рейтингом компаний, география деятельности которых затрагивает Арктическую зону России. В рейтинге участвовали 17 крупных компаний, представленных в Арктике. Среди них 6 нефтегазовых, 3 металлургических, 3 нефтехимических, 2 энергетических концерна, по одной алмазодобывающей и транспортной компании, а также государственная корпорация по атомной энергии. Согласно Указу Президента (№ 296, 2014 г.) «О сухопутных территориях Арктической зоны Российской Федерации», в нее входят полностью Мурманская область, Ненецкий, Ямало-Ненецкий и Чукотский автономные округа, а также отдельные территории Красноярского края, Архангельской области и республик Саха (Якутия), Карелия и Коми.

В состав критериев, определяющих Полярный индекс, входят три составляющие: экономические, экологические и социальные критерии. Только равноценное внимание, уделяемое предприятием всем трем составляющим, позволяет повышать устойчивость компании в сложных условиях развития Арктической зоны РФ.

Арктика — стратегическая ресурсная база, зона экономики промышленного развития с высоким потенциалом, требующим неукоснительного соблюдения принципов устойчивого развития. В основе концепции устойчивого развития находится согласование и уравнивание экономической, социальной и экологической составляющих самого процесса развития общества, ориентируя рост экономики в долгосрочной перспективе на достижение социальных и экологических целей, обеспечивающих рост уровня и качества жизни людей.

Сложность освоения Арктической зоны имеет следующие предпосылки: слабая освоенность территорий, низкая плотность населения, социально-экономические проблемы, связанные с высокой стоимостью и низким уровнем жизни, плохо развитыми транспортно-логистическими связями и неудовлетворительным состоянием инфраструктуры.

Решение данных проблем требует системного применения новых современных технологий и методов. Такой подход виден во многих проектах, направленных на улучшение жизни в Арктической зоне РФ.

В первую очередь обращают на себя внимание моногорода, целиком и полностью зависящие от единственного градообразующего предприятия, состояние которого полностью определяет судьбу города. Ресурсный потенциал (кадровый, интеллектуальный, технологический, инфраструктурный), ориентированный на природные ресурсы и военную отрасль, как было заведено прежде, в настоящее время подходит к концу.

Возникают новые проекты в виде адаптивной концепции «Умный арктический город», позволяющей формировать устойчивые распределенно-автономные единицы в АЗ РФ. Предполагается, что такие города будут органично интегрированы в экономическую структуру региона, формируя уникальные сетевые экосистемы кластерного типа. Обязательным условием успешности решения подобных проектов является адаптация к условиям и особенностям АЗ РФ, учет специфики Арктики.

В условие устойчивого развития предприятий нужно с самого начала закладывать экологический подход. Мощная экономика с высоким уровнем технологий позволяет комплексно подходить к использованию альтернативной электроэнергетики в АЗ РФ, моделировать механизм частно-государственного партнерства, сохранять с помощью концессионных механизмов особо охраняемые природные территории в АЗ РФ. По аналогии с другими странами возможно использовать рекреационный потенциал северных территорий, привлекая десятки тысяч туристов со всего мира. Существенные ресурсы необходимы для ликвидации уже накопленного

экологического ущерба. Создание и использование экотехнопарков может помочь решить проблему утилизации бытовых и промышленных отходов.

Геополитическая напряженность в Арктической зоне диктует необходимость социально-экономического развития арктических территорий РФ [Додин, 2005]. Только использование новейших многопараметрических методов анализа данных может помочь в обосновании коммерческой эффективности и экономической рентабельности, в оценивании социально-экономического эффекта от внедрения современных технологий, поможет дать рекомендации по размещению производства, его оснащению технологическим оборудованием и персоналом, обеспечению необходимой инженерной инфраструктурой, а также расходными материалами.

Новый этап освоения Арктики диктует применение систем искусственного интеллекта (ИИ) в условиях потока цифровой информации, собираемой о многогранных составляющих хозяйствования и природопользования в условиях Крайнего Севера. Сбор и анализ информации, ее обработка, создание наглядного итогового аналитического отчета о проведенных исследованиях — необходимые составляющие устойчивого развития региона, улучшения качества жизни северян и нового витка промышленного освоения макрорегиона. Более чем актуально внедрение технологий с возможностью их практического использования в хозяйственной деятельности экономических субъектов. Необходимо способствовать образованию и развитию интеллектуального и технического творчества молодежи, чтобы содействовать освоению данных подходов не только редкими специалистами в области ИИ.

К новым инструментам финансирования относятся такие понятия, как ответственное инвестирование и зеленое финансирование. То есть для бенефициара проекта важна не только прибыль от его реализации, но и социальные выгоды, а самое главное — чтобы проект не только не вредил экологии, а происходили позитивные изменения в экологической обстановке. Чрезвычайно актуально привлечение средств на проекты, связанные с возобновляемой энергией, повышением энергоэффективности, экологически чистым транспортом или низкоуглеродной экономикой. Инструменты оценки объектов инвестирования не могут обойтись без применения развивающегося рынка новых методов анализа. Это актуально и для отбора муниципальных образований по реализации инвестиционных проектов. Стремление достичь максимальной прозрачности при отборе региона под пилотные, инфраструктурные и инновационные проекты, реализуемые инвестиционными фондами, требует создания особых объективных правил реализации проекта в регионе, победившего в конкурсе. В этом вопросе также не обойтись без новых методов анализа данных.

К задачам, требующим использования ИИ, относится и вопрос здоровья человека при жизни в Арктическом регионе. Существуют исследования, в которых планируется определить факторы, наиболее сильно воздействующие на здоровье и хронические заболевания, возникающие в условиях жизни на Крайнем Севере. Также планируется определение генов, задействованных в адаптации к условиям жизни в высоких широтах. Полученные данные будут использованы для диагностики и профилактики заболеваний, в том числе онкологических, у людей, живущих в Арктическом регионе.

Все перечисленные проблемы и задачи требуют строгого системного аналитического подхода с применением самых современных информационных технологий, в том числе и новейших отечественных методов анализа данных.

Проработанная методология, касающаяся устойчивого развития предприятия, систематизирует критерии, необходимые для этого. В обзоре Кокина и Яковлевой предложено следующее определение: «Финансовая устойчивость любого хозяйствующего субъекта — это способность осуществлять основные и прочие виды деятельности в условиях предпринимательского риска и изменяющейся среды бизнеса с целью максимизации благосостояния собственников, укрепления конкурентных преимуществ организации» [Кокин, Яковлева, 2010; Савицкая, 2005].

Наряду с критериями финансовой устойчивости (1-й вид: инфляция, требования кредиторов, банкротство должников, изменение налоговой системы, экономическая политика государ-

ства, качество выпускаемой продукции, колебание валютных курсов, сезонность поступления денежных потоков) рассматривают также и другие виды:

2-й вид — маркетинговая устойчивость (реклама, уровень доходов потребителей, наличие спроса и предложения);

3-й вид — производственная устойчивость (запасы сырья, объем производства и выпуска продукции, использование новых технологий, реализация продукции);

4-й вид — социальная устойчивость (уровень образования коллектива, рынок рабочей силы, оплата и условия труда, демографические проблемы);

5-й вид — экологическая устойчивость (природоохранная деятельность, техногенные катастрофы).

Была разработана стратегия развития Арктической зоны Российской Федерации и обеспечения национальной безопасности на период до 2020 года, утвержденная Президентом Российской Федерации 20 февраля 2013 г. (№ Пр-232).

Специалисты РАН принимают активное участие в обосновании и создании многоцелевой космической системы «Арктика» для мониторинга обстановки в северных широтах и работают над созданием системы дистанционного (космического) мониторинга естественных и техногенных выходов углеводородов на поверхность акваторий России.

В перспективе было бы интересно проанализировать данные этого мониторинга для исследования экологической устойчивости предприятий АЗ.

В рамках сотрудничества между Центром исследования экономических проблем развития Арктики экономического факультета МГУ имени М. В. Ломоносова и Проектным офисом развития Арктики (ПОРА) был создан Полярный индекс Баренц-региона. В его рамках были созданы:

- 1) первый рейтинг устойчивого развития внутригосударственных территориальных единиц (провинций) Баренц-региона;
- 2) рейтинг устойчивого развития компаний, ведущих деятельность на территории Баренц-региона [Никоноров, Уткина, 2018].

В настоящей работе делается акцент на исследование возможностей устойчивого развития именно компаний Арктической зоны РФ.

Использование инновационных технологий в Арктике может принести широкие социальные и экономические выгоды как в Арктическом регионе, так и за его пределами. В работе [Никоноров, Уткина, 2018] использовали принцип светофора, позволивший проранжировать 26 анализируемых компаний Баренц-региона Арктики из РФ, Финляндии, Швеции, Норвегии по рейтингу устойчивого развития, рассматривая только интегральные показатели по оригинальной методике.

Современные компьютерные технологии и методы дают возможность анализировать многопараметрические базы данных (Big Data). В предлагаемой работе использованы данные отчетности компаний России, полученные из базы «Бюро-ван-Дайк» (формат Ruslana — подробный глобальный). Проведен сравнительный анализ экономических показателей предприятий Арктической зоны РФ. В группы сравнения вошли предприятия из топ-перечня с высоким рейтингом Полярного индекса и предприятия, в него не вошедшие.

Методами машинного обучения на обучающей выборке выявлен набор наиболее информативных показателей, так или иначе влияющих на прогноз того, в какой класс будет отнесено предприятие: в первый класс (высокий Полярный индекс) или во второй. В ходе исследований использованы различные методы машинного обучения, в том числе оригинальные логико-статистические методы (Data Science) [Кузнецова, Сенько, 2005], оценена достоверность найденных закономерностей, основанных на построении оптимальных достоверных разбиений признакового пространства [Senko et al., 2011]. Логико-статистический подход позволяет проводить анализ, не делая априорных предположений о виде вероятностных распределений, а также такой подход эффективен в условиях выборки любого размера и большого количества плохо структурированных признаков.

2. Материалы и методы

В исследование были включены числовые данные 193 предприятий АЗ РФ. Количество объектов первого класса: 42 — дочерних предприятия 17-ти компаний, вошедших в топ-перечень с высоким рейтингом Полярного индекса (см. [Никоноров, Уткина, 2018]). Количество объектов второго класса: 151 предприятие, не вошедшее в этот перечень. Число показателей в исходной обучающей выборке было 64. Показатели взяты из базы «Бюро-ван-Дайк» (формат Ruslana — подробный глобальный), любезно предоставленной Финансовому университету при Правительстве РФ. Финансовые показатели были взяты из разделов: «Баланс», «Отчет о прибылях и убытках», «Биржевая информация» — «Годовая оценка компании».

Для прогноза топ-рейтинга Полярного индекса по экономическим (финансовым) показателям использовали систему «Распознавание», включающую набор методов машинного обучения. Также были использованы хорошо зарекомендовавшие себя подходы: оригинальный метод распознавания — метод статистически взвешенных синдромов [Кузнецова, Сенько, 2005; Senko et al., 2011; Senko, Kuznetsova, 2010; Кузнецова и др., 2011; Kuznetsova et al., 2011; Kirilyuk et al., 2017; Кузнецова и др., 2018; Кузнецова и др., 2018]. Данный метод основан на принятии коллективных решений по областям оптимальных разбиений, найденных с помощью метода оптимально достоверных разбиений (ОДР) [Senko et al., 2011]. Прогностическая способность оценивалась с помощью метода скользящего контроля, дающего объективную несмещенную оценку точности. Под прогностической способностью в данном случае понималось правильное отнесение предприятия к классу из анализируемой выборки по совокупности экономических показателей.

Разбиения с максимальным значением функционала качества ищутся внутри нескольких семейств различного уровня сложности. В настоящем исследовании использовалось простейшее семейство всевозможных одномерных разбиений с одной граничной точкой, а также семейство всевозможных двумерных разбиений с одной граничной точкой по каждой из двух объясняющих переменных. Иными словами, последнее семейство может быть описано как множество двумерных разбиений с двумя границами, каждая из которых параллельна только одной из координатных осей. Оптимальное разбиение считается закономерностью только после положительного результата статистической верификации. Верификация производится с помощью перестановочных тестов, основанных на сравнении качества оптимальных разбиений на исходной анализируемой выборке с качеством разбиений на большом числе (2-3 тысячи) случайных выборок, полученных из исходной обучающей выборки с помощью случайных перестановок значений прогнозируемой величины относительно фиксированных позиций векторов прогностических переменных. Перестановочные тесты позволяют не только оценивать достоверность выявленных закономерностей, но также исключать из множества выходных закономерностей разбиения с избыточной сложностью.

3. Результаты и обсуждение

На первом этапе были проведены исследования по выявлению экономических показателей, наиболее информативных с точки зрения разделения двух сравниваемых классов. Использование метода ОДР на одномерных показателях выявило закономерности, связывающие номер класса с экономическими показателями.

3.1. Одномерные закономерности

В таблице 1 приведены в порядке убывания информативности экономические показатели, которые вносят наибольший вклад в разделение двух исследуемых классов предприятий, то есть информативны при распознавании. В первый класс вошли предприятия с высоким Полярным индексом, второй класс составляют предприятия, не вошедшие в топ рейтинга по Полярному индексу.

Таблица 1. Одномерные разбиения, полученные методом оптимальных достоверных разбиений

№*	Показатель	Граница	Ниже		Выше		F	p <
			1 класс**	2 класс	1 класс	2 класс		
1	Основные средства FIAS	14.28	12 (28.6 %)	0 (0 %)	30 (71.4 %)	151 (100 %)	45.76	0.0005
2	Совокупные активы	15.14	11 (26.2 %)	0 (0 %)	31 (73.8 %)	151 (100 %)	41.72	0.0005
3	Итого капитал и обязательства	15.14	11 (26.2 %)	0 (0 %)	31 (73.8 %)	151 (100 %)	41.72	0.0005
4	Выручка от реализации / оборот	18.09	25 (59.5 %)	145 (96 %)	17 (40.5 %)	6 (4 %)	41.5	0.0005
5	Себестоимость продукции	18.56	31 (73.8 %)	150 (99.3 %)	11 (26.2 %)	1 (0.7 %)	36.53	0.0005
6	Прочий собствен. капитал (включая резервы)	17.08	22 (52.4 %)	137 (90.7 %)	20 (47.6 %)	14 (9.3 %)	33.12	0.0005
7	Прочий собственный капитал	16.08	32 (76.2 %)	150 (99.3 %)	10 (23.8 %)	1 (0.7 %)	32.59	0.0005
8	Собственный капитал	17.23	20 (47.6 %)	133 (88.1 %)	22 (52.4 %)	18 (11.9 %)	32.57	0.0005
9	Прочие операционные расходы	15.98	28 (66.7 %)	145 (96 %)	14 (33.3 %)	6 (4 %)	30.34	0.0005
10	Нераспределенная прибыль	18.42	33 (78.6 %)	150 (99.3 %)	9 (21.4 %)	1 (0.7 %)	28.69	0.0005
11	Резервы предстоящих расходов	14.25	31 (73.8 %)	148 (98 %)	11 (26.2 %)	3 (2 %)	28.46	0.0005
12	Прочие текущие обязательства	14.87	30 (71.4 %)	146 (96.7 %)	12 (28.6 %)	5 (3.3 %)	25.97	0.0005
13	Резервы	13.57	32 (76.2 %)	148 (98 %)	10 (23.8 %)	3 (2 %)	24.78	0.0005
14	Кредиторы (поставщики и подрядчики)	16.88	29 (69 %)	144 (95.4 %)	13 (31 %)	7 (4.6 %)	24.37	0.0005
15	Материальные основные средства	17.68	24 (57.1 %)	135 (89.4 %)	18 (42.9 %)	16 (10.6 %)	23.44	0.0005
16	Проценты уплаченные	9.752	27 (64.3 %)	37 (24.5 %)	15 (35.7 %)	114 (75.5 %)	23.34	0.0005
17	Налоги	14.38	27 (64.3 %)	140 (92.7 %)	15 (35.7 %)	11 (7.3 %)	22.67	0.001
18	Валовая прибыль	16.25	22 (52.4 %)	130 (86.1 %)	20 (47.6 %)	21 (13.9 %)	22.21	0.000333
19	Управленческие расходы	15.83	36 (85.7 %)	151 (100 %)	6 (14.3 %)	0 (0 %)	22.15	0.0005
20	Прочие финансовые расходы	11.09	25 (69.4 %)	42 (28.6 %)	11 (30.6 %)	105 (71.4 %)	20.7	0.0005
21	Оборотные средства	17.18	26 (61.9 %)	136 (90.1 %)	16 (38.1 %)	15 (9.9 %)	19.23	0.000333
22	Ссуды	3.841	25 (59.5 %)	37 (24.5 %)	17 (40.5 %)	114 (75.5 %)	18.39	0.001

* Граница представлена в логарифмической шкале, то есть в таблице указан натуральный логарифм соответствующего показателя в тыс. руб.

** 1 класс — группа предприятий, вошедших в топ рейтинга Полярного индекса; 2 класс — группа предприятий, не вошедших в него.

Представлены закономерности, выявленные в рамках простейшей одномерной модели с одной граничной точкой [Senko, Kuznetsova, 2010], со значимостью не хуже чем $p < 0.001$. Для всех показателей, принимающих вещественные числовые значения, граница ставилась автоматически по базе данных с применением метода «хи-квадрат». Информативность и статистическая достоверность связи с прогнозируемой величиной (высокий или низкий уровень Полярного индекса) в обоих случаях оценивались с помощью метода ОДР.

F-коэффициент показывает степень достоверности закономерности. Он указывает, во сколько раз значение функционала, описывающего качество разбиения на исходной выборке, превышает максимальное значение функционала качества разбиений на случайных выборках (перестановочный тест). F-коэффициенты позволяют сравнить достоверность для двух закономерностей, в которых значение функционала качества на случайных выборках ни разу не превысило значение на реальных данных.

Как видно из таблицы 1, наиболее информативными стали следующие показатели: «Основные средства» — источник FIAS, «Совокупные активы», «Итого капитал и обязательства». Для них ниже границы находятся значения первого класса, и нет совсем значений второго класса. А вот для следующих показателей значения 1-го класса преобладают выше границы: «Выручка от реализации / оборот», «Себестоимость продукции», «Прочий собственный капитал» (включая резервы), «Прочий собственный капитал», «Собственный капитал», «Прочие операционные расходы», «Нераспределенная прибыль», «Резервы предстоящих расходов», «Прочие текущие обязательства», «Резервы», «Кредиторы» (поставщики и подрядчики), «Материальные основные средства» и др.

На диаграммах рассеяния (рис. 1, 2) наглядно представлены значения для предприятий 1 класса (красные крестики) и 2 класса (зеленые кружки) по соответствующим показателям. Границы разбиения поставлены таким образом, чтобы с одной стороны от границы преобладали объекты одного класса.

На рис. 1 для показателя «Совокупные активы» видно, что ниже границы, равной 15.14, преобладают объекты первого класса — 11 (26.2 %), объекты 2 класса отсутствуют. Выше границы, соответственно, преобладают объекты 2 класса — 151 объект (100 %) против 31 объекта 1 класса (73.8 %).

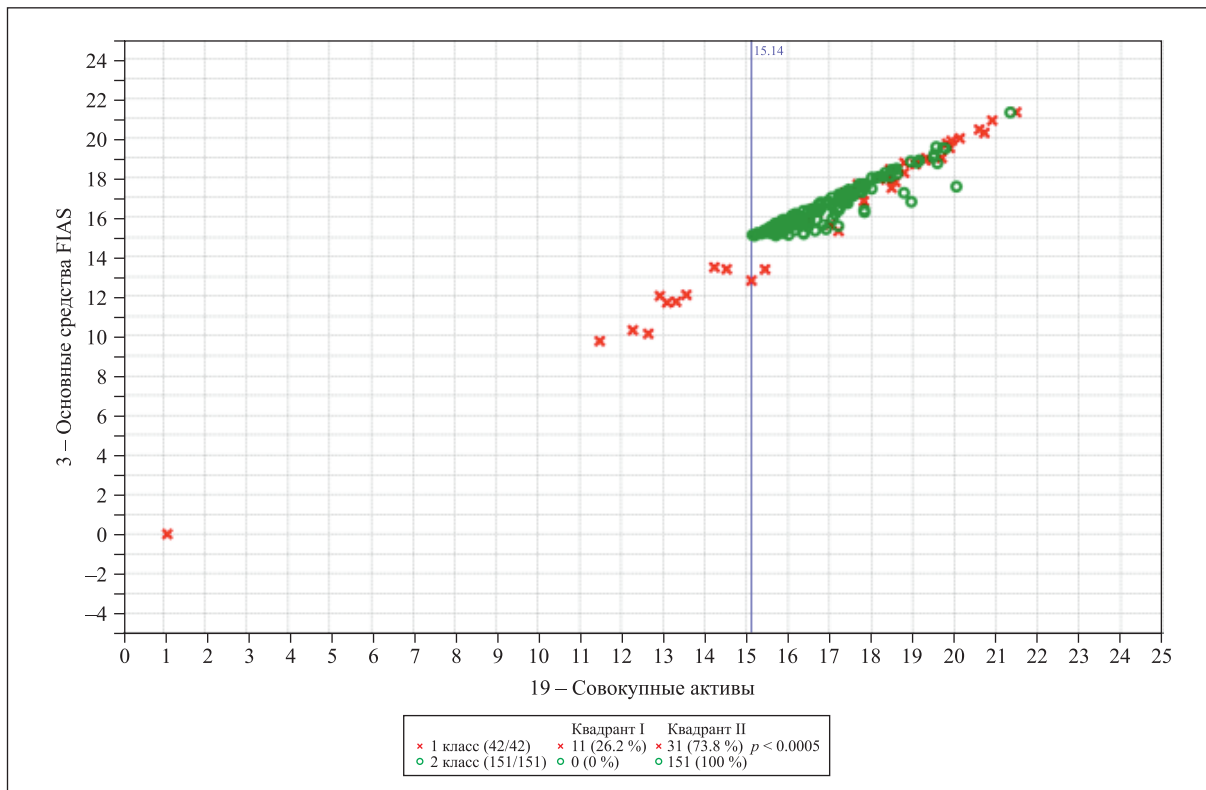


Рис. 1. По оси абсцисс отложен натуральный логарифм показателя «Совокупные активы», измеренного в тысячах рублей. По оси ординат — вспомогательный показатель. В первом квадранте (слева от границы) преобладают значения первого класса. Во втором квадранте (справа от границы) преобладают объекты второго класса

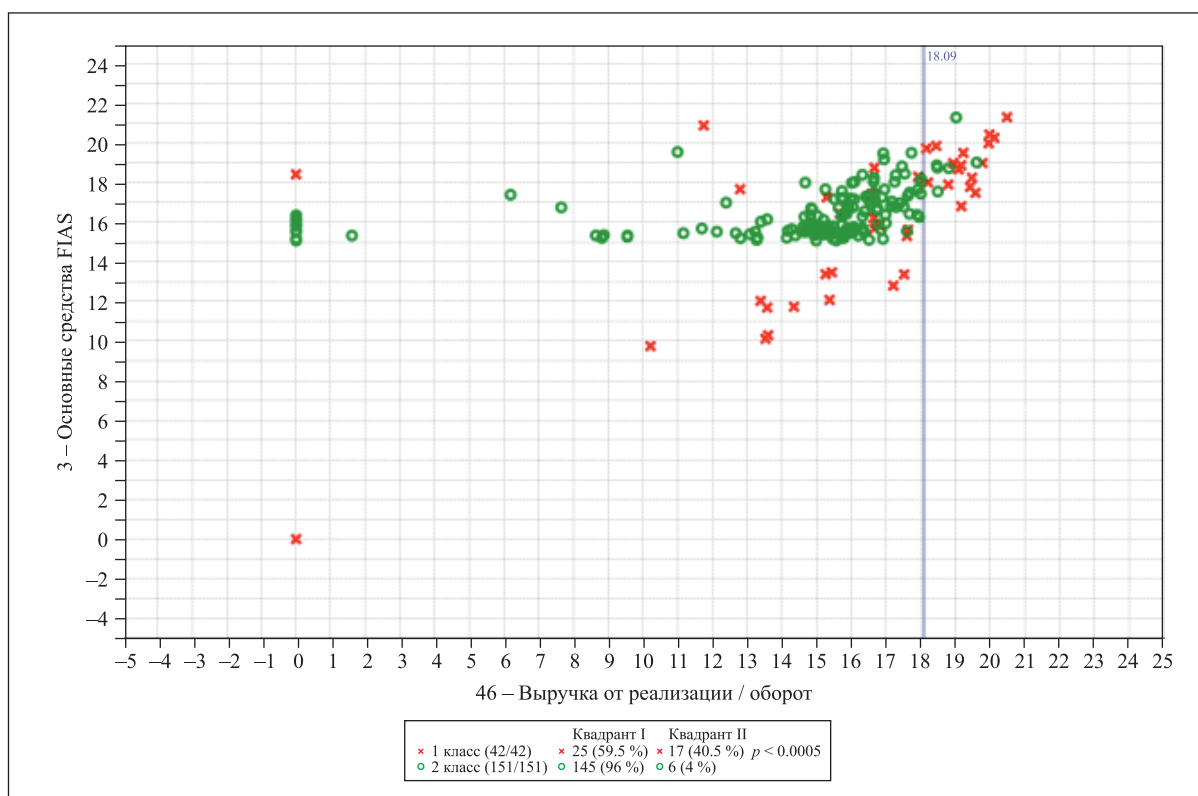


Рис. 2. По оси абсцисс отложен натуральный логарифм показателя «Выручка от реализации / оборот», выраженного в тыс. руб. Второй показатель вспомогательный. В первом квадранте преобладает второй класс (зеленые кружки). Во втором квадранте преобладает первый класс (красные крестики)

На рис. 2 видно, что слева от границы 18.09 преобладают объекты второго класса — 145 объектов (96 %). Выше границы преобладают 17 значений — 40.5 % — объектов класса высокого рейтинга Полярного индекса. В то время как значений второго класса только 4 %. Достоверность данной закономерности на уровне $p < 0.0005$ получена с помощью перестановочного теста.

3.2. Двумерные закономерности

На парах показателей выявлено намного больше достоверных разбиений. Как видно на диаграмме рассеяния (рис. 3), границы разбиений расположены параллельно осям координат, образуя 4 квадранта, пронумерованных по часовой стрелке. При этом верхний левый квадрант имеет номер I.

На рис. 3 можно видеть, что первый класс преобладает во втором и четвертом квадрантах. То есть шанс попасть в топ рейтинга Полярного индекса получают предприятия со значениями логарифма «Основных средств» (FIAS) ниже 14.277 и логарифма «Себестоимости продукции» выше 17.985.

3.3. Применение методов машинного обучения (методов распознавания)

Методы распознавания были использованы для исследования связи экономических показателей компаний с их позицией в рейтинге Полярного индекса (компаний). Для достоверной оценки диагностической способности использовался метод скользящего контроля. В результате выделялся набор методов, обладающих достаточной эффективностью (табл. 2). Структура алгоритма распознавания позволяет осуществлять диагностику в смысле отнесения предприятия

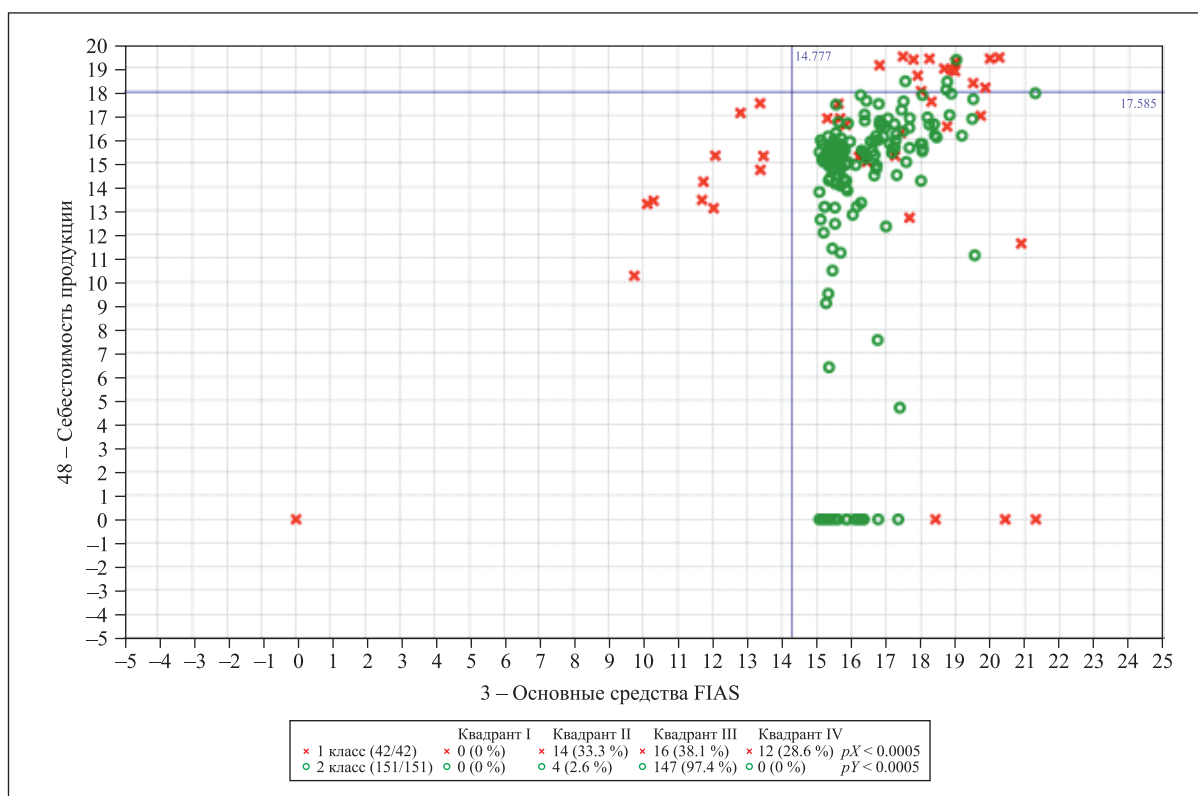


Рис. 3. Показатели: ось абсцисс — «Основные средства, источник FIAS», ось ординат — «Себестоимость продукции». Вдоль осей отложены натуральные логарифмы показателей, выраженных в тысячах рублей. В первом квадранте (левый верхний) нет значений. Во втором квадранте (правый верхний) преобладает первый класс. В третьем квадранте (правый нижний) преобладает первый класс. В четвертом квадранте (левый нижний) преобладает первый класс

к одной из групп [Кузнецова, Сенько, 2005; Senko et al., 2011; Senko, Kuznetsova, 2010; Кузнецова и др., 2011; Kuznetsova et al., 2011; Kirilyuk et al., 2017; Кузнецова и др., 2018; Кузнецова и др., 2018]. Точность алгоритма определяется по общему числу соответствия автоматического диагноза и реального номера класса.

В любом методе распознавание произвольного объекта может быть представлено в виде последовательного выполнения двух операций. На первом шаге вычисляются значения, отражающие меру близости объекта к одному из классов. Это значение называется «оценка за класс». На втором шаге производится собственно распознавание. Обычно используют некий порог d . Для многих методов машинного обучения оценки обычно вычисляются таким образом, что они принадлежат отрезку $[0, 1]$. Если «оценка за класс» объекта ниже порога d , то его относят к первому классу, а если оценка выше порога, то его относят ко второму классу. Все показатели эффективности, кроме ROC AUC, зависят от величины порога d . В таблице 2 представлены показатели эффективности для двух значений порога d : 0.4 и 0.5.

Из таблицы 2 видно, что наилучшим и наиболее сбалансированным является коллективное решение с максимальным значением по Ассигасу и F-мере, а также высокими значениями по другим показателям.

Как видно из таблицы 2, наилучший результат был получен при использовании метода «Решающий лес», особенно высока специфичность (распознавание второго класса) — 96.7 %. Метод статистически взвешенных синдромов (SWS) тоже дал хороший результат, метод опорных векторов (МОВ) лучший результат показал в категориях «точность» и «специфичность».

Таблица 2. Результаты методов машинного обучения и коллективное решение

Метод	Порог	Accuracy	Чувствительность	Специфичность	Precision	F-мера
Коллективное решение	0.4	0.886	0.619	0.96	0.813	0.702
МОВ	0.4	0.808	0.738	0.828	0.544	0.626
СВС	0.4	0.725	0.762	0.715	0.427	0.548
Решающий лес	0.4	0.87	0.619	0.94	0.743	0.676
Коллективное решение	0.5	0.87	0.429	0.993	0.948	0.590
МОВ	0.5	0.881	0.524	0.98	0.88	0.656
СВС	0.5	0.839	0.643	0.894	0.628	0.636
Решающий лес	0.5	0.876	0.548	0.967	0.821	0.658

Для получения коллективного решения вычислялась средняя величина оценок вероятности отнесения объектов к классам, рассчитанным различными методами. Коллективный метод компенсирует ошибки отдельных алгоритмов, входящих в коллектив. Теоретически доказано, что ошибка коллективного решения в задачах прогнозирования не превышает среднюю ошибку, вычисленную для методов по отдельности. Все это позволяет говорить о высокой надежности коллективного решения [Senko, Kuznetsova, 2010].

Таблица 3. Результаты ROC AUC для методов машинного обучения

Метод машинного обучения	Значение ROC AUC
Решающий лес (Decision forest)	0.895
МОВ (SVM)	0.872
СВС (SWS)	0.847
Метод ближайших соседей (KNN)	0.748

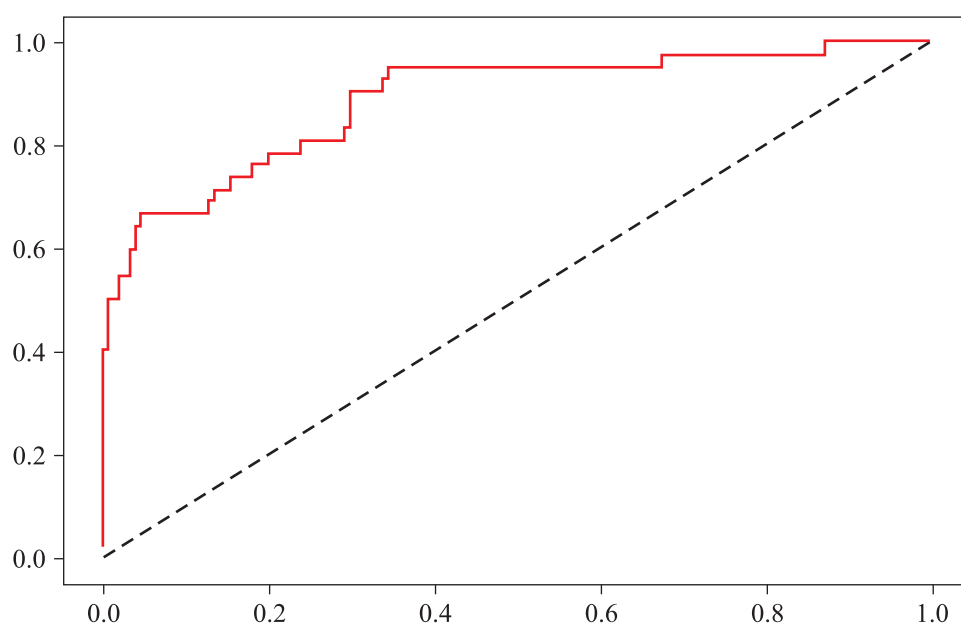


Рис. 4. ROC-кривая для коллективного решения, AUC = 0.888

ROC-кривая имеет тем более выпуклую форму, чем лучше результат распознавания, полученный с применением метода машинного обучения или коллектива методов МО (ансамбля). В данном случае имеем высокий результат распознавания. Площадь под кривой AUC приближается к 0.9.

Обучающая выборка оказалась достаточно информативной для того, чтобы отделить один исследуемый класс от другого.

По результатам распознавания второго класса можно сделать вывод о предприятиях, которые были ошибочно распознаны как 1 класс: у них большой потенциал занять высокую позицию в рейтинге Полярного индекса.

Вместе с тем для какой-то части компаний, которую эксперты относят к высокому Полярному индексу (1 класс), при распознавании получена ошибка. Последнее свидетельствует об определенных отклонениях показателей отчетности таких компаний от типичных значений для предприятий топ рейтинга Полярного индекса. Например, из рис. 5 видно, что величинам логарифма нераспределенной прибыли выше 15.14 (квадрант II) соответствуют 9 объектов первого класса и только один — второго. Вклад признака «Нераспределенная прибыль» для всех объектов из квадранта II будет соответствовать отнесению в первый класс. Для единственного из предприятий из второго класса в квадранте II целесообразно проведение дополнительного анализа причин, по которым оно не попало в топ рейтинга.

Метод статистически взвешенных синдромов основан на голосовании по базовым множествам, определяемым границами разбиений, полученным методом ОДР. При попадании значения объекта одного класса в базовое множество с преобладанием значений объектов другого класса велика вероятность того, что он будет отнесен в другой класс. Отсюда и возникают ошибки распознавания в данном методе.

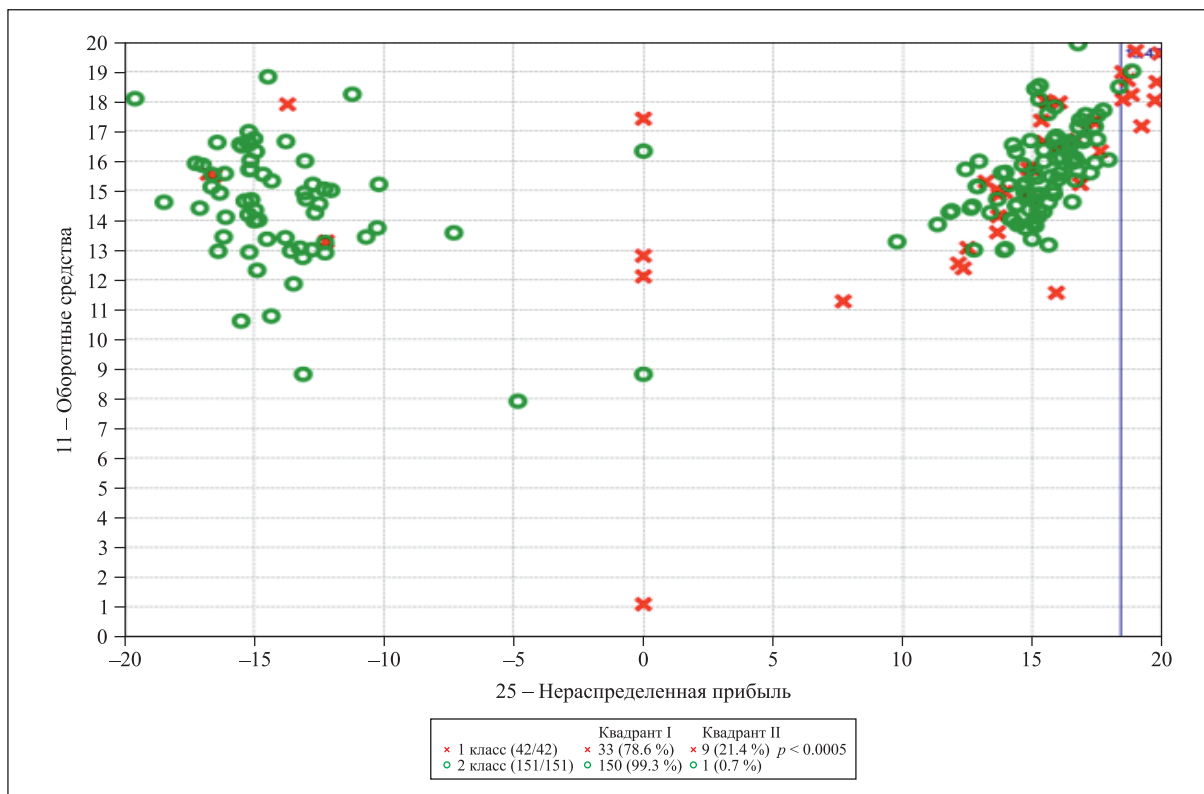


Рис. 5. Пример возможной ошибки при распознавании. По показателю «Нераспределенная прибыль» слева от границы, равной 18.42, 33 объекта 1 класса (x) находятся среди 150 объектов 2 класса (o). Справа от границы одно предприятие 2 класса находится среди 9 объектов 1 класса. В этом случае существует большая вероятность, что объект 2 класса будет ошибочно отнесен алгоритмом распознавания в 1 класс

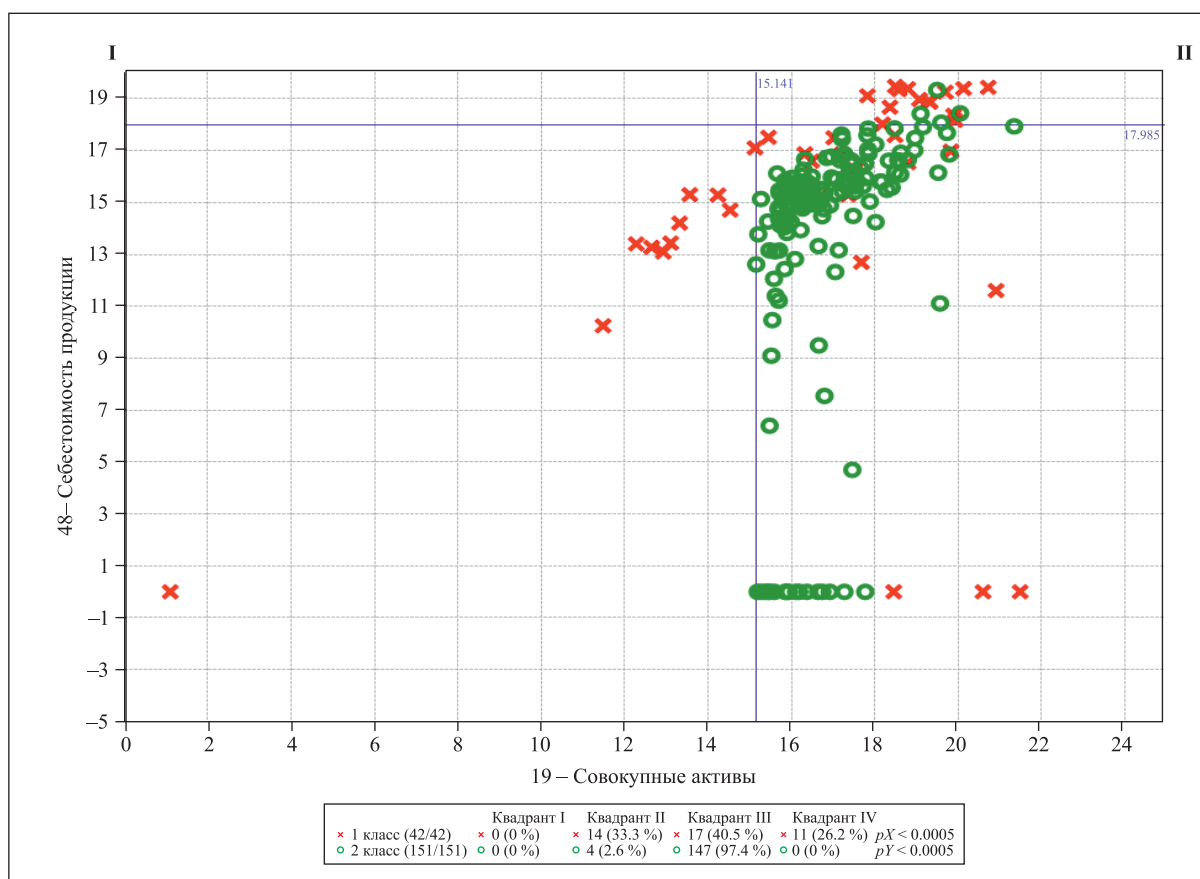


Рис. 6. Двумерная диаграмма рассеяния. Ось X — логарифмы совокупных активов, ось Y — логарифмы себестоимости продукции

На рис. 6 представлена двумерная диаграмма рассеяния. Можно выделить относительно небольшое число компаний, не вошедших в топ рейтинга (2 класс), для которых финансовые показатели соответствуют первому классу. На рис. 6 во втором квадранте с высокими значениями совокупных активов и себестоимости продукции преобладают значения предприятий 1 класса (высокий рейтинг Полярного индекса) — 14 объектов. Но в то же время там присутствуют 4 предприятия из 2 класса. В то же время в третьем квадранте среди большинства значений второго класса (147, 97.4 %) присутствуют 17 объектов первого класса.

Для таких компаний можно попытаться провести дополнительный анализ факторов, отличных от использовавшихся показателей финансовой отчетности и определивших низкую оценку по рейтингу Полярного индекса. Коррекция таких факторов потенциально позволила бы повысить эффективность деятельности компаний с точки зрения ее устойчивости.

4. Выводы

Наше исследование показало, что предприятия, вошедшие в топ рейтинга Полярного индекса, в целом распознаются по финансовым показателям среди компаний Арктической зоны. Для коллективного решения AUC = 0.888. Чувствительность — 62 %. Специфичность — 96 %.

Таким образом, полученное решающее правило может быть использовано в качестве прогностического алгоритма для оценки потенциала компании с точки зрения рейтинга Полярного индекса (компаний).

Ввиду того, что пилотная версия программы оценки устойчивости предприятия по экономическим критериям показала свою эффективность, можно рекомендовать добавить в базу

данных числовую информацию по социальным и экологическим критериям. Это позволит создать решающее правило для прогноза устойчивого развития предприятия по всем трем критериям Полярного индекса: экономическим, социальным и экологическим.

5. Благодарности

Большое спасибо АНО «ЭКСПЕРТНЫЙ ЦЕНТР ПОРА» за поддержку данного исследовательского Проекта в виде гранта №153-Г.

Выражаем благодарность младшему научному сотруднику Института экономики РАН Игорю Леонидовичу Кирилкову за помощь при обсуждении материалов статьи.

Список литературы (References)

- Додин Д. А.* Устойчивое развитие Арктики: проблемы и перспективы. — СПб.: Наука, 2005. — 283 с.
Dodin D. A. Ustoichivoye razvitiye 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 universiteta 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 informatsionnyye 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](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). — Available at: URL: [http://www.matbio.org/2011/Senko2011\(6_115\).pdf](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 tehnologii. Iskusstvennyy intellekt v zdravookhraneni. — 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. Prozrachnyy interfeys dlya prognoza v mashinnom obucheni [A transparent interface for forecasting in machine learning] // M. MGU. Sbornik nauchnykh trudov XIX Mezhdunarodnoy konferentsii DAMDID/RCDL. — 2017. — P. 493–495 (in Russian).

Никонов 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.

Nikonov 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. Никонов. — 2018.

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