

Научная статья**УДК 332.1****DOI: 10.25683/VOLBI.2025.70.1182****Anna Viktorovna Kuznetsova**

Candidate of Biology,
Senior Researcher at the Laboratory of Mathematical Biophysics,
Institute of Biochemical Physics
of the Russian Academy of Sciences
Moscow, Russian Federation
azforus@yandex.ru

Lyudmila Robertovna Borisova

Candidate of Physics and Mathematics,
Associate Professor of the Department of Mathematics
and Data Analysis of the Faculty of Information Technology
and Big Data Analysis,
Financial University under the Government of the Russian Federation
Moscow, Russian Federation
lrborisova@fa.ru

Naum Sevelevitch Kremer

Candidate of Economics,
Associate Professor of the Department of Mathematics
and Data Analysis of the Faculty of Information Technology
and Big Data Analysis,
Financial University under the Government of the Russian Federation
Moscow, Russian Federation
NSKremer@fa.ru

Mira Nisonovna Fridman

Associate Professor of the Department of Mathematics
and Data Analysis of the Faculty of Information Technology
and Big Data Analysis,
Financial University under the Government of the Russian Federation
Moscow, Russian Federation
MNFridman@fa.ru

Анна Викторовна Кузнецова

канд. биол. наук,
старший научный сотрудник
лаборатории математической биофизики,
Институт биохимической физики Российской академии наук
Москва, Российская Федерация
azforus@yandex.ru

Людмила Робертовна Борисова

канд. физ.-мат. наук, доцент,
доцент кафедры математики и анализа данных факультета
информационных технологий и анализа больших данных,
Финансовый университет
при Правительстве Российской Федерации
Москва, Российская Федерация
lrborisova@fa.ru

Наум Шевелевич Кремер

канд. экон. наук, доцент,
доцент кафедры математики и анализа данных факультета
информационных технологий и анализа больших данных,
Финансовый университет
при Правительстве Российской Федерации
Москва, Российская Федерация
NSKremer@fa.ru

Мира Нисоновна Фридман

доцент кафедры математики и анализа данных факультета
информационных технологий и анализа больших данных,
Финансовый университет
при Правительстве Российской Федерации
Москва, Российская Федерация
MNFridman@fa.ru

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ДОТАЦИОННЫХ РЕГИОНОВ РОССИЙСКОЙ ФЕДЕРАЦИИ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ ПО ШИРОКОМУ НАБОРУ ПОКАЗАТЕЛЕЙ ОСНОВНЫХ ФОНДОВ

5.2.3 — Региональная и отраслевая экономика

Аннотация. В работе представлен оригинальный метод поиска связи дотационности регионов Российской Федерации с факторами основных фондов, предоставленных отчетами Росстата. В рамках предложенного метода проведен анализ показателей за 2021 г. выделен набор наиболее значимых показателей основных фондов, по которому группа дотационных регионов Российской Федерации отличается от субъектов федерации без дотаций. Методами машинного обучения в сравниваемых группах выявлены следующие различия: более низкие значения в дотационных регионах по таким показателям, как стоимость основных фондов на конец года по полной учетной стоимости (млн руб.), водоснабжение, водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений, водоснабжение; водоотведение, организация сбора и утилизации

отходов, ввод в действие предприятий с деятельностью по ликвидации загрязнений, обрабатывающие производства, ввод в действие основных фондов по видам экономической деятельности, торговля оптовая и розничная. В то же время более высокие значения в дотационных регионах по структуре деятельности наблюдаются в области информации и связи, предприятиях в сфере сельского, лесного хозяйства, охоты, рыболовства и рыбоводства. Данный подход в анализе данных ранее никем не использовался. Найденные закономерности позволяют разработать план мероприятий по выводу регионов из зоны дотационности.

Ключевые слова: методы машинного обучения, статистика, дотационные регионы, основные средства, коэффициент корреляции, кластерный анализ, математическое моделирование, признаки, разбиения, дотации

Благодарности: авторы приносят искреннюю благодарность директору Государственного предприятия ЦНИИтэтиязмаш канд. экон. наук В. М. Хадарцеву за постановку задачи и плодотворные обсуждения полученных результатов.

© Кузнецова А. В., Борисова Л. Р., Кремер Н. Ш., Фридман М. Н., 2025

© Kuznetsova A. V., Borisova L. R., Kremer N. S., Fridman M. N., 2025

Для цитирования: Кузнецова А. В., Борисова Л. Р., Кремер Н. Ш., Фридман М. Н. Сравнительный анализ дотационных регионов Российской Федерации методами машинного обучения по широкому набору показателей основных фондов // Бизнес. Образование. Право. 2025. № 1(70). С. 20—28. DOI: 10.25683/VOLBI.2025.70.1182.

Original article

COMPARATIVE ANALYSIS OF SUBSIDIZED REGIONS OF THE RUSSIAN FEDERATION USING MACHINE LEARNING METHODS FOR A WIDE RANGE OF INDICATORS OF FIXED ASSETS

5.2.3 — Regional and sectoral economy

Abstract. *The paper presents an original method for finding a connection between the subsidization of the regions of the Russian Federation and the factors of fixed assets provided by Rosstat reports. Within the framework of the proposed method, an analysis of indicators for 2021 was carried out. A set of the most significant indicators of fixed assets was identified, according to which the group of subsidized regions of the Russian Federation differs from the subjects of the Federation without subsidies. The following differences were revealed by machine learning methods in the compared groups: lower values in subsidized regions for such indicators as the cost of fixed assets at the end of the year at full book value (million rubles), water supply, wastewater disposal, organization of waste collection and disposal, pollution elimination activities, water*

supply; wastewater disposal, organization of waste collection and disposal, commissioning of enterprises with pollution elimination activities, manufacturing, commissioning of fixed assets by type of economic activity, wholesale and retail trade. At the same time, higher values in subsidized regions in terms of activity structure are observed in the field of information and communications, enterprises in the field of agriculture, forestry, hunting, fishing and fish farming. This approach has not been used by anyone in data analysis before. The found patterns will allow us to develop an action plan for the withdrawal of regions from the subsidy zone.

Keywords: *machine learning methods, statistics, subsidized regions, fixed assets, correlation coefficient, cluster analysis, mathematical modeling, features, partitions, subsidies*

Acknowledgements: We sincerely thank the director of the State Enterprise Tsniiteityazhmash, Candidate of Economics, V. M. Khadartsev for setting the task and fruitful discussions of the results obtained.

For citation: Kuznetsova A. V., Borisova L. R., Kremer N. S., Fridman M. N. Comparative analysis of subsidized regions of the Russian Federation using machine learning methods for a wide range of indicators of fixed assets. *Biznes. Obrazovanie. Pravo = Business. Education. Law.* 2025;1(70):20—28. DOI: 10.25683/VOLBI.2025.70.1182.

Введение

Актуальность. В настоящий момент существует заметная дифференциация регионального развития в стране, многие регионы обладают низкой самообеспеченностью доходов ввиду недостаточного и неэффективного использования экономического, управленческого, территориального, инвестиционного потенциала. Ввиду этого возникают диспропорции в региональных бюджетах, характеризующиеся дефицитом консолидированного бюджета, в структуре которого значимую часть доходных поступлений составляют дотации. Вследствие этого возникает необходимость изучения проблемы распределения дотаций субъектам Российской Федерации, а также разработки подходов к группировке и кластеризации дотационных регионов с выявлением их ключевых особенностей, учитывая факторы и условия, оказывающие прямое и косвенное воздействие на увеличение или длительное сохранение дотационности региональных бюджетов России, что и обуславливает актуальность данного исследования. Можно без преувеличения сказать, что методы машинного обучения сегодня используются везде. И разумеется важным является использование их в области государственного управления и планирования.

Изученность проблемы и целесообразность. Основные средства — это основа любого производства. Основные средства оцениваются по стоимости их приобретения за вычетом накопленной амортизации. Исследования обеспеченности предприятий основными фондами, динамику основных средств принято проводить с использованием статистического анализа [1]. Предполагается производить расчеты по предложенной Т. А. Журкиной и Т. В. Сабел-

товой [1] формуле общего риска через определенные промежутки времени, чтобы следить за динамикой изменения этого показателя, и в зависимости от направления изменения вносить коррективы в экономическую деятельность с целью уменьшения общего риска.

Есть немало литературы, в которой предлагаются простые способы экономии основных средств. В частности, можно отметить книгу R. Hänggi, A. Fimpel, R. Siegenthaler [2], в которой систематизированы проекты по бережливому производству за последние 20 лет во многих странах. Что касается вопросов, связанных с математическим моделированием оптимизации производственных затрат разных предприятий в различных областях деятельности, то прежде всего используются методы линейного, нелинейного и стохастического программирования, как отмечено в научной статье И. В. Бачериковым с соавторами [3]. Ш. Рахимовым с соавторами [4] на основе критериев и методов оптимального управления, математических моделей и алгоритмов обеспечения экономии водных ресурсов представлена разработка оптимального управления водными ресурсами крупных магистральных каналов с каскадами насосных станций.

В обзорной работе М. В. Мильчакова [5] подчеркивается важность использования экономического потенциала территорий.

В последние годы появилось много работ, которые используют методы машинного обучения для экономических прогнозов. Среди таких работ стоит отметить статью М. Гареева [6], в которой представлена модель ежегодных прогнозов темпов роста валового накопления основного капитала в России.

О. Чистик с соавторами [7] предложен информационно-методический подход — авторский вариант формирования блоков факторных показателей регионального развития по содержательному критерию: факторы экономического развития региона; демографические факторы; социальные факторы; структурные и производственные факторы; факторы цифрового развития; финансовые факторы. На основе применения метода кластерного анализа были выделены шесть качественно однородных групп регионов России по факторным показателям развития инвестиций в основной капитал. Выявлены кластеры с различными условиями формирования: «очень благоприятный», «благоприятный», «хороший», «средний», «плохой», «неблагоприятный». Выявлена географическая особенность: в первых двух кластерах представлены города федерального значения и регионы европейской части России; в последних двух — регионы Крайнего Севера, Дальний Восток и Северный Кавказ; реализован общедоказательный подход к анализу динамики показателей инвестиций в основной капитал, а также их прогнозированию. Анализ инвестиционных тенденций показал их взаимосвязь с экономическими циклами. В годы кризиса инвестиционная активность в России замедлилась, чтобы впоследствии усилиться, но рост активности со временем стал менее существенным. Однако в этой работе не исследована связь основных средств и дотационности регионов, которые авторами названы как «неблагоприятные».

J. Svoboda, J. Lososová, R. Zdeněk [8] сравнили инвестиционные субсидии в сельском хозяйстве в странах — членах Европейского Союза за период с 2004 по 2013 г. на основе базы данных *FADN*. Низкий уровень инвестиций влияет на стоимость и эффективность сельскохозяйственного производства и, следовательно, на общую конкурентоспособность сельскохозяйственной продукции. Европейские программы поддержки инвестиций в сельское хозяйство направлены на повышение конкурентоспособности сельского хозяйства. Развитие субсидий на инвестиции, недвижимост и чистый доход фермерских хозяйств, скорректированный с учетом размера экономики предприятия с помощью корреляционного анализа, сравниваются в каждой стране Европейского Союза.

J. A. Pérez-Méndez, M. Pérez-Urdiales, D. Roibas [9] для оценки эффекта субсидий, предоставляемых в лесохозяйственных фирмах, использовали стохастический подход. Представлена формула, по которой вычисляется уровень выпуска продукции в разные моменты времени в зависимости от переменных показателей. Исследуется эффективность этих показателей. Получен результат, что положительное влияние на эффективность в лесохозяйственных фирмах оказывают возраст персонала и объем продаж за пределами региона.

В последних двух работах решалась задача исследования роли субсидий в деятельности небольших предприятий, таких как сельскохозяйственные и лесохозяйственные хозяйства.

Стоит отметить, что не было ни одной работы по классификации регионов по основным фондам с использованием методов математического моделирования, поэтому проблема применения математического моделирования для анализа региональной экономики остается актуальной.

Проблема изучения связи дотационности регионов с основными средствами является актуальной. В настоящее время не существует единой методики изучения дотационности региона в зависимости от основных средств, от инвестиций.

Все изученные нами литературные источники с математическими моделями не позволяют дать прогноз и рекомендации по выводу региона из дотационного кластера, а представленный в работе подход позволяет это сделать. В этом состоит **научная новизна** выполненного исследования. Алгоритмы машинного обучения и статистические методы позволяют выявить взаимосвязи и зависимости между различными параметрами, определить ключевые факторы, влияющие на уровень дотации в каждом конкретном регионе. На основе этих данных можно разработать модели оценки и прогнозирования дотаций, которые будут учитывать не только текущую ситуацию, но и потенциальные изменения и тенденции развития. Благодаря этому подходу можно повысить эффективность управления бюджетными ресурсами и сделать процесс принятия решений более обоснованным и прозрачным.

В сборниках Росстата представлены данные по учетной стоимости основных фондов по видам экономической деятельности. Эти величины были использованы для классификации регионов.

Используемый в работе подход (классификация методами оптимально достоверных разбиений и статистически взвешенных синдромов) только начинает использоваться в данной области. Найденные закономерности позволяют более точно обрисовать портрет каждого региона Российской Федерации с возможностью дальнейшего прогнозирования его дотационного статуса. Набор значимых характеристик позволит повысить точность прогноза и предложить план по выходу из дотационной группы в группу самодостаточных субъектов Российской Федерации.

Целью исследования было с помощью методов машинного обучения выявить ключевые факторы, коррелирующие с экономическим развитием региона, и определить, какие из них связаны с увеличением или уменьшением потребности в дотациях. Такой подход позволяет более глубоко и объективно изучить проблему дотационности регионов, выявить скрытые закономерности и оптимизировать процессы принятия решений в этой области.

Теоретическую основу составили научные статьи за последние годы, посвященные математическому моделированию региональной экономики, использованию методов базовой статистики и машинного обучения с использованием, в том числе, оригинальных методов.

Теоретическая значимость заключается в применении методов машинного обучения для анализа малых выборок (не более ста элементов) с целью возможности их использования для составления планов регионального развития.

Практическая значимость. Методы оптимально достоверных разбиений и статистически взвешенных синдромов позволяют создать план перевода объекта из группы с неблагоприятным прогнозом в благоприятную группу, показывая, на какие ключевые показатели надо в первую очередь обращать внимание и каких границ требуется достичь для того, чтобы увеличить вероятность благоприятного прогноза.

Для каждого региона из контрольной группы, которые были распознаны как дотационные, можно дать четкий план мероприятий по переводу его в благоприятную группу в виде набора 20 значимых показателей с границами разбиения, которые нужно достигнуть и перейти. После достижения этой цели вероятность того, что регион сможет отказаться от дотаций из центра станет вполне реальной. Это выведет практику работы с дотационными регионами на по-настоящему научный уровень.

Основная часть

Методология. Всего в базу данных вошли 66 показателей, общее количество объектов — 47, количество объектов первой группы (дотационные регионы) — 22, количество объектов второй группы (регионы, обходящиеся без дотаций) — 25. Всего было выбрано 47 регионов, а не 85, как в работе [10], исходя из проведенного анализа используемых данных Росстата, т. к. данные были представлены не по всем регионам.

Были использованы хорошо зарекомендовавшие себя методы машинного обучения, подробно описанные в работах [11—14].

Данная работа является продолжением предыдущей нашей работы [15]. Ранее нами было отмечено, что направления дальнейших исследований предполагаются следующими: поскольку предложенный математический подход показал эффективность в поиске ключевых факторов, можно применить аналогичную технологию для анализа более полной базы данных, содержащей показатели промышленного сектора, предприятия различных сфер деятельности, образовательные учреждения, социальные службы и объекты, привлекательные для инвестиций. Особенный интерес представляют показатели, отражающие основные фонды регионов, поскольку они оказались одними из самых значимых показателей.

Режим автоклассификации представляет собой набор 9 методов машинного обучения, которые могут быть объединены в ансамбли из трех, пяти, семи или девяти наиболее эффективных методов. В данной работе использовали 7 методов из-за малого количества объектов. Нейронные сети неэффективны в подобном случае, поскольку требуют для обучения большое число объектов — десятки и сотни тысяч. Валидация работы методов машинного

обучения проводилась в режиме скользящего контроля *Leave-one-Out*. Метод оптимально достоверных разбиений вычисляет границу между объектами групп сравнения таким образом, чтобы с обеих сторон от границы разбиения преобладали объекты разных групп. Аналогично ставятся границы для двумерного разбиения в пространстве признаков. Значимость найденных закономерностей вычисляется с помощью перестановочного теста, в котором происходит сравнение распознавания исходной обучающей выборки с результатами распознавания на нескольких тысячах случайных таблиц. Малое число верных распознаваний на случайных таблицах говорит о том, что закономерность, найденная на обучающей выборке достоверная.

Второй авторский метод машинного обучения — это метод статистически взвешенных синдромов, в котором проводится голосование по полученным базовым множествам с дальнейшим отнесением объекта в одну из сравниваемых групп.

Результаты. Основные выводы из использования методов базовой статистики. В табл. 1 приведены значимые показатели в порядке возрастания *p-value* и убывания значимости показателя Хи-квадрат (*F*) — правый столбец. Граница, выставленная методом ОДР [15], и число значений каждой группы (с процентным содержанием) слева от границы — 1-й квадрант, и число значений справа от границы — 2-й квадрант. Все показатели — статистически значимые, т. к. величина *p-value* близка к нулю, за исключением показателей: строительство; сельское, лесное хозяйство, рыболовство и рыбоводство. Это связано прежде всего с тем, что регионы России очень гетерогенны по этим показателям. Отметим, что названия показателей соответствуют их названиям на сайте Росстата.

Таблица 1

Одномерные разбиения

№	Показатель	Граница	1 квадрант		2 квадрант		F	p-value
			1-я группа	2-я группа	1-я группа	2-я группа		
1	Обрабатывающие производства	236689	20 (90,9 %)	3 (12 %)	2 (9,1 %)	22 (88 %)	28,54	0,0005
2	Ввод в действие основных фондов по видам экономической деятельности	118080	17 (77,3 %)	1 (4 %)	5 (22,7 %)	24 (96 %)	26,02	0,0005
3	Обрабатывающие производства — ввод в действие	19796	20 (90,9 %)	4 (16 %)	2 (9,1 %)	21 (84 %)	25,72	0,0005
4	стоимость основных фондов на конец года; по полной учетной стоимости; миллионов рублей	2172619	16 (72,7 %)	1 (4 %)	6 (27,3 %)	24 (96 %)	23,43	0,0005
5	Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	17688	19 (86,4 %)	5 (20 %)	3 (13,6 %)	20 (80 %)	20,19	0,0005
6	Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений — ввод в действие	1958	21 (95,5 %)	8 (32 %)	1 (4,5 %)	17 (68 %)	19,52	0,0005
7	Обеспечение электрической энергией, газом и паром; кондиционирование воздуха — ввод в действие	4917	14 (63,6 %)	1 (4 %)	8 (36,4 %)	24 (96 %)	18,75	0,0005
8	Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов	25998	15 (68,2 %)	2 (8 %)	7 (31,8 %)	23 (92 %)	17,97	0,0005
9	Строительство — ввод в действие	1488	13 (59,1 %)	1 (4 %)	9 (40,9 %)	24 (96 %)	16,62	0,0005
10	Транспортировка и хранение — ввод в действие	15438	18 (81,8 %)	6 (24 %)	4 (18,2 %)	19 (76 %)	15,32	0,001

Окончание табл. 1

№	Показатель	Граница	1 квадрант		2 квадрант		F	p-value
			1-я группа	2-я группа	1-я группа	2-я группа		
11	Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов — ввод в действие	2783	15 (68,2 %)	3 (12 %)	7 (31,8 %)	22 (88 %)	15,3	0,001
12	Транспортировка и хранение	340146	16 (72,7 %)	4 (16 %)	6 (27,3 %)	21 (84 %)	15,08	0,001
13	Деятельность в области информации и связи	26661	12 (54,5 %)	1 (4 %)	10 (45,5 %)	24 (96 %)	14,62	0,001
14	Обрабатывающие производства — структура	8,25	20 (90,9 %)	9 (36 %)	2 (9,1 %)	16 (64 %)	14,61	0,001
15	Добыча полезных ископаемых	96208	19 (86,4 %)	8 (32 %)	3 (13,6 %)	17 (68 %)	13,85	0,002
16	Обеспечение электрической энергией, газом и паром; кондиционирование воздуха	190850	17 (77,3 %)	6 (24 %)	5 (22,7 %)	19 (76 %)	13,01	0,002
17	Строительство	21805	15 (68,2 %)	4 (16 %)	7 (31,8 %)	21 (84 %)	12,95	0,004
18	Деятельность в области информации и связи — ввод в действие	4744	20 (90,9 %)	10 (40 %)	2 (9,1 %)	15 (60 %)	12,86	0,004
19	Деятельность в области информации и связи — структура	1,15	2 (9,1 %)	14 (56 %)	20 (90,9 %)	11 (44 %)	11,22	0,004
20	Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	13,7	14 (63,6 %)	25 (100 %)	8 (36,4 %)	0 (0 %)	10,72	0,008
21	Сельское, лесное хозяйство, охота, рыболовство и рыбоводство — структура	2,55	6 (27,3 %)	19 (76 %)	16 (72,7 %)	6 (24 %)	10,92	0,008
22	Обеспечение электрической энергией, газом и паром; кондиционирование воздуха — структура	5,85	6 (27,3 %)	19 (76 %)	16 (72,7 %)	6 (24 %)	10,92	0,009
23	Добыча полезных ископаемых — ввод в действие	17166	19 (86,4 %)	10 (40 %)	3 (13,6 %)	15 (60 %)	10,42	0,011
24	Деятельность в области информации и связи	1,85	5 (22,7 %)	17 (68 %)	17 (77,3 %)	8 (32 %)	9,428	0,012
25	Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	44,8	6 (28,6 %)	18 (72 %)	15 (71,4 %)	7 (28 %)	8,439	0,026
26	Сельское, лесное хозяйство, охота, рыболовство и рыбоводство — ввод в действие	6655	13 (59,1 %)	4 (16 %)	9 (40,9 %)	21 (84 %)	9,211	0,027
27	Строительство — структура	1,35	14 (63,6 %)	24 (96 %)	8 (36,4 %)	1 (4 %)	7,748	0,03
28	Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	6,2	10 (45,5 %)	2 (8 %)	12 (54,5 %)	23 (92 %)	8,45	0,034
29	Строительство	24,85	11 (50 %)	22 (88 %)	11 (50 %)	3 (12 %)	7,908	0,045

В табл. 2—5 представлены результаты качества распознавания, проведенного с использованием различных методов и комбинации части методов в ансамбли с целью выбора самого лучшего метода для распознавания исследуемых данных.

Разные методы на разных массивах дают разное качество распознавания. Выбрали не просто лучший метод для дальнейшего анализа, а именно ансамбль из методов, что-

бы в дальнейшем получить самое высокое качество распознавания данных и классификацию регионов. Качество классификации характеризуется величиной безразмерной площади под ROC-кривой. Чем ближе эта величина к 1, тем лучше распознавание.

Наилучший результат распознавания показал ансамбль из пяти лучших методов с $AUC = 0,9036$.

Таблица 2

Результаты методов машинного обучения, участвовавших в автоклассификации

Методы	Правильность	Точность	Чувствительность	Специфичность	F-Оценка	AUC
Градиентный бустинг	0,851	0,826	0,864	0,840	0,844	0,892
Статистически взвешенные синдромы	0,830	0,850	0,773	0,880	0,810	0,864
Метод опорных векторов	0,851	0,941	0,727	0,960	0,821	0,838
Метод ближайших соседей	0,809	0,783	0,818	0,800	0,800	0,809
Деревья решений	0,766	0,720	0,818	0,720	0,766	0,769
Линейный дискриминантный анализ	0,468	0,468	1,000	0,000	0,638	0,500
Адаптивный бустинг	0,532	1,000	0,000	1,000	0,000	0,500

Таблица 3

Сводная таблица ансамблей

Ансамбль	Правильность	Точность	Чувствительность	Специфичность	F-Оценка	AUC
Ансамбль 3	0,8298	0,8500	0,7727	0,8800	0,8095	0,8900
Ансамбль 5	0,8511	0,8571	0,8182	0,8800	0,8372	0,9036
Ансамбль 7	0,8511	0,8571	0,8182	0,8800	0,8372	0,9036

Таблица 4

Сводная таблица результата для Ансамбля 5

Параметр	Объектов	Правильно	Ошибочно	Из группы 1	Из группы 2
1-я группа	22 (46,8 %)	18 (81,8 %)	4 (18,2 %)	18 (85,7 %)	4 (15,4 %)
2-я группа	25 (53,2 %)	22 (88,0 %)	3 (12,0 %)	3 (14,3 %)	22 (84,6 %)
Итого	47 (100,0 %)	40 (85,1 %)	7 (14,9 %)	21 (44,7 %)	26 (55,3 %)

Таблица 5

Результат автоклассификации для Ансамбля 5

Объекты	KNN	SVM	DT	SWS	XGB	Ансамбль	Верная группа
Республика Северная Осетия – Алания	1	1	1	1	1	1	1
Кабардино Балкарская Республика	1	1	1	1	1	1	1
Забайкальский край	1	1	1	1	1	1	1
Ставропольский край	1	2	2	2	1	2	1
Республика Ингушетия	1	1	1	1	1	1	1
Республика Саха (Якутия)	2	2	2	2	2	2	1
Республика Бурятия	1	1	1	1	1	1	1
Кировская область	2	1	1	1	1	1	1
Курганская область	1	1	1	1	1	1	1
Алтайский край	1	2	1	1	1	1	1
Республика Крым	1	2	1	2	1	1	1
Чеченская Республика	1	1	1	1	1	1	1
Республика Башкортостан	2	2	2	2	2	2	1
Республика Алтай	1	1	1	1	1	1	1
Чувашская Республика	1	1	1	1	1	1	1
Камчатский край	1	1	1	1	1	1	1
Брянская область	1	1	1	1	1	1	1
Карачаево Черкесская Республика	1	1	1	1	1	1	1
Республика Тыва	1	1	1	1	1	1	1
Чукотский автономный округ	2	1	1	1	1	1	1
Республика Дагестан	1	1	1	1	1	1	1
Ростовская область	1	2	2	2	2	2	1
Красноярский край	2	2	1	2	2	2	2
Иркутская область	2	2	2	2	2	2	2
Тюменская область	2	2	2	2	2	2	2
Ярославская область	2	2	2	2	2	2	2
Тюменская область без автономных округов	2	2	2	2	2	2	2
Пермский край	2	2	1	2	2	2	2
Свердловская область	1	2	2	2	2	2	2
Нижегородская область	2	2	2	2	2	2	2
Липецкая область	2	2	2	2	2	2	2
<i>Мурманская область</i>	1	2	1	1	1	1	2
Ленинградская область	2	2	1	2	2	2	2
г. Москва	2	2	2	2	2	2	2
Вологодская область	2	2	2	2	1	2	2

Объекты	KNN	SVM	DT	SWS	XGB	Ансамбль	Верная группа
Сахалинская область	2	2	1	1	1	1	2
Ханты Мансийский автономный округ – Югра	2	2	2	2	2	2	2
Тульская область	2	2	2	2	2	2	2
Краснодарский край	1	2	2	2	2	2	2
Самарская область	2	2	2	2	2	2	2
Республика Татарстан	2	2	1	2	2	2	2
Ненецкий автономный округ	2	1	1	1	1	1	2
г. Санкт Петербург	2	2	2	2	2	2	2
Калужская область	2	2	2	2	2	2	2
Белгородская область	1	2	2	2	2	2	2
Ямало-Ненецкий автономный округ	2	2	2	2	2	2	2
Московская область	1	2	2	2	2	2	2

В табл. 5 выделены регионы, которые алгоритм ошибочно отнес в другую группу: полужирным — отнесенные к бездотационным, курсивом — распознанные как дотационные, — что позволяет для них делать определенные прогнозы.

В табл. 6 можно видеть, что наиболее значимые показатели для сравнения двух групп в данном исследовании — это основные фонды (млн. руб.); обрабатывающие производства, отвечающие за обеспечение электрической энергией, газом и паром, кондиционирование воздуха; водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений; строительство; торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов; транспортировка и хранение; деятельность в области информации и связи; добыча полез-

ных ископаемых. Именно эти показатели характеризуют основную экономическую деятельность, необходимую для жизнеобеспечения регионов. Мы посчитали парные коэффициенты корреляции между всеми рассмотренными выше признаками. Ввиду громоздкости таблицы все результаты не представляем. Все эти числа, кроме одного, больше 0,9, т. е. между рассмотренными показателями существует прямая сильная связь. Этот результат подтверждает наличие очень тесной связи между показателями, характеризующими основные средства. Отрицательная умеренная связь с коэффициентом корреляции, равным $-0,503$ наблюдается между несвязанными между собой показателями: добычей полезных ископаемых и торговлей, ремонтом автотранспортных средств.

Таблица 6

Результаты теста Манна—Уитни—Уилкоксона (*U-test*)

Показатель	N1	Mean1	N2	Mean2	p-value
Стоимость основных фондов на конец года; по полной учетной стоимости; миллионов рублей	22	2 028 896,636	25	12 171 176,480	0,000
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	22	73 925,045	25	137 241,840	0,094
Добыча полезных ископаемых	22	119 952,773	25	1 808 522,720	0,004
Обрабатывающие производства	22	107 453,409	25	819 647,280	0,000
Обеспечение электрической энергией, газом и паром; кондиционирование воздуха	22	141 769,273	25	499 154,800	0,000
Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	22	12 657,727	25	87 837,520	0,000
Строительство	22	21 900,727	25	101 480,760	0,001
Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов	22	32 742,909	25	161 410,760	0,000
Транспортировка и хранение	22	340 121,500	25	1 978 134,480	0,000
Деятельность в области информации и связи	22	31 031,955	25	217 059,360	0,000
Ввод в действие основных фондов по видам экономической деятельности	22	119 660,318	25	724 144,560	0,000
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство — ввод в действие	22	9 375,682	25	18 341,840	0,103
Добыча полезных ископаемых — ввод в действие	22	12 989,864	25	139 748,000	0,001
Обрабатывающие производства — ввод в действие	22	9 798,318	25	79 025,440	0,000
Обеспечение электрической энергией, газом и паром; кондиционирование воздуха — ввод в действие	22	9 003,545	25	30 851,400	0,000

Показатель	N1	Mean1	N2	Mean2	p-value
Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений — ввод в действие	22	623,864	25	7 291,760	0,000
Строительство — ввод в действие	22	2 926,318	25	17 549,840	0,000
Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов — ввод в действие	22	3 067,818	25	20 751,040	0,000
Транспортировка и хранение — ввод в действие	22	13 769,773	25	71 733,960	0,000
Деятельность в области информации и связи — ввод в действие	22	2 324,182	25	24 773,880	0,001
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство — структура	22	4,714	25	2,744	0,029
Обрабатывающие производства — структура	22	4,594	25	10,232	0,009
Обеспечение электрической энергией, газом и паром; кондиционирование воздуха — структура	22	7,845	25	5,111	0,017
Обрабатывающие производства	22	7,600	25	17,461	0,011
Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	21	50,405	25	39,808	0,013
Обрабатывающие производства	22	14,829	25	19,722	0,039

Таблица 6 демонстрирует, что выбранные показатели можно использовать для машинного обучения. Полученное высокое значение качества распознавания — площадь под РОК-кривой равна 0,9 — подтверждает это предположение.

Заключение

Наше исследование показало, что дотационные регионы в целом распознаются по показателям основных средств среди регионов Российской Федерации. Для коллективного решения $AUC = 0.90$. Чувствительность — 82 %. Специфичность — 84 %.

При распознавании были допущены следующие ошибки. Ставропольский край, Ростовская область, Республика Саха (Якутия) и Башкортостан были отнесены к бездотационным регионам, хотя таковыми они не являются. Видимо, в этих регионах наблюдаются недостаточные инвестиции в основные фонды, а существенную роль в бездотационности играет прибыль компаний, расположенных в этих регионах, что не учитывалось в данном исследовании. Несмотря на влияние других финансово-экономических и географических показателей, не учтенных при исследовании, можно уверенно сделать вывод, что эти четыре региона располагают потенциалом выйти из группы дотационных субъектов Российской Федерации.

Три региона: Мурманская область, Сахалинская область, Ненецкий автономный округ — наоборот, были распознаны как дотационные, хотя они не являются регионами-реципиентами, т. к. в этих регионах положитель-

ный баланс (доходы больше расходов), прежде всего из-за добычи и обработки полезных ископаемых, которыми они обладают. Это регионы — доноры. Эти результаты свидетельствуют в пользу того, что руководителям данных регионов надо обратить существенное внимание на состояние основных средств и инвестиции в них.

Можно сделать вывод, что, используя диаграммы рассеяния для пар показателей, характеризующих основные средства, можно получить решающее правило, которое может быть использовано в качестве прогностического алгоритма для оценки потенциала региона с точки зрения дотационности.

Результаты проведенного исследования могут быть использованы органам государственной исполнительной власти в процессе актуализации государственных программ развития конкретных регионов в направлении устойчивого перехода регионов из дотационного пула в бездотационный с точки зрения основных средств.

В заключение отметим, что целью работы было не развитие методов оценки необходимости дотирования регионов, а обоснование включения новых факторов в методику оценки необходимости дотирования регионов на основе самых перспективных методов машинного обучения, таких как метод статистически взвешенных синдромов и метод оптимальных достоверных разбиений, позволяющие выявить наиболее информативные признаки и на основе полученных границ разрабатывать планы перевода объекта из неблагоприятной группы (дотационной) в бездотационную.

СПИСОК ИСТОЧНИКОВ

- Журкина Т. А., Сабетова Т. В. Совершенствование методики анализа основных средств предприятия // Вестник Воронежского государственного университета инженерных технологий. 2018. Т. 80. № 1. С. 273—282. DOI: 10.20914/2310-1202-2018-1-273-282.
- Hänggi R., Fimpel A., Siegenthaler R. LEAN Production – Easy and Comprehensive : A practical guide to lean processes explained with pictures. Vieweg ; Berlin ; Heidelberg : Springer, 2022. XXV, 191 p. DOI: 10.1007/978-3-662-64527-7.
- Бачериков И. В., Свойкин Ф. В., Бирман А. П., Соколова В. А. Стохастическая модель оптимизации затрат при планировании технологических процессов лесозаготовок // Системы. Методы. Технологии. 2017. № 4(36). С. 182—186.
- Algorithms for solving the problems of optimizing water resources management on a reservoir seasonal regulation / S. Rakhimov, A. Seytov, M. Sherbaev et al. // AIP Conference Proceedings. 2022. Vol. 2432. Iss. 1. Art. 060023. DOI: 10.1063/5.0090412.
- Мильчаков М. В. Высокодотационные регионы России: условия формирования бюджетов и механизмы государственной поддержки // Финансовый журнал. 2017. № 1. С. 22—38.

6. Гареев М. Использование методов машинного обучения для прогнозирования инвестиций в России // Деньги и кредит. 2020. Т. 79. № 1. С. 35—56. DOI: 10.31477/rjmf.202001.35.
7. Investments in fixed assets in Russia: analysis and forecast / O. Chistic, O. Ovchinnikov, A. Volgin et al. // E3S Web of Conferences. 2023. Vol. 389. Art. 09016. DOI: 10.1051/e3sconf/202338909016.
8. Svoboda J., Lososová J, Zdeněk R. Subsidies on Investments in the EU Member States // Agris on-line Papers in Economics and Informatics. 2016. Vol. 8. No. 4. Pp. 153—162. DOI: 10.7160/aol.2016.080414.
9. Pérez-Méndez J. A., Pérez-Urdiales M., Roibas D. Evaluating the effect of subsidies for rural development on agri-food and forestry firms: Technical progress and efficiency // Applied Economic Analysis. 2019. Vol. 27. No. 80. Pp. 150—167. DOI: 10.1108/AEA-06-2019-0004.
10. Юсупов Р. А. Формализованный подход в расчете дотаций регионам // Экономика и управление. 2022. № 3(166). С. 61—67.
11. Борисова Л. Р., Кузнецова А. В. Использование работающего компьютерного тренажера Data Master Azforus для обучения методам машинного обучения // Цифровая трансформация социальных и экономических систем : материалы Междунар. науч.-практ. конф. М. : Моск. ун-т им. С. Ю. Витте, 2022. С. 264—270.
12. Кузнецова А. В., Сенько О. В., Кузнецова Ю. О. Преодоление проблемы черного ящика при использовании методов машинного обучения в медицине // Врач и информационные технологии. 2018. № 7. С. 74—80.
13. Кузнецова Ю. О., Борисова Л. Р., Кузнецова А. В., Сенько О. В. Прозрачный интерфейс для прогноза в машинном обучении // Аналитика и управление данными в областях с интенсивным использованием данных : сб. науч. тр. XIX Междунар. конф. DAMDID / RCDL'2017. М. : ФИЦ ИУ РАН, 2017. С. 493—495.
14. Борисова Л. Р., Сенько О. В., Кузнецова А. В., Сергеева Н. В. Применение методов машинного обучения для сравнения компаний арктической зоны РФ по экономическим критериям в соответствии с рейтингом полярного индекса // Компьютерные исследования и моделирование. 2020. Т. 12. № 1. С. 201—215. DOI: 10.20537/2076-7633-2020-12-1-201-215.
15. Кузнецова А. В., Борисова Л. Р., Хадартцев В. М. Применение многопараметрических методов Data Science для классификации субъектов Российской Федерации по признаку дотационности // Управление. 2024. Т. 12. № 3. С. 58—73. DOI: 10.26425/2309-3633-2024-12-3-58-73.

REFERENCES

1. Zhurkina T. A., Sabetova T. V. Improvement of the methods for company's fixed assets analysis. *Vestnik Voronezhskogo gosudarstvennogo universiteta inzhenernykh tekhnologii = Proceedings of the Voronezh State University of Engineering Technologies*. 2018;80(1):273—282. (In Russ.) DOI: 10.20914/2310-1202-2018-1-273-282.
2. Hänggi R., Fimpel A., Siegenthaler R. LEAN Production – Easy and Comprehensive. A practical guide to lean processes explained with pictures. Vieweg, Berlin, Heidelberg, Springer, 2022. XXV, 191 p. DOI: 10.1007/978-3-662-64527-7.
3. Bacherikov I. V., Svoynkin F. V., Birman A. R., Sokolova V. A. Stochastic model of cost optimization in the planning of logging processes. *Sistemy. Metody. Tekhnologii = Systems. Methods. Technologies*. 2017;4(36):182—186. (In Russ.)
4. Rakhimov S., Seytov A., Sherbaev M. et al. Algorithms for solving the problems of optimizing water resources management on a reservoir seasonal regulation. *AIP Conference Proceedings*. 2022;2432(1):060023. DOI: 10.1063/5.0090412.
5. Milchakov M. V. Heavily Non-Purpose Grants-Financed Russian Regions: Budgetary Conditions and Tools of State Support. *Finansovyi zhurnal = Financial journal*. 2017;1:22—38. (In Russ.)
6. Gareev M. Use of Machine Learning Methods to Forecast Investment in Russia. *Den'gi i kredit = Russian Journal of Money and Finance*. 2020;79(1):35—56. (In Russ.) DOI: 10.31477/rjmf.202001.35.
7. Chistic O., Ovchinnikov O., Volgin A. et al. Investments in fixed assets in Russia: analysis and forecast. *E3S Web of Conferences*. 2023;389:09016. DOI: 10.1051/e3sconf/202338909016.
8. Svoboda J., Lososová J, Zdeněk R. Subsidies on Investments in the EU Member States. *Agris on-line Papers in Economics and Informatics*. 2016;8(4):153—162. DOI: 10.7160/aol.2016.080414.
9. Pérez-Méndez J. A., Pérez-Urdiales M., Roibas D. Evaluating the effect of subsidies for rural development on agri-food and forestry firms: Technical progress and efficiency. *Applied Economic Analysis*. 2019;27(80):150—167. DOI: 10.1108/AEA-06-2019-0004.
10. Yusupov R. A. Formalized Approach to Calculating Subsidies to Regions. *Ekonomika i upravlenie = Economics and Management*. 2022;3(166):61—67. (In Russ.)
11. Borisova L. R., Kuznetsova A. V. Using a working computer simulator Data Master Azforus for teaching machine learning methods. *Tsifrovaya transformatsiya sotsial'nykh i ekonomicheskikh sistem = Digital transformation of social and economic systems. Materials of the International Scientific and Practical Conference*. Moscow, Moscow Witte University publ., 2022:264—270. (In Russ.)
12. Kuznetsova A. V., Senko O. V., Kuznetsova Ju. O. Black box problem overcoming in medical applications of machine learn. *Vrach i informatsionnye tekhnologii = Medical doctor and IT*. 2018;7:74—80. (In Russ.)
13. Kuznetsova Ju. O., Borisova L. R., Kuznetsova A. V., Senko O. V. Transparent interface for prediction in machine learning. *Analitika i upravlenie dannymi v oblastiakh s intensivnym ispol'zovaniem dannyykh = Data Analytics and Management in Data Intensive Domains. Collection of Scientific Papers of the XIX International Conference DAMDID / RCDL'2017*. Moscow, FRC CSC RAS publ., 2017:493—495. (In Russ.)
14. Borisova L. R., Senko O. V., Kuznetsova A. V., Sergeeva N. V. Comparison of Arctic zone RF companies with different polar index ratings by economic criteria with the help of machine learning tools. *Komp'yuternye issledovaniya i modelirovanie = Computer Research and Modeling*. 2020;12(1):201—215. (In Russ.) DOI: 10.20537/2076-7633-2020-12-1-201-215.
15. Kuznetsova A. V., Borisova L. R., Khadartsev V. M. Application of multiparametric methods of data science for the classification of Russian subjects on the basis of subsidisation. *Upravlenie = Management (Russia)*. 2024;12(3):58—73. (In Russ.) DOI: 10.26425/2309-3633-2024-12-3-58-73.

Статья поступила в редакцию 30.10.2024; одобрена после рецензирования 24.11.2024; принята к публикации 25.11.2024.
The article was submitted 30.10.2024; approved after reviewing 24.11.2024; accepted for publication 25.11.2024.