

Анализ некоторых аспектов цифровизации в области оказания образовательных услуг населению методами статистического анализа и машинного обучения

Л.Р. Борисова, М.Н. Фридман

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

АННОТАЦИЯ

Цифровая трансформация во всех сферах жизни значительно ускорилась после COVID-19. Во всем мире и в России, в частности, в последние пять лет наблюдается существенный прогресс в области профессиональных навыков в цифровых технологиях. Это одновременно вызов и возможности, когда вопросы интеграции и роста пересекаются. В нашей стране цифровизация в сфере государственных услуг населению развивается опережающими темпами. России необходимы образовательные системы, соответствующие цифровому веку, а также дополнительные программы для обучения и переподготовки той части взрослого населения, которая давно прекратила учебу. Это двусторонняя задача, касающаяся как базовых цифровых навыков, так и современного понимания грамотности. Цель развития цифровых образовательных услуг населению — формирование доступной и эффективной системы обучения для всех категорий граждан. Задача исследования — выявить объективными методами математического анализа данных (без какого-либо использования субъективных оценок) наличие или отсутствие взаимосвязи между социально-экономическими параметрами развития региона Российской Федерации и платными образовательными услугами населению. В работе использованы методы кластеризации и классификации однородных объектов по широкому набору социально-экономических и финансовых показателей, важных для исследования развития цифровизации в сфере услуг для населения. Получено четкое разделение регионов на две большие, практически равные группы по этому набору показателей. Представлен сравнительный анализ методов иерархической кластеризации и классической кластеризации с использованием метода К-средних. Отдано предпочтение в пользу иерархической кластеризации. Получены непротиворечивые результаты с использованием обоих методов. Сделан вывод о важности географического фактора и однородности в экономическом развитии регионов, сгруппированных в кластеры. Москва выделяется по уровню развития цифровизации по сравнению с другими регионами, что подтверждают оба метода кластеризации. Использование методов машинного обучения позволило сделать вывод о большей значимости абсолютных показателей по сравнению с относительными для классификации регионов по уровню развития цифровых образовательных услуг населению.

Ключевые слова: методы машинного обучения; цифровизация; статистика; регрессия; иерархическая кластеризация; метод К-средних; дендрограмма; услуги

Для цитирования: Борисова Л.Р., Фридман М.Н. Анализ некоторых аспектов цифровизации в области оказания образовательных услуг населению методами статистического анализа и машинного обучения. *Цифровые решения и технологии искусственного интеллекта*. 2026;2(3):67-78. DOI: 10.26794/3030-7097-2026-2-3-67-78

ORIGINAL PAPER

Analysis of Certain Aspects of Digitalization in the Field of Providing Educational Services to the Population Using Statistical Analysis and Machine Learning Methods

L.R. Borisova, M.N. Fridman

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

ABSTRACT

Digital transformation in all areas of life has accelerated significantly since COVID-19. Significant progress in the field of professional skills in digital technologies has been observed throughout the world and in Russia, in particular, in the last five years. It is a challenge and an opportunity when issues of integration and growth coincide with each other.



Digitalization in the field of public services is developing rapidly in our country. Russia needs educational systems appropriate to the digital age, as well as additional programs for the education and retraining of that part of the adult population that has long abandoned their studies. This is a two-pronged task, dealing with both basic digital skills and a modern understanding of literacy. The goal of developing digital educational services for the population is to create an accessible and effective learning system for all categories of citizens. The paper uses clustering and classification methods to group homogeneous objects based on a wide range of socio-economic and financial indicators that are important for studying the development of digitalization in the services provided to the population. The objective of the study is to identify, using objective methods of mathematical data analysis (without any use of subjective assessments), the presence or absence of a relationship between the socio-economic parameters of a region's development in the Russian Federation and paid educational services provided to the population. The regions are clearly divided into two large, almost equal groups based on this set of indicators. A comparative analysis of hierarchical clustering and classical clustering using the K-means method is presented. Hierarchical clustering is preferred over classical clustering. Consistent results were obtained using both methods. The importance of the geographical factor and homogeneity in the economic development of regions grouped into clusters was concluded. Moscow stands out for its digitalization development compared to other regions, as evidenced by both clustering methods.

Keywords: machine learning methods; digitalization; statistics; regression; hierarchical clustering; K-means method; dendrogram; services

For citation: Borisova L.R., Fridman M.N. Analysis of certain aspects of digitalization in the field of providing educational services to the population using statistical analysis and machine learning methods. *Digital Solutions and Artificial Intelligence Technologies*. 2026;2(3):67-78. DOI: 10.26794/3030-7097-2026-2-3-67-78

ВВЕДЕНИЕ

Стремительный технический и технологический процесс бросает вызов не только различным отраслям экономики, но также и сфере оказания услуг населению.

Пандемия, наряду с общими тенденциями технического прогресса, стимулировала развитие цифровизации в различных областях оказания услуг населению во всех развитых странах. Россия оказалась одной из самых передовых в применении цифровых технологий для оказания широкого спектра социальных услуг: взаимодействие с государственными органами, с медицинскими и образовательными учреждениями, услуги в области культуры, туризма и спорта, в банковской и финансовой сферах, в области социальной и психологической помощи, в организации досуга и даже поиска партнеров [1, 2].

Достаточно вспомнить, как практически без потерь вузы и школы перешли на онлайн-обучение в разгар эпидемии. В области медицинских услуг Россия оперативно перевела взаимоотношения пациента и медицинского учреждения на полную цифровизацию [3], в отличие, например, от Германии, где цифровые истории болезней начали вводить только как факультативное предложение в 2021 г., и в тот момент на него подписалось менее 1% населения страны¹. Сегодня в области взаимодействия с государственными органами редко кто обходится без сайта «Госуслуги». Правда, следует отметить, что некоторые исследования указывают на некоторый откат в сторону оффлайн-предпочтений после

окончания пандемии² и с причинами этого следует разобраться.

В чем заключается особая важность цифровизации в области сферы услуг для России? В первую очередь это необходимо для обеспечения доступности и качества широкого спектра услуг во всех районах громадной территории страны, вне зависимости от их удаленности от центра, уровня их экономического развития. Это необходимо также для формирования нового, современного рынка услуг, что приводит в конечном итоге к ускорению соответствующих процессов в экономике [4]. Оуществляется обратная связь, взаимодействие спроса и предоставления услуг, причем бизнес должен проводить широкий анализ и прогнозирование спроса на услуги, используя современный математический аппарат работы с большими данными.

В научной среде достаточно работ, указывающих на важность использования цифровизации для повышения качества образования, где приводится анализ динамики цифровизации общего образования в России и сделан вывод, что развитию цифровизации в деятельности школ будет способствовать разработка единых требований в данной области [5, 6]. Но пока регионы отличаются по уровню использования школьниками электронного обучения³.

² Щербakov P.A. Цифровизация государственных услуг. М.: ИСИЭЗ НИУ ВШЭ; 2025. URL: <https://issek.hse.ru/news/1039474722.html>; Попов Е.В. Цифровые навыки россиян. М.: ИСИЭЗ НИУ ВШЭ; 2025. URL: <https://issek.hse.ru/news/1038822258.html>

³ Отчет Счетной палаты РФ «Анализ реализации мер в сфере цифровизации образования». URL: <https://ach.gov.ru/upload/iblock/0d8/b8pg1tjsapfepfg6r0xvur3tk1mrn0el.pdf?ysclid=mt8j4vee1f579362498>

¹ URL: <https://ausnews.de/digital-healthcare/>



Как уже было отмечено, широкая географическая и экономическая дисперсность является серьезной проблемой в задаче унификации качества и количества предоставляемых населению услуг в России.

Особый интерес для авторов представляет развитие цифровизации образовательных услуг, целью которой является формирование доступной и эффективной системы обучения для всех категорий граждан. Актуальность этого вопроса в условиях России не вызывает сомнений. Благодаря цифровой среде обучение становится возможным независимо от географического положения и финансовых возможностей обучающихся, что особенно важно для удаленных регионов с ограниченной инфраструктурой и экономически слаборазвитых регионов.

Предметом исследования было выявление определенных закономерностей в распределении элементов цифровизации по ряду показателей с помощью формирования территориальных кластеров с целью выработки некоторых рекомендаций для оптимизации процесса цифровизации в области предоставления платных образовательных услуг населению.

МАТЕРИАЛЫ И МЕТОДЫ

База данных. В исследовании использованы открытые данные сайта Росстата⁴ для 85 регионов Российской Федерации (РФ). Данные для Архангельской и Тюменской областей, Чукотского, Ямало-Ненецкого и Ханты-Мансийского автономных округов рассмотрены отдельно. Данные для четырех новых регионов Российской Федерации не представлены в статистике Росстата, поэтому их в статье не учитывали.

Используемые для машинного обучения данные в основном относятся к периоду 2023–2024 гг. Исключения составили два показателя: глобальные информационные сети и системы электронного документооборота, используемые организациями. Эти данные относятся к 2019 г. Оба показателя являются безразмерными и представляют собой процент от всех обследованных организаций.

Всего был проанализирован 31 показатель: 30 зависимых показателей приведены в *табл. 1* и 1 независимый показатель (группирующий признак для методов машинного обучения) — уровень платных образовательных услуг населению за июль 2024 г. Последние 7 показателей также относятся к этому же периоду.

Методы анализа. Были использованы иерархическая кластеризация и метод К-средних

(K-means) — одни из наиболее распространенных подходов в кластерном анализе. Иерархическая кластеризация может быть предпочтительнее в ряде случаев, поскольку позволяет исследователю самостоятельно определить количество кластеров по итогам анализа (например, «разрезав» полученную дендрограмму на нужном уровне), тогда как для К-средних число кластеров необходимо задать до запуска алгоритма⁵. Дендрограмма при использовании иерархической кластеризации была построена с применением функции «hclust» из пакета «cluster» языка программирования R. Двумерные диаграммы, иллюстрирующие разбиение 85 регионов России на кластеры, были построены с использованием программного комплекса компании Azforus [7]. Для классификации 95 регионов РФ по показателю цифровизации образовательных услуг также был применен метод статистически взвешенных синдромов (СВС) [8]. Метод СВС основан на принятии коллективных решений по системам синдромов двумерных областей признаков пространства, в котором преобладают объекты одного из распознаваемых классов. Области задаются с помощью границ, которые находятся через оптимальные разбиения интервалов значений признаков. Подсчет взвешенных голосов проводится как по одномерным, так и по двумерным разбиениям. Подробно метод СВС описан в лекциях разработчика метода О.В. Сенько⁶ и в работе [9].

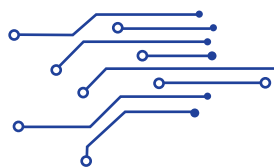
РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Результаты кластерного анализа. Методы кластерного анализа позволяют выделить в многомерном пространстве показателей (признаков) компактные группы близких друг другу объектов. Кластеры отличаются друг от друга по высоте (от англ. height), характеризующей степень различия регионов по исследуемым показателям. Данный подход позволяет выявить скрытые связи и закономерности, которые можно не заметить, если использовать только методы базовой статистики. Отметим, что в предыдущей работе по теме этот метод успешно применен для группировки стран Азиатско-Тихоокеанского региона по значениям инклюзивного экономического роста [10]. Там же подробно описано использование иерархической кластеризации.

⁵ James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning: With Applications in R. Textbook. Springer; 2021.

⁶ Сенько О.В. Математические основы теории прогнозирования. URL: http://www.machinelearning.ru/wiki/images/c/cd/MOTP13_6_bw.pdf?ysclid=mt8disakcn735931691

⁴ URL: <https://data.rosstat.org>



Показатели, используемые для исследования / Indicators Used for the Study

№	Показатели, используемые для исследования / Indicators Used for the Study
	ОРГАНИЗАЦИИ (% от общего числа обследованных) использовавшие
1	Технологии искусственного интеллекта
2	Интернет вещей
3	ПК
4	Серверы
5	Локальные вычислительные сети
6	Глобальные информационные сети
7	Облачные сервисы
8	Технологии сбора, обработки и анализа больших данных
9	Цифровые платформы
10	Мобильный и фиксированный интернет
11	Широкополосной интернет
12	Веб-сайт
13	Системы электронного документооборота
14	Электронный обмен между своими и внешними информационными системами документооборота
Удельный вес домашних хозяйств (% от всех домохозяйств)	
15	Имевших персональный компьютер
16	Имевших доступ к сети Интернет
17	Имевших доступ к широкополосному интернету
18	% населения, использующих интернет
19	% подключенных абонентских устройств мобильной связи на 1000 населения
20	Численность активных абонентов фиксированного широкополосного интернета (на 100 человек)
21	Численность активных абонентов мобильного широкополосного интернета (на 100 человек)
22	Ежемесячные среднедушевые денежные доходы, тыс. руб.
23	Ежемесячные среднедушевые потребительские расходы, тыс. руб.
Платные услуги населению за июль 2024 г. в рублях на чел.	
24	Телекоммуникационные
25	Физической культуры и спорта
26	Почтовой связи и курьеров
27	Фитнес-центров и спортивных клубов
28	Медицинские
29	Ветеринарные
30	Юридические

Источник / Source: составлено авторами по материалам Росстата / Compiled by the authors based on the materials of Rosstat.
URL: <https://rosstat.gov.ru/>



Методы кластерного анализа применены для оценки российских регионов по 31 показателю (30 показателей из табл. 1) и еще одному показателю, который в дальнейшем будет использоваться в качестве зависимой переменной при применении методов машинного обучения. 31-й показатель — это платные образовательные услуги населению за июль 2024 г. в рублях на человека. На рис. 1 представлена дендрограмма иерархической кластеризации.

Проанализируем дендрограмму рис. 1. Для удобства оценки номера регионов с их соответствующей расшифровкой приведены в табл. 2, в которую также записаны кластеры (10), четко разделяемые по высоте. Удобно для анализа разрезать дендрограмму горизонтальной линией на уровне $height = 10$. Тогда число кластеров должно равняться 10 или незначительно отличаться. Количество кластеров выбраны неслучайно. В работе [11] было проанализировано похожее число показателей (32), количество объектов для кластерного анализа — 45 и представлено подробное обоснование, почему опти-

мальное количество кластеров должно находиться в диапазоне 6–12. Как отмечено в работе [12], оптимальное количество кластеров определяется по дендрограмме. Поскольку в исследовании использовали готовую программу из библиотеки программ алгоритмического языка R, то не ставили задачу рассчитать индекс силуэта для оценки качества кластеризации [13].

Проанализируем разделение регионов на 10 кластеров (то же самое число выберем для возможности сравнения однотипных данных) с использованием самого распространенного метода кластеризации — K-средних, в котором число кластеров фиксируется исследователем до проведения кластеризации. Представим двумерную диаграмму из кластерного анализа на рис. 2, где четко прослеживается деление всех 85 регионов на 10 кластеров, отличающихся цветом и геометрическими фигурами (ромбы, квадраты, кружочки, звездочки, крестики) для визуализации кластеризации. Чтобы результаты двух различных методов (иерархической кластеризации и K-средних) можно было сравнить

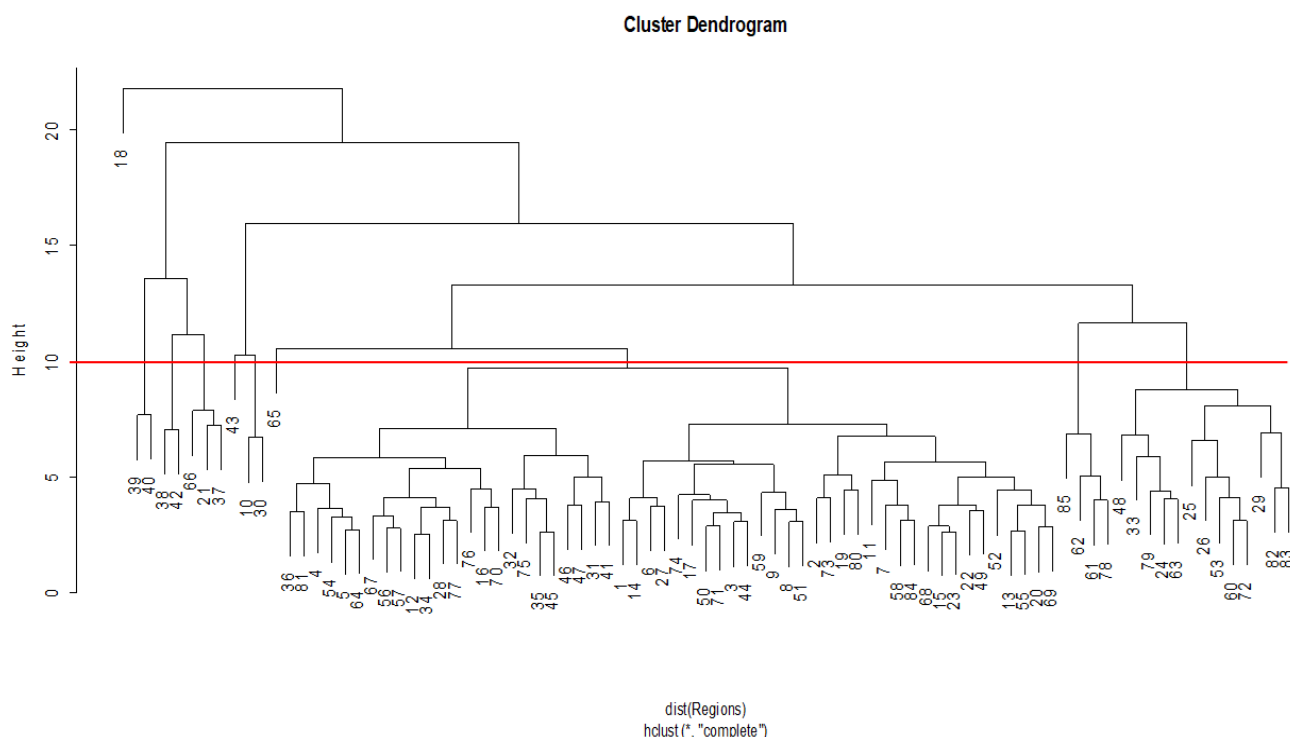
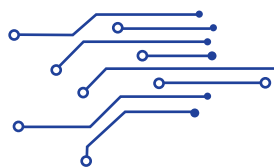


Рис. 1 / Fig. 1. Кластерная дендрограмма, соответствующая разделению РФ по показателям цифровизации услуг населению / Cluster Dendrogram Corresponding to the Division of the Russian Federation by Indicators of Digitalization of Services to the Population

Источник / Source: составлено авторами / Compiled by the authors.

Примечание / Note: Height — высота (ось Y), характеризующая уровень отличия наблюдений (кластеров); ось X — рисунок дендрограммы с условными номерами регионов, расшифрованными в табл. 2 / Height — height (Y-axis), which characterizes the level of difference between observations (clusters); X — axis is a dendrogram drawing with conditional numbers of regions, deciphered in Table 2.

**Нумерация и состав кластеров из регионов по методу иерархической классификации /
Numbering and Composition of Clusters from the Regions Using the Hierarchical Classification Method**

Номер / Number		Название региона / Region name
Кластер / Cluster	Регион / Region	
1	18	Москва
2	39, 40	Республики Ингушетия и Кабардино-Балкария
3	38, 42	Республики Дагестан и Северная Осетия (Алания)
4	66, 21, 37	Республика Тыва, Ненецкий автономный округ, Севастополь
5	43, 10, 30	Чеченская Республика, Московская обл., Республика Адыгея
6	65	Республика Алтай
7	36, 81, 4, 54, 5, 64, 67, 56, 57, 12, 34, 28, 77, 76, 16, 70, 32, 75, 35, 45, 46, 47, 31, 41	Ростовская, Амурская, Воронежская, Оренбургская, Ивановская, Челябинская обл., Республика Хакасия, Самарская, Саратовская, Рязанская, Астраханская, Псковская обл.; Забайкальский край, Республика Саха (Якутия), Тульская, Иркутская обл.; Республики Крым и Бурятия, Волгоградская обл., Республики Башкортостан, Марий Эл, Мордовия, Калмыкия, Карачаево-Черкесская
8	1, 14, 6, 27, 74, 17, 50, 71, 3, 44, 59, 9, 8, 51, 2, 73, 19, 80, 11, 7, 58, 84, 68, 15, 23, 22, 49, 52, 13, 55, 20, 69	Белгородская, Тамбовская, Калужская, Новгородская, Томская, Ярославская обл., Республика Чувашия; Кемеровская, Владимирская обл., Ставропольский край, Курганская, Липецкая, Курская обл., Пермский край, Брянская, Омская обл.; Республика Карелия, Хабаровский край, Орловская, Костромская, Ульяновская обл., Еврейская автономная обл., Алтайский край, Тверская, Волгоградская, Архангельская обл., Республика Удмуртия, Кировская, Смоленская, Пензенская обл., Республика Коми, Красноярский край
9	85, 62, 61, 78, 48	Чукотский, Ямало-Ненецкий, Ханты-Мансийский автономные округа, Камчатский край, Татарстан
10	33, 79, 24, 63, 25, 26, 53, 60, 72, 29, 82, 83	Краснодарский, Приморский края, Калининградская, Тюменская, Ленинградская, Мурманская, Нижегородская, Свердловская, Новосибирская обл., Санкт-Петербург, Магаданская и Сахалинская обл.

Источник / Source: составлено авторами / Compiled by the authors.

между собой, регионы, вошедшие в 10 кластеров с использованием метода К-средних, представим в табл. 3.

Результаты машинного обучения. Выявление скрытых статистически достоверных связей возможно с применением методов классификации. В качестве целевой функции использован показатель объема платных образовательных услуг населению за июль 2024 г. в рублях на человека.

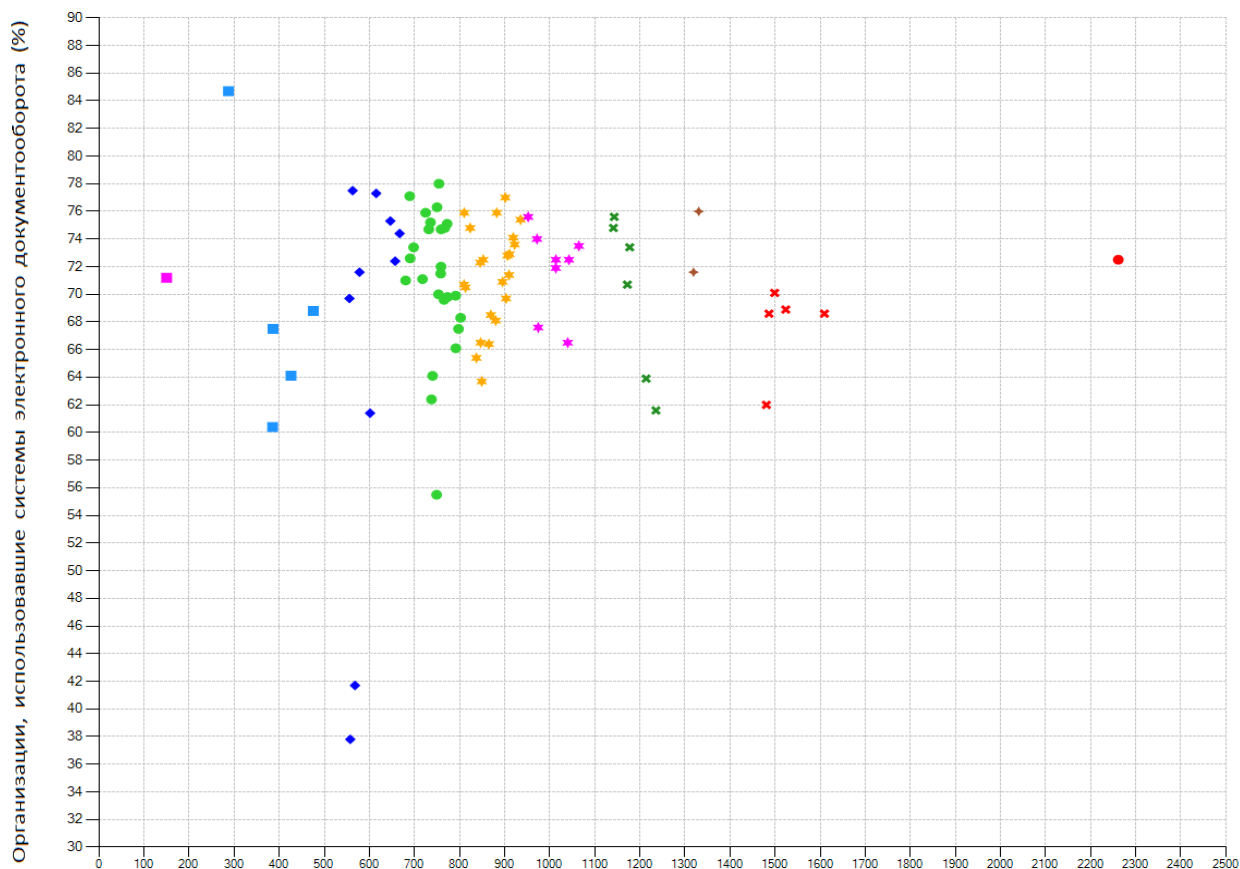
Для деления 85 регионов на две группы было вычислено медианное значение группирующего показателя, составившее 295,88 руб. в месяц на человека. Количество регионов, в которых этот показатель оказался ниже медианного уровня, равно 44, а выше медианного уровня — 41.

Был применен метод распознавания — статистически взвешенные синдромы (СВС), ранее описанный в работе [8]. Зависимые показатели ис-

пользовались те же, что и для кластерного анализа. Их количество составляет 30, и они представлены в табл. 1.

Результаты распознавания характеризуются наличием ошибок при использовании скользящего контроля. В процессе работы программы случайным образом были отобраны 18 регионов из 85 анализируемых, причем количество регионов в каждой группе составило по 9. Значения параметров классификации представлены в табл. 4.

Отметим, что Accuracy score — это классический для машинного обучения метод оценки точности модели, подразумевающий процент совпадения всех предсказанных и реальных результатов. ROC AUC score представляет из себя площадь под графиком рабочей характеристики приемника. График позволяет оценить качество бинарной клас-



Платные телекоммуникационные услуги населению за июль 2024 г. руб. на чел.

Рис. 2 / Fig. 2. Зависимость удельного веса организаций, использующих электронный документооборот (%) от ежемесячных платных телекоммуникационных услуг населению (в руб. за июль 2024 г.) / Dependence of the Share of Organizations Using Electronic Document Management (%) on Monthly paid Telecommunications Services for the Population (in rubles for July 2024)

Источник / Source: составлено авторами / Compiled by the authors.

сификации, отображает соотношение между долей объектов от общего количества носителей признака, верно классифицированных как несущих признак, и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущих признак. Точность линейной классификации можно основывать на 4 типах оценки результата:

TP — (True Positive) истинно-положительное решение;

TN — (True Negative) истинно-отрицательное решение;

FP — (False Positive) ложно-положительное решение;

FN — (False Negative) ложно-отрицательное решение.

Точность рассчитывается по формуле $\frac{TP}{TP + FP}$, полнота — $\frac{TP}{TP + FN}$, чувствительность — $\frac{TP}{TP + FN}$, специфичность — $\frac{TN}{TP + FP}$.

Было зарегистрировано всего четыре ошибки, а именно: три региона, относящиеся к 1-й группе, — Калужская и Иркутская области, Республика Саха (Якутия) (со значением группирующего показателя ниже медианного значения), при использовании скользящего контроля были отнесены ко 2-й группе, а Воронежская область, наоборот, была распознана как регион с показателем платных ежемесячных услуг населению ниже медианы, хотя в июле 2024 г. этот показатель был выше медианы, причем существенно — почти в 2 раза.

Распознавание, выполненное с использованием 30 зависимых показателей, оказалось точнее по основному показателю качества — площади под ROC-кривой, — чем в нашем предыдущем исследовании, когда использовали шесть показателей [9]. Но в предыдущей работе и цель была другая: исследовали влияние образовательных, социальных, информационных, медицинских услуг на показатель «расходы на информационные услуги

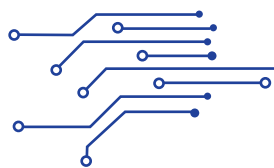


Таблица 3 / Table 3

**Нумерация и состав кластеров из регионов Российской Федерации по методу К-средних /
Numbering and Composition of Clusters from the Russian Federation Regions Using the K-means
Method**

Номер / Number		Название региона / Region name
Кластер / Cluster	Регион / Region	
1	18	Москва (красный кружок на рис. 2)
2	39, 66, 32, 21, 10	Республики Ингушетия, Тыва, Крым, Ненецкий автономный округ, Московская обл. (голубые квадраты)
3	31, 43, 30, 41, 38, 65, 46, 49, 50, 47	Республики Калмыкия, Чеченская, Адыгея, Карачаево-Черкессия, Дагестан, Алтай, Марий Эл, Удмуртия, Чувашия, Мордовия (синие ромбы)
4	84, 5, 14, 44, 58, 48, 34, 28, 40, 57, 42, 1, 8, 27, 6, 23, 17, 15, 4, 52, 73, 37, 56, 16	Еврейская автономная обл., Ивановская и Тамбовская обл., Ставропольский край, Ульяновская обл., Татарстан, Псковская обл., Кабардино-Балкария, Саратовская обл., Северная Осетия (Алания), Белгородская, Курская, Новгородская, Калужская, Ярославская, Тверская, Воронежская, Кировская, Омская обл., Севастополь, Самарская, Тульская обл. (зеленые кружочки)
5	75, 64, 11, 3, 13, 67, 54, 70, 74, 53, 24, 35, 69, 36, 7, 54, 71, 12, 2, 20, 51, 45, 68	Республика Бурятия, Челябинская, Орловская, Владимирская, Смоленская обл., Республика Хакасия, Пензенская, Томская, Иркутская, Нижегородская, Калининградская, Волгоградская обл., Красноярский край, Ростовская, Костромская, Оренбургская, Кемеровская, Рязанская, Брянская обл., Республика Коми, Пермский край, Башкирия, Алтайский край (оранжевые звездочки)
6	81, 63, 79, 19, 72, 59, 22, 9	Амурская, Тюменская обл., Приморский край, Карелия, Новосибирская, Курганская, Архангельская, Липецкая обл. (лиловые звездочки)
7	33, 76, 26, 83, 77, 80	Краснодарский край, Республика Саха (Якутия), Мурманская, Сахалинская обл., Забайкальский и Хабаровский края
8	82, 60	Магаданская и Свердловская обл. (коричневые крестики)
9	85, 62, 61, 78, 29	Чукотский, Ямало-Ненецкий, Ханты-Мансийский автономные округа, Камчатский край, Санкт-Петербург (красные крестики)
10	26	Ленинградская обл. (лиловый квадрат)

Источник / Source: составлено авторами / Complied by the authors.

Таблица 4 / Table 4

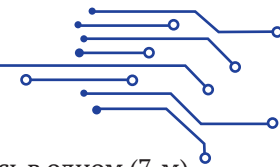
**Результаты классификации по методу статистически взвешенных синдромов /
The Results of Classification by the Method of Statistically Weighted Syndromes**

Правильность / Accuracy	Точность / Precision	Чувствительность / Sensitivity	Специфичность / Specificity	ROCAUC
0,78	0,86	0,67	0,89	0,90

Источник / Source: составлено авторами по результатам исследования / Complied by the authors based on the results of the study.

и сервисы в мае 2024 года». Неудивительно, что показатель чувствительности (табл. 4) для нашего исследования равен 0,67, в отличие от 0,78 для предыдущего исследования, в котором все показатели были однотипными и относились к одному временному периоду.

Как уже было отмечено выше, на сайте Росстата не все данные по удельному весу организаций, характеризующихся тем или иным уровнем цифровизации, были за один год (2023). За этот год были собраны данные на сайте Росстата в сборнике, опубликованном на сайте в 2024 г. Поскольку распозна-



вание только по 23 показателям было недостаточно качественным (в смысле показателей качества распознавания, представленных в табл. 2; данные не приводятся), то добавили информационные показатели ежемесячных трат населения за один выделенный месяц 2024 г.

Интересно посмотреть, изменятся ли существенно результаты исследования, если из анализа исключить влияние двух показателей за 2019 г. (организации, использовавшие глобальные информационные сети, и учреждения, применявшие системы электронного документооборота) на результирующий признак — уровень платных образовательных услуг населению за июль 2024 г. Показатель РОК АУК при этом не изменился. Его значение — 0,90123, что свидетельствует в пользу использования методов машинного обучения для классификации. Ошибки распознавания при использовании контрольной выборки в 20% также остались на прежнем уровне: алгоритм ошибочно отнес Калужскую, Иркутскую области и Якутию ко второму классу, а Воронежскую область — к первому классу.

Остальные результаты не приводятся, так как цель исследования заключалась в поиске наиболее информативных показателей, влияющих на платные ежемесячные образовательные услуги, — на примере зависимости 31 показателя от выбранного зависимого показателя. Месяц июль для группирующего показателя тоже был выбран неслучайно, так как к этому времени уже закончился образовательный процесс в школах, учреждениях СПО и высшего образования. Представлял интерес анализ трат населения на образовательные услуги, так как эти траты в основном были связаны с расходами на повышение квалификации в информационной сфере, цифровизацию разных услуг и использование искусственного интеллекта.

Полученное значение показателя ROC AUC, равное 0,9 (близкое к 1), свидетельствует в пользу использования методов машинного обучения для классификации регионов по показателям, важным для развития цифровизации во всех сферах человеческой жизни.

Отметим, что выполненный в исследовании кластерный анализ продемонстрировал интересные результаты: регионы, отличающиеся по основным экономическим показателям, оказались в одном кластере. Например, Чеченская Республика, Республика Адыгея и Московская область — в 5-м кластере. Санкт-Петербург оказался в одном кластере с самой западной областью РФ — Калининградской — и самыми восточными областями — Магаданской и Сахалинской (10-й кластер). При этом близкие по уровню развития республики Поволжья — Марий Эл,

Мордовия, Башкортостан — оказались в одном (7-м) кластере. Эти результаты свидетельствуют о том, что географический фактор очень важен для развития цифровизации в разных сферах человеческой жизни. Результаты, полученные при помощи метода К-средних (табл. 3), не противоречат этому выводу.

ВЫВОДЫ

Нельзя забывать, что интеллектуальный анализ данных требует больших затрат энергии. Такие факторы, как географический и энергетический, находятся в поле зрения не только властей такой протяженной страны, как Российская Федерация, но и других стран, в частности таких, у которых есть сходные по климату географические зоны. В связи с этим представляет интерес статья [14], в которой было показано, что децентрализация, приватизация и цифровизация — это движущие силы на местном и национальном уровнях. Они меняют распределение инвестиций и расстановку приоритетов.

Полученные в работе результаты свидетельствуют, что регионы по услугам цифровизации не сильно отличаются друг от друга, на что указывает большое количество кластеров. Это подчеркивает, что в настоящее время практически во всех регионах Российской Федерации доступны услуги интернета, в том числе в сфере образования. В статье [15] описано использование иерархической кластеризации для поиска и выявления аномальных наблюдений, что важно для возможности управления развитием социально-экономических систем. Работа интересна и тем, что в ней применяются методы иерархической и классической классификации.

Обсуждая результаты кластерного анализа, полученные в настоящей работе, можно сделать вывод, что методы иерархической кластеризации и метод К-средних не противоречат друг другу. Разумеется, если представить диаграмму для пары других признаков, используемых для кластерного анализа по методу К-средних, то изменится и наполнение кластеров (табл. 3), поэтому иерархическая кластеризация предпочтительнее для анализа с большим числом показателей.

Отметим, что нормализация показателей в кластерном методе проводится.

Методы машинного обучения позволяют выявить наиболее информативные признаки, используемые не для кластеризации, а для классификации изучаемых объектов, в частности регионов РФ по группирующему признаку — затраты на образовательные услуги. В работе получено распознавание с высоким качеством (ROC AUC = 0,9), что свидетельствует в пользу использования данных методов для количественного анализа цифровизации в сфере



услуг населению. Самыми информативными признаками стали абсолютные количественные, а не удельные показатели.

При проверке разделения всех регионов на две группы использование теста Манна–Уитни–Уилкоксона продемонстрировало наименьшие значения основного статистического показателя p -value (0,005 и меньше) для следующих показателей: платные телекоммуникационные услуги, услуги физической культуры и спорта, фитнес-центров и спортивных клубов, медицинские и ветеринарные услуги населению за июль 2024 г. Также статистически значимым оказался единственный удельный показатель — процент организаций, использующих серверы.

Интересно отметить, что ежемесячный среднедушевой доход и ежемесячные среднедушевые расходы не вошли в число наиболее значимых показателей, что свидетельствует в пользу того, что человек — объект безразмерной бесконечности, и цифровизация не в состоянии с вероятностью, близкой к 1, выявить набор самых важных показателей, так как за рамками исследования всегда остается много качественных индикаторов, которые невозможно формализовать.

Наибольшее значение p -value (0,7) было для показателя — процент организаций, использующих технологии искусственного интеллекта, что говорит о недостаточном внедрении методов машинного обучения, основанных на регрессионных моделях. Разумеется, в будущем данные методы найдут применение при анализе массивов данных, не превышающих 100 объектов. Так же широко распространились нейросетевые методы при анализе выборок с числом элементов, превышающим миллион объектов, для качества распознавания, близкого к 100% (ROC AUC $\approx$ 1).

Полученные результаты свидетельствуют о необходимости комплексного подхода к цифровизации образовательных услуг, учитывающего как количественные, так и качественные показатели развития цифровой инфраструктуры.

Проведенный в статье анализ цифровизации услуг на территории Российской Федерации, включая образовательные услуги населению, подтвердил, что цифровизация в целом оказывает положитель-

ное влияние на различные сферы жизни, а цифровизация образовательных услуг обеспечивает доступность обучения и его качество для всех категорий граждан, независимо от их географического положения и финансовых возможностей. Это особенно важно для удаленных и экономически слабо развитых регионов. Исследование показало, что регионы с высоким уровнем цифровизации образовательных услуг характеризуются широким применением электронного документооборота; значительной долей организаций, использующих цифровые платформы; существенным уровнем доступа населения к интернету.

Для оптимизации процесса внедрения цифровых решений в сфере образовательных услуг во всех регионах страны необходимо нарастить и форсировать развитие инфраструктуры широкополосного доступа в интернет в отдаленных регионах; поощрять освоение цифровых технологий в образовательных учреждениях; разработать программы профессионального совершенствования для педагогов в области цифровых технологий; задействовать методы машинного обучения для анализа и прогнозирования спроса на образовательные услуги.

Метод статистически взвешенных синдромов не требует проверки используемых данных на нормальный закон распределения, а также проверки мультиколлинеарности и исключения из дальнейшего анализа сильно коррелируемых признаков, поэтому может использоваться при анализе безразмерных и размерных величин при цифровизации образовательных услуг в дальнейшем.

Интересно, что выполненный на тестовой выборке контроль полученного решающего правила (модели) показал неустойчивое положение с точки зрения влияния всего комплекса анализируемых относительных и абсолютных признаков на показатель платных услуг в области образования в четырех совершенно разных с точки зрения географии и экономического развития регионах — Якутии, Калужской, Иркутской и Воронежской областях.

Полученный результат свидетельствует в пользу дальнейшего использования методов машинного обучения в области цифровизации.

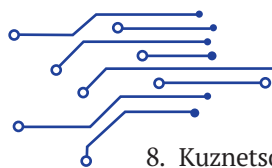
СПИСОК ИСТОЧНИКОВ

1. Сафонов С.А., Зандер Е.В. Оценка цифровой трансформации социальной сферы и региональные особенности цифровизации социальных услуг. *Цифровая экономика*. 2022;(4):53–62. URL: <https://doi.org/10.34706/DE-2022-04-07>
2. Кучумов И.В. Теоретические основы методов кластеризации данных в интеллектуальном анализе. *Цифровая экономика*. 2024;(1):69–77. URL: <https://doi.org/10.34706/DE-2024-01-08>.
3. Афан А.И., Полозова Д.В., Гордеева А.А. Цифровая трансформация государственной системы здравоохранения России: возможности и противоречия. *Цифровое право*. 2021;2(4):20–39. URL: <https://doi.org/10.38044/2686-9136-2021-2-4-20-39>

4. Нырков А.С. Влияние цифровизации на доступность и качество социальных услуг в Российской Федерации. *Молодой ученый*. 2025;20(571):522-524. URL: <https://www.elibrary.ru/xsfosr>
5. Назаров В.Л., Жердев Д.В., Буйначева А.В. Актуальные проблемы цифровой трансформации среднего образования. *Образование и наука*. 2023;25(4):109-166. URL: <https://doi.org/10.17853/1994-5639-2023-4-109-166>
6. Тищенко А.С., Токарева Г.С. Цифровая трансформация общего образования: анализ изменений. *Вестник Российского университета дружбы народов. Серия: Информатизация образования*. 2026;23(1):105-119. URL: <http://doi.org/10.22363/2312-8631-2026-23-1-105-119>
7. Сенько О.В., Кузнецова А.В., Воронин Е.М., Кравцова О.А., Борисова Л.Р., Кирилюк И.Л., и др. Методы интеллектуального анализа данных в исследованиях эпидемии COVID-19. *Журнал Белорусского государственного университета. Математика. Информатика*. 2022;1:83-96. URL: <https://doi.org/10.33581/2520-6508-2022-1-83-96>
8. Кузнецова Ю.О., Борисова Л.Р., Кузнецова А.В., Сенько О.В. Прозрачный интерфейс для прогноза в машинном обучении. В кн.: *Аналитика и управление данными в областях с интенсивным использованием данных*. М.: Федеральный исследовательский центр «Информатика и управление» Российской академии наук; 2017:493-495. URL: <https://elibrary.ru/ougrxt>
9. Кузнецова А.В., Борисова Л.Р., Демина И.А. Методы машинного обучения для прогнозирования течения заболевания на примере развития тяжелой степени пневмонии. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(1):6-19. URL: <https://www.elibrary.ru/rzxxxy>
10. Матвиевский С.С., Борисова Л.Р. Кластеризация стран Азиатско-Тихоокеанского региона по значениям инклюзивного экономического роста. *Вестник университета*. 2023;1:112-121. URL: <https://doi.org/10.26425/1816-4277-2024-1-112-121>
11. Борисова Л.Р. Сравнительный анализ регионов Российской Федерации методами машинного обучения по набору показателей электронных услуг и сервисов. *Цифровая социология*. 2024;7(40):33-43. URL: <https://doi.org/10.26425/2658-347X-2024-7-4-33-43>
12. Репина С.И. Проверка качества кластеров с помощью силуэтного анализа. *Экономика и социум*. 2024;9(124):958-975. URL: <https://doi.org/10.5281/zenodo.13918451>
13. Rousseeuw P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987;20(1):53-65. URL: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
14. Romanello M. et al. The 2022 report of the Lancet Countdown on health and climate change: health at the mercy of fossil fuels. *The Lancet 400 (suppl)*. 2022. URL: [https://doi.org/10.1016/S0140-6736\(22\)01540-9](https://doi.org/10.1016/S0140-6736(22)01540-9)
15. Кисляков А.Н., Поляков С.В. Иерархические методы кластеризации в задаче поиска аномальных наблюдений на основе групп с нарушенной симметрией. *Управленческое консультирование*. 2020;5:116-127. URL: <https://doi.org/10.22394/1726-1139-2020-5-116-127>

REFERENCES

1. Safonov S.A., Zander E.V. Assessment of digital transformation of the social sphere and regional features of digitalization of social services. *The Digital Economy*. 2022;(4):53-62. (In Russ.). URL: <https://doi.org/10.34706/DE-2022-04-07>
2. Kuchumov I.V. Theoretical foundations of data clustering methods in intellectual analysis. *The Digital Economy*. 2024;(1):69-77. (In Russ.). URL: <https://doi.org/10.34706/DE-2024-01-08>
3. Afyan A.I., Polozova D.V., Gordeeva A.A. Digital transformation of the Russian state healthcare system: opportunities and contradictions. *Digital Law*. 2021;2(4):20-39. (In Russ.). URL: <https://doi.org/10.38044/2686-9136-2021-2-4-20-39>
4. Nyrkov A.S. The impact of digitalization on the availability and quality of social services in the Russian Federation. *A Young Scientist*. 2025;20(571):522-524. (In Russ.). URL: <https://www.elibrary.ru/xsfosr>
5. Nazarov V.L., Zherdev D.V., Buinacheva A.V. Current problems of digital transformation in secondary education. *The Education and science journal*. 2023;25(4):109-166. (In Russ.) URL: <https://doi.org/10.17853/1994-5639-2023-4-109-166>
6. Tishchenko A.S., Tokareva G.S. Digital Transformation of General Education: Analysis of Changes. *RUDN Journal of Informatization in Education*. 2026;23(1):105-119. (In Russ.). URL: <http://doi.org/10.22363/2312-8631-2026-23-1-105-119>
7. Senko O.V., Kuznetsova A.V., Voronin E.M., Kravtsova O.A., Borisova L.R., Kirilyuk I.L., et al. Methods of data mining in research on the COVID-19 epidemic. *Journal of the Belarusian State University. Mathematics. Computer science*. 2022;1:83-96. (In Russ.). URL: <https://doi.org/10.33581/2520-6508-2022-1-83-96>



8. Kuznetsova Yu.O., Borisova L.R., Kuznetsova A.V., Senko O.V. Transparent interface for forecasting in machine learning. In: Data Analytics and Management in Data-intensive Areas. Moscow: Federal Research Center "Informatics and Management" of the Russian Academy of Sciences; 2017:493-495. (In Russ.). URL: <https://elibrary.ru/ougrxt>
9. Kuznetsova A.V., Borisova L.R., Demina I.A. Machine learning methods for predicting the course of the disease using the example of the development of severe pneumonia. *Digital Solutions and Artificial Intelligence Technologies*. 2025;1(1):6-19. (In Russ.). URL: <https://www.elibrary.ru/rzxxxyo>
10. Matviyevsky S.S., Borisova L.R. Clusterization of the countries of the Asia-Pacific region according to the values of inclusive economic growth. *Bulletin of the University*. 2023;1:112-121. (In Russ.). URL: <https://doi.org/10.26425/1816-4277-2024-1-112-121>
11. Borisova L.R. Comparative analysis of the Russian regions using machine learning methods for a set of indicators of electronic services. *Digital Sociology*. 2024;7(40):33-43. (In Russ.). URL: <https://doi.org/10.26425/2658-347X-2024-7-4-33-43>
12. Repina S.I. Quality verification of clusters using silhouette analysis. *Economics and Society*. 2024;9(124):958-975. (In Russ.). URL: <https://doi.org/10.5281/zenodo.13918451>
13. Rousseeuw P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987;20(1):53-65. (In Russ.). URL: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
14. Romanello M. et al. The 2022 report of the Lancet Countdown on health and climate change: health at the mercy of fossil fuels. *The Lancet 400 (suppl)*. 2022. URL: [https://doi.org/10.1016/S0140-6736\(22\)01540-9](https://doi.org/10.1016/S0140-6736(22)01540-9)
15. Kislyakov A.N., Polyakov S.V. Hierarchical clustering methods in the problem of searching for anomalous observations based on groups with broken symmetry. *Management Consulting*. 2020;5:116-127. (In Russ.). URL: <https://doi.org/10.22394/1726-1139-2020-5-116-127>

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Людмила Робертовна Борисова — кандидат физико-математических наук, доцент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Lyudmila R. Borisova — Cand. Sci. (Phys. and Math.) Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation
<http://orcid.org/0000-0002-5757-0341>

Автор для корреспонденции / Corresponding author:
lrborisova@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

Mira N. Fridman — Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation
<http://orcid.org/0000-0001-9314-5986>
mnfridman@fa.ru

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Статья поступила 16.02.2026; после рецензирования 02.07.2026; принята к публикации 31.07.2026.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was submitted on 16.02.2026; revised on 02.07.2026 and accepted for publication on 31.07.2026.

The authors read and approved the final version of the manuscript.