



## Типология регионов России по образовательному профилю

К. С. Краюшкин

Аспирант,

[ks.krayushkin@gmail.com](mailto:ks.krayushkin@gmail.com)

Академия труда и социальных отношений,  
Москва, Россия

**Аннотация:** Современная система высшего образования Российской Федерации характеризуется выраженной региональной гетерогенностью, обусловленной различиями в неравномерном распределении по регионам направлений подготовки кадров, соответствующих местным социально-экономическим потребностям. В статье представлены результаты кластерного анализа, проведенного на основе данных федерального статистического наблюдения ВПО-1 за 2024 год по всем 83 субъектам РФ и 56 направлениям подготовки.

С применением метода главных компонент (РСА) и алгоритма *k-means* выявлена устойчивая двухкластерная структура: первый кластер объединяет научно-технические и мультидисциплинарные центры, ориентированные на STEM-дисциплины и креативные индустрии; второй — периферийные регионы с практико-ориентированным профилем, где доминируют социально-гуманитарные и локально значимые направления.

Показано, что специальности в области информационных технологий не формируют отдельного «цифрового» кластера, а органично встроены в более широкую научно-техническую экосистему. Результаты подтверждают необходимость дифференцированного подхода к региональной образовательной политике.

**Ключевые слова:** региональная дифференциация, высшее образование, кластерный анализ, РСА, образовательный профиль.

**Для цитирования:** Краюшкин К.С. Типология регионов России по образовательному профилю. Путеводитель предпринимателя. 2026. Т. 19. № 2. С. 61–67. <https://doi.org/10.24182/2073-9885-2026-19-2-61-67>.

## A typology of Russian regions by educational profile

K. S. Krayushkin

Postgraduate student,

[ks.krayushkin@gmail.com](mailto:ks.krayushkin@gmail.com)

Academy of Labour and Social Relations,  
Moscow, Russia

**Abstract:** The contemporary system of higher education in the Russian Federation exhibits pronounced regional heterogeneity, driven by differences in the structure of specialist training. This article presents the results of a cluster analysis based on the 2024 Federal Statistical Observation VPO-1, covering all 83 federal subjects and 56 fields of study.

*Using Principal Component Analysis (PCA) and the k-means algorithm, a stable two-cluster structure was identified: the first cluster comprises scientific-technical and multidisciplinary centers, focused on STEM disciplines and creative industries; the second includes peripheral regions with a practice-oriented profile, dominated by socio-humanitarian and locally relevant fields.*

*The study demonstrates that Information Technology (IT) fields do not constitute a separate «digital» cluster but are inherently embedded within a broader scientific-technical ecosystem. These findings underscore the need for a differentiated approach to regional higher education policy.*

**Keywords:** regional differentiation, higher education, cluster analysis, PCA, educational profile.

**For citation:** Krayushkin K.S. A typology of Russian regions by educational profile. *Entrepreneur's Guide*. 2026. T. 19. № 2. P. 61–67. <https://doi.org/10.24182/2073-9885-2026-19-2-61-67>.

**Введение и обзор научного опыта.** Современная система высшего образования Российской Федерации характеризуется выраженной региональной гетерогенностью, проявляющейся не только в масштабах образовательной инфраструктуры, но и в качественной структуре подготовки специалистов.

Как известно, одномерные показатели такие, как индекс энтропии Шеннона или индекс Херфиндаля–Хиршмана<sup>1</sup>, позволяют количественно оценить степень диверсификации или концентрации, но они не раскрывают содержательную природу специализации регионов. Два субъекта могут демонстрировать схожий уровень разнообразия, но принципиально различаться по профилю: первый — за счёт сбалансированного сочетания STEM-дисциплин, второй — за счёт доминирования социально-гуманитарных направлений. В этих условиях возникает необходимость перехода от агрегированных метрик к многомерной структурной типологии, способной выявить устойчивые и интерпретируемые модели регионального высшего образования.

Настоящее исследование направлено на выявление таких моделей с помощью кластерного анализа, применённого к полным данным формы федерального статистического наблюдения ВПО-1 за 2024 год по всем 83 субъектам РФ и 56 детализированным направлениям подготовки. Использование метода главных компонент (РСА) и алгоритма k-means, дополненного валидацией на основе коэффициента силуэта, позволило выделить содержательно значимые кластеры, отражающие фундаментальные различия в образовательных стратегиях регионов — от мультидисциплинарных научно-технических центров до практико-ориентированных периферийных систем.

В основе эмпирической базы исследования лежат официальные статистические данные, в первую очередь, форма федерального наблюдения ВПО-1, представляющая собой комплексный источник информации о деятельности высших учебных заведений. В ней фиксируются ключевые параметры функционирования вузов: контингент студентов, профессорско-преподавательский состав, перечень реализуемых образовательных программ, сведения о финансировании и кадровом обеспечении. Подобные массивы информации широко используются в академических работах для диагностики территориальной неоднородности и анализа структурных тенденций в системе высшего образования.

Так, ряд работ посвящен анализу дифференциации на основе плотности вузов и числа студентов относительно рабочего возраста, что позволяет получить представление о распространенности системы ВПО в разных регионах<sup>2</sup>.

Другие исследователи применяют более комплексные подходы, используя множество социально-экономических показателей из баз данных Росстата, для кластеризации регионов и выявления их типологических различий<sup>3</sup>.

<sup>1</sup> Shannon C.E. A mathematical theory of communication / C.E. Shannon. The Bell system technical journal. 1948. T. 27. № 3. С. 379–423. Rhoades S.A. The herfindahl-hirschman index / S.A. Rhoades. Fed. Res. Bull. 1993. T. 79. С. 188.

<sup>2</sup> Regional differentiation of higher education in Russian regions in 2020 / I.S. Pinkovetskaia [и др.]. Revista de la Universidad del Zulia. 2021. T. 12. № 35. С. 428–445.

<sup>3</sup> Ketova K.V. Clustering Russian Federation regions according to the level of socio-economic development with the use of machine learning methods / K.V. Ketova, E.V. Kasatkina, D.D. Vavilova. Ekonomicheskie i Sotsialnye Peremeny. 2021. T. 14. № 6. С. 70–85.

Такие работы позволяют не просто описать различия, но и связать их с общей экономической и социальной структурой региона. Особую ценность представляют исследования, предлагающие авторские методологии для расчета комплексных индексов, таких как «Индекс образовательной деятельности» или «Интегральный индекс развития региональной высшей школы», которые позволяют ранжировать регионы по совокупности параметров и выявлять лидеры и аутсайдеры<sup>4</sup>.

Также важен вклад исследований, анализирующих доступность образования, использующих данные лонгитюдных опросов (например, TrEC) для моделирования барьеров, с которыми сталкиваются абитуриенты в зависимости от их дохода, места проживания и пола<sup>5</sup>.

Зарубежные источники предоставляют богатый материал по методологии кластерного анализа и его валидации, что имеет прямое отношение к цели данного исследования. Исследования, сравнивающие различные алгоритмы (k-means, k-medoids, agglomerative hierarchical clustering), помогают выбрать наиболее подходящий метод для конкретной задачи.

Например, один из таких анализов показал, что иерархический агломеративный метод с правилом Уорда может иметь лучшую предсказательную силу для оценки конечных результатов обучения, чем k-means.<sup>6</sup> Важнейшим выводом является необходимость двухэтапного подхода: сначала использование иерархического метода для определения оптимального числа кластеров, а затем применения k-means для получения окончательного разбиения. Кроме того большое внимание уделяется процедурам валидации результатов: коэффициент силуэта (Silhouette Coefficient) предлагается как надежная метрика для оценки качества кластеризации, поскольку она показывает, насколько объект лучше соответствует своему собственному кластеру, чем любому другому. График силуэта позволяет не только найти оптимальное число кластеров, но и визуально оценить качество каждого из них, выявляя «слабые» или «плохо сгруппированные» кластеры<sup>7</sup>.

Для данных временных рядов критически важно проводить статистическую проверку, чтобы убедиться, что выявленная структура не является случайной («spurious clustering»). Методология на основе Monte Carlo симуляций, в которой реальные данные сравниваются с множеством искусственных наборов данных, генерируемых по известным статистическим законам (например, белый шум или случайное блуждание), является наиболее надежным способом такой проверки<sup>8</sup>.

В отличие от предшествующих исследований, опиравшихся на укрупненные категории специальностей или ограниченную выборку регионов, проводимое исследование использует полные и актуальные данные формы федерального статистического наблюдения ВПО-1 за 2024 год по всем 83 субъектам РФ и 56 детализированным направлениям подготовки. Это позволяет не только избежать агрегационного смещения, но и выявить скрытые паттерны специализации.

**Методология исследования.** Как упоминалось ранее, одномерные индексы (индекс энтропии Шеннона или индекс Херфиндала–Хиршмана), несмотря на свою интерпретируемость, обладают принципиальным ограничением: они сводят многокомпонентную структуру образовательного ландшафта к единственному числу, игнорируя качественное содержание специализации субъекта.

<sup>4</sup> Долгих Е.А. Статистическое изучение региональных различий образовательного потенциала населения / Е.А. Долгих, Т.А. Першина. МИР (Модернизация. Инновации. Развитие). 2024. Т. 15. № 4. С. 558–575.

<sup>5</sup> Anfalova A.V. Inequality in Access to Higher Education in Russia / A.V. Anfalova. Population and Economics. 2025. Т. 9. № 2. С. 86–116. Bugakova P. Regional Accessibility of Higher Education in Russia / P. Bugakova, I. Prakhov. Higher School of Economics Research Paper No. WP BRP. 2020. Т. 58.

<sup>6</sup> Luo Y. Application and Comparison of Clustering Methods to Educational Process Data / Y. Luo. University of Washington, 2022.

<sup>7</sup> Репина С.И. ПРОВЕРКА КАЧЕСТВА КЛАСТЕРОВ С ПОМОЩЬЮ СИЛУЭТНОГО АНАЛИЗА / С.И. Репина. Экономика и социум. 2024. № 9 (124). С. 958–975.

<sup>8</sup> Кирилюк И.Л. Оценка качества кластеризации панельных данных с использованием методов Монте-Карло (на примере данных российской региональной экономики) / И.Л. Кирилюк, О.В. Сенько. Компьютерные исследования и моделирование. 2020. Т. 12. № 6. С. 1501–1513. Репина С.И. ПРОВЕРКА КАЧЕСТВА КЛАСТЕРОВ С ПОМОЩЬЮ СИЛУЭТНОГО АНАЛИЗА / С.И. Репина. Экономика и социум. 2024. № 9 (124). С. 958–975. Luo Y. Application and Comparison of Clustering Methods to Educational Process Data / Y. Luo. University of Washington, 2022. Higher Education: Handbook of Theory and Research : Higher Education: Handbook of Theory and Research. Vol. 20. Higher Education / ed. J.C. Smart. — Berlin / Heidelberg: Springer-Verlag, 2005.

Действительно, два региона могут иметь схожие значения энтропии, но принципиально различаться по профилю: один — за счет сбалансированного сочетания ИТ, инженерии и медицины, другой — за счет равномерного распределения между педагогикой, юриспруденцией и социально-гуманитарными дисциплинами. Аналогично, высокий индекс Херфиндаля–Хиршмана может отражать как технологическую специализацию (например, в атомной или аэрокосмической отрасли), так и зависимость от одного массового направления (например, «Педагогическое образование» в малом регионе).

В связи с этим для выявления устойчивых и содержательно интерпретируемых типов регионов был применен многомерный кластерный анализ, основанный на векторе долей студентов по 56 детализированным направлениям подготовки. Такой подход позволяет перейти от агрегированных метрик к структурной типологии, учитывающей не только степень разнообразия, но и композицию образовательного профиля.

Учитывая высокую размерность признакового пространства (56 переменных) и возможную мультиколлинеарность между направлениями, для улучшения интерпретируемости и устойчивости кластеризации был применен метод главных компонент (Principal Component Analysis, PCA). Анализ показал, что 22 главных компоненты объясняют более 90% общей дисперсии исходных данных, что было признано достаточным компромиссом между сохранением информации и сокращением шума. Дальнейшая кластеризация проводилась в пространстве этих 22 компонент.

Для определения оптимального числа кластеров был использован метод силуэтов (silhouette method), который оценивает степень компактности и разделимости кластеров<sup>9</sup>. Силуэт каждого объекта вычисляется на основе среднего внутрикластерного расстояния и минимального среднего межкластерного расстояния; его значение лежит в диапазоне от  $-1$  до  $+1$ , где более высокие значения указывают на лучшее качество кластеризации.

Расчет среднего коэффициента силуэта для числа кластеров от 2 до 10 показал, что максимальное значение достигается при  $k = 2$ , что свидетельствует в пользу двухкластерной структуры, как наиболее устойчивой и содержательно обоснованной.

В качестве основного алгоритма кластеризации был выбран  $k$ -means — классический метод разбиения, минимизирующий сумму квадратов расстояний от точек до центроидов их кластеров. Несмотря на известные ограничения (чувствительность к выбросам, предположение о сферичности кластеров),  $k$ -means показал высокую воспроизводимость результатов и согласованность с содержательной интерпретацией профилей регионов. Инициализация центроидов выполнялась с использованием метода  $k$ -means++, что позволило избежать неустойчивых локальных минимумов.

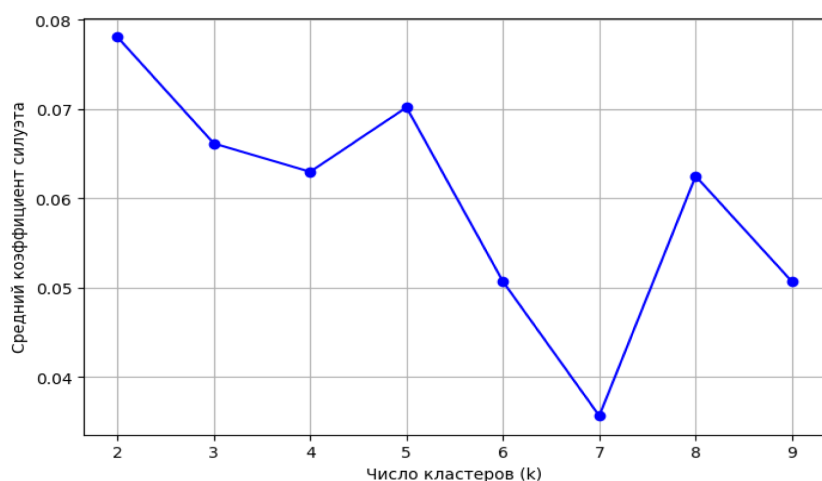


Рис. 1. График среднего коэффициента силуэта в зависимости от числа кластеров<sup>10</sup>

Полученные кластеры были подвергнуты содержательной интерпретации на основе сравнения средних значений долей по направлениям подготовки, а также с учетом социально-экономического и географического контекста регионов (рис. 1).

Рассмотрим получившиеся кластеры (табл. 1).

Первый кластер (43 региона) объединяет крупнейшие научно-образовательные центры страны: города федерального значения (Москва, Санкт-

<sup>9</sup> Rousseeuw P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis / P.J. Rousseeuw. Journal of computational and applied mathematics. 1987. Т. 20. Silhouettes. С. 53–65.

<sup>10</sup> Разработано автором.

Таблица 1

Сравнительная характеристика кластеров модели высшего образования<sup>11</sup>

УГСН	Кластер 1	Кластер 2	Разница
Информатика и вычислительная техника	0.057	0.017	+ 235%
Компьютерные и информационные науки	0.019	0.005	+ 280%
Информационная безопасность	0.021	0.004	+ 425%
Электроника, радиотехника	0.018	0.006	+ 200%
Математика и механика	0.022	0.008	+ 175%

Петербург), а также регионы с сильной исследовательской и инновационной инфраструктурой — Новосибирскую, Томскую, Свердловскую, Нижегородскую области и Республику Татарстан.

Для этого кластера характерна высокая доля обучающихся по направлениям, связанным с точными и естественными науками, информационными технологиями и инженерными дисциплинами. В частности, средняя доля студентов по направлению «Информатика и вычислительная техника» (09.00.00) в этом кластере в 3,4 раза выше, чем в кластере 1 (5,7% против 1,7%). Аналогичная картина наблюдается по таким направлениям, как «Компьютерные и информационные науки» (02.00.00), «Информационная безопасность» (10.00.00), «Математика и механика» (01.00.00) и «Электроника, радиотехника и системы связи» (11.00.00). Кроме того в данных регионах значительно выше представленность клинической и фундаментальной медицины, что отражает наличие крупных медицинских университетов и научно-клинических комплексов.

Важно отметить, что в кластере 1 также наблюдается высокая доля по элитарным и творческим направлениям — «Дизайн» (54.00.00), «Экранные искусства» (55.00.00), «Музыка» (53.00.00), что свидетельствует о концентрации культурных и креативных индустрий в мегаполисах.

Таким образом, кластер 1 представляет собой интегрированную научно-техническую и мультидисциплинарную экосистему, где высшее образование ориентировано на фундаментальные исследования, цифровые технологии и инновационное развитие.

Второй кластер включает 40 субъектов РФ, преимущественно периферийные и малонаселенные регионы, включая Дальний Восток, Северный Кавказ и часть Центральной России. Для этих территорий характерна иная структура высшего образования.

Во-первых, в кластере 2 доминируют социально-гуманитарные и традиционные прикладные направления: «Педагогическое образование» (44.00.00) — 8,9% против 5,5% в кластере 0, «Юриспруденция» (40.00.00) — 6,3% против 3,8%, «Филология» (45.00.00) — 4,1% против 2,6%. Во-вторых, значительную роль играют аграрные и локально значимые специальности: «Агрономия» (35.00.00) представлена почти вдвое чаще, чем в кластере 0.

В ряде регионов (например, Чукотский АО, Магаданская область, Еврейская АО) наблюдается крайне узкая специализация: доля одного-двух направлений превышает 50–60% от общего объема подготовки. Это указывает на зависимость региональной системы высшего образования от одного вуза или филиала, ориентированного на решение локальных кадровых задач.

Следовательно, кластер 2 отражает практико-ориентированную, социально-адаптированную модель высшего образования, сфокусированную на обеспечении базовых кадровых потребностей региона и воспроизводстве традиционных профессий.

Географическое распределение кластеров по субъектам РФ приведено на рисунке 2.

Несмотря на изначальное ожидание выделения отдельного ИТ-кластера, анализ показал, что направления, связанные с информационными технологиями, не существуют изолированно. Они органично встроены в более широкий профиль научно-технических центров и тесно коррелируют с другими STEM-дисциплинами (области науки (физика и др.), технологии, инженерии и математики). Это позволяет сделать вывод: ИТ-специальности в России концентрируются не в отдельных «цифровых» регионах, а в комплексных научно-образовательных центрах, где обеспечивается междисциплинарная синергия.

<sup>11</sup> Разработано автором.



Рис. 2. Географическое распределение кластеров по субъектам РФ<sup>12</sup>

**Выводы и обсуждение результатов.** Кластерный анализ, проведенный в пространстве главных компонент на основе долей студентов по детализированным направлениям, выявил устойчивую двухкластерную структуру, подтвержденную максимальным значением коэффициента силуэта. Первый кластер объединяет регионы с развитой STEM- и креативной инфраструктурой, где доминируют точные науки, информационные технологии, инженерия и элитарные гуманитарные дисциплины. Второй кластер включает преимущественно периферийные территории с практико-ориентированным профилем, где преобладают социально-гуманитарные направления (педагогика, юриспруденция, филология) и локально значимые специальности (агрономия, социология). Важно отметить, что ИТ-направления не формируют отдельного кластера, а органично встроены в более широкий профиль научно-технических центров, что свидетельствует о междисциплинарной природе цифровой специализации в российском контексте.

Полученные результаты подтверждают гипотезу о том, что высшее образование в России не является однородной системой, а представляет собой совокупность региональных моделей, тесно связанных с социально-экономическим, демографическим и институциональным контекстом. Концентрация образовательных, научных и креативных ресурсов в федеральных столицах и крупных научных центрах создаёт устойчивый «якорь диверсификации», в то время как периферийные регионы оказываются в положении структурной уязвимости, зависящей от узкого набора профессий и ограниченной мобильности кадров.

Таким образом, политика в области высшего образования должна учитывать эту поляризацию и избегать универсальных решений. Для мультидисциплинарных центров целесообразно развивать междисциплинарные и международные программы, укреплять связи с инновационной экономикой. Для периферийных регионов — поддерживать базовую образовательную инфраструктуру, способствовать диверсификации профилей подготовки и создавать механизмы компенсации структурных дисбалансов (например, через дистанционное обучение, мобильность студентов и преподавателей, региональные образовательные консорциумы).

#### Список литературы

1. Долгих Е.А. Статистическое изучение региональных различий образовательного потенциала населения / Е.А. Долгих, Т.А. Першина. МИР (Модернизация. Инновации. Развитие). 2024. Т. 15. № 4. С. 558–575.

<sup>12</sup> Составлено автором.

2. Кирилук И.Л. Оценка качества кластеризации панельных данных с использованием методов Монте-Карло (на примере данных российской региональной экономики) / И.Л. Кирилук, О.В. Сенько. Компьютерные исследования и моделирование. 2020. Т. 12. № 6. С. 1501–1513.
3. Репина С.И. Проверка качества кластеров с помощью силуэтного анализа / С.И. Репина. Экономика и социум. 2024. № 9 (124). С. 958–975.
4. Anfalova A.V. Inequality in Access to Higher Education in Russia / A.V. Anfalova. Population and Economics. 2025. Т. 9. № 2. С. 86–116.
5. Bugakova P. Regional Accessibility of Higher Education in Russia / P. Bugakova, I. Prakhov. Higher School of Economics Research Paper No. WP BRP. 2020. Т. 58.
6. Ketova K.V. Clustering Russian Federation regions according to the level of socio-economic development with the use of machine learning methods / K.V. Ketova, E.V. Kasatkina, D.D. Vavilova. Ekonomicheskie i Sotsialnye Peremeny. 2021. Т. 14. № 6. С. 70–85.
7. Luo Y. Application and Comparison of Clustering Methods to Educational Process Data / Y. Luo. University of Washington, 2022.
8. Regional differentiation of higher education in Russian regions in 2020 / I.S. Pinkovetskaia [и др.]. Revista de la Universidad del Zulia. 2021. Т. 12. № 35. С. 428–445.
9. Rhoades S.A. The herfindahl-hirschman index / S.A. Rhoades. Fed. Res. Bull. 1993. Т. 79. С. 188.
10. Rousseeuw P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis / P.J. Rousseeuw. Journal of computational and applied mathematics. 1987. Т. 20. Silhouettes. С. 53–65.
11. Shannon C.E. A mathematical theory of communication / C.E. Shannon. The Bell system technical journal. 1948. Т. 27. № 3. С. 379–423.
12. Higher Education: Handbook of Theory and Research : Higher Education: Handbook of Theory and Research. Vol. 20. Higher Education / Ed. J.C. Smart. – Berlin / Heidelberg: Springer-Verlag, 2005.

#### References

1. Dolgikh, E.A., & Pershina, T.A. (2024). Statistical Analysis of Regional Differences in the Educational Potential of the Population. MIR (Modernization. Innovations. Development), 15(4), 558–575.
2. Kirilyuk, I.L., & Senko, O.V. (2020). Assessing the Quality of Clustering Panel Data Using Monte Carlo Methods (with an Application to Russian Regional Economic Data). Computer Research and Modeling, 12(6), 1501–1513.
3. Repina, S.I. (2024). Evaluating Cluster Quality Using Silhouette Analysis. Economics and Society, 9(124), 958–975.
4. Anfalova A.V. Inequality in Access to Higher Education in Russia / A.V. Anfalova. Population and Economics. 2025. Т. 9. № 2. С. 86–116.
5. Bugakova P. Regional Accessibility of Higher Education in Russia / P. Bugakova, I. Prakhov. Higher School of Economics Research Paper No. WP BRP. 2020. Т. 58.
6. Ketova K.V. Clustering Russian Federation regions according to the level of socio-economic development with the use of machine learning methods / K.V. Ketova, E.V. Kasatkina, D.D. Vavilova. Ekonomicheskie i Sotsialnye Peremeny. 2021. Т. 14. № 6. С. 70–85.
7. Luo Y. Application and Comparison of Clustering Methods to Educational Process Data / Y. Luo. University of Washington, 2022.
8. Regional differentiation of higher education in Russian regions in 2020 / I.S. Pinkovetskaia [и др.]. Revista de la Universidad del Zulia. 2021. Т. 12. № 35. С. 428–445.
9. Rhoades S.A. The herfindahl-hirschman index / S.A. Rhoades. Fed. Res. Bull. 1993. Т. 79. С. 188.
10. Rousseeuw P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis / P.J. Rousseeuw. Journal of computational and applied mathematics. 1987. Т. 20. Silhouettes. С. 53–65.
11. Shannon C.E. A mathematical theory of communication / C.E. Shannon. The Bell system technical journal. 1948. Т. 27. № 3. С. 379–423.
12. Higher Education: Handbook of Theory and Research: Higher Education: Handbook of Theory and Research. Vol. 20. Higher Education / ed. J.C. Smart. – Berlin / Heidelberg: Springer-Verlag, 2005.

*Статья поступила в редакцию 12.02.2026; одобрена после рецензирования 16.03.2026; принята к публикации 23.03.2026.*

*The article was submitted 12.02.2026; approved after reviewing 16.03.2026; accepted for publication 23.03.2026.*