

# Применение методов кластеризации в экономическом анализе регионов

Application of clustering methods in the economic analysis of regions

doi 10.26310/2071-3010.2021.271.5.005



**Н. М. Логачева,**  
д. э. н., доцент, ведущий научный  
сотрудник, Челябинский филиал  
Института экономики УрО РАН  
✉ logacheva.nm@uiec.ru

**N. M. Logacheva,**  
doctor of economics, associate professor,  
leading researcher, Chelyabinsk branch  
of the Institute of economics, Ural branch  
of the Russian academy of sciences



**А. К. Петрова,**  
старший преподаватель, кафедра инновационного  
менеджмента, Институт инновационного  
проектирования и технологического  
предпринимательства, Санкт-Петербургский  
государственный электротехнический университет  
«ЛЭТИ» им. В. И. Ульянова (Ленина)  
✉ ak72p@yandex.ru

**A. K. Petrova,**  
senior teacher, department of innovative  
management, Institute of innovation design  
and technological entrepreneurship,  
Saint-Petersburg electrotechnical university «LETI»  
by Ulyanov (Lenin)

Технологическая и цифровая трансформация, проникающая во все сферы экономики, предъявляет требования к уровню готовности объектов социальной инфраструктуры. Цель статьи оценить готовность к цифровой трансформации инфраструктуры образования регионов, выявить дифференциацию регионов РФ методами машинного обучения. Проведена кластеризация регионов РФ по степени цифровой оснащенности образовательных организаций. Кластеризация осуществлена при помощи самоорганизующихся карт Кохонена и классических методов кластеризации, в частности метода Linkage. Полученные результаты показали возможность распределить регионы на четыре неравномерных кластера, имеющих существенные отличия. Отмечено, что в регионах РФ в среднем наиболее высокая обеспеченность компьютерами присутствует в организациях, осуществляющих образовательную деятельность по программам высшего образования, наименьшая – в школах. Данные исследования могут быть использованы для дальнейшего анализа и принятия решений, повышающих эффективность цифровизации регионов РФ, в том числе для поиска экономических решений в области развития инфраструктуры образования.

Technological and digital transformation, penetrating all economies, makes demands on the social infrastructure readiness level. The purpose of the article is to evaluate regions' education infrastructure readiness for digital transformation, to identify the differentiation of regions of the Russian Federation using machine learning methods. Clustering of the Russian Federation regions was carried out according to the degree of digital equipment of educational institutions. Clustering was carried out using self-organizing Kohonen maps and classical clustering methods, in particular the Linkage method. The results obtained showed the possibility of distributing the regions into four uneven clusters that have significant differences. It is noted that in the regions of the Russian Federation, the organizations carrying out educational activities in higher education programs have the highest computer availability, while the schools have the lowest one. These studies can be used for further analysis and decision-making that increase the efficiency of digital development of the regions of the Russian Federation, including for the search for economic solutions in the field of infrastructure development.

**Ключевые слова:** цифровая трансформация, кластеризация, дифференциация регионов, инфраструктура образования, машинное обучение, нейронные сети, самоорганизующиеся карты Кохонена.

**Keywords:** digital transformation, clustering, differentiation of regions, education infrastructure, machine learning, neural networks, self-organizing Kohonen maps.

Технологическая трансформация предъявляет требования не только к готовности людей ее воспринять, но и к социальным объектам, способным ее внедрять и предоставлять услуги нового формата и качества. Сектор образования, развивающий человеческий потенциал и влияющий на качество жизни населения, в наибольшей степени может способствовать запуску трансформационных процессов в других отраслях социальной сферы.

На данный момент «в России отсутствует система массового формирования компетенций использования высоких технологий. Нельзя модернизировать производство, если нет массового формирования умений пользоваться современными технологиями в образовательной и профессиональной деятельности» [1].

Еще в 2016 г. в Послании Федеральному собранию Президент России В. В. Путин указал на важнейшие направления: фундаментальная наука, исследовательская инфраструктура, IT-индустрия, собственные передовые разработки, цифровые технологии и предложил «запустить масштабную системную программу развития экономики нового технологического поколения — цифровой экономики» [2].

Обозначенные тенденции коррелируют с мировыми трендами в социально-экономическом развитии, представленными Клаусом Швабом в книге «Четвертая промышленная революция» [3]. Более детальная направленность для России была представлена и продолжает актуализироваться в Стратегии научно-технологического развития РФ [4].

При таком подходе очевидно, что инфраструктура образования (для всех его уровней) крайне актуальна, но она не просто «должна быть». Теперь значение имеет современная, оборудованная инфраструктура с кадрами, подготовленными для использования цифровых технологий.

В работах отечественных исследователей можно встретить разные подходы к изучению социальной инфраструктуры. Одни рассматривают ее, как важный элемент региональной экономики, влияющий на конкурентоспособность и устойчивость регионов [5-7], другие делают акцент на особенностях и возможности городских и сельских территориях [8, 9]. Иностранцы также подчеркивают многогранное воздействие социальной инфраструктуры как на человека, так и регион [10-12].

Развитие инфраструктуры является одним из самых дорогих направлений модернизации, при этом спрос на образовательную инфраструктуру будет расти, так как к 2024 г. количество детей и молодых людей в возрасте от 3 лет до 21 года (включительно) увеличится на 12%, а это значит, что со строительством объектов не стоит затягивать.

В этой связи выделяется несколько важных моментов, которые нужно учитывать:

- 1) новые здания должны обеспечить возможности для цифровизации образования, осуществления образовательной, проектной и исследовательской деятельности с применением современных технологий, а это значит необходимо сформулировать новые требования к современным зданиям и образовательным помещениям;
- 2) необходим поиск баланса между цифровыми и традиционными технологиями, основанными на реальных коммуникативных практиках (общение «лицом к лицу») и взаимодействии участников образовательного процесса [13];
- 3) проникновение цифровых технологий в школы, техникумы, вузы должно быть разным.

С одной стороны, исследование, проведенное с участием 9279 учителей и администраторов из 89 разных стран, выявило, что в подавляющем большинстве респонденты считают, что цифровое обучение положительно влияет на обучающихся и преподавателей [14]. С другой стороны, онлайн-занятия и цифровые «нововведения в системе школьного образования не обеспечили обещанных эффектов, так как по оценкам регулярно проводящегося международного исследования PISA результаты российских школьников год от года ухудшаются» [15], при этом отмечается «их эффективность на правах дополнительного инструмента обучения» [16].

Таким образом, в инфраструктуре образования должны использоваться новые технологии, цифровые и информационные ресурсы и возможности, но в разной степени в зависимости от уровня образования.

Так, например, в профессиональном образовании цифровые технологии способствуют переходу на принципиально новый уровень в освоении конкретных навыков с помощью симуляторов и различных тренажеров, а «технологии виртуальной реальности создают возможность использования цифровых тренажеров для освоения буквально любой профессии и профессиональной квалификации, в том числе не привязанной к конкретному рабочему месту» [1]. Отметим, что скорость изменений в производственных технологиях и используемых материалах на современном этапе делает модель технологического переоснащения организаций, осуществляющих образовательную деятельность по программам подготовки квалифицированных рабочих, служащих, по программам подготовки специалистов среднего звена, реальным оборудованием (т. е. без цифровых технологий симуляторов) содержательно и экономически нерациональной.

«Аналитики ABI Research прогнозируют, что объем рынка VR-технологий для профессионального обучения и тренировок составит... в 2022 г. ... до \$6,3 млрд. По их мнению, развитие и совершенствование

VR-гарнитур, сопутствующих аксессуаров и программного обеспечения приведет к тому, что среднегодовые темпы роста (CAGR) на данном направлении в ближайшие пять лет составят 140%» [17].

Что касается высшего образования, то «Россия занимает лидирующие позиции в мире по доле взрослого населения, имеющего высшее образование, а также по среднему количеству лет, проведенных человеком в системе формального образования. Однако, компетенции взрослого населения, имеющего высшее образование, недостаточно сильно отличаются от уровня компетенций населения с более низким образованием, что является уникальной ситуацией для стран ОЭСР (по данным исследования PIAAC)» [1].

В указе Президента РФ «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 г.» сразу четыре национальные цели развития непосредственно связаны с обновлением технологического образования в школе, СПО и вузах, а значит, не могут быть выполнены без внедрения цифровых технологий [18].

Цель данной статьи — показать подходы к оценке инфраструктуры образования регионов РФ с акцентом на ее готовность к цифровой трансформации.

Цифровая трансформация — это «технологическое изменение на всех уровнях организации, которое включает как использование цифровых технологий для улучшения существующих процессов, так и исследование цифровых инноваций, которые определяются как сочетание цифровых технологий и физических компонентов для создания инновационных продуктов» [19]. В процессе разработки стратегии цифровой трансформации при принятии решений об эффективных преобразованиях для промышленных предприятий применяется модели цифровой зрелости [20], среди которых выделяются модели одномерной зрелости (SMM, single maturity model) и многомерной зрелости (MMM, multiple maturity model) [19].

Говоря об инфраструктуре образования, безусловно важным для регионов является ее наличие на территории, но для оценки готовности к цифровой трансформации необходимо изучить технологическую обеспеченность объектов инфраструктуры.

В контексте исследования возможностей образовательных организаций «отвечать» на цифровые вызовы времени, в статье использованы данные по обеспеченности инфраструктуры различного уровня персональными компьютерами (далее — ПК). Было бы актуально также включить данные по подключению школ к Интернету со скоростью, позволяющей 100% школьников одновременно активно использовать современные ресурсы; по формированию современной цифровой инфраструктуры в образовательных организациях (например, сетевые ресурсы, компьютерные классы и пр.), что позволило бы провести многомерную оценки готовности инфраструктуры образования к технологической трансформации, но на момент исследования это не представлялось возможным, так как не доступна информация по всем регионам РФ. Поэтому на данном этапе был выбран показатель «Число персональных компьютеров, используемых в учебных целях в государственных и муниципальных организациях

Таблица 1

Обеспеченность персональных компьютеров, используемых в учебных целях, в государственных и муниципальных организациях осуществляющих образовательную деятельность по образовательным программам начального, основного и среднего общего образования, по программам подготовки квалифицированных рабочих, служащих, по программам подготовки специалистов среднего звена и по программам высшего образования на 1000 обучающихся (студентов), нормированное относительно соответствующего максимального значения, 2019 г.

№	Регионы	ПК <sub>1</sub>	ПК <sub>2</sub>	ПК <sub>3</sub>	№	Регионы	ПК <sub>1</sub>	ПК <sub>2</sub>	ПК <sub>3</sub>
1	Белгородская область	0,17	0,58	0,25	44	Ставропольский край	0,08	0,38	0,16
2	Брянская область	0,15	0,35	0,15	45	Республика Башкортостан	0,17	0,34	0,20
3	Владимирская область	0,20	0,50	0,17	46	Республика Марий Эл	0,10	0,41	0,17
4	Воронежская область	0,21	0,46	0,12	47	Республика Мордовия	0,17	0,49	0,16
5	Ивановская область	0,16	0,50	0,20	48	Республика Татарстан	0,30	0,74	0,15
6	Калужская область	0,26	0,33	0,15	49	Удмуртская Республика	0,13	0,34	0,17
7	Костромская область	0,14	0,41	0,14	50	Чувашская Республика	0,15	0,57	0,12
8	Курская область	0,19	0,48	0,13	51	Пермский край	0,15	0,43	0,18
9	Липецкая область	0,16	0,42	0,13	52	Кировская область	0,15	0,51	0,16
10	Московская область	0,24	0,50	0,18	53	Нижегородская область	0,23	0,47	0,14
11	Орловская область	0,14	0,41	0,16	54	Оренбургская область	0,15	0,39	0,15
12	Рязанская область	0,22	0,53	0,19	55	Пензенская область	0,22	0,39	0,12
13	Смоленская область	0,14	0,41	0,12	56	Самарская область	0,18	0,42	0,15
14	Тамбовская область	0,20	0,44	0,21	57	Саратовская область	0,14	0,33	0,16
15	Тверская область	0,16	0,48	0,14	58	Ульяновская область	0,15	0,37	0,14
16	Тульская область	0,26	0,51	0,14	59	Курганская область	0,13	0,56	0,10
17	Ярославская область	0,18	0,48	0,16	60	Свердловская область	0,21	0,40	0,19
18	г. Москва	0,32	0,95	0,15	61	ХМАО	0,26	0,69	0,16
19	Республика Карелия	0,20	0,50	0,20	62	ЯНАО	1,00	0,80	1,00
20	Республика Коми	0,20	0,51	0,16	63	Тюменская область	0,21	0,62	0,15
21	Ненецкий АО	0,25	0,50	0,00	64	Челябинская область	0,14	0,48	0,19
22	Архангельская область	0,18	0,52	0,21	65	Республика Алтай	0,13	0,30	0,18
23	Вологодская область	0,15	0,45	0,21	66	Республика Тыва	0,10	0,35	0,34
24	Калининградская область	0,23	0,56	0,19	67	Республика Хакасия	0,21	0,58	0,26
25	Ленинградская область	0,21	0,57	0,27	68	Алтайский край	0,15	0,45	0,21
26	Мурманская область	0,20	0,40	0,23	69	Красноярский край	0,16	0,43	0,22
27	Новгородская область	0,15	0,54	0,24	70	Иркутская область	0,16	0,49	0,13
28	Псковская область	0,16	0,40	0,12	71	Кемеровская область	0,12	0,51	0,14
29	г. Санкт-Петербург	0,16	0,50	0,14	72	Новосибирская область	0,14	0,46	0,16
30	Республика Адыгея	0,11	0,25	0,07	73	Омская область	0,11	0,34	0,17
31	Республика Калмыкия	0,16	0,45	0,09	74	Томская область	0,17	0,47	0,26
32	Республика Крым	0,10	0,31	0,15	75	Республика Бурятия	0,11	0,54	0,14
33	Краснодарский край	0,13	0,36	0,13	76	Республика Саха (Якутия)	0,19	0,41	0,16
34	Астраханская область	0,13	0,40	0,14	77	Забайкальский край	0,13	0,40	0,21
35	Волгоградская область	0,19	0,30	0,14	78	Камчатский край	0,39	0,51	0,25
36	Ростовская область	0,22	0,43	0,18	79	Приморский край	0,17	0,40	0,13
37	г. Севастополь	0,11	0,27	0,11	80	Хабаровский край	0,21	0,45	0,22
38	Республика Дагестан	0,12	0,25	0,14	81	Амурская область	0,17	0,49	0,23
39	Республика Ингушетия	0,06	0,19	0,10	82	Магаданская область	0,25	0,80	0,10
40	Кабардино-Балкарская Республика	0,11	0,29	0,15	83	Сахалинская область	0,20	0,76	0,15
41	Карачаево-Черкесская Республика	0,14	0,37	0,13	84	Еврейская автономная область	0,14	0,58	0,22
42	Республика Северная Осетия – Алания	0,15	0,42	0,10	85	Чукотский автономный округ	0,24	1,00	0,46
43	Чеченская Республика	0,11	0,20	0,12					

Примечание: здесь ПК<sub>1</sub> — соответствует индексу обеспеченности ПК общеобразовательных организаций; ПК<sub>2</sub> — индексу обеспеченности ПК организаций, осуществляющих образовательную деятельность по программам подготовки квалифицированных рабочих, служащих, по программам подготовки специалистов среднего звена; ПК<sub>3</sub> — индексу обеспеченности ПК организаций, осуществляющих образовательную деятельность по программам высшего образования.

осуществляющих образовательную деятельность по образовательным программам начального, основного и среднего общего образования, по программам подготовки квалифицированных рабочих, служащих, по программам подготовки специалистов среднего звена и по программам высшего образования на 1000 обуча-

ющихся (студентов)» и поставлена исследовательская цель — оценить возможность кластеризации регионов по уровню развития социальной инфраструктуры образования с позиции обеспеченности ПК.

Для этого использованы официальные статистические данные [21], на этапе подготовки которых

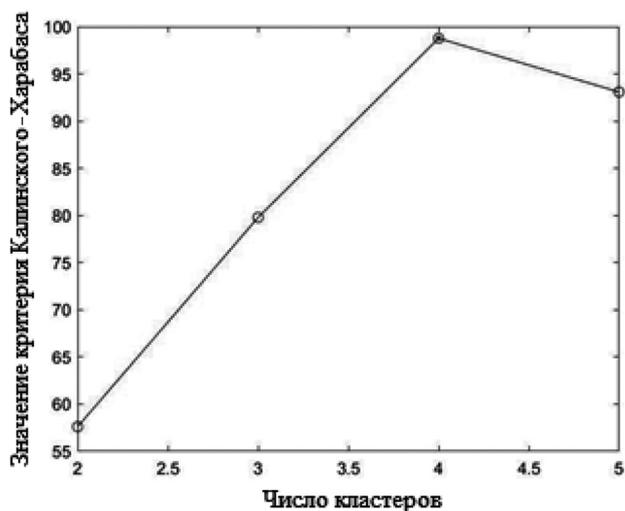


Рис. 1. Распределение критерия Калинского–Харабаса в зависимости от числа кластеров

выполнена нормализация каждой переменной на ее максимальное значение (табл. 1).

Учитывая тот факт, что в качестве одного из приоритетов развития России на период 10-15 лет определен «переход к передовым цифровым, интеллектуальным производственным технологиям, роботизированным системам, новым материалам и способам конструирования, создание систем обработки больших объемов данных, машинного обучения и искусственного интеллекта» [22], без наличия современного оборудования это будет невозможным.

Для выявления дифференциации регионов по уровню обеспеченности образовательных организаций ПК, применены методы машинного обучения. Для сравнения эффективности различных подходов к

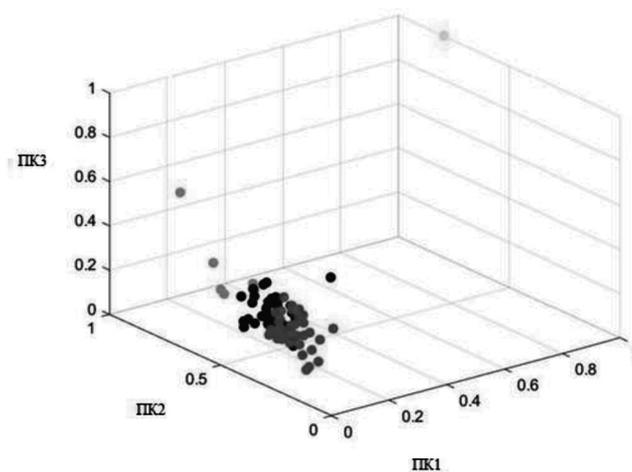


Рис. 2. Визуальное представление кластеризации регионов по обеспеченности ПК образовательных учреждений разного уровня

кластеризации, исследована возможность применения метода Linkage и метода самоорганизующихся карт (СОК) Кохонена.

Кластеризация, представляющая собой автоматическое разбиение элементов некоторого множества на группы в зависимости от их схожести, позволяет:

- 1) разработать типологии или классификации;
- 2) исследовать схемы группирования объектов;
- 3) представлять и проверять гипотезы [23].

Первоначально на основе расчетов в среде Matlab было определено оптимальное число кластеров для исследуемого массива данных. Для этой цели могут быть применены методы:

1. Метод Калинского–Харабаса (Calinski–Harabasz) или критерий отношения дисперсий. Четко опре-

Таблица 2

Число персональных компьютеров, используемых в учебных целях в государственных и муниципальных организациях осуществляющих образовательную деятельность по образовательным программам начального, основного и среднего общего образования, по программам подготовки квалифицированных рабочих, служащих, по программам подготовки специалистов среднего звена и по программам высшего образования на 1000 обучающихся (студентов) в зависимости от кластера, 2019 г.

Среднее, минимум, максимум	Число ПК, используемых в учебных целях в государственных и муниципальных организациях осуществляющих образовательную деятельность по образовательным программам начального, основного и среднего общего образования на 1000 обучающихся	Число ПК, используемых в учебных целях в государственных и муниципальных организациях осуществляющих образовательную деятельность по программам подготовки квалифицированных рабочих, служащих, по программам подготовки специалистов среднего звена на 1000 обучающихся (студентов)	Число ПК, используемых в учебных целях в государственных и муниципальных организациях осуществляющих образовательную деятельность по программам высшего образования на 1000 студентов
<b>1 кластер</b>			
Среднее число	147	189	286
Минимум/максимум	88/306	167/228	160*/442
<b>2 кластер</b>			
Среднее число	117	135	261
Минимум/максимум	46/201	69/165	113/556
<b>3 кластер</b>			
Среднее число	205	301	321
Минимум/максимум	156/249	251/366	157/763
<b>4 кластер</b>			
Число ПК**	784	291	1643

Примечание: \* – Ненецкий автономный округ, попавший в 1 кластер не имеет организаций осуществляющих образовательную деятельность по программам высшего образования и, следовательно, не может иметь ПК для этих организаций, поэтому указан минимум для регионов, имеющих организации данного типа; \*\* – так как кластер содержит 1 регион, указано одно значение по каждому показателю.

деленные кластеры имеют большую дисперсию между кластерами и небольшую дисперсию внутри кластера. Чем больше коэффициент отличия между дисперсиями, тем лучше разделение данных [24].

2. Коэффициент Силуэт (Silhouette): вычисляется с помощью среднего внутрикластерного расстояния и среднего расстояния до ближайшего кластера по каждому образцу [25].
3. Индекс Дэвиса–Болдина (DBI): является средним отношением внутриклассовых разбросов к расстояниям между кластерами [26].

Методы показали, что оптимальное число критериев варьируется в диапазоне от трех до пяти. В качестве примера, результат оценки обоснованности кластеров с помощью метода Калинского–Харабаса представлен на рис. 1. По осям на рис. 1 представлены: ОХ – число кластеров, ОУ – значение критерия Калинского–Харабаса. График показывает, что наибольшее значение критерия Калинского–Харабаса приходится на 4 кластера, что позволяет предположить, что оптимальное количество кластеров – 4.

При этом в качестве методов кластеризации применялись:

1. Метод Kmeans – двухфазный итерационный алгоритм для минимизации суммы расстояний от точки до центроида, суммированных по всем кластерам [27].
2. Метод Gmdistribution: модель гауссовой смеси, в ходе применения которой результат классификации представлен вероятностью, и тот элемент массива данных, у которого более высокая вероятность, считается принадлежащим к тому или иному кластеру [28].
3. Метод Linkage: агломерационное иерархическое кластерное дерево, метод иерархической кластеризации, основывающийся на процессе агломерации (AGNES, Agglomerative Nesting), т. е. последовательного объединения индивидуальных объектов

или их групп во все более крупные подмножества, или обратном по смыслу процессе разбиения (DIANA, Divide Analysis), который начинается с корня и на каждом шаге делит образующие группы по степени их гетерогенности [29, 30].

Результаты применения одного из названных выше подходов, Linkage, при котором регионы РФ наиболее явно распределились на четыре кластера, показаны на рис. 2. Применение других методов не дало такого явного разделения на кластеры.

На данном графике (рис. 2) кластер 1 (темная заливка) содержит 34 региона, находится недалеко от центра левой части графика и включает в себя регионы, в которых объекты образовательной инфраструктуры всех типов в среднем лучше обеспечены ПК, чем регионы, формирующие 2 кластер.

Кластер 2 (полутемная заливка), самый большой по численности, содержит 44 региона с наименьшим количеством ПК в организациях осуществляющих образовательную деятельность. Кластер 3 (светлая заливка) содержит 6 регионов, тяготеет к левому верхнему углу и содержит регионы, имеющие более высокую обеспеченность ПК<sub>2</sub> и ПК<sub>3</sub>. График показывает, что кластер 4 (желтая заливка), состоящий из одного региона, а именно ЯНАО, находится в правом верхнем углу, полностью отделенный от других кластеров. В этом регионе количество компьютеров во всех трех группах образовательных организаций максимально.

Визуализация данных наглядно показывает, что большинство российских регионов по уровню обеспеченности ПК в образовательных организациях разного уровня значительно уступают наилучшим значениям в регионе-лидере (4 кластер – ЯНАО). Скопление точек, представляющих регионы, входящие в разные кластеры, находится в зоне, отдаленной от единицы по всем осям.

Для принятия управленческих решений и для прогнозирования попадания региона в определенный кластер целесообразно вернуться к абсолютным значениям исследуемых показателей (табл. 2).

Таблица 3

Распределение регионов РФ по уровню обеспеченности ПК образовательных организаций разного уровня по 4 кластерам, определенным методом Linkage, 2019 г.

Номер кластера	Регионы, входящие в кластер
1 кластер (34 региона)	Белгородская обл., Владимирская обл., Воронежская обл., Ивановская обл., Курская обл., Московская обл., Рязанская обл., Тверская обл., Тульская обл., Ярославская обл., Республика Карелия, Республика Коми, Ненецкий АО, Архангельская обл., Калининградская обл., Ленинградская обл., Новгородская обл., г. Санкт-Петербург, Республика Мордовия, Чувашская Республика, Кировская обл., Нижегородская обл., Курганская обл., Тюменская обл., Челябинская обл., Республика Хакасия, Иркутская обл., Кемеровская обл., Новосибирская обл., Томская обл., Республика Бурятия, Камчатский край, Амурская обл., Еврейская АО.
2 кластер (44 региона)	Брянская обл., Калужская обл., Костромская обл., Липецкая обл., Орловская обл., Смоленская обл., Тамбовская обл., Вологодская обл., Мурманская обл., Псковская обл., Республика Адыгея, Республика Калмыкия, Республика Крым, Краснодарский край, Астраханская обл., Волгоградская обл., Ростовская обл., г. Севастополь, Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия – Алания, Чеченская Республика, Ставропольский край, Республика Башкортостан, Республика Марий Эл, Удмуртская Республика, Пермский край, Оренбургская обл., Пензенская обл., Самарская обл., Саратовская обл., Ульяновская обл., Свердловская обл., Республика Алтай, Республика Тыва, Алтайский край, Красноярский край, Омская обл., Республика Саха (Якутия), Забайкальский край, Приморский край, Хабаровский край
3 кластер (6 регионов)	г. Москва, Республика Татарстан, ХМАО, Магаданская обл., Сахалинская обл., Чукотский автономный округ
4 кластер (1 регион)	ЯНАО

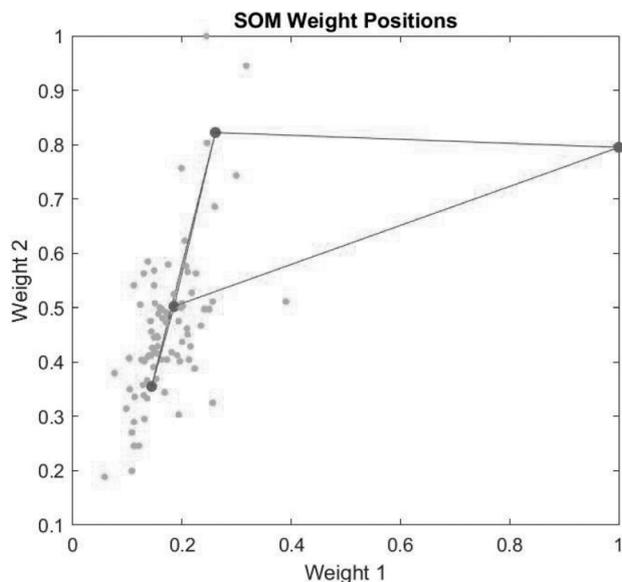


Рис. 3. Отображение входных векторов (светлые точки) на веса нейронов (темные точки)

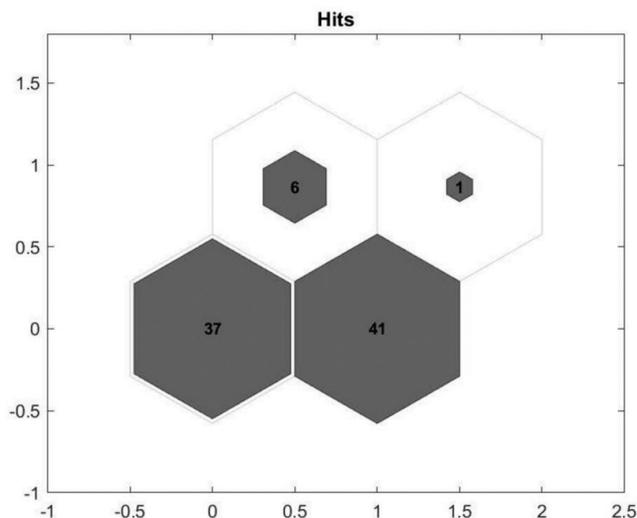


Рис. 4. Карта совпадений результатов обучения СОК

Обращает на себя внимание тот факт, что независимо от попадания в тот или иной кластер, у российских регионов более высокая обеспеченность ПК в образовательных организациях высшего образования, меньшая — в школах.

Распределение регионов, входящих в выделенные 4 кластера, представлено в табл. 3.

Поскольку задача поиска закономерностей в исследуемых несбалансированных и нелинейных данных сложных систем относительно нова (хотя классические методы кластеризации и машинного обучения использовались при исследовании экономики регионов и ранее [31]), кластеризация регионов РФ в зависимости от обеспеченности ПК образовательных организаций разного уровня была проведена также с использованием инструментария самоорганизующихся карт Кохонена (СОК).

В качестве выборки для нейронной сети Кохонена были рассмотрены те же нормированные статистические данные по всем регионам РФ за 2019 г. (ПК1, ПК2, ПК3), которые использовались ранее для классических методов кластеризации (Linkage).

В сети Кохонена обучающее множество состоит лишь из значений входных переменных, в процессе обучения нет сравнения выходов нейронов с эталонными значениями. В процессе последовательной подачи на вход сети обучающих примеров определяется нейрон, у которого скалярное произведение весов и поданного на вход вектора минимально. Этот нейрон объявляется победителем и является центром при подстройке весов у соседних нейронов. Обучение при этом заключается не в минимизации ошибки, а в подстройке весов (внутренних параметров нейронной сети) для наибольшего совпадения с входными данными.

Процесс обучения сети Кохонена:

- 1) инициализация весов;
- 2) случайный выбор входной переменной;
- 3) выбор нейрона-победителя;
- 4) обновление весов;

- 5) возврат к шагу 2 до момента достижения требуемой точности или минимизации ошибки [32].

Рис. 3 содержит проекцию пространства входных векторов (в виде светлых точек) на каждый нейрон (темные точки), соседние нейроны соединены линиями.

Такие результаты выявили, что возможно сгруппировать данные на 4 группы, в соответствии с заданным количеством нейронов сети, равным двум, что совпадает с решением классическим методом кластеризации Linkage, который был использован ранее.

Визуальное представление кластеров, полученных с использованием инструментария самоорганизующихся карт Кохонена (СОК) представлено на рис. 4. «Карта совпадений» («hits map»), получившаяся в результате применения сети Кохонена с двумя нейронами и показывающая, сколько раз каждый нейрон был «победителем» для каждого входного значения дает представление о количестве данных, которые собираются в каждом кластере и позволяет оценить его важность.

Рис. 4 показывает местоположения нейрона в топологии и указывает, сколько из обучающих данных

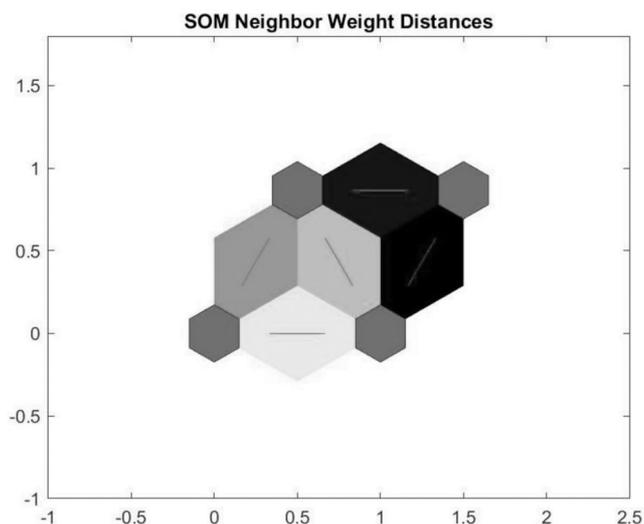


Рис. 5. Карта расстояний между соседними кластерами

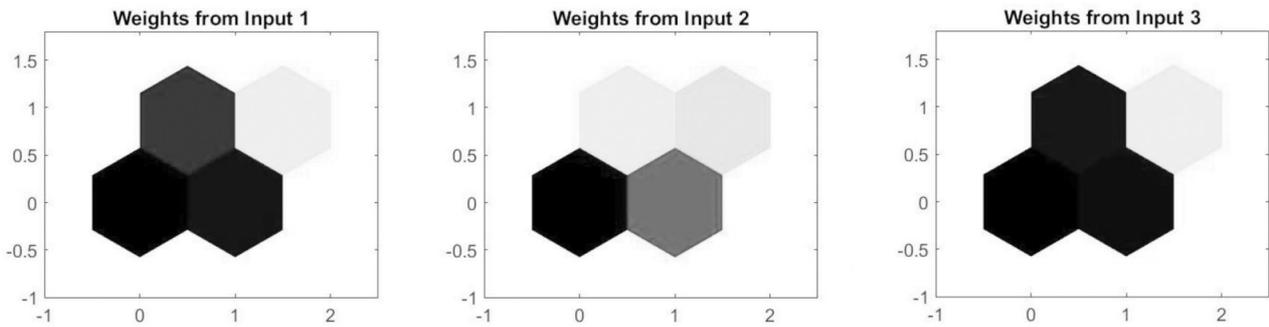


Рис. 6. Веса компонент ПК<sub>1</sub>, ПК<sub>2</sub>, ПК<sub>3</sub> в каждом из четырех кластеров

сопоставлено с каждым из нейронов (кластерные центры). Топология из двух нейронов (по горизонтальной и вертикальной осям) образует сетку 2×2, таким образом, образуется 4 нейрона. Максимальное количество данных, сопоставленных с одним из нейронов, равняется 37. Таким образом в этот кластер направлены 37 входных векторов.

Результаты кластеризации показывают, что сформированные кластеры не только количественно, но и содержательно идентичны тем, которые получились при кластеризации методом Linkage. Кластеры 3 (6 регионов) и 4 (1 регион), расположенные в более высокой области, представлены теми же регионами РФ, а 1 кластер, включающий при использовании метода Linkage 34 региона, при кластеризации нейросетью содержит 37 регионов, в связи с этим 2 кластер включает уже не 44, а 41 регион. Тем не менее, следует признать достаточно точным проведенную 2 способами кластеризацию регионов РФ по уровню обеспеченности ПК образовательных организаций разного уровня.

Карта расстояний соседних кластеров друг от друга представлена на рис. 5.

На рис. 5 нейроны представлены в виде угловых (крайних) пятен. Соседние участки окрашены от темного к светлому, чтобы показать, насколько близок вектор веса каждого нейрона к его соседям. Видно, что кластеры 2 и 3 наиболее близки к 4.

Инструментарий среды Matlab самоорганизующихся карт Кохонена дает дополнительную возможность оценить вес конкретного компонента ПК<sub>1</sub>, ПК<sub>2</sub>, ПК<sub>3</sub> (рис. 6).

Визуализация выражена цветовым представлением. В частности, чем более темным цветом на рис. 6 представлена заливка кластера, тем больше вес фактора. Например, на попадание в 1 кластер влияют сразу 3 фактора (на всех 3 графиках 1 кластер имеет черную заливку). Попадание региона во 2 кластер определяют ПК<sub>1</sub> (индекс обеспеченности ПК общеобразовательных организаций) и ПК<sub>3</sub> (индекс

обеспеченности ПК организаций, осуществляющих образовательную деятельность по программам высшего образования). Влияние ПК<sub>2</sub> (индекса обеспеченности ПК организаций, осуществляющих образовательную деятельность по программам подготовки квалифицированных рабочих, служащих, по программам подготовки специалистов среднего звена) ниже, это показано заливкой более светлого тона. Для 3 кластера наиболее значимым для оценки является ПК<sub>3</sub> (индекс обеспеченности ПК организаций, осуществляющих образовательную деятельность по программам высшего образования — черная заливка).

Исследование показало, что кластеризация регионов методами машинного обучения позволяет прогнозировать попадание конкретного региона в определенный кластер и оценить готовность региона к применению цифровых технологий в образовательном процессе с точки зрения оснащенности образовательных организаций персональными компьютерами, что в свою очередь, является одним из важных факторов цифровой трансформации региона в целом. Это позволит принимать более взвешенные управленческие решения по развитию регионов в области развития инфраструктуры образования (с позиции обеспеченности ПК).

Дальнейшей задачей исследования является проведение многомерной оценки на основе изучение дополнительных факторов, влияющих на причинно-следственные связи и определяющих попадания регионов в соответствующие кластеры. На данном этапе можно предположить, что на высокую обеспеченность ПК в ряде регионов РФ влияют значительные инвестиции и целевая направленность в региональной социально-экономической политике.

### *Благодарности*

Статья подготовлена в соответствии с Планом НИР ФГБУН «Институт экономики УрО РАН» на 2021 г.

### Список использованных источников

1. Я. И. Кузьмин, И. Д. Фрумин, И. В. Абанкина и др. Как сделать образование двигателем социально-экономического развития? М.: Издательский дом НИУ ВШЭ, 2019. <https://publications.hse.ru/books/287219806>.
2. Послание Президента РФ В. В. Путина Федеральному Собранию, 1 декабря 2016 г. <http://kremlin.ru/events/president/news/copy/53379>.
3. К. Schwab. The fourth industrial revolution. Penguin, 2017. 192 p.
4. Стратегия научно-технологического развития Российской Федерации: утверждена указом Президента РФ № 642 от 01.12.2016 г. (ред. от 15.03.2021 г.). [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_207967](http://www.consultant.ru/document/cons_doc_LAW_207967).
5. Е. А. Алпеева, Е. А. Мерзлякова, А. В. Сысоев/ Теоретические подходы к исследованию социально ориентированной инфраструктуры региона//Экономика в промышленности. 2018. Т. 11. № 4. С. 412-417.

6. З. Ф. Гарипова, Л. Р. Халитова. Развитие социальной инфраструктуры как приоритетное направление повышения конкурентоспособности и конкурентоустойчивости территориальной системы//Финансовая экономика. 2020. № 2. С. 263-267.
7. Ю. А. Кузнецова. Социальная инфраструктура в рамках концепции конкурентоустойчивости территории//Международный журнал прикладных и фундаментальных исследований. 2017. № 8-2. С. 333-337.
8. Н. П. Кузьмич. Развитие социальной инфраструктуры сельских территорий региона в целях улучшения качества жизни населения//Экономика: вчера, сегодня, завтра. 2019. Т. 9. № 4А. С. 392-399.
9. А. Л. Сабинина, В. В. Соколовский, Н. А. Шульженко, Н. А. Сычева. О стратегии развития многофункциональных комплексов социальной инфраструктуры в парадигме «умный город»//Финансы и кредит. 2020. Т. 26. № 7 (799). С. 1469-1495.
10. O. A. Busari. The role of economic and social infrastructure in economic development: a global view. [https://www.academia.edu/1566979/The\\_role\\_of\\_economic\\_and\\_social\\_infrastructure\\_in\\_economic\\_development\\_a\\_global\\_view\\_by](https://www.academia.edu/1566979/The_role_of_economic_and_social_infrastructure_in_economic_development_a_global_view_by).
11. B. Ibama, S. S. Owukio, C. Wocha. Comparative Study of Social Infrastructure Provision in Ikwere and Etche Local Government Areas of Rivers State Using Geographic Information System//Scientific Research Journal (SCRJ), Vol. III. Iss. V. May 2015. [https://www.academia.edu/27563581/Comparative\\_Study\\_of\\_Social\\_Infrastructure\\_Provision\\_in\\_Ikwere\\_and\\_Etche\\_Local\\_Government\\_Areas\\_of\\_Rivers\\_State\\_Using\\_Geographic\\_Information\\_System](https://www.academia.edu/27563581/Comparative_Study_of_Social_Infrastructure_Provision_in_Ikwere_and_Etche_Local_Government_Areas_of_Rivers_State_Using_Geographic_Information_System).
12. G. Torrisi. Public infrastructure: definition, classification and measurement issues. University Library of Munich, Germany, MPRA Paper, 2009. [https://www.researchgate.net/publication/23935428\\_Public\\_infrastructure\\_definition\\_classification\\_and\\_measurement\\_issues](https://www.researchgate.net/publication/23935428_Public_infrastructure_definition_classification_and_measurement_issues).
13. Н. С. Ильющенко. Digital learning: Перспективы и риски цифрового поворота в образовании//Проектирование будущего. Проблемы цифровой реальности: труды 2-й Международной конференции (7-8 февраля 2019 г., Москва). М.: ИПМ им. М. В. Келдыша, 2019. С. 215-225. <https://keldysh.ru/future/2019/20.pdf>. doi: 10.20948/future-2019-20.
14. L. Davis. Digital Learning: What to Know in 2019/Evolving Ed. January 25, 2019. <https://www.schoology.com/blog/digital-learning-whatknow-2019>.
15. Г. В. Новикова. Вопросы целесообразности применения технологий виртуальной реальности в образовании школьников и студентов// Проектирование будущего. Проблемы цифровой реальности: труды 4-й Международной конференции (4-5 февраля 2021 г., Москва). М.: ИПМ им. М. В. Келдыша, 2021. С. 276-286. <https://keldysh.ru/future/2021/24.pdf>. <https://doi.org/10.20948/future-2021-24>.
16. Л. В. Мезенцева. Онлайн-курсы не менее эффективны, чем офлайн-форматы. Доказано экспериментально. IQ, научно-образовательный портал НИУ ВШЭ, 2018. <https://iq.hse.ru/news/217043836.html>.
17. Дело в шлеме. [Rspectr.com](https://www.rspectr.com/articles/413/delo-v-shleme) 21.05.2018. <https://www.rspectr.com/articles/413/delo-v-shleme>.
18. Указ Президента Российской Федерации от 07.05.2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 г.» <http://www.kremlin.ru/acts/bank/43027>.
19. T. Bierhold. For a better understanding of Industry 4 — An Industry 4 maturity model. University Of Twente, Enschede The Netherlands, 2018. [https://essay.utwente.nl/75330/1/Bierhold\\_BA\\_BMS.pdf](https://essay.utwente.nl/75330/1/Bierhold_BA_BMS.pdf).
20. А. К. Петрова, Н. В. Лашманова, А. Б. Жернаков. Подход к оценке цифровой зрелости промышленных предприятий на основе нечеткой логики//Инновации. 2020. № 10. С. 75-82.
21. Регионы России. Социально-экономические показатели. 2020: стат. сб. М.: Росстат, 2020. 1242 с.
22. А. А. Зацаринный, К. К. Колин. Теория и практика цифровой трансформации общества в рамках приоритетов научно-технологического развития России//Проектирование будущего. Проблемы цифровой реальности: труды 2-й Международной конференции (7-8 февраля 2019 г., Москва). М.: ИПМ им. М. В. Келдыша, 2019. С. 29-39. <https://keldysh.ru/future/2019/3.pdf>. doi: 10.20948/future-2019-3.
23. Н. Б. Паклин, В. И. Орешков. Бизнес-аналитика: от данных к знаниям. М.: Издательство «Питер», 2013. 704 с.
24. T. Calinski, J. Harabasz. A dendrite method for cluster analysis//Communications in Statistics. Vol. 3. № 1. 1974. P. 1-27.
25. L. Kaufman, P. J. Rousseeuw. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, NJ: John Wiley & Sons, Inc., 1990.
26. D. L. Davies, D. W. Bouldin. A Cluster Separation Measure//IEEE Transactions on Pattern Analysis and Machine Intelligence. 979. Vol. PAMI-1. № 2. P. 224-227.
27. D. Arthur, S. Vassilvitskii. K-means++: The Advantages of Careful Seeding//SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. 2007. P. 1027-1035.
28. G. McLachlan, D. Peel. Finite Mixture Models. Hoboken, NJ: John Wiley & Sons, Inc., 2000.
29. M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set//Journal of Statistical Software. 2014. 61 (6). P. 1-36.
30. C. Fraley, A. E. Raftery, T. B. Murphy, L. Scrucca. Mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical Report № 597, Department of Statistics, University of Washington. 2012.
31. Е. Д. Игнатъева, О. С. Мариев. Совершенствование методологии и инструментария анализа потенциала саморазвития социально-экономических подсистем региона// Вестник УрФУ. Серия «Экономика и управление». 2011. № 5. С. 105-114.
32. T. Kohonen. Self-Organizing Maps (Third Extended Edition). New York, 2001. 501 p.

## References

1. Ya. I. Kuzminov, I. D. Frumin, I. V. Abankina et al. How to make education an engine of social and economic development? М.: NRU HSE, 2019. (In Russ.) <https://publications.hse.ru/books/287219806>.
2. Poslanie prezidenta Rossiyskoy-Federacii V. V. Putina Federalnomu Sobraniyu [Message from the President of the Russian Federation V. V. Putin to the Federal Assembly], December 1, 2016. (In Russ.) <http://kremlin.ru/events/president/news/copy/53379>.
3. K. Schwab. The fourth industrial revolution. Penguin, 2017. 192 p.
4. Strategiya-nauchno-tehnologicheskogo-razvitiya-Rossiyskoy-Federacii [Strategy of scientific and technological development of the Russian Federation] approved by the Decree of the President of the Russian Federation № 642 of 01.12.2016. (In Russ.) [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_207967](http://www.consultant.ru/document/cons_doc_LAW_207967).
5. E. A. Alpeeva, E. A. Merzlyakova, A. V. Sysoev. Theoretical approaches to the study of socially oriented infrastructure of the region//Ekonomika v Promyshlennosti [Industrial Economics]. 2018. Vol. 11. № 4. P. 412-417. (In Russ.)
6. Z. F. Garipova, L. R. Khalitova. Development of social infrastructure as a priority direction for increasing the competitiveness and competitiveness of the territorial system// Finansovaya ekonomika [Financial Economics]. 2020. № 2. P. 263-267. (In Russ.)
7. Yu. A. Kuznetsova. Social infrastructure within the framework of the concept of competitiveness of the territory//Mezhdunarodny zhurnal prikladnyh i fundamentalnyh issledovaniy [International Journal of Applied and Fundamental Research]. 2017. № 8-2. P. 333-337. (In Russ.)
8. N. P. Kuzmich. Development of social infrastructure of rural areas of the region in order to improve the quality of life of the population//Ekonomika-vchera-segodnya-zavtra [Economy: yesterday, today, tomorrow]. 2019. Vol. 9. № 4A. P. 392-399. (In Russ.)
9. A. L. Sabinina, V. V. Sokolovsky, N. A. Shulzhenko, N. A. Sycheva. On the development strategy of multifunctional complexes of social infrastructure in the «Smart City» paradigm//Finansy-i-kredit [Finance and Credit]. 2020. Vol. 26. № 7 (799). P. 1469-1495. (In Russ.)
10. O. A. Busari. The role of economic and social infrastructure in economic development: a global view. [https://www.academia.edu/1566979/The\\_role\\_of\\_economic\\_and\\_social\\_infrastructure\\_in\\_economic\\_development\\_a\\_global\\_view\\_by](https://www.academia.edu/1566979/The_role_of_economic_and_social_infrastructure_in_economic_development_a_global_view_by).
11. B. Ibama, S. S. Owukio, C. Wocha. Comparative Study of Social Infrastructure Provision in Ikwere and Etche Local Government Areas of Rivers State Using Geographic Information System//Scientific Research Journal (SCRJ), Vol. III. Iss. V. May 2015. [https://www.academia.edu/27563581/Comparative\\_Study\\_of\\_Social\\_Infrastructure\\_Provision\\_in\\_Ikwere\\_and\\_Etche\\_Local\\_Government\\_Areas\\_of\\_Rivers\\_State\\_Using\\_Geographic\\_Information\\_System](https://www.academia.edu/27563581/Comparative_Study_of_Social_Infrastructure_Provision_in_Ikwere_and_Etche_Local_Government_Areas_of_Rivers_State_Using_Geographic_Information_System).
12. G. Torrisi. Public infrastructure: definition, classification and measurement issues. University Library of Munich, Germany, MPRA Paper, 2009. [https://www.researchgate.net/publication/23935428\\_Public\\_infrastructure\\_definition\\_classification\\_and\\_measurement\\_issues](https://www.researchgate.net/publication/23935428_Public_infrastructure_definition_classification_and_measurement_issues).
13. N. S. Ilyushenko. [Digital learning: Prospects and risks of the digital turn in education]//Trudy 2 Mezhdunarodnoy konferencii Proektirovanie buduschego. Problemy-cifrovoy-realnosti. IPM im. M. V. Keldysh [Proc. of the 2nd International Conference Designing the future. Problems of Digital Reality of Keldysh Institute of Applied Mathematics]. Moscow, 2019. P. 215-225. (In Russ.) <https://keldysh.ru/future/2019/20.pdf>.
14. L. Davis. Digital Learning: What to Know in 2019/Evolving Ed. January 25, 2019. <https://www.schoology.com/blog/digital-learning-whatknow-2019>.
15. G. V. Novikova. [Questions of the expediency of using virtual reality technologies in the education of schoolchildren and students]//Trudy 4 Mezhdunarodnoy konferencii Proektirovanie buduschego. Problemy-cifrovoy-realnosti. IPM im. M. V. Keldysh [Proc. of the 2nd International Conference Designing the future. Problems of Digital Reality of Keldysh Institute of Applied Mathematics]. Moscow, 2021. P. 276-286. (In Russ.) <https://keldysh.ru/future/2021/24.pdf>.

16. L. V. Mezentseva. Online courses are no less effective than offline formats. Proven experimentally. IG nauchno obrazovatelnyy portal NIU VSE [IQ, a research and educational portal of the NRU HSE], 2018. (In Russ.) <https://iq.hse.ru/news /217043836.html>.
17. The case in the helmet. *Rspectr.com* 21.05.2018. (In Russ.) <https://www.rspectr.com/articles/413/delo-v-shleme>.
18. Ukaz Pprezidenta Rossiyskoy Federacii ot 07.05.2018 № 204 o nacionalnyh celyah i strategicheskikh zadachah razvitiya-Rossiyskoy Federacii na period do 024 goda [Decree of the President of the Russian Federation of 07.05.2018 № 204 «On national goals and strategic objectives of the development of the Russian Federation for the period up to 2024»]. (In Russ.) <http://www.kremlin.ru/acts/bank/43027>.
19. T. Bierhold. For a better understanding of Industry 4 — An Industry 4 maturity model. University Of Twente, Enschede The Netherlands, 2018. [https://essay.utwente.nl/75330/1/Bierhold\\_BA\\_BMS.pdf](https://essay.utwente.nl/75330/1/Bierhold_BA_BMS.pdf).
20. A. K. Petrova, N. V. Lashmanova, A. B. Zhernakov. An approach to assessing the digital maturity of industrial enterprises based on fuzzy logic//*Innovatsii* [Innovation]. 2020. № 10. P. 75-82. (In Russ.)
21. Regions of Russia. Socio-economic indicators. 2020: statistics. M.: Rosstat, 2020. 1242 p. (In Russ.)
22. A. A. Zatsarinny, K. K. Kolin. [Theory and practice of digital transformation of society in the framework of the priorities of scientific and technological development of Russia]//*Trudy 2 Mezhdunarodnoy konferencii Proektirovanie buduschego. Problemy-cifrovoy-realnosti. IPM im. M. V. Keldysh* [Proc. of the 2nd International Conference Designing the future. Problems of Digital Reality of Keldysh Institute of Applied Mathematics]. Moscow, 2019. P. 29-39. (In Russ.) <https://keldysh.ru/future/2019/3.pdf>.
23. N. B. Paklin, V. I. Oreshkov. *Biznes analitika ot dannyh k znaniyam* [Business analytics: from data to knowledge]. M.: Peter, 2013. 704 p. (In Russ.)
24. T. Calinski, J. Harabasz. A dendrite method for cluster analysis//*Communications in Statistics*. Vol. 3. № 1. 1974. P. 1-27.
25. L. Kaufman, P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons, Inc., 1990.
26. D. L. Davies, D. W. Bouldin. A Cluster Separation Measure//*IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979. Vol. PAMI-1. № 2. P. 224-227.
27. D. Arthur, S. Vassilvitskii. K-means++: The Advantages of Careful Seeding//*SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. 2007. P. 1027-1035.
28. G. McLachlan, D. Peel. *Finite Mixture Models*. Hoboken, NJ: John Wiley & Sons, Inc., 2000.
29. M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs. *NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set*//*Journal of Statistical Software*. 2014. 61 (6). P. 1-36.
30. C. Fraley, A. E. Raftery, T. B. Murphy, L. Scrucca. *Mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*. Technical Report № 597, Department of Statistics, University of Washington. 2012.
31. E. D. Ignatieva, O. S. Mariev. [Improving the methodology and tools for analyzing the potential for self-development of socio-economic subsystems in the region]//*Vestnik UrFU seriya Ekonomika i Upravlenie* [Vestnik UrFU. Economics and Management Series]. 2011. № 5. P. 105-114. (In Russ.)
32. T. Kohonen. *Self-Organizing Maps (Third Extended Edition)*. New York, 2001. 501 p.