

# Сравнительный анализ регионов Российской Федерации методами машинного обучения по набору показателей электронных услуг и сервисов

УДК 332.1 DOI 10.26425/2658-347X-2024-7-4-33-43

Получено 30.07.2024

Доработано после рецензирования 07.10.2024

Принято 20.10.2024

## Борисова Людмила Робертовна

Канд. физ.-мат. наук, доц. каф. математики и анализа данных

ORCID: 0000-0002-5757-0341

E-mail: lrborisova@fa.ru

Финансовый университет при Правительстве Российской Федерации, г. Москва, Россия

## АННОТАЦИЯ

Цифровизация важна не только для работы профессионалов на своих рабочих местах, но и для благополучия специалистов и адаптации к изменениям в информационной сфере, а также для предоставления населению качественных образовательных, медицинских и социальных услуг. В работе использованы методы машинного обучения в целях классификации регионов по набору показателей электронных услуг и сервисов. Получено четкое разделение регионов на две большие, практически равные группы по этому набору показателей. Использование известных статистических критериев продемонстрировало статистическую значимость такого разделения. Построены диаграммы рассеяния как

пример взаимосвязи подобных показателей. Множественный коэффициент корреляции между показателями электронных услуг и сервисов равен 0,71, что свидетельствует в пользу тесной связи между индикаторами цифровизации услуг. Кроме того, получено разделение регионов на кластеры с помощью иерархической кластеризации, из которой следует, что Москва в предоставлении электронных услуг и сервисов значительно обогнала другие регионы России, а остальные субъекты гетерогенны по этому показателю с учетом рассмотренных индикаторов цифровизации регионов по данным Федеральной службы государственной статистики за май 2024 г.

## Ключевые слова

Методы машинного обучения, цифровизация, статистика, регрессия, корреляция, кластеризация, дендрограмма, классификация регионов, электронные услуги

## Для цитирования

Борисова Л.Р. Сравнительный анализ регионов Российской Федерации методами машинного обучения по набору показателей электронных услуг и сервисов // Цифровая социология. 2024. Т. 7. № 4. С. 33–43.

© Борисова Л.Р., 2024.

Статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0. всемирная (<http://creativecommons.org/licenses/by/4.0/>).



# Comparative analysis of the Russian regions using machine learning methods for a set of indicators of electronic services

Received 30.07.2024

Revised 07.10.2024

Accepted 20.10.2024

**Ludmila R. Borisova**

Cand. Sci. (Phys. and Math.), Assoc. Prof. at the Mathematics and Data Analysis Department

ORCID: 0000-0002-5757-0341

E-mail: lrborisova@fa.ru

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

## ABSTRACT

Digitalisation is important not only for the work of professionals in their workplaces, but also for the well-being of specialists and adaptation to changes in the information sphere as well as for providing high-quality educational, medical and social services to the population. The paper uses machine learning methods to classify regions according to a set of indicators of electronic services and services. A clear division of the regions into two large, almost equal groups according to this set of indicators has been obtained. The use of well-known statistical criteria has demonstrated the statistical significance of such a division. Scattering diagrams are constructed as an example

of the relationship of such indicators. The multiple correlation coefficient between the indicators of electronic services and services is 0.71, which indicates a close relationship between the indicators of digitalisation of services. In addition, the division of regions into clusters has been obtained using hierarchical clustering, which implies that Moscow has significantly overtaken other regions of Russia in providing electronic services and services, and the remaining regions are heterogeneous in this indicator, considering the considered indicators of digitalisation of the subjects according to the Federal State Statistics Service data for May 2024.

## Keywords

Machine learning methods, digitalisation, statistics, regression, correlation, clustering, dendrogram, classification of regions, electronic services

## For citation

Borisova L.A. (2024) Comparative analysis of the Russian regions using machine learning methods for a set of indicators of electronic services. *Digital sociology*. Vol. 7, no 4, pp. 33–43. DOI: 10.26425/2658-347X-2024-7-4-33-43

© Borisova L.A., 2024.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).



## ВВЕДЕНИЕ / INTRODUCTION

К настоящему времени в Российской Федерации (далее – РФ, Россия) применяются цифровые технологии практически повсеместно: в здравоохранении, образовании, промышленности, связи и т.д. Искусственный интеллект, роботизация производства, призванные существенно повысить производительность труда, находят все большее применение в разных сферах повседневной жизни человека. Цифровизация современной жизни изначально была связана с переходом от аналоговой формы передачи информации к цифровой. Однако социальные изменения не могут быть только причиной такого перехода. Важно учитывать совместный эффект интернета в эволюции информационных технологий (далее – ИТ) [Бабинцев, Серкина, 2022].

## ОСНОВЫ РАБОТЫ / RESEARCH FOUNDATIONS

Включение инноваций в жизнь общества на основе цифровизации необратимо. Надо уметь распознавать и защищаться от манипуляторов в сети [Ашманов, Касперская, 2021]. Особенно нужно знать, как защитить детей. В этой части интересен опыт Китая. Как отмечают О.В. Бальчиндоржиева и О.М. Золхоева, несмотря на повсеместное расширение использования цифровых технологий, учет традиций позволяет сохранять собственную идентичность, прежде всего культурную [Бальчиндоржиева, Золхоева, 2022]. В Китае большое внимание уделяется избавлению от зависимости от гаджетов детей младшего школьного и дошкольного возраста. Для этого существуют специальные учреждения, в которых лечат подобную зависимость.

Внедрение цифровых технологий позволяет получать конкурентные преимущества. В России в лидерах по использованию цифровых технологий прежде всего финансовые учреждения – «Сбербанк» и «Т-банк» [Петрова, Овечкина, 2021]. При применении ИТ-технологий происходят изменения в сферах здравоохранения и образования [Романова, 2020]. Цифровизация социальных услуг включает в себя не только цифровые технологии, но и самих людей.

В настоящее время появляется все больше исследований, посвященных успешному применению ИТ-технологий в медицине. Так, В.О. Исаенко и М.Н. Рыбина описали пример цифровизации при ранней диагностике сердечно-сосудистой патологии, когда лечащий врач в любой момент времени и в любой точке Земли, в которой доступен

интернет, может просматривать результаты кардиограммы пациента [Исаенко, Рыбина, 2018]. Тем не менее, как отмечают Э.Н. Кропивецкий, Ю.А. Дубоглазова, Д.А. Заборовский, единая система социальной поддержки для всех регионов еще не сформирована [Кропивецкий, Дубоглазова, Заборовский, 2022].

Ранее И.В. Грошев и Ю.Д. Красовский указывали, что проблематика цифровизации социологии сложна и противоречива [Грошев, Красовский, 2018]. Возможные допустимые погрешности в социологических исследованиях не должны превышать 2 %, по мнению этих авторов.

Традиционно в них большую роль играло анкетирование. В настоящее время благодаря новым техническим возможностям существуют бесконтактные формы исследований [Крыштановская, 2018]. О.В. Крыштановская использовала кластерный анализ для получения четырех кластеров по результатам работы для 841 аккаунта и показала, что социальные взаимоотношения, особенно в области политической жизни, не связаны с активностью аккаунтов в социальных сетях.

Очевидно, что невозможно обойтись без использования математических моделей для анализа социально-экономических взаимодействий и систем. Однако в силу сложности таких систем стохастическое моделирование социально-экономических процессов полностью осуществить невозможно, так как не существует формул, описывающих структуру организации [Мамченко, Акимочкина, Половникова, Шаповалова, 2019]. Для таких процессов больше подходит метод анализа иерархий Т. Саати [Саати, 1993], так как математическая правомочность решающего правила в нем прозрачна и базируется на методе собственного значения и принципе иерархической композиции, имеющих четкое математическое обоснование. Заметим, что метод анализа иерархий состоит в дроблении проблемы на более простые составляющие части и в дальнейшей математической обработке последовательности суждений эксперта по парным сравнениям, оформленным в виде матриц. Использование этого метода приведено в работе Т. Дивиной, Е. Петраковой и А. Черкова на примере разработки балльно-рейтинговой системы оценки знаний с применением метода иерархий [Divina, Petrakova, Cherkov, 2022].

Предлагаемая система оценки результатов обучения позволит преподавателю в дальнейшем объективно оценивать типы заданий, выполняемых учащимися, с учетом их сложности, трудоемкости, интенсивности, важности с точки зрения приобретения новых компетенций.

Студенты, изучая критерии оценки знаний по такой системе, познакомятся с методом иерархий Т. Саати и при желании смогут применять его в будущем в своей профессиональной деятельности.

В работе О. Хлыстовой и Ю. Калюжной предложена математическая модель, основу которой составляет система из двух уравнений регрессии для анализа отраслевых показателей Европейского союза с использованием 1 397 наблюдений региональной статистики Евростата и Европейского социального исследования [Khlystova, Kalyuzhnova, 2023]. Результаты их работы показывают, что регионы с более высокой долей занятых в креативных индустриях с большей вероятностью выдержат краткосрочный экономический шок, чем регионы с более низкой долей занятых в креативных индустриях. Итоги этого исследования указывают на важность цифровизации в период выхода из кризиса и демонстрируют, что данные индустрии играют важную роль в этом процессе.

Ц. Чжан, Д. ван Горп, Х. Киевит показали, что уровень развития цифровых технологий положительно связан с результатами деятельности национальных предпринимательских экосистем, и эта положительная взаимосвязь укрепляется в странах с благоприятной культурой, высококачественными учреждениями, поддерживающей политикой, доступными ресурсами и хорошо развитой отраслью услуг [Zhang, Gorp, Kievit, 2022]. Этот вывод был сделан на основе анализа социально-экономических данных для 101 страны за период с 2001 г. по 2018 г. В работе считаются коэффициенты корреляции между изучаемыми показателями и проверяется значимость этих коэффициентов.

Регрессионный анализ был применен в работе Г. Тота, З. Элекса, А. Уиттла, Ч. Ли [Toth, Elekes, Whittle, Lee, 2022]. Результаты регрессионного анализа указывают на то, что регионы Италии с более развитой структурой сети технологических знаний демонстрируют более высокий уровень устойчивости к изменениям в уровне занятости.

В.Е. Конюхова и Д.А. Вафоев считают, что математическое моделирование является основным методом в исследовании цифровой экономики [Конюхова, Вафоев, 2023]. В работе подчеркивается, что огромную роль играют методы машинного обучения, которые обрабатывают и классифицируют данные.

Нейронные сети (далее – нейросеть, один из методов машинного обучения) были использованы в работе Е.В. Щекотина, В.Л. Гойко, П.А. Басиной, В.В. Бакулина для изучения качества жизни населения регионов РФ [Щекотин, Гойко,

Басина, Бакулин, 2022]. Однако оценка качества модели ( $f1\text{-macro} = 0,545$ ) невысокая.

Стоит отметить, что нейросети показывают хороший результат при классификации, если используется значительный объем выборки (сотни тысяч, миллионы) [Вакуленко, Жихарева, 2018].

В большинстве случаев базы данных при социологических исследованиях содержат десятки, сотни, реже тысячи элементов (объектов). Тогда нейросети использовать нельзя (например, для анализа данных, содержащихся в открытых сборниках Федеральной службы государственной статистики (далее – Росстат). Л.Р. Борисова и А.В. Кузнецова подробно описали преимущества алгоритма анализа таких баз данных с применением программного комплекса, разработанного компанией «Азфорус» [Борисова, Кузнецова, 2024].

В нашей работе представлены результаты распознавания регионов РФ по набору показателей электронных услуг и сервисов за май 2024 г.<sup>1</sup>

## МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ / RESEARCH METHODOLOGY

### МАТЕРИАЛЫ ИССЛЕДОВАНИЯ / RESEARCH MATERIALS

Всего в базу данных вошли 6 показателей, общее количество объектов – 83, количество объектов первой группы – 42; это регионы, в которых на электронные услуги и сервисы в мае 2024 г. приходилось менее медианного уровня, равного 12,4 руб. на душу населения; количество объектов второй группы – 41; это субъекты России с расходами на электронные услуги и сервисы выше медианного уровня. Новые четыре региона РФ не представлены в данных Росстата, поэтому их в статье не учитывали<sup>2</sup>.

### МЕТОДЫ ИССЛЕДОВАНИЯ / RESEARCH METHODS

Для сравнения двух выборок малого размера был использован классический непараметрический критерий Манна-Уитни-Вилкоксона. Для разделения регионов на кластеры без указания количества кластеров применили пошаговый компилятор – язык программирования R, в библиотеке статистических программ которого есть и программа для применения кластерного анализа.

При использовании методов машинного обучения применяли метод статистически взвешенных синдромов (далее – СВС), разработанный

<sup>1</sup> Федеральная служба государственной статистики. Официальный сайт. Режим доступа: <https://rosstat.gov.ru/> (дата обращения: 25.07.2024).

<sup>2</sup> Там же.

компанией Azforus [Кузнецова, Сенько, 2005; Кузнецова, Борисова, Кузнецова, Сенько, 2017; Senko, Kuznetsova, 2010]. Он основан на процедуре взвешенного голосования по системам так называемых синдромов – областей признакового пространства, содержащих преимущественно объекты одного из классов [Сенько, Кузнецова, Воронин, Кравцова, Добролюбова, Борисова и др., 2022]. Данный метод относится к числу логико-статистических методов, дающих наиболее высокий процент распознавания при наличии пропусков в небольших выборках, что является преимуществом в цифровых исследованиях на региональном уровне, когда число объектов исследования не превышает 100. Логико-статистический подход позволяет проводить анализ, не делая априорных предположений о виде вероятностных распределений, а также такой подход эффективен в условиях выборки любого размера и большого количества плохо структурированных признаков [Борисова, Кузнецова, Сергеева, Сенько, 2020].

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ / RESEARCH RESULTS

### ПРИМЕНЕНИЕ МЕТОДОВ БАЗОВОЙ СТАТИСТИКИ И МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ДАННЫХ / APPLICATION OF BASIC STATISTICS AND MACHINE LEARNING METHODS FOR DATA ANALYSIS

В табл. 1 приведены значимые результаты применения непараметрического статистического критерия Манна-Уитни-Вилкоксона для разделения 83 регионов (в которых были данные по показателю «электронные услуги и сервисы, май 2024 г., руб. на душу населения»), выбранного в качестве группирующего показателя. Все показатели –

статистически значимые, так как величина p-value близка к нулю. Отметим, что их названия соответствуют названиям на сайте Росстата.

Значимый коэффициент множественной регрессии между группирующим показателем и показателями табл. 1, равен 0,71, что указывает на тесную связь между изучаемыми признаками (p-value = 0). На скользящем контроле правильно распознано 15 объектов из общего количества 18 (в контрольную выборку из общей были случайно отобраны 18 объектов). В результате распознавания получены три ошибки: Тамбовская область была отнесена к первому классу (с тратами населения на электронные услуги и сервисы в мае ниже медианного уровня), Свердловская область и Забайкальский край – ко второму, хотя значение группирующего показателя лишь немного отличается от медианного, поэтому эти две ошибки не удивительны. Результаты по качеству классификации представлены в табл. 2. Высокий уровень площади под ROC-кривой (англ. receiver operating characteristic – рабочая характеристика приемника), равный 0,88, существенно больше порогового значения 0,5, когда распознавание провести не удастся. Остальные показатели качества распознавания тоже близки к 1, поэтому регионы РФ, в зависимости от расходов на электронные сервисы и услуги, хорошо разделяются на две группы. В процессе подготовки статьи к публикации на сайте Росстата появились свежие данные по цифровизации услуг для населения за летний период. Эти данные (за июль 2024 г.) были проанализированы на предмет сравнения множественного коэффициента цифровизации электронных услуг и сервисов. Величина 0,71 не изменилась.

Таблица 1. Тест Манна-Уитни-Вилкоксона

Table 1. Mann-Whitney-Wilcoxon test

Показатель	Среднее в первой группе	Стандартное отклонение в первой группе	Среднее во второй группе	Стандартное отклонение во второй группе	p-value
Численность исследователей, имеющих ученую степень, на 10 тыс. чел. населения	2,93	3,2	5,32	5,5	0,007
Медицинские услуги, тыс. руб. на душу населения	0,56	0,16	0,81	0,39	0,0004
Услуги учреждений культуры, май 2024 г., тыс. руб. на душу населения	0,08	0,04	0,15	0,1	0,000

Окончание табл. 1

Показатель	Среднее в первой группе	Стандартное отклонение в первой группе	Среднее во второй группе	Стандартное отклонение во второй группе	p-value
Услуги телекоммуникационные, май 2024 г., тыс. руб. на душу населения	0,78	0,2	0,99	0,32	0,003
Услуги системы образования, май 2024 г., тыс. руб. на душу населения	0,387	0,1	0,53	0,27	0,007

Составлено автором по материалам исследования / Compiled by the author on the materials of the study

**Таблица 2. Результаты классификации по методу СВС**

Table 2. Results of classification by the method of statistically weighted syndromes

Правильность	Точность	Чувствительность	Специфичность	AUC
0,8333	0,8750	0,7778	0,8889	0,88

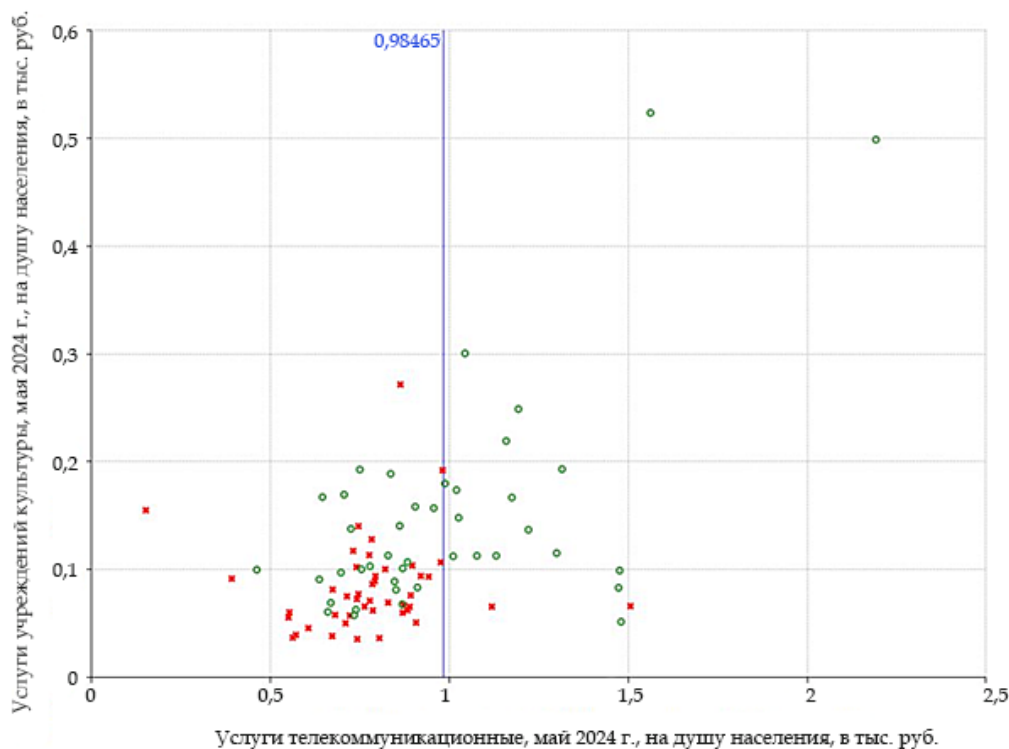
Примечание: AUC – area under curve (англ. площадь под кривой)

Составлено автором по материалам исследования / Compiled by the author on the materials of the study

На диаграмме рассеяния (рис. 1) наглядно представлены значения для регионов первой группы (красные крестики) и второй (зеленые кружки) по соответствующим показателям. Границы разбиения поставлены таким образом, чтобы с одной стороны от границы преобладали объекты одного класса. Чтобы перевести регион из первой

группы во вторую, нужно увеличить расходы населения на телекоммуникационные услуги (они должны превышать значение 0,98 тыс. руб. на душу населения в месяц).

<sup>3</sup> Федеральная служба государственной статистики. Официальный сайт. Режим доступа: <https://rosstat.gov.ru/> (дата обращения: 25.07.2024).



Примечание: в первом квадранте преобладает первый класс, во втором квадранте – второй

Составлено автором по материалам источника<sup>3</sup> / Compiled by the authors on the materials of the source<sup>3</sup>

**Рис. 1. Диаграмма рассеяния по предоставлению электронных услуг**

Fig. 1. Scattering diagram for the provision of electronic services

## ИССЛЕДОВАНИЕ ЦИФРОВИЗАЦИИ УСЛУГ С ПРИМЕНЕНИЕМ КЛАСТЕРНОГО АНАЛИЗА / RESEARCH ON DIGITALISATION OF THE SERVICES USING CLUSTER ANALYSIS

Применим кластерный анализ для этих данных. На рис. 2 представлена кластерная диаграмма, полученная при использовании необходимой библиотеки library (cluster) в пошаговом компиляторе – языке программирования R, который является очень популярным для компьютерного анализа данных<sup>4</sup>.

Если разрезать кластерную диаграмму на условной высоте, равной 6, то получим три региона (под номерами 28, 30, 43) на рис. 2, которые выделяются среди всех остальных регионов. Отметим, что условная высота, представленная на оси у на рис. 2 характеризует уровень отличия наблюдений: чем больше разница в высоте между ними, тем больше отличия [Матвиевский, Борисова, 2023]. Указанные ранее регионы – это Санкт-Петербург, Краснодарский край и Татарстан, что не удивительно, так как в них заметен уровень цифровизации. Конечно же, Москва выделяется отдельно (№ 28 на дендрограмме). Если разрезать дендрограмму на высоте 10, то получим всего два кластера: Москву в одном кластере и все остальные регионы России – в другом.

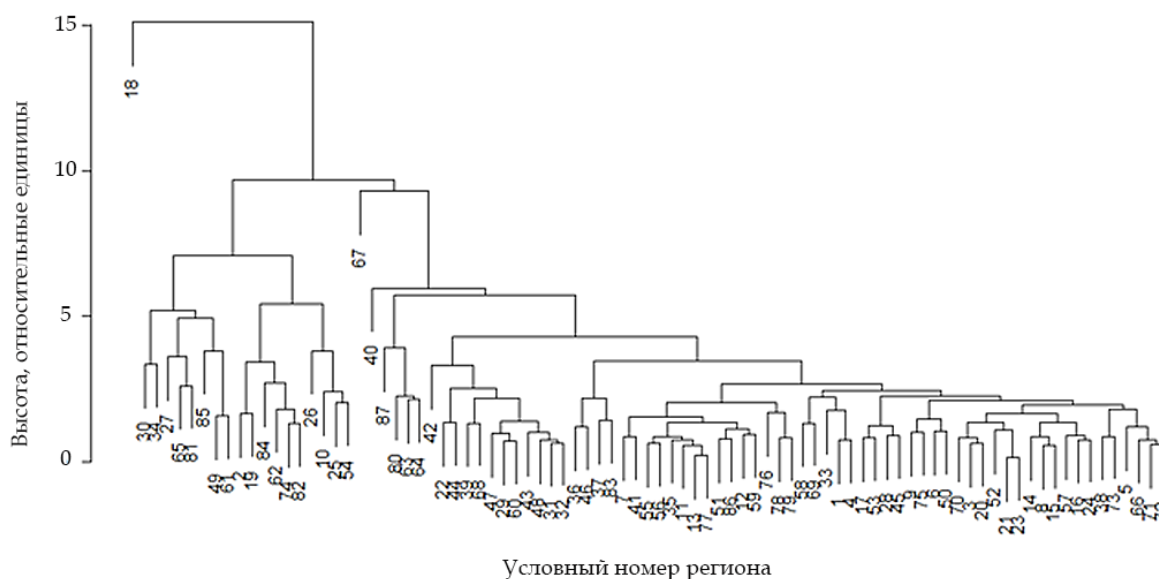
<sup>4</sup> Федеральная служба государственной статистики. Официальный сайт. Режим доступа: <https://rosstat.gov.ru/> (дата обращения: 25.07.2024).

## ВЫВОДЫ / FINDINGS

Наше исследование показало, что регионы в целом распознаются по показателям цифровых услуг среди регионов РФ. Индикатор качества распознавания ROC AUC = 0,88. Чувствительность – 78 %. Специфичность – 89 %.

Отметим следующие регионы, в которых на электронные услуги и сервисы в мае 2024 г. на душу населения было потрачено меньше медианного уровня средств: Чеченская республика, Челябинская, Курская, Тверская, Ленинградская, Тульская, Кемеровская, Смоленская, Астраханская, Архангельская, Томская, Орловская, Владимирская, Ульяновская, Брянская, Саратовская, Калининградская области, Республика Карелия, Республика Мордовия, Еврейская автономная область, Ставропольский, Алтайский края и др. Возможно, из-за наличия большого числа удаленных от районных центров сел и деревень в этих регионах недостаточно развит цифровой сервис.

Можно предположить, что на пользование цифровыми услугами влияет уровень среднедушевого дохода. Статистически значимый коэффициент корреляции (по критерию Фишера) между показателями «электронные услуги и сервисы в мае 2024 г. на душу населения» и «среднедушевые денежные доходы населения за один месяц по данным за первый квартал 2024 г.» равен 0,39 (значимость критерия – 0,002), что говорит о наличии такой связи, но достаточно умеренной.



Составлено автором по материалам источника<sup>4</sup> / Compiled by the authors on the materials of the source<sup>4</sup>

**Рис. 2. Кластерная диаграмма, соответствующая данным по электронным услугам и сервису в мае 2024 г. в регионах РФ**

Fig.2. Cluster diagram corresponding to the data on electronic services and service in May 2024 in the Russian regions

Сделаем вывод, что, используя диаграммы рассеяния для пар показателей, характеризующих цифровой сервис, можно получить пороговые значения, перейдя через которые мы в состоянии улучшить статистику.

Заметим, что показатель Росстата «численность исследователей, имеющих ученую степень, на 10 тыс. населения» был неслучайно использован при машинном обучении. Таким образом мы пытались учесть при анализе влияние цифровизации на наукоемкую сферу деятельности.

В исследовании основное внимание было уделено использованию статистических методов обработки оцифрованных данных, включая методы машинного обучения, которыми предлагается завершать математическую обработку данных. Начинать исследование всегда лучше с корреляционного и регрессионного анализа, чтобы выявить наиболее достоверные (с точки зрения математической статистики) признаки, которые в дальнейшем необходимо использовать в популярных методах цифровой классификации объектов. В работе было предложено для этой цели применять непараметрический тест Манна-Уитни-Вилкоксона и множественную регрессию. Подход оказался удачным, так как качество распознавания регионов с использованием метода СВС было высоким, о чем свидетельствует величина площади под ROC-кривой, равная 0,88, то есть близкая к 1. Заметим, что в соответствии с математическими терминами подразумевается, что площадь под ROC-кривой дается в безразмерных единицах или в %. Интересно отметить: несмотря на высокое качество распознавания цифровых данных по электронным услугам и сервисам, на кластерной дендрограмме такого четкого разделения всех регионов РФ по исследуемым признакам не просматривается. Это свидетельствует о том, что одной иерархической кластеризации недостаточно – следует использовать более сложные математические алгоритмы, а именно методы машинного обучения.

Развитие информационных цифровых технологий в регионах должно способствовать решению главной цели концепции региональной информатизации – повышению уровня жизни граждан. Повсеместное использование таких основных методов математической обработки цифровых данных, как методы машинного обучения, может благоприятствовать решению этой проблемы.

Однако отметим полезность кластерного анализа для быстрой визуализации оцифрованных данных. На дендрограмме выделяются четыре региона: Москва, Санкт-Петербург, Краснодарский край и Татарстан. Они отличаются от всех

остальных субъектов по более успешному использованию цифровых услуг и сервисов. Безусловно, в области цифровой социологии в России уже немало сделано, особенно в указанных выше регионах. Тем не менее, если в сфере образования в интернете можно найти много доступных (в том числе бесплатных) курсов для повышения квалификации, например Coursera, то в сфере здравоохранения больше всего упор был сделан на создание электронных очередей, что, без сомнения, очень хорошо. Наверное, без преувеличения можно сказать, что данная проблема решена во всех регионах. Однако для большинства населения регионов, особенно Северо-Западного, Уральского, Сибирского, Дальневосточного федеральных округов, было бы полезным увидеть цифровые медицинские платформы для самостоятельного изучения полученных анализов с целью постановки предварительного диагноза или неподтверждения маловероятных заболеваний на основе цифровых данных, если поход к врачу с этой целью из-за удаленности проблематичен. Конечно, лучше создавать медицинские цифровые платформы постановки первичного диагноза с использованием методов машинного обучения. Возможно, в недалеком будущем мы будем свидетелями таких платформ.

## ЗАКЛЮЧЕНИЕ / CONCLUSION

Что касается цифровизации остальных услуг в области социальной сферы, то, как справедливо написал А.Н. Сухов, в этой сфере есть немало не до конца решенных вопросов теоретического и практического плана [Сухов, 2023]. Они связаны с неразработанностью показателей, объективностью полученных данных, с неадекватностью методик, процедурой проведения опросов и т.д.

Целями данной статьи выступали демонстрация использования методов математической статистики и машинного обучения, сравнение их между собой, выработка рекомендаций для их применения в различных областях цифровой социологии, что и было достигнуто.

Результаты проведенного исследования могут быть использованы органами государственной исполнительной власти в процессе актуализации государственных программ развития конкретных регионов в направлении развития цифровой экономики.

## СПИСОК ЛИТЕРАТУРЫ

- Ашманов И.С., Касперская Н.И. Цифровая гигиена. СПб.: Питер; 2021. 400 с.
- Бабинцев В.П., Серкина Я.И. «Цифровизация» и «дигитализация» социальной реальности в предметном поле социологии: проблема адекватности понятий. Знание. Понимание. Умение. 2022;4:80–91. <http://dx.doi.org/10.17805/zpu.2022.4.7>
- Бальчиндоржиева О.В., Золхоева М.В. Цифровая культура vs культурная уникальность? (к вопросу о сохранении китайской культурной идентичности). Социологические исследования. 2022;3:90–97. <https://doi.org/10.31857/S013216250016858-6>
- Борисова Л.Р., Кузнецова А.В. Анализ многопараметрических датасетов методами машинного обучения для создания рекомендательных систем в HR практике. В кн.: Цифровая трансформация социальных и экономических систем: материалы международной научно-практической конференции, Москва, 26 января 2024 г. М.: Московский университет имени С.Ю. Витте; 2024. С. 195–203.
- Борисова Л.Р., Кузнецова А.В., Сергеева Н.В., Сенько О.В. Применение методов машинного обучения для сравнения компаний арктической зоны РФ по экономическим критериям в соответствии с рейтингом полярного индекса. Компьютерные исследования и моделирование. 2020;1(12):201–215. <https://doi.org/10.20537/2076-7633-2020-12-1-201-215>
- Вакуленко С.А., Жихарева А.А. Практический курс по нейронным сетям: учебное пособие. СПб.: Университет информационных технологий, механики и оптики; 2018. 71 с.
- Грошев И.В., Красовский Ю.Д. Цифровая трансформация социальных исследований и разработок. Цифровая социология. 2018;1:9–17. <https://doi.org/10.26425/2658-347X-2018-1-9-17>
- Исаенко В.О., Рыбина М.Н. Внедрение системы CardioQVARK для персонализации лечения в условиях формирования экосистемы цифрового здравоохранения. Цифровая социология. 2018;1:35–40. <http://dx.doi.org/10.26425/2658-347X-2018-1-35-40>
- Конюхова В.Е., Вафоев Д.А. Роль математического моделирования в цифровой экономике. Международный журнал теоретических и прикладных вопросов цифровых технологий. 2023;1(3):121–126.
- Кропивецкий Э.Н., Дубоглазова Ю.А., Заборовский Д.А. Цифровизация услуг в социальной сфере: проблемы и перспективы. Экономические науки. 2022;4(209):258–264. <https://doi.org/10.14451/1.209.258>
- Крыштановская О.В. Бесконтактная социология: новые формы исследований в цифровую эпоху. Цифровая социология. 2018;1(1):4–8. <https://doi.org/10.26425/2658-347X-2018-1-4-8>
- Кузнецова А.В., Сенько О.В. Возможности использования методов Data Mining при медико-лабораторных исследованиях для выявления закономерностей в массивах данных. Врач и информационные технологии. 2005;1:49–56.
- Кузнецова А.В., Сенько О.В., Кузнецова Ю.О. Преодоление проблемы «черного ящика» при использовании методов машинного обучения в медицине. Врач и информационные технологии. 2018;7:74–80.
- 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: сборник научных трудов XIX Международной конференции DAMDID/RCDL 2017, Москва, 10–13 октября 2017 г. М.: Федеральный исследовательский центр «Информатика и управление» Российской академии наук; 2017. С. 493–495.
- Мамченко О.П., Акимочкина Т.А., Половникова Е.С., Шаповалова С.В. Моделирование социально-экономических процессов в условиях цифровизации экономики. Управление экономическими системами. 2019;5(123).
- Матвиевский С.С., Борисова Л.Р. Кластеризация стран Азиатско-Тихоокеанского региона по значениям инклюзивного экономического роста. Вестник университета. 2023;1:112–121. <https://doi.org/10.26425/1816-4277-2024-1-112-121>
- Петрова Н.П., Овечкина А.И. К вопросу о цифровой трансформации российской экономики. Известия Санкт-Петербургского государственного экономического университета. 2021;1:57–62.
- Романова Н.В. Цифровизация услуг в социальной сфере: проблемы и перспективы. Вестник УГНТУ. Наука, образование, экономика. Серия: Экономика. 2020;1(31):58–65.
- Саати Т. Принятие решений. Метод анализа иерархий. Пер. с англ. Р.Г. Вачнадзе. М.: Радио и связь; 1993. 278 с.
- Сенько О.В., Кузнецова А.В., Воронин Е.М., Кравцова О.Д., Добролюбова О.А., Борисова Л.Р. и др. Методы интеллектуального анализа данных в исследованиях эпидемии COVID-19. Журнал Белорусского государственного университета. Математика. Информатика. 2022;1:83–96. <https://doi.org/10.33581/2520-6508-2022-1-83-96>
- Сухов А.Н. Цифровизация социальной сферы и ее последствия. Социальная политика и социальное партнерство. 2023;8. <https://doi.org/10.33920/pol-01-2308-01>
- Щекотин Е.В., Гойко В.Л., Басина П.А., Бакулин В.В. Использование машинного обучения для изучения качества жизни населения: методологические аспекты. Цифровая социология. 2022;1(5):87–97. <https://doi.org/10.26425/2658-347X-2022-5-1-87-97>
- Divina T., Petrakova E. Cherskov A. Applying the hierarchy analysis method to evaluate learning outcomes. In: Proceedings Proceedings of the II International Scientific Conference on Advances in Science, Engineering and Digital Education (ASEDU-II), Krasnoyarsk, 28 October 2021. New York: AIP Publishing; 2022. <https://doi.org/10.1063/5.0104789>

- Khlystova O., Kaluzhnova Y.* The impact of the creative industries and digitalization on regional resilience and productive entrepreneurship. *The Journal of Technology Transfer*. 2023;48:1654–1695. <https://doi.org/10.1007/s10961-023-10020-2>
- Senko O., Kuznetsova A.* A recognition method based on collective decision making using systems of regularities of various types. *Pattern Recognition and Image Analysis*. 2010;2(20):152–162. <http://dx.doi.org/10.1134/S1054661810020069>
- Toth G., Elekes Z., Whittle A., Lee Ch.* Technology network structure conditions the economic resilience of regions. *Economic Geography*. 2022;4(98). <http://dx.doi.org/10.1080/00130095.2022.2035715>
- Zhang J, Gorp D. van, Kievit H.* Digital technology and national entrepreneurship: an ecosystem perspective. *The Journal of Technology Transfer*. 2022;48(3):1077–1105. <http://dx.doi.org/10.1007/s10961-022-09934-0>

## REFERENCES

- Ashmanov I.S., Kaspersky N.I.* Digital hygiene. St. Petersburg: Piter; 2021. 400 p. (In Russian).
- Babintsev V.P., Serkina Ya.I.* “Digital transformation” and “digitalization” of social reality in the subject field of sociology: the problem of the adequacy of concepts Knowledge. Understanding. Ability. 2022;4:80–91. (In Russian). <http://dx.doi.org/10.17805/zpu.2022.4.7>
- Balchindorzhieva O.V., Zolkhoeva M.V.* Digital culture vs cultural uniqueness? (on the issue of preserving Chinese cultural identity). *Sociological Studies*. 2022;3:90–97. (In Russian). <https://doi.org/10.31857/S013216250016858-6>
- Borisova L.R., Kuznetsova A.A.* Analysis of multiparametric datasets using machine learning methods to create recommendation systems in HP practice. In: *Digital transformation of social and economic systems: Proceedings of the International Scientific and Practical Conference, Moscow, January 2024*. Moscow: Moscow Witte University; 2024. Pp. 195–203. (In Russian).
- Borisova L.R., Kuznetsova A.V., Sergeeva N.V., Senko O.V.* Application of machine learning methods for comparing companies in the Arctic zone of the Russian Federation according to economic criteria in accordance with the rating of the polar index. *Computer Research and Modeling*. 2020;1(12):201–215. (In Russian). <https://doi.org/10.20537/2076-7633-2020-12-1-201-215>
- Divina T., Petrakova E. Cherskov A.* Applying the hierarchy analysis method to evaluate learning outcomes. In: *Proceedings Proceedings of the II International Scientific Conference on Advances in Science, Engineering and Digital Education (ASEDU-II), Krasnoyarsk, 28 October 2021*. New York: AIP Publishing; 2022. <https://doi.org/10.1063/5.0104789>
- Groshev I.V., Krasovsky Yu.D.* Digital transformation of social research and development. *Digital Sociology*. 2018;1:9–17. (In Russian). <https://doi.org/10.26425/2658-347X-2018-1-9-17>
- Isaenko V.O., Rybina M.N.* The introduction of the CardioQVARK system for the personalization of treatment in the context of the formation of a digital healthcare ecosystem. *Digital Sociology*. 2018;1:35–40. (In Russian). <http://dx.doi.org/10.26425/2658-347X-2018-1-35-40>
- Khlystova O., Kaluzhnova Y.* The impact of the creative industries and digitalization on regional resilience and productive entrepreneurship. *The Journal of Technology Transfer*. 2023;48:1654–1695. <https://doi.org/10.1007/s10961-023-10020-2>
- Konyukhova V.E., Vafoev D.A.* The role of mathematical modeling in the digital economy. *International Journal of Theoretical and Applied Issues of Digital Technologies*. 2023;1(3):121–126. (In Russian).
- Kropivetsky E.N., Duboglazova Yu.A., Zaborovsky D.A.* Digitalization of services in the social sphere: problems and prospects. *Economic Sciences*. 2022;4(209):258–264. (In Russian). <https://doi.org/10.14451/1.209.258>
- Kryshтанovskaya O.V.* Contactless sociology: new forms of research in the digital age. *Digital Sociology*. 2018;1(1):4–8. (In Russian). <https://doi.org/10.26425/2658-347X-2018-1-4-8>
- Kuznetsova A.V., Senko O.V.* Possibilities of using Data Mining methods in medical laboratory research to identify patterns in data arrays. *Doctor and information technologies*. 2005;1:49–56. (In Russian).
- Kuznetsova A.V., Senko O.V., Kuznetsova Yu.O.* Overcoming the black box problem when using machine learning methods in medicine. *Doctor and information technologies*. 2018;7:74–80. (In Russian).
- 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: Proceedings of the XIX International Conference DAMDID/RCDL’ 2017, Moscow, October 10–13, 2017*. Moscow: Federal Research Centre “Informatics and Management” of the Russian Academy of Sciences; 2017. Pp. 493–495. (In Russian).
- Mamchenko O.P., Akimochkina T.A., Polovnikova E.S., Shapovalova S.V.* Modeling of socio-economic processes in the context of digitalization of the economy. *Management of economic systems*. 2019;5(123). (In Russian).
- Matviyevsky S.S., Borisova L.R.* Clustering of the countries of the Asia-Pacific region according to the values of inclusive economic growth. *Vestnik universiteta*. 2023;1:112–121. (In Russian). <https://doi.org/10.26425/1816-4277-2024-1-112-121>
- Petrova N.P., Ovechkina A.I.* On the issue of the digital transformation of the Russian economy. *Proceedings of the St. Petersburg State University of Economics*. 2021;1:57–62. (In Russian).
- Romanova N.V.* Digitalization of social services: problems and prospects. *Bulletin of the USPTU. Science, education, economics. Economics series*. 2020;1(31):58–65. (In Russian).

- Saati T.* Decision-making. Method of hierarchy analysis. Trans. from Eng. R.G. Vachnadze. Moscow: Radio i svyaz; 1993. 278 p. (In Russian).
- Senko O., Kuznetsova A.* A recognition method based on collective decision making using systems of regularities of various types. *Pattern Recognition and Image Analysis*. 2010;2(20):152-162. <http://dx.doi.org/10.1134/S1054661810020069>
- Senko O.V., Kuznetsova A.V., Voronin E.M., Kravtsova O.D., Dobrolyubova O.A., Borisova L.R. et al.* Methods of data mining in studies of the COVID-19 epidemic. *Journal of the Belarussian State University. Mathematics. Computer Science*. 2022;1:83–96. (In Russian). <https://doi.org/10.33581/2520-6508-2022-1-83-96>
- Shchekotin E.V., Goiko V.L., Basina P.A., Bakulin V.V.* Using machine learning to study the quality of life of the population: methodological aspects. *Digital Sociology*. 2022;1(5):87–97. (In Russian). <https://doi.org/10.26425/2658-347X-2022-5-1-87-97>
- Sukhov A.N.* Digitalization of the social sphere and its consequences. *Social Policy and Social Partnership*. 2023;8. (In Russian). <https://doi.org/10.33920/pol-01-2308-01>
- Toth G., Elekes Z., Whittle A., Lee Ch.* Technology network structure conditions the economic resilience of regions. *Economic Geography*. 2022;4(98). <http://dx.doi.org/10.1080/00130095.2022.2035715>
- Vakulenko S.A., Zhikhareva A.A.* Practical course on neural networks: textbook. St. Petersburg: University of Information Technologies, Mechanics and Optics; 2018. 71 p. (In Russian).
- Zhang J, Gorp D. van, Kievit H.* Digital technology and national entrepreneurship: an ecosystem perspective. *The Journal of Technology Transfer*. 2022;48(3):1077–1105. <http://dx.doi.org/10.1007/s10961-022-09934-0>