Пространственная Экономика
2024. Том 20. № 2. С. 71—92

JEL: C21, O18, R12
УДК 332.1+911.3
https://dx.doi.org/10.14530/se.2024.2.071-092

Пространственная алгоритмическая предвзятость в социально-экономической кластеризации российских регионов

В.И. Блануца

Блануца Виктор Иванович
доктор географических наук
ведущий научный сотрудник
Институт географии им. В.Б. Сочавы СО РАН, ул. Улан-Баторская, 1, Иркутск, 664033, Российская Федерация
E-mail: blanutsa@list.ru
ORCID: 0000-0003-3958-216X

Аннотация. Принятие решений на основе сложных человеко-машинных алгоритмов может привести к дискриминации граждан по полу, расе и другим признакам. Однако в мировой науке отсутствует представление об алгоритм-обусловленной дискриминации граждан по месту их проживания. Это относится и к принятию алгоритмических решений по социально-экономическому развитию регионов. Поэтому целью нашего исследования стало обнаружение алгоритмической предвзятости в результатах социально-экономической кластеризации российских регионов. Для достижения цели потребовалось выявить в кластерном анализе чувствительные операции, способные привести к пространственной несправедливости, сформировать массив статей по социально-экономической кластеризации субъектов РФ (регионов), проанализировать все статьи на возможность существования алгоритмической предвзятости и идентифицировать российские регионы с потенциально предвзятым отношением к ним в результате кластеризации. Предложен термин «пространственная алгоритмическая предвзятость». С помощью авторского алгоритма семантического поиска в библиографических базах данных выявлено 604 статьи с эмпирическими результатами кластерного анализа российских регионов по социально-экономическим показателям. Приведена характеристика выявленных статей. Анализ всех статей показал, что алгоритмическая предвзятость наиболее проявляется в четырех операциях алгоритма кластеризации – развертывании концептуальной модели в оптимальный набор показателей, отборе регионов, выборе способа объединения регионов в кластеры и определении количества кластеров. По каждой операции представлены примеры дискриминируемых российских регионов. Указаны три направления дальнейших исследований: выявление пространственной несправедливости в стратегических документах и национальных целях развития, определение новых видов алгоритмической предвзятости, разработка рекомендаций по справедливой кластеризации регионов.

© Блануца В.И., 2024

Ключевые слова: региональное социально-экономическое развитие, кластерный анализ, дискриминация, пространственная несправедливость, регион, Российская Федерация

Благодарности. Исследование выполнено за счет средств государственного задания (номер регистрации темы АААА-А21-121012190018-2).

Для цитирования: Блануца В.И. Пространственная алгоритмическая предвзятость в социально-экономической кластеризации российских регионов // Пространственная экономика. 2024. Т. 20. № 2. С. 71–92. https://dx.doi.org/10.14530/se.2024.2.071-092

Spatial Algorithmic Bias in Socio-Economic Clustering of Russian Regions

V.I. Blanutsa

Viktor Ivanovich Blanutsa
Doctor of Sciences (Geography)
Leading Researcher
V.B. Sochava Institute of Geography SB RAS, 1 Ulan-Batarskaya St., Irkutsk, 664033, Russian Federation
E-mail: blanutsa@list.ru
ORCID: 0000-0003-3958-216X

Abstract. Decision-making based on complex human-machine algorithms can lead to discrimination of citizens based on gender, race and other grounds. However, in world science there is no idea of algorithmically conditioned discrimination of citizens by their place of residence. This also applies to the adoption of algorithmic decisions on the socio-economic development of regions. Therefore, the purpose of our study was to detect algorithmic bias in the results of socio-economic clustering of Russian regions. To achieve this goal, it was necessary to identify sensitive operations in cluster analysis that could lead to spatial injustice, form an array of articles on socio-economic clustering of subjects (regions) of the Russian Federation, analyze all articles for the possibility of algorithmic bias and identify Russian regions with potentially biased attitudes towards them as a result of clustering. The term ‘spatial algorithmic bias’ is proposed. Using the author’s semantic search algorithm in bibliographic databases, six hundred articles with empirical results of cluster analysis of Russian regions by socio-economic indicators were identified. The characteristics of the identified articles are given. The analysis of all the articles showed that algorithmic bias is most evident in the four operations of the clustering algorithm – deploying a conceptual model into an optimal set of indicators, selecting regions, choosing a way to combine regions into clusters and determining the number of clusters. Examples of discriminated Russian regions are presented for each operation. Three directions of further research are indicated. Practical significance may be associated with the adoption of unbiased decisions on regional socio-economic development based on fair clustering of the Russian Federation’s subjects.

Keywords: regional socio-economic development, cluster analysis, discrimination, spatial injustice, region, Russian Federation

Acknowledgements. The research was carried out at the expense of the state assignment (topic registration number AAAA-A21-121012190018-2).

For citation: Blanutsa V.I. Spatial Algorithmic Bias in Socio-Economic Clustering of Russian Regions. *Prostranstvennaya Ekonomika = Spatial Economics*, 2024, vol. 20, no. 2, pp. 71–92. https://dx.doi.org/10.14530/se.2024.2.071-092 (In Russian)

ВВЕДЕНИЕ

Кластерный анализ предназначен для выявления скоплений объектов (кластеров) в многомерном признаковом пространстве. Возможна также кластеризация признаков в пространстве всех объектов, но она не рассматривалась в нашем исследовании. Термин «кластерный анализ» предложен американским психологом Р. Трайоном (Tryon, 1939), а сам метод постоянно усовершенствовался и применялся в различных научных областях, включая региональные социально-экономические исследования (López-Villuendas, del Campo, 2023; Soares, Coutinho, 2010). В таких работах объектом является регион (в современных отечественных исследованиях это субъект РФ). Другие территориальные объекты – населенные пункты, муниципальные образования, страны – не анализировались.

Еще одно ограничение связано с рассмотрением только традиционного кластерного анализа, в котором задается метрика (расстояние, мера различия или сходства между объектами) в многомерном признаковом пространстве и используются статистические методы выделения кластеров (Giordani et al., 2020; Mirkin, 1996; Xu, Tian, 2015). Поэтому обнаружение кластеров регионов при оценке пространственной автокорреляции (Акбердина и др., 2023; Тимирьянова и др., 2018) и машинном обучении (Блануца, 2020; Сорокина и др., 2017) не анализировалось. Также исключались работы, в которых группы регионов назывались кластерами, но их идентификация осуществлялась без применения методов кластерного анализа.

Социально-экономическая кластеризация регионов получила широкое распространение благодаря предоставлению необходимого программного обеспечения (пакеты программ SPSS, STATISTICA, STATGRAPHICS и др.), что позволяло минимизировать необъективные мнения исследователей. Однако при изучении этических проблем автоматического анализа данных выявилась «алгоритмическая предвзятость», когда «алгоритм неравномерно распределяет выгоды и бремя между различными индивидами или группами» (Kordzadeh, Ghasemaghaei, 2022, p. 394). В большинстве случаев это относилось к алгоритмам машинного обучения (Gerdon et al.,

2022; Van Giffen et al., 2022), но проявилось и при использовании традиционных статистических методов, включая кластерный анализ (Pan, Zhong, 2023).

Выявленные случаи алгоритмической предвзятости связаны с дискриминацией граждан по полу, расе, национальности, языку, возрасту, образованию и благосостоянию (Gerdon et al., 2022; Jackson, 2021; Johnson, 2021; Kordzadeh, Ghasemaghaei, 2022; Zou, Schiebinger, 2018). Несмотря на возрастающее количество публикаций по этой проблематике, в мировой науке еще не анализировалась потенциально возможная дискриминация граждан по месту их проживания, вызванная алгоритмическим принятием решений (в том числе при кластеризации регионов). Ближайшими научными направлениями, связанными с известными случаями алгоритмической предвзятости, являются выявление несправедливости в рекомендательных системах, основанных на местоположении (Sánchez et al., 2023), оценка искажений в геопарсинге, извлекающем топонимы из текста и привязывающем их к географическим местоположениям (Liu et al., 2022), идентификация предубеждений в геоинформационных системах (Wagner et al., 2021) и анализ процессов «расообразования пространства», локализующих бедность (Safransky, 2020).

Исходя из приведенных ограничений и уточнений, целью нашего исследования стало обнаружение алгоритмической предвзятости в результатах социально-экономической кластеризации российских регионов, опубликованных в научных журналах. Для достижения этой цели потребовалось решить следующие задачи: выявить в кластерном анализе чувствительные операции, способные привести к пространственной несправедливости; сформировать массив статей по социально-экономической кластеризации субъектов РФ; проанализировать все статьи на возможность существования алгоритмической предвзятости; идентифицировать российские регионы с потенциально предвзятым отношением к ним в результате кластеризации. Основные итоги проделанной работы представлены в следующих трех разделах данной статьи. В связи с необходимостью позиционирования нашего исследования среди других разработок по выявлению алгоритм-обусловленной дискриминации граждан предлагается новый термин – «пространственная алгоритмическая предвзятость», обозначающий несправедливое отношение к отдельным территориям (регионам), способное привести к нарушению (ограничению) прав, свобод и интересов жителей этих территорий при принятии решений по результатам использования алгоритма (кластерного анализа).

КЛАСТЕРНЫЙ АНАЛИЗ

Алгоритм кластеризации подробно описан в специальной литературе (Aldenderfer, Blashfield, 1984; Everitt et al., 2001; Giordani et al., 2020; Hartigan, 1975; Mirkin, 1996; Xu, Wunsch, 2005) и состоит из следующих последовательных действий (операций, процедур, методов): формирование концептуальной модели (что и для чего изучается), определение набора показателей (признаков, индикаторов изучаемого процесса), отбор анализируемых объектов (регионов), сбор исходных данных (фиксирование значения каждого показателя в каждом регионе), оптимизация данных (нормализация, стандартизация, исключение коррелируемых показателей, снижение размерности признакового пространства), установление меры различия (сходства) объектов, выбор способа идентификации кластеров, распределение объектов по кластерам, верификация и интерпретация результатов кластеризации. При этом десять основных операций сопровождаются различными вспомогательными (неосновными) процедурами, набор которых зависит от специфики исходных данных и способов их обработки.

Среди основных операций наиболее разнообразный инструментарий характерен для оценки различия (расстояния) между объектами и идентификации кластеров. Многомерность изучаемых объектов обуславливает необходимость введения некоторой метрики, позволяющей судить о степени различия этих объектов. Выбор метрики зависит от типа шкалы измерения (номинальная, порядковая, интервальная, абсолютная) или сочетания типов для смешенных данных (Foss et al., 2019). Наиболее используемыми расстояниями являются евклидово, квадратичное евклидово, манхэттенское, Чебышева и Махаланобиса (Everitt et al., 2001; Hartigan, 1975; Mirkin, 1996; Xu, Wunsch, 2005). Расстояния между кластерами и критерии выделения кластеров определяются с помощью различных методов (Xu, Tian, 2015) – иерархических (объединяющих, разделяющих) или неиерархических (итеративных, плотностных, модельных и др.), четких (жестких) или нечетких (мягких), многомерно-статистических или теоретико-графовых. Наиболее часто применяются (Hadi, 2022; Soares, Coutinho, 2010; Xu, Tian, 2015) методы k-средних (MacQueen, 1967), c-средних (нечеткая кластеризация), Уорда (Ward, 1963), DBSCAN (Ester et al., 1996), одиночной, полной и средней связи (Soares, Coutinho, 2010). Возможно также параллельное или последовательное применение нескольких методов (Oduntan, Thulasiraman, 2022).

Обобщение публикаций по кластерному анализу показало, что в каждой основной операции возможна алгоритмическая предвзятость. На чувствительные операции кластерного анализа, выполнение которых может привести к дискриминации некоторых групп граждан, указывается в публикаци-

ях по машинному обучению, посвященных «справедливой кластеризации» (Chabra et al., 2021; Gupta et al., 2023; Liu et al., 2023). Предлагаются разные способы учета интересов «защищенных групп» (выделяются по полу, расе и т. д.) и обеспечения их представительства в полученных кластерах. При этом игнорирование интересов некоторых групп возможно до, во время и после машинного обучения, а наиболее чувствительными действиями являются формирование обучающей выборки (исходных данных), объединение объектов в кластеры и интерпретация результатов. Здесь следует отметить, что эти чувствительные операции относятся к существующему пониманию алгоритм-обусловленной несправедливости, а для определения пространственной предвзятости, что делается впервые, необходимо проанализировать опыт кластеризации регионов с позиции имеющегося знания о проблемах кластерного анализа непространственных объектов.

КЛАСТЕРИЗАЦИЯ РЕГИОНОВ РОССИИ

С помощью авторского алгоритма семантического поиска необходимых публикаций, опирающегося на идеи машинного обучения (Бланца, 2020), в отечественных и зарубежных библиографических базах данных обнаружено 604 научные статьи. Отбирались статьи с эмпирическими результатами традиционного кластерного анализа субъектов РФ по социально-экономическим показателям (публикации только по социальным показателям не учитывались). Несмотря на некоторые колебания ежегодного количества статей (*рис. 1*), имел место восходящий тренд, свидетельствующий о росте научного интереса к рассматриваемой кластеризации (в 2019–2023 гг. опубликовано в 16,5 раза больше статей, чем в 2004–2008 гг.). В выявленных публикациях обнаружен 781 результат кластеризации, так как в некоторых статьях приводились итоги кластерного анализа по разным годам, наборам показателей и методам группировки регионов.

Статьи опубликованы в 306 научных журналах. Больше всего публикаций размещено в журналах «Региональная экономика: теория и практика» (23 статьи), «Экономика региона» (20) и «Вопросы статистики» (14). Авторами статей стали 902 исследователя из 253 организаций, среди которых выделялись Российский экономический университет им. Г. В. Плеханова (30,92 статьи; если были авторы из разных организаций, то статья как единица счета делилась на количество соавторов), Воронежский государственный университет (26,58) и Финансовый университет при Правительстве Российской Федерации (20,02). Эти организации находились в 87 городах, а больше всего статей опубликовали авторы из Москвы (132,49 статьи), Воронежа (30,92), Екатеринбурга (23,39), Новосибирска (21,67) и Уфы (20,42).

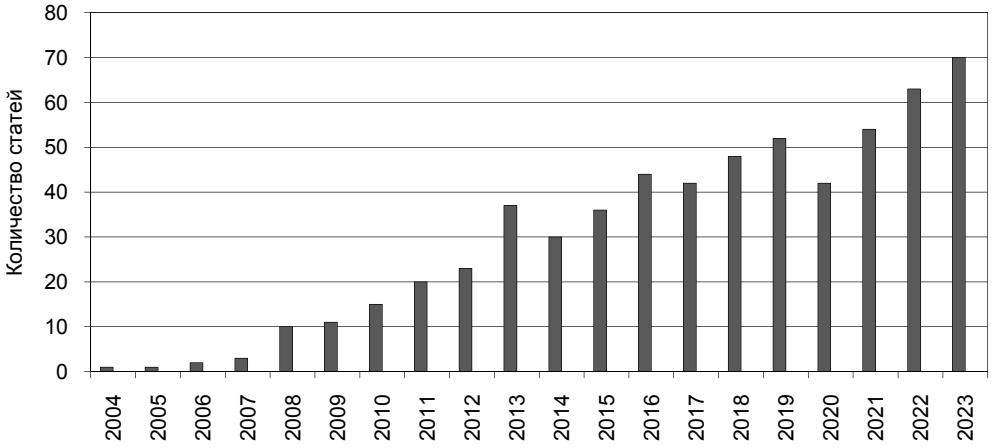

Рис. 1. Изменение ежегодного количества статей по социально-экономической кластеризации регионов России, опубликованных в отечественных и зарубежных научных журналах в 2004–2023 гг.

Fig. 1. The change in the annual number of articles on socio-economic clustering of Russian regions published in domestic and foreign scientific journals in 2004–2023

Источник: составлено автором.

В связи с анализом разного количества регионов все публикации и результаты кластеризации разделены на три массива, обозначенные «А» (учитывались все регионы России; 414 статей, 515 результатов), «Б» (регионы одного федерального округа, иногда двух-трех округов; 144, 187) и «В» (выборка регионов; 46, 79). Последний массив объединил публикации по кластерному анализу регионов из разных федеральных округов, но значительно меньшего количества относительно всех субъектов РФ (на 1 января 2022 г. их было 85). К массиву «В» относились как произвольные наборы регионов, так и специализированные группы (например, северные, угледобывающие, нечерноземные регионы). В качестве целей кластеризации российских регионов выступали, к примеру, определение уровня социально-экономического развития, оценка инвестиционной привлекательности, формирование инновационного потенциала, обеспечение экономической безопасности, обоснование приоритетов региональной политики, идентификация источников экономического роста, измерение пространственного неравенства, диагностика конкурентоспособности и выявление межрегиональной интеграции.

АЛГОРИТМ-ОБУСЛОВЛЕННАЯ ПРОСТРАНСТВЕННАЯ НЕСПРАВЕДЛИВОСТЬ

Кластерный анализ не должен приводить к дискриминации граждан Российской Федерации, проживающих в тех или иных субъектах РФ. Однако обобщение выявленных статей показало, что существует как минимум четыре операции кластеризации с потенциальной опасностью несправедливого отношения к отдельным российским регионам. Возможно, в ходе дальнейших исследований определятся и другие алгоритмически предвзятые действия, но для начала изучения этой проблематики ограничимся краткой характеристикой четырех операций и некоторыми примерами дискриминируемых регионов.

Развертывание концептуальной модели в оптимальный набор показателей. В многомерном статистическом анализе весьма важно, чтобы существовало понимание специфики изучаемого социально-экономического явления и обоснованное представление этого явления в виде необходимого и достаточного набора показателей. Присутствие некоторых случайных (второстепенных) показателей или отсутствие ряда классификационных (первостепенных) признаков может привести к выделению кластеров, не соответствующих явлению. Следует отметить, что данная операция выполнялась в каждом исследовании по кластеризации регионов, но при публикации результатов не всегда подробно излагалась. Поэтому вызывает удивление подбор показателей в отдельных статьях. Например, при изучении межрегиональной миграции рабочей силы выборка из 20 регионов была разделена на 9 кластеров (Ларченко, 2023), в пяти из которых уровень привлекательности региона среди прочего объяснялся плотностью населения (это «привлекало», к примеру, в Чукотский автономный округ, Камчатский край и Магаданскую область).

Для выявления потенциально дискриминируемых регионов в каждом случае кластеризации необходимо провести контрольный кластерный анализ по оптимальному набору показателей. Однако из выявленных статей не представляется возможным получить все данные, необходимые для такой проверки. Поэтому ограничимся косвенной верификацией, заключающейся в определении преобладающей (инвариантной) схемы распределения регионов по кластерам. Если учесть, что в большинстве случаев получается упорядоченная последовательность кластеров, то наиболее значимо попадание региона в число лидеров или аутсайдеров. Тогда имеет смысл свести все выделяемые группы регионов к трем кластерам, с некоторой условностью названными «верхним» (группа с наилучшей ситуацией), «средним» (группы с промежуточными ситуациями или отсутствие групп при 2-кластерном

решении) и «нижним» (группа с наихудшей ситуацией). Сравнение конкретной кластеризации с подобным инвариантом позволит выявить регионы, лишенные статуса лидера или аутсайдера.

В качестве примера возьмем определение уровня социально-экономического развития регионов из массива «А» по статистическим данным за 2018–2022 гг. (валовой региональный продукт, его составляющие, а также дополнительные показатели). Этому посвящено 14 статей (не рассматривались публикации с неполным списком распределения регионов по кластерам). На их основе по частоте встречаемости региона (не учитывались субъекты РФ, выведенные авторами статей за рамки исследования) в верхнем, среднем и нижнем кластерах выделены ядра (не менее 0,75) и переходные зоны (0,51–0,74). Получились следующие результаты: ядро верхнего кластера представлено городом Москва (частота отнесения к верхнему, среднему и нижнему кластерам составляет, соответственно, 0,92; 0,08; 0), а в ядро нижнего кластера вошли Республика Калмыкия (0; 0,07; 0,93), Кабардино-Балкарская Республика (0; 0,14; 0,86), Карачаево-Черкесская Республика (0; 0,14; 0,86), Республика Северная Осетия – Алания (0; 0,14; 0,86), Республика Ингушетия (0; 0,15; 0,85) и Республика Дагестан (0; 0,21; 0,79); переходная зона между ядрами верхнего и нижнего кластеров представлена городом Санкт-Петербург (0,62; 0,38; 0), а между ядрами нижнего и среднего кластеров разместились Республика Тыва (0,07; 0,21; 0,72), Чеченская Республика (0; 0,29; 0,71), Еврейская автономная область (0; 0,29; 0,71), Республика Крым (0; 0,31; 0,69), Республика Алтай (0,07; 0,29; 0,64), г. Севастополь (0; 0,42; 0,58), Забайкальский край (0; 0,43; 0,57) и Псковская область (0; 0,43; 0,57). Таким образом, несмотря на различия в наборах показателей, отражающих уровень развития, имеется 16 регионов, в той или иной мере относимых исследователями к верхнему и нижнему кластерам.

На основе данных о частоте встречаемости регионов в верхнем и нижнем кластерах можно проводить предварительную идентификацию потенциально дискриминируемых регионов. Например, при изучении экономического неравенства (Лясковская и др., 2023) в кластер «наиболее богатых субъектов» не вошла столица России, а при анализе экономической безопасности (Быкова, 2023) в кластер с «наиболее низкими параметрами» не включены Карачаево-Черкесская Республика, Республика Северная Осетия – Алания и Республика Дагестан. Это дает основание для более глубокой проверки упомянутых исследований с целью определения возможной несправедливости в отношении отмеченных регионов.

Отбор регионов. Исключение из анализа любого региона нарушает целостность изучаемого экономического пространства – страны или федерального округа – и приводит к кластеризации, не соответствующей

действительности. Обобщение выявленных статей позволило установить четыре причины удаления региона – «второстепенность» субъекта РФ (исключение трех автономных округов, входящих в Тюменскую и Архангельскую области), отсутствие данных, аномальные значения имеющихся данных («выбросы») и неправильные априорные представления исследователей о кластерах (например, удаление кластеров, содержащих один регион). Это происходило на этапах формирования матрицы исходных данных, выделения кластеров и верификации результатов кластеризации. В итоге получилась не потенциальная, а реальная дискриминация ряда российских регионов.

Невключение региона в результирующее кластерное решение может привести к эффекту «сенсорной пустыни» (Robinson, Franklin, 2020), когда отсутствие сведений по некоторой территории приводит к отказу от выделения средств на развитие данной территории. Особое значение это имеет для дотационных регионов, но и для регионов-лидеров отсутствие в различных верхних кластерах может негативно повлиять на их имидж. В этом контексте весьма странно выглядит исключение трех автономных округов из числа всех субъектов РФ, что противоречит российской конституции. Весьма спорны также утверждения следующего типа: «Из расчетов исключены <…> Москва и Санкт-Петербург в силу их особого экономического и социального положения» (Трещевский и др., 2021, с. 89–90).

Анализ публикаций из массива «А» показал, что по разным причинам дискриминации подверглись 75 (из 85) регионов. По одному разу удалялись 19 регионов. Более одного раза исключались следующие регионы: Ненецкий (166 статей), Ямало-Ненецкий (162) и Ханты-Мансийский (158) автономные округа, города Москва (89) и Санкт-Петербург (65), Чеченская Республика (51), Чукотский автономный округ (43), г. Севастополь (41), Республика Ингушетия (39), Республика Крым (35), Московская (28) и Тюменская (18) области, Республика Тыва (17), Еврейская автономная область (16), Республика Дагестан (13), Республика Калмыкия (11), Магаданская (10), Сахалинская (10) и Ленинградская (9) области, Республика Татарстан (7), Волгоградская (6) и Архангельская (5) области, Республика Башкортостан (5), Нижегородская (5) и Новосибирская (5) области, Республика Саха (Якутия) (5), Свердловская область (5), Республики Сев. Осетия – Алания (5), Адыгея (4) и Алтай (4), Краснодарский (4) и Красноярский (4) край, Мурманская (4), Самарская (4), Челябинская (4) и Белгородская (3) области, Забайкальский край (3), Карачаево-Черкесская Республика (3), Кемеровская область – Кузбасс (3), Пермский край (3), Тамбовская (3), Амурская (2) и Астраханская области (2), Республика Бурятия (2), Воронежская область (2), Кабардино-Балкарская Республика (2), Калининградская область

(2), Республика Карелия (2), Курганская (2), Липецкая (2), Новгородская (2), Оренбургская (2), Ростовская (2) и Томская (2) области, Хабаровский край (2) и Республика Хакасия (2). Ни из одной кластеризации не исключались Республика Марий Эл, Удмуртская Республика, Владимирская, Иркутская, Кировская, Омская, Пензенская, Саратовская, Смоленская и Тульская области. В массиве «Б» удалялись 11 регионов, а больше всех – г. Москва (26) и Московская область (12). При выборке регионов (массив «В») необоснованные исключения не обнаружены.

Выбор способа объединения регионов в кластеры. В матрице исходных данных существует некоторая скрытая (латентная) структура связи регионов, которую можно представить как скопления точек (регионов) в многомерном признаковом пространстве в виде выпуклых (между любыми двумя точками скопления можно проложить линию внутри скопления; гиперсферы, эллипсоиды, кубоиды и др.) и невыпуклых (линия между двумя точками может выходить за пределы скопления; кольцеобразные, крестообразные и др.) форм (Hadi, 2022; Lorimer et al., 2017; Xu, Tian, 2015; Xu, Wunsch, 2005). Визуализация этих структур возможна за счет снижения размерности (до двух- или трехмерного) пространства с помощью метода главных компонент или факторного анализа. При этом не исключена ситуация равномерного или случайного расположения точек в пространстве, что указывает на отсутствие кластеров. Игнорирование или недостаточный учет латентной структуры может привести к неправильному выбору способа выделения кластеров и, как следствие, предвзятой кластеризации (Lorimer et al., 2017).

Из всех выявленных статей в 54 публикациях не упоминался метод идентификации кластеров (в том числе 37 статей в массиве «А» и 12 в «Б»), а делалась ссылка на пакет прикладных программ (проблема в том, что в пакетах предлагается на выбор несколько методов). В остальных статьях отмечено применение от одного до шести методов (724 случая в 550 статьях). Среди них выделялись пять способов объединения регионов в кластеры (*табл. 1*). При этом использовались различные многомерные расстояния между регионами (301 случай в 288 статьях), которые имели свои особенности и влияли на результат кластеризации. Предпочтение отдавалось четырем метрикам (*табл. 2*).

При социально-экономической кластеризации российских регионов латентные структуры не определялись, что заставляет усомниться в правильности полученных кластеров. Идентифицировать такие структуры по текстам статей не представляется возможным. Поэтому ограничимся примером, в котором результат кластерного анализа может быть связан с неправильным выбором способа формирования кластеров. Известно, что наиболее распространенный метод k-средних выявляет только выпуклые

и хорошо разделенные кластеры, чувствителен к априори заданному количеству кластеров и выбросам, а также зависит от начальных условий итерационного процесса, что может привести к локальному, а не глобальному оптимуму (Xu, Tian, 2015; Xu, Wunsch, 2005). Косвенным свидетельством присутствия невыпуклых образований в признаковом пространстве является наличие множества выбросов. В одном из исследований (Беркович и др., 2013) количество аномальных регионов достигло 20%. Возможно, в этом примере при использовании другого метода, определяющего любые формы кластеров (DBSCAN и др.), не было бы дискриминации 16 из 80 регионов России.

Таблица 1

Количество статей с различными методами выделения кластеров в массивах «А» (все регионы России), «Б» (регионы федерального округа) и «В» (выборка регионов)

Table 1

Number of articles with various methods of cluster identification in arrays «A» (all Russian regions), «Б» (regions of the federal district) and «В» (sample of regions)

Метод	Массивы статей			Всего
	«А»	«Б»	«В»	
K-средних	235	56	16	307
Уорда	154	50	20	224
Одиночной связи	24	22	4	50
Полной связи	36	22	12	70
Средней связи	30	12	9	51
Другие методы	14	5	3	22

Источник: составлено автором.

Таблица 2

Количество статей с различными многомерными расстояниями между регионами в массивах «А» (все регионы России), «Б» (регионы федерального округа) и «В» (выборка регионов)

Table 2

Number of articles with various multivariate distances between regions in arrays «A» (all regions of Russia), «Б» (regions of the federal district) and «В» (sample of regions)

Многомерное расстояние	Массивы статей			Всего
	«А»	«Б»	«В»	
Евклидово	142	63	12	217
Квадратичное евклидово	25	10	7	42
Манхэттенское	17	8	6	31
Чебышева	4	1	1	6
Другие расстояния	1	4	0	5

Источник: составлено автором.

В случае выбора многомерного расстояния, являющегося частью процедуры объединения регионов в кластеры, также необходимо учитывать специфику каждой метрики. Например, известно, что квадратичное евклидово расстояние усиливает различия в исходных данных, способствуя в некоторых случаях образованию выбросов (Xu, Wunsch, 2005). Тем не менее эта метрика применялась в социально-экономической кластеризации регионов (см. *табл. 2*). В качестве примера можно привести результат кластерного анализа неравномерности распределения национального богатства с использованием отмеченной метрики и метода Уорда (Лимонова, 2014) – два кластера и два выброса (г. Москва и Тюменская область с автономными округами). Некоторые исследователи пытались снизить предвзятость кластеризации за счет использования двух (10 статей) или четырех (1) расстояний с последующим сравнением полученных результатов. Более распространенным стало применение нескольких способов формирования кластеров – двух (104 статьи), трех (14), четырех (7), пяти (4) или шести (1).

Определение количества кластеров. Одной из наиболее проблемных операций кластерного анализа является идентификация оптимального числа кластеров. В методе k-средних это число задается до начала работы алгоритма, а в иерархических методах оно рассчитывается по дендрограмме (древовидная диаграмма последовательности объединения объектов в кластеры) или матрице расстояний между объектами. Существует много способов определения оптимального количества кластеров. Например, в NbClust (программная среда R) предлагается 30 индексов (Charrad et al., 2014). Однако в региональных исследованиях нередко встречается произвольный выбор числа кластеров. В качестве иллюстрации приведем два умозаключения: «Число кластеров было решено взять равное трем, чтобы выявить сильные, средние и слабые по своему социально-экономическому развитию подгруппы областей ЦФО» (Жолудева и др., 2014, с. 145); «Количество кластеров определяем из предположения, что 5 регионов ЦЧР должны сформировать отдельный кластер. Следовательно, получаем приближенно 3 группы, разделив 18 субъектов ЦФО на 5» (Ляхова, Григорян, 2017, с. 6). Таким образом, число кластеров обосновывалось не структурой исходных данных или статистическими индексами, а априорными предубеждениями исследователей.

В проанализированных статьях выделялось от 2 до 18 кластеров, но больше всего результатов имело три кластера (*рис. 2*). Если 3-кластерное решение для небольшого количества регионов в массивах «Б» (*рис. 3, Б*) и «В» (*рис. 3, В*) не является чем-то необычным, то деление всех российских регионов на три кластера (*рис. 3, А*) выглядит несколько странным при разных показателях, многомерных расстояниях и способах формирования кластеров. Скорее

всего, отсутствие обоснования числа кластеров во многих случаях приводит к интуитивному делению на верхний, средний и нижний кластеры. Для выявления такой предвзятости в каждом конкретном случае необходимо провести повторную кластеризацию по исходным данным с использованием тех же многомерных расстояний и способов образования кластеров, но с обязательным вычислением оптимального числа кластеров. Сравнение первоначальных и повторных результатов позволит определить алгоритмическую предвзятость в виде смещения регионов из одних в другие кластеры.

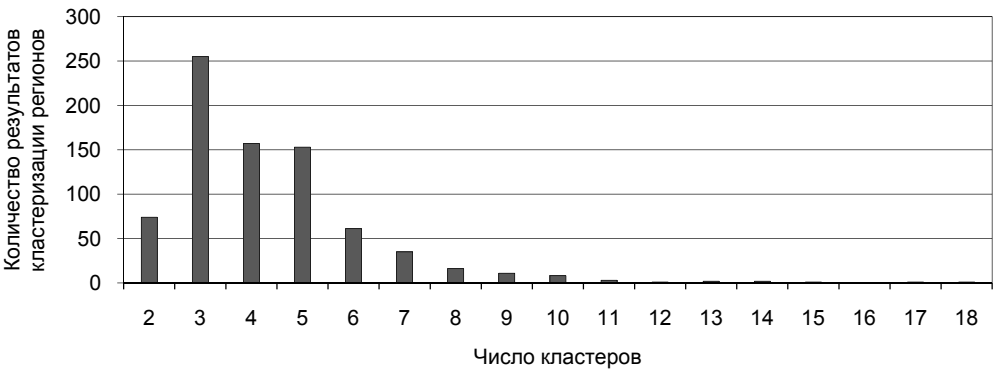

Рис. 2. Количество результатов социально-экономической кластеризации российских регионов с разным числом кластеров (по публикациям в научных журналах в 2004–2023 гг.)

Fig. 2. The number of results of socio-economic clustering of Russian regions with different numbers of clusters (according to publications in scientific journals in 2004–2023)

Источник: составлено автором.

Для иллюстрации смещения возьмем результаты кластеризации регионов Дальневосточного федерального округа по показателям развития малого предпринимательства в 2018 г. (Шпак и др., 2019). Используя манхэттенское расстояние и метод Уорда, авторы оцениваемого исследования объединили 9 регионов в 4 кластера: Камчатский край, Амурская и Сахалинская области; Республика Саха (Якутия); Приморский и Хабаровский края; Магаданская область, Еврейская автономная область и Чукотский автономный округ. Поскольку не приведены все исходные и промежуточные данные, то будем судить о проделанной работе по итоговой дендрограмме. Для этого воспользуемся теоретико-графовым подходом к определению оптимального числа кластеров (Бланца, 2021б). Количественная оценка сложности каждого яруса древовидного графа позволила установить, что 4-кластерное решение имеет меньшую значимость (0,3296), чем деление на три (0,3498) и два (0,3668) кластера. Поэтому первый и второй первоначальные кластеры должны объединиться в один, а третий и четвертый – в другой кластер.

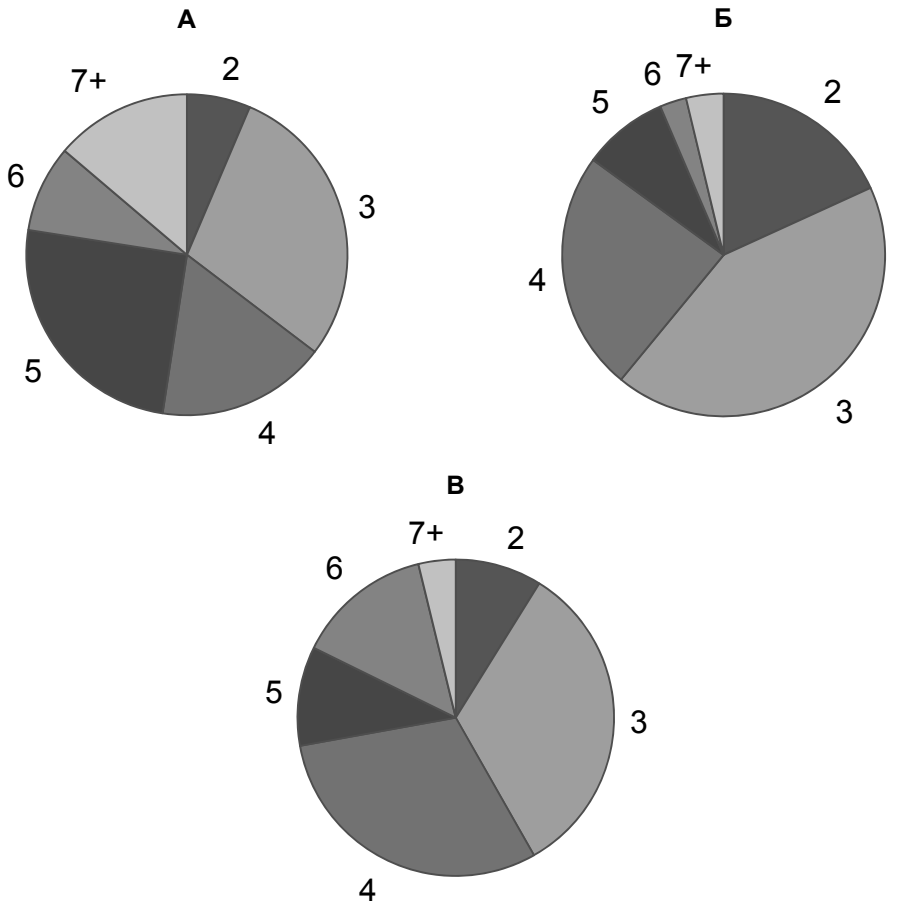

Рис. 3. Удельный вес результатов социально-экономической кластеризации российских регионов с разным числом кластеров (от 2 до 7+; в последнюю группу входит от 7 до 18) в массивах публикаций «А» (все регионы России), «Б» (регионы федерального округа) и «В» (выборка регионов)

Fig. 3. The share of the results of socio-economic clustering of Russian regions with a different number of clusters (from 2 to 7+; the last group includes from 7 to 18) in the arrays of publications «А» (all regions of Russia), «Б» (regions of the federal district) and «В» (a sample of regions)

Источник: составлено автором.

В данном случае имеет место смещение, которое можно трактовать как излишнее дробление (четыре вместо двух кластеров). При разном числе кластеров возможно также излишнее объединение, а при одинаковом количестве групп и неодинаковом наборе регионов в группах – смещение региона из одного в другой кластер. Не исключено излишнее дробление (объединение) со смещением регионов.

ЗАКЛЮЧЕНИЕ

Главным результатом проведенного исследования можно считать определение контуров ранее неизвестной научной проблемы – пространственной алгоритмической предвзятости. Обобщение 604 журнальных статей по социально-экономической кластеризации российских регионов позволило выявить четыре операции кластерного анализа, способные привести к дискриминации отдельных регионов. Выполнение двух операций – развертывание концептуальной модели в оптимальный набор показателей и выбор способа объединения регионов в кластеры – могло сопровождаться потенциальной несправедливостью. Остальные операции приводили к реальной дискриминации (при отборе регионов) и своего рода сегрегации (отделение региона от своего кластера и смещение в кластер с другим статусом при неправильном определении количества кластеров).

Дальнейшие исследования по рассматриваемой проблематике целесообразно направить в первую очередь на (а) выявление пространственной несправедливости в стратегических документах (Блануца, 2021а) и национальных целях развития (Блануца, 2022), (б) определение новых видов алгоритмической предвзятости и (в) разработку рекомендаций по справедливой кластеризации регионов. Параллельно направлению «в» проводятся исследования по созданию справедливого алгоритма k-средних (Pan, Zhong, 2023), но они никак не связаны с дискриминацией регионов. Возможно, пространственные неопределенности с кластерами заставят перейти к более широкому применению алгоритмов мягкой кластеризации (Majewska, Truskolaski, 2019; Miyamoto et al., 2008), когда каждый регион в той или иной степени будет входить во все кластеры.

СПИСОК ЛИТЕРАТУРЫ

Акбердина В.В., Наумов И.В., Красных С.С. Цифровое пространство регионов Российской Федерации: оценка факторов развития и взаимного влияния на социально-экономический рост // Journal of Applied Economic Research. 2023. Т. 22. Вып. 2. С. 294–322. https://doi.org/10.15826/vestnik.2023.22.2.013

Беркович М.И., Боженка С.В., Брут-Бруляко А.А. Оценка социально-экономического развития субъектов Российской Федерации: факторно-кластерный подход // Известия высших учебных заведений. Серия «Экономика, финансы и управление производством». 2013. № 2. С. 41–49.

Блануца В.И. Географическая экспертиза стратегий экономического развития России. М.: ИНФРА-М, 2021а. 198 с.

Блануца В.И. Дендрограммы в региональном социально-экономическом анализе: интерпретация и верификация // Научная визуализация. 2021б. Т. 13. № 5. С. 1–15. https://doi.org/10.26583/sv.13.5.01

Блануца В.И. Кластеризация регионов Сибири и Дальнего Востока по достижению национальных целей развития // Российский экономический журнал. 2022. № 3. С. 63–83. https://doi.org/10.33983/0130-9757-2022-3-63-83

Блануца В.И. Региональные экономические исследования с использованием алгоритмов искусственного интеллекта: состояние и перспективы // Вестник Забайкальского государственного университета. 2020. Т. 26. № 8. С. 100–111. https://doi.org/10.21209/2227-9245-2020-26-8-100-111

Быкова М.Л. Кластеризация как инструмент управления экономической безопасностью регионов Российской Федерации // Beneficium. 2023. № 4. С. 6–12. https://doi.org/10.34680/BENEFICIUM.2023.4(49).6-12

Жолудева И.И., Мельниченко Н.Ф., Козлов Г.Е. Применение кластерного анализа для оценки социально-экономического развития регионов на примере ЦФО и Ярославской области // Экономика, статистика и информатика. 2014. № 1. С. 144–148.

Ларченко Ю.Г. Кластеризация регионов на рынке труда по уровню привлекательности для иногородних соискателей // Ученые записки Комсомольского-на-Амуре государственного технического университета. 2023. № 8. С. 108–113.

Лимонова Г.Г. Статистический анализ неравномерности распределения национального богатства по территории России // Вестник Оренбургского государственного университета. 2014. № 3. С. 137–141.

Лясковская Е.А., Просвирина И.И., Кучина Е.В. Экономическое неравенство в России: анализ региональных особенностей // Вестник ЮУрГУ. Серия «Экономика и менеджмент». 2023. Т. 17. № 3. С. 77–87. https://doi.org/10.14529/em230307

Ляхова Н.И., Григорян Д.Р. Обоснование выделения Центрально-Черноземного района для разработки общей стратегии развития // Региональная экономика и управление: электронный научный журнал. 2017. № 2. 13 с.

Сорокина Н.Ю., Гагарина Г.Ю., Губарев Р.В. Диагностика социально-экономического развития регионов Российской Федерации с применением технологии нейросетевого моделирования // Плехановский научный бюллетень. 2017. № 2. С. 205–209.

Тимирьянова В.М., Зимин А.Ф., Жилина Е.В. Пространственная составляющая в изменении розничного рынка товаров // Экономика региона. 2018. Т. 14. Вып. 1. С. 164–175. https://doi.org/10.17059/2018-1-13

Трещевский Ю.И., Кособуцкая А.Ю., Гарин Л.К. Экономико-статистический анализ локализации эколого-экономической активности российских регионов // Социально-политические исследования. 2021. № 2. С. 87–99. https://doi.org/10.20323/2658-428X-2021-2-11-87-99

Шпак А.С., Шаталова А.С., Сальников К.Н. Оценка развития малого предпринимательства в Дальневосточном федеральном округе // Фундаментальные исследования. 2019. № 12-1. С. 211–217.

Aldenderfer M.S., Blashfield R K. Cluster Analysis. Beverly Hills: Sage Publications, 1984. 88 p.

Chabra A., Masalkovaité K., Mohapatra P. An Overview of Fairness in Clustering // IEEE Access. 2021. Vol. 9. Pp. 130698–130720. https://doi.org/10.1109/ACCESS.2021.3114099

Charrad M., Ghazzali N., Boiteau V., Niknafs A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set // Journal of Statistical Software. 2014. Vol. 61. Issue 6. Pp. 1–36. https://doi.org/10.18637/jss.v061.i06

Ester M., Kriegel H.-P., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise / Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). Portland: AAAI Press, 1996. Pp. 226–231.

Everitt B., Landau S., Leese M. Cluster Analysis. London: Arnold, 2001. 237 p.
Foss A.H., Markatou M., Ray B. Distance Metrics and Clustering Methods for Mixed-Type Data // International Statistical Review. 2019. Vol. 87. Issue 1. Pp. 80–109. https://doi.org/10.1111/insr.12274
Gerdon F., Bach R.L., Kern C., Kreuter F. Social Impact of Algorithmic Decision-Making: A Research Agenda for the Social Sciences // Big Data & Society. 2022. Vol. 9. Issue 1. Pp. 1–13. https://doi.org/10.1177/20539517221089305
Giordani P., Ferraro M. B., Martella F. An Introduction to Clustering with R. Singapore: Springer, 2020. 340 p.
Gupta S., Ghalme G., Krishnan N.C., Jain S. Efficient Algorithms for Fair Clustering with a New Notion of Fairness // Data Mining and Knowledge Discovery. 2023. Vol. 37. Pp. 1959–1997. https://doi.org/10.1007/s10618-023-00928-6
Hadi A.S. A New Distance Between Multivariate Clusters of Varying Locations, Elliptical Shapes, and Directions // Pattern Recognition. 2022. Vol. 129. 108780. https://doi.org/10.1016/j.patcog.2022.108780
Hartigan J.A. Clustering Algorithms. New York: Wiley, 1975. 351 p.
Jackson M.C. Artificial Intelligence & Algorithmic Bias: The Issues with Technology Reflecting History & Humans // Journal of Business and Technology Law. 2021. Vol. 16. Issue 2. Pp. 299–316.
Johnson G.M. Algorithmic Bias: On the Implicit Biases of Social Technology // Synthese. 2021. Vol. 198. Pp. 9941–9961. https://doi.org/10.1007/s11229-020-02696-y
Kordzadeh N., Ghasemaghaei M. Algorithmic Bias: Review, Synthesis, and Future Research Directions // European Journal of Information Systems. 2022. Vol. 31. Issue 3. Pp. 388–409. https://doi.org/10.1080/0960085X.2021.1927212
Liu Z., Janowicz K., Cai L., Zhu R., Mai G., Shi M. Geoparsing: Solved or Biased? An Evaluation of Geographic Biases in Geoparsing / Proceedings of the 25th AGILE Conference on Geographic Information Science. Vilnus: AGILE, 2022. Pp. 1–13. https://doi.org/10.5194/agile-giss-3-9-2022
Liu Z., Zhang X., Jiang B. Active Learning with Fairness-Aware Clustering for Fair Classification Considering Multiple Sensitive Attributes // Information Sciences. 2023. Vol. 647. 119521. https://doi.org/10.1016/j.ins.2023.119521
López-Villuendas A.M., del Campo C. Regional Economic Disparities in Europe: Time-Series Clustering Of NUTS 3 Regions // International Regional Science Review. 2023. Vol. 46. Issue 3. Pp. 265–298. https://doi.org/10.1177/01600176221125703
Lorimer T., Held J., Stoop R. Clustering: How Much Bias Do We Need? // Philosophical Transactions of the Royal Society A. 2017. Vol. 375. 20160293. https://doi.org/10.1098/rsta.2016.0293
MacQueen J.B. Some Methods for Classification and Analysis of Multivariate Observations / Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1976. Pp. 281–297.
Majewska J., Truskolaski S. Cluster-Mapping Procedure for Tourism Regions Based on Geostatistics and Fuzzy Clustering: Example of Polish Districts // Current Issues in Tourism. 2019. Vol. 22. Issue 19. Pp. 2365–2385. https://doi.org/10.1080/13683500.2018.1467883
Mirkin B. Mathematical Classification and Clustering. Dordrecht: Kluwer Academic Publisher, 1996. 580 p.
Miyamoto S., Ichihashi H., Honda K. Algorithms for Fuzzy Clustering: Methods in c-Means Clustering with Applications. Berlin; Heidelberg: Springer, 2008. 248 p. https://doi.org/10.1007/978-3-540-78737-2
Oduntan O.I., Thulasiraman P. Blending Multiple Algorithmic Granular Components:

A Recipe for Clustering // Swarm Intelligence. 2022. Vol. 16. Pp. 305–349. https://doi.org/10.1007/s11721-022-00219-8

Pan R., Zhong C. Fairness First Clustering: A Multi-Stage Approach for Mitigating Bias // Electronics. 2023. Vol. 12. Issue 13. 2969. https://doi.org/10.3390/electronics12132969

Robinson C., Franklin R.S. The Sensor Desert Quandary: What Does It Mean (Not) to Count in the Smart City? // Transactions of the Institute of British Geographers. 2020. Vol. 46. Issue 2. Pp. 238–254. https://doi.org/10.1111/tran.12415

Safransky S. Geographies of Algorithmic Violence: Redlining the Smart City // International Journal of Urban and Regional Research. 2020. Vol. 44. Issue 2. Pp. 200–218. https://doi.org/10.1111/1468-2427.12833

Sánchez P., Bellogín A., Boratto L. Bias Characterization, Assessment, and Mitigation in Location-Based Recommender Systems // Data Mining and Knowledge Discovery. 2023. Vol. 37. Pp. 1885–1929. https://doi.org/10.1007/s10618-022-00913-5

Soares J.O., Coutinho C.C. Cluster Analysis in Regional Science // Advances and Applications in Statistical Science. 2010. Vol. 1. Issue 2. Pp. 311–325.

Tryon R.C. Cluster Analysis: Correlation Profile and Orthometric (Factor) Analysis for the Isolation of Unities in Mind and Personality. Ann Arbor: Edwards Brothers, 1939. 122 p.

Van Giffen B., Herhausen D., Fahse T. Overcoming the Pitfalls and Perils of Algorithms: A Classification of Machine Learning Biases and Mitigation Methods // Journal of Business Research. 2022. Vol. 144. Pp. 93–106. https://doi.org/10.1016/j.jbusres.2022.01.076

Wagner B., Winkler T., Human S. Bias in Geographic Information Systems: The Case of Google Maps // Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS54). Honolulu, 2021. Pp. 837–847. https://doi.org/10.24251/HICSS.2021.103

Ward J.H. Hierarchical Grouping to Optimize an Objective Function // Journal of the American Statistical Association. 1963. Vol. 58. Issue 301. Pp. 236–244.

Xu D., Tian Y.A. Comprehensive Survey of Clustering Algorithms // Annals of Data Science. 2015. Vol. 2. Pp. 165–193. https://doi.org/10.1007/s40745-015-0040-1

Xu R., Wunsch D. Survey of Clustering Algorithms // IEEE Transactions on Neural Networks. 2005. Vol. 16. Issue 3. Pp. 645–678. https://doi.org/10.1109/TNN.2005.845141

Zou J., Schiebinger L. AI Can be Sexist and Racist – It's Time to Make It Fair // Nature. 2018. Vol. 559. Pp. 324–326. https://doi.org/10.1038/d41586-018-05707-8

REFERENCES

Akberdina V.V., Naumov I.V., Krasnykh S.S. Digital Space of Regions: Assessment of Development Factors and Influence on Socio-Economic Growth. *Journal of Applied Economic Research*, 2023, vol. 22, issue 2, pp. 294–322. https://doi.org/10.15826/vestnik.2023.22.2.013 (In Russian).

Aldenderfer M.S., Blashfield R K. *Cluster Analysis*. Beverly Hills: Sage Publications, 1976, 88 p.

Berkovich M.I., Bozhenka S.V., Brut-Brulyako A.A. Assessment of the Socio-Economic Development of the Subjects of the Russian Federation: A Factor-Cluster Approach. *Izvestiya Vysshih Uchebnyh Zavedenij. Seriya: Ekonomika, finansy i Upravlenie Proizvodstvom = Izvestia of Higher Educational Institutions. Series: Economics, Finance and Production Management*, 2013, no. 2, pp. 41–49 (In Russian).

Blanutsa V.I. Clustering the Regions of Siberia and the Far East to Achieve National

Development Goals. *Rossijskij Ekonomicheskij Zhurnal = Russian Economic Journal*, 2022, no. 3, pp. 63–83. https://doi.org/10.33983/0130-9757-2022-3-63-83 (In Russian).

Blanutsa V.I. Dendrograms in Regional Socio-Economic Analysis: Interpretation and Verification. *Scientific Visualization*, 2021b, vol. 13, issue 5, pp. 1–15. https://doi.org/10.26583/sv.13.5.01 (In Russian).

Blanutsa V.I. *Geographical Expertise of Russia's Economic Development Strategies*. Moscow, 2021a, 198 p. (In Russian).

Blanutsa V.I. Regional Economic Research Using Artificial Intelligence Algorithms: State and Prospects. *Vestnik Zabajkal'skogo Gosudarstvennogo Universiteta = Bulletin of the Transbaikal State University*, 2020, vol. 26, issue 8, pp. 100–111 https://doi.org/10.21209/2227-9245-2020-26-8-100-111 (In Russian).

Bykova M.L. Clustering as a Tool for Managing the Economic Security of the Russian Federation's Regions. *Beneficium*, 2023, no. 4, pp. 6–12. (In Russian). https://doi.org/10.34680/BENEFICIUM.2023.4(49).6-12 (In Russian).

Chabra A., Masalkovaité K., Mohapatra P. An Overview of Fairness in Clustering. *IEEE Access*, 2021, vol. 9, pp. 130698–130720. https://doi.org/10.1109/ACCESS.2021.3114099.

Charrad M., Ghazzali N., Boiteau V., Niknafs A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 2014, vol. 61, issue 6, pp. 1–36. https://doi.org/10.18637/jss.v061.i06

Ester M., Kriegel H.-P., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. Portland: AAAI Press, 1996, pp. 226–231.

Everitt B., Landau S., Leese M. *Cluster Analysis*. London: Arnold, 2001, 237 p.

Foss A.H., Markatou M., Ray B. Distance Metrics and Clustering Methods for Mixed-Type Data. *International Statistical Review*, 2019, vol. 87, issue 1, pp. 80–109. https://doi.org/10.1111/insr.12274

Gerdon F., Bach R. L., Kern C., Kreuter F. Social Impact of Algorithmic Decision-Making: A Research Agenda for the Social Sciences. *Big Data & Society*, 2022, vol. 9, issue 1, pp. 1–13. https://doi.org/10.1177/20539517221089305

Giordani P., Ferraro M. B., Martella F. *An Introduction to Clustering with R*. Singapore: Springer, 2020, 340 p.

Gupta S., Ghalme G., Krishnan N.C., Jain S. Efficient Algorithms for Fair Clustering with a New Notion of Fairness. *Data Mining and Knowledge Discovery*, 2023, vol. 37, pp. 1959–1997. https://doi.org/10.1007/s10618-023-00928-6

Hadi A.S. A New Distance between Multivariate Clusters of Varying Locations, Elliptical Shapes, and Directions. *Pattern Recognition*, 2022, vol. 129, 108780. https://doi.org/10.1016/j.patcog.2022.108780

Hartigan J.A. *Clustering Algorithms*. New York: Wiley, 1975, 351 p.

Jackson M.C. Artificial Intelligence and Algorithmic Bias: The Issues with Technology Reflecting History and Humans. *Journal of Business and Technology Law*, 2021, vol. 16, issue 2, pp. 299–316.

Johnson G.M. Algorithmic Bias: On the Implicit Biases of Social Technology. *Synthese*, 2021, vol. 198, issue 1, pp. 9941–9961. https://doi.org/10.1007/s11229-20-02696-y

Kordzadeh N., Ghasemaghaei M. Algorithmic Bias: Review, Synthesis, and Future Research Directions. *European Journal of Information Systems*, 2022, vol. 31, issue 3, pp. 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Larchenko Yu.G. Clustering of Regions in the Labor Market by the Level of Attractiveness for Nonresident Applicants. *Uchenye Zapiski Komsomol'skogo-na-Amure Gosudarst-*

vennogo Tekhnicheskogo Universiteta = Scientific Notes of Komsomolsk-on-Amur State Technical University, 2023, no. 8. pp. 108–113. (In Russian).

Limonova G.G. Statistical Analysis of the Uneven Distribution of National Wealth Across the Territory of Russia. *Vestnik Orenburgskogo Gosudarstvennogo Universiteta = Bulletin of the Orenburg State University*, 2014, no. 3, pp. 137–141. (In Russian).

Liu Z., Janowicz K., Cai L., Zhu R., Mai G., Shi M. Geoparsing: Solved or Biased? An Evaluation of Geographic Biases in Geoparsing. In: *Proceedings of the 25th AGILE Conference on Geographic Information Science*. Vilnus, AGILE, 2022, pp. 1–13. https://doi.org/10.5194/agile-giss-3-9-2022

Liu Z., Zhang X., Jiang B. Active Learning with Fairness-Aware Clustering for Fair Classification Considering Multiple Sensitive Attributes. *Information Sciences*, 2023, vol. 647, 119521. https://doi.org/10.1016/j.ins.2023.119521

López-Villuendas A.M., del Campo C. Regional Economic Disparities in Europe: Time-series Clustering of NUTS 3 Regions. *International Regional Science Review*, 2023, vol. 46, issue 3, pp. 265–298. https://doi.org/10.1177/01600176221125703

Lorimer T., Held J., Stoop R. Clustering: How Much Bias do we Need? *Philosophical Transactions of the Royal Society A*, 2017, vol. 375, 20160293. https://doi.org/10.1098/rsta.2016.0293

Lyakhova N.I., Grigoryan D.R. Justification of the Allocation of the Central Chernozem Region for the Development of a Common Development Strategy. *Regional'naya ekonomika i upravlenie: elektronnyj nauchnyj zhurnal = Regional Economics and Management: Electronic Scientific Journal*, 2017, no. 2, p. 26. (In Russian).

Lyaskovskaya E.A., Prosvirina I.I., Kuchina E.V. Economic Inequality in Russia: An Analysis of Regional Features. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya 'Ekonomika i menedzhment'* = Bulletin of the South Ural State University. The series 'Economics and Management', 2023, vol. 17, issue 3, pp. 77–87. https://doi.org/10.14529/em230307 (In Russian).

MacQueen J.B. Some Methods for Classification and Analysis of Multivariate Observations. In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1976, pp. 281–297.

Majewska J., Truskolaski S. Cluster-Mapping Procedure for Tourism Regions Based on Geostatistics and Fuzzy Clustering: Example of Polish districts. *Current Issues in Tourism*. 2019, vol. 22, issue 19, pp. 2365–2385. https://doi.org/10.1080/13683500.2018.1467883

Mirkin B. *Mathematical Classification and Clustering*. Dordrecht; Boston; London: Kluwer Academic Publisher, 1996. 429 p.

Miyamoto S., Ichihashi H., Honda K. *Algorithms for Fuzzy Clustering: Methods in c-Means Clustering with Applications*. Berlin; Heidelberg: Springer, 2008, 258 p.

Oduntan O.I., Thulasiraman P. Blending Multiple Algorithmic Granular Components: A Recipe for Clustering. *Swarm Intelligence*, 2022, vol. 16, pp. 305–349. https://doi.org/10.1007/s11721-022-00219-8

Pan R., Zhong C. Fairness First Clustering: A Multi-Stage Approach for Mitigating Bias. *Electronics*, 2023, vol. 12, issue 13, e2969. https://doi.org/10.3390/electronics12132969

Robinson C., Franklin R.S. The Sensor Desert Quandary: What Does it Mean (not) to Count in the Smart City? *Transactions of the Institute of British Geographers*, 2020, vol. 46, issue 2, pp. 238–254. https://doi.org/10.1111/tran.12415

Safransky S. Geographies of Algorithmic Violence: Redlining the Smart City. *International Journal of Urban and Regional Research*, 2020, vol. 44, issue 2, pp. 200–218. https://doi.org/10.1111/1468-2427.12833

Sánchez P., Bellogín A., Boratto L. Bias Characterization, Assessment, and Mitigation

in Location-Based Recommender Systems. *Data Mining and Knowledge Discovery*, 2023, vol. 37, pp. 1885–1929. https://doi.org/10.1007/s10618-022-00913-5

Shpak A.S., Shatalova A.S., Salnikov K.N. Assessment of the Development of Small Business in the Far Eastern Federal District. *Fundamental'nye issledovaniya = Fundamental Research*, 2019, no. 12-1, pp. 211–217. (In Russian).

Soares J.O., Coutinho C.C. Cluster Analysis in Regional Science. *Advances and Applications in Statistical Science*, 2010, vol. 1, issue 2, pp. 311–325.

Sorokina N.Yu., Gagarina G.Yu., Gubarev R.V. Diagnostics of Socio-Economic Development of the Regions of the Russian Federation Using Neural Network Modeling Technology. *Plekhanovskij nauchnyj byulleten' = Plekhanov Scientific Bulletin*, 2017, no. 2, pp. 205–209. (In Russian).

Timiryanova V.M., Zimin A.F., Zhilina E.V. The Spatial Change of the Indicators of Consumer Market. *Ekonomika Regiona = Economy of Region*, 2018, vol. 14, issue 1, pp. 164 – 175. https://doi.org/10.17059/2018-1-13 (In Russian).

Treshchevsky Yu.I., Kosobutskaya A.Yu., Garin L.K. Economic and Statistical Analysis of Localization of Ecological and Economic Activity of Russian Regions. *Socialno-Politicheskie Issledovaniya = Socio-Political Studies*, 2021, no. 2, pp. 87 – 99. https://doi.org/10.20323/2658-428X-2021-2-11-87-99 (In Russian).

Tryon R.C. *Cluster Analysis: Correlation Profile and Orthometric (Factor) Analysis for the Isolation of Unities in Mind and Personality*. Ann Arbor: Edwards Brothers, 1939, 122 p.

Van Giffen B., Herhausen D., Fahse T. Overcoming the Pitfalls and Perils of Algorithms: A Classification of Machine Learning Biases and Mitigation Methods. *Journal of Business Research*, 2022, vol. 144, pp. 93–106. https://doi.org/10.1016/j.jbusres.2022.01.076

Wagner B., Winkler T., Human S. Bias in Geographic Information Systems: The Case of Google Maps. In: *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS54)*. Honolulu, 2021, pp. 837–847. https://doi.org/10.24251/HICSS.2021.103

Ward J.H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 1963, vol. 58, issue 301, pp. 236–244.

Xu D., Tian Y. A. Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 2015, vol. 2, pp. 165–193. https://doi.org/10.1007/s40745-015-0040-1

Xu R., Wunsch D. Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 2005, vol. 16, issue 3, pp. 645–678. https://doi.org/10.1109/TNN.2005.845141

Zholudeva I.I., Melnichenko N.F., Kozlov G.E. Application of Cluster Analysis to Assess the Socio-Economic Development of Regions on the Example of the Central Federal District and the Yaroslavl Region. *Ekonomika, Statistika i Informatika = Economics, Statistics and Informatics*, 2014, no. 1, pp. 144–148. (In Russian).

Zou J., Schiebinger L. AI Can Be Sexist and Racist – It's Time to Make It Fair. *Nature*, 2018, vol. 559, pp. 324–326. https://doi.org/10.1038/d41586-018-05707-8

Поступила в редакцию / Submitted: 18.06.2024
Одобрена после рецензирования / Approved after reviewing: 20.06.2024
Принята к публикации / Accepted for publication: 20.06.2024
Доступно онлайн / Available online: 01.07.2024