УДК 004.62
https://www.doi.org/10.47813/mip.7.2026.20.4016
EDN KWHOKA

Особенности применения методов интеллектуального анализа данных для оценки уровня жизни городского населения

В.О. Михайлов*, Е.А. Макарова

Уфимский университет науки и технологий, ул. Заки Валиди 32, Уфа, 450076, Россия

*E-mail: mikhailovvo23@gmail.com

Аннотация. В статье рассматриваются подходы к оценке уровня жизни населения в условиях многомерности региональных и муниципальных социально-экономических данных, а также усиления роли городов в расселенческой структуре страны. Проведён обзор применения методов главных компонент, кластерного анализа и построения случайного леса в задачах исследования уровня жизни населения и устойчивого развития территорий, выявлены особенности и ограничения существующих подходов к профилированию городских кластеров, в том числе недостаточная проработка вопросов интерпретации получаемых типологий и учёта специфики городской среды. На основе проведённого анализа литературы обоснована необходимость разработки информационно-аналитической системы, обеспечивающей комплексную поддержку процедур многомерного анализа данных об уровне жизни городского населения. Сформулированы функциональные возможности такой системы, включающие загрузку исходных данных, формирование выборки признаков по тематическим блокам, выполнение компонентного и кластерного анализа, а также последующую интерпретацию выделенных кластеров методом случайного леса с построением деревьев решений. Предлагаемое решение обеспечивает выявление типологических групп городов и определение факторов, обусловливающих различия между ними, что повышает обоснованность исследовательских выводов и поддерживает последующие процессы принятия решений органами государственного и муниципального управления в области социальной политики.

Ключевые слова: уровень жизни, многомерный анализ данных, поддержка принятия решений, метод главных компонент, кластерный анализ, случайный лес.

Для цитирования: Михайлов, В. О. Особенности применения методов интеллектуального анализа данных для оценки уровня жизни городского населения / В. О. Михайлов, Е. А. Макарова // Модернизация, Инновации, Прогресс (MIP ENGINEERING-VII-2026): материалы VII Всероссийской (национальной) научной конференции (Красноярск, 11–12 сентября 2026 г.). – Красноярск: Красноярский краевой Дом науки и техники Российского Союза научных и инженерных общественных объединений, 2026. – Т. 20. – С. 316-324. – https://doi.org/10.47813/mip.7.2026.20.4016. – EDN: KWHOKA.

Features of applying data mining methods to assess the standard of living of the urban population

V.O. Mihaylov*, E.A. Makarova

Ufa University of Science and Technology, 32 Zaki Validi Street, Ufa, 450076, Russia

*E-mail: mikhailovvo23@gmail.com

Abstract. The article examines approaches to assessing the standard of living of the population under conditions of multidimensional regional and municipal socioeconomic data and the growing role of cities within the settlement structure of the country. It reviews the application of principal component analysis, cluster analysis, and random forest methods in studies of living standards and sustainable territorial development, identifying features and limitations of existing approaches to profiling urban clusters, including insufficient elaboration of typology interpretation and urban environment specifics. Based on this literature analysis, the necessity of developing an information-analytical system providing comprehensive support for multidimensional data analysis on the standard of living of the urban population is substantiated. The functional capabilities of such a system are formulated, including loading source data, forming a feature sample by thematic blocks, performing component and cluster analysis, and interpreting the identified clusters using the

random forest method along with decision tree construction. The proposed solution ensures identification of typological groups of cities and determination of factors underlying differences between them, thereby increasing the validity of research conclusions and supporting subsequent decision-making by state and municipal government bodies in the field of social policy.

Keywords: standard of living, multidimensional data analysis, decision support, principal component analysis, cluster analysis, random forest.

For citation: Mihaylov, V. O. Features of applying data mining methods to assess the standard of living of the urban population / V. O. Mihaylov, E. A. Makarova // Modernization, Innovations, Progress (MIP ENGINEERING-VII-2026): Proceedings of the VII All-Russian (National) Scientific Conference (Krasnoyarsk, September 11–12, 2026). – Krasnoyarsk: Krasnoyarsk Regional Science and Technology City Hall of the Russian Union of Scientific and Engineering Public Associations, 2026. – Vol. 20. – P. 316-324. – https://doi.org/10.47813/mip.7.2026.20.4016. – EDN: KWHOKA.

1. Введение

Оценка уровня жизни населения остаётся одной из приоритетных задач социально-экономической политики в условиях сохраняющейся межрегиональной дифференциации доходов, а также трансформации городской среды и усиления роли городов в расселенческой структуре. Для принятия обоснованных управленческих решений в сфере социально-экономической политики на федеральном, региональном и муниципальном уровнях необходимо проведение исследований различных сценариев реализации управленческих политик с помощью модельных аналитических инструментариев. Для разработки таких инструментариев применяются различные стратегии анализа многомерных данных с использованием различных методов. Результаты анализа многомерных данных в виде кластеров городов и регионов позволяют формировать доказательную базу для поддержки принятия различных решений, таких, например, как корректировка минимального размера оплаты труда (МРОТ), установление прожиточного минимума [1, 2].

2. Постановка задачи (Цель исследования)

Целью исследования является разработка функциональных основ информационно-аналитической системы поддержки процедур анализа многомерных данных для оценки уровня жизни городского населения, обеспечивающей выявление типологических групп городов и определение факторов, влияющих на различия между ними.

2.1. Литературный обзор

Решению проблем оценки уровня жизни населения посвящены исследования многих исследователей, при этом применяются методы компонентного, кластерного, нейросетевого анализа с последующим решением задачи индукции правил классификации. Достаточно

широкое применение получила технология конвейерного анализа данных, основанная на последовательном применении методов компонентного и кластерного анализа данных.

В аналитическом центре Правительства Москвы проведено исследование, в котором с применением метода главных компонент к 141 статистическому показателю по регионам России за 2015-2017 годы выделено 25 обобщённых факторов, характеризующих неформальную занятость и социально-экономическое развитие [3]. Предложено интегрированное применение метода главных компонент и кластерного анализа (метод Уорда) для классификации регионов по структуре неформальной занятости и уровню социально-экономического развития. Особенностью подхода является выделение пяти групп регионов, ранжированных относительно среднероссийского уровня (от низкого до высокого), что позволяет использовать полученную классификацию в качестве информационной основы для принятия управленческих решений по регулированию рынка труда.

Также в аналитическом центре при Правительстве Москвы было проведено исследование субъектов Сибирского федерального округа, в котором с использованием метода главных компонент была показана возможность сокращения размерности 20 индикаторов экономической безопасности до трёх главных компонент в каждом тематическом блоке. Полученные результаты продемонстрировали, что структура выделяемых компонент изменяется во времени, что позволяет выявлять структурные сдвиги в экономике и использовать их при корректировке управленческих решений [4].

В исследовании китайских ученых с применением метода главных компонент к данным по 98 странам показано, что большинство из 12 альтернативных показателей социального прогресса сильно коррелируют с ВВП на душу населения [5].

В институте вычислительного моделирования Сибирского отделения СО РАН проведены исследования Красноярского края, в котором применен метод главных компонент для оценки показателей качества жизни по 12 национальным проектам [6]. Предложено взаимосвязанное применение метода главных компонент и кластерного анализа, для выявления структуры данных и получения интегрального рейтинга регионов. Особенностью алгоритма является возможность итеративного уточнения состава кластеров с учётом взаимосвязей между результатами различных методов интеллектуального анализа данных.

Другим вариантом реализации конвейерного анализа данных для оценки уровня жизни является технология использования иерархического метода построения кластеров с последующим применением регрессионного анализа данных. Так, в Мордовском государственном университете им. Н.П. Огарёва проведено исследование, в котором с

применением иерархического кластерного анализа (метод полных связей, евклидово расстояние) к 19 показателям уровня жизни по 79 регионам России выделено 5 кластеров, различающихся по составу федеральных округов и уровню социально-экономического развития [7]. Предложено интегрированное применение кластерного и регрессионного анализа для выявления факторов, влияющих на среднедушевые денежные доходы населения в каждом кластере. Особенностью подхода является построение регрессионных моделей для каждого выделенного кластера (R^2 от 0,908 до 0,967), позволяющих определить специфические драйверы доходов: в одних кластерах это размер пенсий и потребительские расходы, в других – жилищные условия или структура потребления.

Известны исследования, основанные на построении деревьев решений и случайного леса. В исследовании Чжэцзянского технологического университета метод случайного леса применён для анализа влияния факторов урбанизации (расширение доходов правительства, искажение распределения земельных ресурсов, корректировка промышленной структуры) на устойчивое развитие 292 городов Китая за 2000–2019 годы. [8]. Предложено использование модели случайного леса для моделирования нелинейных зависимостей между характеристиками городской трансформации и интегральным индексом устойчивого развития, включающим семь показателей экономической, социальной и экологической сфер. Особенностью подхода является выявление относительной важности факторов: установлено, что наибольшее влияние на устойчивое развитие оказывает корректировка промышленной структуры, тогда как искажение распределения земельных ресурсов и расширение доходов правительства более значимы для крупных городских агломераций. Качество аппроксимации модели случайного леса превзошло множественную линейную регрессию и градиентный бустинг.

На кафедре технической кибернетики УУНИТ проводятся исследования, основанные как на применении метода главных компонент для различных конфигураций данных о доходах населения в региональном разрезе и для последующей оценки уровня жизни населения субъектов Российской Федерации с учетом иерархических построений, так на применении разработанного метода конвейерного кластеризации на главных компонентах [9, 10]. Особенностью подхода является интеграция методов главных компонент и кластерного анализа на различных структурах данных для построения целостного представления об уровне жизни населения страны с учётом целей устойчивого развития, что позволяет повысить обоснованность принятия стратегических решений при управлении макроэкономическими

системами. Однако не рассматривались аспекты уровня жизни городского населения и не решены были проблемы профилирования кластеров с помощью деревьев решений.

3. Методы и материалы исследования

Исследование основано на применении комплекса методов интеллектуального анализа многомерных социально-экономических данных, предназначенного для оценки уровня жизни населения городов, выявления сходных по условиям жизнедеятельности территорий и определения факторов, обусловливающих различия между выделенными группами городов.

Материалами исследования выступают открытые статистические данные, характеризующие социально-экономическое положение городов Российской Федерации.

Объектами исследования являются города, рассматриваемые как территориальные социально-экономические системы. Единицей наблюдения выступает отдельный город в заданном временном периоде. Признаковое пространство формируется из совокупности количественных показателей, отражающих различные стороны уровня жизни населения.

4. Полученные результаты

Разрабатываемая информационно-аналитическая система оценки уровня жизни городского населения должна обеспечивать поддержку процедур анализа многомерных данных о различных аспектах жизни населения городов и предоставлять возможность получения интерпретируемых результатов для последующей исследовательской работы по подготовке принятия управленческих решений. Функциональные особенности системы состоят в следующем (рисунок 1).

Во-первых, необходимо обеспечение загрузки исходного датасета, выбор временного периода и формирование рабочей выборки признаков в соответствии с задачей исследования. Анализ должен выполняться как по полной совокупности показателей, так и по тематическим блокам, сформированным в рамках триадного подхода, что позволяет рассматривать уровень жизни населения города через взаимосвязанные аспекты как уровня экономического благосостояния населения, так и доступности условий проживания, а также качества городской среды [9, 10].

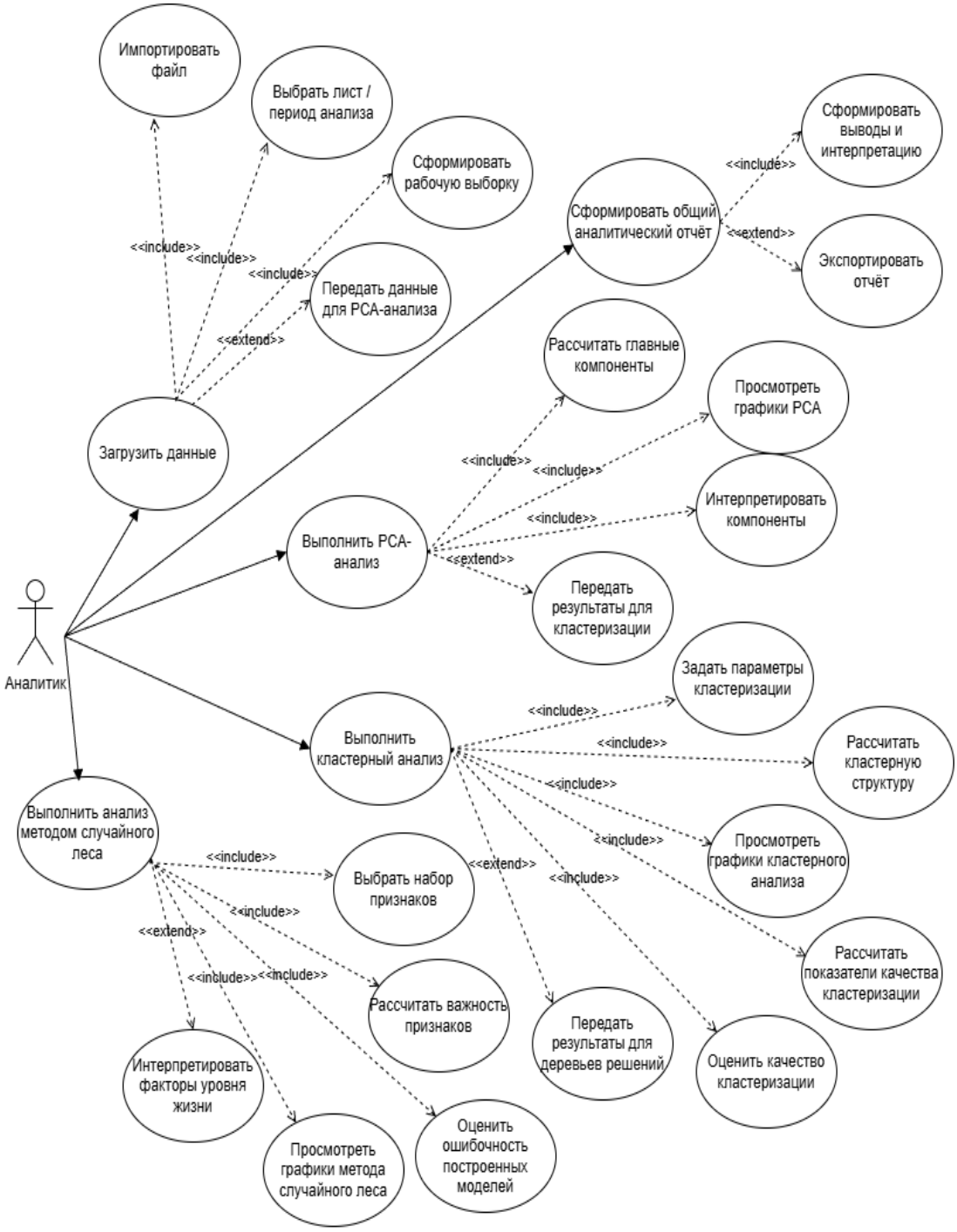

Рисунок 1. Диаграмма прецедентов информационно-аналитической системы оценки уровня жизни городов.

Во-вторых, в рамках аналитического контура система должна поддерживать выполнение компонентного анализа, проведение кластеризации в условиях снижения размерности и, при необходимости, выполнять интерпретация кластеров методом построения случайного леса. Особенностью применения метода главных компонент в системе является

многоуровневый анализ данных и построение интегральных признаков для выделенных блоков данных. Для каждого блока выполняется компонентный анализ и формируются интегральные признаки, отражающие ключевые характеристики рассматриваемого аспекта. Обобщение на основе метода главных компонент позволяет выявить структуру различий между городами и обеспечить их последующую кластеризацию и интерпретацию. При выполнении кластерного анализа система должна обеспечивать формирование типологии городов и визуализацию кластерной структуры, а при построении случайного леса — оценку значимости признаков и выявление факторов, влияющих на различия между городами.

В-третьих, необходима реализация визуально-аналитического представления результатов, которое включает просмотр графиков, сопоставление городов и кластеров, а также интерактивную работу с компонентным пространством и выделенными группами объектов. Итоговым действием является подготовка аналитического отчёта, содержащего описание исходных данных, результаты применённых методов и их обобщающую интерпретацию.

В результате проведённого исследования показано, что анализ уровня жизни населения городов требует применения не одного универсального метода анализа данных, а комплекса методов, позволяющих учитывать многомерный характер городской среды и социально-экономических условий проживания населения. Разрабатываемая система поддержки процедур анализа данных об уровне жизни городского населения, выполняющая функции кластеризации на главных компонентах с последующим профилированием городских кластеров и индукцией правил на деревьях решений позволит обеспечить поддержку в принятии решений органами государственного и муниципального управления в области социальной политики.

5. Выводы

В результате проведённого исследования показано, что анализ уровня жизни населения городов требует применения не одного универсального метода анализа данных, а комплекса методов, позволяющих учитывать многомерный характер городской среды и социально-экономических условий проживания населения. Разрабатываемая система поддержки процедур анализа данных об уровне жизни городского населения, выполняющая функции кластеризации на главных компонентах с последующим профилированием городских кластеров и индукцией правил на деревьях решений позволит обеспечить поддержку в

принятии решений органами государственного и муниципального управления в области социальной политики.

Список литературы

1. Гуляева, Т. И. Оценка уровня жизни населения регионов России на основе применения кластерного анализа / Т. И. Гуляева, Е. В. Такмакова // Экономический анализ: теория и практика. – 2021. – Т. 20, № 5(512). – С. 810-828. – EDN: ZIMSNN. – https://doi.org/10.24891/ea.20.5.810.
2. Оценка качества жизни населения регионов РФ на основе цифровых данных: методологические аспекты / Е. В. Щекотин, Г. Ю. Коварж, В. Л. Гойко [и др.] // Векторы благополучия: экономика и социум. – 2020. – № 3(38). – С. 138-156. – EDN: SPGDQH. – https://doi.org/10.18799/26584956/2020/3(38)/1020.
3. Иванова, И.А. Кластерный и регрессионный анализ уровня жизни населения регионов Российской Федерации / И.А. Иванова, С.А. Орешкина, М.В. Базайкина // Экономика и социум. – 2014. – № 3-2(12) – С. 64-69. – EDN: TENNJL.
4. Дубравская, Э. И. Классификации регионов России с учетом структуры неформальной занятости и уровня социально-экономического развития / Э. И. Дубравская // Статистика и Экономика. – 2020. – Т. 17, № 6. – С. 31-43. – EDN: NDIZYG. – https://doi.org/10.21686/2500-3925-2020-6-31-43.
5. Wang, B. Social Progress beyond GDP: A Principal Component Analysis (PCA) of GDP and Twelve Alternative Indicators / B. Wang, T. Chen // Sustainability. – 2022. – Vol. 14, No. 11. – P. 6430. – EDN: HRVQYC. – https://doi.org/10.3390/su14116430.
6. Логинов, К. К. Применение метода главных компонент в задаче оценки региональной экономической безопасности (на примере субъектов Сибирского федерального округа) / К. К. Логинов, А. Г. Бреусова, А. А. Кораблева // Научные труды Вольного экономического общества России. – 2025. – Т. 255, № 5. – С. 329-352. – EDN: FUJRMF. – https://doi.org/10.38197/2072-2060-2025-255-5-329-352.
7. Peng, Y. Using a Random Forest Model to Study the Impact of Local Government-Led Urbanization on Urban Sustainable Development / Y. Peng // Journal of Environmental and Public Health. – 2022. – Vol. 2022. – P. 3393532. – EDN: NNPFSB. – https://doi.org/10.1155/2022/3393532.
8. Machine-learning model predicting quality of life using multifaceted lifestyles in middle-aged South Korean adults: a cross-sectional study / Ju. Kim, K. Jeong, S. Lee, Y. Baek // BMC Public

Health. – 2024. – Vol. 24, No. 1. – P. 159. – EDN: TQUSSL. – https://doi.org/10.1186/s12889-023-17457-y.

9. Оценка данных о доходах населения в региональном разрезе методом главных компонент / Б. Г. Ильясов, Е. А. Макарова, Е. Ш. Закиева, Э. С. Гиздатуллина // Экономика региона. – 2019. – Т. 15, № 2. – С. 601-617. – EDN: FELIMW. – https://doi.org/10.17059/2019-2-22.
10. Ilyasov, B. Data mining for estimating living standards in the constituent entities of the Russian Federation / E. Makarova, E. Zakieva, E. Gabdullina, T. Teregulov // 2nd Siberian scientific workshop on data analysis technologies with applications (SibDATA-2021), 25 june 2021. – 2021. – Vol. 3047. – P. 51-56. – EDN: LQNJOP.