

ОРИГИНАЛЬНАЯ СТАТЬЯ

DOI: 10.26794/3030-7097-2026-2-1-45-51
УДК 004.852:330.43:311.14:339.9(045)

Методы интеллектуального анализа данных в исследовании экономического развития стран

Л.Р. Борисова

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

АННОТАЦИЯ

Цель исследования – выявить взаимосвязи между относительными экономическими показателями (индексами развития) стран мира на основе данных Всемирного банка за 2023 г. Для решения задачи применены методы машинного обучения и анализа структурных уравнений. **Материалы и методы.** Исследование охватило 205 стран – при анализе методом взвешенных синдромов; 180 стран – при одно- и двухфакторном подтверждающем анализе. В качестве исходных данных использованы восемь ключевых относительных экономических показателей развития стран. Группирующим критерием выступил темп роста ВВП в 2023 г. (в %). Для анализа применены метод взвешенных синдромов; одно- и двухфакторный подтверждающий анализ; оценка качества моделей по индикатору ROC AUC, индексам сравнительного соответствия и Такера-Льюиса (TLI). **Результаты.** 1. Модель на основе метода взвешенных синдромов показала высокое качество распознавания: ROC AUC = 0,92. Диаграммы рассеяния подтвердили четкое разделение стран на две группы по анализируемому показателю. 2. При факторном анализе построены: однофакторная модель с низкими нагрузками ($\lambda=0,258$; $\lambda=0,131$), что указывает на слабую связь наблюдаемых переменных с латентным фактором; двухфакторная модель с неудовлетворительным индексом соответствия TLI = 0,474 (при пороге > 0,90–0,95). **Выводы.** Методы машинного обучения продемонстрировали преимущество при работе с данными, содержащими пропуски. Факторный анализ дал неудовлетворительные результаты: модели не объясняют дисперсию переменных и не соответствуют данным. Содержательная экономическая интерпретация латентных факторов не проведена из-за низкого качества факторных моделей. Сравнение методов машинного обучения и факторного анализа затруднено из-за разного объема пригодных данных.

Ключевые слова: машинное обучение; факторный анализ; экономические показатели; индексы развития; ВВП; ROC AUC; индекс Такера-Льюиса; взвешенные синдромы; структурные уравнения

Для цитирования: Борисова Л.Р. Методы интеллектуального анализа данных в исследовании экономического развития стран. *Цифровые решения и технологии искусственного интеллекта.* 2026;2(1):45-51. DOI: 10.26794/3030-7097-2026-2-1-45-51

ORIGINAL PAPER

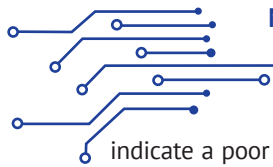
Methods of Data Mining in the Study of Economic Development of Countries

L.R. Borisova

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

ABSTRACT

The paper presents a demonstration of the analysis of a specific data set using machine learning methods to solve the research problem of finding a link between relative economic indicators (development indices) of the world's countries according to the World Bank data for 2023 using machine learning methods and the use of structural equations. This approach was applied to the analysis of 205 countries using the weighted syndrome method and 180 countries using one- and two-factor confirmatory analysis. The advantage of using machine learning methods is the ability to use data with gaps, unlike regression models, which are the basis of factor analysis, when data gaps are not acceptable. To use the weighted syndromes method, eight main economic relative indicators of the countries' development were used. The % GDP growth rate for 2023 was chosen as the grouping indicator. The quality of the model was assessed by the ROC AUC indicator. This indicator is 0.92, which indicates that the selected features really make it possible to divide the countries. Scattering diagrams showing a clear division of countries into two groups according to the analyzed indicators also illustrate the quality of recognition. The use of factor analysis made it possible to build two models (one-factor and two-factor) using not eight indicators, but only four, so that the models measured by the indices of comparative conformity and Tucker-Lewis were statistically significant. However, the loads of the one-factor model (λ) are extremely low (0.258, 0.131), which indicates a weak relationship between the observed variables and the latent factor. The factor practically does not explain the variance of the variables. These results



indicate a poor correspondence of the model to the data, especially for the two-factor model, since the value of the TLI index is 0.474, which also cannot be a satisfactory result with an adequate threshold of $>0.90-0.95$). A meaningful economic interpretation of the latent factors obtained is not provided due to the poor correspondence of the results of factor analysis and the inability to compare machine learning and factor analysis methods with simultaneous analysis, since there was less data suitable for factor analysis than when using machine learning methods that demonstrated their adequacy.

Keywords: machine learning; factor analysis; economic indicators; development indices; GDP; ROC AUC; Tucker Lewis index; weighted syndromes; structural equations

For citation: Borisova L.R. Methods of data mining in the study of economic development of countries. *Digital solutions and artificial intelligence technologies*. 2026;2(1):45-51. DOI: 10.26794/3030-7097-2026-2-1-45-51

ВВЕДЕНИЕ

Мировой опыт свидетельствует о существовании значительных диспропорций в развитии отдельных стран. Проблема оценки рационального развития национальной экономики становится актуальной, поскольку такие оценки могут предотвратить диспропорции в развитии между странами. Вследствие этого возникает необходимость использования методов математического моделирования для анализа относительных показателей экономического развития. Чаще всего применяются методы математической статистики: регрессионные модели, кластерный и факторный анализы [1].

Обзор литературы. В работе [2] использован кластерный анализ. Количество кластеров определено заранее — остановились на трех кластерах. Полученные результаты показывают, что странам кластера 1 требуется существенная поддержка в области управления и инфраструктуры, в то время как странам кластера 2 выгодны инвестиции в институциональную сферу и логистику. Кластер 3 демонстрирует примерные показатели управления и социально-экономических стандартов, предлагая модели устойчивого развития.

В работе [3] с помощью методов машинного обучения удалось установить, что учет ежемесячных данных о суровых климатических условиях — особенно информации, содержащейся в значениях, превышающих среднее (выше среднего) по экстремальным погодным условиям, — значительно превосходит не только базовую авторегрессионную модель, но и эконометрическую структуру, включающую макрофинансовые факторы, при прогнозировании темпов роста квартального реального валового накопления основного капитала.

В работе [4], внедряя принципы энергетической иерархии, обратной связи и самоорганизации Юджина Одума¹, предложена модель для диагностики системных дисбалансов и разработки экологически обоснованных стратегий развития. Модель предлагает более последовательную альтернативу

традиционным экономически ориентированным показателям.

Помимо статистических методов, в математическом моделировании очень часто используется аппарат дифференциальных уравнений. В частности, в работе [5] представлена новая математическая модель для исследования влияния развития навыков на управление циклической безработицей. Сформирована и проанализирована система нелинейных дифференциальных уравнений. На основании полученных результатов сделан вывод, что развитие навыков у циклически безработных — эффективное средство снижения безработицы.

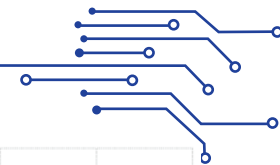
В работе [6] исследована взаимосвязь между экономическим ростом и теневой экономикой. Результаты моделирования показали, что качество институтов тесно связано с взаимосвязью между экономическим ростом и теневой экономикой: более высокий ВВП на душу населения соответствует меньшему уровню теневой экономики в странах с хорошим качеством институтов. Однако в странах с низким качеством институциональной среды увеличение ВВП на душу населения не влияет на масштабы теневой экономики.

В работе [7] выполнен учет безработицы и других экономических факторов при исследовании развития территориальных образований. Рассмотрена математическая модель линейной и многофакторной регрессии, описывающая взаимосвязь между количеством затраченных ресурсов и объемом произведенной продукции. Созданная модель четко показывает, что сельское хозяйство в местных сообществах может стать экономическим двигателем развития с учетом региональных условий.

Отечественные ученые в исследованиях делают основной акцент на использование методов математической статистики и эконометрики. Стоит отметить в этой связи работы В. С. Степанова с соавторами [8, 9], в которых построены интересные регрессионные модели, связывающие уровень жизни населения и уровень онкозаболеваемости с индикаторами социально-экономического развития в регионах России.

Также заслуживают внимания работы С. С. Матвеевского и Л. Р. Борисовой, в которых исследуется

¹ Американский ученый, занимающийся моделированием в экологии.



классификация и кластеризация отдельных регионов или стран по значениям инклюзивного роста [10] и электронных услуг и сервисов [1].

МАТЕРИАЛЫ И МЕТОДЫ

Для исследования взяты данные с сайта Всемирного банка² за 2023 г., так как за 2024–2025 гг. данных недостаточно.

В базу данных для использования методов машинного обучения вошли 205 стран, для которых представлено большинство из следующих восьми показателей: рост ВВП в %; рост ВВП на душу населения в %; инфляция, измеряемая ростом потребительских цен (% в годовом исчислении); годовой прирост в промышленности в %; % от ВВП добавленной стоимости в промышленности; % от ВВП добавленной стоимости в обрабатывающей промышленности; импорт и экспорт товаров и услуг в виде % от ВВП; доступ к электричеству (% населения).

В качестве группирующего показателя при использовании методов машинного обучения был взят рост ВВП в % за 2023 г. В качестве метода распознавания применен метод статистически значимых синдромов (СВС) [11]. Метод СВС основан на процедуре взвешенного голосования по системам так называемых синдромов — областей признакового пространства, содержащих преимущественно объекты одного из классов [1]. Методы машинного обучения реализованы в системе анализа данных Data Master Azforus, использованной ранее в работе [12].

Подтверждающий факторный анализ завершен с применением статистического языка программирования R³.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

При выполнении скользящего контроля в результатах распознавания наблюдаются ошибки. Были отобраны случайным образом 42 страны из 205 (20%), число стран в каждой группе было по 21. Полученное значение показателя ROC AUC, равное 0,92 (близкое к 1), свидетельствует в пользу использования методов машинного обучения для классификации стран по показателям, важным для экономического развития. Заметим, что страны четко разделяются по таким основным показателям, как ВВП на душу населения и экспорт товаров и услуг (диаграмма рассеяния представлена на *рис. 1*). Гра-

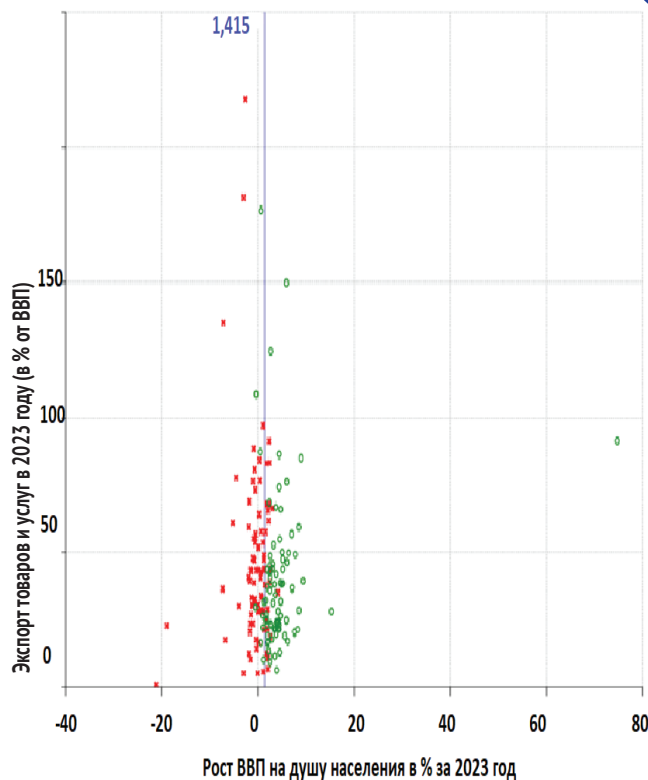


Рис. 1 / Fig. 1. Различие между двумя группами стран по зависимости экспорта товаров и услуг от роста ВВП в 2023 г. / The Difference Between the Two Groups of Countries in the Dependence of Exports of Goods and Services on GDP Growth in 2023

Источник / Source: составлено автором / Compiled by the author.

ница, разделяющая страны с ростом ВВП за 2023 г. с уровнем, ниже медианного от стран с ростом ВВП выше медианного уровня, проходит на уровне 1,415%. Интересно заметить, что эта граница не изменится, если представить диаграмму зависимости импорта товаров и услуг от роста ВВП или зависимости годового прироста в промышленности (в %) от роста ВВП; % от ВВП добавленной стоимости в промышленности от роста ВВП.

Факторный анализ

Моделирование структурными уравнениями охватывает широкий спектр линейных моделей, а именно линейную и многомерную регрессии, анализ траекторий, подтверждающий факторный анализ и структурную регрессию [1]. Из огромного количества экономических показателей стран, размещенных на сайте Всемирного банка, выбраны:

- показатель инфляции в виде роста потребительские цены (% в годовом исчислении) — x_1 ;
- показатель добавленной стоимости в промышленности (включая строительство) (% от ВВП) — x_2 ;

² URL: <https://www.worldbank.org/>

³ Introduction to Structural Equation Modeling (SEM) in R with lavaan. URL: <https://stats.oarc.ucla.edu/r/seminars/rsem/>

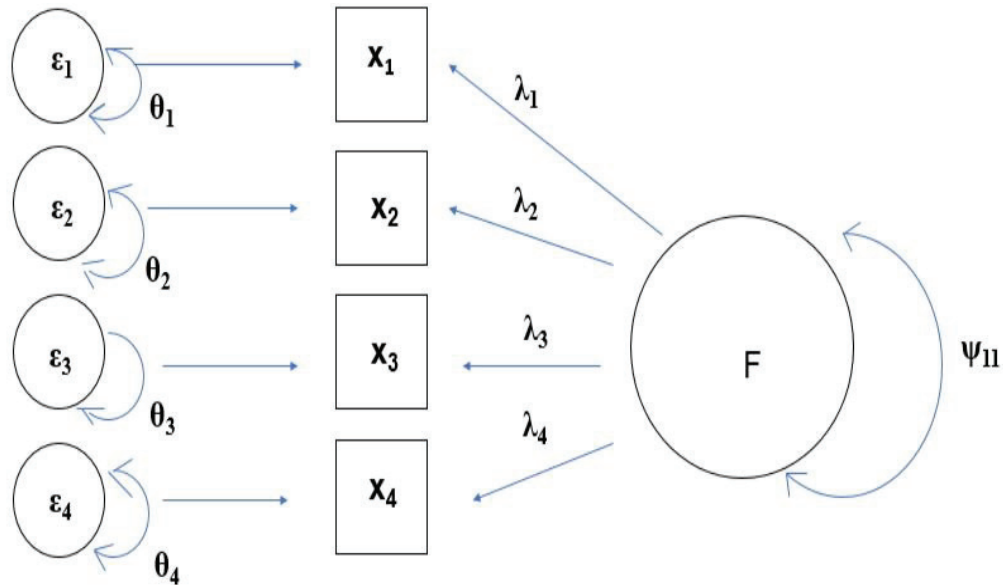


Рис. 2 / Fig. 2. Схема однофакторного подтверждающего анализа с четырьмя позициями /
 Scheme of One-factor Confirmatory Analysis with four Positions

Источник / Source: составлено автором / Complied by the author.

- процент населения, имеющих доступ к электричеству, — x_3 ;
- экспорт товаров и услуг (% от ВВП) — x_4 .

Все показатели — за 2023 г. Факторный анализ используется для ответа на вопрос: насколько общая дисперсия является общей между элементами?

Однофакторная модель

Факторный анализ — это многомерная модель, в которой на каждого субъекта приходится столько результатов, сколько и элементов. В линейной регрессии для каждого субъекта существует только один исход. Наиболее фундаментальной моделью в подтверждающем факторном анализе является однофакторная модель, предполагающая, что ковариация между элементами обусловлена одним общим фактором [1]. Рассмотрим четыре показателя и один фактор (рис. 2).

Параметры:

λ_i ($i = 1, \dots, 4$) — нагрузки, которые можно интерпретировать как корреляцию элемента с фактором F;
 ε_i ($i = 1, \dots, 4$) — остатки факторной модели, то, что остается после учета фактора; θ_i ($i = 1, \dots, 4$) — дисперсии (ковариации) остатков.

Основными индикаторами качества факторной модели являются хи-квадрат (χ^2 -распределение):

- статистика, получаемая из статистики максимального правдоподобия;
- индекс сравнительного соответствия CFI, значения которого могут варьироваться от 0 до 1 (значения больше 0,90 указывают на хорошее соответствие);

- индекс Такера-Льюиса TLI, также находящийся в диапазоне от 0 до 1 (значения больше 0,90 указывают на хорошее соответствие);

- среднеквадратичная ошибка аппроксимации RMSEA (если эта ошибка меньше 0,05, то соответствие модели данным хорошее).

Для анализируемых данных получены следующие значения индикаторов качества: CFI = 0,967, TLI = 0,901, RMSEA = 0,039. Как видим, модель статистически значимая. Осталось проанализировать корреляцию фактора с переменными. В табл. 1 представлены значения полученных при моделировании параметров [1].

Заметим, что в данной модели $\psi_{11} = 1$. Для всех дисперсий, кроме x_4 , величина p-value = 0. Переменная x_4 — экспорт товаров и услуг (в % к ВВП). В этом случае величина p-value = 0,006, т.е. близко к нулю. Если под фактором мы понимаем некий интегральный социально-экономический показатель, определяющий величину замедления роста экономики в регионах, то интересно отметить, что положительно на этот фактор влияют только два показателя; инфляция и показатель добавленной стоимости в промышленности (в % от ВВП). И наоборот, доступ к электричеству и экспорт товаров способствуют росту экономики (параметры λ_1 и λ_2 — положительные, λ_3 и λ_4 — отрицательные).

Двухфакторная модель

Рассмотрим четыре показателя и два фактора. Схема модели имеет вид, представленный на рис. 3.

Значения параметров однофакторной модели / Values of the Parameters of the One-factor Model

λ	θ	λ	θ
$\lambda_1 = 0,258$	$\theta_1 = 0,928$	$\lambda_3 = -0,446$	$\theta_3 = 0,795$
$\lambda_2 = 0,131$	$\theta_2 = 0,977$	$\lambda_4 = -0,595$	$\theta_4 = 0,641$

Источник / Source: составлено автором / Complied by the author.

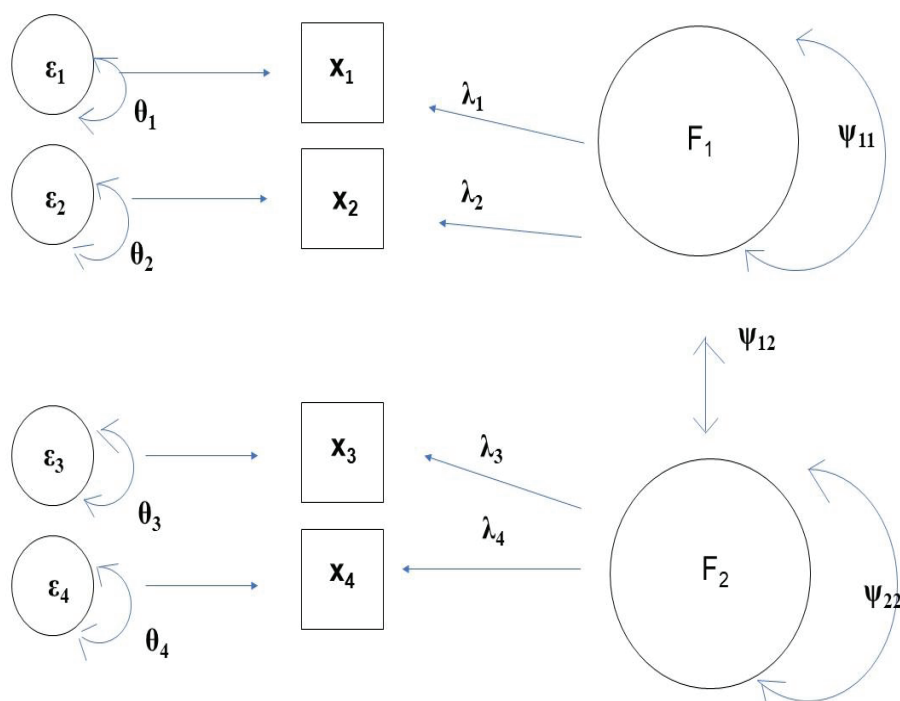


Рис. 3 / Fig. 3. Схема двухфакторного анализа с четырьмя позициями / The Scheme of Two-factor Confirmatory Analysis with Four Positions

Источник / Source: составлено автором / Complied by the author.

Как и в случае однофакторного анализа, моделирование в случае предположения о существовании двух скрытых факторов, влияющих на экономический региональный рост, было выполнено с использованием программы lavaan, имеющейся в библиотеке программ языка R. Для двухфакторного анализа зафиксировали две группы показателей: в первую вошли показатель инфляции и показатель добавленной стоимости в промышленности, во вторую — доступ к электричеству и экспорт

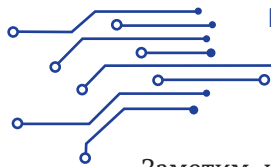
товаров. Оказалось, что два латентных фактора имеют отрицательную ковариацию $-\psi_{12} = -1,592$. Для анализируемых данных получены следующие значения индикаторов качества: CFI = 0,912, TLI = 0,474, RMSEA = 0,09. Как видим, модель статистически значимая. Осталось проанализировать корреляцию двух скрытых факторов с исследуемыми переменными. В табл. 2 представлены значения полученных при моделировании параметров [1].

Таблица 2 / Table 2

Значения параметров двухфакторной модели / Values of the Two-factor Model Parameters

λ	θ	λ	θ
$\lambda_1 = 0,164$	$\theta_1 = 0,968$	$\lambda_3 = 0,442$	$\theta_3 = 0,799$
$\lambda_2 = 0,087$	$\theta_2 = 0,987$	$\lambda_4 = 0,597$	$\theta_4 = 0,638$

Источник / Source: составлено автором / Complied by the author.



Заметим, что в данной модели $\psi_{11} = 1, \psi_{22} = 1$. Интересно отметить, что наличие второго скрытого фактора изменило направление связи исследуемых признаков и фактора: все параметры $\lambda_i (i = 1, \dots, 4)$ — положительные, но сами латентные факторы между собой отрицательно коррелируют ($\psi_{12} = -1,592$).

ВЫВОДЫ

В работе проанализированы данные за 2023 г. по экономическому развитию стран мира, представленные на сайте Всемирного банка двумя разными методами: методом машинного обучения и факторным анализом. В качестве метода классификации был использован метод статистически взвешенных синдромов, который продемонстрировал хорошее качество распознавания на контрольной выборке, составившей 20% от генеральной совокупности (площадь под РОК-кривой равна 0,92, что достаточно близко к 1). Следует отметить, что при применении более распространенных ал-

горитмов классификации — например, случайного леса или моделей бустинга — точность распознавания оказалась ниже по сравнению с методом статистически взвешенных синдромов.

Следовательно, применение технологий машинного обучения подтвердило, что выбранный набор экономических показателей пригоден для классификации стран.

Преимущество данного подхода заключается в том, что он позволяет включать в анализ страны, имеющие пропуски в данных.

После уменьшения числа показателей с 8 до 4 и уменьшения стран в базе данных для исследования были использованы структурные уравнения. Оказалось, что применение факторного анализа также возможно при получении статистически значимых результатов, которые заключаются в том, что существуют некоторые латентные факторы, участвующие во взаимосвязи между исследуемыми экономическими показателями.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Кузнецова А.В., Борисова Л.Р., Кремер Н.Ш., Фридман М.Н. Сравнительный анализ дотационных регионов Российской Федерации методами машинного обучения по широкому набору показателей основных фондов. *Бизнес. Образование. Право*. 2025;1(70):20-28. DOI: 10.25683/VOLBI.2025.70.1182
Kuznetsova A.V., Borisova L.R., Kremer N.S., Friedman M.N. Comparative analysis of subsidized regions of the Russian Federation using machine learning methods for a wide range of indicators of fixed assets. *Biznes. Obrazovanie. Pravo = Business. Education. Law*. 2025;1(70):20-28. (In Russ.). DOI: 10.25683/VOLBI.2025.70.1182
2. Visbal-Cadavid D., Delahoz-Dominquez E., Mendoza-Mendoza A. A multiple factor analysis and hierarchical clustering of global logistics governance and development. *Decision Analytics Journal*. 2025;(15):100579. DOI: 10.1016/j.dajour.2025.100579
3. Sheng X., Cepni O., Gupta R., Markovski M. Mixed Frequency Machine Learning Forecasting of the Growth of Real Gross Fixed Capital Formation in the United States: The Role of Extreme Weather Conditions 2025: Working Papers University of Pretoria. 2025;202520. URL: <https://EconPapers.repec.org/RePEc:pre:wpaper:202520>
4. Langa E.S., Giannetti B.F., Sevegnani f., Agostinho F., Almeida C. From theory to application: measuring development disparities in Mozambique through an Odum-inspired emergy framework. *Ecological Modeling*. 2025;(510):111287. DOI: 10.1016/j.ecolmodel.2025.111287
5. Yahyaoui M.El., Amine S. Mathematical modeling of unemployment dynamics with skills development and cyclical effects. *Partial differential equations in applied mathematics*. 2024; (911):100800. DOI: 10.1016/j.padiff.2024.100800
6. Baklouti N., Boujelbene Y. A simultaneous equation model of economic growth and shadow economy: Is there a difference between the developed and developing countries? *Economic Change and Restructuring*, Springer. 2020;53(1):151-170. DOI: 10.1007/s10644-018-9235-8
7. Bihun R., Lytvyn V., Oleksiv N. Mathematical modeling and analysis of the development of territorial communities. *Technology Audit and Production Reserves*. 2021;3(2(59)):6-12. DOI: 10.15587/2706-5448.2021.232788
8. Степанов В.С., Бобков В.Н., Шамаева Е.Ф., Одинцова Е.В. Построение модели, связывающей индикатор уровня жизни населения с комплексом показателей социально-экономической политики в регионах России. *Уровень жизни населения регионов России*. 2022;18(4):450-465. (In Russ.). DOI: 10.19181/Ispr.2022.18.4.3



- Stepanov V.S., Bobkov V.N., Shamaeva E.F., Odintsova E.V. Building a model linking the indicator of the standard of living of the population with a set of indicators of socio-economic policy in the regions of Russia. *Living Standard of the Population in the Regions in Russia*. 2022;18(4):450-465. (In Russ.). DOI: 10.19181/lsprr.2022.18.4
9. Степанов В.С. Прогноз общей онкологической заболеваемости в регионах и муниципалитетах России на основе многофакторной модели. *Моделирование, оптимизация и информационные технологии*. 2023;11(1):1-17. DOI: 10.26102/2310-6018/2023.40.1.022
Stepanov V.S. The forecast of cancer prevalence in the regions and municipalities of Russia based on a multivariate model. *Modeling, Optimization and Information Technology*. 2023;11(1):1-17. DOI: 10.26102/2310-6018/2023.40.1.022 (In Russ.).
10. Матвиевский С.С., Борисова Л.Р. Кластеризация стран Азиатско-Тихоокеанского региона по значениям инклюзивного экономического роста. *Вестник университета*. 2023;(1):112-121. DOI: 10.26425/1816-4277-2024-1-112-121
Matviyevsky S.S., Borisova L.R. Clusterization of the countries of the Asia-Pacific region according to the values of inclusive economic growth. *Bulletin of the University*. 2023; (1):112-121. (In Russ.). DOI: 10.26425/1816-4277-2024-1-112-121
11. Борисова Л.Р. Сравнительный анализ регионов Российской Федерации методами машинного обучения по набору показателей электронных услуг и сервисов. *Цифровая социология*. 2024;7(4):33-43. DOI: 10.26425/2658-347X-2024-7-4-33-43
Borisova L.R. Comparative analysis of the regions of the Russian Federation using machine learning methods for a set of indicators of electronic services and services. *Digital Sociology*. 2024;7(4):33-43. (In Russ.). DOI: 10.26425/2658-347X-2024-7-4-33-43
12. Senko O.V., Kuznetsova A.V. A recognition method based on collective decision-making using systems of regularities of various types. *Pattern Recogn. Image Anal.* 2010;20(2):152-162. DOI: 10.1134/S1054661810020069

ИНФОРМАЦИЯ ОБ АВТОРЕ / ABOUT THE AUTOR

Людмила Робертовна Борисова — кандидат физико-математических наук, доцент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Lyudmila R. Borisova — Cand. Sci. (Phys. And Math.) Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<http://orcid.org/0000-0002-5757-0341>

lrborisova@fa.ru

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Conflicts of Interest Statement: The author has no conflicts of interest to declare.

Статья поступила 23.09.2025; после рецензирования 31.10.2025; принята к публикации 15.12.2025.

Автор прочитала и одобрила окончательный вариант рукописи.

The article was submitted on 23.09.2025; revised on 31.10.2025 and accepted for publication on 15.12.2025.

The author read and approved the final version of the manuscript.