

УДК 004.855.5:330.59

<http://doi.org/10.35854/1998-1627-2025-3-348-358>

О применении алгоритмов машинного обучения для прогнозирования индекса качества жизни населения

Хакимджон Иномджонович Аминов¹✉, Анастасия Владиславовна Кардаш²

¹ Санкт-Петербургский государственный экономический университет, Санкт-Петербург, Россия

² Филиал АО «ДРТ», Санкт-Петербург, Россия

¹ h_aminov@unecon.ru✉, <https://orcid.org/0009-0004-0183-280X>

² nastykardash@mail.ru

Аннотация

Цель. Исследовать возможности применения различных алгоритмов машинного обучения для прогнозирования индекса качества жизни населения.

Задачи. Разработать прогнозные модели для анализа индекса качества жизни населения выбранных стран (России, Германии, Индии, Нидерландов), используя различные алгоритмы машинного обучения на основе исторических данных веб-сайта Numbeo с 2012 по 2025 г.; систематизировать и провести анализ полученных результатов моделей машинного обучения для указанных государств.

Методология. В процессе исследования применены такие модели машинного обучения, как случайный лес, линейная регрессия, градиентный бустинг, k -ближайших соседей и метод опорных векторов. Прогнозирование индекса качества жизни населения основано на данных о социально-экономических факторах для различных стран, представленных в базе данных Numbeo.

Результаты. Проведен сравнительный анализ результатов прогнозирования индекса качества жизни населения выбранных стран с использованием алгоритмов машинного обучения на базе исторических данных с 2012 по 2025 г. Особое внимание уделено настройке гиперпараметров моделей и кросс-валидации для повышения точности предсказаний. Анализ свидетельствует о том, что наиболее достоверные результаты можно получить с применением ансамбля моделей машинного обучения без учета прогнозов линейной регрессии.

Выводы. Выполненные расчеты показали, что модель градиентного бустинга демонстрирует наилучшие результаты. Однако для улучшения точности и уменьшения отклонений рекомендуется использовать ансамбль моделей. Применение машинного обучения в прогнозировании открывает новые возможности для разработки социальных государственных программ, направленных на улучшение качества жизни населения.

Ключевые слова: прогнозирование, индекс качества жизни, факторы, машинное обучение, алгоритмы, модели

Для цитирования: Аминов Х. И., Кардаш А. В. О применении алгоритмов машинного обучения для прогнозирования индекса качества жизни населения // *Экономика и управление*. 2025. Т. 31. № 3. С. 348–358. <http://doi.org/10.35854/1998-1627-2025-3-348-358>

The application of machine learning algorithms for forecasting the quality of life index of the population

Khakimdzhon I. Aminov^{1✉}, Anastasiya V. Kardash²

¹ St. Petersburg State University of Economics, St. Petersburg, Russia

² Branch of JSC DRT, St. Petersburg, Russia

¹ h_aminov@unecon.ru ✉, <https://orcid.org/0009-0004-0183-280X>

² nastykardash@mail.ru

Abstract

Aim. The work aimed to investigate the possibilities of applying various machine learning algorithms to forecast the quality of life index of the population.

Objectives. The work seeks to develop predictive models for analyzing the quality of life index of the population of selected countries (Germany, India, the Netherlands, Russia) using various machine learning algorithms based on historical data from the Numbeo website from 2012 to 2025; as well as to systematize and analyze the results of machine learning models for these countries.

Methods. The study used machine learning models such as random forest, linear regression, gradient boosting, k-nearest neighbors, and support vector machine. Forecasting the quality of life index of the population is based on data on socio-economic factors for various countries presented in the Numbeo database.

Results. A comparative analysis of the results of forecasting the quality of life index of the population of selected countries was performed using machine learning algorithms based on historical data from 2012 to 2025. Particular attention is paid to adjusting the hyperparameters of the models and cross-validation to improve the accuracy of predictions. The analysis demonstrated that the most reliable results can be obtained using an ensemble of machine learning models without taking into account linear regression forecasts.

Conclusion. The calculations performed revealed that the gradient boosting model demonstrates the best results. However, in order to improve accuracy and reduce deviations, it is recommended to use an ensemble of models. The use of machine learning in forecasting offers new opportunities for the development of social government programs aimed at improving the quality of life of the population.

Keywords: forecasting, quality of life index, factors, machine learning, algorithms, models

For citation: Aminov Kh.I., Kardash A.V. The application of machine learning algorithms for forecasting the quality of life index of the population. *Ekonomika i upravlenie = Economics and Management*. 2025;31(3): 348-358. (In Russ.). <http://doi.org/10.35854/1998-1627-2025-3-348-358>

Введение

Вопросы прогнозирования качества жизни населения остаются актуальными в современном обществе с учетом динамично изменяющихся социально-экономических условий и вызовов, с которыми сталкиваются различные регионы мира. Качество жизни населения является сложным и многогранным показателем, который измеряется с использованием различных индексов [1; 2; 3]. К наиболее популярным в мире индикаторам социально-экономического благополучия населения относится индекс качества жизни населения от веб-сайта Numbeo [4]. Данный индекс характеризуется широким охватом стран, использованием краудсорсинга и учетом множества факторов, что

предоставляет возможность его применения для разработки государственной политики и социальных программ, направленных на улучшение благосостояния граждан. В эпоху развития цифровых технологий и накопления больших данных использование методов машинного обучения для анализа этих факторов открывает новые возможности в прогнозировании.

В области применения алгоритмов машинного обучения для прогнозирования активно проводятся исследования, появляются новые решения схожих проблем. Рассмотрим некоторые особенности существующих исследований [5; 6; 7], их результаты.

Е. А. Петрова, в частности, анализирует процесс использования нейронных сетей для прогнозирования уровня жизни населения

на основе набора социально-экономических и экологических показателей в регионах России [5]. Утверждается, что именно такой подход способен обрабатывать зашумленные и неполные данные, предоставить более надежный и точный подход к прогнозированию уровня жизни, что имеет важное значение для регионов, в которых происходят быстрые экономические изменения [5].

Б. Липеса, Э. Оканго, Б. Омоло и Э. Омонди провели обзор применения моделей машинного обучения для прогнозирования ожидаемой продолжительности жизни населения [6]. В качестве одного из решений предложено использование модели машинного обучения на основе XGBoost. В точности предсказаний эта модель превосходит традиционные, открывая новые горизонты для эффективного анализа. Авторы также анализируют различные регрессионные подходы, включая модель случайного леса, которая показала наибольшую точность среди изученных методов. В статье отдельно раскрыт вопрос о применении нейронных сетей для прогнозирования ожидаемой продолжительности жизни [6].

Е. В. Щекотин, В. Л. Гойко, П. А. Басина и В. В. Бакулин исследовали использование машинного обучения для изучения качества жизни населения на основе анализа цифровых следов пользователей социальных сетей [7]. Проанализированы различные подходы к векторизации текстов, включая методы представления данных и нейросетевые модели. Особое внимание уделено выбору алгоритма классификации, при этом предпочтение отдано Rubert-tiny ввиду его высокой скорости обучения и классификации [7].

Опубликованные исследования свидетельствуют о том, что использование алгоритмов машинного обучения в прогнозировании является актуальным в настоящее время и имеет потенциал для развития в связи с возможностью применения новых моделей машинного обучения. Новые подходы к обработке данных позволяют увидеть то, что раньше было скрыто, выявляя неочевидные тренды и зависимости. Исходя из этого, становится понятным, что выбор темы настоящего исследования обусловлен необходимостью применения современных алгоритмов машинного обучения, которые позволяют точнее прогнозировать изменения такого социально-значимого показателя, как индекс качества жизни населения.

Материалы и методы

В нашей статье прогнозирование индекса качества жизни населения осуществлено на основании исторических данных выбранных стран (России, Германии, Индии, Нидерландов) за 2012–2025 гг., представленных в базе Numbeo [4]. Индекс качества жизни населения определяют в зависимости от ряда факторов. К основным можно отнести значение «индекса покупательной способности, уровень безопасности, уровень медицинского обслуживания, стоимость жизни, соотношение цены недвижимости к доходу, показатель времени в пути на транспорте, уровень загрязнения окружающей среды и значение климатического индекса» [4].

В ходе исследования сформирован исходный набор данных, содержащий 1 845 комбинаций наблюдений с рассчитанным значением индекса качества жизни населения. Он разделен на две части: 80 % — обучающая выборка, 20 % — тестовая выборка.

Для анализа данных использованы такие модели машинного обучения, как случайный лес [8], линейная регрессия [9], градиентный бустинг [10], k -ближайших соседей (KNN) [11] и метод опорных векторов (SVR) [12]. Все переменные нормализованы с использованием минимальной и максимальной нормализации в целях улучшения работы моделей. Набор данных разделен на обучающие и проверочные подмножества, а также использована шестикратная перекрестная проверка для повышения надежности результатов. Поскольку исходные данные могли быть несбалансированными, применена техника передискретизации SMOTE [13], позволяющая устранить дисбаланс между классами.

В работе с моделями случайного леса, линейной регрессии, градиентного бустинга, k -ближайших соседей и SVR использована библиотека для моделей машинного обучения Scikit-learn [14]. К тому же применены библиотеки Pandas для обработки табличных данных [15], NumPy — для математических операций [16], Matplotlib — для визуализации результатов [17], SciPy (Stats) — для оценки качества [18]. Код на Python запущен в среде Jupyter Notebook [19].

Ключевой в исследовании стала модель случайного леса [8], которая использована для анализа важности переменных и выявления факторов, наиболее значимых при прогнозировании индекса качества жизни

Лучшие параметры для случайного леса:

Максимальная глубина деревьев — None
 Минимальное число объектов для разбиения узла — 2
 Количество деревьев — 100

Лучшие параметры для градиентного бустинга:

Темп обучения — 0.1
 Максимальная глубина деревьев — 5
 Количество деревьев — 300

Лучшие параметры для SVR:

Коэффициент регуляризации — 100
 Параметр ядра — scale
 Тип ядра — rbf

Рис. 1. Результат выбора наилучших значений гиперпараметров модели

Fig. 1. The result of selecting the best values of the model hyperparameters

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

Линейная регрессия — Среднеквадратичная ошибка (MSE): 427.68402996944144, Коэффициент детерминации (R^2): 0.8284384139671333
 Random Forest — Среднеквадратичная ошибка (MSE): 233.8987164742543, Коэффициент детерминации (R^2): 0.9661736423213164
 Gradient Boosting — Среднеквадратичная ошибка (MSE): 204.7373954033033, Коэффициент детерминации (R^2): 0.9178513353526323
 K-Neighbors Regressor — Среднеквадратичная ошибка (MSE): 264.2353994579946, Коэффициент детерминации (R^2): 0.3939342933769501
 Support Vector Regressor — Среднеквадратичная ошибка (MSE): 346.1217340857438, Коэффициент детерминации (R^2): 0.3611563722313858

Рис. 2. Значения среднеквадратичной ошибки и коэффициента детерминации R^2

Fig. 2. Values of the root mean square error and the determination coefficient R^2

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

населения. Оценка эффективности модели проведена с применением таких метрик, как F1, точность, чувствительность, специфичность и AUC.

Для настройки гиперпараметров моделей случайного леса, градиентного бустинга, метода опорных векторов и с целью улучшения их точности использован подход, при котором для каждой модели задан диапазон значений гиперпараметров, влияющих на ее работу. Для случайного леса настраивали количество деревьев (50, 100, 200), максимальную глубину деревьев (None, 10, 20, 30) и минимальное количество объектов для разбиения узла (2, 5, 10). Для градиентного бустинга настраивали темп обучения (0.01, 0.1, 0.2), количество деревьев (100, 200, 300) и максимальную глубину деревьев (3, 4, 5). Для метода опорных векторов настраивали коэффициент регуляризации (0.1, 1, 10, 100), параметр ядра ('scale', 0.01, 0.1, 1) и тип ядра (rbf). Применительно к каждой модели создан объект соответствующего класса с базовыми настройками.

Далее использован метод кросс-валидации GridSearchCV для подбора оптимальных значений гиперпараметров [20]. GridSearchCV перебирает комбинации гиперпараметров из заранее заданных сеток для каждой модели

и оценивает их с помощью кросс-валидации ($cv = 5$). Оценка производится по метрике средней квадратичной ошибки.

После обучения модели на тренировочных данных и с помощью GridSearchCV найдены оптимальные значения гиперпараметров для каждой модели на базе заданной метрики. В результате выбраны значения, представленные на рисунке 1.

После обучения всех моделей рассчитаны значения среднеквадратичной ошибки и коэффициента детерминации R^2 . Полученные значения указаны на рисунке 2.

На основе полученных данных можно сделать ряд промежуточных выводов.

1. Модель градиентного бустинга показывает наилучшие результаты среди всех моделей. У нее выявлена самая низкая среднеквадратичная ошибка (MSE), в частности 204.74, что говорит о высокой точности предсказаний. Коэффициент детерминации (R^2) равен 0.918. Это означает, что модель объясняет 91.8 % вариаций целевой переменной.

2. Модель случайного леса также демонстрирует высокие результаты: MSE — 233.899, R^2 — 0.966. Эта модель несколько уступает градиентному бустингу, оставаясь очень точной и понятно объясняющей данные.

3. Модель k -ближайших соседей (KNN) занимает третье место: MSE — 264.235, R^2 — 0.394. Ее показатели хуже, чем у бустинга и случайного леса, но все-таки высоки.

4. Модель SVR показывает неоднозначные результаты: MSE — 346.122 (выше, чем у трех предыдущих моделей), но R^2 — 0.361. Тем не менее это указывает на приемлемую объясняющую способность. Однако по точности предсказаний она уступает остальным моделям.

5. Линейная регрессия имеет наихудшие показатели среди всех моделей: MSE — 427.684 (самая высокая ошибка), R^2 — 0.828. Это говорит о том, что линейная модель хуже улавливает сложные зависимости в данных по сравнению с нелинейными алгоритмами.

Несмотря на то, что модели показывают различные результаты после обучения, для прогнозирования предлагаем применить все пять алгоритмов и обратить внимание на адекватность значений. Для получения значений используемых факторов, на основе которых будет выявлен прогноз индекса качества жизни населения, предлагаем использовать доступный для понимания алгоритм.

1. Для каждого показателя необходимо построить авторегрессионный прогноз на три периода вперед с использованием модели ARIMA [21], которая учитывает тенденции и сезонность данных. Иными словами, если была проанализирована покупательная способность страны X за 2012–2025 гг., модель предсказывала его значения на период со второй половины 2025 г. по конец 2026 г.

2. Полученные прогнозы по всем факторам объединены в один набор данных, в котором каждый показатель соответствовал будущему периоду. Этот набор использован как вход для уже обученных моделей машинного обучения, которые оценивали итоговый индекс качества жизни с учетом спрогнозированных значений.

Далее в целях получения прогнозных значений целевой переменной применяем следующий алгоритм:

1. Для каждого шага прогноза используем пять ранее описанных моделей машинного обучения. Для каждого шага модели предсказываем целевую переменную. Результаты сохраняем в словарь прогнозов `predictions_country`.

2. Все предсказания для каждой страны (по каждой модели и для каждого шага про-

гноза) сохраняем в словарь `predictions_all_countries`, ключом к которому служит название страны, а значением — результаты для указанной страны.

После получения прогнозных значений целевой переменной необходимо визуализировать динамику изменения индекса качества жизни населения для конкретной страны в целях определения адекватности полученных данных и наглядного сравнения прогнозов, выявленных с помощью разных моделей машинного обучения.

Результаты и обсуждение

Рассмотрим результаты прогноза индекса качества жизни населения для избранных стран. На рисунке 3 показана визуализация прогнозных значений для Германии. Очевиден тот факт, что данные, полученные по итогам предсказания линейной регрессии, существенно отличаются от остальных.

На рисунке 4 находит отражение аналогичная визуализация для России, при этом данные близки друг к другу.

Получен прогноз для Индии, отраженный на рисунке 5. Как и в случае результатов применительно к России, можно заключить, что данные не слишком отличаются друг от друга.

После визуализации результатов становится очевидным, что модели дают адекватные результаты, но некоторые из них слишком отличаются от остальных. Несмотря на общую точность модели градиентного бустинга, значительные отклонения в прогнозах могли возникнуть вследствие резких колебаний индекса покупательной способности и уровня безопасности, которые чувствительны к внешним макроэкономическим факторам.

В качестве попытки получить наиболее достоверные результаты можно внести предложение относительно того, чтобы объединить модели случайного леса, градиентного бустинга, k -ближайших соседей и SVR в ансамбль, организованный по принципу стэкинга. Результаты расчета значений среднеквадратичной ошибки и коэффициента детерминации R^2 для ансамбля представлены на рисунке 6.

Полученные значения среднеквадратичной ошибки и коэффициента детерминации для ансамблевой модели являются наилучшими среди использованных моделей и указывают на то, что комбинация нескольких

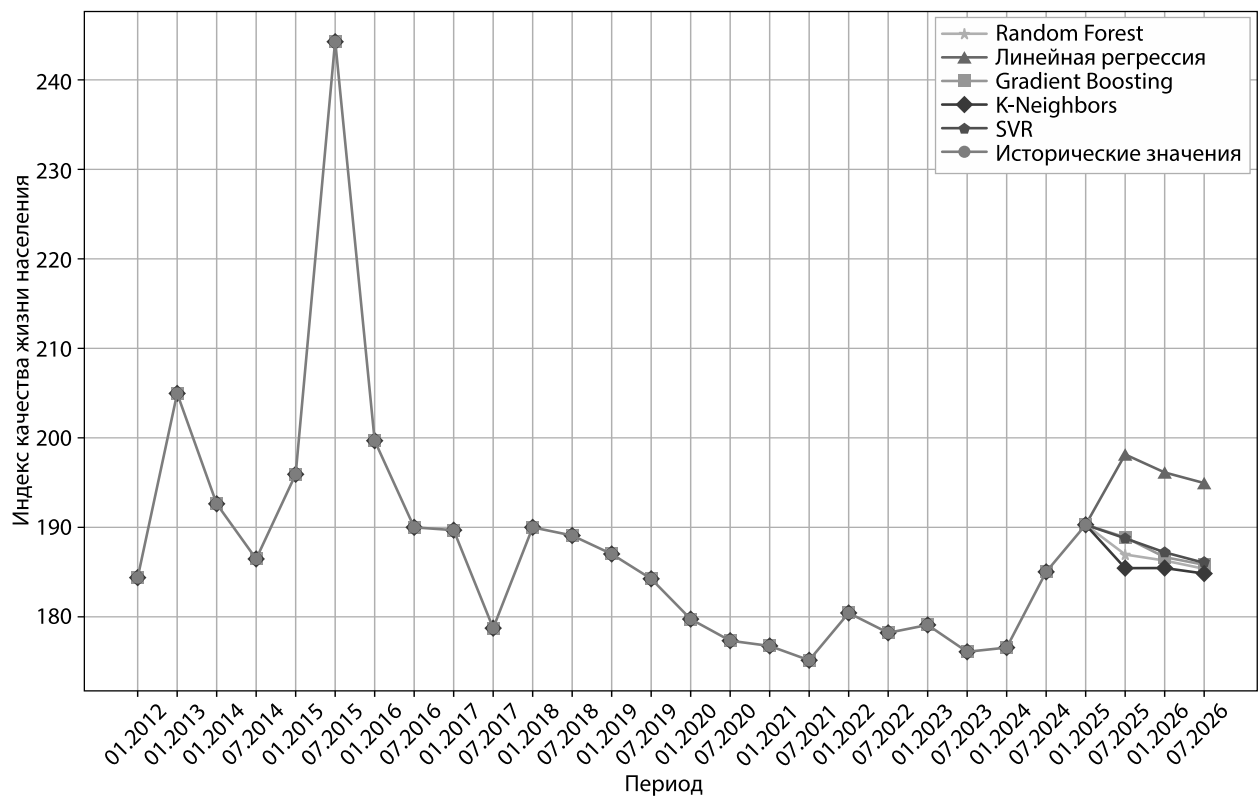


Рис. 3. Динамика изменения индекса качества жизни населения для Германии
Fig. 3. Dynamics of changes in the quality of life index of the population for Germany

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

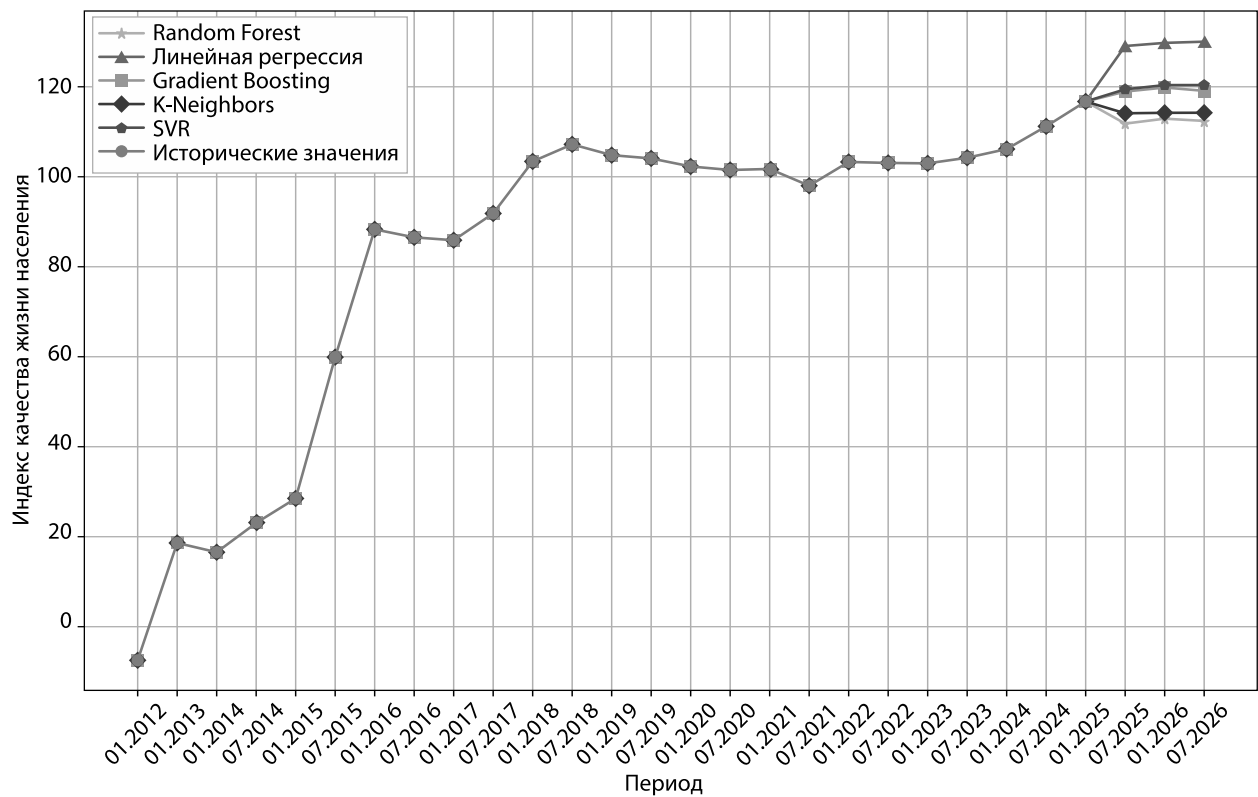


Рис. 4. Динамика изменения индекса качества жизни населения для России
Fig. 4. Dynamics of changes in the quality of life index of the population for Russia

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

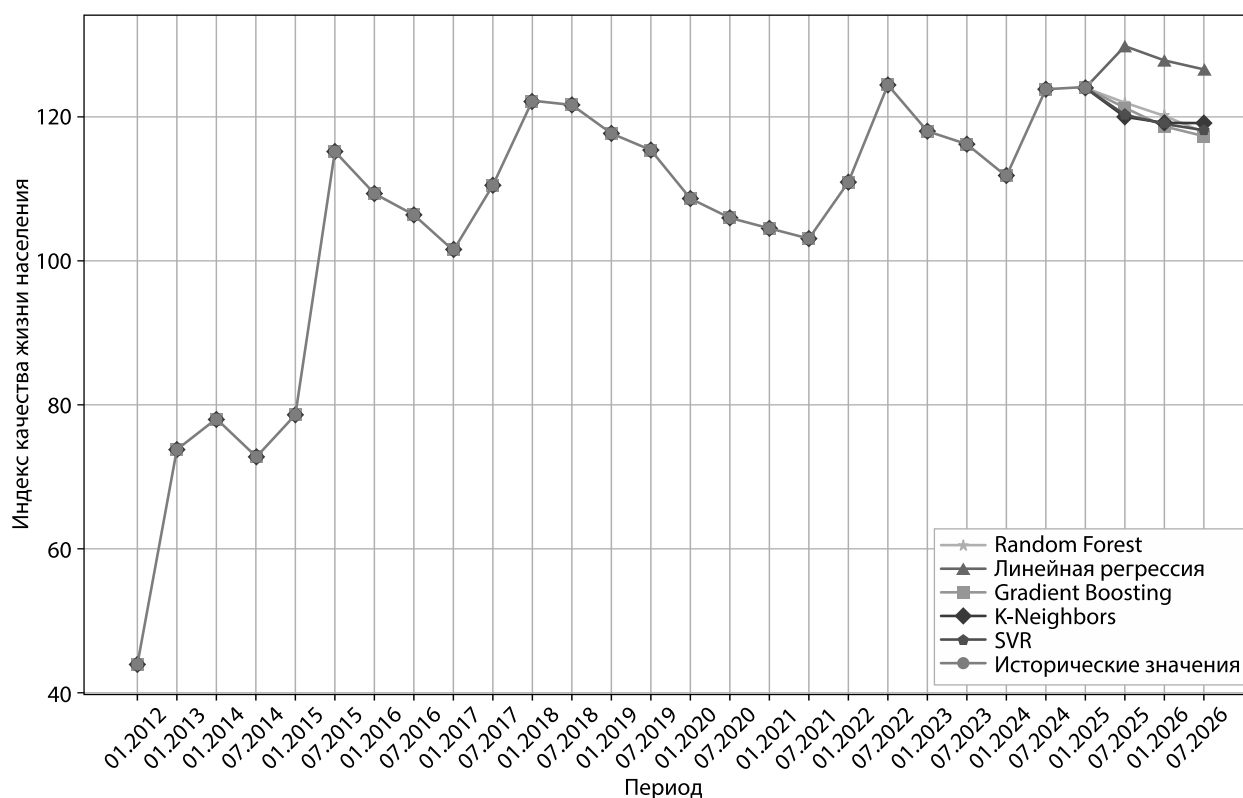


Рис. 5. Динамика изменения индекса качества жизни населения для Индии
Fig. 5. Dynamics of changes in the quality of life index of the population for India

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

Stacking Ensembler — Среднеквадратичная ошибка (MSE): 185.7401589711473,
Коэффициент детерминации (R^2): 0.9254920127240601

Рис. 6. Значения среднеквадратичной ошибки и коэффициента детерминации R^2 для ансамбля
Fig. 6. Values of the root mean square error and the coefficient of determination R^2 for the ensemble

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

алгоритмов позволяет улучшить прогноз за счет компенсации индивидуальных слабостей каждой модели.

Результат применения такого подхода для данных относительно Германии показан на рисунке 7.

На рисунке 8 представлены результаты визуализации прогнозных и усредненных значений для Нидерландов с использованием ансамбля.

Результаты расчета критерия Фридмана ($\chi^2 = 85.759$, $p < 0.001$) демонстрируют наличие статистически значимых различий в точности рассматриваемых моделей машинного обучения. Последующий post-hoc анализ с поправкой Немеи [22] позволил детально исследовать структуру этих различий и выявить значимые пары моделей, как показано на рисунке 9.

Проведенный анализ свидетельствует о том, что модели градиентного бустинга (GBR) и случайного леса (RF) образуют группу лидеров, не демонстрируя значимых различий между собой ($p\text{-value} = 0.9999$). При этом обе модели существенно превосходят по точности линейную регрессию (LR) и метод опорных векторов (SVR). Особенно показательным является сравнение GBR с LR, при котором низкое p -значение подтверждает существенное преимущество более сложных алгоритмов.

Ансамблевая модель (Stacking) занимает промежуточное положение в рейтинге. Она не показывает статистически значимого отличия от лидирующих GBR и RF ($p\text{-value} > 0.99$), но тем не менее наблюдается превосходство над методом k -ближайших соседей (KNN) с p -значением 0.0798.

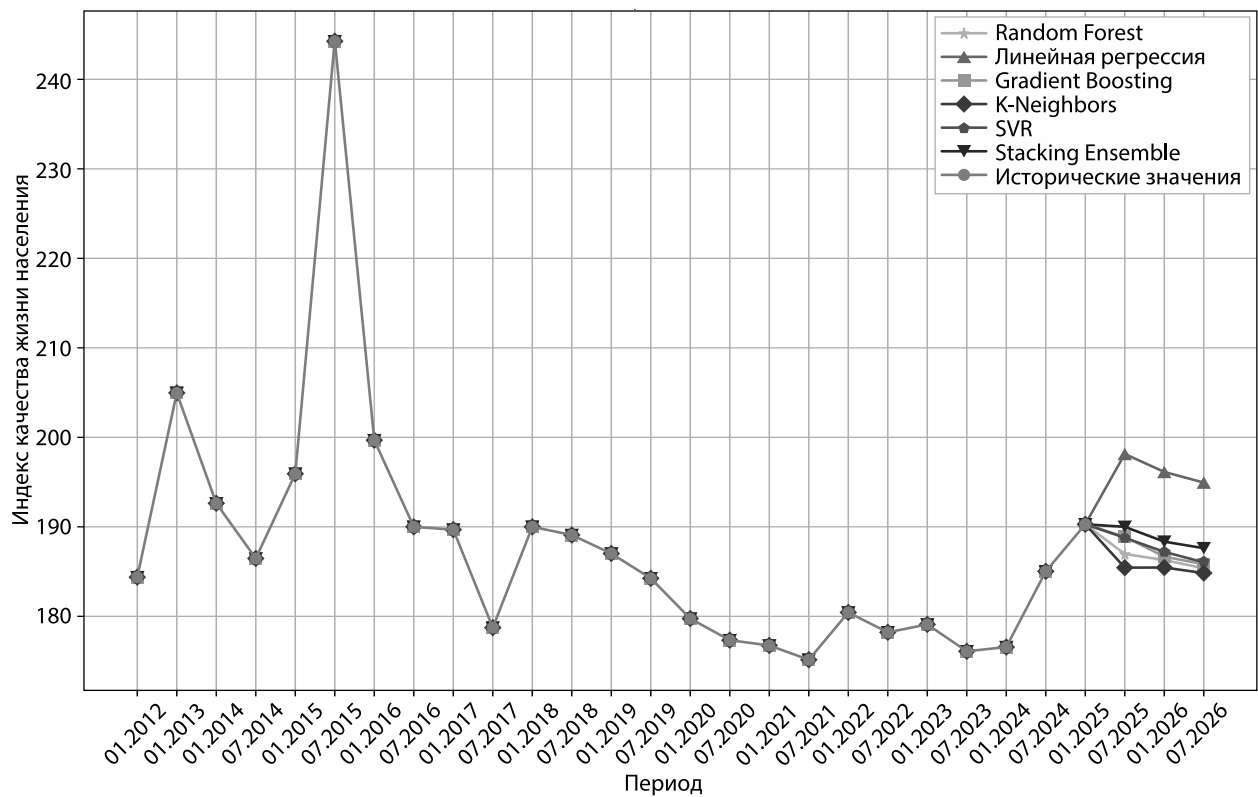


Рис. 7. Динамика изменения индекса качества жизни населения для Германии с прогнозом ансамбля
Fig. 7. Dynamics of changes in the quality of life index for Germany with the ensemble forecast

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

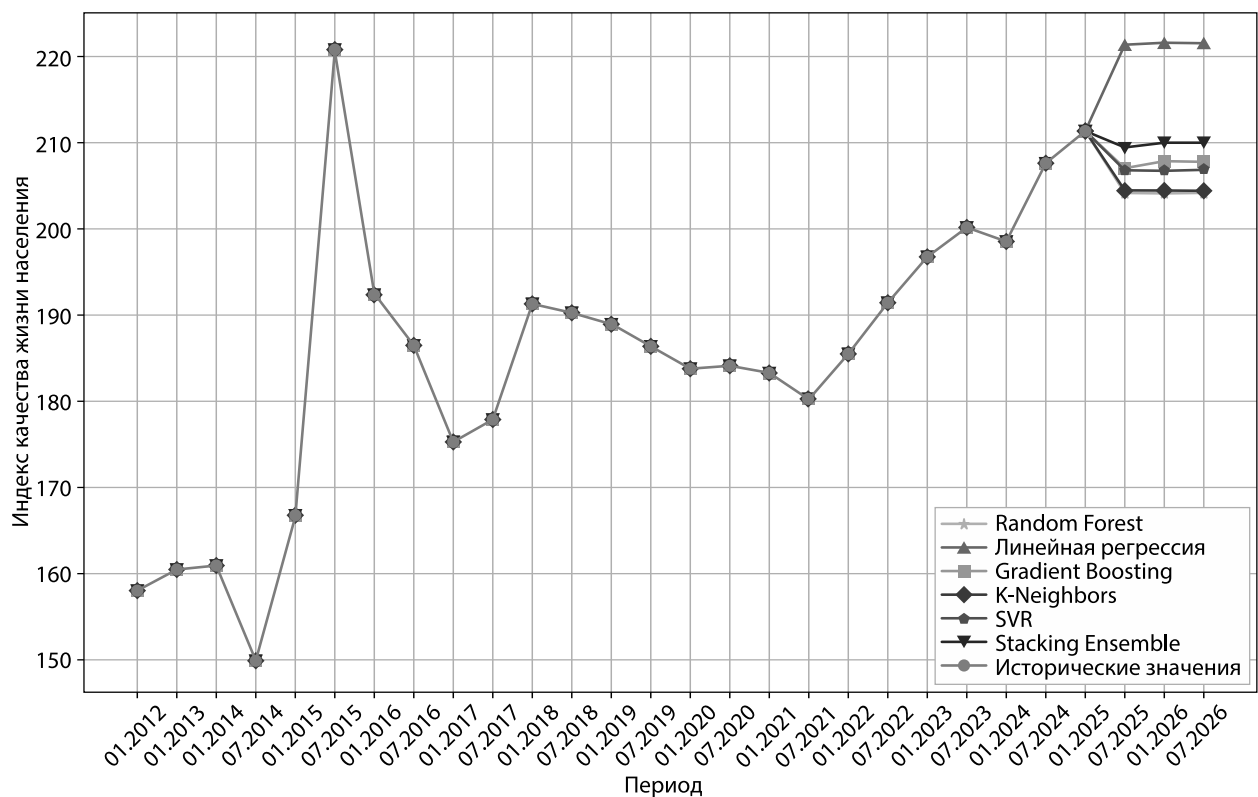


Рис. 8. Динамика изменения индекса качества жизни населения для Нидерландов с прогнозом ансамбля
Fig. 8. Dynamics of changes in the quality of life index for the Netherlands with the ensemble forecast

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

	GBR	KNN	LR	RF	SVR	Stack
GBR	1.0000	0.2973	0.1085	0.9999	0.7309	0.9939
KNN	0.2973	1.0000	0.0000	0.1899	0.9865	0.0798
LR	0.1085	0.0000	1.0000	0.1845	0.0006	0.3640
RF	0.9999	0.1899	0.1845	1.0000	0.5882	0.9994
SVR	0.7309	0.9865	0.0006	0.5882	1.0000	0.3596
Stack	0.9939	0.0798	0.3640	0.9994	0.3596	1.0000

Рис. 9. Матрица попарных сравнений p -значений (p -value)
Fig. 9. Matrix of pairwise comparisons of p -values

Источник: разработано авторами на основе данных веб-сайта Numbeo [4].

Наихудшие результаты продемонстрировала линейная регрессия, которая значительно уступает практически всем альтернативным моделям. Особенно показательным видится сравнение LR с KNN (p -value < 0.001) и SVR (p -value = 0.0006), при котором преимущество более сложных методов не вызывает сомнений. При этом SVR, несмотря на свою сложность, не смогла существенно превзойти KNN (p = 0.9865).

Приведенные результаты подтверждают эффективность алгоритмов на основе деревьев (GBR и RF) для задач данного типа, а также демонстрируют ограниченную применимость линейной регрессии и ставят под сомнение целесообразность использования ресурсоемкого алгоритма SVR. Особый интерес представляет анализ ансамблевой модели, которая, несмотря на отсутствие статистически значимого преимущества перед GBR и RF, имеет минимальное значение среднеквадратичной ошибки и наибольшее значение коэффициента детерминации.

В настоящей статье нами проведен комплексный сравнительный анализ потенциала применения моделей машинного обучения для прогнозирования индекса качества жизни населения на базе различных социально-экономических факторов. Основой исследования послужили данные о показателях, таких как уровень безопасности, доступность медицинских услуг, загрязнение окружающей среды, стоимость жизни и ряд других параметров. Рассмотрено и протестировано несколько моделей, вклю-

чая градиентный бустинг, случайный лес, метод ближайших соседей, линейную регрессию и метод опорных векторов.

Выводы

Таким образом, изложенный подход к прогнозированию индекса качества жизни населения с использованием ансамбля моделей машинного обучения может быть применен для анализа и предсказания изменений социально-экономического положения населения на базе исторических данных. Это дает возможность государственным и региональным органам управления принимать более обоснованные решения для улучшения условий жизни, повышения уровня благополучия населения.

Прогнозирование индекса качества жизни с учетом факторов, уникальных применительно к каждому региону, может быть полезным не только для России, но и для стран с различными экономическими и социальными условиями, в том числе США, государств Европейского союза, стран Азии и др. В будущем дополнение факторов макроэкономической стабильности, показателей социальной интеграции и развитие цифровых технологий расширит возможности моделей и улучшит их точность. Такие данные помогут учитывать в полной мере волатильность социально-экономических факторов и адаптировать модели для анализа быстро изменяющихся условий.

Список источников

1. Суворов А. В. Методологические проблемы прогнозирования уровня жизни населения // Проблемы прогнозирования. 2000. № 1. С. 22–38.
2. Деркаченко О. В. Кластеризация регионов России по качеству жизни населения и уровню цифровой экономики // Вопросы региональной экономики. 2020. № 2. С. 59–65. DOI: 10.21499/2078-4023-2020-43-2-59-65

3. Еникеева Л. А., Шишкикова М. С. Модели прогнозирования качества жизни на основе международных индексов // *Современные проблемы науки и образования*. 2015. № 1-1. С. 656–664.
4. Quality of life index by country // Numbeo. URL: https://www.numbeo.com/quality-of-life/rankings_by_country.jsp? (дата обращения: 01.02.2025).
5. Петрова Е. А. Управление качеством жизни населения региона: моделирование факторов внешней и внутренней среды // *Вестник Волгоградского государственного университета. Серия 3: Экономика. Экология*. 2018. Т. 20. № 1. С. 30–39. DOI: 10.15688/jvolsu3.2018.1.4
6. Lipesa B. A., Okango E., Omolo B. O., Omondi E. O. An application of a supervised machine learning model for predicting life expectancy // *SN Applied Sciences*. 2023. Vol. 5. No. 7. Article 189. DOI: 10.1007/s42452-023-05404-w
7. Щекотин Е. В., Гойко В. Л., Басина П. А., Бакулин В. В. Использование машинного обучения для изучения качества жизни населения: методологические аспекты // *Цифровая социология*. 2022. Т. 5. № 1. С. 87–97. DOI: 10.26425/2658-347X-2022-5-1-87-97
8. Чистяков С. П. Случайные леса: обзор // *Труды Карельского научного центра Российской академии наук*. 2013. № 1. С. 117–136.
9. Ротова О. М., Шибанова А. Д. Основные принципы метода линейной регрессии // *Теория и практика современной науки*. 2020. № 1. С. 483–487.
10. Дружков П. Н., Золотых Н. Ю., Половинкин А. Н. Реализация параллельного алгоритма предсказания в методе градиентного бустинга деревьев решений // *Вестник Южно-Уральского государственного университета. Серия: Математическое моделирование и программирование*. 2011. № 37. С. 82–89.
11. Пяткова А. А. Применение метода k-Nearest Neighbours к представлению гистограммных временных рядов // *Достижения вузовской науки*. 2014. № 8. С. 206–210.
12. Вэньлун И., Герасимов И. В., Кузьмин С. А., Хоцзяо Хэ, Хунюнь Ян. Применение метода опорных векторов для задачи восстановления регрессии в обработке данных листьев поливного риса // *Известия СПбГЭТУ ЛЭТИ*. 2016. № 2. С. 17–22.
13. SMOTE // Data Science. URL: <https://datascience.eu/ru/программирование/smote/> (дата обращения: 01.02.2025).
14. Machine Learning in Python // Scikit-learn. URL: <https://scikit-learn.org/> (дата обращения: 01.02.2025).
15. About pandas // Pandas. URL: <https://pandas.pydata.org/about/> (дата обращения: 01.02.2025).
16. NumPy user guide // NymPy. URL: <https://numpy.org/doc/stable/user/> (дата обращения: 01.02.2025).
17. Visualization with Python // Matplotlib. URL: <https://matplotlib.org/> (дата обращения: 01.02.2025).
18. About SciPy // SciPy. URL: <https://scipy.org/about/> (дата обращения: 01.02.2025).
19. Project Jupyter Documentation // Project Jupyter. URL: <https://docs.jupyter.org/> (дата обращения: 01.02.2025).
20. GridSearchCV // Scikit-learn developers. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (дата обращения: 01.02.2025).
21. Noble J. What are ARIMA models? // IBM. URL: <https://www.ibm.com/think/topics/arima-model> (дата обращения: 01.02.2025).
22. How can we perform the Nemenyi Post-Hoc Test in Python? // *Psychological Scales*. URL: <https://scales.arabpsychology.com/stats/how-can-we-perform-the-nemenyi-post-hoc-test-in-python/> (дата обращения: 01.02.2025).

References

1. Suvorov A.V. Forecasting the standard of living: Methodological problems. *Studies on Russian Economic Development*. 2000;11(1):1-10. (In Russ.: *Problemy prognozirovaniya*. 2000;(1):22-38.).
2. Derkachenko O.V. Clustering of regions of Russia by the quality of life of the population and the level of the digital economy. *Voprosy regional'noi ekonomiki = Problems of Regional Economy*. 2020;(2):59-65. (In Russ.). DOI: 10.21499/2078-4023-2020-43-2-59-65
3. Enikeeva L.A., Shirshikova M.S. Prognosis models of quality of life on the basis of international indexes. *Sovremennye problemy nauki i obrazovaniya = Modern Problems of Science and Education*. 2015;(1-1):656-664. (In Russ.).
4. Quality of life index by country. Numbeo. URL: https://www.numbeo.com/quality-of-life/rankings_by_country.jsp (accessed on 01.02.2025).
5. Petrova E.A. Living standards control in the region: Modeling of factors of external and internal environment. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 3: Ekonomika. Ekologiya = Science Journal of VolSU. Global Economic System*. 2018;20(1): 30-39. (In Russ.). DOI: 10.15688/jvolsu3.2018.1.4
6. Lipesa B.A., Okango E., Omolo B.O., Omondi E.O. An application of a supervised machine learning model for predicting life expectancy. *SN Applied Sciences*. 2023;5(7):189. DOI: 10.1007/s42452-023-05404-w

7. Shchekotin E.V., Goiko V.L., Basina P.A., Bakulin V.V. Using machine learning to study the population life quality: Methodological aspects. *Tsifrovaya sotsiologiya = Digital Sociology*. 2022;5(1):87-97. (In Russ.). DOI: 10.26425/2658-347X-2022-5-1-87-97
8. Chistiakov S.P. Random forests: An overview. *Trudy Karelskogo nauchnogo tsentra Rossiiskoi akademii nauk = Transactions of the Karelian Research Centre of the Russian Academy of Sciences*. 2013;(1):117-136. (In Russ.).
9. Rotova O.M., Shibanova A.D. Basic principles of the linear regression method. *Teoriya i praktika sovremennoi nauki = Theory and Practice of Modern Science*. 2020;(1):483-487. (In Russ.).
10. Druzhkov P.N., Zolotikh N.Yu., Polovinkin A.N. Parallel implementation of prediction algorithm in gradient boosting trees method. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Matematicheskoe modelirovanie i programmirovaniye = Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming & Computer Software*. 2011;(37):82-89. (In Russ.).
11. Pyatkova A.A. Application of k-Nearest Neighbours method to representation of histogram time series. *Dostizheniya vuzovskoi nauki*. 2014;(8):206-210. (In Russ.).
12. Yi W., Gerasimov I.V., Kuzmin S.A., He H., Yang H. Application of support vector regression in modeling and data processing of rice leaves. *Izvestiya SPbGETU LETI = Proceedings of Saint Petersburg Electrotechnical University*. 2016;(2):17-22. (In Russ.).
13. SMOTE. Data Science. URL: <https://datascience.eu/computer-programming/sMOTE/> (accessed on 01.02.2025).
14. Machine Learning in Python. Scikit-learn. URL: <https://scikit-learn.org/> (accessed on 01.02.2025).
15. About pandas. Pandas. URL: <https://pandas.pydata.org/about/> (accessed on 01.02.2025).
16. NumPy user guide. NymPy. URL: <https://numpy.org/doc/stable/user/> (accessed on 01.02.2025).
17. Visualization with Python. Matplotlib. URL: <https://matplotlib.org/> (accessed on 01.02.2025).
18. About SciPy. SciPy. <https://scipy.org/about/> (accessed on 01.02.2025).
19. Project Jupyter Documentation. Project Jupyter. URL: <https://docs.jupyter.org/> (accessed on 01.02.2025).
20. GridSearchCV. Scikit-learn developers. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (accessed on 01.02.2025).
21. Noble J. What are ARIMA models? IBM. URL: <https://www.ibm.com/think/topics/arima-model> (accessed on 01.02.2025).
22. How can we perform the Nemenyi Post-Hoc Test in Python? Psychological Scales. URL: <https://scales.arabpsychology.com/stats/how-can-we-perform-the-nemenyi-post-hoc-test-in-python/> (accessed on 01.02.2025).

Сведения об авторах

Хакимджон Иномджонович Аминов

доктор экономических наук, доцент,
доцент кафедры информационных систем
и технологий

Санкт-Петербургский государственный
экономический университет

191023, Санкт-Петербург,
наб. канала Грибоедова, д. 30–32а

Анастасия Владиславовна Кардаш

аналитик данных

Филиал АО «ДРТ»

199004, Санкт-Петербург, Средний пр. В. О.,
д. 38, корп. 1, лит. К

Поступила в редакцию 24.02.2025
Прошла рецензирование 12.03.2025
Подписана в печать 10.04.2025

Information about the authors

Khakimdzhon I. Aminov

D.Sc. in Economics, Associate Professor,
Associate Professor at the Department
of Information System and Technologies

St. Petersburg State University
of Economics

30–32 Griboedov Channel Emb., St. Petersburg
191023, Russia

Anastasiya V. Kardash

data analyst

Branch of JSC DRT

38/1 Sredny Ave V.I., St. Petersburg 199004,
Russia

Received 24.02.2025
Revised 12.03.2025
Accepted 10.04.2025

Конфликт интересов: авторы декларируют отсутствие конфликта интересов,
связанных с публикацией данной статьи.

Conflict of interest: the authors declare no conflict of interest
related to the publication of this article.