

Машинное обучение в комплексном исследовании индекса человеческого развития

Орлова Екатерина Андреевна

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Северо-Западный институт управления, Санкт-Петербург, Российская Федерация

студентка 4-го курса бакалавриата, факультет экономики и финансов

E-mail: eorlova-22@edu.ranepa.ru

Научный руководитель:

Полянская Светлана Владимировна

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Северо-Западный институт управления, Санкт-Петербург, Российская Федерация

E-mail: polyanskaya-sv@ranepa.ru

Аннотация

Введение: актуальной задачей региональной политики является не только мониторинг индекса человеческого развития, но и выявление ключевых социально-экономических факторов, определяющих его уровень. Понимание этих зависимостей необходимо для разработки научно обоснованных мер по сокращению межрегионального неравенства и повышению качества жизни.

Методы: исследование построено на анализе официальных статистических данных. На основе медианного значения целевой показатель был преобразован в бинарную переменную для задачи классификации. После обработки пропусков и генерации новых признаков было обучено и оптимизировано шесть алгоритмов машинного обучения, включая линейные и ансамблевые методы. Для интеграции их сильных сторон был дополнительно создан стекинг-ансамбль. Качество моделей оценивалось по комплексным метрикам на тестовой выборке.

Результаты: корреляционный анализ выявил сильную отрицательную связь индекса с уровнем бедности и безработицы и сильную положительную связь с показателями демографической нагрузки и экономической активности. Среди моделей наилучшую способность к прогнозированию продемонстрировал градиентный бустинг, достигший точности 0,8824 и значения ROC-AUC 0,9523. Стекинг-ансамбль позволил получить сбалансированный классификатор с конкурентными показателями.

Обсуждение: результаты подтверждают эффективность современных ансамблевых методов для моделирования комплексных социально-экономических зависимостей. Построенные модели служат инструментом для аналитической поддержки управленческих решений, позволяя оценивать потенциальное влияние изменений в отдельных сферах на интегральный показатель развития региона. Перспективой работы является углубление интерпретации моделей для выработки конкретных рекомендаций.

Ключевые слова: индекс человеческого развития; машинное обучение; градиентный бустинг; социально-экономические показатели; регионы России; классификация, стекинг-ансамбль.

Для цитирования: Орлова Е. А. Машинное обучение в комплексном исследовании индекса человеческого развития // Новизна. Эксперимент. Традиции (Н.Экс.Т). – 2026. – Т. 12. – № 2 (34). – С. 66–77.

Machine Learning in the Comprehensive Study of the Human Development Index

Ekaterina A. Orlova

Russian Presidential Academy of National Economy and Public Administration, North-West Institute of Management (Saint Petersburg, Russian Federation)

BA student, Faculty of Economics and Finance

E-mail: eorlova-22@edu.ranepa.ru

Academic Supervisor:

Svetlana V. Polyanskaya

Russian Presidential Academy of National Economy and Public Administration, North-West Institute of Management (Saint Petersburg, Russian Federation)

E-mail: polyanskaya-sv@ranepa.ru

Abstract

Introduction: An urgent task of regional policy consists not only in monitoring the human development index, but also in identifying key socio-economic factors determining its level. Realizing these dependencies necessitates the development of evidence-based measures to reduce inter-regional inequality and improve the quality of life.

Methods: The study is based on the analysis of official statistical data. The target indicator was transformed into a binary variable for classification tasks based on the median value. After processing the gaps and generating new features, six machine learning algorithms, including linear and ensemble methods, were trained and optimized. A stacking ensemble was additionally developed to integrate their strengths. The quality of the models was assessed using complex metrics in a test sample.

Results: Correlation analysis has revealed a strong negative relationship between the index and the level of poverty and unemployment, and a strong positive relationship with the indicators of demographic burden and economic activity. Among the models, gradient boosting demonstrated the best predictive ability, reaching an accuracy of 0.8824 and a ROC-AUC value of 0.9523. The stacking ensemble has enabled us to obtain a balanced classifier with competitive indicators.

Discussion: The results have confirmed the effectiveness of modern ensemble methods for modeling complex socio-economic dependencies. The constructed models serve as a tool for analytical support of management decisions, providing the opportunity to assess the potential impact of changes in individual areas on the integral indicator of the region's development. The future work will focus on providing a more profound interpretation of the models to develop specific recommendations.

Keywords: human development index; machine learning; gradient boosting; socio-economic indicators; regions of Russia; classification, stacking ensemble.

For citation: Orlova, E. A. (2026) Machine Learning in a Comprehensive Human Development Index Study. *Novelty. Experiment. Traditions (N.Ex.T)*. Vol. 12, no. 2 (34). Pp. 66–77. (In Russ.)

ВВЕДЕНИЕ

Индекс человеческого развития (ИЧР) представляет собой интегральный показатель, характеризующий уровень благосостояния и качество жизни населения на основе ожидаемой продолжительности жизни, уровня образования и дохода на душу населе-

ния^{1, 2, 3}. В условиях нарастающей дифференциации социально-экономического развития субъектов Российской Федерации актуальной задачей становится не только мониторинг значений ИЧР⁴, но и выявление определяющих его факторов, что необходимо для формирования научно обоснованных стратегий региональной политики и сокращения неравенства.

Цель исследования — построение и сравнительный анализ моделей машинного обучения для прогнозирования уровня человеческого развития на основе социально-экономических показателей российских регионов. В рамках работы проведены сбор и предварительный анализ данных по 39 признакам для 85 субъектов РФ, обработка пропущенных значений, генерация новых признаков и бинарное разделение регионов на группы с высоким и низким ИЧР. Основное внимание уделено обучению и настройке гиперпараметров алгоритмов классификации (логистическая регрессия, случайный лес, градиентный бустинг, CatBoost), а также построению стекинг-ансамбля для повышения качества прогнозирования.

Научная новизна работы заключается в применении комплексного подхода к моделированию ИЧР с использованием современных методов машинного обучения и инженерии признаков. Практическая значимость полученных моделей состоит в возможности их использования для поддержки управленческих решений и оценки влияния социально-экономических факторов на интегральный показатель регионального развития.

МЕТОДЫ

Исследование проводилось на основе данных официальной статистики (с сайта <https://rosstat.gov.ru>) по 85 субъектам РФ. Исходный набор данных содержал 255 наблюдений с 40 признаками, включая целевой показатель — индекс человеческого развития. Целевая переменная была преобразована в бинарный признак для задачи классификации⁵. Порог разделения был выбран как медианное значение ИЧР по всем регионам (0,840). Регионы со значением ИЧР, равным или превышающим медиану, были отнесены к классу «высокий ИЧР» (131 наблюдение), остальные — к классу «низкий ИЧР» (124 наблюдения).

Предварительный анализ данных выявил наличие пропущенных значений в нескольких признаках: «Число зарегистрированных преступлений на 100 тыс. жителей»

¹ Буценко И. Н. Индекс человеческого развития — инструмент сравнения глобальных тенденций развития человечества // Актуальные проблемы и перспективы развития экономики : Труды XXIII Международной научно-практической конференции. Симферополь – Гурзуф, 17–19 октября 2024 года. Симферополь: ИП Зуева Т. В., 2024. С. 72–74. EDN: НОСТКУ

² Кузнецова М. В., Ивашина Н. С. Социально-экономическое развитие через призму индексов человеческого развития и развития человеческого капитала // Устойчивое развитие территорий: теория и практика : Материалы X Всероссийской научно-практической конференции с международным участием. В 2-х томах. Сибай, 14–16 ноября 2019 года. Т. 2. Сибай: Сибайский информационный центр-филиал ГУП РБ ИД «Республика Башкортостан», 2019. С. 357–359. EDN: EUUTNA

³ Каменева А. И., Карпухно И. А. Зависимость индекса человеческого развития от роста валового внутреннего продукта на примере Норвегии // Экономическая теория в условиях глобализации экономики : Тезисы докладов и выступлений XI Международной научно-практической конференции студентов и молодых ученых. Донецк, 13–14 марта 2019 года / Под общ. ред. Л. И. Дмитриченко. Донецк: Донецкий национальный университет, 2019. С. 172–173. EDN: NGLVAN

⁴ Камалдинова И. М., Мухаметшина Г. Р. Анализ индекса человеческого развития России и его роль в темпах экономического развития страны // Вестник Ростовского государственного экономического университета (РИНХ). 2023. Т. 28, № 2. С. 106–114. DOI: 10.54220/v.rsue.1991-0533.2023.2.28.013. EDN: AYMPQE

⁵ Михайлова С. С., Гринева Н. В. Разработка модели бинарной классификации на малых данных с использованием методов машинного обучения // Проблемы экономики и юридической практики. 2024. Т. 20, № 1. С. 129–140. DOI: 10.33693/2541-8025-2024-20-1-129-140. EDN: WFJKOK

(10 пропусков), «Численность студентов высшего образования на 10 000 человек населения» (3 пропуска) и «Численность зрителей театров на 1000 человек» (9 пропусков). Для сохранения объема выборки пропущенные значения были заполнены медианными значениями соответствующих признаков. Для выявления первичных взаимосвязей был проведен корреляционный анализ между ИЧР и всеми независимыми переменными. По абсолютному значению коэффициента корреляции Пирсона был составлен рейтинг топ-10 наиболее влиятельных признаков. Наиболее сильные отрицательные связи были выявлены с уровнем бедности и уровнем безработицы, а положительные — с численностью занятых, приходящихся на одного пенсионера, и оборотом розничной торговли на душу населения.

Для повышения предсказательной силы моделей была применена инженерия признаков. На основе исходных переменных, показавших высокую корреляцию с целевым показателем, были созданы новые признаки. Методами генерации выступили попарное умножение (взаимодействие) и деление наиболее информативных показателей, таких как численность населения, оборот розничной торговли и объем транспортных услуг. В результате было создано 20 новых синтетических признаков, а общее количество признаков для моделирования увеличилось до 52.

Далее данные были стандартизированы для приведения всех признаков к единому масштабу, что является важным этапом для многих алгоритмов машинного обучения, в частности для логистической регрессии и методов, основанных на расчете расстояний. Выборка была разделена на обучающую и тестовую части в соотношении 80 % к 20 % (204 и 51 наблюдение соответственно).

Для построения моделей зависимости был использован набор базовых алгоритмов машинного обучения, репрезентативных для разных подходов. Логистическая регрессия (Logistic Regression) была выбрана в качестве базовой линейной модели. Ее принцип работы основан на оценке вероятности принадлежности к классу путем построения линейной комбинации признаков. Модель интерпретируема, а сила влияния каждого признака выражается через коэффициенты. Однако она предполагает линейную зависимость между признаками и логарифмом шансов, что может быть ограничением для сложных данных.

Случайный лес (Random Forest) представляет собой ансамбль решающих деревьев. Его ключевая идея — построение множества некоррелированных деревьев на различных подвыборках данных с использованием случайного подмножества признаков в каждом узле. Итоговый прогноз формируется путем голосования всех деревьев. Этот подход эффективно снижает переобучение, характерное для отдельных деревьев, и повышает устойчивость модели.

Экстра-деревья (Extra Trees) — это вариант ансамбля деревьев, который вносит дополнительную рандомизацию в процесс обучения. В отличие от случайного леса, где в узле выбирается лучшее разделение из случайного подмножества признаков, в экстра-деревьях разделение выбирается совершенно случайно. Это делает обучение быстрее.

Адаптивный бустинг⁶ (AdaBoost) — один из первых практических алгоритмов бустинга. На каждой итерации он увеличивает веса неверно классифицированных объектов, заставляя последующую базовую модель (чаще всего «пни» — деревья глубины 1) концентрироваться на сложных случаях. Итоговый прогноз — взвешенное голосование всех моделей. Однако алгоритм может быть чувствителен к шуму в данных.

⁶ *Рукомин М. А.* Обзор ансамблевых моделей предиктивной аналитики и их сравнение с традиционными ML-подходами // Вестник науки. 2025. Т. 1, № 8 (89). С. 368–373. EDN: GJBKYS

Градиентный бустинг (Gradient Boosting) — это мощный ансамблевый метод, основанный на принципе бустинга. Он последовательно строит множество слабых моделей (обычно неглубоких деревьев), где каждая новая модель обучается фокусироваться на ошибках, допущенных предыдущими. Цель — минимизировать функцию потерь с помощью градиентного спуска. Этот метод особенно эффективен для выявления сложных нелинейных зависимостей и взаимодействий признаков.

Также использован CatBoost — реализация градиентного бустинга от Yandex, использующая упорядоченное бустинговое дерево и специальные методы обработки категориальных переменных, что часто приводит к высокой точности «из коробки» без обширной предобработки.

Для каждого алгоритма были определены гиперпараметры для оптимизации^{7,8}. В качестве основной метрики для оптимизации использовалась доля правильных ответов (ДПО — accuracy). После выбора лучшей конфигурации модель заново обучалась на всей обучающей выборке и оценивалась на независимой тестовой выборке.

Качество окончательных моделей оценивалось по набору метрик, включающему accuracy, precision (точность), recall (полнота), F1-мера (гармоническое среднее точности и полноты) и ROC-AUC (площадь под ROC-кривой, характеризующая способность модели к разделению классов). Для интеграции сильных сторон отдельных алгоритмов был применен метод построения обобщенной модели на основе данных, полученных на базовых моделях (метод стекинга — Stacking). В качестве базовых моделей были выбраны три алгоритма, показавшие наилучшие результаты по accuracy на тестовой выборке: экстра-деревья (Extra Trees), градиентный бустинг (Gradient Boosting⁹) и CatBoost¹⁰. Все этапы анализа, предобработки данных, обучения и валидации моделей были реализованы с использованием языка программирования Python.

РЕЗУЛЬТАТЫ

Результаты предварительного анализа данных показали, что индекс человеческого развития в исследуемой выборке регионов России варьируется в диапазоне от 0,771 до 0,940, со средним значением 0,842 и медианой 0,840. Распределение регионов по бинарной целевой переменной, основанной на медианном пороге, оказалось сбалансированным (рис. 1).

Анализ корреляционной матрицы выявил наиболее сильные линейные связи ИЧР с другими показателями. Наибольшую отрицательную корреляцию продемонстрировали уровень бедности (-0,686) и уровень безработицы населения (-0,627). Наиболее сильные положительные связи наблюдались с показателем численности занятых, приходящихся на одного пенсионера (0,586), и оборотом розничной торговли на душу населения (0,514). Топ-10 значимых признаков представлен в таблице 1.

⁷ Астапов Р. Л., Мухаммадеева Р. М. Автоматизация подбора параметров машинного обучения и обучение модели машинного обучения // Актуальные научные исследования в современном мире. 2021. № 5-2 (73). С. 34–37. EDN: GJEUNW

⁸ Колосов В. С. Возможности оптимизации гиперпараметров для разработки моделей градиентного бустинга на примере открытого набора данных // Медицина будущего 2025 : Материалы Всероссийского научного форума с международным участием. Тюмень, 27–29 марта 2025 года. Тюмень: РИЦ «Айвекс», 2025. С. 149–150. EDN: FSBJOT

⁹ Макаров С. М. Прогнозирование уровня отклонений в производственном процессе с помощью регрессии на основе градиентного бустинга // Парадигма. 2025. № 5-2. С. 137–144. EDN: IFPBFR

¹⁰ Шуткина Е. В., Курашкин С. О. Применение метода CATBOOST для задачи прогнозирования СХД // Актуальные проблемы авиации и космонавтики : сборник материалов X Международной научно-практической конференции, посвященной 100-летию академика М. Ф. Решетнева и Дню космонавтики: в 3-х томах. Красноярск, 08–12 апреля 2024 года. Красноярск: Сибирский государственный университет науки и технологий им. акад. М. Ф. Решетнева, 2024. С. 251–253. EDN: LDGCVK

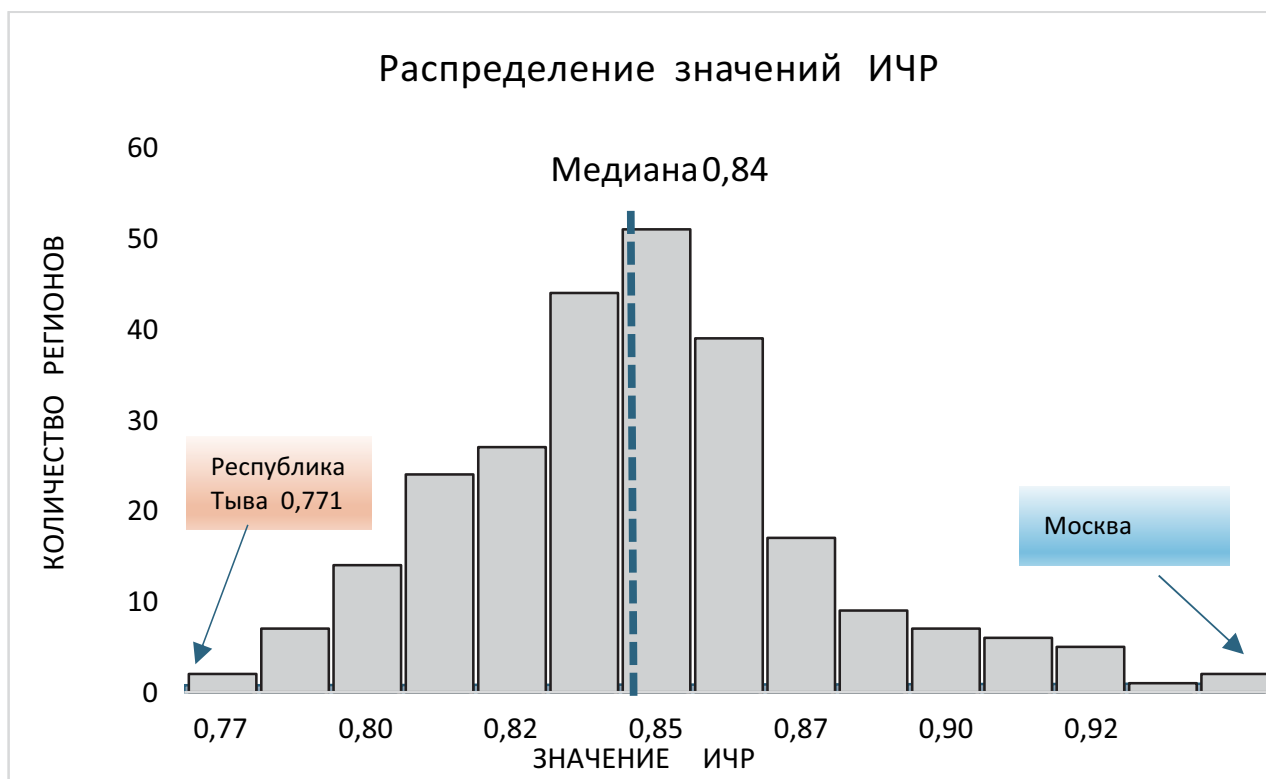


Рис. 1. Распределение ИЧР по регионам (составлено автором по данным с сайта <https://rosstat.gov.ru>)

Таблица 1

Топ-10 признаков по абсолютной корреляции с ИЧР

№	Показатель	Положительная/отрицательная	Значение
1	Уровень бедности, в %	отрицательная	-0,686 (сильная)
2	Уровень безработицы, в %	отрицательная	-0,627 (сильная)
3	Численность занятых на одного пенсионера, чел.	положительная	0,586 (сильная)
4	Оборот розничной торговли на душу населения	положительная	0,514 (сильная)
5	Объем транспортных услуг на душу населения	положительная	0,421 (средняя)
6	Уровень инновационной активности организаций, в %	положительная	0,378 (средняя)
7	Удельный вес организаций, осуществлявших технологические инновации, в %	положительная	0,340 (средняя)
8	Численность студентов высшего образования на 10 000 чел. населения	положительная	0,327 (средняя)
9	Число посещений музеев на 1000 чел. населения	положительная	0,309 (средняя)
10	Объем бытовых услуг на душу населения	положительная	0,292 (слабая)

Источник: составлено автором по данным с сайта <https://rosstat.gov.ru>.

Оценка качества моделей на тестовой выборке показала следующие результаты. По точности наивысший и идентичный результат (0,8824) продемонстрировали три модели: экстра-деревья (Extra Trees), градиентный бустинг (Gradient Boosting) и CatBoost (рис. 2).

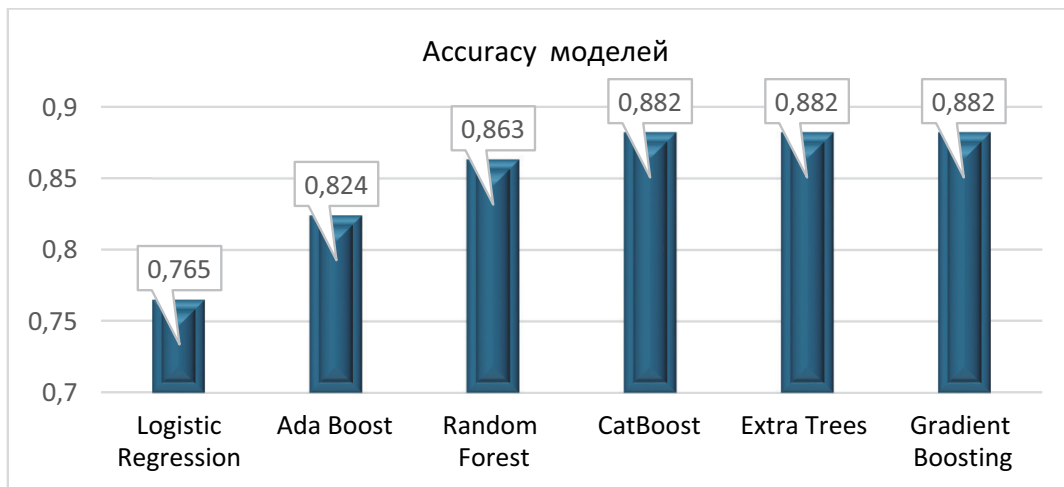


Рис. 2. Точность (Accuracy) моделей на тестовой выборке
(составлено автором по данным с сайта <https://rosstat.gov.ru>)

Случайный лес показал ассурасу 0,8627, адаптивный бустинг — 0,8235, а логистическая регрессия — 0,7647. Анализ метрики ROC-AUC, более устойчивой к дисбалансу классов, выявил иное ранжирование (рис. 3).

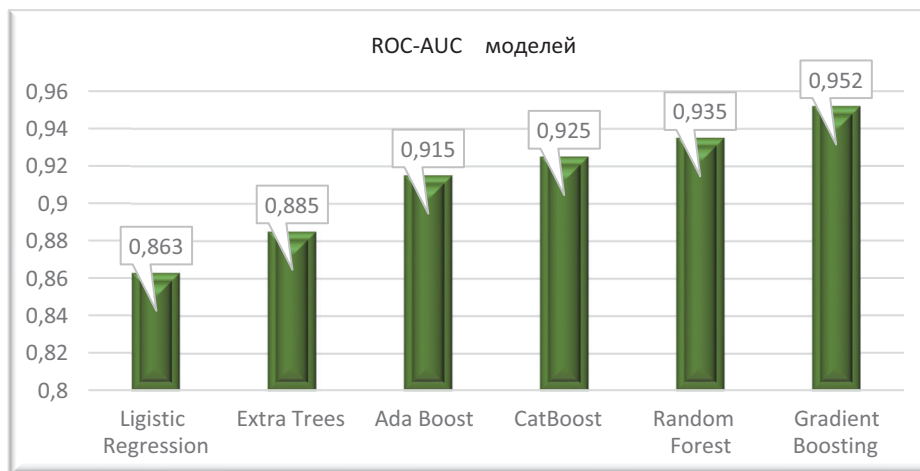


Рис. 3. ROC-AUC моделей на тестовой выборке
(составлено автором по данным с сайта <https://rosstat.gov.ru>)

Наилучшей способностью к разделению классов обладала модель градиентного бустинга с ROC-AUC 0,9523. За ней следовали случайный лес (0,9354) и CatBoost (0,9246). Модель экстра-деревьев, несмотря на высокую точность (ассурасу), показала относительно низкое значение ROC-AUC (0,8846), что может указывать на менее уверенное разделение классов. Логистическая регрессия показала ROC-AUC 0,8631.

Сравнение метрик на обучающей и тестовой выборках выявило признаки переобучения для ряда ансамблевых моделей. Алгоритмы экстра-деревьев, адаптивного бустинга, градиентного бустинга и CatBoost достигли на тренировочных данных ассурасу и F1-меры, равных 1,0, что указывает на идеальную подгонку. Однако их результаты на тестовой выборке были существенно ниже, особенно это было заметно для экстра-деревьев, у которых разница между тренировочным и тестовым ROC-AUC составила более 0,115 пункта. Наименьший разрыв между качеством на обучении и тесте наблюдался у логистической

регрессии и случайного леса, что свидетельствует о лучшей обобщающей способности этих моделей в рамках данного эксперимента. Сводные данные метрик для всех моделей приведены в таблице 2.

Таблица 2

Метрики для всех моделей

Модель	Выборка	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	Обучающая	0,9559	0,9615	0,9524	0,9569	0,9910
Logistic Regression	Тестовая	0,7647	0,8182	0,6923	0,7500	0,8631
Random Forest	Обучающая	0,9902	0,9813	1,0000	0,9906	0,9999
Random Forest	Тестовая	0,8627	0,8800	0,8462	0,8627	0,9354
Extra Trees	Обучающая	1,0000	1,0000	1,0000	1,0000	1,0000
Extra Trees	Тестовая	0,8824	0,8571	0,9231	0,8889	0,8846
AdaBoost	Обучающая	1,0000	1,0000	1,0000	1,0000	1,0000
AdaBoost	Тестовая	0,8235	0,8696	0,7692	0,8163	0,9154
Gradient Boosting	Обучающая	1,0000	1,0000	1,0000	1,0000	1,0000
Gradient Boosting	Тестовая	0,8824	0,8846	0,8846	0,8846	0,9523
CatBoost	Обучающая	1,0000	1,0000	1,0000	1,0000	1,0000
CatBoost	Тестовая	0,8824	0,8846	0,8846	0,8846	0,9246

Источник: составлено автором по данным с сайта <https://rosstat.gov.ru>.

Построение стекинг-ансамбля на основе трех лучших по ассурасу моделей (экстра-деревья, градиентный бустинг, CatBoost) позволило получить финальную модель. Ее точность (ассурасу) на тестовой выборке составила 0,8824, что соответствует результатам лучших базовых моделей. При этом значение ROC-AUC (0,9308) превысило соответствующий показатель модели экстра-деревьев (0,8846) на 0,0462, а также превзошло результат CatBoost (0,9246), хотя и не достигло значения градиентного бустинга (0,9523). Таким образом, построение ансамбля позволило получить модель, которая по комплексной метрике ROC-AUC улучшила результат одного из базовых алгоритмов, сохранив при этом максимально достижимую в эксперименте точность классификации (ассурасу).

ОБСУЖДЕНИЕ

Результаты исследования подтверждают, что современные ансамблевые методы машинного обучения, особенно градиентный бустинг, способны с высокой точностью прогнозировать уровень человеческого развития в регионах России. Выявленные различия в эффективности алгоритмов обусловлены их архитектурными особенностями и способностью улавливать нелинейные зависимости в социально-экономических данных.

Наилучшие результаты градиентного бустинга (ROC-AUC = 0,9523, ассурасу = 0,8824) объясняются его ключевым принципом работы: последовательным построением ансамбля слабых моделей, каждая из которых корректирует ошибки предыдущих. Такая итеративная минимизация ошибки позволяет эффективно выявлять сложные нелинейные взаимосвязи, характерные для региональных социально-экономических систем.

Низкая эффективность логистической регрессии (ассурасу = 0,7647, ROC-AUC = 0,8631) закономерна: как линейный классификатор она неспособна адекватно отразить нелинейный характер зависимостей между ИЧР и комплексом социально-экономических факторов, что подтверждается успехом нелинейных методов.

Показательно сравнение случайного леса и экстра-деревьев. При сопоставимой точности (0,8627 и 0,8824 соответственно) экстра-деревья продемонстрировали значительно более низкий ROC-AUC (0,8846 против 0,9354), что указывает на «неуверенность» прогнозов с вероятностями, близкими к пороговым значениям. Это следствие повышенной рандомизации при построении деревьев, приводящей к менее калиброванным вероятностным оценкам. Случайный лес показал наименьший разрыв между обучающими и тестовыми метриками, подтверждая лучшую обобщающую способность.

Результаты CatBoost и адаптивного бустинга согласуются с общей эффективностью бустинговых алгоритмов. CatBoost незначительно уступил градиентному бустингу по ROC-AUC при равной точности, тогда как адаптивный бустинг показал более скромные результаты, что может объясняться его чувствительностью к шумам в данных.

Для сравнения моделей график ROC-кривых приведен на рисунке 4.

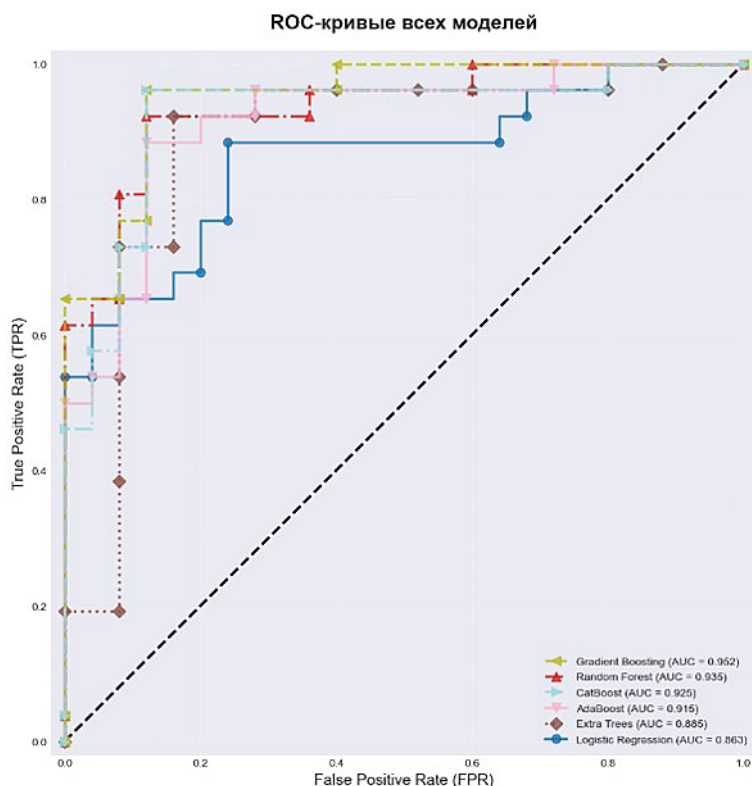


Рис. 4. ROC-кривые моделей на тестовой выборке (составлено автором по данным с сайта <https://rosstat.gov.ru>)

Успех стекинг-ансамбля, повысившего ROC-AUC по сравнению с экстра-деревьями на 0,0462, подтверждает перспективность комбинирования разнородных алгоритмов.

Основным ограничением работы выступает размер выборки (255 наблюдений при 52 признаках), что повышает риск переобучения сложных моделей. Перспективным направлением дальнейших исследований является применение методов интерпретируемого искусственного интеллекта (SHAP, LIME) для оценки вклада отдельных факторов в прогноз, что имеет высокую практическую ценность для органов управления.

ВЫВОД

Проведенное исследование подтвердило эффективность применения методов машинного обучения для построения прогностических моделей уровня человеческого развития на основе социально-экономических показателей регионов РФ. Анализ данных по

85 субъектам выявил, что ИЧР наиболее тесно связан с уровнем бедности, уровнем безработицы, демографической нагрузкой (численность занятых на одного пенсионера) и экономической активностью населения (оборот розничной торговли).

Сравнительный анализ шести алгоритмов классификации показал превосходство нелинейных ансамблевых методов над линейными подходами. Наилучшие результаты продемонстрировал градиентный бустинг (ROC-AUC = 0,9523, accuracy = 0,8824), что подтверждает существенно нелинейный характер зависимости ИЧР от комплекса социально-экономических факторов. Построение стекинг-ансамбля на основе лучших моделей позволило получить сбалансированный классификатор, сочетающий сильные стороны различных алгоритмов.

Практическая значимость работы заключается в возможности использования разработанных моделей для мониторинга и прогнозирования регионального развития, а также для оценки влияния управленческих решений на интегральный показатель качества жизни населения. Перспективными направлениями дальнейших исследований являются расширение временного горизонта анализа, применение методов интерпретируемого искусственного интеллекта и включение дополнительных социально-экономических индикаторов.

Материал выверен, цифры, факты, цитаты сверены с первоисточником, материал не содержит сведений ограниченного распространения. Научное направление статьи — экономика и управление народным хозяйством (региональная экономика).

Список источников

1. Астапов Р. Л. Автоматизация подбора параметров машинного обучения и обучение модели машинного обучения / Р. Л. Астапов, Р. М. Мухамадеева // Актуальные научные исследования в современном мире. 2021. № 5-2 (73). С. 34–37. EDN: GJEUNW
2. Буценко И. Н. Индекс человеческого развития — инструмент сравнения глобальных тенденций развития человечества / И. Н. Буценко // Актуальные проблемы и перспективы развития экономики : Труды XXIII Международной научно-практической конференции, Симферополь – Гурзуф, 17–19 октября 2024 года. Симферополь: ИП Зуева Т. В., 2024. С. 72–74. EDN: НОСТКУ
3. Камалдинова И. М. Анализ индекса человеческого развития России и его роль в темпах экономического развития страны / И. М. Камалдинова, Г. Р. Мухаметшина // Вестник Ростовского государственного экономического университета (РИНХ). 2023. Т. 28, № 2. С. 106–114. DOI: 10.54220/v.rsue.1991-0533.2023.2.28.013. EDN: AYMPQE
4. Каменева А. И. Зависимость индекса человеческого развития от роста валового внутреннего продукта на примере Норвегии / А. И. Каменева, И. А. Карпухно // Экономическая теория в условиях глобализации экономики : Тезисы докладов и выступлений XI Международной научно-практической конференции студентов и молодых ученых. Донецк, 13–14 марта 2019 года / Под общ. ред. Л. И. Дмитриченко. Донецк: Донецкий национальный университет, 2019. С. 172–173. EDN: NGLVAN
5. Колосов В. С. Возможности оптимизации гиперпараметров для разработки моделей градиентного бустинга на примере открытого набора данных / В. С. Колосов // Медицина будущего 2025 : Материалы Всероссийского научного форума с международным участием. Тюмень, 27–29 марта 2025 года. Тюмень: РИЦ «Айвекс», 2025. С. 149–150. EDN: FSBJOT
6. Кузнецова М. В. Социально-экономическое развитие через призму индексов человеческого развития и развития человеческого капитала / М. В. Кузнецова, Н. С. Ивашина // Устойчивое развитие территорий: теория и практика : Материалы X Всероссийской научно-практической конференции с международным участием. В 2-х тт. Сибай, 14–16

ноября 2019 года. Т. 2. Сибай: Сибайский информационный центр-филиал ГУП РБ ИД «Республика Башкортостан», 2019. С. 357–359. EDN: EUUTNA

7. Макаров С. М. Прогнозирование уровня отклонений в производственном процессе с помощью регрессии на основе градиентного бустинга / С. М. Макаров // Парадигма. 2025. № 5-2. С. 137–144. EDN: IFPBFR

8. Михайлова С. С. Разработка модели бинарной классификации на малых данных с использованием методов машинного обучения / С. С. Михайлова, Н. В. Гринева // Проблемы экономики и юридической практики. 2024. Т. 20, № 1. С. 129–140. DOI: 10.33693/2541-8025-2024-20-1-129-140. EDN: WFJKOK

9. Рукомин М. А. Обзор ансамблевых моделей предиктивной аналитики и их сравнение с традиционными ML-подходами / М. А. Рукомин // Вестник науки. 2025. Т. 1, № 8 (89). С. 368–373. EDN: GJBKYS

10. Шуткина Е. В. Применение метода CATBOOST для задачи прогнозирования СХД / Е. В. Шуткина, С. О. Курашкин // Актуальные проблемы авиации и космонавтики : сборник материалов X Международной научно-практической конференции, посвященной 100-летию академика М. Ф. Решетнева и Дню космонавтики: в 3-х тт. Красноярск, 08–12 апреля 2024 года. Красноярск: Сибирский государственный университет науки и технологий им. акад. М. Ф. Решетнева, 2024. С. 251–253. EDN: LDGCVK

Сведения об авторе:

Орлова Екатерина Андреевна, студентка 4-го курса бакалавриата, факультет экономики и финансов, Северо-Западный институт управления, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации (Санкт-Петербург, Российская Федерация); e-mail: eorlova-22@edu.ranepa.ru

Научный руководитель: Полянская Светлана Владимировна, кандидат технических наук, доцент кафедры бизнес-информатики, Северо-Западный институт управления, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации (Санкт-Петербург, Российская Федерация); e-mail: polyanskaya-sv@ranepa.ru

References

1. Astapov, R. L., Mukhamadeeva, R. M. (2021) Selection's Automatization of Machine Learning Parameters and Training a Machine Learning Model. *Current scientific research in the modern world*. No. 5-2 (73). P. 34–37. EDN: GJEUNW (In Russ.)

2. Butsenko, I. N. (2024) Human Development Index — a tool for comparing global trends in human development. *Current problems and prospects for economic development: Proceedings of the XXIII International Scientific and Practical Conference, Simferopol – Gurzuf, October 17–19, 2024*. Simferopol: IP Zueva T. V. Pp. 72–74. EDN: HOCTKY (In Russ.)

3. Kamaldinova, I. M., Mukhametshina, G. R. (2023) Analysis of Russian Human Development Index and Its Role in The Pace of Economic Development of Country. *Bulletin of the Rostov State Economic University (RINH)*. Vol. 28, no. 2. Pp. 106–114. DOI: 10.54220/v.rsue.1991-0533.2023.2.28.013. EDN: AYMPQE (In Russ.)

4. Kameneva, A. I., Karpukhno, I. A. (2019) Dependence of the Human Development Index on the Growth of Gross Domestic Product: The Case of Norway. *Economic Theory in the Context of Economic Globalization: Abstracts of Reports and Presentations at the XI International Scientific and Practical Conference of Students and Young Scientists*. Donetsk, March

- 13–14, 2019 / General Editor L. I. Dmitrichenko. Donetsk: *Donetsk National University*. Pp. 172–173. EDN: NGLVAN (In Russ.)
5. Kolosov, V. S. (2025) Possibilities of Hyperparameter Optimization for Developing Gradient Boosting Models: The Case of an Open Dataset. *Medicine of the Future 2025: Proceedings of the All-Russian Scientific Forum with International Participation*. Tyumen, March 27–29, 2025. Tyumen: *RIC "Aivex"*. Pp. 149–150. EDN: FSBJOT (In Russ.)
6. Kuznetsova, M. V., Ivashina, N. S. (2019) Socio-economic development through the prism of human development indices and human capital development. Sustainable development of territories: theory and practice: Proceedings of the 10th All-Russian scientific and practical conference with international participation. In 2 vol. Sibay, November 14–16, 2019. Vol. 2. Sibay: *Sibay Information Center-branch of the State Unitary Enterprise of the Republic of Bashkortostan Publishing House*. Pp. 357–359. EDN: EUUTNA (In Russ.)
7. Makarov, S. M. (2025) Forecasting the Level of Deviations in the Production Process Using Regression Based on Gradient Boosting. *Paradigm*. No. 5-2. Pp. 137–144. EDN: IFPBFR (In Russ.)
8. Mikhailova, S. S., Grineva, N. V. (2024) Development of a Binary Classification Model Based on Small Data Using Machine Learning Methods. *Economics Problems and Legal Practice*. Vol. 20, no. 1. Pp. 129–140. DOI: 10.33693/2541-8025-2024-20-1-129-140. EDN: WFJKOK (In Russ.)
9. Rukomin, M. A. (2025) Review of Ensemble Models in Predictive Analytics and Their Comparison with Traditional Machine Learning Approaches. *Vestnik nauki*. Vol. 1, no. 8 (89). Pp. 368–373. EDN: GJBKYS (In Russ.)
10. Shutkina, E. V., Kurashkin, S. O. (2024) Application of the CATBOOST Method to the Problem of Predicting CCHD Prediction. Actual Problems of Aviation and Cosmonautics: Collection of Materials of the X International Scientific and Practical Conference Dedicated to the 100th Anniversary of Academician M. F. Reshetnev and Cosmonautics Day: in 3 vol. Krasnoyarsk, April 8–12, 2024. Krasnoyarsk: *Siberian State University of Science and Technology named after academician M. F. Reshetnev*. Pp. 251–253. EDN: LDGCVK (In Russ.)

About the Author:

Ekaterina A. Orlova, BA student, Faculty of Economics and Finance, North-West Institute of Management, Russian Presidential Academy of National Economy and Public Administration, (Saint Petersburg, Russian Federation); e-mail: eorlova-22@edu.ranepa.ru

Academic Supervisor: Svetlana V. Polyanskaya, PhD of Technical Sciences, Associate Professor of the Department of Business Informatics, North-West Institute of Management, Russian Presidential Academy of National Economy and Public Administration, (Saint Petersburg, Russian Federation); e-mail: polyanskaya-sv@ranepa.ru