

УДК 004.852

DOI: [10.26102/2310-6018/2026.52.1.005](https://doi.org/10.26102/2310-6018/2026.52.1.005)

## Модель и метод оценки весовых коэффициентов ансамбля машинного обучения в задаче прогнозирования провозной платы железнодорожных грузовых перевозок

К.А. Бухарова 

*Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург, Российская Федерация*

**Резюме.** Настоящая статья посвящена вопросу оценки эффективности разработанной ансамблевой модели машинного обучения, применяемой для прогнозирования провозной платы на железнодорожном транспорте. В качестве эмпирической базы использованы данные ОАО «РЖД» за трехлетний период, включающие порядка 50 миллионов записей о грузовых перевозках. Такой массив обеспечивает репрезентативность выборки и позволяет учитывать отраслевое разнообразие данных. На основе алгоритмов Random Forest, XGBoost, LightGBM и CatBoost разработана ансамблевая модель, метауровень которой реализован в виде многомерной линейной регрессии. Для вычисления весовых коэффициентов применены метод наименьших квадратов и регуляризация Тихонова. Данный подход позволил стабилизировать решение и снизить влияние коррелированных выходов базовых моделей. Результаты вычислительных экспериментов показали, что объединение разнородных моделей в ансамбль обеспечивает повышение точности прогнозирования по сравнению с отдельными алгоритмами. Средняя абсолютная ошибка снизилась на 7–13 %, а средняя абсолютная процентная ошибка на 6–12 %, при этом коэффициент детерминации увеличился до 0,942. Дополнительно проведена оценка устойчивости ансамбля методом скользящего окна, что позволило определить горизонты прогнозирования, при которых сохраняется стабильность результатов. Расширенный анализ поведения ансамбля при варьировании входных признаков показал, что модель демонстрирует устойчивость к умеренным искажениям данных и сохраняет высокую воспроизводимость расчетов. Полученные результаты подтверждают практическую значимость предложенного подхода для задач транспортной аналитики, планирования перевозочного процесса и формирования экономически обоснованной тарифной политики.

**Ключевые слова:** машинное обучение, ансамблевые модели, регуляризация Тихонова, метод наименьших квадратов, точность модели, устойчивость модели, железнодорожный транспорт.

**Для цитирования:** Бухарова К.А. Модель и метод оценки весовых коэффициентов ансамбля машинного обучения в задаче прогнозирования провозной платы железнодорожных грузовых перевозок. *Моделирование, оптимизация и информационные технологии*. 2026;14(1). URL: <https://moitvivt.ru/ru/journal/pdf?id=2123> DOI: 10.26102/2310-6018/2026.52.1.005

## Model and method for evaluating the weighting coefficients of an ensemble machine learning model in the task of forecasting railway freight rates

К.А. Bukharova 

*Emperor Alexander I St. Petersburg State Transport University, St. Petersburg,  
the Russian Federation*

**Abstract.** This article examines the effectiveness of a developed ensemble machine learning model for forecasting rail freight rates. Russian Railways data for a three-year period, comprising approximately 50 million freight shipment records, serves as the empirical base. This dataset ensures a representative

sample and accounts for industry-specific data diversity. An ensemble model is developed using the Random Forest, XGBoost, LightGBM, and CatBoost algorithms, with a meta-level implemented as a multivariate linear regression. The ordinary least squares method and Tikhonov regularization are used to calculate the weighting coefficients. This approach stabilizes the solution and reduces the impact of correlated outputs from the base models. Results of computational experiments have shown that combining heterogeneous models into an ensemble improves forecasting accuracy compared to individual algorithms. The average absolute error decreased by 7–13 %, and the average absolute percentage error by 6–12 %, while the determination coefficient increased to 0.942. Additionally, the ensemble's stability was assessed using a sliding window method, which allowed us to determine forecasting horizons that maintain stable results. An extended analysis of the ensemble's behavior with varying input features showed that the model is robust to moderate data distortions and maintains high calculation reproducibility. The obtained results confirm the practical significance of the proposed approach for transport analytics, transportation planning, and the development of economically sound pricing policies.

**Keywords:** machine learning, ensemble models, Tikhonov regularization, least squares method, model accuracy, model stability, railway transport.

**For citation:** Bukharova K.A. Model and method for evaluating the weighting coefficients of an ensemble machine learning model in the task of forecasting railway freight rates. *Modeling, Optimization and Information Technology*. 2026;14(1). (In Russ.). URL: <https://moitvivt.ru/ru/journal/pdf?id=2123> DOI: 10.26102/2310-6018/2026.52.1.005

## Введение

Современная железнодорожная отрасль переживает этап интенсивной цифровой трансформации, в рамках которой точность прогнозных расчетов становится ключевым условием для принятия обоснованных управленческих решений. Одним из значимых показателей, определяющих эффективность перевозочного процесса и обоснованность тарифной политики, является провозная плата. Недостаточная точность ее прогнозирования приводит кискажению плановых расчетов, дополнительным затратам и снижению конкурентоспособности компаний<sup>1</sup>.

Традиционные эконометрические модели, которые применялись для расчета тарифов, не всегда обеспечивают необходимую точность. Они не способны в полной мере учитывать нелинейные зависимости и сложное взаимодействие факторов, характерное для транспортных данных. В научных исследованиях все чаще подчеркивается, что методы машинного обучения позволяют существенно повысить качество прогноза.

В ряде отечественных исследований отмечается, что для повышения точности прогнозирования в транспортных системах целесообразно использовать адаптивные и ансамблевые подходы [1]. Развитие транспортных моделей связано с необходимостью учитывать стохастическую природу входных факторов, что требует применения методов, способных адаптироваться к изменению внешних условий [2]. Отмечается, что объединение нескольких алгоритмов позволяет компенсировать индивидуальные недостатки каждого из них и обеспечивать более устойчивые оценки при ограниченности и вариативности данных [3].

Схожие тенденции прослеживаются и в зарубежных работах, где значительное внимание уделяется вопросам повышения устойчивости прогнозных систем, внедрению метаобучения и применению методов регуляризации [4]. Кроме того, подчеркивается, что объединение нескольких алгоритмов позволяет увеличить точность прогнозов и снизить их чувствительность к случайным колебаниям данных [5].

<sup>1</sup> Матюшин Л.Н. Тарифная политика на железнодорожном транспорте России. Москва: Учебно-методический центр по образованию на железнодорожном транспорте; 2021. 208 с.

В целом современные исследования подтверждают, что применение методов машинного обучения в задачах прогнозирования на транспорте позволяет заметно повысить точность по сравнению с традиционными подходами. Ансамблевые алгоритмы зарекомендовали себя как эффективный инструмент для анализа сложных зависимостей за счет комбинирования моделей с разными свойствами. Несмотря на накопленный опыт, вопросы количественной оценки устойчивости ансамблей и определения горизонта их корректного применения в транспортной аналитике остаются изученными не в полной мере.

В связи с этим, настоящее исследование направлено на экспериментальную проверку эффективности ансамблевой модели, построенной на основе четырех методов машинного обучения, а также анализ ее поведения при изменении временного диапазона и структуры данных.

## Материалы и методы

Задача прогнозирования провозной платы относится к классу регрессионных задач, в которых требуется построить модель, способную по заданным характеристикам конкретной перевозки определить ожидаемую стоимость тарифа. Подобные модели широко используются в транспортной аналитике как при формировании планов перевозочного процесса, так и при разработке тарифных решений и оценке экономической эффективности различных сценариев.

Эмпирической основой исследования стал массив данных ОАО «РЖД» за тридцать шесть месяцев, включающий порядка пятидесяти миллионов записей. Объем данных такого размера обеспечивает высокую репрезентативность и позволяет учитывать разнообразие условий работы.

Для проведения вычислительного эксперимента был выбран один из наиболее загруженных маршрутов железнодорожной сети. Это позволило оценить работу модели в условиях выраженной вариативности исходных параметров, что характерно для больших грузопотоков.

В качестве целевой переменной рассмотрена фактическая провозная плата, формирующаяся под влиянием множества факторов. Ее величина зависит не только от расстояния и массы груза, но и от тарифной категории, типа груза, сезонных коэффициентов и инфраструктурных особенностей.

В модели учитывались расстояние перевозки, масса груза, количество вагонов, категория отправки, а также временные признаки, отражающие календарную нестационарность тарифной среды. В качестве временных факторов использовались год перевозки и сезон, кодируемый набором фиктивных переменных. Введение этих признаков позволяет учитывать изменения тарифных условий и сезонные колебания грузопотоков, приводящие к систематическим сдвигам уровня провозной платы даже при близких технологических параметрах перевозки. Роль временных факторов дополнительно оценивалась при скользящем тестировании по годам, что дает возможность анализировать переносимость модели на следующий календарный период.

Перед тем как построить модель, была проведена подготовка исходных данных. На данном этапе выборка очищалась от пропусков, числовые показатели приводились к единому масштабу, а категориальные признаки преобразовывались таким образом, чтобы сохранить их содержательное значение.

Для оценки силы связи между исходными факторами и целевой переменной был проведен корреляционный анализ. Наиболее выраженная линейная зависимость наблюдалась у признаков масса груза, количество вагонов и расстояние перевозки, что

соответствует экономической логике формирования тарифа. В то же время между самими признаками не выявлено избыточной взаимозависимости.

Для объективной оценки качества моделей применялся метод перекрестной проверки (out-of-fold-валидация), который позволил приблизить условия эксперимента к реальным условиям, где модель регулярно сталкивается с данными, ранее не встречавшимися в обучении. На каждом шаге базовые алгоритмы обучались на части выборки и тестировались на другой, заранее исключенной из процесса обучения. Прогнозы, полученные на таких независимых подвыборках, формировали новый набор признаков для метамодели. Такой подход исключает утечку информации между уровнями ансамбля и предотвращает искусственное повышение точности [6].

Подобная организация обучения оказалась важной и для корректного взаимодействия между алгоритмами. Поскольку каждый из них формировал прогнозы на собственных подмножествах данных, метамодель получала информацию, отражающую реальное различие подходов, а не особенности одной и той же обучающей выборки. Это позволило ей точнее выделять устойчивые закономерности и компенсировать ошибки отдельных моделей [7]. На практике это обеспечило более ровное и стабильное поведение ансамбля при варьировании исходных данных.

В качестве базовых моделей применены четыре популярных метода машинного обучения. Выбор комбинации Random Forest, XGBoost, LightGBM и CatBoost обусловлен спецификой задачи прогнозирования провозной платы, характеризующейся неоднородностью наблюдений, наличием числовых и категориальных факторов и выраженными нелинейными зависимостями. Random Forest используется как робастный алгоритм, устойчивый к локальным выбросам и редким сочетаниям параметров перевозки. Градиентные бустинги (XGBoost и LightGBM) обеспечивают высокую точность на больших выборках и эффективно аппроксимируют сложные нелинейные зависимости. CatBoost включен как алгоритм, устойчивый к особенностям категориальных признаков и их взаимодействиям, что критично для тарифных задач. Совместное использование указанных моделей формирует разнородный набор предикторов, за счет чего достигается повышение точности ансамбля по сравнению с каждым базовым алгоритмом [8].

Гиперпараметры базовых моделей фиксировались по результатам предварительных вычислительных экспериментов на обучающей выборке с использованием out-of-fold-валидации и подбирались в логике контроля сложности (предотвращение переобучения) при сохранении различий между алгоритмами. Для Random Forest использовались 700 деревьев, схема отбора признаков sqrt, минимальный размер листа 2 и минимальный размер разбиения 5. Для XGBoost и LightGBM применялись 900 итераций, глубина деревьев 7 и скорость обучения 0,05 при подвыборке объектов и признаков 0,9. Для CatBoost использовались 900 итераций, глубина 7 и скорость обучения 0,05. Такая настройка обеспечивает сопоставимый уровень сложности базовых моделей и формирует устойчивый набор предикторов для метамодели.

На втором уровне применена метамодель в виде многомерной линейной регрессии. Она принимает на вход результаты базовых алгоритмов и формирует из них общий прогноз. Ее задача сводится к тому, чтобы не ограничиваться простым усреднением полученных значений, а подбирать такое сочетание вкладов разных моделей, при котором усиливаются их преимущества и сглаживаются возможные недостатки. Формально итоговое значение провозной платы  $\hat{y}$  задается выражением:

$$\hat{y}(x) = w_1 \cdot \hat{y}_1(x) + w_2 \cdot \hat{y}_2(x) + w_3 \cdot \hat{y}_3(x) + w_4 \cdot \hat{y}_4(x), \quad (1)$$

где  $\hat{y}_1(x)$  – прогноз Random Forest,  $\hat{y}_2(x)$  – прогноз XGBoost,  $\hat{y}_3(x)$  – прогноз LightGBM,  $\hat{y}_4(x)$  – прогноз CatBoost,  $\hat{y}(x)$  – итоговый прогноз провозной платы,  $w_i$  – весовые коэффициенты каждой модели,  $x$  – вектор признаков.

Для определения весовых коэффициентов применен метод наименьших квадратов. Однако из-за частичной коррелированности выходов отдельных моделей классический метод наименьших квадратов может приводить к неустойчивым решениям. Для того, чтобы избежать этого, применена регуляризация Тихонова, позволяющая стабилизировать веса и ограничить влияние избыточных коэффициентов [9]. Такой подход позволил снизить влияние коррелированных признаков, повысить устойчивость вычислений и добиться более стабильного распределения весов между моделями.

Параметр регуляризации  $\alpha$  подбирался на обучающей выборке с использованием K-fold перекрестной проверки при числе разбиений  $K=5$ . Поиск оптимального значения выполнялся в логарифмическом диапазоне  $\alpha \in [10^{-10}; 10^2]$  по критерию минимизации средней абсолютной ошибки на валидационных подвыборках. В результате вычислительного эксперимента для варианта с регуляризацией Тихонова оптимальное значение параметра составило  $\alpha = 10^{-10}$ . В целях сопоставления дополнительно рассматривался метод Lasso, для которого оптимальное значение параметра было получено на уровне  $\alpha = 10^{-6}$ . Сравнительный анализ показал, что применение метода Lasso приводит к более разреженному распределению весовых коэффициентов, однако не обеспечивает преимущества по показателям точности по сравнению с вариантом регуляризации Тихонова.

В качестве ключевых метрик оценки качества ансамблевой модели были приняты средняя абсолютная ошибка (MAE), средняя абсолютная процентная ошибка (MAPE) и коэффициент детерминации ( $R^2$ ) [10]. Результаты представлены в Таблице 1.

Таблица 1 – Метрики качества базовых алгоритмов и итоговой метамодели ансамбля  
Table 1 – Performance metrics of base algorithms and the final ensemble meta-model

Модель	MAE (руб.)	MAPE (%)	$R^2$
Random Forest	537 982	9,0	0,913
XGBoost	503 771	8,5	0,931
LightGBM	501 143	8,5	0,937
CatBoost	499 762	8,4	0,936
<b>Ансамблевая модель</b>	<b>465 874</b>	<b>7,9</b>	<b>0,942</b>

Ансамблевая модель показала снижение MAE на 7–13 % по сравнению с отдельными алгоритмами, а также повышение  $R^2$  до уровня 0,94.

Для проверки устойчивости выявленного улучшения была проведена парная статистическая оценка разности ошибок ансамблевой модели и лучшего базового алгоритма. Использовался бутстрэп-метод с повторной выборкой наблюдений и анализом распределения разности средних абсолютных ошибок. В результате среднее снижение MAE составило 33 891,5 руб., при этом 95-процентный доверительный интервал равен [32 817,9; 34 918,5] руб. Поскольку доверительный интервал не содержит нулевого значения, преимущество ансамбля по точности прогнозирования является статистически значимым и воспроизводимым.

## Результаты

Полученные результаты подтвердили, что ансамблевая модель обеспечивает более точный и устойчивый прогноз. Когда модель проверялась на разных частях

выборки, оказалось, что ее преимущество сохраняется не только на общей совокупности наблюдений, но и в отдельных диапазонах провозной платы, что особенно важно для практического применения.

Дополнительный анализ показателей MAE, MAPE и  $R^2$  для разных интервалов тарифов также подтвердил это наблюдение. Ошибки уменьшались и на средних значениях, и в зоне более высоких тарифов, где разброс данных обычно заметно больше. Такой результат демонстрирует, что модель корректно улавливает основные закономерности и не дает систематических смещений при прогнозировании.

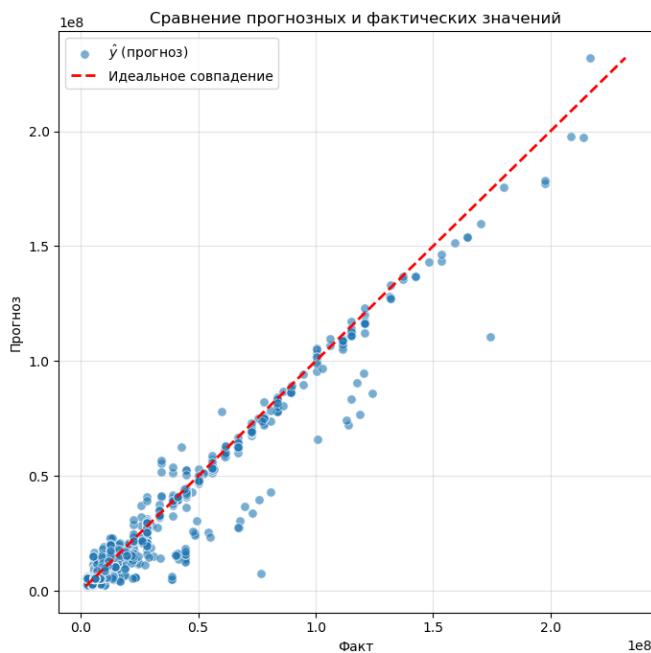


Рисунок 1 – Сравнение фактических и прогнозных значений провозной платы  
Figure 1 – Comparison of actual and predicted freight rate values

На Рисунке 1 в виде точечной диаграммы представлено сопоставление фактических и прогнозных значений провозной платы. Плотность точек вдоль диагонали указывает на то, что модель не только отличается достаточной точностью, но и формирует равномерные прогнозы во всем диапазоне тарифов.

Качество прогноза существенно не снижается даже при колебаниях исходных данных. Наблюдаемые небольшие отклонения отражают влияние рыночных факторов, сезонных колебаний загрузки и специфики тарифного регулирования, но они не образуют выраженных групповых смещений.

Таким образом, графическое представление полностью согласуется с расчетными данными и подтверждает правильность выбранного подхода.

## Обсуждение

Отдельного внимания заслуживает разработанный метод оценки весовых коэффициентов. Применение регуляризации методом Тихонова заметно стабилизировало распределение весов между моделями, за счет чего исчезли выраженные перекосы в сторону коррелированных алгоритмов, а итоговое решение стало более ровным и предсказуемым. Такой эффект особенно полезен при работе с большими массивами транспортных данных, где часть признаков связана друг с другом. В результате ансамбль демонстрирует не только более высокую точность, но и устойчивость к небольшим изменениям входных данных.

Интерпретируемость ансамбля обеспечивалась на двух уровнях. На уровне архитектуры вклад базовых моделей оценивался по весовым коэффициентам линейной комбинации, что позволяет анализировать структуру ансамбля и ее изменения во времени. На уровне признаков применялись методы SHAP и LIME для базовых моделей CatBoost и LightGBM. Результаты анализа на основе SHAP-подхода показали, что наибольшее влияние на прогноз провозной платы оказывают временной фактор (год перевозки), параметры объема перевозки (количество вагонов, масса груза), категория отправки и сезонные индикаторы. Структура значимости факторов согласуется между CatBoost и LightGBM, что подтверждает воспроизводимость интерпретации. Метод LIME использовался для локального объяснения отдельных прогнозов и подтвердил доминирование факторов, непосредственно связанных с формированием тарифа, на уровне отдельных наблюдений.

Устойчивость ансамбля во времени оценивалась методом скользящего окна с годовым шагом тестирования. На первом этапе модель обучалась на данных первого года и проверялась на данных второго года. На следующем этапе обучающая выборка расширялась за счет включения данных за первый и второй годы, после чего выполнялось тестирование на данных третьего года. Для каждого окна параметр регуляризации подбирался отдельно в рамках используемой процедуры обучения. При расширении обучающего интервала наблюдалось снижение ошибки прогноза и рост коэффициента детерминации, а значения MAE при тестировании на третьем году оставались сопоставимыми с базовой оценкой точности ансамбля (порядка 466 тыс. руб.), что подтверждает устойчивость ансамблевого метода при переходе к последующим временным периодам.

Надежность ансамблевой модели оценивалась с помощью проведенной серии стресс-тестов, представленной на Рисунке 2.

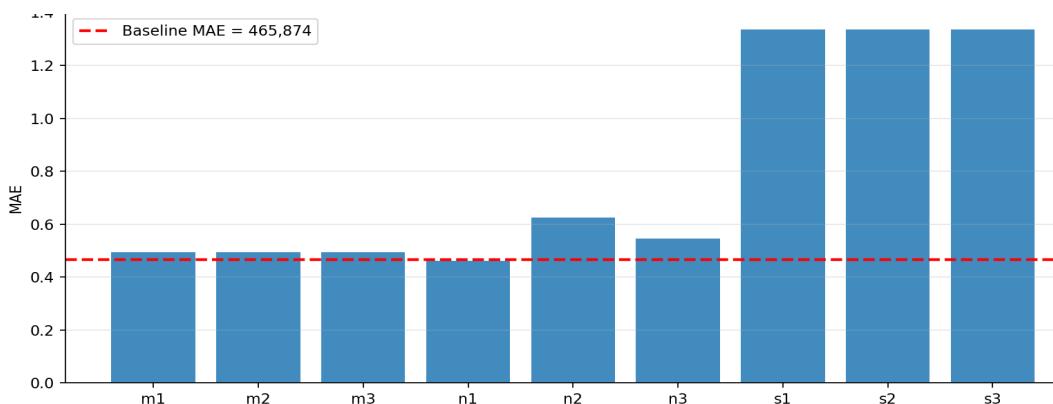


Рисунок 2 – Результаты стресс-теста  
Figure 2 – Stress test results

Частичное исключение признаков (10–20 %) или внесение умеренного шума (до  $\pm 5\%$ ) привело лишь к незначительному росту MAE и небольшому снижению  $R^2$ , который оставался на уровне 0,92–0,93. Такой результат позволяет сделать вывод о высокой устойчивости ансамбля к умеренным структурным нарушениям данных. Только значительные отклонения, такие как смещение распределений, заметно ухудшали качество прогноза. Наиболее выраженное ухудшение качества наблюдалось при моделировании структурного сдвига распределений признаков, что соответствует сценарию концептуального дрейфа. В этом случае средняя абсолютная ошибка возрастила до 1 319 716 руб., а коэффициент детерминации снижался до 0,603 по сравнению с базовым сценарием ( $MAE \approx 466$  тыс. руб.,  $R^2 \approx 0,942$ ). Данный результат

отражает пределы применимости модели, обученной на исторических данных, и указывает на необходимость периодического переобучения при изменении тарифной политики, структуры грузопотоков либо появлении новых категорий перевозок.

Исходя из результатов проведенного вычислительного эксперимента, можно сделать вывод о том, что предложенный подход работает заметно надежнее, чем отдельные методы машинного обучения.

В то же время полученные данные показали, что при существенных изменениях структуры выборки поведение модели становится менее предсказуемым. Это указывает на необходимость расширять набор данных и дополнительно изучать, как именно алгоритм реагирует на такие изменения.

## Заключение

Проведенное исследование подтвердило, что использование ансамблевого подхода способствует повышению точности прогнозирования провозной платы. Этот вывод опирается на анализ реальных отраслевых данных, что придает результатам практическую значимость.

Предложенная схема объединения моделей отличается простой структурой, однако она обеспечивает содержательный эффект за счет рационального использования сильных сторон отдельных алгоритмов. На первом этапе формируются прогнозы четырех независимых моделей, после чего их результаты объединяются в общий прогноз с помощью линейной регрессии, выполняющей роль метамодели.

Отдельно стоит отметить разработанный метод определения весовых коэффициентов. Применение регуляризации по Тихонову позволило избежать неустойчивых решений, уменьшить влияние коррелированных прогнозов и добиться более обоснованного распределения вкладов между моделями. Благодаря этому итоговое значение становится менее чувствительным к случайным колебаниям данных, а поведение модели более предсказуемым и интерпретируемым для аналитика.

Важным этапом стала проверка работы модели на последовательных временных интервалах. Использование метода скользящего окна показало, что по мере накопления информации ансамбль сохраняет стабильность и демонстрирует улучшение показателей. При переходе к новым периодам наблюдается снижение ошибок, при этом коэффициент детерминации остается на высоком уровне. Такой результат указывает на независимость модели от конкретного временного отрезка и ее пригодность для оперативного и тактического планирования.

Дополнительные выводы об устойчивости ансамбля были получены в ходе стресс-тестирования. Исключение части признаков и внесение умеренного уровня шума привели лишь к небольшому ухудшению значений метрик, что подтверждает корректность выбранного набора факторов и устойчивость модели к вариациям исходных данных.

С методологической точки зрения показано, что даже относительно простое линейное объединение прогнозов может быть эффективным при корректной организации этапов обучения и оценки весовых коэффициентов.

Использование большой выборки отраслевых данных позволило проверить модель в условиях, максимально близких к реальным. Это повышает надежность полученных результатов и создает основу для дальнейших исследований в направлении оценки устойчивости ансамблей и их поведения при изменении входных факторов.

С практической точки зрения предложенный подход может рассматриваться как инструмент поддержки тарифных решений и планирования перевозочного процесса. Модель позволяет заранее оценивать возможный уровень провозной платы при разных

комбинациях параметров перевозки и тем самым уменьшать неопределенность при принятии экономически целесообразных управлеченческих решений. В перспективе такая система может быть интегрирована в корпоративные аналитические платформы железнодорожных компаний и использоваться на регулярной основе.

Дальнейшее развитие работы может идти по нескольким направлениям. Во-первых, за счет расширения состава признаков, связанных с внешними экономическими условиями и состоянием инфраструктуры. Во-вторых, за счет более детального анализа объяснения прогнозов и разработки средств интерпретации вклада отдельных факторов и моделей, удобных для практиков. В-третьих, за счет адаптации предложенного ансамбля для других задач транспортной аналитики, например, для оценки загрузки инфраструктуры.

В заключение отметим, что результаты исследования показывают, что ансамбль прогнозных моделей может стать эффективным инструментом для аналитиков. Сочетание точности, устойчивости к изменениям данных и понятной структуры делает такую модель востребованной в условиях цифровизации железнодорожной отрасли и возрастающих требований к качеству аналитических расчетов.

## СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Загидуллин Р.Р., Хайбуллин А.Н. Прогнозирование транспортной загруженности с использованием методов машинного обучения. *Транспорт и информационные технологии*. 2025;15(2):202–216. <https://doi.org/10.12731/2227-930X-2025-15-2-347>  
Zagidullin R., Khaybullin A. Traffic Congestion Forecasting Using Machine Learning Methods. *Transportation and Information Technologies in Russia*. 2025;15(2):202–216. (In Russ.). <https://doi.org/10.12731/2227-930X-2025-15-2-347>
2. Побирченко В.Р., Побирченко В.В. Теоретические аспекты моделирования социально-экономического развития региона. *Ученые записки Крымского федерального университета имени В.И. Вернадского. Экономика и управление*. 2025;11(2):170–182.
3. Вегера Ж.Г. Применение генеративного искусственного интеллекта (ИИ) для анализа образовательных данных и прогнозирования академической успеваемости студентов. *Управление образованием: теория и практика*. 2024;14(8-1):116–125. <https://doi.org/10.25726/j2473-1350-7803-t>  
Vegera Z.G. The Use of Generative Artificial Intelligence (AI) to Analyze Educational Data and Predict Student Academic Performance. *Education Management Review*. 2024;14(8-1):116–125. (In Russ.). <https://doi.org/10.25726/j2473-1350-7803-t>
4. Zhou Zh.-H. *Ensemble Methods: Foundations and Algorithms*. New York: Chapman & Hall/CRC; 2012. 236 p. <https://doi.org/10.1201/b12207>
5. Hyndman R.J., Athanasopoulos G. *Forecasting: Principles and Practice*. Melbourne: OTexts; 2021. 442 p.
6. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer; 2009. 745 p. <https://doi.org/10.1007/978-0-387-84858-7>
7. Zhang Yu., Ma J., Liang Sh., Li X., Liu J. A Stacking Ensemble Algorithm for Improving the Biases of Forest Aboveground Biomass Estimations from Multiple Remotely Sensed Datasets. *GIScience & Remote Sensing*. 2022;59(1):234–249. <https://doi.org/10.1080/15481603.2021.2023842>
8. Рукомин М.А. Обзор ансамблевых моделей предиктивной аналитики и их сравнение с традиционными ML-подходами. *Вестник науки*. 2025;1(8):368–373.

- Rukomin M.A. Review of Ensemble Models in Predictive Analytics and Their Comparison with Traditional Machine Learning Approaches. *Science Bulletin.* 2025;1(8):368–373. (In Russ.).
9. Тихонов А.Н., Арсенин В.Я. *Методы решения некорректных задач.* Москва: Наука; 1979. 288 с.
10. Харитонова А.Е. Прогнозирование налоговой нагрузки сельскохозяйственных предприятий методами машинного обучения. *Налоги и налогообложение.* 2023;(4):28–38. <https://doi.org/10.7256/2454-065X.2023.4.43917>
- Kharitonova A.E. Forecasting the Tax Burden of Agricultural Enterprises by Machine Learning Methods. *Taxes and Taxation.* 2023;(4):28–38. (In Russ.). <https://doi.org/10.7256/2454-065X.2023.4.43917>

#### ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

**Бухарова Ксения Алексеевна**, аспирант, **Ksenya A. Bukharova**, Postgraduate, Emperor Peterburgskiy государственный университет Alexander I St. Petersburg State Transport путей сообщения Императора Александра I, University, St. Petersburg, the Russian Federation. Санкт-Петербург, Российская Федерация.

e-mail: [bukharova@pgups.ru](mailto:bukharova@pgups.ru)

*Статья поступила в редакцию 06.11.2025; одобрена после рецензирования 29.12.2025;  
принята к публикации 13.01.2026.*

*The article was submitted 06.11.2025; approved after reviewing 29.12.2025;  
accepted for publication 13.01.2026.*