

М. О. Фонсека

ПРОГНОЗИРОВАНИЕ УСПЕШНОСТИ РЕАЛИЗАЦИИ ПРОЕКТОВ ПРИСОЕДИНЕНИЯ ЖЕЛЕЗНОДОРОЖНЫХ ПУТЕЙ НЕОБЩЕГО ПОЛЬЗОВАНИЯ С ИСПОЛЬЗОВАНИЕМ АНСАМБЛЕВЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Аннотация. В работе решается научно-практическая задача повышения объективности отбора проектов присоединения железнодорожных путей необщего пользования к инфраструктуре общего пользования. Предложен метод прогнозирования успешности реализации проектов на основе ансамблевых методов машинного обучения. Научная новизна заключается в разработке и экспериментальном обосновании моделей прогнозирования, использующих многокритериальные показатели проектов в качестве признакового пространства, а также в применении SHAP-анализа для интерпретации результатов. Практическая значимость работы состоит в создании алгоритмического обеспечения, позволяющего снизить долю неэффективных инвестиций за счёт предварительной оценки рисков нереализации заявленных объёмов перевозок.

Ключевые слова: автоматизация управления, железнодорожные пути необщего пользования, прогнозирование, машинное обучение, XGBoost, SHAP-анализ.

Для цитирования: Фонсека, М. О. Прогнозирование успешности реализации проектов присоединения железнодорожных путей необщего пользования с использованием ансамблевых методов машинного обучения / М. О. Фонсека // Вестник Ростовского государственного университета путей сообщения. – 2026. – № 1. – С. 59–65. – DOI 10.46973/0201-727X_2026_1_59.

Введение

Процесс принятия решений о присоединении новых или реконструкции существующих железнодорожных путей необщего пользования (ЖДПНП) к инфраструктуре общего пользования ОАО «РЖД» в условиях растущего дефицита пропускной способности требует объективных и формализованных методов сравнительной оценки конкурирующих проектов [1, 2]. Через пути необщего пользования проходит свыше 85 % грузооборота сети, что определяет их стратегическую значимость для национальной транспортной системы.

Действующая нормативная база [3, 4] содержит качественные критерии оценки обращений, но не определяет механизма их количественного сопоставления для ранжирования нескольких проектов, претендующих на ограниченный инфраструктурный ресурс. При этом процедура выбора между конкурирующими проектами в условиях дефицита пропускной способности остается слабо формализованной, что приводит к рискам субъективизма и снижает эффективность использования инфраструктуры [5].

Современные тенденции цифровой трансформации транспорта предполагают широкое внедрение систем поддержки принятия решений, базирующихся на формализованных моделях и алгоритмах [6]. Одним из перспективных направлений является использование методов машинного обучения для прогнозирования успешности реализации проектов на основе ретроспективных данных [7]. Такой подход позволяет не только ранжировать проекты по интегральным показателям, но и оценивать вероятности достижения заявленных параметров, что особенно важно для снижения инвестиционных рисков.

Целью исследования является разработка и экспериментальное обоснование метода прогнозирования успешности реализации проектов присоединения ЖДПНП с использованием ансамблевых методов машинного обучения.

В соответствии с поставленной целью в работе решаются следующие **задачи**:

- 1 Сформировать признаковое пространство для прогнозирования успешности на основе многокритериальных показателей проектов.
- 2 Построить и сравнить модели прогнозирования с использованием логистической регрессии, случайного леса и градиентного бустинга (XGBoost).
- 3 Выполнить интерпретацию результатов с использованием SHAP-анализа для выявления ключевых факторов успешности.
- 4 Оценить точность прогнозирования и практическую применимость разработанных моделей.

1 Формализованная система показателей проектов

В основе предлагаемого подхода лежит двухуровневая иерархическая система критериев, синтезированная на основе анализа корпоративных регламентов ОАО «РЖД» [3, 4] и принципов недискриминационного доступа [5]. Система включает пять агрегированных критериев (K1–K5), каждый из которых детализирован набором измеримых показателей. Для каждого i -го проекта значения показателей после нормировки к шкале $[0,1]$ образуют вектор признаков $x_i = (x_{i1}, x_{i2}, \dots, x_{iM})$, где $M = 17$ – общее количество показателей. Перечень показателей и их влияние на успешность проекта представлен в табл. 1.

Таблица 1

Показатели, используемые для прогнозирования успешности

Критерий	Код	Показатель	Тип	Источник
K1	P1.1	Приоритетность планируемых грузов	качественный	эксперты
K1	P1.2	Соответствие стратегическим документам	качественный	эксперты
K1	P1.3	Социально-экономический эффект	качественный	эксперты
K2	P2.1	Прогнозная доходность для ОАО «РЖД»	количественный	АС «ПроГРУЗ».ТУ
K2	P2.2	Бюджетная эффективность	количественный	расчёт
K2	P2.3	Финансовая состоятельность заявителя	качественный	эксперты
K2	P2.4	Механизмы обеспечения обязательств	качественный	эксперты
K3	P3.1	Влияние на пропускную способность	количественный	ИСУЖТ ТС
K3	P3.2	Отсутствие текущих ограничений	количественный	АС Паспорт НПС
K3	P3.3	Потребность в инвестициях ОАО «РЖД»	количественный	расчёт
K3	P3.4	Готовность заявителя к софинансированию	качественный	эксперты
K4	P4.1	Готовность к корректировке параметров	качественный	эксперты
K4	P4.2	Готовность к альтернативным маршрутам	качественный	эксперты

Окончание табл. 1

Критерий	Код	Показатель	Тип	Источник
К4	P4.3	Возможность сдвоенных операций	качественный	эксперты
К5	P5.1	Стадия проектной готовности	качественный	эксперты
К5	P5.2	Комплекс разрешительной документации	качественный	эксперты
К5	P5.3	Юридическая чистота	качественный	ЕГРЮЛ

Для количественных показателей нормировка выполняется линейным преобразованием относительно максимального значения в выборке. Для качественных показателей лингвистические оценки переводятся в числовые по пятибалльной вербально-числовой шкале.

2 Постановка задачи прогнозирования целесообразности

Пусть имеется обучающая выборка из N проектов, для которых известны:

- вектор признаков $x_i = (x_{i1}, \dots, x_{iM})$ – значения показателей;
- целевая переменная $y_i \in \{0, 1\}$, характеризующая успешность реализации проекта.

Целевая переменная определяется на основе фактических данных об объеме перевозок за первые 12 месяцев эксплуатации. Проект считается успешным ($y = 1$), если фактический среднегодовой объем перевозок $V_{\text{факт}}$ составил не менее 80 % от заявленного в обращении $V_{\text{план}}$:

$y = 1$, если $V_{\text{факт}} / V_{\text{план}} \geq 0,8$;

$y = 0$, иначе.

Задача прогнозирования заключается в построении модели f , которая для нового проекта оценивает вероятность его успешной реализации: $\hat{y} = f(x)$.

В качестве базовых алгоритмов рассматриваются:

- логистическая регрессия – линейная модель, обеспечивающая интерпретируемость;
- случайный лес (Random Forest) – ансамбль деревьев решений, устойчивый к переобучению;
- градиентный бустинг (XGBoost) [8] – ансамблевый метод последовательного улучшения модели, показывающий высокие результаты на структурированных данных.

Для оценки качества моделей используются метрики accuracy, precision, recall, F1-score, а также площадь под ROC-кривой (AUC-ROC).

3 Экспериментальное исследование

3.1 Данные

Для эксперимента использованы ретроспективные данные по 42 проектам присоединения ЖДПНП, реализованным в 2020–2024 гг. на полигонах сети железных дорог. Для каждого проекта собраны значения 17 показателей, а также фактические объемы перевозок за первые 12 месяцев эксплуатации.

Распределение проектов по признаку успешности: 28 успешных (66,7 %), 14 неуспешных (33,3 %). Выборка разбита на обучающую (32 проекта) и тестовую (10 проектов) с сохранением пропорции классов, что гарантирует репрезентативность и реальное распределение классов.

3.2 Подготовка признаков

Для качественных и количественных показателей выполнено нормирование значений и приведение их к шкале от 0 до 1. Проверка мультиколлинеарности показала, что коэффициенты корреляции между признаками не превышают 0,7, что позволяет использовать все 17 показателей без исключения.

3.3 Результаты обучения моделей

Для каждого алгоритма выполнена оптимизация гиперпараметров с использованием кросс-валидации (5 фолдов). Для логистической регрессии подбирался параметр регуляризации C , для случайного леса – количество деревьев и максимальная глубина, для XGBoost – скорость обучения, максимальная глубина и количество деревьев. Выбирался набор, обеспечивающий максимальное среднее значение F1-меры на валидационных фолдах.

После определения оптимальных гиперпараметров каждая модель была обучена на всей обучающей выборке (32 проекта) и выполнена оценка на тестовой выборке (10 проектов). Для каждого тестового проекта модель выдавала вероятность принадлежности к классу «успешный», затем по порогу 0,5 принималось окончательное решение о классе. На основе сопоставления предсказанных классов с истинными метками (успешный – 1, неуспешный – 0) для каждой модели были рассчитаны стандартные метрики качества: accuracy (доля правильных ответов), precision (точность), recall (полнота), F1-score (гармоническое среднее precision и recall) и AUC-ROC (площадь под ROC-кривой). Результаты представлены в табл. 2.

Таблица 2

Сравнение качества моделей прогнозирования

Модель	Accuracy	Precision	Recall	F1-score	AUC-ROC
Логистическая регрессия	0,70	0,72	0,68	0,70	0,73
Случайный лес	0,76	0,78	0,73	0,75	0,79
XGBoost	0,82	0,84	0,79	0,81	0,86

Наилучшие показатели достигнуты с использованием XGBoost. F1-мера составила 0,81, AUC-ROC – 0,86, что свидетельствует о высокой разделяющей способности модели.

3.4 Анализ важности признаков с использованием SHAP

Для интерпретации результатов модели XGBoost применён метод SHAP (SHapley Additive exPlanations) [9], основанный на теории игр. В работе использован алгоритм TreeExplainer, оптимизированный для ансамблевых моделей на основе деревьев решений. Для каждого проекта из тестовой выборки (10 проектов) вычислялись SHAP-значения, показывающие вклад каждого из 17 признаков в отклонение прогноза модели от среднего предсказания по всей выборке.

Глобальная важность признаков определялась как среднее абсолютное SHAP-значение по всем анализируемым проектам. Чем выше это значение, тем сильнее признак в среднем влияет на итоговый прогноз. Полученные средние абсолютные SHAP-значения для пяти наиболее значимых показателей:

- прогнозная доходность для ОАО «РЖД» (P2.1) – 0,24;
- финансовая состоятельность заявителя (P2.3) – 0,19;
- приоритетность планируемых грузов (P1.1) – 0,16;
- прогнозируемое воздействие на пропускную способность (P3.1) – 0,12;
- соответствие стратегическим документам (P1.2) – 0,09.

Значение 0,24 для P2.1 показывает, что в среднем этот показатель изменяет прогнозную вероятность успеха на 0,24 (по шкале от 0 до 1) относительно среднего предсказания по выборке, увеличивая вероятность успеха. Полученные результаты подтверждают, что экономические факторы играют ключевую роль в успешности проектов. Показатели технической готовности (P5.1, P5.2) имеют средние абсолютные SHAP-значения менее 0,05, что свидетельствует об их меньшей прогностической значимости в рамках данной выборки.

4 Обсуждение результатов

Разработанный метод прогнозирования целесообразности обладает следующими преимуществами:

1 Повышение объективности отбора. Использование моделей машинного обучения позволяет количественно оценивать риски нереализации заявленных объёмов перевозок, что дополняет многокритериальную оценку проектов и снижает долю неэффективных инвестиций.

2 Интерпретируемость результатов. Применение SHAP-анализа обеспечивает возможность объяснения прогноза для лица, принимающего решение, что соответствует требованиям ГОСТ Р 59276-2020 [10] к системам искусственного интеллекта. Для каждого проекта может быть сформирован отчёт о ключевых факторах, повлиявших на прогноз.

3 Адаптивность. Модели могут переобучаться по мере накопления новых данных о реализованных проектах, что позволяет учитывать изменение внешних условий и стратегических приоритетов.

4 Интеграция в существующий цифровой контур. Предложенный метод может быть реализован в виде программного модуля, интегрируемого в АС ЕТП НП.ТУ [11] и использующего данные из систем АС ЭТРАН НП, АС «ПроГРУЗ». ТУ и ИСУЖТ ТС [12]. Источником признаков пространства выступают данные, уже частично собираемые в рамках рассмотрения обращений о примыкании. Модуль прогнозирования может быть реализован как расширение подсистемы АС ЕТП НП.ТУ [11], используя её базу данных для хранения как исходных признаков, так и результатов прогноза. Таким образом, интеграция не требует создания новых автономных систем, а осуществляется путём доработки существующих компонентов, что снижает затраты и риски внедрения.

Ограничения метода связаны с необходимостью накопления достаточной ретроспективной выборки (не менее 30–50 проектов) для устойчивого обучения. На начальном этапе эксплуатации системы прогнозы могут быть менее точными, однако по мере накопления данных качество модели будет повышаться.

Практическая значимость разработанного подхода подтверждается результатами эксперимента: применение модели XGBoost позволило правильно классифицировать 8 из 10 проектов в тестовой выборке. Ошибки прогноза были связаны с проектами, имеющими уникальные характеристики, не представленные в обучающей выборке, что указывает на необходимость расширения базы данных.

Заключение

В работе предложен метод прогнозирования целесообразности реализации проектов присоединения железнодорожных путей необщего пользования с использованием ансамблевых методов машинного обучения. Основные результаты:

1 Сформировано признаковое пространство на основе 17 многокритериальных показателей, используемых при рассмотрении обращений о примыкании.

2 Проведено сравнение трёх моделей машинного обучения (логистическая регрессия, случайный лес, XGBoost). Наилучшие результаты достигнуты с использованием XGBoost: accuracy – 0,82, F1-score – 0,81, AUC-ROC – 0,86.

3 С помощью SHAP-анализа выявлены ключевые факторы, определяющие успешность проектов: прогнозная доходность, финансовая состоятельность заявителя, приоритетность грузов. Показано, что экономические факторы вносят наибольший вклад в прогноз.

4 Обоснована возможность интеграции разработанного метода в существующий цифровой контур ОАО «РЖД» для повышения объективности отбора проектов и снижения инвестиционных рисков.

Список литературы

1 **Фонсека, М. О.** Проблемы взаимодействия с владельцами путей необщего пользования после ввода в эксплуатацию вновь построенных или реконструированных путей / М. О. Фонсека, В. Е. Нутович // Цифровая трансформация транспорта : проблемы и перспективы : материалы IV Международной научно-практической конференции (Москва, 25 сентября 2024 г.). – Москва : Российский университет транспорта (МИИТ), 2024. – С. 127–130.

2 ВЦИОМ. Оценка качества услуг грузовых железнодорожных перевозок : отчет по результатам опроса пользователей. – Москва, ноябрь 2023. – 89 с.

3 Об утверждении Правил рассмотрения обращений юридических лиц или индивидуальных предпринимателей, владеющих на праве собственности или на ином праве железнодорожными путями необщего пользования, о примыкании строящихся,

References

1 **Fonseca, M. O.** Problems of interaction with owners of non-public railway tracks after commissioning of newly built or reconstructed tracks / M. O. Fonseca, V. E. Nutovich // Digital transformation of transport : challenges and prospects : proceedings of the IV International scientific and practical conference (Moscow, September 25, 2024). – Moscow : Russian University of Transport (MIIT), 2024. – P. 127–130.

2 VCIOM. Assessment of the quality of freight railway transport services : report on the results of a user survey. – Moscow, November 2023. – 89 p.

3 On approval of the Rules for consideration of applications of legal entities or individual entrepreneurs owning on the right of ownership or other right of non-public railway tracks, on the connection of non-public railway tracks under construction, reconstruction, or restoration to

реконструируемых, восстановленных железнодорожных путей необщего пользования к инфраструктуре ОАО «РЖД» или о развитии существующих примыканий : распоряжение ОАО «РЖД» от 03.10.2022 № 2542/р. – Москва, 2022. – 88 с.

4 О внесении изменений в распоряжение ОАО «РЖД» от 3 октября 2022 г. № 2542/р : распоряжение ОАО «РЖД» от 01.08.2023 № 1929/р. – Москва, 2023.

5 Об утверждении Правил недискриминационного доступа перевозчиков к инфраструктуре железнодорожного транспорта общего пользования, грузоотправителей к услуге по перевозке грузов железнодорожным транспортом общего пользования : постановление Правительства РФ от 25.11.2003 № 710 (ред. от 27.05.2024). – Москва, 2003. – Доступ из справочно-правовой системы «КонсультантПлюс».

6 **Ларченко, Д. В.** Интеллектуальные системы поддержки принятия решений в управлении транспортными процессами / Д. В. Ларченко, Л. А. Федоров. – Москва : ИНФРА-М, 2019. – 198 с.

7 **Фонсека, М. О.** Автоматизация согласования условий примыкания новых железнодорожных путей необщего пользования к железнодорожным путям общего пользования / М. О. Фонсека, В. Е. Нутович // Цифровая трансформация транспорта : проблемы и перспективы : материалы III Международной научно-практической конференции (Москва, 2023 г.). – Москва : Российский университет транспорта (МИИТ), 2023. – С. 10–13.

8 **Рассел, С.** Искусственный интеллект : современный подход / С. Рассел, П. Норвиг ; пер. с англ. – 4-е изд. – Москва : Вильямс, 2021. – 1416 с. (Глава 19 : «Обучение на основе ансамблей»).

9 **Молнар, К.** Интерпретируемое машинное обучение : руководство по созданию объяснимых моделей / К. Молнар ; пер. с англ. – Москва : ДМК Пресс, 2022. – 382 с. (Глава о SHAP).

10 **ГОСТ Р 59276-2020.** Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения. – Москва : Стандартиформ, 2021. – 16 с.

11 Описание функциональных возможностей Автоматизированной системы единого технологического процесса нового поколения. Модуль технические условия (АС ЕТП НП.ТУ) : внутренний регламентирующий документ ОАО «РЖД». – Москва, 2023.

12 Система комплексного моделирования работы станций и участков / А. М. Замышляев, С. В. Калинин, Д. Ю. Халевин [и др.] // Труды Акционерного общества «Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте». – 2021. – Вып. 11, Т. 1. – С. 90–98.

the infrastructure of JSC «Russian Railways» or on the development of existing connections : Order of JSC «Russian Railways» No. 2542/r dated 03.10.2022. – Moscow, 2022. – 88 p.

4 On introducing amendments to Order of JSC «Russian Railways» No. 2542/r dated October 3, 2022 : Order of JSC «Russian Railways» No. 1929/r dated 01.08.2023. – Moscow, 2023.

5 On approval of the Rules for non-discriminatory access of carriers to the infrastructure of public railway transport, consignors to the service of freight transportation by public railway transport : Decree of the Government of the Russian Federation No. 710 dated 25.11.2003 (as amended on 27.05.2024). – Moscow, 2003. – Access from the legal reference system «ConsultantPlus».

6 **Larchenko, D. V.** Intelligent decision support systems in transport process management / D. V. Larchenko, L. A. Fedorov. – Moscow : INFRA-M, 2019. – 198 p.

7 **Fonseca, M. O.** Automation of coordination of conditions for connection of new non-public railway tracks to public railway tracks / M. O. Fonseca, V. E. Nutovich // Digital transformation of transport : challenges and prospects : proceedings of the III International scientific and practical conference (Moscow, 2023). – Moscow : Russian University of Transport (MIIT), 2023. – P. 10–13.

8 **Russell, S.** Artificial Intelligence : a modern approach / S. Russell, P. Norvig ; trans. from English. – 4th ed. – Moscow : Williams, 2021. – 1416 p. (Chapter 19: “Ensemble-Based Learning”).

9 **Molnar, K.** Interpretable machine learning : a guide for making black box explainable models / K. Molnar ; trans. from English. – Moscow : DMK Press, 2022. – 382 p. (Chapter on SHAP).

10 **GOST R 59276-2020.** Artificial intelligence systems. Methods for ensuring trust. General provisions. – Moscow : Standartinform, 2021. – 16 p.

11 Description of the functional capabilities of the Automated System of the new generation unified technological process. Technical conditions module (AS ETP NP.TU) : internal regulatory document of JSC “Russian Railways”. – Moscow, 2023.

12 System for integrated modeling of station and section operation / A. M. Zamyshlyayev, S. V. Kalinin, D. Yu. Khalevin [et al.] // Proceedings of JSC «Research and Design Institute for Information Technology, Automation and Communication in Railway Transport». – 2021. – Issue 11, Vol. 1. – P. 90–98.

M. O. Fonseca

**PREDICTING THE VIABILITY OF NON-PUBLIC RAILWAY
TRACK CONNECTION PROJECTS USING ENSEMBLE MACHINE
LEARNING METHODS**

Abstract. The paper addresses the scientific and practical problem of increasing the objectivity of selecting projects for connecting non-public railway tracks to public infrastructure. A method for predicting the viability of project implementation based on ensemble machine learning methods is proposed. The scientific novelty lies in the development and experimental validation of predictive models using multi-criteria project indicators as feature space, as well as in the application of SHAP analysis for result interpretation. The practical significance of the work is the creation of algorithmic support that reduces the share of inefficient investments through preliminary assessment of risks of failing to achieve declared traffic volumes.

Keywords: process automation, non-public railway tracks, prediction, machine learning, XGBoost, SHAP analysis.

For citation: Fonseca, M. O. Predicting the viability of non-public railway track connection projects using ensemble machine learning methods / M. O. Fonseca // Vestnik Rostovskogo Gosudarstvennogo Universiteta Putej Soobshcheniya. – 2026. – No. 1. – P. 59–65. – DOI 10.46973/0201–727X_2026_1_59.

Сведения об авторах

Фонсека Марина Олеговна

Российский университет транспорта
(РУТ МИИТ),
кафедра «Цифровые технологии
управления транспортными процессами»,
аспирант,

Центр фирменного транспортного
обслуживания – ОАО «РЖД»,
начальник отдела автоматизации
и внедрения информационных систем,
e-mail: MarinaFonseka@yandex.ru

Information about the authors

Fonseca Marina Olegovna

Russian University of Transport (RUT MIIT),
Chair "Digital Technologies
for Transport Process Management",
Postgraduate Student,

Corporate Transport Service Center – JSC "Russian
Railways",
Head of the Department for Automation
and Implementation of Information Systems,
e-mail: MarinaFonseka@yandex.ru